A quest for disease genes that unleash inflammation p. 1430

Tree rings authenticate musical instruments p. 1434

A chemoenzymatic cascade turns carbon dioxide into starch p. 1523

Science

$15
24 SEPTEMBER 2021
science.org

AAAS



SPECIAL ISSUE

THE HUMAN GENOME
EXPANDING INSIGHTS

Explore what's possible

With premium flow cytometry technology for small particle detection

Developing a robust and high-resolution flow cytometry method for small particle analysis can be challenging. With BD, you can take advantage of premium BD FACSymphony™ instrument technology and our complete workflow for small particle analysis—including reagents, software and system quality control. The new BD FACSymphony™ A1 Cell Analyzer with the BD® Small Particle Detector option is designed to fit on a benchtop and enables independent detection of large and small particles on a single instrument.

With our expert-level support and flow cytometry domain expertise, you can focus on your research and be confident in your results.



Independent small particle side scatter detector for **resolving particles as small as 90 nm**

Discover the difference at **bdbiosciences.com/SmallParticleAnalyzer**

BD

Class 1 Laser Product.
For Research Use Only. Not for use in diagnostic or therapeutic procedures.

BD, the BD Logo and FACSymphony are trademarks of Becton, Dickinson and Company or its affiliates.
© 2021 BD. All rights reserved. BD-41270 (v1.0) 0821

Institute of Chemistry, Chinese Academy of Sciences (ICCAS):

65 years of accomplishments based on collaborative and interdisciplinary research

Since its founding in 1956, the Institute of Chemistry, Chinese Academy of Sciences (ICCAS) has played a major role in the rapid advances China has made in the field of chemistry. ICCAS scientists have initiated or led progress in China on a wide range of chemistry subdisciplines and relevant areas, including salt lake chemistry, chemical thermodynamics, polymer science, molecular reaction dynamics, organic solids, nanoscience and nanotechnology, and frontier fields of molecular science. For instance, ICCAS scientists were the first in China to design and fabricate scanning tunneling microscopy and atomic force microscopy, which accelerated the country's research on nanoscience and nanotechnology.

In recent years, ICCAS scientists have made remarkable achievements in areas such as two-dimensional (2D) carbon materials, controllable molecular self-assembly and nanomaterials for energy storage, organic optoelectronic and thermoelectric materials and devices, organic nanophotonics, green and environmental chemistry, high-performance polymeric materials, and analytical chemistry for living biosystems.

For example, researchers at ICCAS created a novel sp-hybridized 2D carbon allotrope called graphdiyne, which is the first all-carbon material obtained through synthetic chemistry at low temperature and atmospheric pressure. Graphdiyne's structure, its abundant carbonation bonds, its intrinsic band gap, and the nonuniformity of its surface charge distribution make it well suited for rapid ion, proton, and charge transport. Graphdiyne has been demonstrated to be a promising material for catalysis, energy storage, photoelectric conversion, and application in information intelligence.

ICCAS researchers pursue scientific breakthroughs, but they also want their findings to have practical benefits outside the laboratory. For instance, ICCAS began researching organic photoconductor (OPC) drums in the 1980s. These drums are core components of laser printers and digital copiers. In 2000, industrial partners leveraged ICCAS's research findings to create China's first automatic production line for OPC drums. Since then, Chinese OPC manufacturers have captured one-third of domestic market share and risen to third place in global OPC drum sales.

"ICCAS has first-rate scientific staff, excellent research platforms, high-level technical support, open cooperation at both the domestic and international levels, plus efficient service and managerial innovation," says Deqing Zhang, director of ICCAS.

Bringing answers into the light

Zhang studied at ICCAS as a graduate student in the late 1980s. Today, in addition to his responsibilities as director of ICCAS, he is also a research professor in the Key Laboratory of Organic Solids, where he and his team design and synthesize new organic functional molecules and explore their optoelectronic functions for high-performance organic optoelectronic functional materials.

"The infrastructure and research conditions at ICCAS have improved remarkably, with more young PIs [principal investigators] working here than ever before," says Zhang, reflecting on how things have changed at ICCAS since he studied there more than 30 years ago. On the other hand, Zhang says that the spirit of the institute has stayed the same, with ICCAS continuing to value and prioritize innovation, cooperation, devotion, and seeking truth.

PHOTOS PROVIDED BY ICCAS





ICCAS campus (**top**) and Director Deqing Zhang (**bottom**)

The dynamic advantage

Compared to other Chinese research institutions, ICCAS stands out for its interdisciplinary nature, Zhang says, with faculty members coming from a wide range of different research backgrounds.

Additionally, ICCAS has a dynamic and flexible organizational structure. "This collaboration is essential to make progress on innovative research projects or achieve industrial transformation," Zhang explains.

ICCAS is a center of both research and education for the next generation of chemists. Each year, more than 200 graduate students receive doctoral and Master's degrees.

Collaborating at the frontier

"ICCAS is in a period of rapid development," says Zhang. "I believe that ICCAS researchers will continue to achieve original innovations at the forefronts of molecular science."

For those interested in working in an interdisciplinary and rigorous, yet open and inclusive environment, ICCAS is currently seeking young scientists who are excellent researchers to join research teams working at the frontiers of molecular science. Team leaders should be very active scientists in their fields with good leadership and organizational skills, plus the capacity to grasp strategic development trends in their fields. For candidates who have these qualifications, ICCAS stands ready to provide the managerial, financial, and logistical support that will enable them to pursue innovations and seek answers to some of the most interesting and challenging questions in chemistry.

To learn more about opportunities at ICCAS, visit **http://english.ic.cas.cn/.**

Sponsored by

中国科学院化学研究所
INSTITUTE OF CHEMISTRY CHINESE ACADEMY OF SCIENCES



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

24 SEPTEMBER 2021
VOLUME 373
ISSUE 6562

SPECIAL SECTION

Human Genomics

INTRODUCTION

1458 The human genome

NEWS

1460 Baby steps *By J. Kaiser*

REVIEWS

1464 From variant to function in human disease genetics
T. Lappalainen and D. G. MacArthur

1468 Discovery and implications of polygenicity of common diseases
P. M. Visscher et al.

1474 Mapping genomic and epigenomic evolution in cancer ecosystems
T. Ushijima et al.

1479 Insights into human history from the first decade of ancient human genomics *Y. Liu* et al.

ON THE COVER

Visualization of the human genome as a spiral (scale: 1 million bases per centimeter), with colored segments representing different chromosomes. Circle size denotes the number of genes associated with Mendelian disorders; hollow circles indicate the number of mutation clusters from a pan-cancer analysis. This special issue explores lessons from past studies of human genomics, with an eye toward future research efforts.



See page 1458.
Graphic: Martin Krzywinski/Canada's Michael Smith Genome Sciences Center; Data: OMIM; E. Pleasance et al., Nat. Cancer ***1***, *452 (2020)*

SEE ALSO EDITORIAL p. 1419; NEWS STORY p. 1430; PERSPECTIVES pp. 1440 & 1442; POLICY FORUM p. 1444; BOOKS p. 1448; RESEARCH ARTICLE p. 1499; PODCAST

NEWS

IN BRIEF

1420 News at a glance

IN DEPTH

1424 Costs balloon for huge new U.S. physics project
Department of Energy says it has no plan to downsize $3 billion neutrino experiment
By A. Cho

1425 Scientists have high hopes for Peru's leftist new president
Pedro Castillo, a former teacher, has promised to support research and bolster the country's fight against COVID-19
By D. Meza

1426 Footprints support claim of early arrival in the Americas
If dates hold, tracks would put people in New Mexico at the peak of the ice age, thousands of years earlier than thought *By L. Wade*
REPORT p. 1528

1427 Threat of drug-resistant malaria looms in Africa
Analyses of patient blood samples suggest artemisinin effectiveness is slipping
By D. Normile

1428 Israel's high case plateau may offer worrisome portent
Opposing forces keep outsize case count steady: school openings, holidays work against boosters *By M. Wadman*

1429 Large study will probe Long Covid
With up to 40,000 participants, U.S. study will investigate causes of lingering symptoms
By J. Kaiser

FEATURES

1430 Hot pursuit
Daniel Kastner's genetic sleuthing has helped explain—and treat—unusual disorders marked by fever and inflammation
By M. Leslie
HUMAN GENOMICS SECTION p. 1458

INSIGHTS

PERSPECTIVES

1434 Tree-ring dating of musical instruments
Dendrochronology detects fraudulent art but with some caveats *By P. Cherubini*
PODCAST

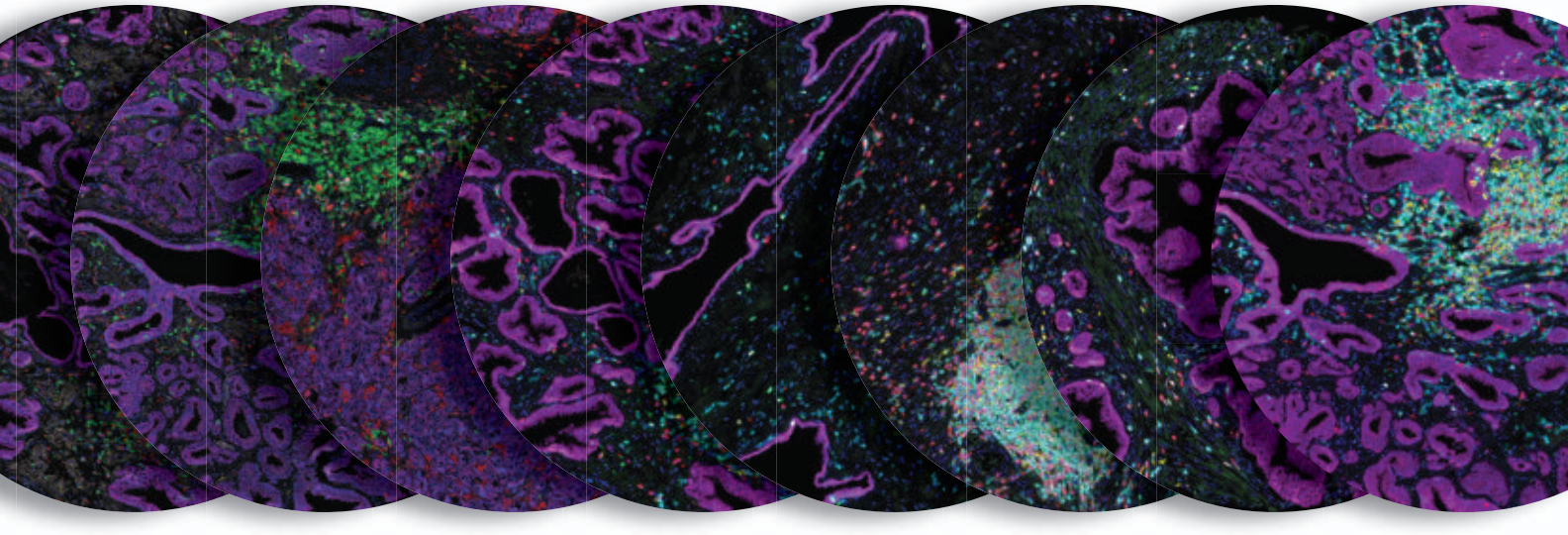
1436 Photons and qubits get a better connection
Quantum networks will require versatile quantum interconnects *By A. M. Kaufman*
REPORT p. 1511

1438 Strange bedfellows inside a superconductor
An unexpected link is found between the "strange metal" phase and charge density waves *By M. Le Tacon*
RESEARCH ARTICLE p. 1506

1439 Reversing aging for heart repair
Short-term induction of pluripotency gene expression allows adult mouse cardiomyocytes to regenerate
By Y. X. Wang and H. M. Blau
REPORT p. 1537

1440 Missing heritability may be hiding in repeats
Unexplored variable number tandem repeats make a large contribution to complex traits *By M. Gymrek and A. Goren*
HUMAN GENOMICS SECTION p. 1458; RESEARCH ARTICLE p. 1499

GRAPHIC: MARTIN KRZYWINSKI/CANADA'S MICHAEL SMITH GENOME SCIENCES CENTER; DATA: OMIM; E. PLEASANCE *ET AL.*, *NAT. CANCER* **1**, 452 (2020)



Akoya's Phenoptics™ multiplex immunofluorescence (mIF) platform provides a deeper understanding of cancer biology.

Rethinking a fundamental concept about cancer

The structure of tumors found in cancer patients is as diverse as the disease. Akoya Biosciences Phenoptics™ platform allows one research team to reveal new details that could impact future treatments.

To treat tumors, scientists need a better idea of what drives them to grow. Some studies suggest that a tumor that includes significant amounts of stroma–a collection of surrounding cells other than cancerous ones–will be more aggressive. Recent research indicates that a tumor's prognosis is more complicated than that.

"Some evidence suggested that simple metrics, like the area of stroma versus tumor, provide predictive diagnostic information," says Artur Mezheyeuski, a researcher at Uppsala University, Sweden, and the founder and CEO of HistoOne, a services company that offers pathology, image analysis, and bioinformatics support for the academia, pharma, and biotech industry. "So, we wanted to apply an unbiased approach to cover many tumor types and see how this simple method behaves."

For that purpose, Mezheyeuski and his colleagues used image analysis that could be trained to measure the stroma content of tumors from the most common cancers–something lacking in previous work.

Collecting spatial and cellular details

As Mezheyeuski describes it, "We wanted to distinguish different elements of the stroma." To do that, he turned to Akoya's Phenoptics platform.

Phenoptics 2.0 provides spatial and phenotypic information from formalin-fixed, paraffin-embedded (FFPE) samples. A sample is stained for biomarkers, imaged, and analyzed. Starting with resected specimens from patient tumors, Mezheyeuski's team used the platform to categorize sample areas as tumor or stroma.

"The platform's machine learning tool can be used to train the software to distinguish tumor and stroma," Mezheyeuski explains. The time required for training depends on the tumor type. "If the tumor is very simple, the training can take just 1 or 2 hours," he says. "More complex tumors take several hours of training."

With the spatial data and molecular information about the cells provided by biomarkers for more than 2,700 tumors, Mezheyeuski found something new.

No simple metric

For one thing, the amount of stroma varies across cancer types. Mezheyeuski found 25% stroma in renal tumors and as much as 70% in some pancreatic tumors. In addition, there can be a lot or a little stroma even in the same kinds of cancers from different patients.

To get an idea of a tumor's aggression, Mezheyeuski and his colleagues compared the amount of stroma in patients' tumors to their survival. In 13 of 16 types of cancer, multivariable analysis revealed no significant correlation between the amount of stroma and how long a patient lived. Even when there was a correlation, a higher level of stroma did not necessarily connect to an aggressive cancer. In one of the 16 cancers studied, a form of intestinal tumor, a higher level of stroma did correlate with a poor prognosis. But, in two others–one kind of pancreatic cancer and a type of breast cancer–more stroma indicated a better prognosis.

Some of the new findings come from the Phenoptics platform, which allows the team to combine information from the image with cellular staining. As these scientists noted in a study recently published in *EBioMedicine*, "This allowed a reliable comparison between the different tumor types."

From this work, Mezheyeuski's team concluded that a "crude measurement of the tumor stroma content might be insufficient for the purpose of prognostication, and it is likely that other stromal qualities, including mechanical features, extracellular matrix components or immune cell composition have overriding importance in the tumorigenic process."

Still, as scientists learn more about the stroma in tumors, cancer can be more thoroughly examined. That requires the right equipment for the job. Speaking of Akoya's Phenoptics platform, Mezheyeuski says, "It's a really versatile tool."

Reference

1. P. Micke *et al.*, *EBioMedicine* **65**, 103269 (2021), doi: https://doi.org/10.1016/j.ebiom.2021.103269.

Sponsored by

AKOYA BIOSCIENCES®
THE SPATIAL BIOLOGY COMPANY®

IMAGES PROVIDED BY AKOYA BIOSCIENCES



1442 The legacy of the Human Genome Project
The initiative forever altered biomedicine, but work remains to fulfill its true potential
By J. E. Rood and A. Regev
HUMAN GENOMICS SECTION p. 1458

POLICY FORUM

1444 Regulating forensic genetic genealogy
Maryland's new law provides a model for others *By N. Ram* et al.
HUMAN GENOMICS SECTION p. 1458

BOOKS *ET AL.*

1448 The allure of the ordered Universe
A science writer's ode to patterns packs a visual punch *By S. H. Ali*

1449 Biology meets public policy
The tenuous portability of polygenic scores undermines an attempt to link genetics and justice *By A. Panofsky*
HUMAN GENOMICS SECTION p. 1458

LETTERS

1451 Retired satellites: A chance to shed light
By E. C. Stokes et al.

1452 Protect minorities in genetic research
By V. Lipphardt et al.

1452 True informed consent requires more than forms
By L. H. Glantz

1453 Technical Comments abstracts

RESEARCH

IN BRIEF

1485 From *Science* and other journals

RESEARCH ARTICLES

1488 Coronavirus
Vaccine nationalism and the dynamics and control of SARS-CoV-2
C. E. Wagner et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABJ7364

1489 Enzyme mechanism
Hybrid radical-polar pathway for excision of ethylene from 2-oxoglutarate by an iron oxygenase *R. A. Copeland* et al.

1494 Batteries
Carbon-free high-loading silicon anodes enabled by sulfide solid electrolytes
D. H. S. Tan et al.

1499 Human genomics
Protein-coding repeat polymorphisms strongly shape diverse human phenotypes *R. E. Mukamel* et al.
HUMAN GENOMICS SECTION p. 1458; PERSPECTIVE p. 1440

1506 Superconductivity
Restored strange metal phase through suppression of charge density waves in underdoped $YBa_2Cu_3O_{7-\delta}$ *E. Wahlberg* et al.
PERSPECTIVE p. 1438

REPORTS

1511 Quantum optics
Entanglement transport and a nanophotonic interface for atoms in optical tweezers
T. Đorđević et al.
PERSPECTIVE p. 1436

1514 Topological optics
Topological insulator vertical-cavity laser array *A. Dikopoltsev* et al.

1518 Catalysis
Steam-created grain boundaries for methane C–H activation in palladium catalysts *W. Huang* et al.

1523 Synthetic biology
Cell-free chemoenzymatic starch synthesis from carbon dioxide *T. Cai* et al.

1528 Paleoanthropology
Evidence of humans in North America during the Last Glacial Maximum
M. R. Bennett et al.
NEWS STORY p. 1426

1532 Plant science
A plant lipocalin promotes retinal-mediated oscillatory lateral root initiation
A. J. Dickinson et al.

1537 Heart development
Reversible reprogramming of cardiomyocytes to a fetal state drives heart regeneration in mice *Y. Chen* et al.
PERSPECTIVE p. 1439

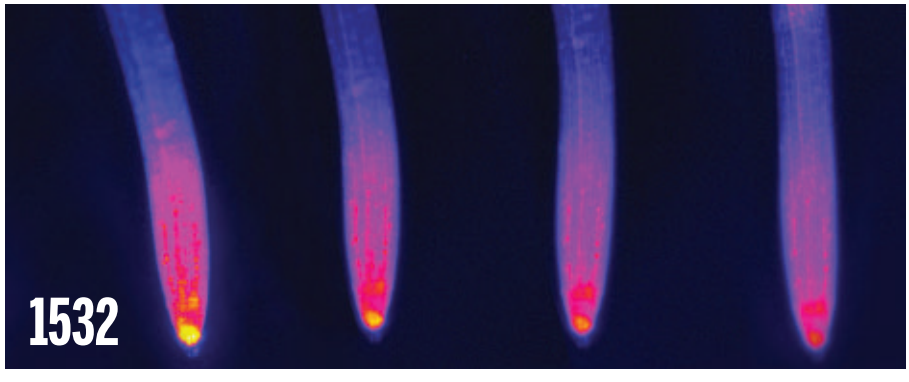
DEPARTMENTS

1417 Editorial
Build back wiser *By Meagan S. Mauter*

1419 Editorial
A cautionary history of eugenics
By Adam Rutherford
HUMAN GENOMICS SECTION p. 1458

1546 Working Life
What my coach gave me *By Laura Rossi*

AAAS News & Notes 1456
New Products 1544
Science Careers 1545



CREDITS (TOP TO BOTTOM): KATLEHO SEISA/ISTOCK.COM; A. J. DICKINSON *ET AL.*

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

Build back wiser

The US Infrastructure Investment and Jobs Act (IIJA), which is winding its way through Congress, makes progress toward President Biden's campaign promise to "build back better" by investing $0.55 trillion to repair the nation's transportation, water, and energy systems. But this allocation is only a fraction of the American Society of Civil Engineers estimated $2.5 trillion infrastructure investment deficit. To bridge this value gap, the nation must build back "wiser" by investing IIJA dollars in digitized, versatile, distributed, and inclusive infrastructure systems.

Wise infrastructure creates value by integrating digital and physical systems to sense, analyze, control, and adapt in real time. Services expand and costs decrease by building less and coordinating more. Although the IIJA dedicates funding for smart roadways that reduce congestion without costly expansion, the greater value proposition is in systems that integrate diverse infrastructure sectors. Those same smart roadways could anticipate power demand from electric vehicles, signal for increased electricity generation, and coordinate powerplant dispatch with air-quality sensors to protect community health. The scientific community will be instrumental in identifying opportunities for cyber-physical systems integration and mitigating the risk that coupled systems precipitate cascading system failures.

> "...wise infrastructure is inclusive of communities that have been bypassed..."

Wise infrastructure investments also maximize asset versatility, making infrastructure multifunctional and repurposing underutilized assets on demand. Imagine sidewalks that also house a low-cost operating system for last mile infrastructure management or water utilities that exploit excess water storage capacity to modulate electricity demand and stabilize the electric power grid. The scientific community should engage communities and manufacturers to expand multifunctionality and anticipate implications for asset life span.

Wise infrastructure also maintains value over time by augmenting aging or inadequate centralized systems with distributed, small-scale systems in a hybrid network model. Rather than repairing and expanding centralized water distribution networks, for example, investment should flow into distributed, fit-for-purpose water treatment systems that enhance the resiliency of centralized networks while reducing the cost of hedging against drought, water quality fluctuations, population growth, and emerging contaminants. The scientific community must help identify new coordination, maintenance, and equity challenges raised by hybrid systems. Federal agencies should fund test beds and demonstration activities that help researchers, developers, and communities to uncover and correct new vulnerabilities.

Finally, wise infrastructure is inclusive of communities that have been bypassed, bullied, and bulldozed in past efforts to modernize. The inadequacy of current infrastructure in these neighborhoods magnifies the potential returns on investment and innovation and amplifies the importance of community-led, equity-driven infrastructure design and operation. This makes social scientists crucial partners in developing tailored, scalable, and secure digital infrastructure platforms on which future physical infrastructure operates. It also makes workforce training and public-private partnerships essential for sustaining community engagement and ensuring that these new infrastructure assets have enduring value and utility.

Critics may argue that wise infrastructure projects haven't been proven at scale, aren't shovel-ready job makers, or don't fall under the siloed jurisdictional authority of existing federal agencies. Federal program offices should be reassured by the vast number of private, municipal, and international efforts to deploy such systems and might invest in documentation tools to distill lessons from failed demonstration projects and disseminate best practices from successful ones. Translating such insights for diverse community settings and scaling these systems nationally will create jobs for a new workforce skilled in designing, building, and operating next-generation infrastructure systems. The IIJA should also institutionalize interagency coordination across diverse federal agencies to manage the proposed spending. The federal budget, carbon budget, and water budget all depend on exploiting second-order, systems-of-systems efficiencies across infrastructure systems.

Strategies that the United States pursues under the IIJA will determine service quality, equity, resiliency, cybersecurity, climate preparedness, and costs for generations. The nation must not squander this opportunity to maximize returns on better infrastructure by prioritizing wiser investments.

– Meagan S. Mauter



PHOTO: DAVID COCHRAN

Meagan S. Mauter is an associate professor in the Department of Civil and Environmental Engineering at Stanford University, Stanford, CA, USA and the Research Director of the National Alliance for Water Innovation, Berkeley, CA, USA. mauter@stanford.edu

10.1126/science.abm3438

eppendorf

New epT.I.P.S.®
Box 2.0



Clever and Smart

Time for a fresh reboot

Often copied but unrivaled, the epT.I.P.S.® Box has been in laboratories worldwide since 2002. The fresh new epT.I.P.S Box 2.0 design features a bright white tone, clean lines and a redesigned closure button in typical Eppendorf blue, as well as the bluish-tinted transparent lid with indentations on the sides.

> No slits on the back of boxes to reduce risk of contamination
> Autoclavable up to 100 x for a long lifetime
> Lid with indentations on sides to improve stackability
> Small silicone feet in the bottom of the box reduce risk of slipping



www.eppendorf.com/epTIPS-News

Eppendorf®, the Eppendorf Brand Design, epT.I.P.S.® and PhysioCare Concept® are registered trademarks of Eppendorf AG, Germany.
All rights reserved, including graphics and images. Copyright © 2021 by Eppendorf AG.

EDITORIAL

A cautionary history of eugenics

A century ago this week, 300 scientists, policymakers, and campaigners gathered at the American Museum of Natural History in New York to discuss their work about heredity and eugenics—the political ideology designed to sculpt societies through biological methods of population control (the meeting was highlighted in *Science* a week later). The aims of eugenics were to nurture the propagation of people deemed "desirable" and to reduce the number of "undesirable" or "defective" people, primarily through enforced sterilization. Although recognized as toxic now, back then, eugenics enjoyed popular and bipartisan support and would grow to be one of the defining ideas of the 20th century.

The meeting had been coordinated by the Eugenics Record Office at Cold Spring Harbor Laboratory in New York, founded by the prominent eugenicist Charles Davenport, and by the office's director, the equally zealous Harry Laughlin. Laughlin was the author of a "Model Eugenical Sterilization Law" to standardize state legislation to prevent people with "undesirable" characteristics from having babies. It would eventually be translated and adopted by the Third Reich. The conference treasurer was Madison Grant, best known for his popular book *The Passing of the Great Race*—a treatise on white supremacy that was pivotal in developing the policies of racial hygiene in Nazi Germany. These zealots and other scientists carved out specious ideas of human population control, which would lead to the coerced sterilization of some 70,000 in the United States and to genocide in Nazi Germany.

It is sobering to look back from today's perspective and see how mainstream such jarring ideas were only a century ago, and how integrated they were with the nascent science of human genetics. Eugenics policies emerged from racist, classist, and ablist beliefs and co-opted vague, nebulous definitions and pseudo-clinical categories—feebleminded, defective, imbecile. From its inception, eugenics was a political creed, but one that was wedded to a science that was immature and frequently wrong. Ultimately, in the US, forced sterilization primarily targeted the poor and those with disabilities and was deployed against African Americans, Indigenous Americans, and other marginalized groups. These policies lasted late into the 20th century, but coercive sterilization still occurs in the US today. Globally, this practice continues as a strategy for oppression and population control.

> "It is sobering to...see how mainstream such jarring ideas were only a century ago..."

After the atrocities of the Holocaust were exposed, popular support for eugenics waned, and it became a poisonous idea. Many eugenics offices around the world were shut or evolved into genetics research laboratories. The majority of scientists abandoned this now reviled ideology and applied their growing knowledge of heredity to better ends—techniques such as preimplantation diagnosis, which identifies possible genetic defects in embryos created through in vitro fertilization. Our understanding of polygenic traits and diseases has skyrocketed in the past few years, revealing the complex relationship between nature and nurture—a phrase coined by Francis Galton, the man who also invented the term "eugenics."

As science makes advances in reproductive biology, it is incumbent upon scientists and society to firmly grasp this cautionary history and to understand how a political idea transmogrified into policy for many countries. With modern genetic techniques, including precision gene editing, we are inventing unprecedented possibilities for control of human biology, and society should proceed with a clear understanding not just of the limitations of this science, but of its grim history. Today, we hear frequent discussions about ideas such as embryo selection—not just to reduce disease risk but to enhance traits—and indeed, companies are emerging in the US with this as a potential service offered by their future businesses. Before embracing such technologies, it's critical to remember that these techniques are both scientifically dubious and share an ancestry with the racist history of eugenics.

Scientific ideas are easily marshalled into political ideologies, regardless of whether the science is well understood or not. Cultures change, and today, we strive to change the culture of science to be more inclusive and to center the voices of those who only a century ago were the targets of an ideology that enlisted an immature science into its arsenal. J. B. S. Haldane, a titan of 20th-century biology, wrote in 1938—in response to the eugenic policies of the US and Nazi Germany—words that nearly 85 years later, are still true: "...I do not believe that our present knowledge of human heredity justifies such steps."

–Adam Rutherford



Adam Rutherford is a Research Fellow at University College London, London, UK. a.rutherford@ucl.ac.uk

PHOTO: STEPHAN JAKUBOWSKI

10.1126/science.abm4415

NEWS

IN BRIEF

Edited by **Jeffrey Brainard**



COVID-19

Aiding pandemic fight, India pledges to export vaccine again

India said this week its vaccine manufacturers will resume exporting COVID-19 shots in October, potentially eliminating a major roadblock for global vaccine equity. Indian officials said the country's producers could churn out 1 billion more coronavirus vaccine doses by the end of the year, but did not specify how many would be exported. The Serum Institute of India (above) and other local producers had stopped exports at the government's behest in the spring, when only about 2% of India's population had been fully vaccinated and daily reported cases in the country soared to nearly 400,000. That deprived the COVID-19 Vaccines Global Access Facility of hundreds of millions of doses it had planned to distribute in lower and middle-income countries. (So far, India has exported fewer than 70 million doses.) Now some 15% of India's population has become fully vaccinated, and daily new cases have fallen to about 30,000. "Our supplies are comfortable, and the pace of local vaccination has risen rapidly, leaving room for exports," Indian health minister Mansukh Mandaviya said ahead of a global COVID-19 vaccine summit this week.

Pfizer says vaccine works in kids

COVID-19 | Pfizer and BioNTech announced their messenger RNA–based COVID-19 vaccine prompts "robust" antibody responses in children ages 5 to 11. In a 20 September news release, the companies said the antibody levels, observed in 2268 children 1 month after a second dose, were on par with those seen in 16- to 25-year-olds, with "generally comparable" side effects. The younger children received one-third of the dose that older teens and adults get. But the companies did not report case numbers among the children in the new study, as they did for a similar trial of 12- to 15-year-olds. "It's frustrating that we're in an age of science by press release," says vaccinologist Paul Offit of the Children's Hospital of Philadelphia. The companies said they will quickly seek permission from U.S. and EU regulators for emergency use in 5- to 11-year-olds. The firms expect results from even younger children soon. U.S. pediatric cases of COVID-19 have grown by about 240% since July.

Second J&J shot adds protection

VACCINES | Johnson & Johnson said on 21 September that a second dose of its COVID-19 vaccine improved protection against moderate and severe disease in a large clinical trial. In February, the U.S. Food and Drug Administration (FDA) authorized a single dose of the vaccine based on earlier evidence of substantial efficacy. In the new trial, which involved more than 30,000 people in 10 countries, the two doses, administered 8 weeks apart, provided 75% efficacy against moderate to severe disease; participants who received the shots were compared with unvaccinated ones who got a placebo. Efficacy against severe to critical disease was 100%, although the number of COVID-19 cases observed was small—only

PHOTO: RAJA SENGUPTA

eight in the placebo arm, and none in the vaccine arm. Among trial participants in the United States, the two-dose regime provided 94% efficacy against moderate to severe disease; the single shot had delivered 67%. The second dose was generally well tolerated, the company said. About 14.6 million U.S. residents have received one dose of the vaccine, which uses a modified, common cold virus. The single-dose vaccine is also authorized in 64 other countries.

Coronavirus spillovers abound?

PUBLIC HEALTH | Viruses similar to SARS-CoV-2 may jump from bats to humans much more frequently than believed, according to work led by the global nonprofit EcoHealth Alliance, whose leader, Peter Daszak, has argued the COVID-19 pandemic started this way. After overlaying a map of the habitats of 23 bat species known to carry severe acute respiratory syndrome (SARS)-related coronaviruses with data on human populations, researchers found that close to 500 million people live in areas where spillovers can occur, including northern India and most of Southeast Asia. Based on antibody surveys and data on human-bat interactions in those areas, they conclude in a preprint that as many as 400,000 infections with SARS-related viruses likely occur annually, without being noticed or causing big outbreaks. But the estimate comes with a huge confidence interval because the underlying data are scant, and David Fisman, an epidemiologist at the University of Toronto, calls the modeling "shaky." Even if the number is off, the study is "important," says virologist Angela Rasmussen of the University of Saskatchewan, because it "provides a road map for thinking about how to quantify zoonotic spillover" to prepare for future pandemics.

Remdesivir draws more doubts

COVID-19 | Despite widespread use for COVID-19 cases and billions of dollars in sales, the antiviral drug remdesivir offers no clinical benefit to hospitalized patients already receiving oxygen support. That's according to the latest large-scale, phase 3 clinical trial published last week in *The Lancet*. In October 2020, after months of allowing emergency use of the drug, the U.S. Food and Drug Administration approved remdesivir (Veklury), which works by inhibiting viral RNA synthesis, for all hospitalized COVID-19 cases. Studies had shown that patients given the intravenous drug early after symptom onset improved more quickly than those given a placebo. But enthusiasm for the drug was quickly tempered by other results showing remdesivir had no effect on reducing mortality. The new trial, part of the European multicenter Discovery trial, evaluated outcomes in 857 patients on oxygen support who received either standard of care or the same care with remdesivir and found no difference in clinical effectiveness or mortality after 15 days.

Pandemic extends NIH preference

WORKPLACE | Reversing its previous practice, the U.S. National Institutes of Health (NIH) last week modified a policy meant to improve the odds for young scientists to win a research grant: Those for whom the agency has extended their eligibility may now request a further extension before the current one expires. The agency gives preference to these "Early Stage Investigators" (ESIs) in awarding its independent research grants; to qualify, they

BY THE NUMBERS

79%

Share of white U.S. Catholics who say they are at least partially vaccinated against COVID-19, versus 57% of white Evangelicals. Although the vaccines don't contain cells from aborted fetuses, many Evangelicals who refuse the shots note that such cells were used to test and make several of them. The Vatican has called the connection too "remote" to justify refusing shots. (Pew Research Center)

345

Number of universities in China offering degrees in artificial intelligence, up from zero in 2018. It's by far the most rapid growth of any degree program there during that period. (Center for Security and Emerging Technology)



Methane collected by pipes at a landfill in Irvine, California, is used to power 26,000 homes.

PHOTO: REUTERS/MIKE BLAKE

CLIMATE POLICY

U.S., EU seek cut in methane to hit 1.5°C target

The United States and the European Union last week proposed cutting record-high global methane emissions by nearly one-third by decade's end to control global warming. The primary component of natural gas, methane is a more potent greenhouse gas than carbon dioxide, but not as long-lived: It breaks down in the atmosphere in about 12 years. By targeting methane emissions from natural gas, landfills, livestock, and some farming, officials hope to keep the goal of limiting global warming to 1.5°C above preindustrial levels within reach. The European Commission and the Biden administration will announce new methane regulations later this year. They want other nations to join the pledge in advance of the U.N. climate summit in Glasgow, U.K., in November.

must have completed advanced degrees or medical residencies within the previous 10 years. But NIH had told women whose ESI status it previously extended because of a maternity leave or other personal reasons that they could not apply for an additional, COVID-19–related extension while the first extension was still in effect. Nearly 500 people signed a petition posted online on 12 September calling that practice discriminatory against women with children. The petition also questioned NIH's policy of tying ESI status to Ph.D. completion, noting that many scientists spend years doing postdoctoral work—leaving some with far fewer than 10 years of ESI eligibility.

Iran sanctions lead to jailing

EXPORT CONTROLS | The United States charged an Iranian-born medical geneticist and two family members last week with conspiring to launder money and violate sanctions on Iran. Mohammad Faghihi ran a Miami-based genetic testing firm, Express Gene, that received nearly $3.5 million in overseas wire transfers between 2016 and 2020, according to the U.S. government, which alleges that the company used some of that money to buy genetic sequencing machines that were shipped to Iran without a required license. Faghihi ran a lab at the University of Miami until leaving in 2020; he is an author on several high-profile papers on antisense transcription and was recently studying rare genetic disorders in Iran. A judge last week deemed Faghihi a flight risk and ordered him detained before trial.

Australia backtracks on preprints

FUNDING | After an outcry from scientists, the Australian Research Council (ARC) last week said it scrapped an unusual policy that forbade grant applicants from mentioning preprints in their funding proposals. After the council introduced the policy in 2020, some researchers said they unwittingly violated it and lost ARC grants, costing them job offers. In a statement, the council said the change "reflects contemporary trends and the emerging significance of preprint acceptance and use across multiple research disciplines."

Marburg outbreak in Guinea ends

INFECTIOUS DISEASES | The first outbreak of Marburg disease in West Africa has ended after just a single case, authorities declared last week. On 3 August, samples from a man who had died 1 day earlier in Guinea tested positive for Marburg virus, a filovirus similar to the Ebola virus that can also cause fatal hemorrhagic fever. Health authorities identified 173 contacts, but none developed symptoms over the 21-day incubation period. Authorities waited the recommended 42 days after the man's death before declaring the outbreak over. Carried by fruit bats, the Marburg virus has also caused outbreaks in central and southern Africa as well as laboratories in Europe. There are no approved vaccines or treatments for the disease.

Chile targets sexual harassment

#METOO | Chile enacted a law last week giving state-funded universities 1 year to develop protocols for preventing and responding to cases of sexual harassment, violence, and gender discrimination. Noncompliant institutions risk losing state accreditation and funding. In 2017, only about 12% of the nation's 60 universities had such protocols in place. The new law, originally proposed by the Chilean Network of Women Researchers, was partly a response to a wave of protests held across the country in 2018; students and faculty members demanded more support for victims of sexual harassment at universities. The new protocols must include plans to provide education aimed at preventing sexual harassment on campus. Institutions must also individually establish policies setting sanctions for offenders and develop plans to protect victims who come forward.



A ladybug is shown next to a 1-millimeter-wide microflier, for scale.

PHOTO: NORTHWESTERN UNIVERSITY

SENSORS

Shaped like seeds, tiny fliers ride the wind, collecting data

Engineers have designed and tested tiny flying machines—spinning, motorless microchips that could record environmental data after their release from airplanes. A team at Northwestern University modeled them after the helicopterlike, wind-borne seeds of maple trees and woody vines, but at a smaller scale. Their diameters range from 0.5 to 40 millimeters, including onboard antennas and transistors. Simulations and wind-tunnel tests suggest the microfliers would spread out as they slowly fell through the atmosphere; their predictable descent paths would make them suitable for monitoring air and water pollution as a networked sensor array, the designers say. Next, the team is working to build the devices from biodegradable materials. The work appears this week in *Nature*.

CONVERSATIONS

Science forum: COVID-19's origin

Science will host an online discussion about how SARS-CoV-2 emerged and sparked the pandemic, featuring several scientists who have carefully evaluated the "lab leak" and natural origin theories, the two leading contenders. At 9 a.m. EDT on 30 September, Senior Correspondent Jon Cohen will discuss the theories with researchers Jesse Bloom of the Fred Hutchinson Cancer Research Center, Alina Chan of the Broad Institute, Linfa Wang of Duke-NUS Medical School, and Michael Worobey of the University of Arizona. The event can be viewed live on *Science*'s Facebook, LinkedIn, and Twitter accounts, and a recording will be archived.

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



BII | Science Prize for Innovation

Apply for the new BII & Science Prize for Innovation today

Behind every life-changing solution is an entrepreneurial scientist–a creative mind who proved an idea in the lab and dared to carry it out in the world.

To encourage more scientists to translate their research, BioInnovation Institute (BII) & Science present a new annual award.

Our three winners will have their essays published in Science magazine and will be invited into BII's entrepreneurial ecosystem. In addition, the Grand Prize winner will receive a prize of USD 25,000 and each runner-up will receive USD 10,000 at a grand award show celebration in Copenhagen, Denmark.

The call for applications has just opened.
Apply before November 1, 2021.
www.bii.dk/scienceprize

Presented by BII & Science

BII BioInnovation Institute Science

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



IN DEPTH

A conveyor is already hauling rock out of South Dakota's Homestake gold mine, where a massive neutrino detector will be housed 1.5 kilometers underground.

PHOTO: MATTHEW KAPUST/SANFORD UNDERGROUND RESEARCH FACILITY

PARTICLE PHYSICS

Costs balloon for huge new U.S. physics project

Department of Energy says it has no plan to downsize $3 billion neutrino experiment

By **Adrian Cho**

The gigantic new experiment on which the future of U.S. particle physics depends is far overbudget and years behind schedule, and a leadership shake-up at Fermi National Accelerator Laboratory (Fermilab), the experiment's host, is adding to the turmoil. The Long-Baseline Neutrino Facility/Deep Underground Neutrino Experiment (LBNF/DUNE), a massive effort to probe nature's most elusive particles, is now expected to cost $3 billion, 60% more than the preliminary estimate, and construction has slipped 4 years, with first data expected in 2029. The rising costs obligate the Department of Energy to reassess the project's design, although DOE officials say they have no plans to downsize it.

"DOE is not planning to implement any scope reductions that would sacrifice the science reach of the experiment," Jim Siegrist, DOE's associate director for high energy physics, wrote in an email to *Science*. Some physicists hope DOE will get the money to cover the overruns from a $3.5 trillion spending plan to reinvigorate the economy, now being debated in Congress. The House of Representatives version of that spending package contains $1.3 billion for LBNF/DUNE. But passage is far from certain.

LBNF/DUNE would fire barely detectable particles called neutrinos from Fermilab, in Illinois, to an underground particle detector 1300 kilometers away in an abandoned gold mine in South Dakota. Neutrinos come in three types—electron, muon, and tau—which morph into one another as they zip along at near–light speed. LBNF/DUNE aims to put the theory of such neutrino oscillations to the definitive test and look for a difference in how neutrinos and antineutrinos oscillate that might help explain how the universe generated more matter than antimatter.

The budget overruns are the latest twist in a 20-year quest to build an enormous neutrino experiment deep in the Homestake Mine, where 1.5 kilometers of rock would block inference from cosmic rays. In 2001, physicists proposed that the National Science Foundation build a vast underground lab at Homestake, but 10 years later NSF's governing board balked. DOE quickly affirmed it wanted the neutrino experiment, but only at half of the $1.9 billion it was then estimated to cost. Physicists kept the project big by inviting international contributions, and in 2013 CERN, the European particle physics laboratory, said it would take part.

On 5 November 2015, DOE approved a conceptual design for the two-part experiment, along with a preliminary cost range of $1.26 billion to $1.86 billion. LBNF includes the buildings, a new neutrino beam at Fermilab, the experimental halls in South Dakota, and the vast refrigerated tubs to hold the neutrino detectors. DUNE consists of just the detectors, including one in South Dakota holding 70,000 tons of frigid liquid argon, and a smaller one at Fermilab to monitor the outgoing beam.

Approval to build the detectors is at least 1 year away, but construction of the facilities began in Illinois in 2019 and South Dakota in 2017. In May, workers tested the conveyor system that will haul 800,000 tons of rock out of the mine. Several factors have driven the steep rise in costs, Siegrist says: The mine's aging infrastructure needed upgrading, installing the delicate detector at Homestake will be more difficult than anticipated, and the smaller detector at Fermilab needs to be more sensitive than originally envisioned. Excavation and the construction of buildings account for roughly $2 billion of the cost, multiple sources say.

DOE's budget request for fiscal year 2022, which begins 1 October, also notes that international contributions have been lower than expected. Still, international partners are paying for half of the first two of an eventual four modules in the Homestake detector, and CERN will build the refrigerators for those two, notes Stefan Söldner-Rembold, a neutrino physicist at the University of Manchester and cospokesperson for the 1300-member DUNE team.

The overrun in the preliminary cost is so large that DOE rules require officials

to rethink the project's conceptual design. But they seem reluctant. Once new cost estimates are finalized, Siegrist says, "DOE would reaffirm the scope of the project and approve the new cost and schedule ranges." That is, DOE envisions maintaining the design and changing the cost, not the other way around.

Within a year that flexibility will vanish, as project leaders must deliver their final detailed costing and construction schedule. Given the work already done and the problems solved, researchers are confident they can formulate a reliable baseline, Söldner-Rembold says. "The difference is that now, everything is on the table," he says.

DOE's tack means LBNF/DUNE will need more money, which physicists hope will come from the spending package. "Having something really excellent that is important for the long-term competitiveness of the United States as an intellectual powerhouse and not finding that money seems really tragic," says Patrick Huber, a theorist at Virginia Polytechnic Institute and State University.

Representative Bill Foster (D–IL), a former Fermilab physicist who serves on the House science committee, says he's confident the spending plan will pass and will include funding for LBNF/DUNE, as U.S. particle physicists made the project their priority in a 2014 long-range plan. "When a project has the full support of the scientific community, if there are some costs increases, it's a mistake for Congress to cut it off," Foster says. If the plan does not pass, DOE may have to cut other high energy physics programs to support the more expensive LBNF/DUNE.

Fermilab may have to push ahead without its current director, Nigel Lockyer, who on 10 September announced he will step down once his replacement has been found. Lockyer declined to discuss his reasons, but DOE appears to want to shake up the lab's management. In a letter to LBNF/DUNE's international collaborators, Siegrist and Stephen Binkley, acting director of DOE's Office of Science, say the office "sees this change as an opportunity to reinvigorate all Fermilab efforts."

DOE officials and researchers credit Lockyer for guiding Fermilab through tough times. He arrived in 2013, 2 years after its flagship atom smasher, the Tevatron, shut down, ceding the 2012 discovery of the famous Higgs boson to CERN's Large Hadron Collider. Lockyer rallied Fermilab and U.S. physicists around the neutrino program, says Michael Turner, a cosmologist at the University of Chicago, who helped recruit Lockyer. "He's got us set up to lead in neutrinos and do the biggest project DOE has ever done." ■

RESEARCH POLICY

Scientists have high hopes for Peru's leftist new president

Pedro Castillo, a former teacher, has promised to support research and bolster the country's fight against COVID-19

By **Daniel Meza**

The victory of leftist candidate Pedro Castillo in Peru's June presidential elections has worried some in the country's business elite. Castillo, a former schoolteacher and union leader, has promised major changes to reduce poverty and inequality, including increasing taxes on the mining sector. But among Peruvian scientists, his early moves have raised hopes.

Castillo appointed a leading scientist as a top adviser earlier this month and has said he will address systemic problems in Peruvian science, including low budgets, a weak governance system, and a lack of prospects for young researchers. He has also vowed to better manage the country's response to the COVID-19 pandemic.

> **"We trust that the government will keep its promises and treat scientists well."**
>
> **Modesto Montoya**, National University of Engineering

Peru only spends 0.12% of its gross domestic product on science, far below the 0.6% average in Latin America. In May, after he had won the first round of the elections, the *amauta* (Castillo's nickname, which means teacher in Quechua) promised to significantly increase health and education budgets, elevate science and technology's role, and create a new science ministry. His plans received an enthusiastic welcome in a 22 May letter from 50 Peruvian scientists, most of them trained outside Peru, who urged Castillo to adopt a new national science, technology, and innovation policy.

That same month, Castillo asked physicist Modesto Montoya of the National University of Engineering (UNI) to create a panel to outline a new career path for scientists and a structure for a future Ministry of Science to replace the leadership role of the National Council of Science and Technology (Concytec). The informal group met with Castillo at the presidential palace on 27 August, and Montoya was appointed presidential adviser on scientific matters on 8 September.

Montoya says Peru's public research centers are "disjointed, competing against each other, duplicating efforts, and wasting money." A dedicated science ministry should improve coordination, he says. "Priorities are defined in the Council of Ministers," he says. "It is time to land on a more solid structure that would ensure that science focuses on the needs of the country," adds Martha Esther Valdivia Cuya, a biologist at the National University of San Marcos, who attended the 27 August meeting.

Researchers also hope the new administration will do more to stop or reverse Peru's brain drain. The country offers financial incentives to bring researchers home and set them up at a Peruvian institution, but does not guarantee good working conditions after their arrival. "There is no good equipment here, and it is not easy to buy materials and reagents," says chemist María Esther Quintana Cáceda, who returned to Peru in 2011 after a doing postdoctoral research on nanomaterials in Sweden. Scientists hope Castillo will be able to create more—and more attractive—research jobs, improve laboratory conditions, and simplify funding procedures.

In his Plan for the Pandemic, released in May, Castillo also promised to strengthen the role of science in fighting COVID-19, which has caused more deaths per capita in Peru than in almost any other country. He wants to streamline pandemic information, educate decision-makers about the use of scientific evidence, and even promote the development of a Peruvian COVID-19 vaccine.

Whether Castillo can make good on his promises remains to be seen. Peru Libre, his party, only has 37 of the 130 seats in Parliament; the first 6 weeks of his presidency have been rocky; and the new administration has many other priorities, foremost the economy. But based on his interactions with Castillo so far, Montoya is optimistic. "We trust that the government will keep its promises and treat scientists well," he says. ■

Daniel Meza is a science journalist in Lima, Peru.



These footprints in New Mexico might rewrite the history of the peopling of the Americas.

PHOTO: BENNET *ET AL.*/*SCIENCE*

ARCHAEOLOGY

Footprints support claim of early arrival in the Americas

If dates hold, tracks would put people in New Mexico at the peak of the ice age, thousands of years earlier than thought

By **Lizzie Wade**

Between 23,000 and 21,000 years ago, people squished through the mud along a lakeshore in what is now New Mexico, alone and in small groups, leaving behind their footprints. Or at least that's the conclusion of a new paper that Oregon State University, Corvallis, archaeologist Loren Davis calls "potentially groundbreaking." If the dates are right, the discovery would be the strongest evidence yet that people reached the Americas during the middle of the last ice age, thousands of years earlier than many archaeologists thought.

"If that's true … it's going to be a revolution in the way that we think about archaeology in the Americas," says Davis, who wasn't involved with the work. It might reignite debates about how people first reached the continent from Asia. But Davis and others would like corroboration of the surprising dates before they rewrite their understanding of when and how people arrived.

During the maximum extent of the last ice age, from about 26,500 to 19,000 years ago, land connected Russia and Alaska, allowing people to settle the now mostly submerged region archaeologists call Beringia. But glaciers covered much of Canada, blocking the way south into what's now the continental United States and beyond. Archaeologists once thought the first people arrived in the Americas by walking through a corridor that opened between the glaciers by about 13,500 years ago. In recent decades, however, data from multiple sites have suggested people were in the Americas at least 16,000 years ago, leading many researchers to suspect that the first arrivals skirted the ice by traveling down the Pacific coast by boat.

A smattering of sites have hinted at even older dates, but the claims have been controversial. For example, last year a *Nature* paper argued that humans left artifacts in a highland cave in Zacatecas, Mexico, at least 27,000 years ago, but many archaeologists doubt the fractured rocks are stone tools.

Footprints are "a whole different level of evidence," says Ciprian Ardelean, an archaeologist at the Autonomous University of Zacatecas who discovered the cave there. "When you have human feet printed on the ground, that's undeniable."

Over years of fieldwork in White Sands National Park in New Mexico, researchers have found thousands of footprints left by humans and animals around a now-dry lakebed, including extinct megafauna such as mammoths and ground sloths. The new paper, on p. 1528, focuses on 60 human footprints found in seven layers of sediment, like a "palimpsest of people walking over a long period of time," says Michael Bennett, an ancient footprint expert at Bournemouth University. Based on the sizes of those prints, he thinks most were left by teenagers and children who were perhaps fetching water or just passing the time. "People spend a lot of time playing. And what better place to play than the edge of a lake?" says team member Daniel Odess, an archaeologist at the National Park Service.

"There is no doubt that these are human footprints," says Kevin Hatala, an ancient footprint expert at Chatham University. Jennifer Raff, an anthropological geneticist at the University of Kansas, Lawrence, agrees. "So the question is then, how old are they?"

To find out, the researchers radiocarbon dated seeds embedded in several layers of earth between the footprints. The dating put the seeds between about 23,000 and 21,000 years old, during the height of glaciation. If the footprints are that old, people must have made it to the Americas before ice sheets blocked the path, meaning an early land journey might have been possible.

"We kept trying to refute our own findings," Odess says, for example checking that the ancient lake water's chemistry didn't skew the dates. "And it kept coming back as yeah, they really are that old."

"From the dating perspective I think the authors have done a very sound job," says Tom Higham, a leading radiocarbon dating expert at the University of Oxford.

But Davis suggests a nagging possibility: that the seeds are older than the footprints because they eroded out of older sediments, then sifted into the mud the team excavated. He'd like to see the team try optically stimulated luminescence dating, a method that reveals when quartz grains were last exposed to light, to date when the sediment around the footprints was buried. "With something so extraordinary, it would be nice if we had multiple lines of evidence," agrees archaeologist Ben Potter of Liaocheng University.

So far, the team has found no artifacts that could shine a light on the culture of the people who left the footprints. But Kim Charlie, an enrolled member of the Pueblo of Acoma in New Mexico, feels a deep connection. "Thousands and thousands of years ago, our ancestors walked this place," says Charlie, who has visited the footprints and even uncovered some herself. Seeing prints of humans together with extinct megafauna such as camels sheds light on why the Acoma language has a word for "camel," she says.

Odess says White Sands bolsters other traces of early occupation of the Americas. The prints "make all those other [very ancient] sites more plausible," he says.

But Potter thinks each site "needs to stand and fall on its own merits." White Sands, he says, is "one of the stronger cases for a very early occupation. It's not definitive. But it's stronger." ■

INFECTIOUS DISEASE

Threat of drug-resistant malaria looms in Africa

Analyses of patient blood samples suggest artemisinin effectiveness is slipping

By **Dennis Normile**

Public health authorities have long worried that the emergence in Africa of malaria resistant to artemisinin was only a matter of time. Parasites that can thwart this key component of front-line treatments appeared in Southeast Asia in the early 2000s and eventually led to clinical failures. The stakes are higher for Africa, which in 2019 saw 94% of both malaria cases and deaths worldwide—estimated at 229 million and 409,000, respectively. Two recent papers provide evidence that the dreaded resistance now has a foothold on the continent.

The findings are "a game changer," says malaria researcher Philip Rosenthal of the University of California, San Francisco. For now, there's no evidence that current treatments are failing outright; patients with the resistant parasites simply take longer to clear them. "I don't think this is a reason to panic," Rosenthal says, "but we should be very concerned that things may be changing, and we definitely need new drugs."

The mosquito-borne parasite *Plasmodium falciparum*, which causes most malaria in Africa, mutates rapidly. It developed resistance to once potent drugs such as chloroquine in the 1950s and '60s, leading to a surge in the world's malaria burden. The arrival in the 1990s of artemisinin, derived from a Chinese herbal remedy, reduced the toll. Artemisinin is now typically paired with one of several secondary drugs in orally administered artemisinin-based combination therapies (ACTs) that deliver a one-two punch to the parasites. It's also administered alone intravenously for severe disease.

By the mid-2000s, the malaria parasite had developed resistance to ACTs in Southeast Asia. Studies traced the resistance to mutations in a gene known as *kelch13*. Several groups launched surveys to watch for these mutations in Africa and in recent years have documented them in parasites isolated from African patients. The two recent studies show those mutations are slowing the rate at which drug treatment eliminates the parasite from patients' bodies.

The first study, published in April in *The Lancet Infectious Diseases*, focused on children brought to three health centers in Rwanda with fevers in 2018. Those confirmed to have malaria were treated with a 3-day course of artemether-lumefantrine, an ACT. After 3 days, most patients had cleared the parasites. But more than 10% of the patients were still infected, and their parasites carried the telltale *kelch13* mutations. The authors, led by Naomi Lucchi, a malaria specialist at the U.S. Centers for Disease Control and Prevention, called their finding "the first documented evidence of artemisinin partial resistance in Africa."

The second team, led by malaria specialist Toshihiro Mita of Juntendo University, has tracked malaria in northern Uganda since 2013. From 2017 through 2019, the team studied the response of malaria patients treated at St. Mary's Hospital Lacor in Gulu with intravenous artesunate, a water-soluble artemisinin derivative. In this study, researchers collected patient blood samples several times a day and checked how quickly the number of parasites declined. In 14 out of 230 patients, it took longer than 5 hours to halve the parasite load, which is a standard indication of delayed clearance. Thirteen of those patients had parasites with *kelch13* mutations, the team reported on 23 September in *The New England Journal of Medicine* (*NEJM*).

"These results are quite obviously important given that artemisinin combination therapies are the only category of drugs currently available for the treatment of *falciparum* malaria," says Ivo Mueller, an infectious disease epidemiologist at the Walter and Eliza Hall Institute of Medical Research.

The genetic details of the two new studies suggest the drug-evading mutations developed locally and were not imported from Southeast Asia. Drawing on previous studies of blood samples from malaria patients, both research groups found that *kelch13* mutations appeared in the mid-2010s and increased by the late 2010s, going from 7.4% to 20% in Rwanda and from 3.9% to 19.8% in Uganda. This suggests artemisinin resistance emerged several years ago, Mita says. But its geographic spread remains unclear. "It is a possibility that this parasite migrated from other regions near northern Uganda," he says, "but we don't have any data from those regions." Also unclear is whether overuse or misuse of artemisinin has driven resistance in Africa, researchers say.

"Fortunately, despite observed delay in parasite clearance, patients eventually clear their infections as long as the ACT partner drug remains effective, and resistance to partner drugs has not been confirmed in Africa," says Betty Balikagala, a Juntendo University epidemiologist and first author of the *NEJM* paper. Only when such resistance emerged did clinical failures start to occur in Southeast Asia, Mueller notes. "The public health relevance of these findings should not be overly dramatized," he says of the new studies.

Still, with few potential new drugs in the pipeline, he and others are urging new treatment strategies. Doctors in Southeast Asia have managed to maintain the efficacy of combination therapies by cycling through different partner drugs. "And if two drugs don't work, try three," Rosenthal says, referring to recent interest in adding another secondary drug to ACTs. The approach "has shown good efficacy in clinical trials," he says.

Artemisinin's diminishing efficacy is a "serious warning sign that we need to react to," says Abdoulaye Djimdé, a genetic epidemiologist at the University of Science, Techniques and Technology of Bamako. "We should not wait until kids start dying from resistant parasites before tackling this problem." ■



A patient receives a malaria test result from a community health worker in Rwanda, where researchers recently identified artemisinin resistance.

PHOTO: NICHOLE SOBECKI/VII/REDUX



A ritual transfer of sins marks the eve of Yom Kippur, a holiday that may have driven up COVID-19 cases.

PHOTO: AP PHOTO/ODED BALILTY

COVID-19

Israel's high case plateau may offer worrisome portent

Opposing forces keep outsize case count steady: school openings, holidays work against boosters

By **Meredith Wadman**

Israel, among the first countries to launch coronavirus vaccinations and the first to roll out booster shots on a large scale, is offering a disturbing glimpse of what could be in store for other rich nations if they begin to give boosters this fall. Israel launched its pioneering booster campaign in late July, prompted by a surge in cases reflecting the extreme contagiousness of the Delta variant, the loosening of restrictions, and an apparent waning of protection from vaccines given in early winter. But cases have risen even higher since, suggesting boosters are far from a panacea when children and others remain unvaccinated.

Since 30 July, Israel has given a third shot of messenger RNA vaccine to more than 3 million people, including a majority of those 40 and older. Yet Israel is "stuck in a status quo of 1000 or 900 new cases per million per day," says Ran Balicer, chief innovation officer at Clalit Health Services, Israel's largest health maintenance organization, "which is a very bad status quo to be stuck at."

Public health experts differ about exactly why a country of 9.3 million that is vaccinating so aggressively still has one of the highest rates of reported infections per capita in the world, more than twice that of the United States. Extensive testing and social factors may play a role. But David Dowdy, an infectious disease epidemiologist at the Johns Hopkins Bloomberg School of Public Health, says it's a sign that "providing booster shots alone does not dramatically change the course or trajectory of transmission at a country level. Because the majority of transmission is still occurring from people who are unvaccinated." Israel's vaccination rate—64% of its population has received at least two doses—puts it ahead of the United States but behind some 30 other countries.

Israel's experience is feeding into booster discussions such as last Friday's meeting of advisers to the U.S. Food and Drug Administration (FDA), which voted unanimously to recommend a booster dose of Pfizer and BioNTech's SARS-CoV-2 vaccine, but only for people 65 and older and those at high risk of severe disease.

Among Israelis 60 and older who received a booster, the risk of infection fell 11-fold in August and the risk of severe disease 20-fold compared with their twice-vaccinated peers, according to a study of 1.1 million Israelis published in *The New England Journal of Medicine* last week and discussed by the FDA committee. At the meeting, Ron Milo, a systems biologist at the Weizmann Institute of Science, presented data showing that during August, the virus' ability to spread—its so-called reproduction number—fell by 30%, until each infected person was infecting slightly less than one additional person, a threshold that's essential for ultimately ending an outbreak.

But Israel's recent case counts don't reflect that trend, noted advisory committee member Amanda Cohn, chief medical officer at the U.S. Centers for Disease Control and Prevention's respiratory disease center. "Why is it that if your [reproduction number] went below one … you're at your highest [new case] rates right now?" she asked at the meeting.

The paradox reflects social realities, Sharon Alroy-Preis, director of public health services for Israel's Ministry of Health, told the meeting. Public schools opened on 1 September and the Jewish High Holy Days, with their attendant travel and family gatherings, stretch from 6 to 27 September. (Many Israelis do not wear masks at indoor family gatherings.)

Balicer, who is chair of the expert committee that advises the Israeli government on pandemic response, agrees. "The combination of [unvaccinated] children meeting in school followed up by large family gatherings is the recipe for mass dissemination of the disease," he told *Science*.

A significant portion of the new infections is occurring in the 2 million Israeli children who are younger than 12, for whom vaccines are not yet approved. Whereas this group represented 24% of new infections on 14 August, that number had risen to 42% in the week ending 16 September. "There are two parallel forces driving this proportion up: school opening, and gradually increased protection of the older population," Balicer says. More than three-quarters of Israelis who are 60 and older have received boosters.

Israel's case count likely reflects a third factor, too—extensive testing, says Dvir Aran, a biomedical data scientist at the Israel Institute of Technology (Technion). "Since school started children must get tested if there was a confirmed case in their

class. This means that many more children are getting tested." (His 2-year-old daughter is now in quarantine with her mother, separated from him and her siblings, because a preschool classmate tested positive.)

Even as infections surge in children, daily new severe cases have declined from between 90 and 100 in late August to about 70 so far in September. Because children are less likely to get severely ill from COVID-19, "We are seeing a lot of new cases but we don't see a lot of new severe patients," says Dror Mevorach, a physician who cares for COVID-19 patients at Hadassah Hospital Ein Kerem and advises the government.

The booster campaign has also dramatically shifted the vaccination status of people who do get seriously ill. In mid-August, after vaccine efficacy had started to wane and before the effects of boosters had taken hold, 59% of severe patients were fully vaccinated. By mid-September, 70% of new cases of severe illness were in the unvaccinated. "All [our] severe and critical patients below 60 are nonvaccinated," Mevorach says of his hospital.

Dowdy cautions that Israel may be an outlier rather than a harbinger for Western nations. Schools have reopened in dozens of other countries that are not experiencing equivalent surges—and many nations with very low levels of vaccination aren't facing outbreaks akin to Israel's. "What's happening in Israel isn't happening really anywhere else in the world," he says. Ruth Hershberg, an evolutionary microbiologist at Technion, thinks Israel's high rate of testing could be a factor. "We are better at detecting cases [than most other countries]," she says.

But Aran worries an Israel-like situation awaits nations that began vaccination campaigns later, even if they begin to offer boosters. "Beware," he says. "The combination of Delta and waning immunity is bound to create very harsh outbreaks in other countries as well, not just in Israel."

Israeli experts hope a continuing booster drive will help the country emerge from the current holding pattern. "The majority of Israelis over 40 have [received] a booster. But it's a minority under 40. And infections all over the world are driven by the young," Hershberg says.

The country is also turning to other measures to contain its surge. Beginning on 3 October, to gain access to sporting and cultural events, restaurants, hotels, gyms, bars, universities, and religious services of 50 or more, people 12 and older will need to show proof they were fully vaccinated no longer than 6 months ago—or that they have received a booster. ■

COVID-19

Large study will probe Long Covid

With up to 40,000 participants, U.S. study will investigate causes of lingering symptoms

By **Jocelyn Kaiser**

Last week, the National Institutes of Health (NIH) announced a nearly $470 million grant to probe one of the mysteries of COVID-19: the debilitating symptoms that many people experience weeks or months after their infection clears. The study will enroll up to 40,000 adults and children newly and previously infected with SARS-CoV-2 to explore the causes of these postinfection effects, dubbed Long Covid, and look for clues to treatments and prevention.

"The only way … we're going to sort this out is with very large studies that collect lots and lots of data about symptoms, physical findings, and laboratory measures," NIH Director Francis Collins said at an online press briefing. But some advocates for the growing ranks of people with Long Covid complain that the study will include too few people now struggling with the disorder.

> **"The only way … we're going to sort this out is with very large studies that collect lots and lots of data."**
>
> **Francis Collins**, National Institutes of Health

Long Covid can include pain, fatigue, "brain fog," trouble sleeping, headaches, shortness of breath, fever, chronic cough, depression, and anxiety that linger or appear more than 4 weeks after an initial infection. The symptoms can be severe enough to interfere with work and daily tasks. Possible explanations include a hidden reservoir of SARS-CoV-2, a misfiring immune system, or a metabolic problem triggered by the infection, Collins said. By some estimates, 10% to 30% of COVID-19 patients develop Long Covid.

Pinning down those numbers is one of the goals of the Researching COVID to Enhance Recovery (RECOVER) Initiative, funded with a $470 million grant to New York University (NYU) Langone Health. NYU Langone will enlist more than 100 researchers at 35 institutions who will start to enroll patients next month using a common protocol. About one-half will be children, including some newborns, study leaders hope.

Although some people who already have Long Covid will be studied, about 60% will have acute infections, said NYU Grossman School of Medicine cardiologist Stuart Katz, the study's principal investigator. The study will include people with serious or milder COVID-19, as well as some people who were vaccinated before becoming ill.

By comparing those who quickly recover with others whose symptoms persist, researchers will look for risk factors as well as molecular processes that may explain the different outcomes and could be blocked with medication.

NIH hopes to find those molecular targets within 18 months and then launch treatment trials using money remaining from $1.15 billion over 4 years that Congress approved in December 2020 for Long Covid research.

Lisa McCorkell, who is part of the Patient-Led Research Collaborative, a Long Covid group started by patients who are also researchers, says if the RECOVER study engages meaningfully with patients, it will be "a welcome step toward finding answers."

But one prominent patient activist is disappointed the federal effort will largely enroll people who haven't yet developed Long Covid. NIH's study, says Diana Berrent, founder of Survivor Corps, which represents people with Long Covid, is "smoke and mirrors. This doesn't get us closer to any answers."

She also thinks RECOVER may struggle to recruit participants who aren't vaccinated—the group that likely accounts for most new cases of the disorder. "The people who are just getting Long Covid now because they're not vaccinated are not the same people who are going to enroll in an NIH longitudinal study," she says.

But NIH says it is seeing "robust" recruitment in trials testing COVID-19 treatments, suggesting enrollment in RECOVER won't be a problem. And Katz says enrolling people who don't yet have Long Covid is the only way to determine its true incidence. "I had Long Covid. It's a mystery. I understand that people are suffering," he says. "But we have to balance looking backwards versus looking forwards." ■

FEATURES



Daniel Kastner (left) diagnosed Sarkis Hagopian with familial Mediterranean fever in 1985, sparking a careerlong search for inflammatory diseases.

HOT PURSUIT

Daniel Kastner's genetic sleuthing has helped explain—and treat—unusual disorders marked by fever and inflammation

By **Mitch Leslie**

The bizarre symptoms had tormented Sarkis Hagopian since he was a young child crawling around his Pittsburgh home. Every so often, one of his knees would balloon or an ankle would swell so much that he couldn't put on his shoe. Fierce cramps would wrench his abdomen, leaving him unable to stand up straight. "I was pitched at a 45° angle," recalls the now-60-year-old printing salesman.

Those mysterious bouts, which usually lasted 2 to 3 weeks, prevented him from playing youth sports. "I could not get through a season without a swollen knee or ankle," he recalls. In third grade, he missed so much school that he was almost held back. Hagopian's parents took him to several doctors, but none could figure out what was wrong. "They would say I must have banged my knee." And after one severe episode during which he was hospitalized for weeks, doctors worried he had muscular dystrophy.

PHOTO: ERNESTO DEL AGUILA III/NHGRI

By the time he was 24, Hagopian had moved to Washington, D.C. He was determined to find someone there who could help. His joints were enlarging to grapefruit size in flare-ups that hampered his sales work. "When these symptoms show up, you are pretty much incapacitated," he says.

The next doctor he saw was as mystified as the others, but had the good sense to refer him to the National Institutes of Health (NIH). Hagopian turned up for his first appointment there in 1985, he recalls, and in walked "this young guy full of energy with a big smile."

The meeting was a watershed for both. Hagopian finally learned he suffered from a rare genetic illness, familial Mediterranean fever (FMF), and started to take the first treatment that salved his symptoms, the drugs colchicine and naproxen.

For the energetic, smiling guy, immunologist and physician Daniel Kastner, the encounter launched a more than 3-decade-long scientific quest into genetic diseases. Even before the human genome had been deciphered, Kastner teased out the DNA flaw behind FMF. He has gone on to define a new category of illnesses, known as autoinflammatory diseases, in which immune system malfunctions unleash inappropriate inflammation, resulting in a diversity of problems that includes arthritis, rashes, strokes, swollen eyes, and unexplained fevers.

Autoinflammatory diseases, which can vary from mild to lethal, affect more people than immunologists initially thought—FMF afflicts as many as one in 500 in certain populations. Researchers have uncovered more than 30 defective genes that trigger those conditions. Kastner and collaborators discovered or co-discovered 13 of them, reporting their most recent find in August.

By revealing what drives those diseases, Kastner's work opened the door for life-changing and even lifesaving treatments for patients like Hagopian, other scientists say. "It's the greatest example of bench to bedside you can think of," says pediatric rheumatologist Ronald Laxer of the Hospital for Sick Children in Toronto.

Autoinflammatory diseases may also yield insight into the organization of the immune system by revealing "chokepoints," or weak links that can fail and lead to illness, suggests Kastner, who rose through NIH's ranks to become scientific director of the National Human Genome Research Institute (NHGRI) in 2010. In addition, the maladies could shed light on some common illnesses, such as Crohn disease, gout, and Alzheimer's disease, that show autoinflammatory characteristics.

Having tracked a disease to its causative gene decades ago, Kastner now thinks the opposite approach could be fruitful: sequencing genomes from many people with unexplained health problems to pinpoint mutations and then determining whether those glitches are responsible for specific symptoms. "Taking these genotype-driven approaches," he says, "maybe we can define diseases that we don't even imagine right now." That strategy has already enabled him and his colleagues to uncover an unusual new autoinflammatory disease, and he plans to search for others.

KASTNER GREW UP IN LOCKPORT, a city near Buffalo, New York. Neither of his parents was involved in science—his factory worker father never graduated from high school—but Kastner says he has been enthralled by it "throughout my conscious life." As an undergraduate at Princeton University, however, he couldn't decide what to major in and eventually settled on philosophy. "You have freedom to think about a lot of things."

"It's the greatest example of bench to bedside you can think of."

Ronald Laxer, Hospital for Sick Children

Still, he ultimately decided to become a physician and enrolled at Baylor College of Medicine. But the mysteries of the immune system soon lured Kastner away from pursuing his M.D. "It was such an exciting time in immunology," Kastner says. "There were lots of black boxes—that was really exciting for somebody with a philosophy background" who liked to ask questions. After completing a Ph.D., however, he returned to medical school and finished his M.D. "I wanted to pursue the dual existence of doing science and seeing patients."

Kastner arrived at NIH in 1985 as a medical staff fellow, ready to begin his two-track career. Hagopian turned up shortly afterward and supplied a black box Kastner could try to pry open. Why FMF patients have periodic bursts of symptoms was a mystery, although some researchers thought the illness was an autoimmune disease, in which people's own immune cells or antibodies attack their tissues.

FMF clearly ran in families, particularly ones with Mediterranean or Middle Eastern backgrounds. An ambitious plan germinated in Kastner's mind. Francis Collins, the geneticist who now heads NIH, and other researchers were laboriously homing in on the gene that is defective in cystic fibrosis by using a technique called positional cloning. The approach involves identifying markers near the target gene by analyzing affected families. Kastner proposed a similar search to ferret out the mutated gene behind FMF.

At the time, the Human Genome Project to sequence our DNA had not started, and the whereabouts and identities of almost all human genes were unknown. Positional cloning was still unproved, and when Kastner proposed the project, his supervisor exploded. "He called me an idiot and other choice names." Kastner eventually won him over, but doubters remained. "Most people looked upon me as a Don Quixote figure."

Even before the project got off the ground, Kastner and colleagues hit a stumbling block—locating the large families with lots of FMF cases that could provide DNA. Hagopian is of Armenian ancestry, and Kastner first tried to recruit Armenian Americans from the Washington, D.C., area, enlisting a local priest to help. The church members who responded to his request rattled off plenty of health complaints, Kastner says, but none had FMF. Kastner eventually teamed with an Israeli doctor, Mordechai Pras of the Chaim Sheba Medical Center at Tel HaShomer, who had tracked down hundreds of patients with the disease. In 1989, Kastner flew back from Israel with 347 cell lines from patients and unaffected family members, which furnished a trove of DNA.

The researchers needed two other ingredients to complete the first step of their project: a set of DNA probes that would latch onto specific nucleotide sequences in the genome and a supply of fresh human placentas. The scientists planned to add the probes individually to DNA from the Israeli subjects and correlate the binding patterns with the pattern of FMF inheritance in the families.

The placentas, meanwhile, were meant to address a shortcoming of the probes: They were embedded in larger pieces of DNA that tended to stick to irrelevant "decoy" sequences scattered around the genome. Other researchers had found that DNA from placental tissue would glom onto those decoys and allow the probes to home in on their intended chromosome locations.

The probes were hard to come by; few researchers in the world could supply a comprehensive set that recognized sequences across the chromosomes. As for the placentas, Kastner picked up a supply every month from a local hospital. Kastner's older son was born during the study, but his wife forbade him from using that placenta, he recalls with amusement.

After 2 years and a wrong turn during which the scientists became convinced that

See more on the human genome in our special section on p. 1458.

the mutant gene resided on chromosome 17, they traced it to chromosome 16's short arm in 1991.

That was the easy part, Kastner says. The next stage—tracking down the gene itself—took another 6 years and required an international collaboration with researchers at the Los Alamos National Laboratory, the University of Adelaide, and other institutions then sequencing chromosome 16 for the Human Genome Project. "We were a microcosm of the genome project," Kastner says. "We had to go through all of the steps" they went through, such as generating high-resolution gene maps.

By June 1997, the researchers had winnowed the candidates to 10 genes near one tip of chromosome 16. Once they cleared several final hurdles, including a crucial mix-up over whether one of the Israeli subjects had the disease, the scientists nabbed the causative gene. Called *MEFV*, it encoded a novel protein of unknown function that they named pyrin, from the ancient Greek word for fire or heat. In a near-tie, a team of French researchers identified the same gene at almost the same time.

WITH THE FMF PROJECT complete, Kastner and colleagues began to search for additional mutated genes that trigger similar illnesses. As sequencing projects deciphered more and more of the human genome, the task grew easier. The researchers also had a bit of luck as they pored over DNA from seven families with unexplained fevers and other symptoms. Evidence suggested the gene fell within a region on chromosome 12, and the first gene the team checked turned out to be the right one. The gene, which they identified on Thanksgiving Day 1998, codes for one of the receptors for tumor necrosis factor alpha (TNF-alpha), a potent inflammation-stimulating molecule, and they named the disease that results from the faulty receptor TRAPS.

The discovery of a second disease-causing gene suggested ailments marked by fevers and inflammation might not be oddballs. Patients with FMF and TRAPS did not carry the self-targeting antibodies characteristic of autoimmune diseases such as lupus and rheumatoid arthritis. So Kastner and his team made an intellectual leap, arguing that TRAPS and FMF belonged to a new category of illnesses, which they dubbed autoinflammatory diseases. The name, he says, was a riposte to the "naysayers" who contended that the conditions had to be autoimmune diseases.

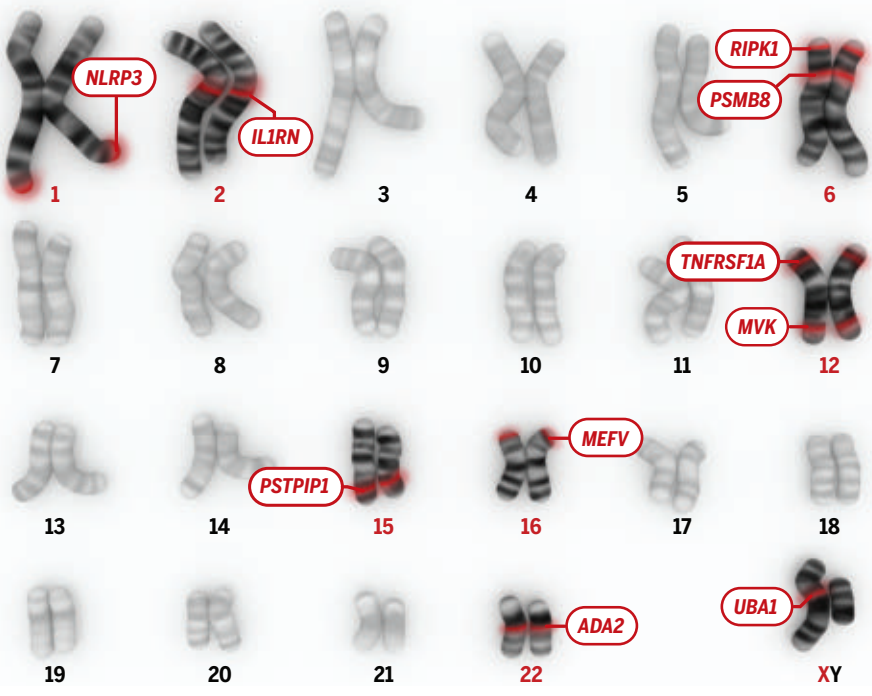
Kastner and other scientists have since identified many other autoinflammatory diseases, which go by an alphabet soup of acronyms, including DIRA and DITRA, CAPS and CAMPS, DADA2, HA20, PLAID, CANDLE, and SAVI. The diseases attack many parts of the body, but they share a general mechanism. The faulty genes behind them perturb the intricate control circuits that keep inflammation in check. That natural surge of white blood cells and the pathogen-fighting substances they release is vital to combating invaders, but the response can tilt from beneficial to damaging if is too intense, lasts too long, or happens in uninfected parts of the body.

In at least nine of the diseases, the inflammasome—an inflammation-managing, pinwheel-shaped protein cluster inside certain cells—occasionally goes into overdrive. A defect in pyrin, which normally responds to pathogens by forming an inflammasome, is often the culprit. In FMF, for example, mutations in *MEFV* cause cells to produce faulty pyrin that provokes inflammation without any apparent pathogenic trigger. CAPS results from defective copies of the protein NLRP3, part of which resembles pyrin and helps form inflammasomes.

People with FMF, CAPS, and most other autoinflammatory diseases inherited the faulty genes that cause their illness from

Heat map

The more than 30 faulty genes that cause autoinflammatory diseases are scattered across human chromosomes. These 10 illnesses, listed by their name and faulty gene (red), typically produce fevers, as well as a diversity of other symptoms. Familial Mediterranean fever (FMF) affects as many as one in 500 people in some populations; the other diseases are rarer.



CAPS *NLRP3*
Rash, joint pain and swelling, deafness, meningitis, brain damage

TRAPS *TNFRSF1A*
Eye inflammation; skin lesions; abdominal, joint, muscle, and chest pain

PAPA *PSTP1P1*
Arthritis, skin ulcers, severe acne

CRIA syndrome *RIPK1*
Swollen lymph nodes, headaches, abdominal and joint pain, mouth ulcers

DADA2 *ADA2*
Muscle and joint pain, blood vessel inflammation often leading to strokes, bone marrow malfunctions leading to reduced white and red blood cell counts, immunosuppression

DIRA *IL1RN*
Rash, bone inflammation, mouth lesions, spleen and liver swelling

FMF *MEFV*
Swollen joints, rash, kidney damage, abdominal and lung inflammation

CANDLE *PSMB8* and several others
Muscle wasting, skin lesions, swollen eyelids, loss of fat on face and upper body, delayed development

HIDS *MVK*
Rash, joint pain, arthritis, liver and spleen swelling, swollen lymph nodes

VEXAS *UBA1*
Blood clots, lung inflammation, cartilage inflammation in nose and ears, low blood cell counts

CREDITS: (GRAPHIC) V. ALTOUNIAN/*SCIENCE*; (DATA) ONLINE MENDELIAN INHERITANCE IN MAN®

one or both parents. But a gene-first approach has now led Kastner and collaborators to an exception. The idea for the research, Kastner says, came from an "impatient young person," NIH genetics fellow David Beck, who proposed sequencing more than 800 genes involved in a process that enables cells to dispose of unneeded proteins. After analyzing those genes in more than 2500 patients who had come to NIH with undiagnosed conditions, the researchers discovered three middle-aged men who suffered from problems such as lung inflammation, skin lesions, and irritated cartilage in the nose and ears. All three carried glitches in a gene called *UBA1*. By searching other groups of undiagnosed patients, the team uncovered 22 more people with mutations in the gene.

The novelty of VEXAS syndrome, as Kastner and colleagues called it in their 2020 report, was that patients had acquired their disease-causing mutations during their lives, in much the same way that sporadic cancer results from new mutations. Because the gene is on the X chromosome, men only have one working copy, so they have a much greater chance of developing VEXAS. Researchers have since reported cases of VEXAS syndrome in older women who had lost one of their two X chromosomes from some of their cells. Other, as-yet-undiscovered autoinflammatory diseases may also stem from acquired mutations.

"MOST CONCEPTS in immunology have a life span of 5 to 10 years," says immunologist Jean-Laurent Casanova of Rockefeller University. "The concept of autoinflammatory diseases is good and will stand the test of time." Pediatric immunologist Hal Hoffman of the University of California, San Diego, adds that the concept helped explain recognized immune disorders that nobody knew how to categorize. What Kastner has done, he says, is "give all those diseases a home and give us a way to think about treating and managing them."

The idea promises to deepen understanding of other conditions as well. Some researchers argue, for example, that autoimmune and autoinflammatory diseases fall along a single continuum of immune system disorders.

Helen Lachmann, a physician who studies autoinflammatory diseases at University College London, says Kastner's impact reflects his ability to pick the right challenges. "He's very good at spotting not just the big problems, but the solvable problems. It's a rare gift." And once he has identified a problem, he knows how to marshal research talent to solve it, she adds. Two scientific talents he helped nurture, his former postdoc Ivona Aksentijevich and former rheumatology fellow Raphaela Goldbach-Mansky, have collaborated with him on most of the gene discoveries and independently probe autoinflammatory diseases as NIH researchers.

Kastner's legacy also includes the first targeted treatments for several types of autoinflammatory diseases. Until his discoveries, the standard therapies were drugs such as colchicine, an anti-inflammatory that doctors have deployed against conditions as diverse as cancer, cirrhosis, and heart arrhythmias. But uncovering the genes that go awry in FMF and related diseases pointed to specific causes for symptoms, such as surges in inflammation-promoting molecules like interleukin 1 and TNF-alpha.

Pharmaceutical companies have developed several drugs, such as anakinra and canakinumab, to block those proinflammatory molecules. Those medications have become standard therapies for autoinflammatory diseases.

The best illustration of the treatments' power may be NOMID, a severe variant of CAPS, Kastner says. Rashes and fevers begin when patients with the illness are infants. But as the patients age, they develop meningitis that by their teen years can lead to deafness and brain damage. Kastner, Goldbach-Mansky, and colleagues gave anakinra to NOMID patients in a 2006 clinical trial. "Within a few hours the fevers go away, the skin rash goes away," he says. Brain scans showed that neuroinflammation subsided in the patients who had meningitis.

However, Kastner notes, children with NOMID must start treatment early in life—and continue on the drug as they age—to gain any benefit. The treatment can halt, but not reverse, brain damage. Now, a few clinical trials are assessing other potential therapies for autoinflammatory diseases.

Kastner has spent his entire working life at NIH. He recently turned 70 and has no plans to retire, though he is stepping down as NHGRI's scientific director. One reason he is leaving, he says, is to give the institute a chance to increase diversity in its leadership. His replacement, announced 13 September, is genetic epidemiologist Charles Rotimi, born in Nigeria.

Kastner plans to continue the physician-scientist's double life. "I think having the combined existence was really important for me," he says. "I will continue to see new patients and hopefully will continue to discover new diseases."

One long-term patient he still sees is Hagopian. The colchicine and naproxen Hagopian started taking in 1985 tamped down his symptoms for 30 years. However, Hagopian says the drugs were deceptive. "When you get relief from a treatment, you feel like you've been cured." Six years ago, he had to be rushed to the hospital because he had developed an intestinal obstruction caused by adhesions, belts of connective tissue in the abdomen that result from inflammation.

Kastner suggested Hagopian start injections of anakinra. The drug banished nagging problems, such as back pain he had attributed to aging, Hagopian says. "I never knew what it felt like to be normal until I was 55 years old." ■



Daniel Kastner (right) chats with a child who has an undiagnosed autoinflammatory disease affecting his liver.

PHOTO: ERNESTO DEL AGUILA III/NHGRI



FORENSIC SCIENCE

Tree-ring dating of musical instruments

Dendrochronology detects fraudulent art but with some caveats

By **Paolo Cherubini**[1,2]

Stringed instruments are among the most appreciated and valuable works of art. For example, the "Vieuxtemps" Guarneri "del Gesù" violin sold for more than $16 million in 2012. Instruments made by luthiers in northern Italy in the 17th and 18th centuries are particularly valued. They are regarded to have superior tonal qualities and are renowned for having been the preferred choice of many leading musicians for centuries, for their resonance and ease of response. This reputation is still a matter of debate (*1*), but it is clear that both the skills of the luthier who made the instrument and the quality of the timber used are key determinants. Because the market value of an instrument is based mainly on its craftsmanship, how can its authenticity be assessed?

Verifying the origin of a stringed instrument is challenging and often questioned because an examination of style, design, construction details, and varnish appearance alone is insufficient to confirm the maker. A dendrochronological analysis of the wood used to make the work of art—a method also used to detect fraudulent paintings—is the only objective nondestructive analysis that can indicate the date after which an instrument could have been made. Climatic variability influences tree growth such that different trees of the same species growing at the same site have similar tree-ring patterns. The cross section of a tree reveals a pattern of concentric rings of wood, each produced during a growing season. All trees growing at a site will form rings of similar widths depending on the meteorological conditions of that specific year. The most basic principle of dendrochronology is cross-dating, or the comparison of ring-width patterns, using statistical methods. This approach provides a level of statistical confidence that gives the results credibility (*2*). Dendrochronological dating of wood is achieved by cross-dating different ring-width series. The year of formation of the last ring on the front of a violin can be identified by comparing the pattern of tree rings visible on the front of the instrument with the ring-width pattern from a chronology built with living trees growing at the site from which the wood was putatively taken. The year of formation of the last ring will be the terminus post quem, the date after which the instrument could have been made.

The back, sides, and neck of most string instruments are usually made from maple. The front is generally made from Norway spruce. The physical and chemical characteristics of the wood used to make an instrument affect its acoustic properties, namely its radiation spectrum (*3*). Typically, wood of trees grown at higher elevations, with regular, narrow rings and a straight grain direction of the wood tracheids, is preferred.

[1]Dendrosciences, Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland. [2]Department of Forest and Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, BC, Canada. Email: paolo.cherubini@wsl.ch

PHOTO: MUSEUM/HERITAGE IMAGES/GETTY IMAGES



The authenticity of the Stradivari violin of 1716, known as the "Messiah," has long been disputed. Its controversial dating created intense confusion in the violin community regarding the reliability of tree-ring dating.

The wood used to make musical instruments is called resonance wood and is produced in several different regions in Europe. The geographical origin of the wood used by the old violin-making schools is surrounded by mystery, legend, and anecdotes, but it is usually thought to have come from forests in the Eastern Italian Alps. The Paneveggio forest in the Fiemme Valley is said to have been one of the most favored forests for producing resonance wood used by the old schools. Legend has it that Antonio Stradivari himself went to the forest to select the most appropriate wood for his violins, where he recognized the right wood by the sound the timber made as it was slid down-valley. Sadly, such reports are not confirmed by any historical records. In Paneveggio, resonance wood is often referred to as hazel wood, characterized by indented growth rings. Sound transmission is thought to be slower in wood with hazel growth, making the wood more resonant. An early observation of this was made by luthier Jean-Baptiste Vuillaume (1798–1875) on the Stradivari "Messiah" violin of 1716 (*4*). It is also commonly seen in instruments made by Giovanni Battista Guadagnini (1711–1786), and it is found in an Antonio Stradivari (1644–1737) cello of 1710. However, it is not found in all classical instruments.

For more than a century, scientists have sought to understand the secret technique used by the old violin-making schools to obtain wood of such high tonal quality (*5*). Many hypotheses have been proposed to explain the unusual sound quality of instruments from these schools. Were these instruments treated with special varnishes or complex mineral preservatives; made of wood that had been stored for a long time before use; soaked in (sea) water to leach chemicals from the wood tissues; or made from trees felled in a certain season? Although several studies have tested these hypotheses, the results remain controversial. Fungal decay has been proposed as a way to improve the acoustic properties of wood by reducing wood density (*6*). Another hypothesis points to the climate that existed between 1645 and 1715, when the most prominent Cremonese masters, such as Stradivari, made their violins. Known as the Little Ice Age, this period was characterized by reduced solar activity and lower temperatures, which induced a notable reduction in tree growth and thus regular narrow rings (*7*). However, this would also imply that all violins from trees that grew during that period should have superior sound, which is not the case.

Corona (*8*) was the first to date two violins made by Giovanni Battista Gabrielli (1716–1771), and he identified the Fiemme Valley in the Italian Eastern Alps as the possible source region for violins made by the old schools. Major advancements were

made by Klein *et al.* (*9*), who confirmed the authenticity or required a reappraisal of the date of manufacture of many instruments. Dendrochronology also enabled the identification of the provenance of the wood of entire collections (*10*, *11*).

The authenticity of the 1716 "Messiah" Stradivari violin has long been disputed. In 1999, Pollens (*12*) reported an apparent contradiction between an inscribed letter in the pegbox and the mold from which it was made. By measuring the ring widths on a photograph of the violin's front, he dated the youngest ring to 1738, a year after Stradivari's death (*12*). The 1738 date was later retracted without a clear explanation, further confounding the controversy. Later, the ring widths were measured directly on the instrument's front, and the resulting chronology was cross-dated against those of two undisputed Stradivari violins. The treble side (between 1581 and 1675) and the bass side (between 1590 and 1682) of the instrument were dated. The terminus post quem was therefore 1682, a date consistent with the attributed date of manufacture, 1716, supporting Stradivari as the maker of the "Messiah." Cross-dating with other instruments revealed that the "Messiah" correlates with the "Archinto" and "Kux-Castelbarco" violas. These instruments have a terminus post quem of 1687, which still supports the attribution to Stradivari. Nevertheless, this controversial dating of the "Messiah" created intense confusion in the violin community regarding the reliability of tree-ring dating.

It is important for musicians and art collectors to know how much confidence they can place in the dendrochronological authentication of fine violins. Dendrochronological dating does not provide an absolute date for the creation of a work of art. Indeed, dendrochronology cannot provide information about how many years passed between the time the tree was felled and when it was crafted into an instrument, nor can it indicate how much of the outermost wood was lost when the wood was crafted. Wood may also be stored and used later. Dendrochronological dating can only identify the date before which the instrument was certainly not made. Furthermore, the utility of dendrochronological methods for dating instruments is limited by the availability of reference ring-width chronologies, which are lacking for many regions. Few chronologies go back to 1500 or 1400. Future studies will hopefully expand on both the quantity and geographic and temporal coverage of such chronologies. Both living trees and the wood of stringed instruments can contribute to this effort (*10*).

Dendrochronology is not the only technique that can be deployed in dating stringed instruments. Although seldom used on musical instruments, wood densitometry has long been superior to ring-width measurements for detecting climatic signals in high-elevation conifers. Today, instruments can be x-rayed or analyzed by newly developed image analysis techniques, such as computer tomography with neutron imaging, or by synchrotron facilities (*13*). Other techniques can be used to assess and confirm wood sources. For example, the variation in tree genetics is useful for assessing population diversity at the regional scale. This approach is promising for establishing the wood source (*14*). Radiocarbon enables wood dating, and stable isotopes offer a promising means of determining wood sources (*15*). Although such methods require some wood for analysis and are therefore destructive, collection curators and luthiers should keep in mind that if some wood pieces do become available—for example, during restoration efforts—such methods may provide precious information.

Dendrochronology enables the objective verification of date attributions made on the basis of art history and stylistic criteria and offers a distinctive nondestructive, scientifically sound analytical technique when correctly applied. Larger datasets of reference ring-width chronologies with open-source availability will facilitate more accurate dendrochronological dating and determination of wood sources. ■

REFERENCES AND NOTES

1. C. Fritz, J. Curtin, J. Poitevineau, F.-C. Tao, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5395 (2017).
2. H. D. Grissino-Mayer *et al.*, *Dendrochronologia* **28**, 149 (2010).
3. V. Bucur, *Handbook of Materials for String Musical Instruments* (Springer, 2017).
4. W. E. Hill, *"The Salabue Stradivari": A History of the Famous Violin Known as "Le Messie"* (W. E. Hill and Sons, 1891).
5. H.-C. Tai, Y.-P. Shen, J.-H. Lin, D.-T. Chung, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 5926 (2018).
6. F. W. M. R. Schwarze, M. Spycher, S. Fink, *New Phytol.* **179**, 1095 (2008).
7. L. Burckle, H. D. Grissino-Mayer, *Dendrochronologia* **21**, 41 (2003).
8. E. Corona, *Ital. For. Mont.* **35**, 112 (1980).
9. P. Klein, H. Mehringer, J. Bauch, *Holzforschung* **40**, 197 (1986).
10. M. Bernabei, J. Bontadi, K. Čufar, A. Baici, *J. Cult. Herit.* **27S**, S55 (2017).
11. A. Lauw, M. Beuting, H. Pereira, *J. Cult. Herit.* **48**, 161 (2021).
12. S. Pollens, *J. Violin Soc. Am.* **16**, 77 (1999).
13. C. Invernizzi *et al.*, *Coatings* **11**, 29 (2021).
14. S. Jansen, H. Konrad, T. Geburek, *J. Environ. Manage.* **233**, 308 (2019).
15. P. Cherubini, G. Battipaglia, J. L. Innes, *Curr. For. Rep.* **7**, 69 (2021).

ACKNOWLEDGMENTS

I thank B. Carlson, W. Talirz, and M. H. Wiener for guidance on luthiery, music, and art for this piece.

10.1126/science.abj3823

QUANTUM ENGINEERING

Photons and qubits get a better connection

Quantum networks will require versatile quantum interconnects

By **Adam M. Kaufman**

A central theme in quantum science and engineering is that each technology, ranging from trapped ions and neutral atoms to superconducting qubits and photonics, has its strengths and challenges. A trapped-ion system might provide high-quality gate operations for quantum computations, but scaling to a large qubit number can be more difficult. Conversely, neutral atoms and superconducting qubits can be scaled to in the range of 100 or more controllable qubits, but their two-qubit gate fidelities are only now beginning to catch up to those of trapped ions. Meanwhile, photons might exhibit exquisite environmental isolation for preserving and transmitting quantum information, but this isolation is afforded by a lack of interactions that also precludes straightforward two-qubit gate operations. One compelling path to realize desired features for quantum engineering leverages and expands the strengths of different physical implementations by coherently connecting them. Quantum networks of devices are created by such quantum interconnects, which convert a quantum state represented in one set of physical qubits into another set of physical qubits. On page 1511 of this issue, Đorđević *et al.* (*1*) report a step forward in realizing a quantum interconnection between neutral-atom qubits and optical photons.

For quantum computing, quantum networks have been proposed as a path for expanding the set of qubits in a quantum processor: Coherently connecting many smaller quantum processors makes it possible to circumvent the qubit scaling limitations of a single device. In quantum communication, photons are a natural choice for rapid transmission of quantum-secure

JILA/Department of Physics, University of Colorado, Boulder, CO 80309, USA. Email: adammicahkaufman@gmail.com

information through quantum key distribution across a network. However, photon losses in fiber over long distances necessitate quantum repeaters, which require two-qubit gate sequences (*2*) that are potentially more easily implemented on physical devices such as solid-state spins, atoms, or ions. In quantum metrology, a distributed network of entangled clocks can be used for a new shared time scale reference and applications such as geodesy but requires entangling protocols between stationary (atoms or ions) and flying qubits (photons) (*3*).

These possibilities have motivated a diverse effort in realizing the technology for quantum networks (*4*). Pioneering efforts demonstrated state transfer between an atomic ensemble and a single photon (*5*); single atomic qubits strongly coupled to optical cavities have been connected with an optical photonic link tens of meters long (*6*). Superconducting qubits have coherently transmitted and received microwave photons, between chips and also between distinct dilution refrigerators (*7*–*9*). In trapped-ion and electron spin systems, information transfer between distinct regions has been realized by means of heralded photon methods (*10*, *11*). Progress has also been made in coherent transduction, in anticipation of the need to transduce photons of vastly different frequency to, for instance, coherently link qubit technologies with different energy splittings.

Đorđević *et al.* used tightly focused beams of light to trap individual single ^{87}Rb atoms, which serve as versatile qubits. The number, geometry, and position of the tweezers can be tuned in real time, capabilities that have been leveraged for a number of directions in quantum science. The authors used two traps to isolate and spatially manipulate two atoms, to precisely position them near a nanophotonic photonic crystal cavity (PCC) (see the figure). The PCC confines an optical mode with a tiny mode area of $<\lambda^2$ that leaks outside of the PCC structure. Light can be guided in and out of the PCC with an optical fiber. The microscale properties of the PCC alleviate some technical constraints of macroscopic cavities and are potentially more scalable.

The small PCC mode volume results in a high atom-photon cooperativity when a tweezer-trapped atom is overlapped with the optical mode that leaks out of the PCC. This means that the time scale for coherent exchange of energy between the atom and the cavity substantially exceeds the time scale of uncontrolled decay paths, such as atomic decay and photon leakage out of the PCC. For a single atom near the PCC, the state of a single-photon incident on the PCC can be phase switched and entangled with the state of the atom, allowing for fast optical readout or atom-photon entanglement (*12*). For more than one atom near the cavity, entangled states can be created by using PCC photons to mediate interactions between the atoms or to perform collective measurements on the qubits.

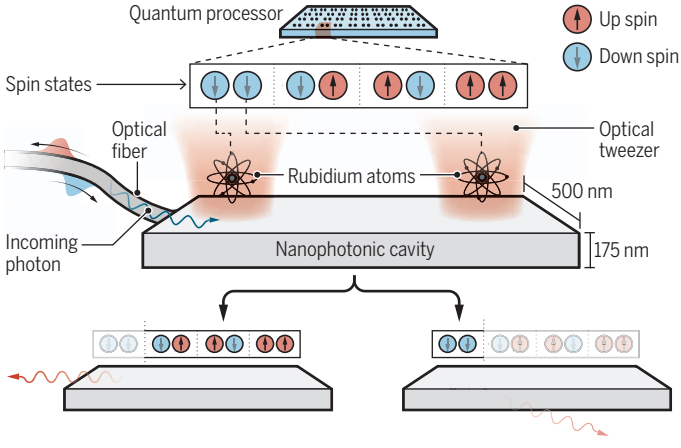
With this toolset, Đorđević *et al.* move toward realizing a fully capable network node between neutral atoms and photons. Using two hyperfine states within the atoms, the authors encode a qubit that can be manipulated with microwave radiation. A target capability for this system is processing quantum information away from the PCC (see the figure) and then transmitting it in flying photonic qubits. The mobility of the optical tweezers allows transport of the qubits into and out of the PCC mode, but the PCC mode imparts large frequency shifts on the qubit, which can imperil the qubit coherence. The authors carefully characterized these shifts and developed a pulse sequence to preserve coherence over the transport, allowing the atomic quantum states to be transported into and out of the PCC mode.

With coherent transport under control, not just single qubits but multiparticle entangled states can also be transported. The authors create two-particle entangled states using a scheme known as "cavity carving" (*13*). The PCC is brought near resonance with a transition between one of the qubit states (the ↑ state) and an optically excited state. When one or both of the atoms is in the ↑ state, the atom-cavity system hybridizes so that the new eigenenergies are distinct from the harmonic energy spectrum of the cavity alone. Consequently, when a photon impinges on the PCC, the PCC properties can be arranged so that detection of a reflected photon removes any quantum amplitude in the ↓↓ state, while maintaining quantum coherence in the remaining two-particle state. An initial two-particle product state can then be projected, or "carved," into an entangled state through this photon detection. The authors performed this protocol and then characterized the resulting two-particle state near the PCC and far away from the PCC structure after transporting it. Their measurements prove the existence of entanglement before and after transport.

With improved fidelities, the interaction between the atoms and the PCC can be used to engineer multiparticle quantum states and protocols that are useful for quantum information processing, such as cluster states and error correction. Using an alkaline-earth atom could allow the PCC to generate large-scale entanglement useful for quantum-enhanced metrology or entangled clock networks (*3*). An alkaline-earth or alkali atom could be used for an atom-based quantum interconnect at a telecom wavelength (*14*, *15*). More broadly, the capabilities set the stage for connecting large-scale neutral atom arrays by means of photonic buses, for improved scaling and distributed entanglement. ■

A quantum interconnection

A pair of optical tweezers allows rubidium atoms to be moved near a nanophotonic cavity. The rubidium atoms each represent qubits through their internal spin. When a photon guided to the cavity is reflected, it indicates that at least one of the qubits is in the "up" state, but not which qubit is in that state. This can be used as a resource to entangle the qubits by carving out a component of the two-particle quantum state.



GRAPHIC: C. BICKEL/*SCIENCE*

REFERENCES AND NOTES

1. T. Đorđević *et al.*, *Science* **373**, 1511 (2021).
2. S. Muralidharan *et al.*, *Sci. Rep.* **6**, 20463 (2016).
3. P. Kómár *et al.*, *Nat. Phys.* **10**, 582 (2014).
4. D. Awschalom *et al.*, *PRX Quantum* **2**, 017002 (2021).
5. D. N. Matsukevich, A. Kuzmuch, *Science* **306**, 663 (2004).
6. S. Ritter *et al.*, *Nature* **484**, 195 (2012).
7. C. J. Axline *et al.*, *Nat. Phys.* **14**, 705 (2018).
8. P. Kurpiers *et al.*, *Nature* **558**, 264 (2018).
9. P. Magnard *et al.*, *Physical Review Letters* **125**, 260502 (2020).
10. D. Hucul *et al.*, *Nat. Phys.* **11**, 37 (2015).
11. H. Bernien *et al.*, *Nature* **497**, 86 (2013).
12. T. G. Tiecke *et al.*, *Nature* **508**, 241 (2014).
13. S. Welte *et al.*, *Phys. Rev. Lett.* **118**, 210503 (2017).
14. J. P. Covey *et al.*, *Physical Review Applied* **11**, 034044 (2019).
15. T. Chanelière *et al.*, *Physical Review Letters* **96**, 093604 (2006).

10.1126/science.abk2617

SUPERCONDUCTIVITY

Strange bedfellows inside a superconductor

An unexpected link is found between the "strange metal" phase and charge density waves

By **Matthieu Le Tacon**

The discovery of superconducting cuprates in 1986 is considered a watershed moment in the study of superconductivity—not only because of their high superconducting temperatures (T_c's) but also on account of their highly exotic properties, which are still largely enigmatic (*1*). On page 1506 of this issue, Wahlberg *et al.* (*2*) bring insights into the intriguing physics of cuprates' nonsuperconducting state by connecting two widely studied phenomena previously believed to be completely independent of each other: the linear resistivity of the strange metallic phase and charge density waves (CDWs).

Although it may appear mundane at first sight, the linear temperature dependence of electrical resistivity in a high-temperature superconductor is no trivial matter and illustrates better than anything else how peculiar these materials can be outside of their superconducting state. Before diving into the details of the new findings, some background is necessary to understand why their observation may come as a surprise.

At very low temperatures, a few kelvin above absolute zero, a normal metal should display resistivity that is mostly constant, before transitioning into a quadratic temperature dependence. This is one of the most emblematic signatures of a Landau Fermi liquid, which is considered the standard model for metals. This model introduces the notion of "quasiparticles," which behave like bare electrons with a higher mass, a reduced lifetime and coherent spectral weight, and whose mutual scattering results in the quadratic temperature dependence. At even higher temperatures, beyond a characteristic Debye temperature θ_D, thermally activated phonons scatter off the quasiparticles, which leads to a linear increase of the metal's resistivity with temperature. However, this linear relationship does not hold indefinitely. The resistivity of metals saturates beyond the Mott-Ioffe-Regel (MIR) limit as the electronic mean free path approaches the de Broglie wavelength of the quasiparticle (*3*).

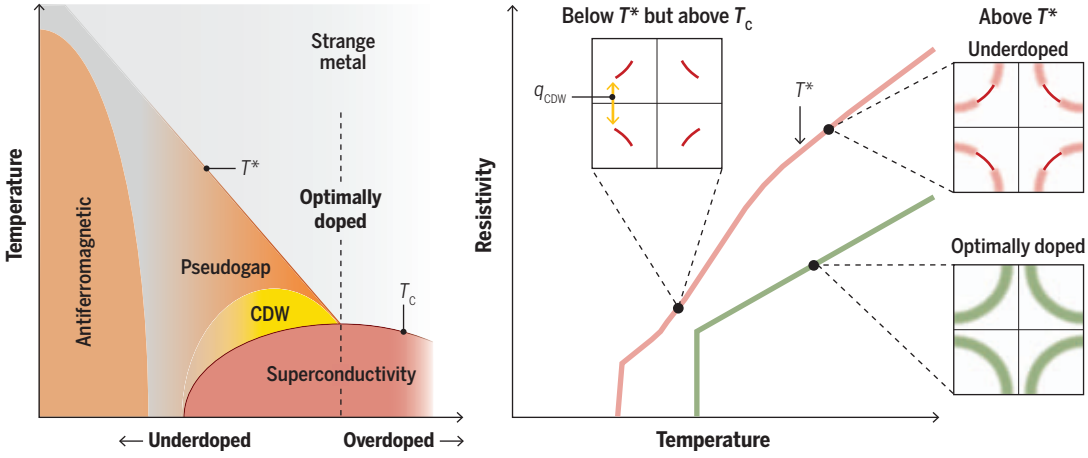
Perhaps counterintuitively, above their superconducting temperature T_c, the optimally doped cuprates (see the figure), where T_c is maximized, are poor metals, with a relatively high resistivity. The resistivity can further increase linearly well beyond the MIR limit, with temperatures as high as 1000 K, without showing any sign of saturation. This bizarre property of cuprates is fundamentally incompatible with the Landau Fermi liquid model (*4*) and ultimately challenges the notion of quasiparticles. As a result, the metallic state of optimally doped cuprates is called a "strange metal" (*1*).

To gain an even fuller picture of the mysteries surrounding cuprates, one would have to consider the overdoped and underdoped cuprates, in addition to the optimally doped cuprates discussed thus far. On the underdoped side, however, coherent quasiparticles are only found in a limited region of the Brillouin zone—that is, at specific locations of the crystal lattice. The quasiparticles form "Fermi arcs" prolonged with incoherent spectral weight up to the edge of the Brillouin zone, and upon cooling, below the pseudogap temperature T^* but still above T_c, this incoherent spectral weight appears to shrink away from the Fermi level. A deviation from the linear temperature dependence of the in-plane resistivity below T^* has been considered as a signature for the pseudogap opening (*5*). The results of Wahlberg *et al.* bring fresh insights to the debate on the interplay between the strange metallic phase, the pseudogap, and the recently discovered CDW.

It has indeed been shown that in the pseudogap regime, there exists a ubiquitous tendency for incommensurate CDWs to form (*6*). These are quite different from those existing in low-dimensional metallic systems and strongly compete with superconductivity. (*7*). A connection between the CDW and the Fermi arcs in the pseudogap regime has been suggested on the basis of the evolution with doping of the CDW ordering wave vector (q_{CDW}), which is found intriguingly close to the vector matching

A match on the Fermi surface

Shown below on the left is a schematic phase diagram of hole-doped cuprates. The diagram on the right shows the temperature dependence of the resistivity of underdoped and optimally doped cuprates, with insets showing the Fermi surfaces at different temperatures. The thin lines represent coherent quasiparticles, and the thicker lines represent the incoherent spectral weight at the Fermi level. For the underdoped cuprate below T^* but above the superconducting temperature T_c, the charge density wave (CDW) ordering wave vector q_{CDW} is found to match the gap between the tips of the Fermi arcs.



GRAPHIC: K. FRANKLIN/*SCIENCE*

Institute for Quantum Materials and Technologies, Karlsruhe Institute of Technology (KIT), 76131 Karlsruhe, Germany. Email matthieu.letacon@kit.edu

the tips of the arcs (*8*). However, the link between the pseudogap and the CDW has previously been downplayed owing to the difference in doping dependence of their respective onset temperatures.

Wahlberg *et al.* go against this assumption by suggesting that the departure from *T*-linear resistivity in underdoped cuprates should be seen as a fingerprint of the CDW. The origin of the CDW is unsettled, and its relevance to the cuprates' physics is still being debated. For instance, some may claim that a necessary condition for the CDW to form is to avoid the strange metal phase and stay below T^*, where the incoherent spectral weight at the Fermi level has been suppressed and only the Fermi arcs remain. Wahlberg *et al.* fuel this debate in an unexpected manner by showing that the suppression of the CDW can instead restore the linear resistivity in underdoped cuprates, and they go even further by suggesting that the suppression of the strange metallic phase below T^* is directly due to the CDW, and not the pseudogap opening.

"...the suppression of the charge density wave can instead restore the linear resistivity in underdoped cuprates..."

Most notably, this connection between the CDW and the strange metal phase has been achieved through the manipulation of the crystal lattice by means of a large epitaxial strain. This provides a strong argument that the lattice should be considered not as a mere spectator but rather as a legitimate and insightful tuning parameter for the high-temperature superconductor (*9*). This result warrants more systematic investigations of the interplay between lattice and electronic degrees of freedom in the cuprates but also more generally in other quantum materials. ■

REFERENCES AND NOTES

1. B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, J. Zaanen, *Nature* **518**, 179 (2015).
2. E. Wahlberg *et al.*, *Science* **373**, 1506 (2021).
3. N. E. Hussey, K. Takenaka, H. Takagi, *Philos. Mag.* **84**, 2847 (2004).
4. C. M. Varma, *Rev. Mod. Phys.* **92**, 031001 (2020).
5. T. Timusk, B. W. Statt, *Rep. Prog. Phys.* **62**, 61 (1999).
6. A. Frano, S. Blanco-Canosa, B. Keimer, R. J. Birgeneau, *J. Phys. Condens. Matter* **32**, 374005 (2019).
7. G. Ghiringhelli *et al.*, *Science* **337**, 821 (2012).
8. R. Comin *et al.*, *Science* **343**, 390 (2014).
9. H. H. Kim *et al.*, *Science* **362**, 1040 (2018).

ACKNOWLEDGMENTS

I thank M. Frachet, R. Heid, and I. Vinograd for critical reading of this Perspective.

10.1126/science.abi9685

REGENERATIVE MEDICINE

Reversing aging for heart repair

Short-term induction of pluripotency gene expression allows adult mouse cardiomyocytes to regenerate

By **Yu Xin Wang** *and* **Helen M. Blau**

Partial reprogramming (PR) is an innovative and promising approach for reversing cellular age through short-term induction of pluripotency factors that rejuvenate aged cells to regain functions of young ones. PR takes advantage of the stepwise nature of the reprogramming process (*1, 2*) by using a controlled short-burst expression of factors that remodel the epigenome and rejuvenate gene expression but do not go as far as erasing cellular identity (*3*). Applications of PR in age-related diseases and models of premature aging suggest that a controlled regimen can reverse the timeworn molecular signature of aging in a range of cell types (*3–5*). Unlike many adult tissues, the heart lacks stem cells and is incapable of regeneration. On page 1537 of this issue, Chen *et al.* (*6*) show that PR can induce repair of the heart after a myocardial infarction by temporarily reactivating a neonatal developmental program in mature cardiomyocytes in mice.

Regenerative potential of tissues is generally maintained by tissue-resident stem cells or progenitors, which are usually committed to the specific cell type they are destined to replace. Despite the importance of the heart in pumping blood to provide oxygen and nutrients to the rest of the body, the heart loses its potential to repair tissue damage shortly after birth (*7*). This is because of the lack of cells capable of generating cardiomyocytes in the myocardium postinjury—e.g., after a heart attack. Instead, mesenchymal stem cells in the heart form a fibrotic scar at the site of damage. Therapies aimed at replacing or forming new cardiomyocytes are highly sought after for restoring cardiac function but are currently lacking. Thus, transient induction of the expression of the pluripotency factor genes *Oct4* (octamer-binding protein 4), *Sox2* (SRY-box 2), *Klf4* (Kruppel-like factor 4), and *Myc* (collectively known as OSKM) to coerce cardiomyocytes to reenter the cell cycle, divide, and form new cardiomyocytes in mice is an important advancement.

Baxter Laboratory for Stem Cell Biology, Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA. Email: hblau@stanford.edu

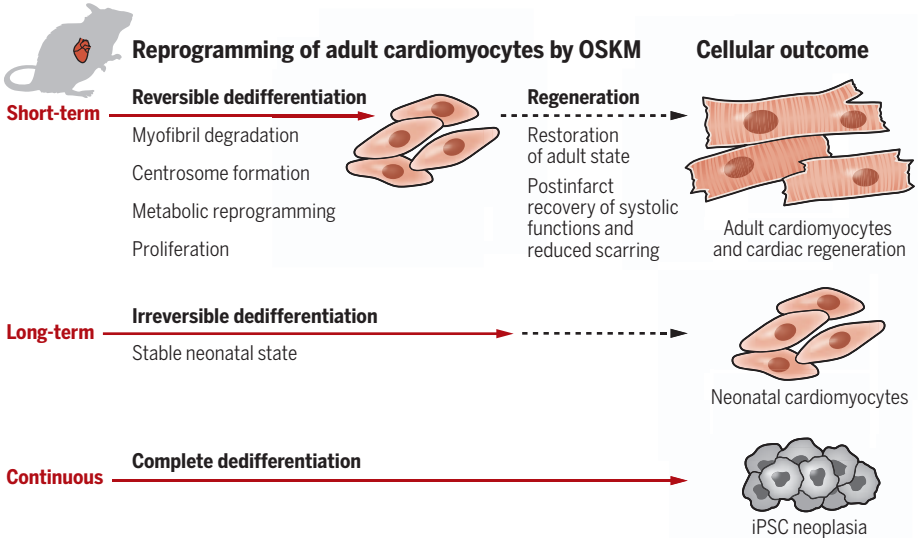
Notably, Chen *et al.* show that inducible expression of OSKM in transgenic mice before or immediately after a surgical arterial ligation–induced myocardial infarct facilitates partial regeneration of the myocardium, reduces scar size, and promotes recovery of left-ventricular ejection fraction, which demonstrates the potential therapeutic relevance of this approach. Moreover, the authors show that OSKM expression in adult cardiomyocytes elicits a neonatal transcriptional program that enables them to dedifferentiate and divide without the loss of cell identity.

A natural dedifferentiation process occurs in newts during skeletal and cardiac muscle tissue regeneration. This entails inactivation of the tumor suppressor retinoblastoma protein (RB) in terminally differentiated cells, which restores the ability to proliferate (*8*). In mammalian cells that have evolved additional tumor suppressors, transient down-regulation of RB as well as the cell cycle inhibitor p19 alternative reading frame (ARF) are required to trigger mitosis in differentiated skeletal myocytes and cardiomyocytes in vitro (*8, 9*). Chen *et al.* show that the proto-oncogene *Myc* is a necessary component of the PR regimen. MYC plays a major role in overcoming RB during OSKM reprogramming to pluripotency (*10*) and metabolic reprogramming of cardiomyocytes under stress (*11*). However, neither *Myc* nor OSK overexpression alone was able to trigger mitosis of cardiomyocytes in vitro, which suggests that MYC requires OSK to remove additional barriers to cell cycle reentry.

One such barrier is the physical constraint of myofibrillar proteins that comprise the cardiomyocyte cytoskeleton. The structural proteins of striated muscles are organized for contraction and mechanotransduction. The centrosome, which is necessary for mitosis, is disassembled in mature skeletal muscle cells and cardiomyocytes. Agrin, a component of the neonatal cardiac extracellular matrix, can stimulate centrosome reassembly, reduce myofibrillar proteins, and enable mature cardiomyocyte division in mice (*12*). Chen *et al.* also found that after OSKM expression, rod-shaped adult cardiomyocytes express the centrosome regulator, Aurora kinase B, and undergo extensive morphological transformation before mitosis and cytokinesis.

Reprogramming of adult cardiomyocytes in mice

The timing of in vivo reprogramming by *Oct4* (octamer-binding protein 4), *Sox2* (SRY-box 2), *Klf4* (Kruppel-like factor 4), and *Myc* (collectively, OSKM) in adult mouse cardiomyocytes is key to the outcome. Prolonged OSKM expression in vivo led to neonatal cardiomyocytes that could not mature or to neoplasias formed from induced pluripotent stem cells (iPSCs) within the heart, whereas short-term induction enabled adult cardiomyocytes to regenerate the heart.



The potency of PR results from the function of OCT4, SOX2, and KLF4 as pioneer factors that can open condensed chromatin (*13*), which activates the transcription of a common set of genes in various cell types. Although PR improves regeneration of aged tissues such as skeletal muscle, the effects have been attributed to improved stem cell function (*3*). Chen *et al.* provide evidence that PR can cause reversible dedifferentiation to a neonatal state, leading to functional recovery in a tissue lacking stem cells. Whether PR can transiently reactivate other postmitotic cell types is unknown. For example, can neurons be dedifferentiated to repair the spinal cord after traumatic injury? PR induced the formation of additional pancreatic islets in a mouse model of diabetes (*3*). Although it is unclear whether this was a result of dedifferentiation, this suggests that other pathologies could benefit from PR.

Chen *et al.* also show that caution is needed for the therapeutic window of OSKM-mediated PR. They found that dosage and duration of the treatment are both critical factors in achieving functional recovery. With long-term treatments, cardiomyocytes either reach an irreversible neonatal state that is unable to sustain function or reach a pluripotent state that forms neoplasms even in the absence of injury, leading to premature death (see the figure). These findings contrast with in vivo reprogramming in skeletal muscle, where interleukin-6 signaling is necessary to facilitate induced pluripotent stem cell formation after injury (*14*, *15*). This is an important reminder that the OSKM genes are potent oncogenic factors that can erase cellular identity in the process of reprogramming, and tight control of their timing and amount is necessary to achieve a therapeutic effect.

There is much to explore in the application of OSKM as a therapy. A gene therapy approach requiring targeting and tight dosage control may be technically challenging for translation to humans. Identification of molecular thresholds that promote dedifferentiation but avoid a point of no return will be critical to engineering a safe therapy. To reduce the potential for neoplastic transformation it would be desirable to replace the OSKM proto-oncogenes, by using RNA-mediated methods or small molecules. Regardless of these concerns, the finding that the regenerative potential of adult tissues can be reactivated has exciting implications for regenerative medicine. ■

REFERENCES AND NOTES

1. J. M. Polo *et al.*, *Cell* **151**, 1617 (2012).
2. S. Yamanaka, H. M. Blau, *Nature* **465**, 704 (2010).
3. A. Ocampo *et al.*, *Cell* **167**, 1719 (2016).
4. T. J. Sarkar *et al.*, *Nat. Commun.* **11**, 1545 (2020).
5. Y. Lu *et al.*, *Nature* **588**, 124 (2020).
6. Y. Chen *et al.*, *Science* **373**, 1537 (2021).
7. E. R. Porrello *et al.*, *Science* **331**, 1078 (2011).
8. H. M. Blau, J. H. Pomerantz, *JAMA* **305**, 87 (2011).
9. K. E. Hatzistergos *et al.*, *Circ. Res.* **124**, 1184 (2019).
10. T. A. Rand *et al.*, *Cell Rep.* **23**, 361 (2018).
11. P. Ahuja *et al.*, *J. Clin. Invest.* **120**, 1494 (2010).
12. E. Bassat *et al.*, *Nature* **547**, 179 (2017).
13. A. Soufi *et al.*, *Cell* **161**, 555 (2015).
14. J. J. Brady *et al.*, *Nat. Cell Biol.* **15**, 1244 (2013).
15. A. Chiche *et al.*, *Cell Stem Cell* **20**, 407 (2017).

10.1126/science.abl8679

GENOMICS

Missing heritability may be hiding in repeats

Unexplored variable number tandem repeats make a large contribution to complex traits

By **Melissa Gymrek**[1,2] *and* **Alon Goren**[1]

Genome-wide association studies (GWASs) have identified tens of thousands of genomic loci associated with complex traits in humans (*1*). However, translating GWAS signals into predictive models and mechanistic insights remains challenging (*2*). Most standard GWAS analysis pipelines focus on common single-nucleotide polymorphisms (SNPs) and do not capture the full range of human genetic variation. Variable number tandem repeats (VNTRs) are a major source of polymorphism whose contribution to the inheritance of complex traits has been difficult to study owing to the analysis challenges they present. On page 1499 of this issue, Mukamel *et al.* (*3*) describe a set of methods for analyzing VNTR copy number in large cohorts, and demonstrate that polymorphic protein-coding VNTRs exhibit some of the strongest effect sizes yet identified for multiple traits, including height and health biomarkers.

A typical GWAS tests for association between common SNPs and a trait of interest. Significantly associated SNPs help pinpoint the regions of the genome that are important for that trait. For most complex traits, there may be tens to hundreds of such signals, each of which explains only a modest amount of trait variability (effect size). Despite many success stories (*4*), common SNPs identified by GWAS still fail to fully explain the heritability of most traits analyzed, even as sample sizes have reached hundreds of thousands of people, a problem dubbed the "missing heritability" (*5*).

Further, uncovering mechanistic biological insights from GWAS signals has proven

[1]Department of Medicine, University of California San Diego, La Jolla, CA, USA. [2]Department of Computer Science and Engineering, University of California San Diego, La Jolla, CA, USA. Email: mgymrek@ucsd.edu; agoren@ucsd.edu

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

exceedingly challenging. A major hurdle is that the lead SNPs identified by GWAS are rarely themselves causal (*6*). Rather, trait-associated SNPs are often simply correlated with (tagging) an underlying causal variant. Hiding within each association signal is a biological story: One or more genetic variants in the population result in a cascade of molecular differences that eventually manifest as variation in a complex trait. Decoding such stories is likely to improve both our understanding of the biological mechanisms of complex traits and our ability to predict phenotypes and disease risk from genotypes. But such mechanistic insights are rarely immediately clear.

Indeed, detailed analyses of individual GWAS signals so far suggest that the underlying story is often more complex than a single causal common SNP. For instance, a multiallelic copy number variant (CNV) drives much of the strongest identified GWAS signal for schizophrenia (*7*), additional signals for psychiatric traits appear to be at least partially driven by VNTRs (*8*, *9*), and an array of CNVs was identified as underlying a GWAS signal for malaria resistance (*10*).

Variants in repetitive regions of the genome have been noticeably missing from most GWASs and other association studies, such as expression quantitative trait loci (eQTL) analyses. Mukamel *et al*. focused on one type of repeat, VNTRs, which consist of sequences of seven or more base pairs, repeated in tandem from several to hundreds of times. As a result of frequent mutations that add or delete copies of the repeat, VNTRs may have multiple distinct common alleles in a population (unlike SNPs, which typically have only two alleles). Although most VNTRs are in noncoding regions, those overlapping protein-coding genes can have profound, and more interpretable, effects by modulating the size and/or copy number of domains within a protein.

However, VNTRs and other repeats present considerable challenges to existing pipelines for association studies. Many VNTRs are longer than next-generation sequencing (NGS) read lengths and are missed by most NGS variant–calling pipelines, which typically focus on SNPs and small indels (insertions and deletions). Further, most large cohorts used for GWAS are only genotyped by SNP arrays, which do not allow for direct analysis of VNTRs. In addition, because VNTRs and other repeat types are often multiallelic, many are only partially tagged by individual common SNPs and are not automatically handled by existing models or software that assume biallelic genotypes.

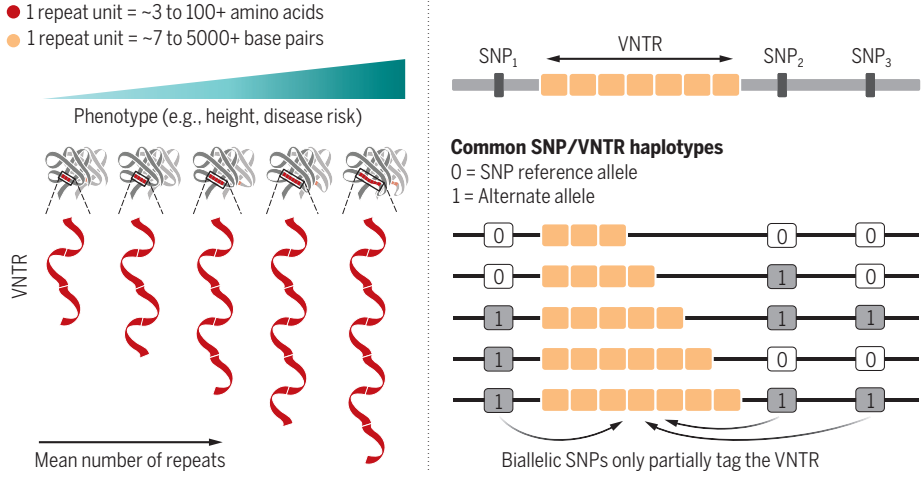
Mukamel *et al*. developed a suite of methods for VNTR analysis to enable accurate imputation of VNTR lengths from SNP array datasets. They applied this pipeline to analyze 118 protein-coding VNTRs in more than 400,000 UK Biobank participants and tested for association with more than 700 traits. They identified five distinct VNTRs that appear to causally drive some of the strongest known GWAS signals. In several cases, the VNTRs show an increasing or decreasing signal of association between the number of repeat units and the trait value across a series of multiple distinct alleles, strongly supporting the VNTR as the causal variant. These multiallelic signals also highlight the potential of repeats to elegantly help explain continuous phenotypes such as height (see the figure). Findings at two of these VNTRs confirm previous observations—a VNTR in the lipoprotein A (*LPA*) gene was associated with LPA concentration (*11*) and another in aggrecan (*ACAN*) was linked to height (*12*). The case of *LPA* is a compelling example of the sometimes complex nature of these association signals: A detailed model incorporating nonlinear effects of VNTR length and interaction of the VNTR with nearby SNPs explains substantially more phenotypic variance compared to models based on SNPs or the VNTR alone.

Intriguingly, though revealing these VNTR associations required innovative computational methods, the authors did not need to generate any new data to arrive at their findings. The information needed to identify these VNTR associations was hiding within existing DNA sequencing and SNP array datasets recently generated as part of the UK Biobank. The VNTR associations identified by Mukamel *et al*. likely represent just the tip of the iceberg of repetitive variants underlying complex traits. Although their study focused on 118 protein-coding VNTRs, hundreds of thousands of polymorphic repeats exist throughout the genome. These can be either coding or noncoding and include VNTRs, microsatellites (short tandem repeats), and CNVs, all of which are still largely missing from the GWAS repertoire. Many of these variants can likely be explored by using existing datasets through development of additional analysis techniques.

Systematic incorporation of these variants into GWAS is an important future goal. It remains to be seen how many GWAS signals are driven by such variants, how much additional heritability will be explained by incorporating them, and whether including these variants will improve our ability to construct informative polygenic risk scores to predict disease susceptibility. Integration of repeats and other complex variants will also present new opportunities to learn about biological mechanisms driving a range of complex traits. ■

Tandem repeats as drivers of complex traits

Variation in repeat copy number at multiple protein-coding variable number tandem repeats (VNTRs) is associated with complex traits such as height, across a range of alleles. These repeats are often multiallelic, and their contribution to a trait may be only partially captured by tagging with biallelic single-nucleotide polymorphisms (SNPs) typically considered in genome-wide association studies (GWASs).



GRAPHIC: V. ALTOUNIAN/*SCIENCE* BASED ON RAHEL WACHS

REFERENCES AND NOTES

1. A. Buniello *et al*., *Nucleic Acids Res.* **47**, D1005 (2019).
2. A. Torkamani *et al*., *Nat. Rev. Genet.* **19**, 581 (2018).
3. R. E. Mukamel *et al*., *Science* **373**, 1499 (2021).
4. P. M. Visscher *et al*., *Am. J. Hum. Genet.* **101**, 5 (2017).
5. B. Maher, *Nature* **456**, 18 (2008).
6. M. van de Bunt *et al*., *PLOS Genet.* **11**, e1005535 (2015).
7. A. Sekar *et al*., *Nature* **530**, 177 (2016).
8. J. H. T. Song *et al*., *Am. J. Hum. Genet.* **103**, 421 (2018).
9. M. Li *et al*., *Nat. Med.* **22**, 649 (2016).
10. E. M. Leffler *et al*., *Science* **356**, eaam6393 (2017).
11. R. Clarke *et al*., *N. Engl. J. Med.* **361**, 2518 (2009).
12. D. Beyter *et al*., *Nat. Genet.* **53**, 779 (2021).

ACKNOWLEDGMENTS

We thank J. Margoliash and A. Massarat for critical comments and acknowledge support from NIH grants DP5OD024577 and R01HG010885.

10.1126/science.abl7794

LANDMARK: GENOMICS

The legacy of the Human Genome Project

The initiative forever altered biomedicine, but work remains to fulfill its true potential

By **Jennifer E. Rood** and **Aviv Regev**

Twenty years after the publication of its first draft sequence, the Human Genome Project (HGP) has had a profound effect on biomedical research—at the bench, on the computer, and across the scientific community. Large datasets and reference maps are now valued resources to guide scientific and clinical research, drug development, and medical practice. The HGP has also left us with a major mission—still incomplete 20 years later—to understand how genomic information leads to the development, function, and malfunction of cells and organisms and to fully leverage this knowledge to promote human health and treat disease.

In the 1980s, scientists used ever-finer genetic mapping tools to pinpoint the genes underlying Mendelian disorders such as cystic fibrosis and Huntington's disease. But to fully understand the genetic basis of common human diseases, where multiple genes are at play, these genomic marker–based mapping techniques were insufficient. Instead, theoretical analysis showed that an infinitely dense physical map would be needed—that is, the full sequence of the human genome. This realization converged with new advances in sequencing technology and other motivations, such as identifying cancer-driving mutations, to propel the 1990 launch of a concerted initiative to sequence the full genome—the HGP—delivering a first draft in 2000 and a "finished" reference sequence in 2003 (*1*, *2*).

Many initially doubted the utility of undertaking this costly effort, claiming that it could undermine attempts to understand the function of individual genes while not directly revealing the disease biology. Yet one of the first direct outcomes of the reference genome was to enable studies of human genetic variation, including in disease, initially through genome-wide association studies (GWASs) that have since identified tens of thousands of loci for specific traits and diseases (www.ebi.ac.uk/gwas). As modern genetics predicted, GWASs mostly find genes with small effect sizes, but these are nonetheless essential clues into causal disease mechanisms that may otherwise remain inaccessible (*3*).

Genentech, 1 DNA Way, South San Francisco, CA, USA.
Email: aviv.regev.sc@gmail.com

The human reference genome also enabled other crucial efforts to decipher disease biology, such as cancer genome analyses to map somatic genetic drivers in tumors, mapping of rare disease genes, and phenome-wide association studies leveraging large-scale biobanks to examine variants linked to multiple traits and diseases in an unselected population (see the figure). The reference further allowed for comparative genomics to annotate functional elements, and ancient DNA techniques to track human migrations and trait development, including susceptibility to disease.

A reference human genome sequence has itself proved a critical research tool. Basic lab experiments today are unimaginable without it: from designing primers for molecular cloning to genetic manipulation by small interfering RNA (siRNA) or CRISPR-Cas9–based guide RNAs. Both the HGP itself and the availability of the reference drove the development of powerful new technologies for rapid and cheap sequencing. Next-generation sequencing is the microscope of the 21st century, enabling an array of powerful sequencing-based functional genomics methods, such as profiling gene expression in individual cells and characterizing the spatial context of cells and tissues. These methods, reliant on a reference genome, play an increasingly important role in fulfilling the HGP's promise of elucidating gene function and enabling individual labs to make key contributions that previously would have required large-scale efforts (*4*).

Building the reference also ushered in the growth of modern computational biology and an appreciation of the power of computational thinking to solve biological problems by drawing on large datasets that previously did not exist. The scale of the data generated during genome projects necessitated new data processing and analysis approaches, drawing a generation of computer scientists into biological problems and driving the development of assembly algorithms (*5*). As the quantity of genomic data has grown rapidly in the ensuing decades, biology has drawn on mathematical, statistical, and computational advances to expand the analyses of these and other data, most recently with deep learning models.

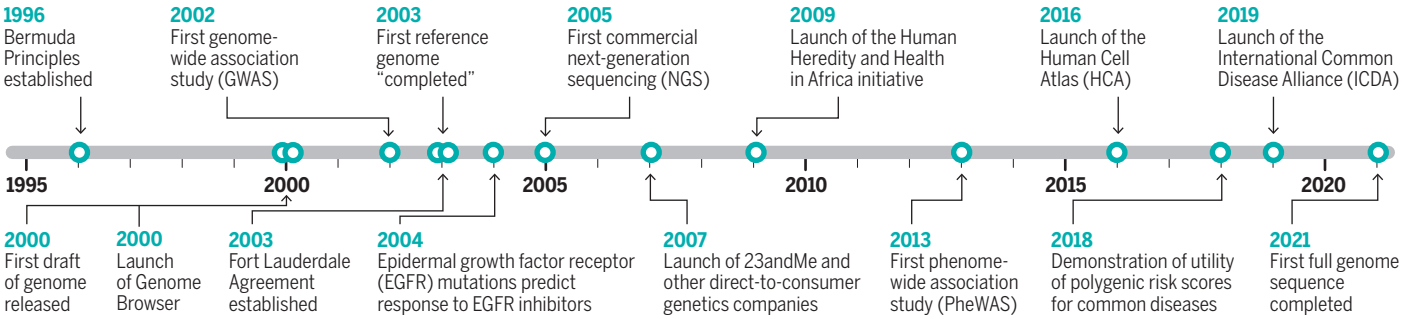
A further computational challenge for the HGP was making the data accessible to researchers, both in its raw form—while maintaining security and confidentiality of personal genomic information—and in a manner that allowed effective querying and visualization. The need for data portals, a relatively late realization in the HGP, drove the creation of some of the first large-scale biological databases and innovative platforms to access, analyze, search, and visualize the data, in particular the University of California, Santa Cruz, "Genome Browser" (*6*). The Genome Browser's "tracks" concept and sustained software engineering set the bar for future endeavors and are among its enduring legacies, both as an essential research resource and as inspiration for other genomic data platforms.

The HGP rapidly changed a climate of mostly small-scale collaboration and limited prepublication data sharing, by demonstrating the value of big data and the reference maps and by pioneering the principles necessary to undertake the creation and operation of a large international consortium. The Bermuda Principles of 1996, and their expansion in 2003 as the Fort Lauderdale Agreement, have served as important guidance around biological data sharing and credit attribution ever since (*7*). The HGP became a blueprint for many consortia that openly share data to accelerate science—such as the 1000 Genomes Project, The Cancer Genome Atlas (TCGA), the International Cancer Genome Consortium, ENCODE, gnomAD, the BRAIN Initiative, and the Human Cell Atlas—each assembling reference maps for different challenges as a public good.

The HGP's ultimate promise was to improve human lives, moving from biological understanding of disease to diagnosis, treatment, or even prevention. The reference genome has already begun to make this impact, most prominently in rare neonatal disorders and cancer. For example, researchers can now more readily identify the specific mutations underpinning rare genetic disorders, and begin to develop mechanistic understanding and treatment options, while enabling families to plan future pregnancies and, through shared data resources, connect them to other similar patients (*8*). In cancer, genetic markers and profiles are leading to improved diagnosis through routine standard assays, precision treatments (*9*), and new ways to monitor disease progression and treatment response less invasively through liquid biopsies, and single-cell profiling is helping reveal tumor

Key outcomes of the Human Genome Project

The Human Genome Project has had a considerable effect on research and society more generally, but questions about what a human genome reference is today and how it can benefit human health remain to be resolved.



heterogeneity in unprecedented detail.

Moreover, human genetic studies of large populations are increasingly affecting disease treatment—gene targets chosen on the basis of human genetics are up to twice as likely to lead to successful drugs (*10*). It is also now possible to use data from large populations to assess risk at the individual level through the development and implementation of polygenic risk scores for several conditions, including prevalent complex diseases.

Despite the indisputable sea change wrought by the HGP in research biology and, more recently, in clinical practice, has it fully achieved its mission and delivered on its promise? The two core questions at the heart of the HGP remain open: What is a complete reference in biology, and what is its value? And how can a reference be translated to benefit human health? Answers are emerging to both questions, extending and reshaping the legacy of HGP.

Charting the reference genome was only the first step in a long road of unpacking its secrets to affect health. Although early HGP critics grossly miscalculated its enormous scientific value—especially for the study of individual gene function—they were correct that deciphering the functions of genes as they act in the human body, both individually and collectively, cannot be solved only with the human genome sequence. Instead, this requires newer approaches such as functional genomics, which generates and combs through large datasets to tease out the function of coding and noncoding sequences, characterize molecular phenotypes, and power functional screens (*11*) (see also www.icda.bio); single-cell biology, to understand the functional impact of genetic variation at the basic unit of human biology; and advances in computation, especially machine learning.

Additionally, the notion of a reference is challenging in biology—because evolution and physiology always lead to a tension between conserved or invariant features, and divergent or dynamic ones. The HGP initially focused on the least-divergent features, while embarking on mapping common (and then increasingly rare) variants in the context of the shared reference. [Of the conserved features, the final 8% of repetitive heterochromatin sequence was only delivered in 2021 (*12*).] Capturing the ever-growing genetic diversity of humans requires profiling a more diverse set of genomes while shifting from the static to the statistical, leveraging algorithms to predict the impact of unobserved combinations of alleles. Similar challenges are faced by current mapping efforts, such as healthy and disease cell atlases.

This ongoing tension poses a societal challenge. The HGP and its legacy must serve humanity as a whole, not neglecting those who are currently underrepresented in biological research, as research subjects, and as scientists. Ultimately, although highly useful, a single reference genome is inherently biased and may undermine this goal. This bias and lack of diversity are currently also more broadly reflected in human genetics studies, where the vast underrepresentation of people of non-European ancestry in existing genomics datasets limits our understanding of human genetic variation, health, and disease. It also undermines health equity because it could lead to diagnostic models or therapies targeted to disease variants that do not occur in most of the world (*13*). Notable efforts to address this situation are underway, such as the Human Heredity and Health in Africa (H3Africa) initiative. But for the HGP to achieve its ultimate goal of serving human health, more conscious effort is required to engage researchers and participants from historically underrepresented communities as equal partners in research, with control over their own data and opportunity to benefit directly from it.

With these concerns in mind, and with genome sequencing in broad use, a single reference may no longer make sense, but making each personal sequence its own reference introduces new privacy concerns, especially around open data sharing, which may require new privacy-preserving algorithms. Already, sequencing technology permeates many aspects of daily life, from clinical diagnostics, including non-invasive prenatal diagnostics, to popular personal genomics for assessing potential disease risk and enabling genealogy tracing through commercial kits, and forensic science. These uses bring with them their own ethical concerns—from the personal to the societal. For example, genome profiling can reveal unexpected familial relationships or late-onset disease risk; it may be accessed by law enforcement without knowledge or consent or misused for tracking and oppression (*14*). During the HGP itself, there was vigorous debate over the patentability of genes or genome sequences (*2*), and current initiatives such as the Global Alliance for Genomics and Health (GA4GH) continue to assess ethical hurdles. National and international legal and ethical frameworks should address these challenges. As genomic data and its uses continue to balloon, it will be critical to curb potential abuse and ensure that the legacy of the HGP contributes to the betterment of all human lives. ■

GRAPHIC: C. SMITH/*SCIENCE*

REFERENCES AND NOTES

1. R. A. Gibbs, *Nat. Rev. Genet.* **21**, 575 (2020).
2. J. Sulston, G. Ferry, *The Common Thread: A Story of Science, Politics, Ethics, and the Human Genome* (Joseph Henry, 2002).
3. V. Tam *et al.*, *Nat. Rev. Genet.* **20**, 467 (2019).
4. A. Regev *et al.*, *eLife* **6**, e27041 (2017).
5. B. Ewing, L. Hillier, M. C. Wendl, P. Green, *Genome Res.* **8**, 175 (1998).
6. W. J. Kent *et al.*, *Genome Res.* **12**, 996 (2002).
7. www.genome.gov/10506537/reaffirmation-and-extension-of-nhgri-rapid-data-release-policies
8. L. D. Smith *et al.*, *CSH Perspect. Med.* **6**, a023168 (2016).
9. S. Morganti *et al.*, *Crit. Rev. Oncol. Hematol.* **133**, 171 (2019).
10. E. A. King, J. W. Davis, J. F. Degner, *PLOS Genet.* **15**, e1008489 (2019).
11. E. Cano-Gamez, G. Trynka, *Front. Genet.* **11**, 424 (2020).
12. S. Nurk *et al.*, *bioRxiv* 2021.05.26.445798 (2021).
13. R. E. Peterson *et al.*, *Cell* **179**, 589 (2019).
14. D. Curtis, F. Balloux, *Ann. Hum. Genet.* **84**, 313 (2020).

ACKNOWLEDGMENTS

We thank A. Hupalowska for help with the figure. A.R. is a cofounder and equity holder of Celsius Therapeutics and Immunitas. Both authors are employees of Genentech.

10.1126/science.abl5403

POLICY FORUM

TECHNOLOGY AND LAW

Regulating forensic genetic genealogy

Maryland's new law provides a model for others

By **Natalie Ram[1], Erin E. Murphy[2], Sonia M. Suter[3]**

In May 2021, Maryland enacted the first law in the United States—and in the world—that comprehensively regulates law enforcement's use of consumer genetic data to investigate crimes (*1*). Until now, the primary restraint on law enforcement has come from consumer genetics platforms themselves, with some declining to cooperate, some covertly cooperating, and a handful working openly with criminal investigators. Courts have largely taken a hands-off approach, and one of the only efforts at oversight emerged from the US Department of Justice (DOJ), which issued an "interim" policy in November 2019. By contrast, Maryland's new law—adopted with near-unanimous, bipartisan support—engaged a broad array of stakeholders and was adopted by elected officials after a transparent and open legislative process. Its success provides a roadmap for regulating genetic genealogy in a way that balances privacy and public safety, and its terms include six critical features that others should model moving forward.

SOLVING CRIMES, PROVOKING ALARM

The technology at issue in the Maryland law came to the public's attention with the high-profile arrest of Joseph James DeAngelo Jr. as the Golden State Killer in 2018. Investigators identified DeAngelo, after decades of dead ends, by uploading a genetic profile derived from crime scene DNA to several consumer genetics platforms. The profile pointed to several distant relatives, and after sleuthing in the resulting family trees, investigators homed in on their suspect (*2*). This method—forensic genetic genealogy (FGG)—has since been used in hundreds of other cases in the United States alone while generating considerable attention and interest in other nations, including Australia, Sweden, and the United Kingdom.

[1]University of Maryland Carey School of Law, Baltimore, MD, USA. [2]New York University School of Law, New York, NY, USA. [3]The George Washington University Law School, Washington, DC, USA. Email: nram@law.umaryland.edu

But FGG has generated concern, not just acclaim. In early 2019, FamilyTreeDNA faced criticism after the public (and site users) learned that the company had secretly been working with the US Federal Bureau of Investigation (FBI) for nearly a year to conduct FGG searches. GEDmatch similarly faced reproach after it secretly permitted law enforcement to search its database to investigate a crime outside the scope of the site's stated list of offenses. In response, both sites unilaterally altered their terms of service and privacy policies to explicitly embrace their challenged conduct (*3*). But even after those efforts, sites have come under attack for violating their own policies, quietly changing their settings, and even failing to delete material as promised.

Law enforcement's conduct in pursuing FGG has also provoked alarm. FGG methods can reveal deeply sensitive information. The specific type of genetic material that FGG analyzes can indicate current or future health risks. As the US Supreme Court has acknowledged, such genetic information "presents additional privacy concerns" (*4*). Moreover, linking the records of biologically related persons who are otherwise disconnected in the social world can disclose previously unknown familial relationships or sensitive information, further heightening privacy considerations.

Relatedly, FGG may raise concerns about data security and bias. Genealogists may retain troves of sensitive data on insecure home laptops or cloud services. The incidental characteristics of genealogical databases may also invite law enforcement to devote scarce resources to a narrow subset of crimes. Consumer genetics platforms are disproportionately composed of profiles from individuals of European descent. Given this tilt, and that biological material is most readily recovered in sexual assault cases, police have already disproportionately used FGG to solve cases involving white female victims. This is so even though black males experience substantially higher rates of serious victimization (*5*).

Law enforcement has also, at times, engaged in deception in pursuing FGG. Police have tricked family members of suspects into "volunteering" their genetic data for investigative use or surreptitiously collected genetic material from nonsuspect family members without their knowledge. Or, as happened in the Golden State Killer case, the genealogist working with investigators can simply pose as an ordinary user to gain access to consumer genetics data at a site that does not cooperate with law enforcement (*6*).

As this short history indicates, the policies governing when and under what circumstances law enforcement may engage in FGG have largely been dictated by entities other than actual lawmakers. Consumer genetics platforms have played a primary role. Yet, these private, often for-profit entities hold immensely valuable data, and their interests may not fully align with or accurately reflect those of their users. That much is evident in GEDmatch's and FamilyTreeDNA's repeated and unilateral alterations in the privacy protections of their platforms—sometimes without their users' knowledge—and inconsistent application of their policies without genuine recourse for misled users. Moreover, as ownership of consumer genetics platforms has changed hands, many have wondered whether promises of privacy are trustworthy. This is so even at sites like AncestryDNA, which so far has steadfastly refused to cooperate with law enforcement. Although it has affirmatively fought law enforcement subpoenas and warrants for access to user genetic data, it was recently acquired by Blackstone. In this sense, the consumer genetics marketplace is similar to other markets in big data, where private companies often act in ways that belie consumer trust.

In addition, law enforcement itself has played a major role in deciding when and how FGG is used. Most notably, in 2019, the DOJ issued an "interim" policy governing the federal government's use of FGG to develop investigative leads. Although this policy established some meaningful limits on FGG's use, its impact has been hampered in at least two ways. First, this policy is only binding on federal law enforcement personnel and on state entities that receive federal grant monies specifically to conduct FGG searches. Yet, federal law enforcement comprises only a small fraction of law enforcement entities and officers nationwide. Thus, DOJ's policy does not bind the great majority of law enforcement officers in the United States.

Second, only one relevant constituency drafted the DOJ's interim policy—law enforcement itself. The DOJ, as the federal government's law enforcement agency, crafted a policy that reflects the limits that

law enforcement is willing to impose upon itself, rather than the limits that a broader coalition of relevant stakeholders might agree to and adopt. It also ignored critical issues like defense access to exculpatory searches, which cannot be conducted in any database without access to crime scene samples held in government control.

Finally, courts have indicated that they are not eager to take center stage in setting limits on FGG. In the only case thus far to consider the constitutionality of an FGG search, the court held that the criminal defendant could not challenge that search at all because it was his relative's DNA, and not his own, that first provided a partial match to crime scene evidence. The defendant, in legal terms, lacked standing to challenge the constitutionality of the FGG search (*7*). Although scholars have argued that this reasoning is problematic (*8*), it indicates that courts may be unlikely to provide relief from overzealous law enforcement efforts.

A MODEL FOR LAWMAKING

Under the leadership of state Senator Charles Sydnor III and Delegate Emily Shetty, Maryland took a different approach. The legislators convened a working group to advise the drafting of legislation to regulate the use of FGG in the state. This working group included representatives from the police, prosecution, and crime labs, but it also included representatives from the office of the public defender, the Innocence Project (which uses DNA evidence to free individuals wrongly convicted of crimes they did not commit), and academics with expertise in genetics, genetic privacy, and criminal justice (including the authors). The working group also met with genetic genealogists and private service providers that conduct FGG searches. Over the course of several months, the group reached relative consensus. The bill that emerged garnered near-unanimous bipartisan support in the legislature, as well as support from a wide array of stakeholders across the criminal justice process.

Maryland has often been a leader in regulating law enforcement use of genetic data, including its early ban on searches of law enforcement databases for persons other than criminal suspects. Its newest law in this area and the process from which it emerged stand as a model for lawmaking when new technology presents tensions between privacy and policing. The Maryland statute arrives at a time when several states have begun to set limits on FGG searches. For example, Montana (*9*) and Utah (*10*) also recently enacted laws addressing FGG searches, albeit more limited in scope and sweep than Maryland's. These laws reinforce the notion that regulating FGG searches has appeal across the political spectrum.

Most crucially, Maryland's new law includes the following six critical features that every legislative effort to regulate FGG—whether in the United States or elsewhere—should include.



ILLUSTRATION: DION MBD

Mandates ex ante judicial supervision of law enforcement's use of FGG

Under the new legislation, an FGG search "may not be initiated without judicial authorization and without certifying before the court that the forensic sample and the criminal case satisfy" specified criteria. Many of those criteria are similar to those of the DOJ's interim policy or earlier policies governing familial searches in official law enforcement DNA databases. These include requirements that the crime at issue be sufficiently serious, that investigators reasonably believe the crime scene DNA comes from the putative perpetrator, that the crime scene sample is of sufficient quality for an effective search, that a DNA search in the law enforcement databases failed to identify a suspect, and that other nongenetic investigative methods have been reasonably exhausted without success.

Under Maryland's law, these criteria permit law enforcement to use FGG to investigate crimes like those committed by the Golden State Killer. The criteria also help to minimize risks that law enforcement will pursue FGG for less serious crimes (such as joyriding or commercial trespass) or in cases where the crime scene evidence may point to an innocent bystander rather than the true perpetrator of the crime (such as when a discarded bottle is found in an alley near the victim, but it has no clear connection to the case).

The introduction of judicial supervision for the initiation of an FGG search, moreover, represents a meaningful new protection for individual privacy. Requiring law enforcement to certify to a court, under penalty of perjury, that these criteria are met will help to ensure that FGG is deployed only where appropriate. It serves to prevent law enforcement from engaging in deception, breaking its own rules, or redefining the permissible scope of searches, as has happened in the past.

Affirms individual control over the investigative use of one's genetic data

The DOJ interim policy permits investigators to use consumer genetics platforms that merely provide "explicit notice" to users regarding law enforcement use of profiles. The Maryland statute, by contrast, permits law enforcement use only when platforms obtain affirmative consent from their users. Law enforcement must attest that this is so in seeking judicial authorization to initiate an FGG search. "Explicit notice," standing alone, arguably includes a disclosure buried deep in a site's terms of service or privacy policy, which people rarely read or understand. Consent to law enforcement matching, by contrast, requires a knowing and voluntary choice to participate in law enforcement efforts. Limiting law enforcement access to platforms that obtain genuine "acknowledgment and consent" will better achieve Maryland's intent that ordinary site users knowingly submit genetic data for law enforcement use.

This constraint echoes newly established protections for users of consumer genetics platforms in Utah. Under a statute enacted there in March, consumer genetics platforms serving Utah-based users must obtain a "consumer's express written consent" (or be served with valid legal process) before disclosing a consumer's genetic data to third parties (*10*). Maryland's law goes further, however, by placing the legal obligation on law enforcement to certify in its search application to a judge that it is

only using services that have obtained affirmative consent from their users for law enforcement searches. This approach minimizes the risk of the deceptive behavior that law enforcement has sometimes engaged in when pursuing FGG searches, including simply disregarding the terms of service imposed by a platform.

Establishes robust protections for third parties who are not suspects in the crime

In some instances, law enforcement will want to obtain genetic information from individuals who are not suspected of a crime but who law enforcement believe may be related to a putative perpetrator and whose DNA may be sought to fill in a genealogical family tree. Maryland's statute sets a presumptive requirement of informed consent and prohibits covert collection if a nonsuspect third party has refused consent. Law enforcement must also explain to the third party that nonconsent is a meaningful option because that refusal must be honored. In the absence of this legislation, officers elsewhere have implied or even directly stated that a refusal will simply lead the officers to collect a sample covertly. Moreover, although a court may authorize covert collection from third parties in certain narrow circumstances, the court cannot grant such authorization solely on the basis that an officer believes a party is likely to refuse to consent to sampling. Maryland's law also mandates that the genealogists conducting FGG be properly licensed and integrates genetic counselors into the informed consent process. The law requires that genetic counselors train law enforcement officers in how to appropriately obtain consent from nonsuspect third parties to ensure that the process is more than pro forma.

Taken together, the stringent requirements for informed consent, as well as the highly limited and judicially supervised exceptions permitting covert collection, provide essential protection for innocent persons against potentially overzealous investigative efforts. They also give due respect to the evolving nature of genetic science and the ways in which recreational decisions to share immutable and heritable genetic information may have impacts down the line in ways that cannot now be anticipated fully. Finally, they treat the value of privacy, often seen as abstract, amorphous, and hard to quantify, as a weighty and important concern for a civilized society.

Ensures that FGG is available to prove either guilt or innocence

The law permits criminal defendants and individuals seeking postconviction relief for certain crimes to seek access to FGG, where appropriate. Permitting equitable access to FGG for criminal defendants, and not just prosecutors, is vital to advancing justice and ensuring that only the guilty are convicted. Yet defense counsel typically lacks the negotiating (or strong-arm) power of government actors like the FBI, and sites with explicit terms generally limit access to "law enforcement." Even a defense attorney willing to disregard a platform's terms of service may have trouble accessing crime scene evidence for testing.

Unfettered access to consumer databases is troubling, whether the search is done by the prosecution or the defense. Maryland's law enables and constrains both. The statutory terms applicable to the defense mirror those applicable to the government.

Imposes consequences for violations

Maryland's statute permits an individual to sue if her "genetic genealogy information, FGG profile, or DNA sample is wrongfully disclosed, collected, or maintained" in violation of the law. An individual experiencing such a violation is entitled to a minimum of $5000 in damages. This private right of action ensures that those whose privacy is unlawfully trammeled have recourse to hold violators accountable and avoids any potential judicial limitations as to whether they would have legal standing. This right to sue also provides a stronger incentive to law enforcement and others engaged in FGG to comply with the standards set out in the statute.

Requires annual public reporting and review

Public reporting must include how often, under what circumstances, and in what manner these searches take place, as well as what share of these searches lead to an arrest, successful prosecution, or exoneration. These reporting requirements establish a transparent process that will allow legislators and the public to learn how consumer genetic information intersects with law enforcement efforts. Among other things, it is important for Marylanders to know whether law enforcement use of FGG continues to focus disproportionately on cases with white victims or leads to disproportionate third-party sampling or arrest of Black or other individuals from marginalized populations.

A further provision requires that an expert panel, including a broad array of relevant stakeholders, reviews this report and makes policy recommendations. This review ensures that Maryland will be able to exercise informed oversight of FGG methods and adapt state policy as needed. Such transparency and accountability will help promote public trust in the process.

SENSIBLE SAFEGUARDS, PUBLIC TRUST

In sum, Maryland's new law enacts a robust statutory scheme for balancing public safety, privacy, and criminal justice. It emerged from a process that sought consensus across divergent constituencies, not merely law enforcement. The conversations were expertly shepherded by the legislator- sponsors, with a commitment to identify key points of agreement and disagreement and work toward meeting "in the middle." The participants also each engaged in good faith, recognizing that the goal was not to win each single argument but to reach consensus around a sensible suite of safeguards. Indeed, law enforcement and privacy advocates both found value in developing guardrails. For example, some members of law enforcement noted that limits on the technology to promote public trust would help ensure that the technology remains available to assist them in their work.

The resulting legislation is more protective than existing policies that have emerged from consumer genetics platforms, law enforcement entities, or courts, while still permitting law enforcement to use FGG in the most appropriate cases. And it establishes a framework for transparency and accountability that will ensure that Maryland's citizens continue to enjoy both privacy and security. Maryland's law is the first of its kind, but other jurisdictions—both in the United States and elsewhere around the world—should take note. Meaningful and equitable restraints are not only possible; they are a worthy pursuit. ■

REFERENCES AND NOTES

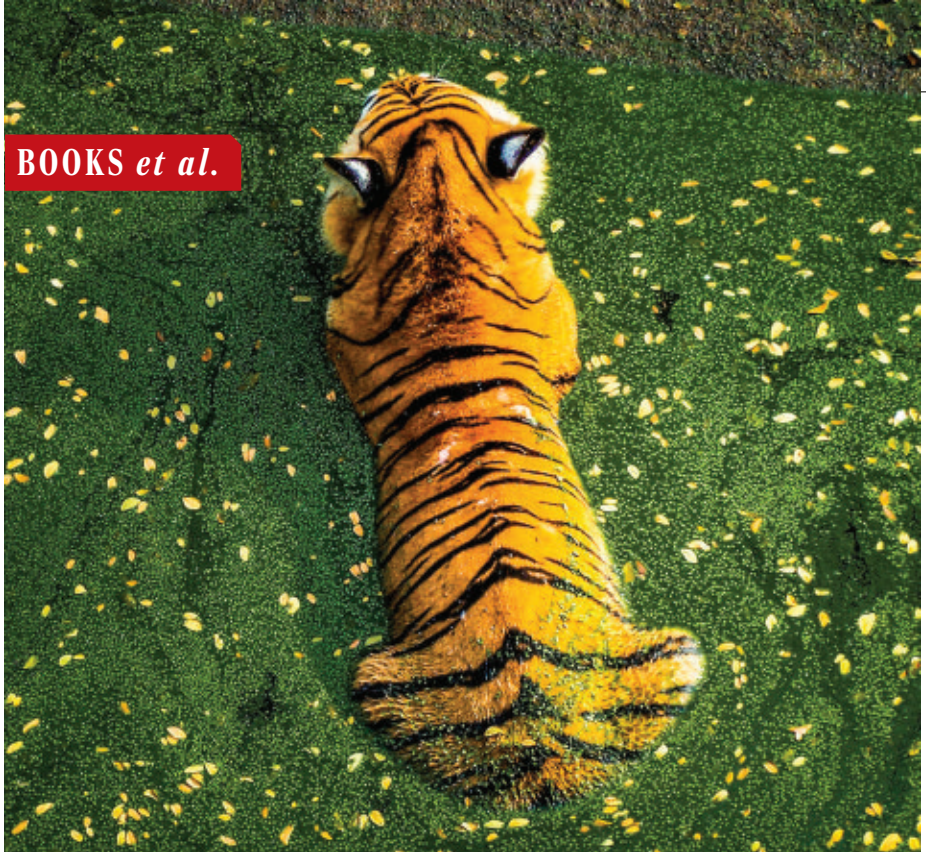
1. H. B. Maryland, 240 (2021); https://mgaleg.maryland.gov/2021RS/chapters_noln/Ch_681_hb0240E.pdf.
2. N. Ram, C. J. Guerrini, A. L. McGuire, *Science* **360**, 1078 (2018).
3. N. Ram, J. L. Roberts, *Nat. Biotechnol.* **37**, 707 (2019).
4. *Maryland* v. *King*, 569 US 435, 464 (2013).
5. J. Stern, S. Zhang, "The victims left behind by genetic genealogy," *Atlantic*, 27 January 2021.
6. P. St. John, "The untold story of how the Golden State Killer was found: A covert operation and private DNA," *Los Angeles Times*, 8 December 2020.
7. *Iowa* v. *Burns*, No. FECR129718, Slip Op. at 7 (Iowa Dist. Ct., 6 February 2020).
8. N. Ram, *SciTech Lawyer* **16**, 18 (2020).
9. H. B. Montana, 602 (2021); https://legiscan.com/MT/text/HB602/id/2384004/Montana-2021-HB602-Enrolled.pdf.
10. S. B. Utah, 227 (2021); https://le.utah.gov/~2021/bills/static/SB0227.html.

ACKNOWLEDGMENTS

Funding was provided by a Greenwall Foundation Faculty Scholars Grant (N.R.).

10.1126/science.abj5724


Ignite Your Passion
Your passion for discovery transforms a precious sample into a flicker of hope.
It takes more than technology to fight disease. Your vision ignites the drive to discover, the freedom to optimize, and the confidence to impact outcomes.
Ignite your passion at explore.agilent.com/ignite-change
Agilent
Trusted Answers
For Research Use Only. Not for use in diagnostic procedures.
© Agilent Technologies, Inc. 2021



The tiger's camouflage striping is a visually powerful example of mirror symmetry.

PHOTO: SANDEEP VR/GETTY IMAGES

LAWS OF NATURE

The allure of an ordered Universe

A science writer's ode to patterns packs a visual punch

By **Saleem H. Ali**

The human brain's proclivity for patterns is the inspiration for *Ten Patterns That Explain the Universe*, Brian Clegg's visually appealing compendium of scientific knowledge. The book's focus on a mere 10 patterns, in conformance with the decimal system, is itself an example of how humanity craves order. But do such patterns underlie our cosmic reality, or are they functional artifacts that help the human mind process information? Catering to a popular audience, Clegg wrestles with these questions as he elaborates on his chosen patterns, which range from periodicity present in cosmic microwave background radiation maps and the double-helix structure of DNA to heuristics that aim to capture the probability of certain events, such as the Feynman diagrams that describe interactions between subatomic particles.



Ten Patterns That Explain the Universe
Brian Clegg
MIT Press, 2021. 224 pp.

This book is a valuable contribution to popular science, providing a kaleidoscope of visuals that will appeal to readers across generational divides. The section on mathematical patterns that are distilled as number lines is particularly well written. Here, Clegg links well-known patterns such as the Fibonacci series to more fundamental tools in mathematical heuristics, such as logarithms. Elusive concepts such as imaginary numbers, cardinality, and infinite sets are presented in a highly readable narrative alongside clear diagrams.

Pithy quotes from classical and contemporary figures including Blaise Pascal and Annie Proulx add literary flare and inspirational interludes to the narrative. Quoting Michio Kaku, for example, the book notes, "To a physicist, beauty means symmetry and simplicity."

Clegg has vast experience in popularizing science. He uses that experience effectively in this book to help the reader visualize complex phenomena from various perspectives. In his discussion of the periodic table, for example, he presents a history of its development but also considers various alternative ways in which the table could be organized, including an elegant looped table with silicon at the center.

Cladograms are a compelling pattern included in this roundup, presenting "the tree of life" in various forms. Yet in his quest to make the material accessible, Clegg neglects to discuss key insights regarding horizontal gene transfer that challenge classical cladograms and help to explain why bacteria are able to develop antibiotic resistance, which have been explored capably elsewhere (*1*).

Complexity science—the subject of Clegg's 2020 book *Everyday Chaos*—is likewise touched on only in passing, as it relates to weather patterns, but its broader connections to pattern emergence and structure in living forms are omitted. Complexity doyen Stuart Kauffman's work on fitness landscapes (*2*, *3*) would have made him a fine addition to the pantheon of scientists profiled elsewhere in the book.

Clegg's concluding pattern is symmetry, a topic that may seem like a captivating coda but is, unfortunately, given short shrift. For more trenchant insights on symmetry, the reader must seek out other sources, such as those in Nobel laureate P. W. Anderson's short and highly readable classic paper "More is different: Broken symmetry and the nature of the hierarchical structure of science" (*4*).

Among the most valuable insights offered in this volume are Clegg's observations in the book's opening overture that caution readers against relying too heavily on patterns, noting how they can sometimes lead us astray. Our brains have evolved to take patterns seriously because they can be useful in mitigating risk, thereby ensuring survival. Yet an overreliance on pattern detection can also lead to negative stereotyping. As Clegg astutely notes, "Prejudice, too, is the result of a misapplication of a pattern, extrapolating from a single example to a wider group."

In the highly crowded market of popular science media, Clegg has provided a useful, if incomplete, distillation of universal patterns that stands out for its ambitious attempt to synthesize themes and observations made across various genres of natural sciences and mathematics. ■

The reviewer is at the Department of Geography and Spatial Sciences, University of Delaware, Newark, DE 19716, USA, and is the author of a book on the laws of nature to be published in 2022 by Oxford University Press. Email: saleem@alum.mit.edu

REFERENCES AND NOTES

1. D. Quammen, *The Tangled Tree* (Simon & Schuster, 2018).
2. S. A. Kauffman, *The Origins of Order* (Oxford Univ. Press, 1993).
3. S. A. Kauffman, *At Home in the Universe* (Oxford Univ. Press, 1996).
4. P. W. Anderson, *Science* **177**, 393 (1972).

10.1126/science.abj8932

SCIENCE AND SOCIETY

Biology meets public policy

The tenuous portability of polygenic scores undermines an attempt to link genetics and justice

By **Aaron Panofsky**

Behavior genetics, its forbear eugenics, and its cousins sociobiology and evolutionary psychology have always been political sciences. When not being used to bar immigration, sterilize, or exterminate, behavior genetics has been an eager ideological partner of what Albert O. Hirschman called the "rhetoric of reaction."

Extrapolating from basic statistical correlations among twins, behavior geneticists have told us that parenting, education, and social supports matter little and have blasted the progressive aims of the welfare state. Reformist liberals, they say, have ignored this correlation structure, futilely wasted resources, and jeopardized our fragile meritocratic, capitalistic society by perversely encouraging bad breeding and behavior.



Genetic conditions such as myopia can have environmental solutions.

PHOTO: KATLEHO SEISA/ISTOCK.COM

In her new book, *The Genetic Lottery*, Kathryn Paige Harden somewhat euphemistically labels this reactionary tradition as "hereditarian pessimism," and she aims to extricate behavior genetics from it. The first half of the book offers an overview of genetics as it relates to social inequality, while the second reframes moral philosophy and reviews behavior genetics studies to make the case for a genetically driven, antieugenic public policy.

The reviewer is director of the UCLA Institute for Society and Genetics, Los Angeles, CA 90095, USA. Email: apanofsky@ucla.edu

Unlike others in this tradition, Harden does not use science to claim "objective" limits to social improvement. For her, behavior genetics is a political science, but not in a negative way. Harden is an unabashed combatant for social justice, and she claims that behavior genetics is an essential weapon for the struggle.

The Genetic Lottery is a good read, peppered with relatable stories and examples. Harden pulls off the trick of simultaneously introducing a technical field to newcomers; addressing deep, specialist debates; and taking seriously the intersection of scientific and philosophical analyses of inequality.

Harden begins by explaining that in the domain of educational attainment, there are no genes for schooling or even intelligence, but there are a huge number of small genetic differences and complex biosocial pathways that shape children's outcomes. Harden wants to attribute causality to correlational behavior genetics studies on the basis of heritability estimates and polygenic risk scores but admits that they are causal "difference makers" only for particular genotypes distributed in particular environments. Most accounts of causality, however, are concerned with external validity or generalizability. Harden slides past this issue.

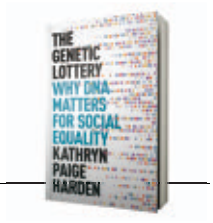
This problem is acute in Harden's discussion of race. On one hand, she accurately distinguishes between social race and genetic ancestry and shows why extending intelligence or education polygenic scores derived in European-ancestry populations to Black or brown populations is motivated by racism. But the so-called portability of polygenic scores within ancestry groups is also problematic. For good reasons, Harden limits her discussion to American inequality among people of European ancestry, but such a discussion that excludes racial inequality undermines the notion that behavior genetics is an essential tool for social justice.

In the book's second half, Harden deploys the philosopher John Rawls to call for interventions into the forms of "luck"—genetic and social—that largely determine social inequalities. She shows that polygenic scores can reveal hidden inequalities (for example, poor children who fail to attend college despite high genetic educational attainment scores) and might help determine which policies are effective. The book concludes with a helpful contrast of eugenic, gene-blind, and antieugenic genetic social policy.

In the end, the book might have been more successful had it targeted the policy pessimism and soft racism of the center-right rather than trying to convince the progressive left of genetics's relevance to social policy. Debates about which forms of luck a decent society is obligated to confront are the province of center-right hand-wringing; the left has plenty of philosophical justification, from human rights and solidarity to systemic critiques of exploitative capitalism and colonialism. Likewise, Harden appeals to the technocratic policy domain of chiseling tweaks and neoliberal nudges. But the progressive left has moved down the field, promoting structural reforms and redistribution that, if politically successful, will alter the very gene–environment correlations on which the tenuous portability of polygenic scores depends. Harden could have served her purposes better by targeting the rhetoric of neoreactionaries rather than trying to yoke the left to genetics. ■

10.1126/science.abl3851

The Genetic Lottery: Why DNA Matters for Social Equality
Kathryn Paige Harden
Princeton University Press, 2021. 312 pp.



SCIENCE & RACE



PODCAST

The Social Life of DNA: Race, Reparations, and Reconciliation After the Genome
Alondra Nelson
Beacon Press, 2016. 216 pp.

Our genome is just one of many facets of our identity, but for descendants of enslaved individuals, genetic genealogies can offer clues about one's cultural and ancestral heritage that have otherwise been erased from history. This week on the *Science* podcast, Alondra Nelson discusses the tensions that exist between the beneficial uses of genetic technologies in this context and their ultimately limited capacity to answer fundamental questions about identity. https://scim.ag/2XpjpCr

10.1126/science.abm1869



Elaine Redding
Brinster Prize
in
Science *or* Medicine

♦

CONGRATULATIONS
TO

C. David Allis, Ph.D.
Rockefeller University

INAUGURAL WINNER

♦

Penn IRM
Institute for Regenerative Medicine

For more information:
irm.med.upenn.edu/brinster

Where Science Gets Social.

AAAS.ORG/COMMUNITY






AAAS' Member Community is a one-stop destination for scientists and STEM enthusiasts alike. It's "Where Science Gets Social": a community where facts matter, ideas are big and there's always a reason to come hang out, share, discuss and explore.

Member COMMUNITY
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE



Decommissioned satellites could be repurposed to capture valuable data at night, such as this image of the Mediterranean Sea.

IMAGE: JOSHUA STEVENS USING SUOMI NPP VIIRS/EARTH OBSERVATORY/NASA DATA: MIGUEL ROMAN/NASA/GSFC

Edited by **Jennifer Sills**

Retired satellites: A chance to shed light

Satellite data are vital to decisions about both environmental and human health. As satellites are replaced and decommissioned, they could be repurposed to spark new scientific insights.

The path and timing of the Suomi National Polar-orbiting Partnership (NPP) and National Oceanic and Atmospheric Administration–20 satellites—part of the larger Joint Polar Satellite System (JPSS)—prioritize the collection of data for numerical weather prediction models (*1*). Also flying on these satellites is the Visible Infrared Imaging Radiometer Suite's (VIIRS) day/night band (DNB) (*2*), the only openly accessible instrument currently measuring lights from Earth's surface at night, globally. These nighttime data have enabled scientists to track urbanization, estimate population and gross domestic product, and monitor the impacts of disasters and conflicts (*3*). However, Suomi-NPP and the JPSS constellation of satellites acquire measurements at about 1:30 am local time, even though human activities substantially decrease after midnight, as do lighting from building windows, vehicle headlights, businesses, and timed streetlights. Late-night lighting disproportionately decreases over developing countries due to energy and cost constraints (*4*). Because the satellites' orbits were selected to meet high-priority demands, such as weather forecasting, they are missing a wealth of potential information that could be gathered from measurements collected at an earlier local time.

Earth and social scientists have long wanted access to earlier nighttime observations. The launch of follow-on JPSS satellites, beginning in 2022, provides a valuable opportunity. The new satellites are expected to continue measurements at 1:30 am. Meanwhile, the decommissioned spacecraft they replace could be shifted to an orbit at an earlier local time. Collecting imagery at around 10:00 a.m. and 10:00 p.m. would minimize the risk of cloud cover and match the orbit of two predecessor sensors—the Terra Moderate Resolution Imaging Spectroradiometer (MODIS), which has collected daytime data at 10:30 a.m. for 20 years (*5*), and the Defense Meteorological Program Operational Line-Scan System (DMSP-OLS), which collected nighttime data between 7:30 and 9:30 p.m. for more than 40 years (*6*).

Additional evening data would deepen understanding of the rhythm of human civilization, much in the same way that daytime sensors—many of which collect data continuously throughout the day—have illuminated phenomena such as seasonal vegetation, sea ice cycles, wind patterns, ocean currents, land surface temperatures, and precipitation patterns (*3*, *7*). Multiple observations per night would better characterize the stability and angular properties of light sources, improving detection of traffic, power outages (*8*), conflict (*9*), illegal fishing (*10*), and demographic trends.

Reprogramming retired satellites is not standard practice and may require additional planning and investment. However, with replacement JPSS satellite launches planned every 4 to 6 years, each launch will offer another opportunity to repurpose a retired satellite. Future JPSS satellites should be designed to include a "retirement plan" that can provide data to fill knowledge gaps. For researchers interested in Earth at night and the broader land science community, repurposing decommissioned VIIRS-DNBs would provide crucial data for disaster monitoring, while advancing the science of human activities, their dynamics, and their interactions with the natural world.

Eleanor C. Stokes[1,2]*, Miguel O. Román[1], Zhuosen Wang[3,4], Christopher C. M. Kyba[5], Steven D. Miller[6], Tobias Storch[7], Kevin R. Gurney[8]

[1]Earth from Space Institute, Universities Space Research Association, Columbia, MD 21046, USA. [2]GESTAR, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [3]Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20740, USA. [4]Terrestrial Systems Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [5]GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany. [6]Cooperative Institute for Research in the Atmosphere, Colorado State University, Fort Collins, CO 80521, USA. [7]Earth Observation Center, German Aerospace Center (DLR), 82234 Oberpfaffenhofen, Germany. [8]School of Informatics, Northern Arizona University, Flagstaff, AZ 86011, USA.
*Corresponding author.
Email: estokes@usra.edu

REFERENCES AND NOTES

1. Joint Polar Satellite System, Mission and Instruments (www.jpss.noaa.gov/mission_and_instruments.html).
2. S. D. Miller *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15706 (2012).
3. N. Levin *et al.*, *Remote Sens. Environ.* **237**, 111443 (2020).
4. S. B. Pritchard, *Environ. Hist.* **22**, 312 (2017).
5. NASA, Terra, MODIS (2021); https://terra.nasa.gov/about/terra-instruments/modis.
6. M. O. Román *et al.*, *Remote Sens. Environ.* **210**, 113 (2018).
7. Z. Zhu *et al.*, *Remote Sens. Environ.* **228**, 164 (2019).
8. NASA, Earth Observatory, Satellite Observes Power Outages in New Orleans (2021); https://earthobservatory.nasa.gov/images/148777/satellite-observes-power-outages-in-new-orleans.
9. NASA, Earth Observatory, Night Lights Change in the Middle East (2016); https://earthobservatory.nasa.gov/images/90100/night-lights-change-in-the-middle-east.
10. NASA, Earth Observatory, Something Fishy in the Atlantic Night (2013); https://earthobservatory.nasa.gov/features/Malvinas.

10.1126/science.abl9965

Protect minorities in genetic research

In his In Depth story "Genetics papers from China face ethical scrutiny" (13 August, p. 727), D. Normile points out that publications based on DNA samples collected from ethnic minorities, including Uyghurs and other minority groups in China, now face ethical scrutiny from journal editors. Such concerns should not be restricted to repressive regimes or to current studies. Scientists have unethically used data collected from Indigenous and vulnerable people across the world for decades.

Large database and biobank projects include data and samples collected over decades from Indigenous peoples and minority groups. These data and materials have been used and shared by researchers around the world and across disciplines. Exploitative research practices in genetic research have been reported in Indigenous communities and vulnerable populations (*1*, *2*), including the San people in Southern Africa (*3*); the Havasupai Tribe (*4*), Navajo Nation (*5*), and other Native American communities in the United States (*6*); and Australia's Indigenous people (*7*).

Since the 1990s, DNA has also been collected from Roma individuals in Europe. Roma have been subjected to a long history of social discrimination, including internment in camps, forced labor, and extermination. Scientists were infamously involved in their persecution (*8*, *9*). Even in the recent past, ethical standards have remained low during data collection from Roma individuals, many of whom may not have been able to read an informed consent form (*10*). Privacy issues persist, and the vocabulary used to label and describe these populations in genetic research is sometimes stigmatizing and racializing (*10*, *11*).

Data and material collected from Roma individuals have been widely shared for decades by geneticists (*10*). Their data have also been uploaded to forensic genetic databases used by law enforcement, even if originally collected for medical research (*12*). Such secondary usage violates contemporary ethical norms and regulations, as set out in Article 31 of the United Nations Declaration on the Rights of Indigenous Peoples.

Scientists must ensure the protection of individuals who have agreed to contribute their DNA, and they must not use DNA obtained without appropriate consent. Failing to do so violates human rights and will likely decrease public trust in genetics.

Veronika Lipphardt[1]*, Gudrun A. Rappold[2], Mihai Surdu[1,3]
[1]Albert-Ludwig-University, Freiburg, Germany. [2]Ruprecht-Karls-University, Heidelberg, Germany. [3]Research Centre on Antigypsyism, University of Heidelberg, Heidelberg, Germany.
*Corresponding author:
Email: veronika.lipphardt@ucf.uni-freiburg.de

REFERENCES AND NOTES

1. M. E. D'Amato *et al.*, *Forens. Sci. Int. Genet.* **48**, 102299 (2020).
2. E. Kowal, J. Radin, J. Reardon, *Soc. Stud. Sci.* **43**, 465 (2013).
3. R. Chennells, A. Steenkamp, in *Ethics Dumping: Case Studies from North-South Research Collaborations*, D. Schroeder, J. Cook, F. Hirsch, S. Fenet, V. Muthuswamy, Eds. (Springer International Publishing, 2018), pp. 15–22.
4. N. A. Garrison, *Sci. Technol. Hum. Values* **38**, 201 (2013).
5. J. Reardon, *The Postgenomic Condition: Ethics, Justice, and Knowledge after the Genome* (University of Chicago Press, 2017).
6. K. TallBear, *Native American DNA: Tribal Belonging and the False Promise of Genetic Science* (University of Minnesota Press, 2013).
7. E. Kowal, *Soc. Stud. Sci.* **43**, 577 (2013).
8. B. Müller-Hill, *Murderous Science: Elimination by Scientific Selection of Jews, Gypsies, and Others: Germany, 1933-1945* (Oxford University Press, 1988).
9. E. Varsa, *National. Pap.* **45**, 114 (2017).
10. V. Lipphardt, G. A. Rappold, M. Surdu, *Sci. Context* 10.13140/RG.2.2.13286.04165 (2021).
11. V. Lipphardt, *Varia Historia* **33**, 109 (2017).
12. Q. Schiermeier, *Nature* **594**, 320 (2021).

10.1126/science.abm2270

True informed consent requires more than forms

In his News Feature "Failure to Protect?" (13 August, p. 729), C. Piller quotes several commentators who suggest that the contents of an informed consent form could have influenced whether parents decided to enroll their child in an asthma study. However, informed consent forms were never intended to be a primary method for educating potential study participants. Such forms merely document an agreement to enroll in research; they have always been poor purveyors of information.

All forms exist to protect the institutions that create them. Research consent forms have increased in length over time and now include copious "boilerplate" language unrelated to informing subjects about the research at hand. The form for the asthma study described by Piller was 15 single-spaced pages. Worse, consent forms tell us nothing about what potential subjects were told by the research staff tasked with enrolling the required number of participants.

Instead of focusing exclusively on the precise wording of consent forms, we should make sure that potential research participants thoroughly understand the benefits and risks of the research project to which they are being recruited. This process should be facilitated by participant advocates who are skilled at providing scientific information to a scientifically naïve population. Documentation that this information was provided could be achieved by digitally recording and archiving these discussions.

No consent form or process would excuse the approval of unethical research by Institutional Review Boards, nor would informed consent to unethical research make it ethical. However, the failure to obtain true informed consent from participants, beyond a signature at the end of an unreadable form, does make otherwise ethical research unethical.

Leonard H. Glantz
Department of Health Law, Policy, and Management, Boston University School of Public Health, Boston, MA 02118, USA.
Email: lglantz@bu.edu

10.1126/science.abl9591

TECHNICAL COMMENT ABSTRACTS

Comment on "Recent global decline of CO_2 fertilization effects on vegetation photosynthesis"

Christian Frankenberg, Yi Yin, Brendan Byrne, Liyin He, Pierre Gentine

Wang *et al*. (Research Articles, 11 December 2020, p. 1295), using photosynthesis proxies from long-term satellite records, report a significant decline in CO_2 fertilization effects. We find that small systematic biases in Advanced Very High Resolution Radiometer (AVHRR) data affect their analysis to the degree that the key finding is not robust. Caution is recommended when using AVHRR to detect changes in near-infrared reflectance of vegetation (NIRv) trends and vegetation indices.

Full text: dx.doi.org/10.1126/science.abg2947

Comment on "Recent global decline of CO_2 fertilization effects on vegetation photosynthesis"

Yuxing Sang, Ling Huang, Xuhui Wang, Trevor F. Keenan, Chenzhi Wang, Yue He

Wang *et al*. (Research Articles, 11 December 2020, p. 1295) reported a large decline in the CO_2 fertilization effect (CFE) on global photosynthesis, on the basis of changing regression coefficients of CO_2 (β) for satellite proxies/products of photosynthesis. We show that their methods cannot robustly estimate CFE, leading to incorrect attribution, and largely neglect CO_2 effects on the photosynthetic enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase.

Full text: dx.doi.org/10.1126/science.abg4420

Comment on "Recent global decline of CO_2 fertilization effects on vegetation photosynthesis"

Zaichun Zhu, Hui Zeng, Ranga B. Myneni, Chi Chen, Qian Zhao, Junjun Zha, Simin Zhan, Ian MacLachlan

Wang *et al*. (Research Articles, 11 December 2020, p. 1295) reported a large decrease in CO_2 fertilization effect (CFE) across the globe during the period 1982–2015 and suggested that ecosystem models underestimate the rate of CFE decline. We find that their claims are artifacts of incorrect processing of satellite data and problematic methods for deriving and comparing CFE between satellite data and model simulations.

Full text: dx.doi.org/10.1126/science.abg5673

Response to Comments on "Recent global decline of CO_2 fertilization effects on vegetation photosynthesis"

Songhan Wang, Yongguang Zhang, Weimin Ju, Jing M. Chen, Alessandro Cescatti, Jordi Sardans, Ivan A. Janssens, Mousong Wu, Joseph A. Berry, Elliott Campbell, Marcos Fernández-Martínez, Ramdane Alkama, Stephen Sitch, William K. Smith, Wenping Yuan, Wei He, Danica Lombardozzi, Markus Kautz, Dan Zhu, Sebastian Lienert, Etsushi Kato, Benjamin Poulter, Tanja G. M. Sanders, Inken Krüger, Rong Wang, Ning Zeng, Hanqin Tian, Nicolas Vuichard, Atul K. Jain, Andy Wiltshire, Daniel S. Goll, Josep Peñuelas

Our study suggests that the global CO_2 fertilization effect (CFE) on vegetation photosynthesis has declined during the past four decades. The Comments suggest that the temporal inconsistency in AVHRR data and the attribution method undermine the results' robustness. Here, we provide additional evidence that these arguments did not affect our finding and that the global decline in CFE is robust.

Full text: dx.doi.org/10.1126/science.abg7484

ERRATA

Erratum for the Research Article "Liver homeostasis is maintained by midlobular zone 2 hepatocytes," by Y. Wei *et al*., *Science* **373**, eabl8195 (2021). Published online 3 September 2021; 10.1126/science.abl8195

Erratum for the Report "Global acceleration in rates of vegetation change over the past 18,000 years," by O. Mottl *et al*., *Science* **373**, eabk2221 (2021). Published online 9 July 2021; 10.1126/science.abk2221



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**

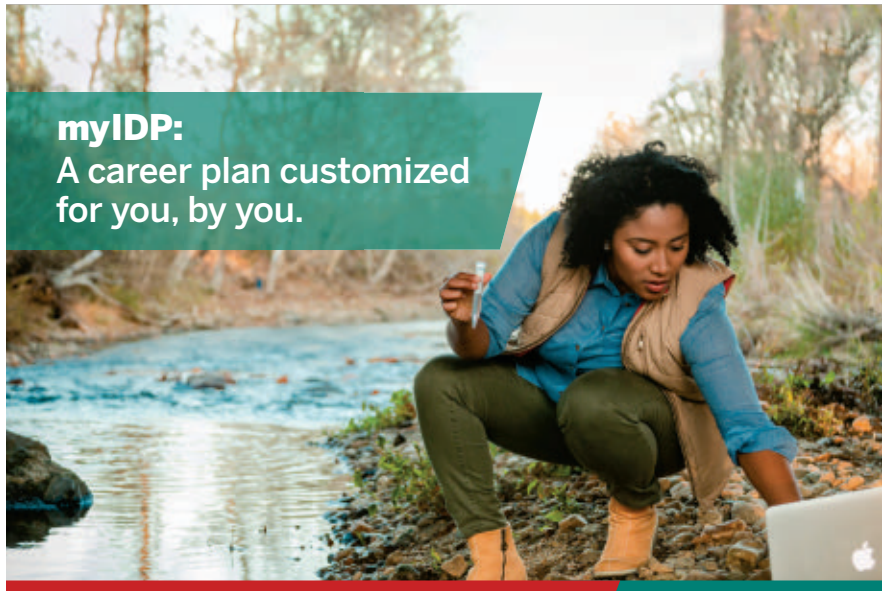
The Most ADVANCED Patch Clamp System Available Today

Visit
www.dpatch.com
to learn more



SUTTER INSTRUMENT

+1.415.883.0128
info@sutter.com
www.sutter.com



For your career in science, 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.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.

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

FASEB Federation of American Societies for Experimental Biology

UMASS University of Massachusetts Medical School

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**



Fellows provide technical talent to a wide range of federal policy-makers.

Fellowships bring much-needed expertise to lawmakers

More than 3400 alumni work in government, academia, nonprofits, and industry

By **Becky Ham**

Since 1973, the Science & Technology Policy Fellowships have provided essential scientific expertise to US policy-makers in Washington, D.C. From its inaugural class of seven fellows, supported by the American Association for the Advancement of Science, the American Physical Society, the American Society of Mechanical Engineers, and the Institute of Electrical and Electronics Engineers, the program now facilitated by AAAS's Science & Technology Policy Fellowships (STPF) brings more than 250 scientists each year to work in congressional offices, executive branch agencies, and the federal judiciary.

Like their predecessors, the 2021–2022 fellowship class that began work on 1 September will bring a much-needed set of technical skills, such as data analysis, along with subject matter expertise in fields from epidemiology to advanced manufacturing, to support the US Congress and other federal policy-makers.

According to a November 2020 report by the Harvard Kennedy School's Belfer Center for Science and International Affairs, US policy-makers lack the resources to maintain congressional committee staff to absorb the complex science and technology information needed to inform policy. In the past two decades, other sources of scientific and technical expertise have disappeared as well, such as the Office of Technology Assessment, defunded in 1995.

In the face of this diminished institutional knowledge and support, the Belfer Center report noted, the STPF program has been "one of the oldest and most respected pathways sending technical talent to Capitol Hill."

A 2019 report by the Goodman Research Group that reviewed the program's impact found that 95% of the government staff who mentor the fellows and were interviewed by Goodman felt that the fellows made important contributions to their offices, including translating scientific content into accessible language, providing expertise on complex problems, and drafting briefing documents, among other work.

"We have long assumed that this program is highly effective. The outcomes can be seen all around us in the thousands of alumni working at the intersection of science and policy, and it serves as a model for virtually every other fellowship that brings science to all levels of government, at home and abroad," said AAAS CEO Sudip Parikh about the Goodman report.

Diana Pankevich, a 2014–2016 Executive Branch Fellow at the White House Office of Science and Technology Policy, said in a 28 May live chat with STPF that her scientific background was immediately useful in her new job—even though she was a neuroscientist now working on climate change issues.

> **"When you hear STPF fellows... talk about their experiences, the lasting impacts... become obvious."**
>
> **Rashada Alexander,** STPF Director

"I was able to use the skills of being able to quickly research topics, being able to dig deep enough into the topic to understand it and figure out what are the challenges and what are the opportunities," recalled Pankevich, now a director for science and innovation policy at Pfizer. "The curiosity that you develop as a scientist, and as a grad student and as a postdoc, thinking about hypotheses and thinking about questions, I think that curiosity for me

PHOTO: ZRFPHOTO/ISTOCK.COM

is what was very helpful and probably was a foundation for the success I had at OSTP."

According to the Belfer Center report, scientists also bring "metacognitive diversity" to the policy-making process, approaching problems and designing solutions in a different way from their new peers in government. Georgia Hartman, a 2020–2022 Executive Branch Fellow, has experienced this phenomenon in her work with the Gender Equality and Women's Empowerment Hub at the US Agency for International Development.

As an anthropologist, she said in the 28 May live chat, it was second nature for her—but not some of her colleagues—to consider the different roles of people in the environmental projects pursued by the agency. "I think part of the reason I'm there, and why they want other social scientists there, is that we have a really different way of understanding the world."

During their tenure, fellows dramatically increase their understanding of policy-making and productive collaboration, according to the Goodman findings. These skills remain valuable career assets even if fellows return to academia. Gillian Bowser, a 2011 Executive Branch Fellow in the US Department of State's Office of Marine Conservation, said in an 18 August STPF live chat that learning how to work on a multidisciplinary and multicultural team was one of the most important skills she took away from the fellowship.

Bowser, an associate professor in ecosystem science and sustainability at Colorado State University, said that the fellowship had broadened her concept of multidisciplinary collaboration. On negotiating teams like those she worked on at the State Department, "you could be talking to social scientists, you could be talking to different countries, talking to other scientists, you're talking to people all over, and you have to come up with a common language." It's a lesson that she carries forward with her students, as she teaches them about the environmental negotiation process.

Paul Lartey, a 2017–2019 Executive Branch Fellow at the National Science Foundation's Office of Legislative and Public Affairs, had worked in the pharmaceutical industry for years and thought that government work would be more "laid back," he said in the 28 May chat. "That was one assumption I had that very quickly disappeared. People do work hard, and I mean hard…I kept telling myself, this country is in good hands, because people will just work until the job is done."

Even after their fellowships end, the program's participants remain a vital and much-needed force in government. Don Ufford, a 2020–2022 Executive Branch Fellow at the Office of Advanced Manufacturing at the National Institute of Standards and Technology, said in the 18 August chat that he had been impressed to see the network of former fellows in action in Washington.

"Some parts of the US government can be siloed between agencies whereas it seems like the fellows can cross over those agencies a little bit more easily," said Ufford, a former director of global vehicle engineering at Ford Motor Company, "and maybe create some opportunities that wouldn't be there otherwise."

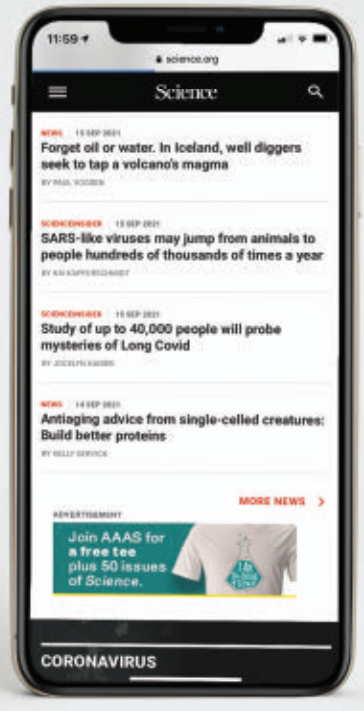
The Goodman report found that 42% of fellows went into federal, state, or local government jobs at the conclusion of their fellowships. This network of fellows encourages science policy cooperation across agencies and offices, while offering support to new fellows entering government service, according to the program's alumni.

"When you hear STPF fellows and alumni talk about their experiences, the lasting impacts across multiple facets of our government and scientific advancement become obvious. Reports like those produced by the Belfer Center and the Goodman Research Group get me very excited, because they help quantify and enhance the story of the impact of STPF," said Rashada Alexander, the program's director. "It also helps folks see how investing in and supporting STPF yield a lot of much-needed value and technical support to policy-makers."

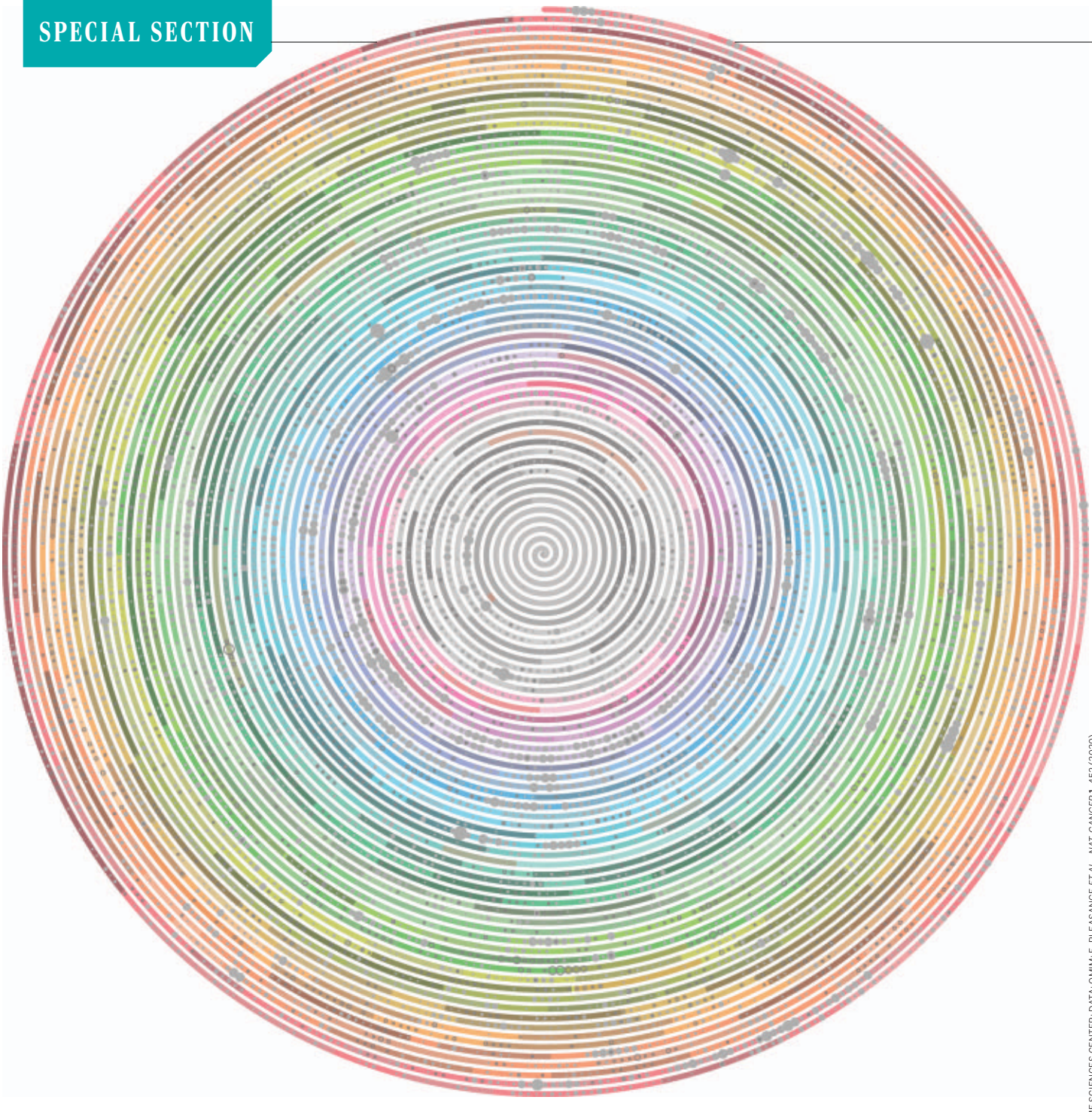
NEWS FROM
Science

Up-to-the-minute research and policy news you won't find in print

Visit us online to read all the news coverage that there just wasn't enough room to print in this issue.



Science.org/news

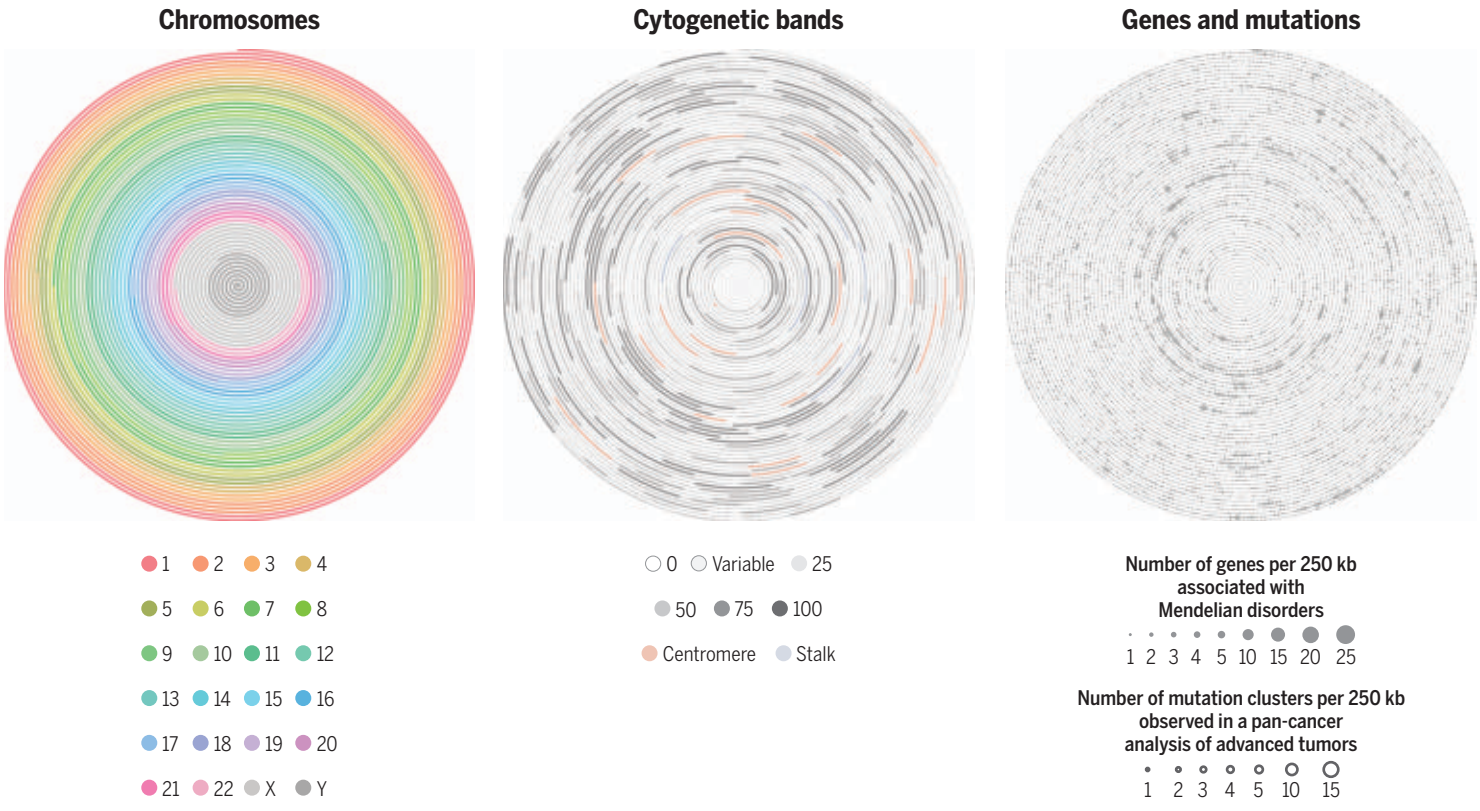


GRAPHIC: MARTIN KRZYWINSKI/CANADA'S MICHAEL SMITH GENOME SCIENCES CENTER; DATA: OMIM; E. PLEASANCE *ET AL. NAT. CANCER* **1**, 452 (2020)

THE HUMAN GENOME

A RESEARCH ULTRAMARATHON

By **Laura M. Zahn**



In the two decades since the publication of the human genome, advances in genetics have raced forward. Progress in human genomics has produced insights applicable to treating cancer, identifying drug targets, and diagnosing disease—especially in infants. However, despite these steps forward, we still lack the ability to predict how an individual's inherited genome contributes to aging, complex disease, and even some monogenic diseases. Furthermore, de novo mutations have increasingly been proposed to affect disease onset and progression.

Although some hoped that having the human genome in hand would fuel a sprint to medical miracles, progress in this field is more accurately characterized as an ongoing relay race of contributions from genomic studies. We now have a better understanding of our history from ancient DNA sequencing, which, along with modern human population genetics, provides an idea of which mutations are tolerated and which are likely pathogenic. The effects of mutations, from single-nucleotide polymorphisms to the loss of full genes, and epigenetic changes that modify gene expression in single cells, tumors, and individuals are being investigated in ever-increasing numbers. These efforts have made it easier to identify how both common and rare variants, which may differ among populations, affect the genetic architecture underlying complex traits and disease.

This special issue examines well-earned successes in applying human genomic research toward understanding human evolution, cancer, polygenic traits, and functional genomics. Yet, there is still much ground ahead to cover.

Yevgeniya Nusinovich, John Travis, and Lila Guterman also contributed to this special issue.

INSIDE

NEWS

Baby steps p. 1460

REVIEWS

From variant to function in human disease genetics p. 1464

Discovery and implications of polygenicity of common diseases p. 1468

Mapping genomic and epigenomic evolution in cancer ecosystems p. 1474

Insights into human history from the first decade of ancient human genomics p. 1479

RELATED ITEMS

EDITORIAL p. 1419
NEWS STORY p. 1430
PERSPECTIVES pp. 1440 & 1442
POLICY FORUM p. 1444
BOOKS *ET AL.* p. 1449
RESEARCH ARTICLE p. 1499
PODCAST

The visualization on the left is the same as on the cover except at a scale of 1,800,000 bases per centimeter. The three panels on the right separate out layers of information: (Left) chromosomes 1 through 22 as well as X and Y, (Middle) chromosome features and staining density, (Right) genes associated with Mendelian disorders and genetic mutations found in cancers.

BABY STEPS



Sequencing every newborn's genome to detect diseases faces ethical and practical obstacles, but the United Kingdom is pushing ahead with a major test

By **Jocelyn Kaiser**

In 2016, a girl named Cora Stetson was born in Boston. Within 48 hours, hospital staff pricked her heel to get a drop of blood to look for molecules that signal dozens of rare genetic diseases—a test required for all U.S. newborns. Because Cora's parents had agreed to enroll her in a study, a researcher also took blood for a much broader test—one that scoured her genome for about 1500 disease genes.

The genetic information proved crucial. Although the standard test flagged a disorder involving a B vitamin–processing enzyme called biotinidase, a follow-up test was negative, and her pediatrician concluded Cora did not have the disorder. But the genome test revealed she indeed had mutations causing biotinidase deficiency—a mild form that nonetheless could result in "bad eyesight and struggling in school," says Cora's mother, Lauren Stetson, a theater educator. Cora now takes a daily biotin tablet and is a "spunky, crazy, sassy" kindergartner, Stetson says.



Genome sequencing showed Cora Stetson (center and opposite page) inherited an enzyme deficiency from her parents. Now 5, she takes a supplement and is thriving.

Cora's case illustrates the promise of sequencing the entire genomes of newborns: uncovering a bounty of genetic information that could identify infants needing treatment and improve health later in life. "Five to 7% of people are born with a rare disease, and many could be treated very early in their life" if the disease is detected, says Richard Scott, chief medical officer of Genomics England, a government-funded company. Genome sequencing could help. "The costs have come down so much that we're now at a tipping point where it's wrong not to."

Genomics England hopes to test that promise in a pilot research project involving up to 200,000 babies. Though it would initially look for the genes of rare childhood diseases, it would also store the genome data for later, when it could help predict drug sensitivities and risks for adult diseases such as cancer. Some U.S. researchers are also eager to add genomewide sequencing to newborn screening. One impetus is to find babies who could benefit from a growing wave of gene-based treatments for devastating, often fatal childhood diseases, such as Sanfilippo syndrome, a metabolic disorder that causes brain damage.

But the U.K. program has an advantage: The country's national health care system already uses whole genomes in clinical care. In the United States, sequencing every newborn's genome is probably still a long way off. Even with inexpensive technology, newborn genome screening on a countrywide scale could take complex infrastructure and hundreds of millions of dollars, if not billions. Some companies already market newborn tests that sequence many genes or the whole genome, at a cost of several hundred to a couple thousand dollars. But those tests are likely to benefit only relatively well-off families.

In every country, ethical and practical questions abound, including which disease genes to test for and whether testing should be done by default. In fact, an ethics group funded by the U.S. National Institutes of Health (NIH) warned in a 2018 report that the evidence to date "does not support genomewide sequencing of all babies at birth." The report noted that the health consequences of many mutations are unknown, and many genetic diseases remain untreatable. Instead of genomewide sequencing, U.S. disease advocacy groups and clinical geneticists have focused on speeding up the sluggish existing national system for screening newborns.

Newborn genome sequencing may become attractive as costs come down, but it's not easy or straightforward, says bioethicist Barbara Koenig of the University of California, San Francisco, a co-author of the 2018 report. "The genome is so, so much more complicated than it appears."

THE IDEA OF READING a newborn's genome dates back at least to the first draft of the human genome, released in 2001. In a TV interview aired that year, Francis Collins, then director of the National Human Genome Research Institute, predicted it would be "feasible" within 20 years to produce a "kind of report card analysis" from a baby's DNA sequence. In 2010, NIH held a workshop to plan four pilot projects to explore newborn genome sequencing.

One project, led by Stephen Kingsmore, now at Rady Children's Hospital-San Diego, has proved an overwhelming success: sequencing critically ill newborns to find out whether they have a genetic disease. For example, in October 2020, a couple brought its inconsolable 5-week-old boy to the Rady emergency room; a computerized tomography scan showed brain abnormalities. Kingsmore's team then found in the baby's genome a mutation for a severe B vitamin metabolic disorder, and soon the boy was quietly drinking vitamin-spiked formula; he's now healthy. The condition likely explained his sister's death as an infant 9 years earlier, Kingsmore's team reported on 3 June in *The New England Journal of Medicine*.

In June, Kingsmore reported at an online

PHOTOS: KEN RICHARDSON; (OPPOSITE PAGE) KEN RICHARDSON

NIH-sponsored meeting on gene therapy that across 23 studies in the past decade by his and other groups, genome sequencing led to a genetic diagnosis for 36% of 1839 seriously ill children, mostly infants. In 533 patients—29% of the total—the findings led to changes in medical care, which saved some babies' lives.

Because of results like those, several countries, including Australia and England, are making genome sequencing for very sick newborns routine, and California and Michigan have approved Medicaid coverage for the test. Genome sequencing "is becoming a new form of care for critically ill neonates," said Diana Bianchi, director of the National Institute of Child Health and Human Development, at the meeting.

Other NIH pilot studies tested genome sequencing as a screening tool for all babies, healthy and sick, comparing it with standard newborn screening. That program began in the United States in the 1960s to identify phenylketonuria (PKU), a metabolic disorder that leads to intellectual disability unless infants eat a special diet. States now screen for PKU and about 30 to 70 other treatable disorders, using mostly biochemical tests. Whole-genome sequencing could look for those single-gene diseases and hundreds more that don't now have a biochemical test, such as neonatal diabetes, hemophilia, and a kidney condition called cystinosis.

But the technique is not infallible. The NIH-funded research and related studies found that sequencing the whole genome or protein-coding DNA misses 12% or more of cases detected by newborn screening. That's because the sequencing misses some genetic changes, and analysts may disregard others—even ones associated with a newborn disease—if that change hasn't been shown to be harmful. But the studies also suggested the two methods could be powerful if combined because sequencing could confirm an ambiguous biochemical test result, as it did with baby Cora.

Cora was part of a pilot project called BabySeq that highlighted another complication of newborn sequencing: It turns up mutations that may never affect health. A team co-led by Robert Green of the Harvard University–affiliated Brigham and Women's Hospital looked for disease mutations in some 1500 genes in 127 healthy babies and 32 sick babies. About 8% of apparently healthy babies and 9% overall had mutations for a childhood genetic disorder—"startling" proportions, Green says. And 88% of the babies were carriers of a genetic disease, a finding that might unsettle parents even though their babies only had one of the two copies of the mutation necessary to get sick.

Among the 10 healthy babies with disease genes, only Cora later had clinical test results indicating she had the condition and should be treated. Seven others had risk genes for heart disease, and a few with slightly abnormal results on heart function tests will be monitored. But many people with those mutations never develop symptoms. "We are plowing new ground here," Green says, by finding mutations whose effects don't always cause illness.

Despite that uncertainty, BabySeq parents, including the 15 families whose babies had an unexpected risk gene for childhood disease, did not experience substantial anxiety or disrupted family bonding from learning that result, according to survey results published in August in *JAMA Pediatrics*. "Many parents would much rather know about these risk factors than not know," Green says. That finding comes with a large caveat, however: Only 7% of parents invited to participate in BabySeq accepted, and—like Cora's parents—most were white and well-educated. To learn how a broader population responds to newborn sequencing, NIH is funding BabySeq2, which will enroll 500 ethnically and socioeconomically diverse families in Boston, New York City, and Birmingham, Alabama.

IN THE UNITED KINGDOM, which has embraced using whole-genome sequencing in routine health care, the public has already weighed in on newborn sequencing. This summer, Genomics England and the U.K. National Screening Committee released the results of a consultation with 130 diverse members of the public. They supported newborn sequencing if parents gave consent—and if they received results only for treatable or preventable childhood diseases.

Genomics England intends to follow those principles in its large pilot project, which could screen for up to 600 genetic diseases that can cause symptoms in early childhood. Those diseases will all have treatments, if not cures, and may include vitamin B6–dependent epilepsy and familial Diamond-Blackfan anemia, a red blood cell disorder. Planners hope to enroll as many as 200,000 newborns over several years, Scott says—a sizable fraction of the 600,000 babies born annually in the country.

The pilot project has the support of Genetic Alliance UK, which includes many groups advocating for patients with rare diseases. Director of Policy Nick Meade says the groups see whole-genome sequencing as a way to rev up the United Kingdom's current screening program, which only tests for nine diseases.

To critics, however, "There are massive ethical and cost issues," as psychiatric geneticist David Curtis of University College London puts it. He is concerned that identifying disease variants that will never make some babies seriously ill will lead to needless testing and family anxiety. He also worries the cost of newborn genome screening—perhaps $900 per baby, or $540 million per year—would be too high for the potential payoffs, and notes newborns can't give consent to storing their genome where it could potentially be accessed by companies as part of agreements with Genomics England. "That baby, in 18 years' time, is he going to be happy that somebody took his whole-genome sequence and put it in a database?" Curtis asks.

ADVOCATES OF GENOMEWIDE screening agree that plenty of uncertainties need to be resolved, including whether to give parents results for mutations that may not be pathogenic and whether to test for incurable diseases, such as fragile X syndrome, which causes intellectual disability. One parent told the U.K. consultation that he was glad *not* to know his son had Duchenne muscular dystrophy before symptoms appeared—they had four happy, worry-free years.

Still, says Tiina Urv, of the National Center for Advancing Translational Sciences, basically everyone at NIH's June meeting on gene therapy "recognized that whole-

A bigger net

Standard newborn screening using biochemical tests flags several dozen genetic disorders. Genome sequencing could screen for thousands more.

	STANDARD SCREENING	GENOME SEQUENCING
What is analyzed?	Mostly metabolites that reflect a disease	All 3 billion DNA base pairs or the 1% encoding proteins
Number of disorders	Ranges from two to about 70 depending on country or U.S. state	Potentially more than 5000 known single-gene diseases
Method	Mass spectrometry. Additional tests for some disorders	Next-generation DNA sequencing
Drawbacks	Can only detect diseases that result in a biochemical change	Misses some cases found with standard screening
Cost (in the United States)	About $30 to $200	Up to $900, not including analysis

ILLUSTRATIONS: K. FRANKLIN/*SCIENCE*



Standard newborn screening, done with blood from a heel prick, could vastly expand with DNA sequencing.

PHOTO: RANDY RISLING/*TORONTO STAR* VIA GETTY IMAGES

genome sequencing [of all infants] is the way to go in the future." They discussed stories like that of a San Diego couple who came home from the hospital in June 2019 with a seemingly healthy baby and then got a phone call that brought terrifying news: Standard newborn screening had found that little Fitz Kettler lacked a functioning immune system. Fitz was referred to Kingsmore's group for genome sequencing, which revealed he had Artemis-deficient severe combined immunodeficiency (SCID), a rare disease that could kill him within a year.

Fortunately, in San Francisco, researchers were testing a gene therapy for Artemis-deficient SCID. Fitz was enrolled in the trial and received a transplant of his own bone marrow cells, genetically modified to correct the mutation. His family still isolates him to avoid COVID-19, but he now has functioning immune cells and is "flourishing," his mother, Christina Eagle-Kettler, said at the June meeting. Early diagnosis through sequencing was key, she said. "I can't imagine not wanting to help make that happen" for other families.

Yet adopting newborn sequencing in the United States has practical obstacles. One is that adding a single new disease to the current screening program takes years, followed by more time to move the disease to state screening lists, dimming prospects for adding hundreds of diseases that genome sequencing could screen for. University of North Carolina, Chapel Hill, clinical geneticist Cynthia Powell chairs the committee that oversees the national list. She says the committee is looking at ways to move more quickly—for example, by adding classes of diseases at once.

Some researchers also fear adding genome sequencing to U.S. newborn screening could cause a backlash against the existing program. Newborn screening is essentially mandatory, like childhood vaccinations. (Only rarely do parents opt out.) The goal of preventing severe childhood diseases across the population "is so important," Koenig says. Adding genes that don't always lead to serious disease or that lack a clearly effective treatment, along with storing the baby's entire genome, could turn parents off the program, Powell says. "I can just see potentially jeopardizing the whole newborn screening system and not even being able to identify conditions such as PKU. I think we have to be very careful," she says.

The United States also lacks the infrastructure for universal newborn sequencing, such as an obvious place to store the genome data for the 3.7 million babies born each year, says clinical geneticist Marc Williams of the Geisinger Health System. And the state labs that perform newborn screening today may not be capable of interpreting whole genomes. The process "would have to be superstreamlined," says clinical geneticist Melissa Wasserstein of the Albert Einstein College of Medicine.

In the United Kingdom, by contrast, much of the infrastructure is already in place. Thanks to the 100,000 Genomes Project, a study that over the past decade used sequencing to diagnose or treat patients with rare diseases, the country already has a research database and seven laboratories that offer whole-genome tests within the National Health Service.

The U.S. health care system would also need to provide genetic counseling and possibly treatment to families who receive alarming sequencing results. "Our health care system is already overwhelmed with the conditions currently tested for with standard newborn screening," Powell says.

For now, U.S. researchers are pushing ahead with demonstration projects. Starting next year, for example, researchers plan to seek parental consent to screen as many as 20,000 North Carolinian newborns for 100 or more genetic diseases by using a DNA test that only sequences specific genes. Wasserstein heads a study called ScreenPlus that over 5 years will test 175,000 babies born at hospitals in ethnically diverse New York City neighborhoods for 14 severe disorders not on the standard screening list. One goal is to see whether the estimated 20 to 40 babies with the diseases have better health outcomes than those diagnosed later. The project will also ask parents for their views on newborn screening, including the use of genomewide sequencing. That input "will hopefully inform how to roll out newborn whole-genome sequencing in a sensitive way," Wasserstein says.

Lauren Stetson says she understands the high stakes. Newborn genome sequencing "has a gravity and weight to it," she says, because it can deliver life-altering news. But the information is important to have. "I want to be part of it. We want to sequence Cole," Cora's 1-year-old brother. "I was like, Dr. Green, I have another one!"

But she cautions that treatment must be easy to find after newborn sequencing. When she and her husband learned Cora's genetic results, they were immediately referred to a genetic counselor and a specialist in biotinidase deficiency. "The doctors, the information was all lined up for us. It made all the difference in the world," she says. If newborn sequencing "does become commonplace, the support system needs to be there." ■

REVIEW

From variant to function in human disease genetics

Tuuli Lappalainen[1,2]* and Daniel G. MacArthur[3,4,5]*

Over the next decade, the primary challenge in human genetics will be to understand the biological mechanisms by which genetic variants influence phenotypes, including disease risk. Although the scale of this challenge is daunting, better methods for functional variant interpretation will have transformative consequences for disease diagnosis, risk prediction, and the development of new therapies. An array of new methods for characterizing variant impact at scale, using patient tissue samples as well as in vitro models, are already being applied to dissect variant mechanisms across a range of human cell types and environments. These approaches are also increasingly being deployed in clinical settings. We discuss the rationale, approaches, applications, and future outlook for characterizing the molecular and cellular effects of genetic variants.

Since the completion and publication of the human genome approximately two decades ago (*1*), the primary bottleneck in human genetics has shifted from the discovery of genetic variation to the large-scale characterization of the associations between genetic variation and phenotype, and more recently to the characterization of the causal mechanisms by which genetic variation influences human biology. Of the more than 200,000 genetic variants confidently linked with human complex traits through genome-wide association studies (GWAS), the majority remain mechanistically uncharacterized (*2*). Furthermore, because the majority of these variants fall in noncoding regions of the genome, there is often uncertainty about which gene is responsible for their biological effects (*3*). Similarly, of the nearly 1 million entries in the ClinVar database (*4*) of variants identified in patients with severe genetic disease, 47% are classified as having either uncertain or conflicting annotations, indicating a lack of clarity about variants' impact on molecular function and disease (*5*).

Resolving this pervasive uncertainty will require more accurate and scalable methods to unravel the molecular processes by which genetic variation influences phenotype. Such approaches will increase the accuracy of genetic diagnosis and prediction, especially for rare or unique disease-causing variants, by enhancing the direct inference of variants likely to disrupt the normal function of critical genes. These approaches will also accelerate therapeutic development, not only by highlighting the gene products directly involved in the causation of disease but also by revealing the direction of effect, the relevant cell types, and the overall biological pathways by which a gene influences disease risk.

[1]Science for Life Laboratory, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden. [2]New York Genome Center, New York, NY, USA. [3]Centre for Population Genomics, Garvan Institute of Medical Research, and UNSW Sydney, Sydney, New South Wales, Australia. [4]Centre for Population Genomics, Murdoch Children's Research Institute, Melbourne, Victoria, Australia. [5]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA.
*Corresponding author. Email: tuuli.lappalainen@scilifelab.se (T.L.); daniel.macarthur@populationgenomics.org.au (D.G.M.)

The genetic architecture of human disease

Over the past decade, it has become clear that the genetic architecture of human traits—the characteristics of the variants that causally influence those diseases—is highly variable. This architecture is the consequence of the interplay between the demographic and selective forces from our species' evolutionary history (*6*) and the differing distribution of the locations of variants within our genome. The action of natural selection means, in general, that variants with large effects on biological function and disease risk will tend to be removed from the population; any such variants will thus tend to be very rare, and if a variant is common, it is thus extremely unlikely to have large effects on disease (*7*). The structure of the genome means that variants with the largest effect will tend to be found in or very close to protein-coding regions, as changes in protein sequence and structure can affect a protein's normal function across all or most cell types where that gene is expressed.

In contrast, variants falling outside protein-coding regions (noncoding variants) are much more likely to be biologically neutral, and those that do affect biology likely do so through more subtle and potentially cell type–specific changes in gene regulation (*8*).

The interplay between evolutionary and biological forces explains the architecture of human traits that has emerged from large-scale genomic analysis. Causal variants across a wide range of disease fall along a spectrum of variant frequency and effect size (*9*) (Fig. 1A). One end of the spectrum represents very rare variants with large effects on human biology, virtually all of which are protein-coding or near-coding, including the majority of variants identified as causal for rare, severe genetic diseases. On the other end of the spectrum are variants that have small individual effects on disease risk, but which [because they experienced weaker selection (*10*)] have been able to reach higher frequencies in the population and generally have an impact on regulatory elements in noncoding sequences (*11*). Because of their higher frequency, these common noncoding variants contribute to a large fraction of the total explained heritability (genetic contribution to risk) of common diseases such as type 2 diabetes (*12*). These trends are not absolute, however: Rare variation, especially in protein-coding regions, nonetheless contributes disproportionately to the genetic architecture of common complex traits (*12*, *13*), and common variants can also contribute to the risk and severity of rare diseases (*14*, *15*).

This genetic architecture has consequences for the optimal approaches for discovering causal variants (Fig. 1A). Rare coding variants

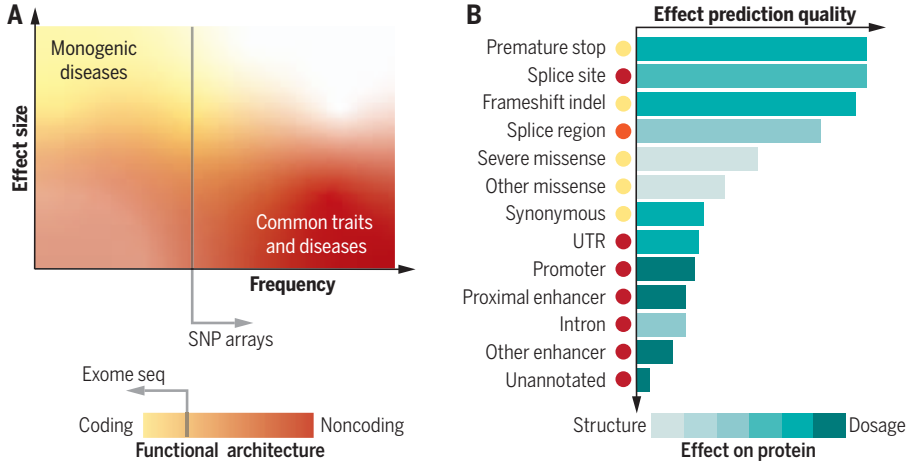


Fig. 1. Genetic variant effects on disease and gene function. (**A**) Functional genetic architecture of human disease ranging from a spectrum of monogenic diseases driven by rare predominantly coding variants with a strong effect on disease to common diseases and traits whose heritability is dominated by noncoding common variants with small effect sizes. However, these classifications are not exclusive: Noncoding causal variants in Mendelian disease exist, as do rare variants affecting common traits and diseases. (**B**) Quality of variant effect predictions by functional annotation class. The color of the bars indicates whether the variant classes tend to affect protein structure or dosage or whether a given class of variants can affect both—for example, via nonsense-mediated decay and splicing. The color of the labels indicates coding and noncoding annotation.

of large effect are currently efficiently identified with exome sequencing studies (an approach that cost-effectively sequences the protein-coding segments of the genome), either in families (for variants with extremely large effect sizes, such as those underlying most "monogenic" disorders) or in large case/control studies (*16*, *17*). In contrast, common variants with smaller effect sizes are cost-effectively discovered with large-scale case/control studies with genotyping arrays or low-coverage whole-genome sequencing. Note that although the differential architecture of common and rare disorders is clear, its degree may have been exaggerated by the different methods predominantly used in each field that have hindered the discovery of noncoding variants in rare disease and rare variants in complex disease (*12*). Over the next decade, deep whole-genome sequencing will be applied to much larger cohorts of patients and population samples, providing an unbiased view of genetic architecture and improving the discovery of its shared features across different types of diseases (*13*, *18*).

Genetic architecture also alters the optimal approaches for using functional data to understand variants' effects on phenotype. In common disease, associated GWAS loci typically contain a large number of variants in linkage disequilibrium—meaning that they are close enough to each other that they are typically inherited together—of which only a small proportion are causal and affect the regulatory function of the genome (*19*). For such cases, a functional dissection typically requires assaying noncoding regulatory elements, often in a highly cell type–specific or environment-specific fashion. For rare disease variants, and for the small proportion of common disease variants found in coding regions, assays for functional impact generally focus on changes in the structure or dosage of mRNA transcripts and their encoded proteins. Below, we discuss high-throughput approaches spanning the complete range of potential biological impacts that will be needed for a comprehensive understanding of human variation.

Observational approaches for characterizing functional variation

The foundation of genome-wide variant interpretation is the functional annotation of the genome, which has been established by projects including ENCODE (*20*) Gencode (*21*), and Roadmap Epigenomics (*22*), which have measured the biochemical activity of the genome in multiple cell types. The resulting annotation of genes, transcripts, and regulatory elements has enabled high-quality predictions of the most severe classes of likely gene-disrupting variants (Fig. 1B), which are particularly important for rare disease interpretation and therapeutic target discovery (*23*).

However, for most variant classes, accurate predictions of variant effects cannot be derived solely from genome annotations. Even missense variants, where the amino acid changes are easy to identify and annotate, typically lack high-quality predictions of their effects on protein structure and function (*24*). For noncoding variants dominating complex disease heritability, predictions of regulatory effects currently have limited practical applicability and performance. This limitation is due to our generally poor knowledge of the complex regulatory code of the genome and the fact that such analyses are dependent on the specific cellular context. Variants affecting transcript splicing—an important class of post-transcriptional effects in both rare and common disease (*25*, *26*)—fall between these extremes of prediction performance. Specifically, variants in canonical splice sites are easy to annotate, and predictive models have reasonable performance (*27*). However, changes in splicing (e.g., due to synonymous variants or variants deep in introns) remain difficult to predict (*27*).

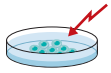
Approach		Biological systems	Genetic variation	Phenotyping	Interpretation
Genetic analysis of human populations 	+	Primary cells in native physiological environments	Exact representation of natural variation	Access to physiological traits	Direct relevance to human physiology and health
	−	Complexity of tissues, limited access to dynamic cellular states	Limited to study of existing genetic variation in the population	Confounders of molecular and cellular phenotypes limit interpretability	Interindividual variation and confounders complicate interpretation of causality
Cellular models in vitro 	+	Access to and control of cell differentiation and state	Perturbation of any loci with different variants of diverse effect sizes	Cleaner data from molecular and cellular phenotyping	Cleaner data of causal molecular and cellular mechanisms
	−	Limited set of cell lines or other models in cell culture media	SNP editing difficult; assays may lack genomic context	Physiological phenotypes not available	Potentially limited generalizability to physiological systems

Fig. 2. Functional characterization approaches relying on analysis of genetic variation in human populations versus experimental perturbations of the genome or its function.

Common genetic effects on gene regulation can be inferred by mapping quantitative trait loci (QTLs) for proximal molecular features. The most common type of molecular QTLs are cis-eQTLs (cis-associated QTLs associated to gene expression levels), but QTLs can also be mapped for splicing, chromatin accessibility, and protein levels, for example (*28*). Human eQTL studies have been pursued for more than 10 years, including by consortia such as GTEx (*25*) and eQTLGen (*29*), and have generated comprehensive catalogs of cis-eQTLs and splicing QTLs (sQTLs) (*30*) in diverse human tissues. Together with chromatin accessibility QTL (caQTL) and protein QTL (pQTL) maps, these data have been valuable in pinpointing potential causal genes and molecular mechanisms in individual loci.

However, the QTL approach has limitations. It can only capture the effects of existing common variation where linkage disequilibrium obscures the identity of the actual causal variants (*31*), and this makes it suboptimal for scanning the functional effects of individual nucleotides. Furthermore, when the same variant or haplotype affects the expression of multiple genes or molecular traits, it is difficult to assess which one(s) have downstream effects that causally contribute to the disease association. The success of large studies in identifying eQTLs or sQTLs for nearly all human genes has made this problem of low specificity painfully clear (*32*). It is further exacerbated by the poor representation of non-European ancestries in QTL datasets (*30*), which also limits the utility of current QTL catalogs in the interpretation of GWAS data from diverse ancestries. In cases of rare variant analysis, the association-based QTL approach is not applicable, but analogous characterization of rare variants with likely molecular effects can identify situations where a rare variant coincides with an individual's status as a population outlier for gene expression, splicing, or allelic expression (*33*).

Currently, QTL datasets are limited by their derivation primarily from a limited number of easily accessible tissues (such as blood) or post mortem tissue samples (*30*). These biospecimens are a mixture of multiple, often unknown, steady-state cell types. This can lead to poor resolution and detection power for genetic regulatory effects that are active in specific cell types or dynamic cell states that may be particularly important in disease risk (*34*). Thus, expanding QTL studies to computationally resolved cell types (*35*) and single-cell datasets from tissue samples is a major ongoing focus of the field (*36*), especially because many cell types are not represented by available cell lines. A complementary approach to analyzing developmental lineages and environmental responses is to extract primary cells or cell lines from donors, induce pluripotency, and grow or differentiate the cells in vitro to study molecular phenotypes by bulk or single-cell sequencing. This has allowed access to key cell types that

are next to impossible to sample at scale, and also provides a renewable and scalable source for molecular assays (*28*, *37*–*39*). However, not all primary cell types can be easily extracted or differentiated, and thus the map of regulatory associations of standing human variation across all cell types is still very incomplete.

Genome perturbation approaches

In addition to analysis of naturally existing genetic variation in human populations, we now can perform scalable experimental introduction of genetic variants in a cellular model system, followed by molecular phenotyping of their effects, often at a single-cell level (*40*). The assays that pair potential regulatory sequences with sequencing barcodes in artificial DNA sequences, so that the ability of the different alleles to drive gene expression can be compared (*41*, *42*), are the best suited to test thousands of variants. However, this approach only detects the cis-regulatory potential of short sequences. Pooled CRISPR-based perturbation of the genome or its regulatory activity, followed by single-cell RNA sequencing and cellular phenotyping (*43*–*46*), has enabled the characterization of target genes in cis and downstream cellular effects of regulatory or coding variants.

These approaches are rapidly expanding in scale and scope, including deeper phenotyping of diverse cellular functions by high-throughput imaging, multi-omics, as well as application not only to cell lines but also to organoid and animal models. However, there are substantial limitations to the scale and cost of precision genome perturbation, cellular phenotyping, and the availability of good biological model systems. The latter represents a major challenge: Easy-to-use cancer cell lines represent a limited range of human cell types and differ from primary cells in many ways. The extent to which different model systems can recapitulate variant effects is largely unknown, poorly defined, and likely to vary not only between the cells used but also between different types of variants (*25*). For example, a variant introducing a premature stop in a constitutively expressed exon of a ubiquitously expressed gene is likely to have much lower context specificity than a variant in a cell state–specific enhancer. Thus, integration of experimental approaches with human genetic studies remains valuable as a way to leverage the benefits of both approaches (Fig. 2).

The clinical impact of variant-to-function approaches

The molecular diagnosis of patients affected by severe genetic diseases has accelerated rapidly over the past decade as a result of the increased availability of exome and genome sequencing, improved methods for variant interpretation, and global data sharing. Current standards for clinical variant annotation (*47*) require the integration of multiple classes of evidence to assess the probability that a variant is pathogenic (disease-causing). These fall broadly into three categories: (i) evidence that the variant is rare in the general population, (ii) evidence that the variant has previously been seen in other patients with similar clinical appearance, and (iii) evidence (either direct or indirect) that the variant has a functional impact on a gene previously implicated in the same class of disorders. Assessment of the first two categories has been empowered by the availability of large databases of population variation (*48*) and of variants previously seen and interpreted in disease patients (*4*, *49*). However, assessing the probability that a variant is biologically damaging is much more challenging and less standardized.

Because most variants found in patients have not yet been subjected to well-calibrated functional assays, most clinical interpretation relies on indirect measures of functional impact, such as evolutionary conservation (*47*). Historically, when functional information was incorporated, it was generated with custom assays for specific variants, often were only weakly quantitative, and lacking information regarding the probability of obtaining the same result from randomly selected variants in the same gene (*50*). However, large-scale genomic approaches to inferring variant function are now feasible, and are beginning to be applied in clinical settings.

Such approaches fall into two broad classes. First, there are direct assays run on patient tissue samples or cell lines, which provide a direct functional readout of variants discovered in that patient. The most widely used functional genomic assay for the diagnosis of rare genetic diseases is bulk transcriptome sequencing. This approach can be used to assess the impact of variants on gene expression and transcript splicing; its use improves diagnosis rates over those derived from DNA sequencing alone, although these improvements vary by disease and tissue type [>30% when applied to muscle biopsies from severe muscle disease patients (*51*); 16% in a study of fibroblast samples from patients with mitochondrial disease (*52*)]. Proteomic analysis has also proven valuable both for the assessment of individual candidate genes (*53*) and in combination with transcriptomics for diagnosis (*54*). Other genome-scale technologies including metabolomics and epigenomics have also been deployed to improve the identification of causal genes (*55*).

These approaches are powerful, but access to disease-relevant tissues and cell types in a clinical setting remains challenging; post mortem tissue collection is of limited use, and differentiation of induced pluripotent stem cells from patient tissue comes with a considerable cost. Other hurdles to widespread clinical adoption include the heterogeneity of clinical samples, the lack of standardization of assays and analysis pipelines, and the need for experienced data interpretation.

A second broad category of approaches consists of in vitro assays in which an artificially created model of the variant is tested for functional impact in a well-established cell line, as outlined above. These approaches can be rapidly scaled into multiplexed assays of variant effect (MAVEs), allowing for the rapid testing of candidate variants, and potentially of all possible genetic variants for a given gene (*40*). Such assays have already proved effective in inferring the pathogenicity of variants for a subset of well-established disease genes, in some cases eclipsing standard approaches to variant interpretation (*56*). For MAVEs to accurately infer pathogenicity, the precise functional assay used must be both highly scalable and tightly correlated with the disease-relevant function(s) of the target gene and the mechanisms of action of pathogenic variants, and thus will be extremely challenging or impossible for a subset of disease-causing variation. Although MAVE data do not yet exist for the vast majority of clinically relevant genes, current approaches include assays exquisitely customized to individual genes, as well as generalizable assays of protein stability (*40*, *57*). Efforts are now under way to systematically develop well-calibrated assays for clinically relevant genes, and to harmonize and release the resulting data (*58*).

The future of variant function

The human population, through explosive growth, has performed a comprehensive saturation mutagenesis experiment on itself. It is now the case that any single base substitution that is compatible with life is expected to be present somewhere among the nearly 8 billion living humans (*59*). Humanity has thus, in effect, done many of the natural experiments required to understand our own genotype-phenotype map; this leaves geneticists to catalog the outcomes of those experiments, and to leverage both observational and experimental approaches to understand the mechanisms by which variants alter biology. Over the next decade, unless hampered by major obstacles to data sharing, increasingly massive cohorts of disease patients and deeply phenotyped population samples should produce the requisite catalog, and sophisticated and scalable tools will be applied to characterize the underlying functional mechanisms in both research and clinical settings (Table 1). These approaches should generate large, standardized datasets that can be combined with machine learning tools to provide predictions of variant effects on biology and disease.

We believe that the consequences of this process will be profound. In the clinical setting, the assessment of the potential effects of

Table 1. Key variant-to-function challenges in common and rare disease, and approaches to address them.

● Common disease ● Rare disease

	Challenge	Approaches
Causal variant	Linkage disequilibrium	Fine-mapping, improvements from diverse populations and functional priors; scalable experimental assays
	Few observations of any given variant ($N \sim 1$)	Gene and variant prioritization with larger datasets and functional priors; scalable experimental assays
Immediate functional effect	"Regulatory code" of variant effects on regulatory elements mostly unknown	Regulatory region annotations, chromatin/splicing QTL mapping, scalable experimental assays, predictive models
	Difficulty of predicting effects on protein 3D structure and function	Predictive models, scalable experimental assays
Causal gene and effect direction	Difficulty of enhancer target prediction	eQTL mapping, predictive models, scalable experimental assays
	Small number of patients with genetic disruption of a given gene	Data aggregation; better priors for gene and variant function
Downstream cellular effects	Small effect sizes of individual loci; relevant cell type/state often unknown	Cellular phenotype characterization in (selected) population samples or model systems
	Few observations of any given variant ($N \sim 1$)	

a sequence variant on disease risk will become increasingly quantitative, driven by access to three critical strands of information: the frequency of that variant across hundreds of millions of humans; the phenotype observed in any other human observed to carry it; and the results of well-calibrated experiments assessing the precise impact of that variant on gene function. In effect, variant interpreters should have access to a quantitative functional look-up table for any possible variant in a wide variety of clinically relevant genes. This will empower the diagnosis of genetic disease while also improving the prediction of risk for clinical phenotypes that have not yet manifested (e.g., in the assessment of genomes sequenced before or at birth, or those with environmental responses).

For complex traits, rapid in silico assessment of likely functional effects should speed the identification of causal variant(s) responsible for multiple uncharacterized risk loci and their likely target genes in cis, and will increase the proportion of accurate predictions of the affected pathways and relevant cell types. It will also enhance individual risk prediction, as genome-wide functional annotation can already improve the accuracy and portability of polygenic scores for complex disease (*60*, *61*), and may allow the partitioning of individual risk into components driven by different physiological mechanisms and potentially amenable to different therapeutic approaches (*62*).

The functional characterization of human variants is likely to continue to increasingly influence the development of new therapeutics or repurposing of already established drugs. Drug targets with genetic evidence have higher success rates (*63*, *64*). For a trait-associated gene affected by multiple functional variants, robust functional data of their effects can provide allelic series, allowing for functional dosages for each gene to be linked to phenotypic outcomes (*65*) and directly support specific therapeutic hypotheses. The convergence of genome-wide disease signals into genetically implicated pathways (*66*) should provide targets beyond individual GWAS genes. Characterization of the cell types where genetic risk effects manifest—which are often different from the organs affected by the disease—is also important for being able to target causal pathways rather than treating consequences. Understanding the molecular mechanisms underlying pathology will also pinpoint pathways that mediate the contribution of environmental and developmental risk factors to disease, and thus advance the integration of genetic and nongenetic factors in predicting and treating disease.

Finally, a shift from pure genetic approaches toward high-throughput functional data as evidence for variant interpretation will also contribute to greater equity in the utility of genomic medicine across communities. Allele frequencies and linkage disequilibrium patterns reflect population history and are thus highly population-specific; hence, they can contribute to biases due to the current underrepresentation of non-European samples (*67*). In contrast, although individual variants affecting disease risk often vary by ancestry, their molecular and cellular mechanisms are more likely to be shared among all humans. This hypothesis should be systematically addressed by performing functional experiments in model systems that represent different ancestries, sexes, ages, and environmental exposures.

Despite these potentialities, the challenges ahead for high-throughput biologists are daunting. Although human genetic datasets are increasingly rich, the power to assess trait associations for all genetic variants—let alone their combinations—has its limits. The complexity of highly correlated and easily confounded human phenotype measurements greatly complicates this task. The paths by which genetic variants lead to phenotypic outcomes are complex and dynamic: They take place across multiple cell types, in response to many environmental conditions and developmental cues. They act through a wide range of molecular and cellular processes, and potentially interact with other genetic variants present in the same individual, including somatic variants that have arisen during cell division over that person's lifetime. As such, developing an understanding of the physiological mechanisms of disease will require deep molecular characterization across interacting cell types and dynamic cell states, as well as access to informative biospecimens and/or biological model systems. Across these, we must harness the power of a diverse and complementary toolkit, ranging from interrogation of natural genetic variation to genome perturbations, followed by multimodal cellular assays. Although we are optimists about the transformative power of these approaches, it is always important to be humble in the face of biology's complexity, and we acknowledge that interactions and higher-order effects can complicate conclusions drawn from simple experimental models.

Different approaches, such as those outlined above, have unique advantages and limitations. Embracing a diversity of scalable approaches, including highly generalizable assays (of gene expression, protein structure, and protein stability), well-calibrated assays of gene-specific functions (such as enzyme activity), and analytical methods for integrating complex information across molecular classes and cell types, will be necessary to deeply characterize the mechanisms by which genome sequence shapes human biology.

REFERENCES AND NOTES

1. International Human Genome Sequencing Consortium, *Nature* **409**, 860–921 (2001).
2. M. Claussnitzer *et al.*, *Nature* **577**, 179–189 (2020).
3. M. T. Maurano *et al.*, *Science* **337**, 1190–1195 (2012).
4. M. J. Landrum *et al.*, *Nucleic Acids Res.* **46**, D1062–D1067 (2018).
5. ClinVar; www.ncbi.nlm.nih.gov/clinvar.
6. G. Sella, N. H. Barton, *Annu. Rev. Genomics Hum. Genet.* **20**, 461–493 (2019).
7. J. Zeng *et al.*, *Nat. Commun.* **12**, 1164 (2021).
8. E. A. Boyle, Y. I. Li, J. K. Pritchard, *Cell* **169**, 1177–1186 (2017).
9. M. I. McCarthy *et al.*, *Nat. Rev. Genet.* **9**, 356–369 (2008).
10. E. C. Glassberg, Z. Gao, A. Harpak, X. Lan, J. K. Pritchard, *Genetics* **211**, 757–772 (2019).
11. H. K. Finucane *et al.*, *Nat. Genet.* **47**, 1228–1235 (2015).
12. S. Gazal *et al.*, *Nat. Genet.* **50**, 1600–1607 (2018).
13. P. Wainschtein *et al.*, bioRxiv 588020 [preprint]. 11 June 2021.

14. A. R. Harper *et al.*, *Nat. Genet.* **53**, 135–142 (2021).
15. S. S. Shringarpure *et al.*, medRxiv 21258643 [preprint]. 16 June 2021.
16. O. Zuk *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E455–E464 (2014).
17. J. E. Posey *et al.*, *Genet. Med.* **21**, 798–812 (2019).
18. E. Turro *et al.*, *Nature* **583**, 96–102 (2020).
19. O. Weissbrod *et al.*, *Nat. Genet.* **52**, 1355–1363 (2020).
20. ENCODE Project Consortium, *Nature* **489**, 57–74 (2012).
21. A. Frankish *et al.*, *Nucleic Acids Res.* **47**, D766–D773 (2019).
22. Roadmap Epigenomics Consortium, *Nature* **518**, 317–330 (2015).
23. E. V. Minikel *et al.*, *Nature* **581**, 459–464 (2020).
24. B. J. Livesey, J. A. Marsh, *Mol. Syst. Biol.* **16**, e9380 (2020).
25. GTEx Consortium, *Science* **369**, 1318–1330 (2020).
26. R. Soemedi *et al.*, *Nat. Genet.* **49**, 848–855 (2017).
27. K. Jaganathan *et al.*, *Cell* **176**, 535–548.e24 (2019).
28. Y. I. Li *et al.*, *Science* **352**, 600–604 (2016).
29. U. Võsa *et al.*, bioRxiv 447367 [preprint]. 19 October 2018.
30. N. Kerimov *et al.*, bioRxiv 924266 [preprint]. 9 January 2021.
31. Q. S. Wang *et al.*, *Nat. Commun.* **12**, 3394 (2021).
32. A. N. Barbeira *et al.*, *Genome Biol.* **22**, 49 (2021).
33. X. Li *et al.*, *Nature* **550**, 239–243 (2017).
34. B. D. Umans, A. Battle, Y. Gilad, *Trends Genet.* **37**, 109–124 (2021).
35. S. Kim-Hellmuth *et al.*, *Science* **369**, eaaz8528 (2020).
36. M. van der Wijst *et al.*, *eLife* **9**, e52155 (2020).
37. D. Liang *et al.*, *Nat. Neurosci.* **24**, 941–953 (2021).
38. M. C. Ward, N. E. Banovich, A. Sarkar, M. Stephens, Y. Gilad, *eLife* **10**, e57345 (2021).
39. B. P. Fairfax *et al.*, *Science* **343**, 1246949 (2014).
40. G. M. Findlay, *Hum. Mol. Genet.* ddab219 (2021).
41. R. Tewhey *et al.*, *Cell* **165**, 1519–1529 (2016).
42. M. Kircher *et al.*, *Nat. Commun.* **10**, 3583 (2019).
43. A. Dixit *et al.*, *Cell* **167**, 1853–1866.e17 (2016).
44. J. A. Morris *et al.*, bioRxiv 438882 [preprint]. 8 April 2021.
45. M. C. Canver *et al.*, *Nature* **527**, 192–197 (2015).
46. B. Adamson *et al.*, *Cell* **167**, 1867–1882.e21 (2016).
47. ACMG Laboratory Quality Assurance Committee, *Genet. Med.* **17**, 405–424 (2015).
48. K. J. Karczewski *et al.*, *Nature* **581**, 434–443 (2020).
49. H. L. Rehm *et al.*, *N. Engl. J. Med.* **372**, 2235–2242 (2015).
50. D. G. MacArthur *et al.*, *Nature* **508**, 469–476 (2014).
51. B. B. Cummings *et al.*, *Sci. Transl. Med.* **9**, eaal5209 (2017).
52. V. A. Yépez *et al.*, medRxiv 21254633 [preprint]. 5 April 2021.
53. N. J. Lake *et al.*, *Am. J. Hum. Genet.* **101**, 239–254 (2017).
54. R. Kopajtich *et al.*, medRxiv 21253187 [preprint]. 12 March 2021.
55. K. Kerr *et al.*, *Orphanet J. Rare Dis.* **15**, 107 (2020).
56. G. M. Findlay *et al.*, *Nature* **562**, 217–222 (2018).
57. K. A. Matreyek *et al.*, *Nat. Genet.* **50**, 874–882 (2018).
58. The Atlas of Variant Effects (AVE) Alliance: Understanding Genetic Variation at Nucleotide Resolution (2021); https://zenodo.org/record/4989960.
59. B. H. Shirts, C. C. Pritchard, T. Walsh, *Trends Mol. Med.* **22**, 925–934 (2016).
60. T. Amariuta *et al.*, *Nat. Genet.* **52**, 1346–1354 (2020).
61. P. M. Visscher, L. Yengo, N. J. Cox, N. R. Wray, *Science* **373**, 1468–1473 (2021).
62. M. S. Udler, M. I. McCarthy, J. C. Florez, A. Mahajan, *Endocr. Rev.* **40**, 1500–1520 (2019).
63. M. R. Nelson *et al.*, *Nat. Genet.* **47**, 856–860 (2015).
64. E. A. King, J. W. Davis, J. F. Degner, *PLOS Genet.* **15**, e1008489 (2019).
65. R. M. Plenge, E. M. Scolnick, D. Altshuler, *Nat. Rev. Drug Discov.* **12**, 581–594 (2013).
66. N. Sinnott-Armstrong, S. Naqvi, M. Rivas, J. K. Pritchard, *eLife* **10**, e58615 (2021).
67. A. Buniello *et al.*, *Nucleic Acids Res.* **47**, D1005–D1012 (2019).

ACKNOWLEDGMENTS

We thank A. Pinho for her contributions to the preparation of this manuscript. **Funding:** Supported by the KTH Royal Institute of Technology, New York Genome Center, and NIH grants R01GM122924, R01HL142028, R01AG057422, and R01MH106842 (T.L.) and the Garvan Institute of Medical Research, the Murdoch Children's Research Institute, and NIH grants U24HG011450 and U01HG011755 (D.G.M.). **Author contributions:** T.L. and D.G.M. contributed equally to the writing of this manuscript; author order was determined by a coin toss. **Competing interests:** T.L. is an advisor for Variant Bio, Goldfinch Bio, and GSK and has stock in Variant Bio. D.G.M. is a founder with equity in Goldfinch Bio and is an advisor to Insitro, Variant Bio, GSK, and Foresite Labs.

10.1126/science.abi8207

REVIEW

Discovery and implications of polygenicity of common diseases

Peter M. Visscher[1]*, Loic Yengo[1], Nancy J. Cox[2], Naomi R. Wray[1,3]

The sequencing of the human genome has allowed the study of the genetic architecture of common diseases: the number of genomic variants that contribute to risk of disease and their joint frequency and effect size distribution. Common diseases are polygenic, with many loci contributing to phenotype, and the cumulative burden of risk alleles determines individual risk in conjunction with environmental factors. Most risk loci occur in noncoding regions of the genome regulating cell- and context-specific gene expression. Although the effect sizes of most risk alleles are small, their cumulative effects in individuals, quantified as a polygenic (risk) score, can identify people at increased risk of disease, thereby facilitating prevention or early intervention.

Competing genetic models of disease

Paradoxically, many common diseases occur in individuals with no family history of the disease, yet relatives of those affected have an increased risk compared with an average person from the population (*1*). These empirical observations can be reconciled with several genetic models that differ in the expected number of contributing causal variants. The ability to directly measure from DNA the degree of genetic variation among individuals has allowed for tests that could support or refute proposed genetic models. Initial experimental designs mimicked approaches that had been successful in identifying causal variants of single-gene Mendelian disorders that assume simple recessive or dominant inheritance, driven by the hypothesis that the genetics of common disease would be explained by relatively few genes harboring major mutations that individually or in combination cause disease. However, by and large, these experimental designs were not successful in leading to robust and replicable associations for polygenic traits despite their success in mapping mutations causing Mendelian disease.

Sequencing of the human genome (*2*, *3*) has led to the characterization of genomic variation between individuals and populations and expanded our genomic technologies, computational methods, and experimental designs. One important innovation has been the development of the single nucleotide polymorphism (SNP) array or SNP-chip technology, in which hundreds of thousands of DNA variants (mostly SNPs) can be scored robustly in a single assay (*4*). The low cost of this technology drove widespread uptake of the experimental design of genome-wide association studies (GWAS), which test for SNP allele frequency differences between affected (cases) and unaffected (controls) individuals and have revolutionized our understanding of the genetic basis of common polygenic diseases (Box 1).

GWAS has illuminated the polygenic nature of common diseases. Hundreds to thousands of genomic loci with a robust association with any one disease or trait have been identified (*5*). As GWAS sample sizes increase, the number of loci detected with statistical significance also increases in a linear manner (*6*). However, as sample sizes grow, the effect sizes of the new significantly associated loci tend to become very small. In combination with the known limitations of GWAS (Box 1), this complicates efforts to use raw results from GWAS to gain insights into disease pathogenesis and pathophysiology.

In addition to GWAS, genome-wide surveys of rare variants have been conducted either by population studies of the protein-coding regions of the genome (the exomes) in case–control studies or by enrichment of de novo variants in cases from family studies (*7*). Empirical results are consistent with those from GWAS in that rare, deleterious mutations in many genes can contribute to disease risk and that, on average, the exome contributes only a small proportion of risk in the population. Because every individual harbors de novo mutations, the disease relevance of the mutations requires the affected gene to have been identified in several families together with annotations of functional relevance, and loss-of-function variants have the most relevance (*8*). The GWAS design is not restricted to disease and has become a powerful tool with which to study phenotypes such as gene expression, DNA methylation, and histone modification at the level of bulk tissue and individual cells.

Two decades ago, the debate about the genetic architecture of common disease was divided between proponents of a Mendelian model, in which a few rare variants of large effect size were segregating in the population, and advocates of a many-gene polygenic model, in which the cumulative burden of risk variants in a person determines their risk (*9*). Halfway models postulated that effect sizes

[1]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland 4072, Australia. [2]Vanderbilt Genetics Institute and Division of Genetic Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN 37232, USA. [3]Queensland Brain Institute, University of Queensland, Brisbane, Queensland 4072, Australia.
***Corresponding author. Email: peter.visscher@uq.edu.au**

would be intermediate and that such variants would likely be found among protein-altering polymorphisms (*9*). Many proponents of sequencing the human genome anticipated a rapid identification of disease genes with a subsequent fast track in developing therapeutics. These predictions were mostly wrong. Instead, GWAS studies provided empirical support for a highly polygenic architecture of common disease and evidence for a role of polygenic variation in rare diseases (Table 1).

A spectrum of genetic architectures

Genetic architecture encompasses the total number of functional variants and the frequency and effect size of each of them that affect a trait or disease. Causal variants of common diseases have frequencies across a wide allelic spectrum, from rare to common. Many genetic effects act in an additive manner with respect to an underlying disease liability, and disease results from the accumulated burden of risk factors (Fig. 1), resulting in nonadditive (e.g., multiplicative) determination of whether disease occurs. Empirical results from GWAS have shown that although common diseases are highly polygenic, their genetic architecture differs. In particular, some diseases (e.g., coronary artery disease and type 2 diabetes) have risk variants of large effects that segregate at appreciable frequency in some populations (*10*) combined with a polygenic tail, whereas for other diseases (e.g., schizophrenia), no such variants have been detected (Table 1). The disease burden (liability) model is consistent with observed risk for diseases in which there are variants of large effect. For example, the consequence of large effect variants can be amplified or attenuated by polygenic risk background and other factors contributing to risk (Table 1). In general, common adult-onset disorders are not caused by rare large effect variants because such mutations often lead to more severe and early clinical symptoms that are diagnosed in childhood. A wide range of polygenic genetic architectures underlie diseases (*11*). Thus, the genetic architecture of disease, whether rare or common, is consistent with the hypothesis that all diseases are polygenic but some have mutations of large effect (Table 1).

Polygenicity and natural selection

Recent studies (*11*–*13*) have used theory, simulations, and empirical analyses to highlight negative selection as a prominent evolutionary force shaping the degree of polygenicity of diseases as well as its variation across genomic loci. Negative selection primarily acts by removing alleles with a strong deleterious effect on fitness from the population. Therefore, negative selection directly controls the number of mutations segregating in the population, thus imposing an upper bound on disease polygenicity depending on a variant's relation to fitness. More generally, negative selection also prevents mildly deleterious alleles on fitness from reaching high frequency in the population (*14*). Thus, a signature of negative selection induces an inverse relationship between the magnitude of SNP effects and their minor allele frequency (*13*, *15*). Statistical methods using GWAS summary statistics have exploited this property to show evidence of negative selection for a wide range of traits and diseases, including late-onset diseases (*12*, *13*, *16*). However, because GWAS only partially inform on causal variants, those methods have limited ability to quantify a direct relationship between the strength of selection and the degree of polygenicity. Indeed, the estimated number of SNPs with nonzero effects on a trait (*12*, *17*, *18*), another commonly used statistic to quantify polygenicity, is correlated with but is not an unbiased estimator of the

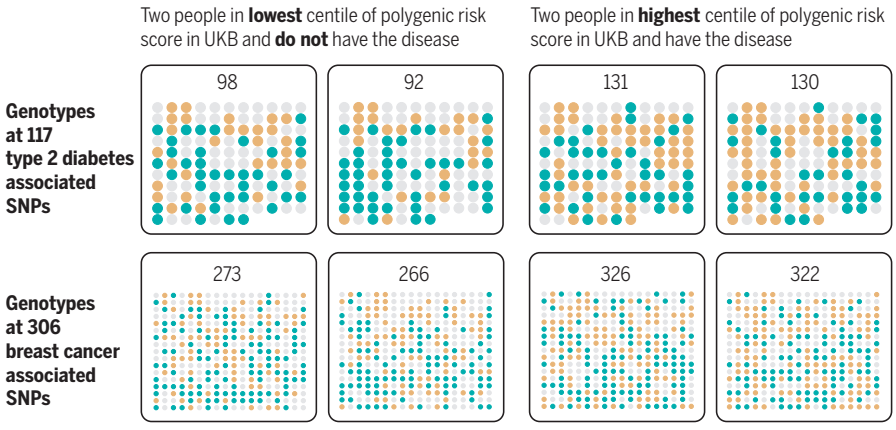


Fig. 1. Visualizing the polygenicity of common disease. Each panel shows the number of risk alleles (gray dots, zero risk alleles; blue dots, one risk allele; and red dots, two risk alleles) for an individual from the UK Biobank (UKB; a volunteer cohort of 500,000 people age 50 to 70 years) at each of 117 type 2 diabetes and 306 breast cancer disease-associated loci. The total number of risk alleles for the individual is listed at the top of each panel. For each disease, two individuals were selected with disease and in the top 1% of the count of the number of risk alleles and two controls without disease from the bottom 1%. The panels show that (i) every individual has a distinct risk profile, (ii) controls carry risk alleles at many loci, and (iii) cases have, on average, only a small increase in the risk variant burden. Selecting the top and bottom 1% of this population on the number of risk variants gives mean numbers of risk variants of 131 and 98 for type two diabetes and 325 and 269 for breast cancer, respectively, with disease prevalence of 7% versus 2% (type 2 diabetes) and 19% versus 2% (breast cancer). Therefore, although the mean number of risk variants between cases and controls is generally small, individuals who have a high burden of risk variants have greatly increased risk of disease. PRSs are not diagnostic; most individuals in the top 1% of the PRS distribution do not have disease and, conversely, some in the bottom 1% do.

Box 1. GWAS.

GWAS is an experimental design to detect population-wide associations between genetic markers and trait loci for applications in plant and animal breeding and for human disease (*22*). In human GWAS, the association between each of the hundreds of thousands to millions of SNPs identified within a sequenced pool of individuals and a complex trait is tested for statistical significance from a sample of individuals in the population. For common diseases, the sample consists of cases and controls, whereas for a quantitative trait, it can comprise a random set of individuals spanning the distribution of the trait.

GWAS relies on population-level LD (the nonrandom association between alleles at different genomic loci) between alleles at observed genetic markers and unobserved disease loci on the same chromosome. LD is created by geographic and evolutionary forces in humans, particularly those caused by finite ancestral population size and natural selection (*53*). The amount of LD in most human populations is such that scoring a few hundred thousand well-chosen loci is sufficient to "tag" the entire genome for unobserved common variants, even when the total number of common variants is on the order of tens of million. By design, GWAS leads to statistically significant associations between loci and traits. However, in practice, neither the variant or variants that cause the statistical association signal nor the gene (or other functional unit) that is their target is identified because the causal variant(s) may not be genotyped or imputed, a noncausal variant in LD (correlated) with a causal variant may have a stronger association signal, or the actual causal variants may be far from the target gene, e.g., 1 Mb for a SNP associated with body mass index (*48*). This is a known but important limitation of the design.

Table 1. Polygenic contributions in a range of genetic architecture contexts.

Genetic architecture context	Disease example	Detail of polygenic role
Dominant Mendelian disease, in which a single variant (rare in the population) leads to high frequency of disease in some families	Huntington's disease, a progressive neurodegeneration disorder defined by elevated number of copies of a CAG trinucleotide repeat located in the first exon of the *HTT* gene	Age of onset is associated with the number of copies of the CAG repeat. The Genetic Modifiers of Huntington's Disease Consortium have used the GWAS paradigm studying age of onset (*42*), identifying to date 21 associated loci, providing evidence of natural compensatory mechanisms with a polygenic architecture that are considered the most likely avenue to identify drug target opportunities for this classic Mendelian disease.
	Lynch syndrome, a colorectal cancer defined by pathogenic variant in one of four genes: *MLH1*, *MSH2*, *MSH6*, or *PMS2*	In the UKB, the probability of colorectal cancer by age 75 was 2% for those without Lynch syndrome pathogenic variants and 35% for those with variants (0.15% in UKB). For those with Lynch syndrome variants, the probability of colorectal cancer by age 75 ranged from 11 to 80% for the lowest to highest percentile of PRS constructed from colorectal cancer GWAS, respectively (*57*).
	Familial hypercholesterolemia (FH), a high-cholesterol disease leading to high risk of coronary artery disease (CAD) defined by pathogenic variant in one of three genes: *LDLR*, *APOB*, or *PCSK9*	In the UKB, the probability of CAD by age 75 was 13% for those without FH pathogenic variants and 41% for those with variants (0.67% of CAD cases). For those with FH variants, the probability of CAD by age 75 ranged from 18 to 78% for the lowest to highest percentile of PRS constructed from CAD GWAS, respectively (*57*).
Presumed monogenic mutations	Developmental delay disorder, children with neurodevelopmental disorders with clinical features sufficiently severe to implicate a de novo or inherited monogenic mutation	In addition to reporting severe, moderate, or mild intellectual disability and high rates of ASD* (17%), patients in the Deciphering Developmental Delay (DDD) study (N = 6987) were on average 0.72 SD shorter, weighed 0.15 SD less, and had a head circumference that was 1.20 SD smaller than the age- and sex-adjusted population (*58*). In addition to these mean differences associated with developmental delay, variation in birth weight, height, and ASD within the DDD cohort was highly associated with their corresponding PRS. Finally, common variants that are not correlated with monogenic mutations explained ~8% of risk of DDD. Therefore, polygenic variation can affect both the risk and clinical features in a disorder that is usually considered to be monogenic.
Large-effect variants in common diseases	Breast, ovarian, and prostate cancers, in which pathogenic or likely pathogenic variants in *BRCA1* and *BRCA2* are found in a small proportion of patients	In the UKB, the probability of breast cancer by age 75 was 10% in those without pathogenic or likely breast cancer pathogenic variants (found in 3.1% of the 17,344 controls), compared with 35% in those with such variants (found in 9.1% of the 1920 women with breast cancer). For those with pathogenic variants, the probability of breast cancer by age 75 ranged from 13 to 76% for the lowest to highest percentile of PRS constructed from breast cancer GWAS, respectively (*57*). Qualitatively similar results have been reported for ovarian and prostate cancers (*59*, *60*).
	Glaucoma, in which pathogenic variants in the myocilin *MYOC* gene account for 2 to 4% of primary open-angle glaucoma, the most common subtype of glaucoma; p.Gln368Ter is the most common variant	In the UKB, of those with the pathogenic Gln368Ter *MYOC* variant (n = 965), those in the lowest tertile of PRS for glaucoma had a 2% risk of glaucoma by age 60, but the risk was sixfold higher in those in the top tertile of PRS (*56*).
Disease definition affecting genetic architecture	ASD and/or developmental delay	For psychiatric and behavioral disorders, diagnoses are allocated on the observable clinical symptoms that overlap only partially with genetic architecture. For example, ASD is three to four times more common in boys, indicating a female protective effect. As a result, girls with ASD have higher mean ASD PRS than boys. Girls with ASD are more likely to have large effect or de novo variants than boys because boys with these variants are more likely to be diagnosed with developmental delay.
	Common disorders of onset in early adulthood, such as schizophrenia and type 2 diabetes	Although large effect variants are observed in late-onset (post-reproductive years) disorders such as Alzheimer's disease (*APOE*), amyotrophic lateral sclerosis (*SOD1*), and age-related macular degeneration (*CFH*), such variants are notably absent from common adult onset behavioural and psychiatric (*61*) or metabolic disorders (*62*). Large effect variants may lead to earlier clinical presentations, resulting in childhood-onset syndromic diagnoses.

*ASD, autism spectrum disorders.

number of causal variants. The bias of that estimator depends on the disease heritability but also on the strength of selection (*12*). Other factors influencing bias are properties of the study design, such as sample size and criteria for selecting SNP density and frequency (*17*). Despite all of these limitations, estimates of the number of nonzero effect SNPs suggests that common diseases are, on average, less polygenic than other traits, including anthropometric and brain-related phenotypes. Schizophrenia is a well-recognized outlier, with one of the largest degrees of polygenicity reported across traits and diseases (*11*, *18*). One feature of the observed polygenicity of complex traits, including common diseases, is that it tends to be distributed widely across the entire genome. This implies that there are many sites in the genome that, when perturbed through mutation, can contribute to a particular trait or disease.

Other methods (*19*–*21*) have also leveraged the availability of large sets of GWAS-associated SNPs to quantify small changes in risk allele frequencies in response to new selection pressures. These methods have shown evidence of directional selection on health-related traits such as blood pressure and body mass index (*20*). Furthermore, the refined characterization of genes that are constrained by natural selection has provided new ways to annotate GWAS findings, thereby improving their biological interpretation.

In summary, recent contributions have shed new light on the causal link between natural selection and polygenicity and improved GWAS-based approaches to refine detection and quantification of the effect of natural selection on the genetic architecture of complex diseases. We anticipate that quantification of polygenicity from whole-genome sequencing (WGS) data will reveal more detailed resolution and more reliable predictions regarding the evolution of complex disease (*7*).

Polygenic risk scores continue to mature

An exciting development in human genetics has been the use of GWAS data to create polygenic predictors for complex traits and quantify their association with disease and other outcomes in independent data. A polygenic risk score (PRS) is a sum of the number of risk variants at multiple genomic loci, weighted by their effect sizes that have been estimated from independent data. Harnessing genome-wide linkage disequilibrium between marker genotypes and trait loci and performing a genomic prediction was first suggested in the context of plant and animal breeding and has been used widely in agriculture (*22*). Like GWAS, polygenic prediction is entirely based upon association. The increase in sample sizes in GWAS has facilitated polygenic predictors that have effect sizes of risk as large as known monogenic mutations that are clinically actionable (*10*) (Fig. 2). Because genetic factors

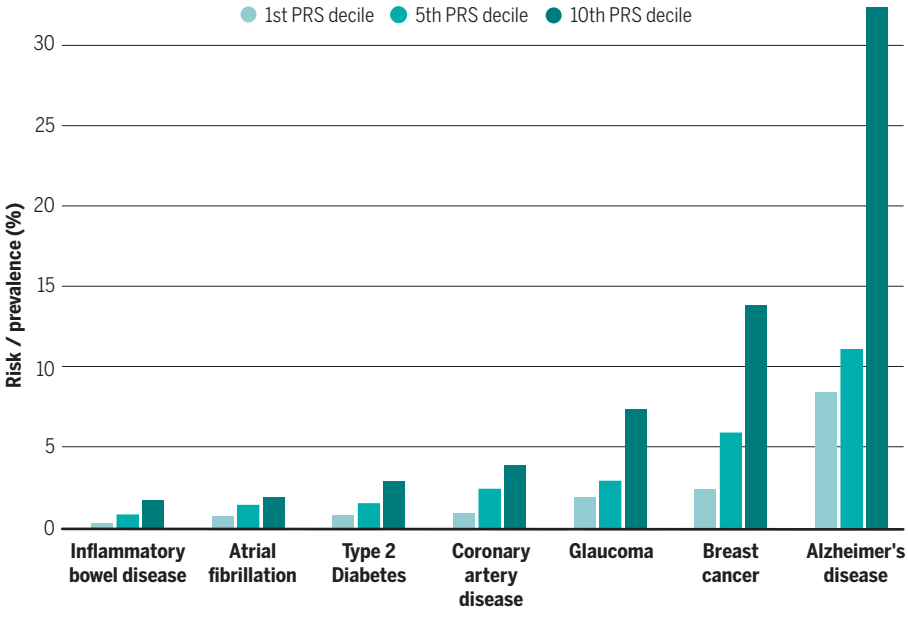


Fig. 2. Risk as a function of polygenic predictor percentiles. Risk or prevalence of inflammatory bowel disease, atrial fibrillation, type 2 diabetes, coronary artery disease, glaucoma, breast cancer, and Alzheimer's disease is shown as a function of the 1st, 5th, and 10th decile of their PRS predictor. Glaucoma and breast cancer prevalence were measured at age 70 and 60 years, respectively. Prevalence of inflammatory bowel disease, atrial fibrillation, type 2 diabetes, and coronary artery disease was measured in the UKB at ages ranging from 40 to 70 years and the prevalence of Alzheimer's disease at age 65 years. Improvements in the accuracy of PRSs will increase the difference in lifetime risks across the percentiles. Data are from (*10*, *54*–*56*). Rare *BCRA1/BRAC2* mutations affecting breast cancer risk are not captured in the common variant PRS. For Alzheimer's disease, the common APOE ε4 variant explains approximately half of the risk in the 10th decile group (*54*).

are not the only contributors to risk of common diseases, polygenic predictors are not expected to be diagnostic. Nonetheless, they have utility similar to other predictors used in medicine. When more of the trait variance is accounted for in polygenic predictors, the difference in lifetime prevalence between individuals at high and low predicted risk should increase. The accuracy of polygenic prediction can be improved by better modeling of the distribution of effect sizes (*23*–*25*). Resources with PRS information have been created to ensure reproducibility in research and translation (*26*, *27*).

Polygenicity spurs new experimental designs

Past studies on human disease genetics relied on family data, whereas the current focus is on populations (e.g., GWAS). However, the distinction is artificial because populations consist of individuals that are related to each other in some way, either with recent or more distant common ancestors. Combining the power of population genomics studies with the naturally occurring controls that pedigree studies offer can address the degree of heritable versus environmental impacts on phenotype (nature versus nurture) with new designs, whereas environment and genetic effects were previously difficult to disentangle. For example, GWAS data on families were used to separate a direct (causal) genetic effect from an indirect genetic effect that reflects the parental "nurturing" environment (*28*). We anticipate that such designs are important for understanding common diseases and their risk factors, especially those with a behavioral component. The availability of tens of thousands of GWAS-associated SNPs offers opportunities to answer scientific questions about natural selection, human behavior, and causality. For example, as the predictive accuracy of PRSs increases, we can now quantify the flux of trait-associated alleles between geographical areas and thereby elucidate some of the correlates of population migration (*29*, *30*). More widely, there is a growing trend to embed genomics into social sciences research (*31*), and this requires careful consideration of the implications and applications of such studies (*32*).

Another exciting new experimental design facilitated by GWAS discoveries is Mendelian randomization, which attempts to mimic a randomized controlled trial by using SNPs that are associated with a modifiable risk factor (e.g., body mass index) to quantify its causal effect on disease risk (e.g., type 2 diabetes) (*33*). Testing for putative causality between exposure and disease is important in epidemiology and public health because it leads to evidence-based policy interventions and health advice. Mendelian randomization is a cost-effective way to prioritize hypotheses that could then be addressed by formal but expensive clinical

trials or to provide causal evidence when such trials would be unethical. More generally, the widespread availability of GWAS summary statistics for diseases and other complex traits, including molecular omics phenotypes, has led to new powerful analysis paradigms by mimicking experiments in which all traits would be measured on all individuals without the necessity for individual-level data (*34*).

GWAS to date have mostly relied on SNP-chip technology combined with statistical imputation using a sequenced reference panel. Although they were successful for populations of European ancestry, they have been less so for other less-sampled populations (*35*), likely because of differences in allele frequencies and linkage disequilibrium (LD) patterns between human populations (*35*). Large biobank-style studies across multiple ancestries are needed to understand human gene-disease associations within and between populations and to improve polygenic predictors across ancestries.

WGS offers more information for GWAS analysis to discover disease loci because all genomic variants in a sample are observed. This is in contrast to GWAS-by-chip, in which only a small fraction of all possible variants are genotyped and additional variants are imperfectly imputed using a statistical algorithm. Therefore, GWAS-by-WGS can, in principle, improve the ability to detect associations between rare sequence variants and disease. However, the relationship between effect size and allele frequency (*12*) implies that sample sizes to detect rare variants from WGS data would need to be very much larger than what would be needed for the detection of common variants through GWAS-by-chip, because detection power is a function of both the effect size and allele frequency and also on the total number of variants tested, which is much larger in GWAS-by-WGS. Burden tests, in which the number and composition of multiple rare variants in a gene are jointly assessed for disease association, may overcome some of these statistical limitations. Sufficiently powered studies performing GWAS-by-WGS will improve polygenic risk scores because the combination of rare and common variants will explain more of the genetic variation contributing to the trait.

From genetic to biological models

Researchers have, understandably, tried to simplify the complexity of polygenicity and proposed biological models that are compatible with a large number of genomic sites associated with common diseases. One such model proposes that the diagnostic label of a given disease can reflect multiple distinct underlying biological pathways, each polygenic (*36*). In this way, having high polygenic risk for any one of the underlying biological pathways is sufficient to lead to disease. This model implies that gene-gene interactions on the scale of liability for the labeled disease are

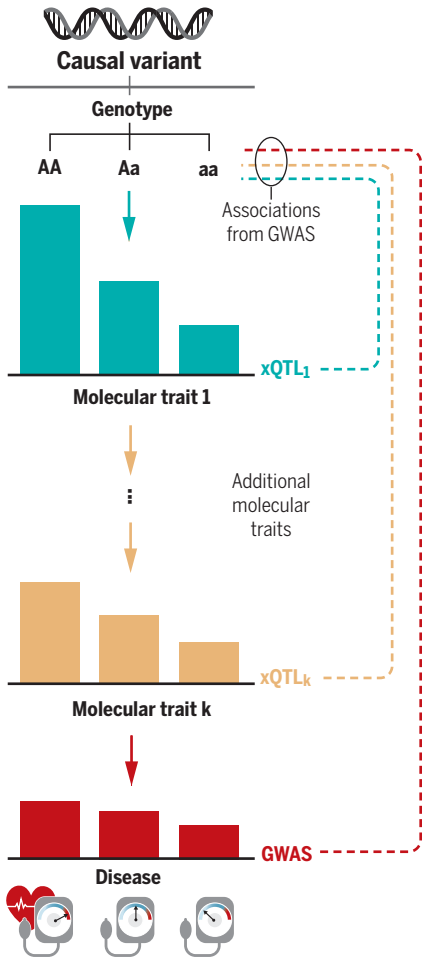


Fig. 3. Causal path from a genotype to disease through multiple molecular mediators. The red dotted line shows that a genotype is associated with disease (hypertension), the primary observation from GWAS. The yellow and blue dotted lines show that the same genotype is associated with multiple molecular mediators (1 to k). The genotype is a quantitative trait locus (QTL) for moderators x, where x can encompass a wide range of mechanisms, including gene expression, splicing, and epigenetic marks such as methylation and protein levels, both in cis and trans. The schematic shows how a causal variant can contribute to disease risk through many different routes. There could be a single molecular mediator (k = 1, e.g., gene expression) or multiple ones, and if there are multiple molecular traits involved, the order may be different for different causal variants. Finally, as observed with empirical data, effect sizes tend to become smaller (from top to bottom, as shown) because at each step additional factors come into play that dilute the contribution from the genetic variant. Researchers use statistical colocalization and mediation analyses to find the most likely causal paths using GWAS data on disease and moderators. Data on additional molecular mediators are being collected to fill in missing links in the genotype to the disease causal pathway.

important, which appears to be inconsistent with subsequent empirical results pointing to additivity on this scale. Additivity on the liability scale is an important property because it implies that effect sizes are transferrable across individuals within a population, so a disease can be studied and predicted irrespective of the specific combination of risk alleles in a person (Fig. 1).

More recently, the omnigenic model was proposed to reconcile polygenicity with biological function (*37*). The model proposes a mechanistic explanation for why so many variants spread across the genome can be responsible for the genetic variation between people observed for complex traits and common diseases. In brief, genetic variants with proximal, cis effects on peripheral genes perturb the regulation of a smaller class of core genes through distal, trans-regulatory networks. Thus, studying peripheral genes helps to unravel the biology of a trait because it explains the trans-regulatory context within which core genes sit; however, disease-specific core genes are the most important for disease-specific research. One component of the hypothesis, the identification of core genes, has been tested for three molecular traits measured in blood, urate, insulin-like growth factor-1, and testosterone, and association signals were interpretable in terms of the physiology of the traits and concentrated near core genes and core pathways (*38*). However, even for these molecular traits, it was not possible to test the trans-regulatory network component of the hypothesis because this requires larger sample sizes than currently available, suggesting that it may be impossible to test the omnigenic model for complex traits such as common diseases.

A disease-specific palette model was proposed for type 2 diabetes (*39*) as a simple visualization of the polygenic model. This model suggests that although there are multiple biological processes involved in diabetes (such as insulin secretion, insulin resistance, and dyslipidemia) and there are multiple underlying environmental risk factors, individuals with diabetes are likely to have a risk burden across many biological and environmental categories. Many common diseases are chronic, relapsing and remitting. A network dyshomeostasis model has been proposed to describe the cumulative contribution of genetic variants and the environmental factors affecting the gut microbiome that together lead to risk of inflammatory bowel disease (*40*). Statistically, the omnigenic, palette, and dyshomeostasis models are consistent with the additive liability threshold model of common disease, in which probability of disease is multiplicative across all risk factors, both genetic and environmental.

From variant discovery to disease biology

Many GWAS results confirm biology that is already known or suspected, for example, the

role of sodium or potassium ATPase in type 2 diabetes (*41*), which drives the expectation that GWAS results can add to our understanding of biology. In general, the biological inference is more straightforward for individual molecular traits such as blood metabolites than for common diseases. We currently have few maps linking mechanism to medicine. For example, whereas the causal role of Huntington's disease was identified as an expanded genomic repeat in 1993, this knowledge has still not led to new treatments. However, GWAS studies are now being used to identify polygenic variants that provide protection for onset of Huntington's disease on the basis of the age of onset (*42*). GWAS of Crohn's disease (*43*, *44*) helped to define the role of autophagy in this disease and catalyzed research to define the roles of key genes in autophagy and the role of autophagy in its pathophysiology. GWAS of inflammatory bowel diseases has led the field in biological understanding and translation (*40*), perhaps because these diseases are well defined, less heterogeneous, and the affected tissue is easier to biopsy than in other common diseases. Such research has identified genetic variants converging on a multitude of cellular and molecular pathways that regulate homeostasis of the mucosal immune system interacting with the gut microbiome (*40*).

In general, however, it has been difficult to learn new biology solely from GWAS results because of the large number of associated variants, their small effect sizes, and the inherent challenges in moving from the associated SNPs to causal variants to causal genes (or other functional units) and toward the driving disease biology. Also, the effects from associated SNPs to trait may transit through intermediate phenotypes such as transcript levels (Fig. 3) (*8*). Effect sizes of SNPs estimated in GWAS have been integrated with molecular phenotypes such as gene expression, chromatin accessibility, and other functional annotations (*45*) in enrichment analyses as a way to learn more about the most proximal molecular mechanisms affected by GWAS signals (Fig. 3). Greater understanding of the nature of DNA secondary structures (*46*) has provided insights into how a SNP thought to be causal for an established association in or near one gene may exert effects on a gene located at a considerable distance within the genome (*47*, *48*). Furthermore, the impacts and interactions of a panoply of RNA-based functional units are still lacking. Thus, the transition from causal SNP to causal gene or protein remains challenging.

There are similar challenges in understanding how genes alter human biology to drive disease susceptibility. Given the number of genes with functions that are unknown or poorly annotated, the nature of some of these challenges is clear. Knowledge of when and where each gene is most needed for the maintenance of healthy homeostasis is required (*49*, *50*). Studies in Mendelian diseases with large effect mutations have revealed complexities in cell- and tissue-specific effects of genes on human biology affecting disease risk; there will be greater challenges in elucidating similar mechanisms for genes in which the genetic variation contributing to disease risk is substantially smaller.

From our current vantage point, it is clear that genetic studies of common human diseases need more tools to directly investigate the biological links among genetic variants, environment, and disease. When individually small effects coalesce to have organism-wide consequences for health and disease, they do so by working through biological systems that we are learning to characterize at all levels of scale from nuclear to cellular to tissue to organism.

Given the preponderance of noncoding variation associated with polygenic liability to human disease, research is needed to elaborate and measure, over a variety of contexts and at scale, the many ways that RNA affects protein abundance. These approaches are already starting to emerge. For example, through an integrated analysis of genome, transcription, and DNA methylome of kidneys, 479 GWAS associations between SNPs and blood pressure could be connected to 1038 genes expressed in kidneys, identifying druggable targets for hypertension (*51*).

There are many drugs currently used with success for common diseases, such as lithium for bipolar disorder, for which the biological mechanism is not understood. Causal inference through Mendelian randomization, particularly with regard to using genetic information to identify potential drug targets (*52*), could lead to drug repurposing and partially sidestep the need to link GWAS SNPs to mechanism.

As genetics and genomics partner with medical investigations in the diagnosis and treatment of patients, we anticipate many medical centers becoming a source of data to probe how genetic variation relates to the broad phenotypic spectrum of humanity. Challenges lie ahead, but tackling these complex problems provides opportunities to improve all human health. Despite the all-too-human tendency to oversimplify the complexity of biology, recent progress reminds us that we can improve, not just our global understanding of the biology of common human diseases, but also our practical ability to contribute meaningfully to their treatment even as we discover and acknowledge greater complexity in the underlying genetic models.

REFERENCES AND NOTES

1. J. Yang, P. M. Visscher, N. R. Wray, *Eur. J. Hum. Genet.* **18**, 1039–1043 (2010).
2. J. C. Venter *et al.*, *Science* **291**, 1304–1351 (2001).
3. International Human Genome Sequencing Consortium, *Nature* **409**, 860–921 (2001).
4. S. Kim, A. Misra, *Annu. Rev. Biomed. Eng.* **9**, 289–320 (2007).
5. P. M. Visscher *et al.*, *Am. J. Hum. Genet.* **101**, 5–22 (2017).
6. O. Canela-Xandri, K. Rawlik, A. Tenesa, *Nat. Genet.* **50**, 1593–1599 (2018).
7. K. J. Karczewski *et al.*, *Nature* **581**, 434–443 (2020).
8. T. Lappalainen, D. MacArthur, *Science* (2021).
9. N. J. Risch, *Nature* **405**, 847–856 (2000).
10. A. V. Khera *et al.*, *Nat. Genet.* **50**, 1219–1224 (2018).
11. L. J. O'Connor *et al.*, *Am. J. Hum. Genet.* **105**, 456–476 (2019).
12. J. Zeng *et al.*, *Nat. Commun.* **12**, 1164 (2021).
13. A. P. Schoech *et al.*, *Nat. Commun.* **10**, 790 (2019).
14. A. Eyre-Walker, *Proc. Natl. Acad. Sci. U.S.A.* **107** (suppl_1), 1752–1756 (2010).
15. S. Gazal *et al.*, *Nat. Genet.* **50**, 1600–1607 (2018).
16. D. Speed, N. Cai, M. R. Johnson, S. Nejentsev, D. J. Balding; the UCLEB Consortium, *Nat. Genet.* **49**, 986–992 (2017).
17. Y. Zhang, G. Qi, J. H. Park, N. Chatterjee, *Nat. Genet.* **50**, 1318–1326 (2018).
18. D. Holland *et al.*, *PLOS Genet.* **16**, e1008612 (2020).
19. Y. Field *et al.*, *Science* **354**, 760–764 (2016).
20. L. Speidel, M. Forest, S. Shi, S. R. Myers, *Nat. Genet.* **51**, 1321–1329 (2019).
21. A. J. Stern, L. Speidel, N. A. Zaitlen, R. Nielsen, *Am. J. Hum. Genet.* **108**, 219–239 (2021).
22. N. R. Wray, K. E. Kemper, B. J. Hayes, M. E. Goddard, P. M. Visscher, *Genetics* **211**, 1131–1141 (2019).
23. G. Kichaev *et al.*, *Am. J. Hum. Genet.* **104**, 65–75 (2019).
24. L. R. Lloyd-Jones *et al.*, *Nat. Commun.* **10**, 5086 (2019).
25. T. Ge, C. Y. Chen, Y. Ni, Y. A. Feng, J. W. Smoller, *Nat. Commun.* **10**, 1776 (2019).
26. S. A. Lambert *et al.*, *Nat. Genet.* **53**, 420–425 (2021).
27. J. Becker *et al.*, *Nat. Hum. Behav.* (2021).
28. A. Kong *et al.*, *Science* **359**, 424–428 (2018).
29. A. Abdellaoui *et al.*, *Nat. Hum. Behav.* **3**, 1332–1342 (2019).
30. S. Haworth *et al.*, *Nat. Commun.* **10**, 333 (2019).
31. K. P. Harden, P. D. Koellinger, *Nat. Hum. Behav.* **4**, 567–576 (2020).
32. P. Turley *et al.*, *N. Engl. J. Med.* **385**, 78–86 (2021).
33. N. M. Davies, M. V. Holmes, G. Davey Smith, *BMJ* **362**, k601 (2018).
34. B. Pasaniuc, A. L. Price, *Nat. Rev. Genet.* **18**, 117–127 (2017).
35. A. R. Martin *et al.*, *Nat. Genet.* **51**, 584–591 (2019).
36. O. Zuk, E. Hechter, S. R. Sunyaev, E. S. Lander, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1193–1198 (2012).
37. E. A. Boyle, Y. I. Li, J. K. Pritchard, *Cell* **169**, 1177–1186 (2017).
38. N. Sinnott-Armstrong, S. Naqvi, M. Rivas, J. K. Pritchard, *eLife* **10**, e58615 (2021).
39. M. I. McCarthy, *Diabetologia* **60**, 793–799 (2017).
40. D. B. Graham, R. J. Xavier, *Nature* **578**, 527–539 (2020).
41. G. Sweeney, A. Klip, *Mol. Cell. Biochem.* **182**, 121–133 (1998).
42. Genetic Modifiers of Huntington's Disease (GeM-HD) Consortium, *Cell* **178**, 887–900.e14 (2019).
43. J. Hampe *et al.*, *Nat. Genet.* **39**, 207–211 (2007).
44. J. D. Rioux *et al.*, *Nat. Genet.* **39**, 596–604 (2007).
45. H. K. Finucane *et al.*, *Nat. Genet.* **47**, 1228–1235 (2015).
46. J. W. Ho *et al.*, *Nature* **512**, 449–452 (2014).
47. S. Smemo *et al.*, *Nature* **507**, 371–375 (2014).
48. M. Claussnitzer *et al.*, *N. Engl. J. Med.* **373**, 895–907 (2015).
49. L. E. González-Serrano, J. W. Chihade, M. Sissler, *J. Biol. Chem.* **294**, 5309–5320 (2019).
50. M. Sissler, L. E. González-Serrano, E. Westhof, *Trends Mol. Med.* **23**, 693–708 (2017).
51. J. M. Eales *et al.*, *Nat. Genet.* **53**, 630–637 (2021).
52. M. C. Hyman *et al.*, *Hypertension* **77**, 376–382 (2021).
53. M. Slatkin, *Nat. Rev. Genet.* **9**, 477–485 (2008).
54. Q. Zhang *et al.*, *Nat. Commun.* **11**, 4799 (2020).
55. N. Mavaddat *et al.*, *Am. J. Hum. Genet.* **104**, 21–34 (2019).
56. J. E. Craig *et al.*, *Nat. Genet.* **52**, 160–166 (2020).
57. A. C. Fahed *et al.*, *Nat. Commun.* **11**, 3635 (2020).
58. M. E. K. Niemi *et al.*, *Nature* **562**, 268–271 (2018).
59. K. B. Kuchenbaecker *et al.*, *J. Natl. Cancer Inst.* **109**, djw302 (2017).
60. J. Lecarpentier *et al.*, *J. Clin. Oncol.* **35**, 2240–2250 (2017).
61. M. J. Gandal, V. Leppa, H. Won, N. N. Parikshak, D. H. Geschwind, *Nat. Neurosci.* **19**, 1397–1407 (2016).
62. C. Fuchsberger *et al.*, *Nature* **536**, 41–47 (2016).

ACKNOWLEDGMENTS

We thank T. Lin and Y. Wu for help with the figures and three anonymous reviewers for helpful comments and suggestions. Figures 1 and 2 were created using data from the UK Biobank under project no. 12505. Due to space limitations, we were not able to cite many other relevant papers in the literature. **Funding:** This work was funded by the Australian Research Council (grants FL180100072 and DE200100425), the Australian National Health and Medical Research Council (grants 1113400 and 1173790), and the National Institutes of Health (grants U01HG0090086, U54MD010722, and R01MH113362). **Competing interests:** The authors declare no competing financial interests.

10.1126/science.abi8206

REVIEW

Mapping genomic and epigenomic evolution in cancer ecosystems

Toshikazu Ushijima[1]*, Susan J. Clark[2,3]*, Patrick Tan[4,5,6,7]*

Cancer is a major cause of global mortality underpinned by genomic and epigenomic derangements. Here, we highlight the importance of multimodal data integration in understanding the molecular evolution of malignant cell states across the cancer life cycle. The widespread presence of driver mutations and epigenetic alterations in normal-appearing tissues is prompting a reassessment of how cancer initiation is defined. In later-stage cancers, studying the roles of clonal selection, epigenomic adaptation, and persister cells in metastasis and therapy resistance is an emerging field. Finally, the importance of tumor ecosystems in driving cancer development is being unraveled by single-cell and spatial technologies at unprecedented resolution. Improving cancer risk assessment and accelerating therapeutic discovery for patients will require robust, comprehensive, and integrated temporal, spatial, and multilevel tumor atlases across the cancer life cycle.

Cancer is a major cause of global disease-related morbidity and mortality—in 2020, cancers resulted in an estimated 19 million new cases and 10 million cancer deaths. These numbers are expected to increase markedly in the future, particularly in developing countries (*1*). Several large-scale consortia have sought to comprehensively identify cancer genetic drivers (*2*, *3*), and when combined with clinical and commercial programs, more than 1 million tumors have been sequenced to date across various tumor types and subtypes using whole-genome, exome, or targeted panels. Collectively, these efforts have revealed that distinct tumor types often exhibit different levels of mutation burden, with a few genes, such as *TP53* and *KRAS*, mutated often and several genes mutated at low recurrence rates (the so-called long tail) (*4*). Although most of the frequently mutated cancer driver genes have likely been identified, rarely mutated drivers are also relevant because they often converge on common oncogenic pathways (*5*). Mutations in chromatin modifier genes can also elicit changes in the cancer epigenome (*2*, *3*). Beyond protein-coding genes, noncoding genomic alterations can act as drivers by influencing gene regulation (*6*). Whole-genome analysis has further revealed that somatic mutations are not randomly distributed throughout the genome, but are influenced by underlying mutational processes and tissue-specific chromatin contexts (*7*).

Despite impressive scientific progress, previous cancer genome studies have largely focused on isolated bulk biological specimens, such as primary tumors, typically representing a single snapshot of the malignant state. To better understand the temporal dynamics of cancer, we highlight the importance of integrating genomic, epigenomic, and transcriptomic information across the entire cancer life cycle, from normal tissues to premalignancy and from primary tumors to metastases and treatment refractory lesions. We further showcase how emerging single-cell and spatial imaging technologies are integrating facets of the tumor ecosystem, clarifying how cancer cells interact with other cell types in the tumor microenvironment to elicit cancer hallmarks.

Precancer evolution and nascent carcinogenesis

Traditionally, histologically normal tissues were believed to harbor no or few somatic mutations. Using sensitive approaches capable of distinguishing bona fide mutations from sequencing errors, recent studies have shown that many normal tissues, such as the esophagus, colon, bronchial epithelium, and urothelium, can harbor somatic mutations in driver and passenger genes (*8*, *9*). Nonmalignant tissues with histological changes, such as gastric tissue with intestinal metaplasia and inflammatory bowel disease (IBD), can also exhibit somatic mutations (*10*–*13*). Driver mutations in these so-called normal tissues can vary by organ type—in some cases, these mutations can also drive cancers in the same organ, such as *TP53* and chromatin modifier genes. In other cases, driver mutations in a normal tissue are rarely present in cancers of the same tissue, such as *NOTCH1* in the esophagus (*14*, *15*) and interleukin-17 (IL-17) signaling genes in IBD (*11*–*13*). The latter may provide context-specific survival advantages in noncancer tissues but not in cancer samples, which emphasizes that the biological consequences of specific driver mutations are influenced by parallel molecular alterations and the cellular ecosystem (*11*–*15*). Although the number of genetic driver mutations is small, analysis of the substantial number of passenger mutations in these tissues [20 to 100 per exome and 1000 to 10,000 per genome in the IBD-exposed colon (*11*, *13*); 1000 to 4000 per genome in the esophagus (*15*)] enables associations between mutational loads with biological factors, such as aging, and lifestyle factors, including smoking and alcohol consumption. Age-associated increases in somatic mutations are observed in most tissues, and analysis of mutational signatures has revealed exposure to specific cancer-causing agents, such as tobacco smoking in the bronchial epithelium (*16*) and aristolochic acid in the urothelium of Chinese urothelial cancer patients (*17*).

Epigenomic alterations, such as aberrant DNA methylation and remodeling of histone modifications, are also important for precancer evolution. Aberrant DNA methylation is induced by aging and accelerated by chronic inflammation (*18*). DNA methylation is inherited with high fidelity in somatic cells, and methylation of promoter CpG islands is consistently associated with transcriptional silencing of downstream genes. Some DNA methylation target genes are credentialed cancer drivers—for example, *CDKN2A* methylation in cancer cells leads to cell cycle deregulation, and *CDH1* methylation influences cell attachment and WNT signaling. However, similar to DNA mutations, most genes displaying aberrant DNA methylation in normal and cancer tissues are likely passengers, because genes susceptible to promoter CpG island DNA methylation are typically unexpressed with regulatory elements exhibiting repressive polycomb histone modifications (H3K27me3). Although it is rare for DNA mutations in specific genes to occur at high cellular proportions in normal tissues, aberrant DNA methylation of specific promoter CpG islands can be readily observed in high fractions of cells in a tissue (*19*, *20*). Tissues are composed of large numbers of clonal patches, where each patch is produced from a stem cell sharing the same mutations and aberrant DNA methylation (Fig. 1). It has been suggested that the accumulation of clonal patches with aberrant DNA methylation may provide a "cancerization field" that precedes driver mutations and predisposes cells to oncogenic mutations that would otherwise induce cellular senescence (*21*), highlighting the importance of interactions between epigenomic and genomic alterations. Other genome-epigenome interactions include *MLH1* methylation producing a mutator phenotype, global hypomethylation leading to chromosomal instability, and mutations in epigenomic modifiers in normal tissues leading to further epigenomic abnormalities (*3*, *10*). Notably, although it was previously thought that the presence of a

[1]Division of Epigenomics, National Cancer Center Research Institute, Tokyo 104-0045, Japan. [2]Epigenetics Research Laboratory, Genomics and Epigenetics Theme, Garvan Institute of Medical Research, Sydney, NSW 2010, Australia. [3]St. Vincent's Clinical School, Faculty of Medicine, University of New South Wales Sydney, Sydney, NSW 2010, Australia. [4]Cancer and Stem Cell Biology, Duke-NUS Medical School Singapore, Singapore 169857, Singapore. [5]Epigenomic and Epitranscriptomic Regulation, Genome Institute of Singapore, Singapore 138672, Singapore. [6]Cancer Science Institute of Singapore, National University of Singapore, Singapore 117599, Singapore. [7]Department of Physiology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore.
***Corresponding author. Email: tushijim@ncc.go.jp (T.U.); s.clark@garvan.org.au (S.J.C.); gmstanp@duke-nus.edu.sg (P.T.)**

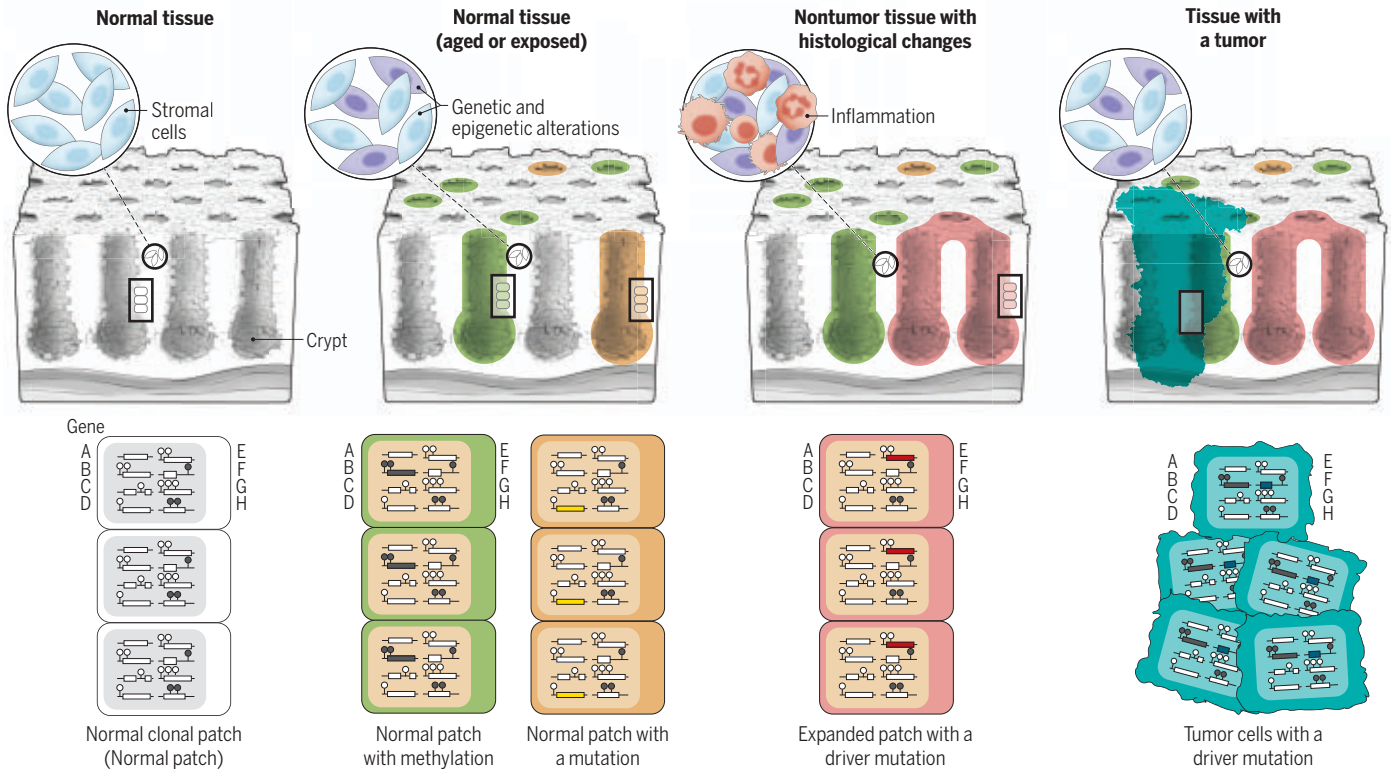


Fig. 1. Precancer evolution in nonmalignant tissues. Normal tissues of young individuals, represented by the colon, have few genomic and epigenomic alterations. Aging or exposure to carcinogenic factors, such as mutagens and chronic inflammation, increases clonal patches with genomic or epigenomic alterations, mainly passengers (mutated genes are shown in colors, methylated CpG islands are shown by closed circles, and crypts are colored reflecting the alterations). Some mutations in driver genes can confer growth advantages (gene E) only in a specific cellular ecosystem, such as an inflamed colon, and not in cancer. Tumors exhibit specific combinations of genomic and epigenomic driver alterations (exemplified by methylation of gene B and mutation of gene F).

driver mutation in cells is sufficient to cause cancer, the presence of genomic and epigenomic alterations in normal tissues raises provocative questions about how cancer initiation is defined. Therefore, integrating effects from multiple driver events, including the combination of specific genetic and epigenetic events in a specific microenvironment, are likely required to truly define a cancer cell state. Mapping the precise temporal order of genomic and epigenomic alterations is critical to gain insights into different pathways driving carcinogenesis.

Cancer evolution during metastasis and therapy

Unfortunately for too many patients, metastasis and therapy resistance often represent end points of the cancer life cycle. These terminal states are facilitated by mutational processes acting over extended time periods on tumor reservoirs in primary and distal organ sites, resulting in highly polyclonal diasporas of genetically distinct cancer cell populations upon which intrapatient selection, adaptation, and expansion can operate (Fig. 2). Cross-sectional studies comparing primary tumors and their metastatic counterparts have revealed similarities and differences in mutational burden, signatures, and driver gene profiles, with metastases generally exhibiting increased rates of structural variants and whole-genome duplication events (*22*, *23*). In complement, the analysis of longitudinal patient-matched primary and recurrent samples, or multiple metastases from the same patient (*24*), has enabled the reconstruction of within-patient clonal and subclonal relationships through shared mutations and structural variations. Generally, increased aneuploidy appears to be a general feature of advanced tumors (*24*), in some cases driving metastatic dissemination (*25*). Data from such studies have also enabled investigations into the timing of metastatic seeding (*26*), with some studies reporting phylogenetic data compatible with early seeding from a common primary progenitor and others indicating that metastasis may be a late event associated with multiple rounds of polyclonal seeding and the possibility that founder clones may alter their metastatic niches to facilitate colonization by secondary clones (*27*). Differences in the timing of analysis may explain apparently conflicting results on the cellular composition of metastases—some investigators have stated that metastatic lesions comprise a single or few clones, possibly reflecting limited tumor evolution after metastatic seeding, whereas others have reported increased clonal diversity (*22*). Because it is not ethically possible to study the evolution of metastasis over time in human patients, insights into metastases are being facilitated by experimental models. For example, CRISPR-based molecular barcoding can be combined with single-cell RNA sequencing to enable evolutionary lineage tracing across thousands of cells over wide temporal periods, capturing functional transcriptional states reflecting discrete cellular transitions in the metastatic program, marking tissue routes traveled by cells during the metastatic journey, and revealing specific genes contributing to dissemination (*28*). Use of such experimental models may allow the capture of critical functional changes related to metastasis that is not possible by purely studying primary and metastatic sites, such as tumor cells undergoing the actual dissemination process including survival in the circulation.

At the genetic level, driver mutations in primary lesions are often retained at recurrence, with metastases-specific driver gene alterations being fairly uncommon, which raises the question whether there are specific cell states that are specific to metastases. In this regard, epigenomic reprogramming is rapidly emerging as an important signature of the primary-to-metastasis transition and is particularly notable because epigenetic changes are inherited and transmitted through

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

subsequent cell divisions, thereby enabling propagation of prometastatic gene expression programs. Studies in multiple tumor types have revealed that metastatic lesions often exhibit radical genome-wide changes in their repertoire of regulatory enhancer elements, concomitant alterations in the binding profiles of master transcriptional regulators (*29*), and reactivation of early developmental gene expression programs and the epithelial-to-mesenchymal transition (*30*). Given the multistep nature of metastasis, involving detachment from primary sites, extravasation to the blood vessels, survival in blood, and reestablishment at distal sites, it is perhaps not surprising that these processes require a high degree of reversible transcriptional plasticity, which is more readily achieved by epigenomic rather than genomic changes. Notably, analysis of epigenetic and transcriptional changes during metastasis is guiding the identification of specific genes required for colonization that may represent potential drug targets (*31*) and cellular processes important for metastases (*32*–*34*). Beyond metastasis in general, emerging work is identifying mechanisms that dictate site-specific patterns of metastasis to organs such as the brain, which may involve lineage-associated metabolic circuits such as lipid metabolism for breast cancer (*35*) and molecular features of circulating tumor cells (CTCs) involving genes such as *SEMA4D* (*25*, *35*, *36*). Integrative genomic and epigenomic analysis of advanced cancers is yielding insights beyond the study of single levels alone. For example, genomic analysis of castration-resistant prostate cancers (CRPCs) identified a recurrently amplified noncoding genomic region distal to the androgen receptor (AR), which was subsequently shown by epigenomic analysis to function as a developmental enhancer driving AR expression and resistance to androgen deprivation therapy (*37*). Another recent CRPC study used whole-genome bisulfite sequencing to identify DNA methylation changes in *TMPRSS2* upstream regions predictive of *ERG* expression (a key driver of prostate cancer), but only in *TMPRSS2-ERG* fusion–positive tumors (*38*). However, despite these advances, current studies may be limited by not considering emerging epigenomic modifications, such as histone lactylation, in influencing cancer gene expression (*39*).

Therapy resistance often develops in many tumors, ultimately causing patient demise. Beyond baseline resistance, genomic and epigenomic analyses of tumors undergoing therapeutic challenge have revealed that although current cancer therapies may eradicate most cancer cells in a bulk population, the same treatments can also promote clonal evolution (*26*), causing acquired drug resistance through the mechanisms described above and beyond previously known secondary mutations in drug targets (*40*). For example, endocrine resistance in breast cancer can result from alterations in epigenetic-associated reprogramming of long-range enhancer-promoter interactions (*41*), and therapeutic resistance of colorectal cancer cells to anti–epidermal growth factor receptor (EGFR) therapy can involve mutations in accessory pathways such as *NF1*, noncanonical *KRAS* and *BRAF* aberrations, and signaling pathways originating from the tumor microenvironment (*42*). One active area of investigation involves the origin of cancer drug resistance—are drug-resistant cancer cells induced by therapy, or do they represent the clonal emergence of rare cell populations with preexisting resistance features? Some studies probing the temporal dynamics of drug resistance have reported tumor cell kinetics consistent with the outgrowth of preexisting drug-resistant populations. However, cancer cells from diverse tumor types, such as colorectal, breast, and lung cancers, can also survive under therapy by adopting drug-tolerant persister (DTP) phenotypes associated with slow-growth population dynamics and specific chromatin configurations, such as histone H3K4 and H3K27 demethylation and H3K9 trimethylation (*43*). Ultimately, both mechanisms may converge on similar genetic resistance mechanisms. Recent studies have also illuminated functional mechanisms through which exogenous therapies can induce faster rates of adaptation and change in cancer cells, such as mammalian target of rapamycin (MTOR)–mediated stress-induced mutagenesis (*44*), down-regulation of DNA repair genes, expression of error-prone polymerases (*45*), and enhancement of plasticity. These mechanisms are likely pivotal to enhancing the emergence of drug-resistant clones (*44*, *45*) and may highlight potential strategies for synthetic lethal combination therapies. Experimentally, studies in cancer drug resistance are also being facilitated by technical approaches combining molecular barcoding with mathematical modeling, which has demonstrated that DTP populations are not clonal but retain markedly high clonal complexity with phenotypic similarities to diapause—a state of suspended embryonic development involving survival mechanisms such as autophagy (*46*). Together, these studies have provided not only insights into how cancer cells can elicit drug resistance but also strategies as to how combinatorial therapeutics may overcome drug-resistance mechanisms while avoiding tumors transiting into accelerated evolution.

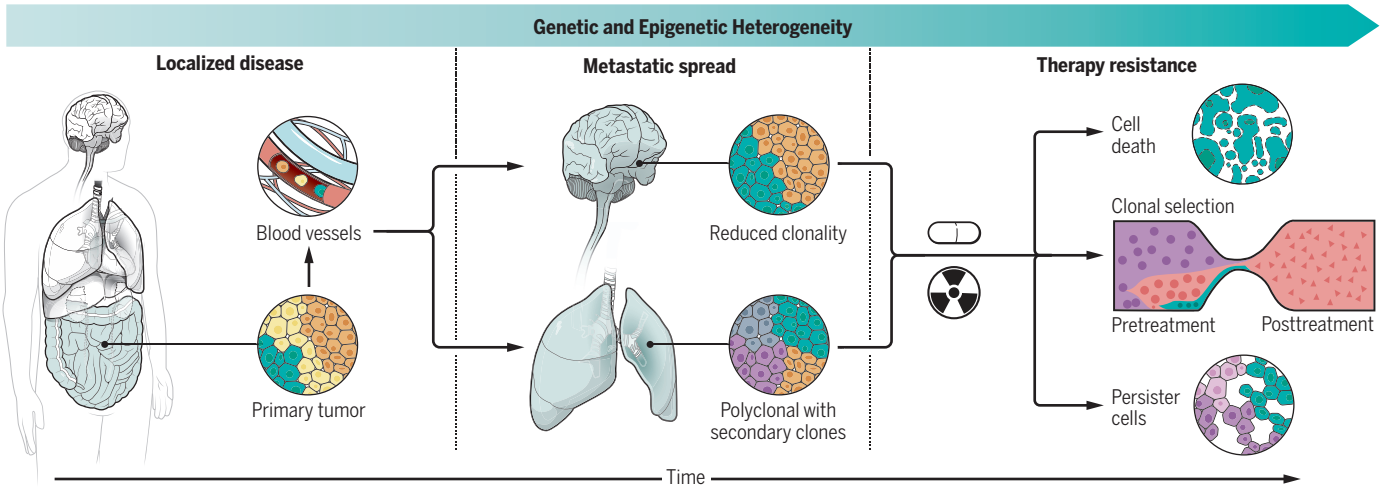


ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

Fig. 2. Cancer progression to metastasis and therapy resistance. (**Left**) Cells from primary tumors disseminate through the circulation to distal organs, such as the brain and lung. (**Middle**) Cases where the metastasis represents a single seeding event can result in reduced clonality (top), whereas increased clonality can result from successive rounds of polyclonal seeding (bottom). (**Right**) Exposure of tumors to therapy can induce cell death, induce the outgrowth of preexisting rare resistant populations, or induce DTP states that undergo further evolution to engender fully drug-resistant populations.

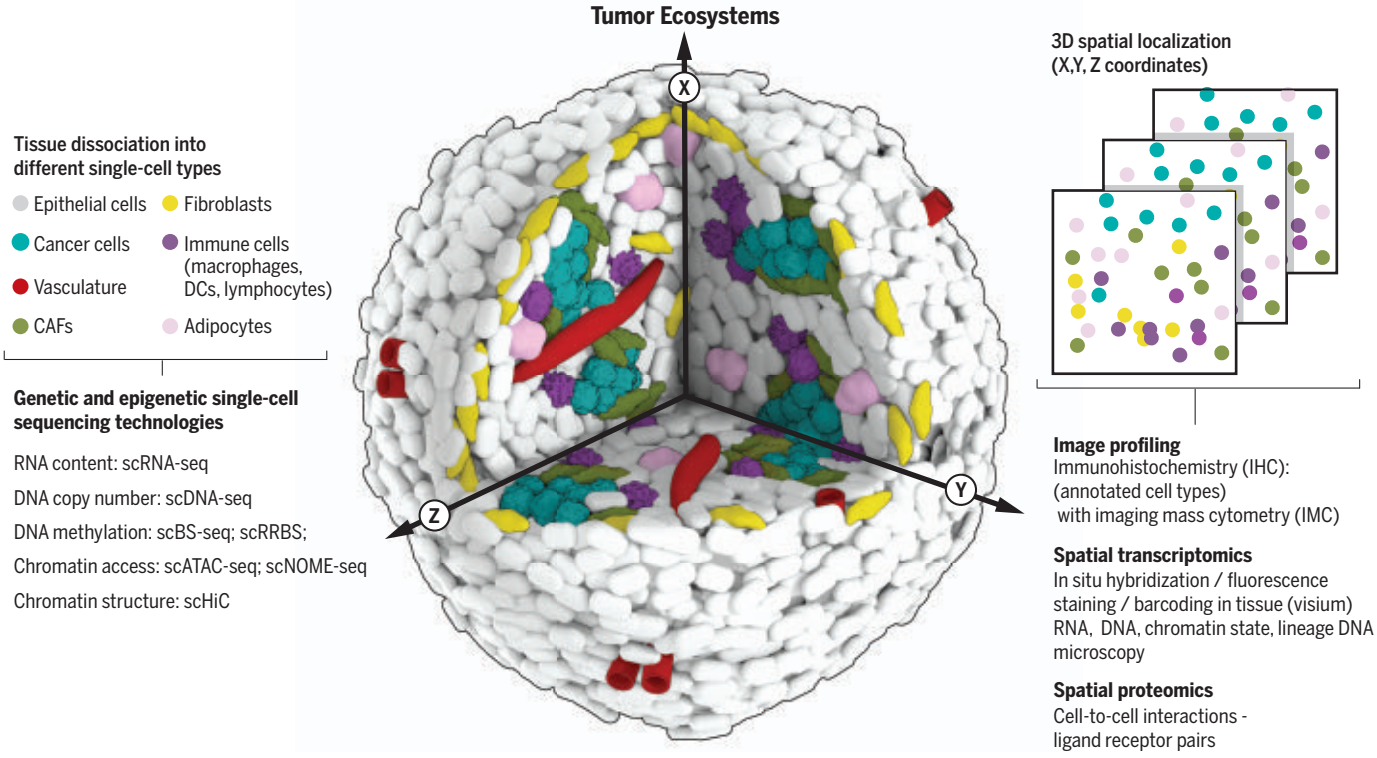


Fig. 3. Visualization of tumor ecosystems. Graphical 3D representation showing some of the main types of cells present in the tumor ecosystems of a solid tumor. After tissue dissociation, a range of single-cell genetic and epigenetic technologies can be used to uncover potential changes in molecular function and heterogeneity of different cell types. Molecular information pertaining to 3D spatial localization and cell-cell interactions can also be harnessed from a variety of imaging and sequencing technologies to better understand tumor progression, clonal outgrowth, and therapy outcome. DCs, dendritic cells; scRNA-seq, single-cell RNA sequencing; scDNA-seq, single-cell DNA sequencing; scBS-seq, single-cell bisulfite sequencing; scRRBS, single-cell reduced-representation bisulfite sequencing; scNOME-seq, single-cell nucleosome occupancy and methylome sequencing; scHiC, single-cell Hi-C.

Genomic and epigenomic alterations in the tumor ecosystem

Cancer epithelial cells reside in a tumor ecosystem with other diverse cell types, including tissue-resident normal cells, infiltrating immune cells, stromal subsets, endothelial cells, and adipocytes. Mapping the exact cellular constitution of the tumor ecosystem, understanding how it evolves across the cancer life cycle, and understanding its contribution to metastatic disease and therapy resistance remain cardinal questions. Previous (epi)genomic studies were primarily limited to bulk tumor profiling because of the microgram-level sample input requirements. However, single-cell and imaging technologies are now delineating genetic and epigenetic changes in tumor ecosystems at unprecedented resolutions (*47*, *48*), providing deeper insights into the role of the tumor microenvironment in promoting cancer cell clonal outgrowth (Fig. 3).

Notably, the coupling of epigenetic information with parallel DNA and RNA data is starting to facilitate biological interpretation of genetic and epigenetic heterogeneity in cancer (*49*, *50*). For example, single-cell triple omics sequencing (scTrio-seq) has been applied to examine mutations, expression, and DNA methylomes in colorectal cancer, revealing that DNA methylation levels are consistent within sublineages but can differ substantially among clones (*49*). Another study using nanopore sequencing compared the epigenomes of mutant and wild-type alleles in cancer cells, demonstrating large-scale differences in epigenetic states associated with somatic structural variants (*50*). Single-cell sequencing is also providing insights into genetic and epigenetic heterogeneity of the tumor stroma. Although it has been appreciated that stromal fibroblasts in cancer are often reprogrammed into cancer-associated fibroblasts (CAFs) that interact with malignant cells to promote or suppress growth, how this occurs has been unclear. Recently, single-cell analysis of pancreatic CAFs during in vivo tumor evolution identified a previously undescribed transforming growth factor–β (TGFβ)–driven, LRRC15^{+} CAF lineage associated with poor immunotherapy outcome (*51*), raising the possibility of boosting pancreatic cancer patient responses to immune checkpoint blockade by targeting stromal cells in the tumor microenvironment. Enduring epigenetic changes also occur in CAFs; for example, subsets of DNA methylation alterations can be shared between prostate cancer cells and CAFs (*52*), which is suggestive of common extrinsic signals conferring tumor-specific epigenome modifications in the tumor and stromal microenvironment. In tumors, endothelial cells are also reprogrammed to tumor endothelial cells (TECs) to promote tumor vasculature and immune cell regulation (*53*). Single-cell studies have recently uncovered new TEC subtypes displaying distinct signaling pathways, which highlights potential responses of stromal cells to antiangiogenic therapies (*54*, *55*).

The immune tumor microenvironment has received intense attention because of cancer immunotherapy, and single-cell sequencing is providing insights into the heterogeneity of immune cell states in tumor ecosystems (*56*, *57*). For example, although tumor-associated macrophages have traditionally been classified as inflammatory or tumor promoting, single-cell approaches are revealing a continuum of macrophage expression programs that reflects a large diversity of cell states. Similarly, new immune cell subtypes with distinct functions or states have been identified, with genes specifically expressed in rare immune cells being linked to tumor immune evasion (*58*, *59*). Single-cell profiling of tumor-infiltrating T cells shows that a suppressive immune microenvironment, with increased T cell exhaustion signatures and decreased activated T cells, is associated with

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

progression in various human cancers (*60*–*62*), whereas the presence of tissue-resident memory T cells, showing higher cytotoxic activity in tumors, is correlated with better prognosis (*61*). Epigenetic profiling using single-cell assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) in cancer models has also shown that exhausted $CD8^+$ T cell dysfunctional states are accompanied by broad remodeling of enhancer landscapes and transcription factor binding (*63*, *64*), and epigenomic mechanisms of cancer immunotherapy resistance are emerging, characterizing chromatin regulators of intratumoral T cell dysfunction before and after immunotherapy (*65*, *66*). Specifically, single-cell ATAC-seq analysis of tumor biopsies before and after programmed cell death protein 1 (PD-1) blockade has identified chromatin regulators of therapy-responsive T cell subsets and shared regulatory programs governing intratumoral $CD8^+$ T cell exhaustion and $CD4^+$ T follicular helper cell development (*67*). These studies highlight that gene-environment interactions may alter expression programs in different tumor-resident cell types through the interplay of genetics and epigenetics. However, most studies are still limited by technology to provide a comprehensive integrated analysis of genomic and epigenomic changes linked to the transcriptional output in individual cells, which will be needed to further garner insights into molecular pathways heralding therapy resistance or response.

Notwithstanding their ability to chart different independent transcriptional and (epi)genomic cell states, most single-cell studies of the tumor microenvironment are restricted by the lack of spatial information linking tumor cells to the microenvironment. Analysis of tissue sections by conventional immunofluorescence-based methods can provide spatial information, but such methods are often constrained by the number of markers that can be measured. Recent technological advances in high-dimensional imaging techniques, such as imaging mass cytometry (IMC) (*68*), multiplexed ion beam imaging (MIBI) (*69*), and co-detection by indexing (CODEX) (*70*), have greatly expanded the number of markers (>50) that can be simultaneously captured on tissue slides. Additionally, approaches now enable the mapping of cellular interactions into three-dimensional (3D) spatial volumes to measure cell types in situ within tumors (*47*, *71*). DNA microscopy analyzing the spatial locations of molecules using barcoded oligonucleotides has uncovered previously inaccessible layers of information within cells and within tissues (*71*), and spatial transcriptomics is increasingly used to study and visualize the distribution of mRNAs within tissue sections of many cancer types, including breast, prostate, and pancreatic cancers and metastatic melanoma (*72*). A systematic analysis of single-cell phenotypic and spatial correlates of genomic alterations in breast tumor ecosystems, performed by coupling IMC to multimodal genomic platforms, has identified diverse tissue phenotypes paralleling the genomic heterogeneity of breast cancer (*73*). In early-stage lung cancer, the spatial context of the immune microenvironment has been shown to exert strong selection on neoantigen expression through promoter methylation or copy number changes (*74*). Recent integration of spatial transcriptomics and proteomics has also revealed a key role for the immune microenvironment in melanoma by conferring distinct T cell phenotypes (*75*). The locations of immune cells within a tumor have been shown to markedly affect their anti- or protumorigenic effects (*76*).

More comprehensive single-cell epigenetic and transcriptional profiling will be required to explore these observations in greater detail. To this end, a number of international research consortia have been launched to further provide single-cell and spatial molecular data in different cancer contexts, such as the Human Tumour Atlas Network (HTAN), a National Cancer Institute (NCI)–funded Cancer Moonshot initiative to generate informative and accessible 3D atlases of cancer transitions for a diverse set of tumor types (*77*), and Imaging and Molecular Annotation of Xenografts and Tumors (IMAXT), a Cancer Research UK Grand Challenge project to build a 3D tumor using virtual reality to show every single different type of cell in the tumor. However, because of the current lack of scalable methods to integrate and standardize multimodal single-cell and imaging spatial datasets with different data distributions across diverse analytical tools and laboratories, these consortia will have to face and tackle formidable data informatic integration challenges so that investigations into the commonalities and differences in cellular landscapes between cancer tissues can be robustly addressed.

Conclusions

Despite substantial progress, recent precision oncology trials targeting patients on the basis of their genomic profiles have reported mixed outcomes (*78*). These sobering results highlight our still-incomplete understanding of the evolving nature of the cancer ecosystem, involving co-occurring genomic and epigenomic alterations, intratumor heterogeneity, and the role of the tumor microenvironment, including fibroblasts and immune cells. Our review has highlighted examples where combining multimodal levels of bioinformation can provide greater insights across the complete cancer life cycle. Beyond these correlative examples, the field is now challenged to achieve a higher degree of true data integration, where information from disparate sources across the spectrum of data-structure types is combined into single datasets to provide a unified view of both individual biological processes and cross-pathway molecular connections. In the future, it will be critical to incorporate additional levels, such as the microbiome, (epi)transcriptomics, and the effect of genetic ancestry on somatic mutation landscapes, to shed further light into the foundational nature of cancer.

Another critical future challenge will be to translate basic insights into tangible clinical applications. From a public health perspective, studies of mutational and epigenetic signatures may facilitate efforts in removing carcinogenic agents and processes from our environment. In the early-detection space, the high levels of aberrant DNA methylation in normal tissues make their accurate quantification feasible, and associations with future cancer risk have already been shown for some cancers (*18*, *79*). Blood-based assays of genomic and epigenomic alterations, which can be harnessed for clinical use, could potentially be used for screening because certain local infections or lifestyle-factor exposures have been shown to elicit common systemic responses, including alterations in bone marrow and thus peripheral blood (*80*). However, current efforts in diagnosing cancer on the basis of molecular patterns will also be challenged by the presence of genomic and epigenomic alterations in normal tissues and the designation of driver versus passenger events. For advanced cancers, integrative and temporal analysis may further pinpoint culprit processes that truly define metastatic states for therapeutic development. Finally, the combination of imaging and single-cell sequencing technologies is providing compelling evidence that mapping the spatial transcriptional landscapes of cancer is cardinal to understanding tumor progression and therapy outcome. However, to fully understand cancer, scientists still need to understand everything about a tumor—what types of cells are in it, how many there are, what they are doing, and where they are located. For future advances, we therefore need the integration of molecular features, such as genetic, transcriptional, and epigenetic states in single cells, and the rendering of these measurements into a 3D context. Through this, we can learn how emergent cancer phenotypes evolve and how changes in regulatory states, cell interactions, and spatial contexts influence cancer differentiation trajectories within the cancer ecosystem.

REFERENCES AND NOTES

1. H. Sung *et al*., *CA Cancer J. Clin.* **71**, 209–249 (2021).
2. The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, *Nature* **578**, 82–93 (2020).
3. L. Ding *et al*., *Cell* **173**, 305–320.e10 (2018).
4. M. A. Reyna *et al*., *Nat. Commun.* **11**, 729 (2020).
5. S. K. Loganathan *et al*., *Science* **367**, 1264–1269 (2020).
6. E. Rheinbay *et al*., *Nature* **578**, 102–111 (2020).
7. K. C. Akdemir *et al*., *Nat. Genet.* **52**, 1178–1188 (2020).
8. N. Kakiuchi, S. Ogawa, *Nat. Rev. Cancer* **21**, 239–256 (2021).
9. N. Wijewardhane, L. Dressler, F. D. Ciccarelli, *Cancer Cell* **39**, 125–129 (2021).
10. K. K. Huang *et al*., *Cancer Cell* **33**, 137–150.e5 (2018).
11. N. Kakiuchi *et al*., *Nature* **577**, 260–265 (2020).

12. K. Nanki *et al.*, *Nature* **577**, 254–259 (2020).
13. S. Olafsson *et al.*, *Cell* **182**, 672–684.e11 (2020).
14. I. Martincorena *et al.*, *Science* **362**, 911–917 (2018).
15. A. Yokoyama *et al.*, *Nature* **565**, 312–317 (2019).
16. K. Yoshida *et al.*, *Nature* **578**, 266–272 (2020).
17. R. Li *et al.*, *Science* **370**, 82–89 (2020).
18. H. Takeshima, T. Ushijima, *NPJ Precis. Oncol.* **3**, 7 (2019).
19. K. Ramachandran, R. Singal, *Epigenomics* **4**, 243–245 (2012).
20. S. Yamashita *et al.*, *Clin. Epigenetics* **11**, 191 (2019).
21. Y. Tao *et al.*, *Cancer Cell* **35**, 315–328.e6 (2019).
22. F. Bertucci *et al.*, *Nature* **569**, 560–564 (2019).
23. P. Priestley *et al.*, *Nature* **575**, 210–216 (2019).
24. F. P. Barthel *et al.*, *Nature* **576**, 112–120 (2019).
25. D. J. H. Shih *et al.*, *Nat. Genet.* **52**, 371–377 (2020).
26. Z. Hu, Z. Li, Z. Ma, C. Curtis, *Nat. Genet.* **52**, 701–708 (2020).
27. S. Yachida *et al.*, *Nature* **467**, 1114–1117 (2010).
28. J. J. Quinn *et al.*, *Science* **371**, eabc1944 (2021).
29. M. M. Pomerantz *et al.*, *Nat. Genet.* **52**, 790–799 (2020).
30. I. Pastushenko *et al.*, *Nature* **556**, 463–468 (2018).
31. J. J. Morrow *et al.*, *Nat. Med.* **24**, 176–185 (2018).
32. R. Y. Ebright *et al.*, *Science* **367**, 1468–1473 (2020).
33. C. K. Lee *et al.*, *Science* **363**, 644–649 (2019).
34. O. G. McDonald *et al.*, *Nat. Genet.* **49**, 367–376 (2017).
35. X. Jin *et al.*, *Nature* **588**, 331–336 (2020).
36. R. Klotz *et al.*, *Cancer Discov.* **10**, 86–103 (2020).
37. D. Y. Takeda *et al.*, *Cell* **174**, 422–432.e13 (2018).
38. S. G. Zhao *et al.*, *Nat. Genet.* **52**, 778–789 (2020).
39. D. Zhang *et al.*, *Nature* **574**, 575–580 (2019).
40. P. Razavi *et al.*, *Cancer Cell* **34**, 427–438.e6 (2018).
41. J. Achinger-Kawecka *et al.*, *Nat. Commun.* **11**, 320 (2020).
42. A. Woolston *et al.*, *Cancer Cell* **36**, 35–50.e9 (2019).
43. S. Shen, S. Vagner, C. Robert, *Cell* **183**, 860–874 (2020).
44. A. Cipponi *et al.*, *Science* **368**, 1127–1131 (2020).
45. M. Russo *et al.*, *Science* **366**, 1473–1480 (2019).
46. S. K. Rehman *et al.*, *Cell* **184**, 226–242.e21 (2021).
47. J. G. Camp, R. Platt, B. Treutlein, *Science* **365**, 1401–1405 (2019).
48. J.Lee, D. Y. Hyeon, D. Hwang, *Exp. Mol. Med.* **52**, 1428–1442 (2020).
49. S. Bian *et al.*, *Science* **362**, 1060–1063 (2018).
50. I. Lee *et al.*, *Nat. Methods* **17**, 1191–1199 (2020).
51. C. X. Dominguez *et al.*, *Cancer Discov.* **10**, 232–253 (2020).
52. R. Pidsley *et al.*, *Genome Res.* **28**, 625–638 (2018).
53. B. Lim, Y. Lin, N. Navin, *Cancer Cell* **37**, 456–470 (2020).
54. J. Goveia *et al.*, *Cancer Cell* **37**, 21–36.e13 (2020).
55. Q. Zhao *et al.*, *Cancer Res.* **78**, 2370–2382 (2018).
56. E. Azizi *et al.*, *Cell* **174**, 1293–1308.e36 (2018).
57. Q. Song *et al.*, *Cancer Med.* **8**, 3072–3085 (2019).
58. L. Gong *et al.*, *Nat. Commun.* **12**, 1540 (2021).
59. D. Lambrechts *et al.*, *Nat. Med.* **24**, 1277–1289 (2018).
60. X. Guo *et al.*, *Nat. Med.* **24**, 978–985 (2018).
61. P. Savas *et al.*, *Nat. Med.* **24**, 986–993 (2018).
62. C. Zheng *et al.*, *Cell* **169**, 1342–1356.e16 (2017).
63. M. Philip *et al.*, *Nature* **545**, 452–456 (2017).
64. D. R. Sen *et al.*, *Science* **354**, 1165–1169 (2016).
65. D. Miao *et al.*, *Science* **359**, 801–806 (2018).
66. D. Pan *et al.*, *Science* **359**, 770–775 (2018).
67. A. T. Satpathy *et al.*, *Nat. Biotechnol.* **37**, 925–936 (2019).
68. C. Giesen *et al.*, *Nat. Methods* **11**, 417–422 (2014).
69. M. Angelo *et al.*, *Nat. Med.* **20**, 436–442 (2014).
70. D. Phillips *et al.*, *Front. Immunol.* **12**, 687673 (2021).
71. J. A. Weinstein, A. Regev, F. Zhang, *Cell* **178**, 229–241.e16 (2019).
72. N. L. de Vries, A. Mahfouz, F. Koning, N. F. C. C. de Miranda, *Front. Oncol.* **10**, 1254 (2020).
73. H. R. Ali *et al.*, *Nat. Can.* **1**, 163–175 (2020).
74. R. Rosenthal *et al.*, *Nature* **567**, 479–485 (2019).
75. R. Cabrita *et al.*, *Nature* **577**, 561–565 (2020).
76. M. Binnewies *et al.*, *Nat. Med.* **24**, 541–550 (2018).
77. O. Rozenblatt-Rosen *et al.*, *Cell* **181**, 236–249 (2020).
78. G. Middleton *et al.*, *Nature* **583**, 807–812 (2020).
79. M. Maeda *et al.*, *Gut* **66**, 1721–1723 (2017).
80. S. Wahl *et al.*, *Nature* **541**, 81–86 (2017).

ACKNOWLEDGMENTS

We thank Z. Xiao for technical assistance and G. Lin and K. Patterson for figure drafting. **Funding:** T.U. is supported by AMED JP20ck0106552h0001 and JP20 ggm1310006h0001. S.J.C. is supported by a National Health and Medical Research (NHMRC) fellowship (no. 1156408) and NHMRC project grants (nos. 1147974 and 1144574). P.T. is supported by a National Medical Research Council grant (NMRC/STaR/0026/2015); an Open Fund-Large Collaborative grant (MOH-OFLCG18May-0003); Duke-NUS Medical School; the Agency for Science, Technology and Research; and the Cancer Science Institute of Singapore. **Competing interests:** The authors declare no competing interests.

10.1126/science.abh1645

REVIEW

Insights into human history from the first decade of ancient human genomics

Yichen Liu[1]†, Xiaowei Mao[1]†, Johannes Krause[2], Qiaomei Fu[1,3]*

Recent advancements in DNA sequencing technologies and laboratory preparation protocols have rapidly expanded the scope of ancient DNA research over the past decade, both temporally and geographically. Discoveries include interactions between archaic and modern humans as well as modern human population dynamics, including those coinciding with the Last Glacial Maximum and the settlement history of most world regions. This new type of data allows us to examine the deep past of human population dynamics and sharpen the current understanding of our present. The continued development in the ancient DNA field has transformed our understanding of human genetic history and will keep uncovering the further mysteries of our recent evolutionary past.

In 2001, a draft sequence of the human genome was published (*1*, *2*), promising a better understanding of human biology and evolution from a molecular perspective. Since the mid-2000s, advancements and the subsequent broad applications of high-throughput sequencing (HTS) techniques enabled rapid and cost-effective sequencing of eukaryotic genomes and paved the way for a genomic era in the field of ancient DNA (aDNA; DNA obtained from remains of past organisms). By providing time-stamped genetic data, aDNA has revealed admixture of modern humans with previously unknown archaic populations (*3*); uncovered the rise of key adaptive mutations (*4*), shedding light on long lasting debates on our recent evolutionary past (*5*); and continuously added new pieces of evidence to reveal the genetic history of past and present human populations over the course of the past few tens of thousands of years.

In this Review, we discuss the global human population movements revealed by aDNA studies, admixture, and replacements of modern and archaic humans from the Middle Paleolithic to the historical period, with a focus on the insights gained from aDNA evidence. In addition to modern humans (*Homo sapiens*, the only extant hominin), this Review also discusses current knowledge about our most closely related archaic relatives, Neanderthals and Denisovans (who once inhabited Eurasia), for whom genome-wide information has been retrieved (*3*, *6*). Most of the hypotheses reviewed here are from a very limited sampling of ancient humans because of the scarcity of well-preserved human subfossils (remains that are not completely fossilized and contain endogenous organic molecules).

[1]Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Center for Excellence in Life and Paleoenvironment, Chinese Academy of Sciences, Beijing, 100044, China. [2]Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. [3]University of the Chinese Academy of Sciences, Beijing, 100049, China.
***Corresponding author. Email: fuqiaomei@ivpp.ac.cn**
†These authors contributed equally to this work.

By the late 1980s, it was shown that DNA can survive long after an organism's death but is inherently highly degraded and often chemically altered (*7*). Although early attempts in characterizing aDNA were limited to short DNA fragments, applying HTS techniques to ancient humans along with an in-depth understanding of DNA degradation led to the publication of three draft ancient genomes [Neanderthal, Denisovan, and a 4-thousand-year (kyr)–old modern human from Greenland] in 2010 (*7*). The Denisovan represents a new archaic lineage that was identified by using only aDNA data (*8*). As improvements in laboratory and computational protocols allowed for the retrieval and identification of extremely short aDNA fragments and methods for removing modern contaminating DNA, the number of sequenced ancient human genomes has grown exponentially (Fig. 1). These ancient genomes allow us to explore the genetic makeup and adaptation of archaic lineages, ancient modern humans, and present-day populations, providing a glimpse into the genetic history of humans across the globe over the past millennia.

Archaic and modern human interactions

The genomes of Neanderthals and Denisovans have been reconstructed in 2010 (*3*, *6*). Analyses of their autosomes indicate that they are more closely related to each other than to modern humans, their sister group (*3*). From the level of shared genetic variants, archaic humans are estimated to have separated from modern humans ~550 thousand years ago (ka), with Neanderthals and Denisovans splitting from each other ~400 ka (*9*–*11*) (Fig. 2). Genetic admixture with Neanderthals and Denisovans has been widely detected in present-day non-Africans, although admixture proportions and geographical distributions differ widely. Neanderthal-derived admixture has been detected in all present-day non-Africans, with East Asian genomes containing a higher amount of Neanderthal ancestries than that of western Eurasians (2.3 to 2.6% and 1.8 to 2.4%,

respectively) (*3*, *6*, *9*), although a recent study suggests a smaller difference in levels of Neanderthal ancestry in non-African populations after correcting for back-migration into African populations (*12*). In addition, it has been hypothesized that some African populations carry ancestry from an even older archaic lineage that predates the divergence of Neanderthals and modern humans (*13*). Furthermore, ancient modern humans mated with extinct archaic lineages as multiple waves of introgression from Neanderthals and Denisovans to non-Africans have been suggested (Fig. 2) (*14*). This is supported by estimates of up to four admixture pulses from Denisovans and three pulses from Neanderthals to modern humans, accounting for the presence of archaic-derived DNA [including some adaptive introgressions, such as *EPAS1* (*endothelial Pas domain protein 1*) gene (*4*)] in present-day populations (*15*–*17*).

Admixture between modern humans and archaic lineages appears to have occurred in both directions. Studies suggest that both the mitochondrial DNA and Y chromosome of Neanderthals were replaced by that of early modern humans, which was likely a result of introgression from modern humans into Neanderthals between ~370 and 220 ka (Fig. 2) (*18*, *19*). Neanderthal–modern human admixture occurring as recently as ~40 ka has been identified, although these lineages apparently did not survive to the present day (*20*, *21*). Also, an ~50-kyr-old archaic individual (Denisova 11) had a Neanderthal mother and a Denisovan father (*22*). Considering the increasing number of ancient individuals identified with recent archaic ancestries, past hominins may have mixed frequently, opening the question of whether archaic and modern human should be regarded as distinct lineages or rather points taken from a continuous spectrum of genetic diversity that was genetically connected throughout the past ~500 kyr similar to that of present-day human populations (*20*, *22*, *23*).

Population dynamics of modern humans

Diversity of early modern human populations

Africa

Genetic data strongly supports a modern human origin in Africa (*24*), and the morphology of early modern human groups suggests geographically dispersed populations in Pleistocene Africa (*25*). However, it remains difficult to determine a single model to characterize the origin of African ancestry (*23*, *26*). Many locations have been proposed from estimates of genetic diversity and divergence time, including in South Africa (*27*), West Africa (*28*), and East and Central Africa (*29*). The earliest evidence of partial modern human morphology was found in present-day Morocco, dating to ~315 ka (*23*, *30*). It is hypothesized that modern humans emerged from an interconnected pan-African population, but further evidence is needed to test this model (*26*). Sometime between ~250 and 200 ka (*29*), five major branches contributing to early modern human ancestry started to split from each other within a short time: (i) a population that contributed ancestry primarily to South African hunter-gatherers (*31*); (ii) another population that contributed ancestry primarily to Central African hunter-gatherers (*29*); (iii) other human populations, which include West Africans, non-Africans, and East African agro-pastoralists (such as the Agaw); (iv) an unsampled modern population that contributed ancestry to West Africans and Mota, a 4.5-kyr-old individual from present-day Ethiopia (*32*); and (v) an ancient North African population that was equally related to East and West Africans that contributed about half of the ancestry of the oldest currently available African genomes from the Pleistocene site of Taforalt in present-day Morocco (*33*).

Alternatively, divergence times for modern human ancestry were dated to ~360 to 260 ka by using a coalescent-based approach (*31*). Between ~80 and 60 ka, a series of divergences that involved the ancestral populations representing non-Africans, East Africans (including East African hunter-gatherers and agro-pastoralists), and West Africans potentially occurred in East Africa (*29*). The majority of all non-African human ancestry stems from a worldwide expansion out of Africa that started ~60 ka (*23*). Several archaeological and fossil sites hint at earlier expansions [for example, (*34*)], but their dates and attribution to modern humans are debated.

Eurasia

Genomic data has been obtained from early modern humans that lived in Eurasia as early as ~45 ka. Although the fossil record across Eurasia remains scattered and fragmented, the genetic data have identified several early modern human lineages (Fig. 2) (*35*). Some of these lineages showed no detectable genetic continuity to later populations, whereas others could be genetically connected with present-day human populations. Three lineages older than ~40 kyr from present-day Russia (*36*),

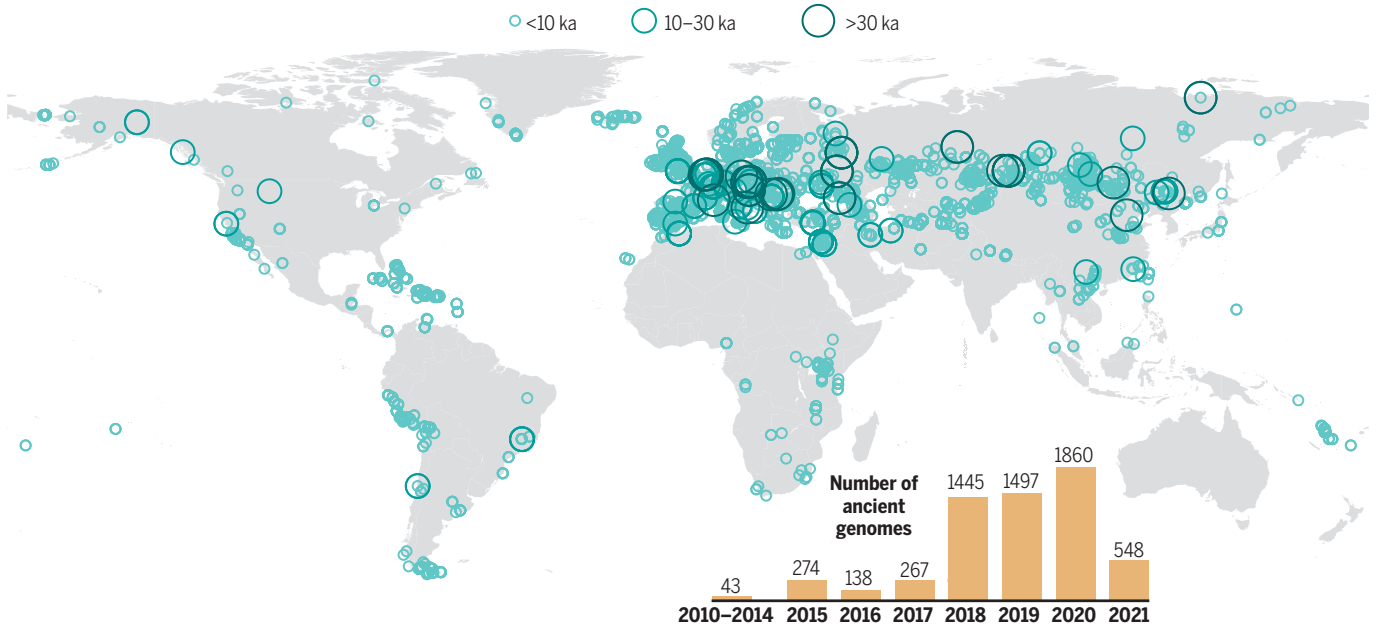


Fig. 1. Published ancient human genomes (including whole-genome sequences, exomes, and genome-wide single-nucleotide polymorphism data) by June 2021. Large dark blue circles indicate ancient individuals older than 30 kyr; medium blue circles indicate ancient individuals younger than 30 kyr and older than 10 kyr; and small light-blue circles indicate ancient individuals younger than 10 kyr.

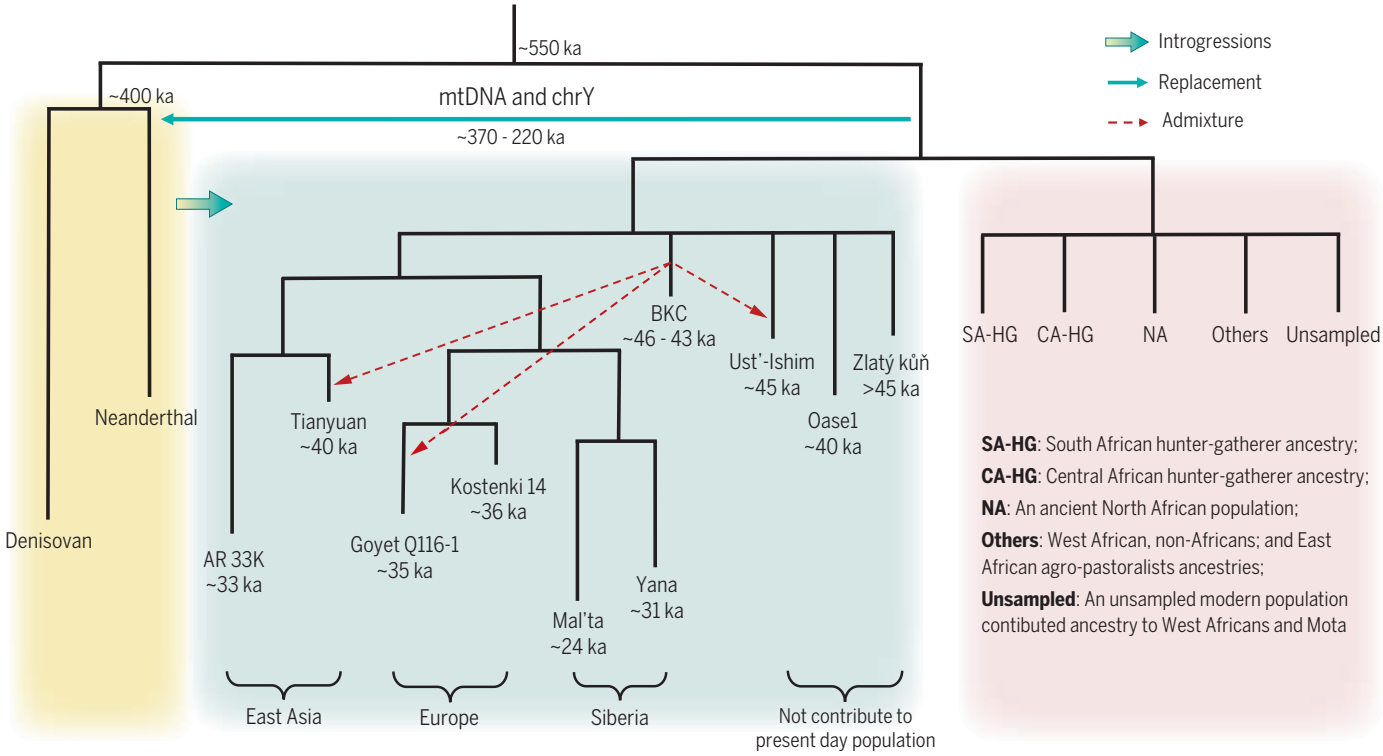


Fig. 2. Schematic illustrating the population history of archaic humans and early modern humans. Modern humans split from archaic humans ~550 ka, and Denisovans split from Neanderthals ~400 ka (*11*–*13*). Multiple pulses of introgression detected from archaic humans to non-Africans and Neanderthal mitochondrial genome and Y chromosome were replaced by modern humans ~370 to 220 ka (*19*, *20*). Distinct populations were identified by using genomic evidence in Eurasia and Africa.

Romania (*20*), and the Czech Republic (*37*), representing the most basal and oldest modern human genomes to date, did not contribute to present-day populations.

Other early modern human lineages from ~24 to 40 ka are genetically connected to present-day populations (*38*–*40*). Ancient Europeans whose ancestry is found in present-day West Eurasians are represented by two 35- to 37-kyr-old individuals from Europe (*38*, *39*). In addition, demographic modeling proposes that Ancient North Siberians (ANS) were found to have diverged from West Eurasians ~39 ka, with affinities to both early West Eurasians and early East Asians. Furthermore, the descendants of ANS, Ancient North Eurasians (ANE), are closely related to Native Americans, as shown by an ~24-kyr-old individual from present-day Russia (*40*, *41*). In East Asia, an ~40-kyr-old individual from present-day China (Tianyuan) represents a population related to present-day Asians (*42*). Despite our current scant knowledge of these earlier periods, it is clear that over time, increasing population structure, greater population interaction, and higher migration occurred across Eurasia (*35*).

With the increased availability of ancient human genomes from early Eurasians, we are starting to piece together the Pleistocene population history in Eurasia. For example, it was unclear how the two populations, represented by Tianyuan from present-day China and GoyetQ116-1 from present-day Belgium, are genetically connected (*42*). Direct gene flow seems unlikely considering the physical distance between the two. Interestingly, a population connected to both Tianyuan and GoyetQ116-1 has recently been revealed through analyses of genomes of ~46- to 43-kyr-old individuals from Bacho Kiro Cave (present-day Bulgaria) (*21*). Although it remains highly challenging to obtain direct supporting evidence considering the scarcity of both the genetic and archeological evidence, population dynamics such as that represented by Tianyuan, GoyetQ116-1, and Bacho Kiro are likely related to the contemporary expansion and transition of Initial Upper Paleolithic cultures, such as the spread of the Aurignacian complex (*43*). At least some early Upper Paleolithic Eurasian populations [before the Last Glacial Maximum (LGM)] showed high genetic diversity and low mutational load (the burden of deleterious variation in a population), whereas bottlenecks associated with a reduction of diversity occurred in northern latitudes with the LGM (*44*).

Human population dynamics during the LGM

We discuss here population changes coinciding with the LGM, a period of harsh environmental conditions in Europe, East Asia, and Siberia. In Europe, one of the early populations represented by an ~35-kyr-old individual from present-day Belgium may have been displaced, although an ~19-kyr-old individual from Iberia carrying this ancestry after the LGM (~19 ka) was identified in southwestern Europe (*38*). After the LGM, admixture was observed among hunter-gatherers from different European refugia (*45*). In East Asia, ancestry related to Tianyuan and a ~33-kyr-old individual from the Amur Region in present-day China was widely distributed across northern East Asia before the LGM (*42*, *46*). At the end of the LGM, a potential population shift occurred related to an ~19-kyr-old individual from the Amur Region, with the appearance of the *EDAR* (*ectodysplasin A receptor*) *V370A* mutation (associated with traits such as thicker hair shafts) (*46*). Similar to those in Europe and East Asia, population changes also occurred in the Pleistocene in Siberia, which was settled by early modern humans by at least ~45 ka (*36*). After the LGM, the populations carrying ANS and ANE ancestries were replaced by Ancient Paleo-Siberians (APS), who were formed by the admixture between ancient East Asians and the ANE and spanned from northeastern Siberia to just south of Lake Baikal (*40*, *47*).

Human population dynamics after the LGM

After the LGM, a warmer and more stable climate emerged, and human population dynamics

from this time are characterized by a series of rapid expansions, migrations, interactions, and replacements. The patterns of population movements differ regionally, with some largely restricted to long-term population continuity and internal interactions (such as mainland East Asia and Australia), and some dominated by recurring population admixture and replacement (such as Europe and the Eurasian steppe). In the next sections, we summarize major demographic events geographically, following a roughly chronological order. However, our understanding of human demographic history is far from complete, and there are still many gaps to be closed in future.

Africa

Africans harbor the highest genetic diversity among present modern human populations (*5*). However, because of poor aDNA preservation in Africa, we are only just beginning to uncover the ancient African population structure and migration patterns (Fig. 3). Ancient genomes reveal that some early Africans are genetically connected with populations in the Near East, supported by ~15- to 5-kyr-old individuals from present-day Morocco (*33*, *48*) and ~3- to 1-kyr-old individuals from eastern and southern Africa (Fig. 3) (*28*). Between ~5 and 1 ka, multiple admixture events among herders and farmers in sub-Saharan Africa gave rise to present-day East African populations (*49*, *50*). Arguably, one of the most important events in recent African population history was the dispersal of West African–related Bantu speakers (Fig. 3). It presents one of the largest known expansion events of farming populations believed to have occurred globally (*51*) and spread agriculture and West African–related ancestry across most parts of sub-Saharan Africa (*29*). Additionally, genetic and archaeological evidence support a pastoralist expansion from eastern to southern Africa ~2 ka (Fig. 3) (*28*). Because of the limited availability of ancient African genomes and complex gene flow among Holocene African groups, the population structure and movements within Africa remain debated [for example, (*28*, *29*)].

Europe

aDNA studies have enhanced both our understanding of the spread of farming into Europe as well as accompanying languages such as Indo-European (the largest present language family, with more than 445 living languages and more than 3 billion native speakers). Archaeological evidence indicates that farming spread over the European continent during the Neolithic period starting ~8.5 ka; however, it was debated whether the spread of agriculture was due to the migration of people or rather a lateral transmission of ideas and culture. Analysis of aDNA showed that Neolithic farming populations from Near East (Anatolia) largely expanded throughout Europe and admixed in the subsequent millennia with Mesolithic hunter-gatherers, demonstrating that the spread of farming was mitigated by the migration of people rather than ideas (*5*). The connection with present-day Near Easterners could already be slightly observed as early as ~14 ka, coinciding with the appearance of the *HERC2* (*HECT and RLD domain-containing E3 ubiquitin protein ligase 2*)–derived allele related to light eye color in present-day Europeans (*38*). Around 4.9 ka, steppe ancestry [a mixture of at least two hunter-gatherer ancestries from present-day Russia (Eastern hunter-gatherers) and the Caucasus (Caucasus hunter-gatherers) (*52*)]—closest related to individuals that were found associated with the the Yamnaya culture, a cultural complex that spread across the entire Pontic-Caspian-Ural steppe region—expanded westward and eastward (Fig. 3) (*53*). This ancestry appeared in Central Europe and formed the population associated with the Corded Ware culture (hallmarked by cord-decorated ceramics) ~4.9 ka (*53*). It was suggested that Indo-European languages spread into Europe together with populations carrying steppe ancestry (*53*). Around 4.6 ka, individuals with steppe ancestry arrived in the British Isles, coinciding with the spread of the Bell Beaker Complex (defined by assemblages of stylized bell-shaped grave goods), replacing within a few hundred years ~90% of the local gene pool (*54*). The genetic evidence suggests that this process was mostly driven by male individuals (*55*) because in both the British Isles and Iberia, almost all Late Neolithic Y chromosomes were replaced by Eastern European steppe–related Y chromosomes (*54*). Ancient genomes thus revealed that present-day European ancestry consists of three major genetic components: European hunter-gatherer ancestry, early farmer ancestry, and steppe ancestry in varying proportions across Europe (*53*).

Steppe, Central Asia, and South Asia

The previously mentioned steppe populations played an important role in the population dynamics of Eurasia. Around 4.9 ka, steppe populations not only expanded into present-day Europe but also eastward, contributing to individuals associated with the Afanasievo culture (in the Altai region) (*53*, *56*). This steppe-related ancestry spread into the eastern steppe in present-day central Mongolia, but not further eastward (*57*). In Bronze Age Central Asia, the ancestries of the Bactria-Margiana Archaeological Complex (found in present-day Uzbekistan and Turkmenistan) were derived from Iranian farmers (60 to 65%) and Anatolian farmers (20 to 25%), with minor contributions from West Siberian hunter-gatherers (~10%) and Andamanese hunter-gatherers (2 to 5%) (*58*).

In South Asia, the genome of an ~5-kyr-old individual from the Indus Valley Civilization (IVC; one of the first large-scale urban societies of the ancient world) presents a population that is the largest ancestry source for present-day South Asians (*59*). Later, this IVC-related population mixed with northwestern groups carrying steppe ancestry and southeastern groups, forming the Ancestral North Indians (ANI) and Ancestral South Indians (ASI), respectively. The mixture of ANI and ASI led to the main genetic cline in South Asia today (*58*). The steppe ancestry in South Asia, which contributed up to 30% of the ancestries of present-day groups in the region, provides additional evidence for the spread of Proto-Indo-European languages through the steppe expansion (*58*).

East Asia and Southeast Asia

In Mongolia, as mentioned above, major genetic turnovers are associated with the expansion and migration of steppe herders (such as the Afanasievo people) and dairy pastoralists (such as the Xiongnu and Mongols) (Fig. 3) (*57*, *60*). Particularly, Neolithic populations (~8 to 6.5 ka) carried almost complete East Asian ancestry, and steppe pastoralist ancestry was introduced with a pastoralist expansion ~5 ka (*57*, *61*). Later, Mongolian populations (~3.4 ka) showed a mix of ancestries derived from individuals associated with the Yamnaya culture and European farmers (*61*). Individuals associated with the Xiongnu confederate, a nomadic confederation of the eastern steppes, were established ~2 ka and carried a genetic of populations from Mongolia and the surrounding regions (*57*), whereas historical Mongols have a high level of eastern Eurasian ancestry similar to those of present-day Mongolic-speaking populations (*57*).

In the East Asian mainland and archipelagos, diverse genetic lineages were present during the Neolithic, including populations that carried northern East Asian ancestry, ancient hunter-gatherer ancestry from Southeast Asia associated with the Hòabìnhian culture, at least two distinct southern East Asian lineages, and ancestry lineage best presented by individuals associated with the Jomon culture present in the Japanese archipelago (*35*, *62*, *63*). Genetic differentiation between northern and southern East Asians, the earliest branches of present-day East Asians, traces back as early as ~19 ka (*46*, *62*). After the Neolithic, northern East Asian ancestry spread across southern East Asia, with increased genetic affinity to northern East Asians occurring in southern East Asians over time. South-to-north gene flow can be seen from southern East Asian ancestry identified in northern Han populations and some northern East Asians (Fig. 3) (*61*, *62*, *64*). In northern East Asia, an ~14-kyr-old individual found in the Amur River Basin

is most closely related to the East Asian source found to date that contributed to APS in Siberia; a genetic continuum is maintained in the Amur River Basin to the present day (*46*).

The connections between southern East Asians and Southeast Asians are complex. Recent genetic evidence suggests a complicated model in which the indigenous hunter-gatherers of Southeast Asia admixed with multiple waves of East Asian farmers (*65*), and admixture between the Southeast Asian and southern East Asian hunter-gatherers was detected in the southern regions of mainland East Asia, well before the spread of farming (~9 to 6 ka) (Fig. 3) (*63*). Asians in mainland East Asia and the Taiwan Strait islands are suggested to be ancestral to Neolithic Austronesians (~11.5 to 4.2 ka) (*62*, *63*). These seafaring Austronesians quickly expanded from islands of Southeast Sea to Near Oceania.

Oceania

Archeological evidence suggests that human populations from Southeast Asia have initially settled in Sahul (present-day Australia, Tasmania, and New Guinea) before ~50 ka (*66*). In the Australian continent, a single rapid migration along the east and west coasts ~49 to 45 ka was followed by the continuous presence of populations, and neither the mitochondrial genomes nor Y chromosomes suggest more recent gene flow into Australia (Fig. 3) (*67*, *68*). In the Oceanian archipelagos, the indigenous Papuans encountered the Austronesians (who likely expanded from the Taiwan Strait islands and nearby areas), and multiple waves of migration and admixture events occurred within the past several thousands of years (Fig. 3). About 3.2 ka, the Lapita culture expanded into remote Oceania, carrying mostly Austronesian-related ancestry (*69*). However, this ancestry was nearly completely replaced by Papuan-related ancestry ~2.7 to 2.3 ka in the most western islands of Oceania such as modern-day Vanuatu, presumably resulting from continuous gene flow from the Bismarck archipelago, northeast of New Guinea (*70*, *71*). Additional migration from present-day Philippines to the Mariana Islands in Micronesia has been proposed, highlighting the complexity of the peopling of Oceania (Fig. 3) (*72*). In the past millennium, an influx of Polynesian-related ancestry was detected in Vanuatu people (*72*). Whether Polynesians admixed with Native Americans remains controversial. Analyses of present-day genomes suggest that Polynesians contacted Native South Americans ~800 years ago (*73*), whereas another study did not support this hypothesis (*71*).

Siberia

In Siberia, as mentioned, the APS lineage is best represented by an ~9.8-kyr-old individual from the Kolyma River site in eastern Siberia and an ~14-kyr-old individual from Lake Baikal (Fig. 3) (*40*, *47*). Besides APS, the other lineage formed through admixture between East Asians and populations associated with the ANE lineage formed the basal American branch, whose descendants crossed Beringia to ultimately arrive in the Americas (*74*). Individuals carrying APS ancestry were ancestral to present-day groups of northeastern Siberia (such as the Chukchi, Koryak, and Itelmen) and represent the closest relatives of Native Americans outside the Americas (*74*). During the early- to mid-Holocene in eastern Siberia, APS-related populations were replaced by Neo-Siberians, who carried mainly East Asian ancestry with varying proportions of West Eurasian steppe ancestry (*40*). From the Early Neolithic to the Bronze Age, a genetic transition occurred in Lake Baikal, related to prolonged gene flow between East Asian ancestries (from Amur River Basin) and ANE-related ancestries (*47*).

Americas

After the formation of the basal American branch in Siberia, the initial peopling of Beringia was followed by a standstill (*74*). The common

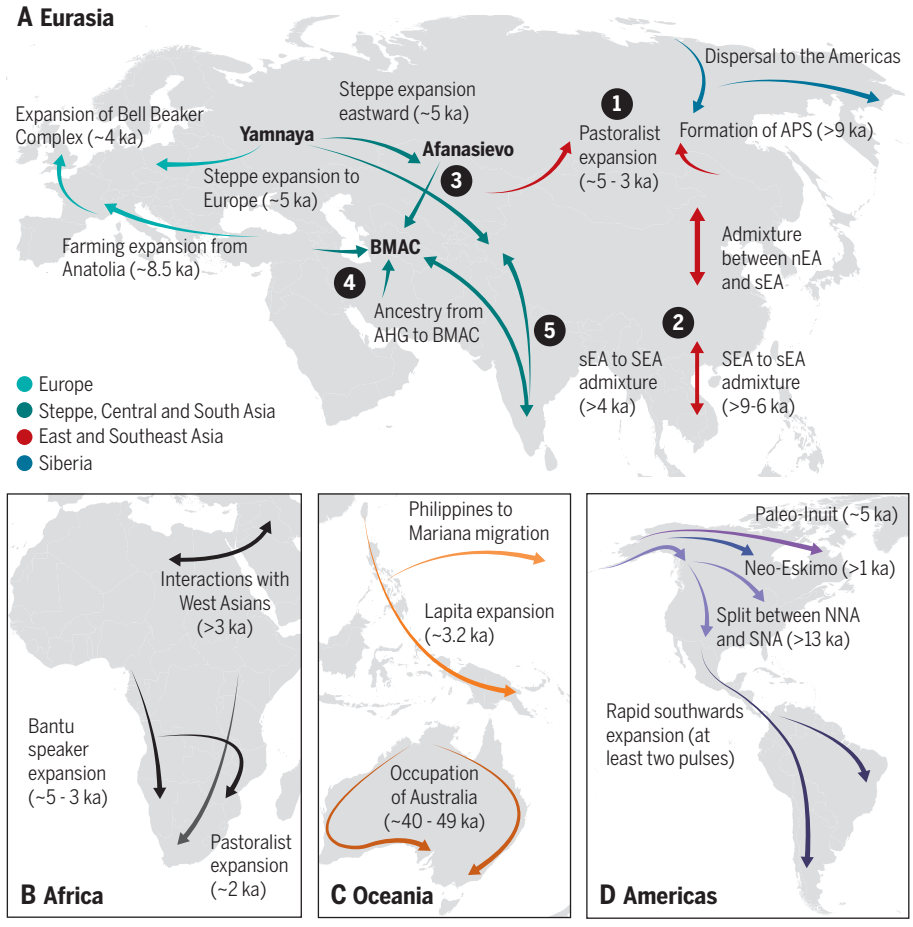


Fig. 3. Major population dynamics across the globe after the LGM. (**A**) Eurasia. In Siberia (blue arrows), APS were formed from the admixture between ancient East Asians and ANE. In East Asia and Southeast Asia (red arrows), (1) Steppe ancestry was introduced through multiple waves of pastoralist expansion ~5 to 3 ka in Mongolia. (2) Gene flow was detected between northern East Asians (nEA) and southern East Asians (sEA) as well as sEA and Southeast Asians (SEA) in Steppe, Central and South Asia (dark green arrows). (3) Steppe expansion to South Asia ~4 ka. (4) The formation of BMAC (Bactria-Margiana Archaeological Complex)–related populations consisted of ancestries from Iranian farmers (60 to 65%) and Anatolian farmers (20 to 25%), with minor contributions from West Siberian hunter-gatherers (~10%) and Andamanese hunter-gatherers (AHG) (2 to 5%). (5) The mixture of ANI and ASI led to the main genetic cline in South Asia. In Europe (light blue arrows), farming expanded from Anatolia, and steppe ancestry expanded from Pontic-Caspian-Ural steppe region, with a subsequent expansion of Bell Beaker Complex to the British Isles. (**B**) Africa. Bantu speaker expansion spread agriculture and West African–related ancestry across Africa, and pastoralists expanded from eastern to southern Africa. Africans also showed connections with West Asians. (**C**) Oceania. Two major expansion routes included the Lapita expansion and a Philippines to Mariana islands migration. Australia has been occupied by modern humans for ~49 kyr, and neither mitochondrial genomes nor Y chromosomes suggest recent geneflow into Australia. (**D**) Americas. Major events during the peopling of the Americas.

ancestor of Northern Native Americans (NNA) and Southern Native Americans (SNA) branched off from Ancient Beringians, best represented by an ~11.5-kyr-old individual from present-day Alaska (Fig. 3) (*74*). Around 17.5 to 14.6 ka, NNA and SNA likely split from each other probably south of the North American ice sheets, as indicated by an ~12.8-kyr-old child from the Anzick site in Montana. This site is associated with the Clovis complex, which is the oldest widespread archaeological complex defined in North America (*75*). After the split, NNA ancestry was restricted to northern North America. In North America, before ~9 ka, admixture between the NNA and SNA groups likely gave rise to the ~9-kyr-old Kennewick man and ancient Algonquians (*74*). SNA dispersed across the Americas widely and rapidly, supported by genetic similarity among ancient genomes from North and South America dating to ~13 to 10 ka (Fig. 3) (*74*). The dispersal of SNA into South America may have occurred with at least two pulses, owing to different levels of affinity that Anzick has to ancient individuals in Argentina, Brazil, and Chile (*74*). Around 5 ka, much of the American Arctic was settled by the Paleo-Inuit, being largely displaced themselves ~800 years ago by a "Neo-Eskimo" lineage—the ancestor of present-day Inuit and Yup'ik (Fig. 3) (*76*).

Future perspectives

The assembly of the first human genome (*1*, *2*) has driven technological progress, allowing us to use aDNA to extend our understanding of human history. More than 6000 ancient genomes have been reconstructed to date, but we have only scratched the surface of human history, and many gaps remain to be filled. Because of variation in access and preservation, the vast majority of ancient genomes (especially those older than 30 kyr) are from Northern Eurasia, with limited sampling in Africa, Asia, Americas, and Oceania. Thus, efforts should be focused on obtaining ancient genomes from these regions.

Technological and computational improvements are also anticipated to extend our abilities to make use of ancient genomes. Imputation of missing information is a powerful tool to complement the fragmentary genetic data, and capturing aDNA from sediments allows for the recovery of aDNA independent of subfossils, which greatly increases the availability of source materials for aDNA research (*7*).

Researchers are also working on retrieving proteomic, isotopic, microbiomic, and epigenetic data from ancient individuals, further complementing the scope of the aDNA field (*7*). With recent successes in obtaining DNA from megafauna older than 1 million years (*77*), it seems increasingly feasible that a deeper evolutionary path can be reconstructed in the near future, for not only modern humans, Neanderthals, and Denisovans but also archaic humans whose genomic information has not yet been obtained. Possible sources of these genomes include newly published fossils from China, for which the species typing remains debated (*78*).

Apart from the increase in knowledge about human history, the past decade of aDNA research also revealed insights into our understanding of human biology. For example, a major genetic risk factor for severe COVID-19 is inherited from Neanderthals (*79*). With advancements in gene-editing techniques, we are now ready to characterize adaptive variants identified from ancient genomes to better appreciating how evolution shaped present-day human genomic architecture (*7*).

Last, our past guides how we should face new challenges and provides lessons for climate change and pandemics. Genetic and archaeological evidence suggests that humans explored, survived, and populated various environments, including extremely harsh arctic conditions. Detailed reconstruction of the past, especially for challenging conditions such as those encountered during the LGM, will aid in revealing human adaptation to the extreme environment. Although this paper focuses on human evolution, aDNA studies of human pathogens inform on pathogenic evolution and host-pathogen interactions through time, showing promise for a better understanding of our adaptations to infectious agents (*80*). The increasing number of ancient genomes being analyzed each year reveals numerous stories of our past, the knowledge of which will also guide us to embrace our future.

REFERENCES AND NOTES

1. International Human Genome Sequencing Consortium, *Nature* **409**, 860–921 (2001).
2. J. C. Venter *et al*., *Science* **291**, 1304–1351 (2001).
3. D. Reich *et al*., *Nature* **468**, 1053–1060 (2010).
4. F. Racimo, S. Sankararaman, R. Nielsen, E. Huerta-Sánchez, *Nat. Rev. Genet.* **16**, 359–371 (2015).
5. P. Skoglund, I. Mathieson, *Annu. Rev. Genomics Hum. Genet.* **19**, 381–404 (2018).
6. R. E. Green *et al*., *Science* **328**, 710–722 (2010).
7. L. Orlando *et al*., *Nat. Rev. Meth. Prim.* **1**, 1–26 (2021).
8. J. Krause *et al*., *Nature* **464**, 894–897 (2010).
9. K. Prüfer *et al*., *Nature* **505**, 43–49 (2014).
10. K. Prüfer *et al*., *Science* **358**, 655–658 (2017).
11. F. Mafessoni *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 15132–15136 (2020).
12. L. Chen, A. B. Wolf, W. Fu, L. Li, J. M. Akey, *Cell* **180**, 677–687. e16 (2020).
13. A. Durvasula, S. Sankararaman, *Sci. Adv.* **6**, eaax5097 (2020).
14. B. Vernot *et al*., *Science* **352**, 235–239 (2016).
15. J. Choin *et al*., *Nature* **592**, 583–589 (2021).
16. F. A. Villanea, J. G. Schraiber, *Nat. Ecol. Evol.* **3**, 39–44 (2019).
17. B. M. Peter, 100,000 years of gene flow between Neandertals and Denisovans in the Altai mountains. bioRxiv [Preprint] 15 March 2020; doi:10.1101/2020.03.13.990523.
18. M. Petr *et al*., *Science* **369**, 1653–1656 (2020).
19. C. Posth *et al*., *Nat. Commun.* **8**, 16046 (2017).
20. Q. Fu *et al*., *Nature* **524**, 216–219 (2015).
21. M. Hajdinjak *et al*., *Nature* **592**, 253–257 (2021).
22. V. Slon *et al*., *Nature* **561**, 113–116 (2018).
23. A. Bergström, C. Stringer, M. Hajdinjak, E. M. L. Scerri, P. Skoglund, *Nature* **590**, 229–237 (2021).
24. C. Posth *et al*., *Curr. Biol.* **26**, 827–833 (2016).
25. P. Gunz *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 6094–6098 (2009).
26. E. M. L. Scerri *et al*., *Trends Ecol. Evol.* **33**, 582–594 (2018).
27. E. K. F. Chan *et al*., *Nature* **575**, 185–189 (2019).
28. P. Skoglund *et al*., *Cell* **171**, 59–71.e21 (2017).
29. M. Lipson *et al*., *Nature* **577**, 665–670 (2020).
30. J.-J. Hublin *et al*., *Nature* **546**, 289–292 (2017).
31. C. M. Schlebusch *et al*., *Science* **358**, 652–655 (2017).
32. M. G. Llorente *et al*., *Science* **350**, 820–822 (2015).
33. M. van de Loosdrecht *et al*., *Science* **360**, 548–552 (2018).
34. K. Harvati *et al*., *Nature* **571**, 500–504 (2019).
35. M. A. Yang, Q. Fu, *Trends Genet.* **34**, 184–196 (2018).
36. Q. Fu *et al*., *Nature* **514**, 445–449 (2014).
37. K. Prüfer *et al*., *Nat. Ecol. Evol.* **5**, 820–825 (2021).
38. Q. Fu *et al*., *Nature* **534**, 200–205 (2016).
39. A. Seguin-Orlando *et al*., *Science* **346**, 1113–1118 (2014).
40. M. Sikora *et al*., *Nature* **570**, 182–188 (2019).
41. M. Raghavan *et al*., *Nature* **505**, 87–91 (2014).
42. M. A. Yang *et al*., *Curr. Biol.* **27**, 3202–3208.e9 (2017).
43. J.-J. Hublin *et al*., *Nature* **581**, 299–302 (2020).
44. E. Svensson *et al*., *Curr. Biol.* **31**, 2973–2983.e9 (2021).
45. V. Villalba-Mouco *et al*., *Curr. Biol.* **29**, 1169–1177.e7 (2019).
46. X. Mao *et al*., *Cell* **184**, 3256–3266.e13 (2021).
47. H. Yu *et al*., *Cell* **181**, 1232–1245.e20 (2020).
48. R. Fregel *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6774–6779 (2018).
49. M. E. Prendergast *et al*., *Science* **365**, eaaw6275 (2019).
50. K. Wang *et al*., *Sci. Adv.* **6**, eaaz0183 (2020).
51. M. Vicente, C. M. Schlebusch, *Curr. Opin. Genet. Dev.* **62**, 8–15 (2020).
52. R. Nielsen *et al*., *Nature* **541**, 302–310 (2017).
53. W. Haak *et al*., *Nature* **522**, 207–211 (2015).
54. I. Olalde *et al*., *Nature* **555**, 190–196 (2018).
55. A. Goldberg, T. Günther, N. A. Rosenberg, M. Jakobsson, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 2657–2662 (2017).
56. M. E. Allentoft *et al*., *Nature* **522**, 167–172 (2015).
57. C. Jeong *et al*., *Cell* **183**, 890–904.e29 (2020).
58. V. M. Narasimhan *et al*., *Science* **365**, eaat7487 (2019).
59. V. Shinde *et al*., *Cell* **179**, 729–735.e10 (2019).
60. C. Jeong *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11248–E11255 (2018).
61. C.-C. Wang *et al*., *Nature* **591**, 413–419 (2021).
62. M. A. Yang *et al*., *Science* **369**, 282–288 (2020).
63. T. Wang *et al*., *Cell* **184**, 3829–3841.e21 (2021).
64. C. Ning *et al*., *Nat. Commun.* **11**, 2700 (2020).
65. H. McColl *et al*., *Science* **361**, 88–92 (2018).
66. J. F. O'Connell, J. Allen, *J. Archaeol. Sci.* **56**, 73–84 (2015).
67. A. Bergström *et al*., *Curr. Biol.* **26**, 809–813 (2016).
68. R. Tobler *et al*., *Nature* **544**, 180–184 (2017).
69. P. Skoglund *et al*., *Nature* **538**, 510–513 (2016).
70. M. Lipson *et al*., *Curr. Biol.* **28**, 1157–1165.e7 (2018).
71. C. Posth *et al*., *Nat. Ecol. Evol.* **2**, 731–740 (2018).
72. I. Pugach *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2022112118 (2021).
73. A. G. Ioannidis *et al*., *Nature* **583**, 572–577 (2020).
74. E. Willerslev, D. J. Meltzer, *Nature* **594**, 356–364 (2021).
75. M. Rasmussen *et al*., *Nature* **506**, 225–229 (2014).
76. P. Flegontov *et al*., *Nature* **570**, 236–240 (2019).
77. T. van der Valk *et al*., *Nature* **591**, 265–269 (2021).
78. Q. Ji, W. Wu, Y. Ji, Q. Li, X. Ni, *Innovation* **2**, 100132 (2021).
79. H. Zeberg, S. Pääbo, *Nature* **587**, 610–612 (2020).
80. M. A. Spyrou, K. I. Bos, A. Herbig, J. Krause, *Nat. Rev. Genet.* **20**, 323–340 (2019).

ACKNOWLEDGMENTS

We thank M. A. Yang, E. A. Bennett, V. Kumar, J. Gardner, S. Qiao, M. Zhang, Z. Chen, W. Wang, B. Miao, and T. Wang for discussions and editing and S. Qiao, J. Ran, and J. Xue for organizing published aDNA data. **Funding:** This work was supported by the Chinese Academy of Sciences (CAS) and the Ministry of Finance of the People's Republic of China (XDB26000000 and YSBR-019), the National Natural Science Foundation of China (41925009), the Tencent Foundation (through the XPLORER PRIZE), the Howard Hughes Medical Institute (grant 55008731), and the Shanghai Qi Zhi Institute. **Author contributions:** Conceptualization: Q.F.; writing, original draft: Y.L., X.M., and Q.F.; writing, review and editing: Q.F., J.K., Y.L., and X.M. **Competing Interests:** The authors declare no competing interests.

10.1126/science.abi8202

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



ARCHAEOLOGY

Early shell beads from North Africa

Shell beads are among the earliest artifacts in human prehistory used to define personal identity. Their presence also signals the increasing cognitive capacity of our hominid ancestors. Sehasseh *et al.* have discovered marine shell beads at the Middle Stone Age site of Bizmoune Cave, Morocco. The beads, made by early *Homo sapiens* and dated to 142,000 years ago, are the earliest yet recovered from the continent. The use of beads across the span of the Middle Stone Age in northern Africa, the authors argue, reflects clear evidence of cultural continuity and suggests a mosaic of cultural development across the African continent during this period that contrasts with contemporary archaeological evidence from South Africa. —MSA *Sci. Adv.* 10.1126/sciadv.abi8620 (2021).

Shell beads from Bizmoune Cave, Morocco, provide insight into early human cultural practices in northern Africa.

IMAGE: SEHASSEH *ET AL.*

SYNTHETIC BIOLOGY

From carbon dioxide to starch: no plants required

Many plants turn glucose from photosynthesis into polymers that form insoluble starch granules ideal for long-term energy storage in roots and seeds. Cai *et al.* developed a hybrid system in which carbon dioxide is reduced to methanol by an inorganic catalyst and then converted by enzymes first to three and six carbon sugar units and then to polymeric starch. This artificial starch anabolic pathway relies on engineered recombinant enzymes from many different source organisms. It can be tuned to produce amylose or amylopectin at excellent rates and efficiencies relative to other synthetic carbon fixation systems and, depending on the metric used, even to field crops. —MAF

Science, abh4049, this issue p. 1523

ALZHEIMER'S DISEASE

Determining Alzheimer's trajectories

The locus coeruleus (LC) has been recognized as the initial site of tau aggregation in Alzheimer's disease (AD). However, how LC pathology relates to disease progression and symptoms remains to be elucidated. Jacobs *et al*. used longitudinal imaging data in cognitively impaired and unimpaired individuals and postmortem data from two large cohorts of patients with AD, and showed that the integrity of the LC is associated with the main pathological and cognitive features of early AD. This suggests that monitoring LC using noninvasive imaging methods might help to predict cognitive trajectories and AD pathology later in life. —MM

Sci. Transl. Med. **13**, eabj2511 (2021).

PALEOANTHROPOLOGY

Early footsteps in the Americas

Despite a plethora of archaeological research over the past century, the timing of human migration into the Americas is still far from resolved. In a study of exposed outcrops of Lake Otero in White Sands National Park in New Mexico, Bennett *et al.* reveal numerous human footprints dating to about 23,000 to 21,000 years ago. These finds indicate the presence of humans in North America for approximately two millennia during the Last Glacial Maximum south of the migratory barrier created by the ice sheets to the north. This timing coincided with a Northern Hemispheric abrupt warming event, Dansgaard-Oeschger event 2, which drew down lake levels and allowed humans and megafauna to walk on newly exposed surfaces, creating tracks that became preserved in the geologic record. —AMS

Science, abg7586, this issue p. 1528

TOPOLOGICAL OPTICS

Topologically locked for emission

The output power from a laser system can be increased by forming an array of lasers; however, because the individual

lasers are independent, the resultant output may not be coherent. Dikopoltsev *et al.* report on the realization of a topological vertical-cavity surface-emitting laser (VCSEL) array. The topological nature of the array-based laser emission was achieved through a combination of topological in-plane propagation of evanescent light linking the vertical cavity surface-emitting lasers of the array. The topological features of the array force injection locking, making all emitters (30 in this case) act as a single coherent laser. This development will be important for realizing large-scale coherent laser arrays. —ISO

Science, abj2232, this issue p. 1514

PLANT SCIENCE

Building a wider root network

As roots grow through the soil, lateral branches develop to broaden the reach toward water and nutrients. Just where on the main root these lateral roots will develop is organized by an oscillatory root clock. Dickinson *et al.* have now identified an initiating signal in the form of the signaling molecule retinal. Oscillatory expression of a retinal-binding protein, a temperature-sensitive lipocalin, defines sites where retinal signals can initiate growth. —PJH

Science, abf7461, this issue p. 1532

BATTERIES

Silicon anode solid-state batteries

Research on solid-state batteries has focused on lithium metal anodes. Alloy-based anodes have received less attention in part due to their lower specific capacity even though they should be safer. Tan *et al.* developed a slurry-based approach to create films from micrometer-scale silicon particles that can be used in anodes with carbon binders. When incorporated into solid-state batteries, they showed good performance across a range of temperatures and excellent cycle life in full cells. —MSL

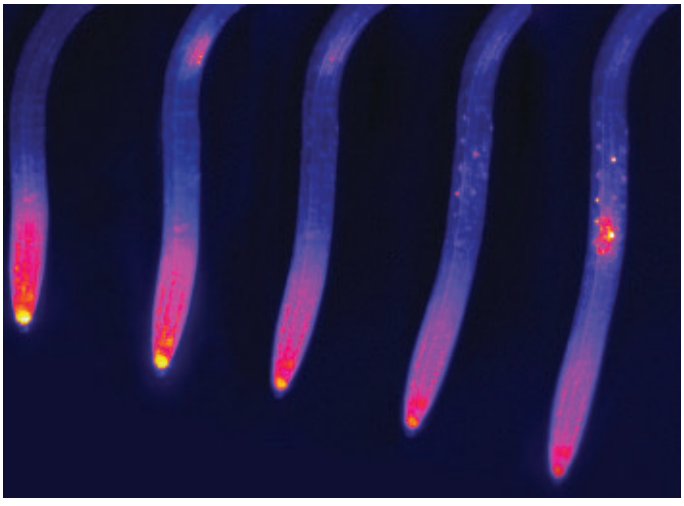
Science, abg7217, this issue p. 1494

CATALYSIS

Going with the grain

Changes in surface structure can make metal nanoparticles supported on oxides more active for certain catalytic reactions. Huang *et al.* show that steam pretreatment of palladium nanoparticles on alumina led to a high density of twin boundaries, unlike other oxidation and reduction pretreatments. The density of these stable grain boundaries at the surface correlated with higher methane oxidation rates and lower temperatures for the initiation of the reaction. The introduction of additional defect sites through laser ablation created even more active catalysts. —PDS

Science, abj5291, this issue p. 1518



Fluorescence microscopy image revealing a nascent lateral root node (far right)



IN OTHER JOURNALS

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

SYNTHETIC BIOLOGY

Three codes are better than one

Within living cells, membrane encapsulation or membraneless organelles can be used to separate distinct molecular processes. A challenge in synthetic biology is how to make enzymes that contain non-natural amino acids without altering the entire make-up of the cell. Reinkemeier *et al.* generated ultrathin, synthetic film–like organelles on the surfaces of the endoplasmic reticulum, Golgi body, plasma membrane, and mitochondria. These subresolution synthetic films allowed the production of functionally distinct enzymes inside different organelles within the same cell. The film-like organelles were used to equip eukaryotic cells with two orthogonal genetic codes, in which stop codons were engineered to incorporate noncanonical amino acids. This manipulation enabled the specific reprogramming of multiple distinct translational machineries, each with single-residue precision. Thus, the stage is set to engineer different proteins with distinct noncanonical functionalities within eukaryotic cells that simultaneously execute three distinct genetic codes. —SMH

Cell 10.1016/j.cell.2021.08.001 (2021).

IMMUNOLOGY

Preterm brain injury

Although medical advances have resulted in the increased survival of extremely premature infants (less than 28 weeks), the chance of severe morbidities, including neurological damage, continues to be high. Seki *et al.* analyzed the gut microbiota, as well as the immune and neurological development, of 60 extremely premature infants to determine the drivers of such damage. Babies with severe brain injury showed higher percentages of brain γδ T cells compared with premature infants with mild to no brain injury. Injury was accompanied by enhanced levels of vascular endothelial growth factor and proinflammatory cytokines such as interleukin-17.

IMAGES (LEFT TO RIGHT): DICKINSON ET AL.; GORODENKOFF/ISTOCK.COM

NEUROSCIENCE

Sleep, cortisol, and consolidation

Reactivation of a memory can make it malleable to subsequent change during reconsolidation. Targeted pharmacological and behavioral manipulations after memory reactivation can modulate reconsolidation and modify the memory. Antypa *et al.* investigated whether changes in stress hormone levels during sleep affected later memory of a reactivated episode. The authors recited a story accompanied by a slide show to a group of male and female subjects. If subjects were given treatment to block cortisol synthesis during early morning sleep, then their 3-day-old memory of the story was more precisely recalled than if the early morning cortisol spike was uncontrolled. However, this improvement only occurred if the subjects had been given a visual cue for the story just before anti-cortisol treatment. —PRS

J. Neurosci. **41**, 7259 (2021).

During sleep, early morning changes in stress hormone levels can alter memory.

Moreover, overgrowth of *Klebsiella* spp. in the gut was strongly predictive for brain injury. Separating cause from effect is tricky, and the possibility remains that brain injury drives these immune cell alterations and gut dysbiosis rather than the other way around. —STS

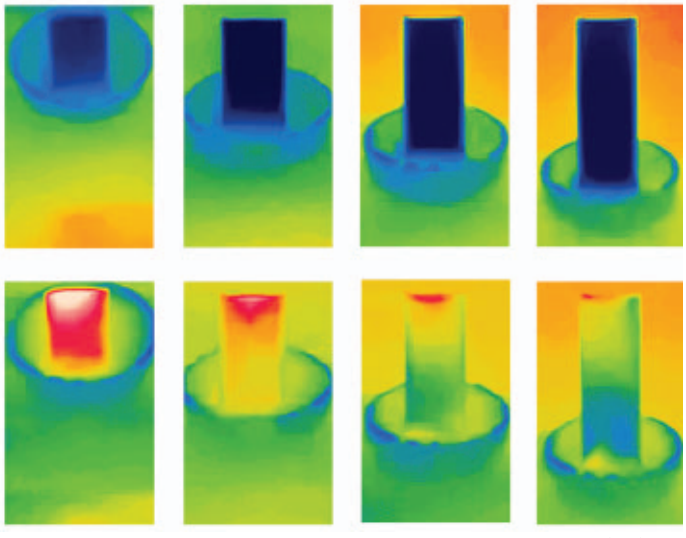
Cell Host Microbe 10.1016/j.chom.2021.08.004 (2021).

MEMBRANES

Solar purification of oily wastewater

Solar evaporation is a potentially low-cost, simple way to obtain clean water using limited resources. Conventional evaporators work well for desalination, but if the feed contains oily emulsions, then these can block water channels and foul the membranes. Wu *et al.* deposited polydopamine onto microporous polypropylene membranes. Polydopamine makes the membranes hydrophilic, which is necessary for pulling water up from the reservoir while also giving them underwater oil repellence. This separation of the evaporation regions from the area rejecting the oily contaminants helps to avoid evaporation of the oily materials and improves heat management. Further, fabrication of the evaporators in a peaked roof structure enhances the surface area for evaporation. Outdoor experiments showed the production of approximately 8 kilograms per square meter per day of clean water. —MSL

ACS Appl. Mater. Interfaces **13**, 39513 (2021).



Infrared images of membrane evaporators of different heights in the dark (top) and in sunlight (bottom)

IMAGE: WU ET AL., ACS APPLIED MATERIALS & INTERFACES, 13, 3951 (2021).

SURFACE POLYMERS

Quickly sowing micelle seeds

The living growth of polymer micelles on surfaces has been controlled through nucleation kinetics. Tang *et al*. created surface micelles from block copolymers of hydrophobic poly(γ-benzyl-L-glutamate) and hydrophilic poly(ethylene glycol) dissolved in organic solution. After slowly adding enough water to pass the critical micelle concentration, small aggregates with swollen hydrophobic cores formed that anchored as seeds to a hydrophobic polystyrene substrate. The remaining solution aggregates quickly de-swelled and could not form new seeds. Instead, elongated surface micelles formed by anchoring of aggregates to seeds through hydrophilic corona interactions, followed by hydrophobic core reorganization. The micelle length could be increased with additional cycles of exposure to dissolved aggregates, and shorter micelles grew on more hydrophilic substrates. —PDS

J. Am. Chem. Soc. 10.1021/jacs.1c06020 (2021).

CHEMICAL KINETICS

Interface formation in lithium-ion batteries

Interfacial phenomena play a determining role in electrochemical systems such as lithium-ion batteries (LIBs), but despite decades of research, their chemical characterization remains a long-standing challenge. Using a combination of graph representation, density functional theory, and machine learning, Xie *et al.* developed an automated reaction network for the formation of lithium ethylene monocarbonate (LEMC) in the solid-electrolyte interphase, a critical process in LIBs. The analysis of this comprehensive reaction network, consisting of millions of hypothetical reactions, identified both existing and previously unknown feasible routes to LEMC and provided insights in the reaction mechanism under various conditions. These findings could be useful for addressing limiting factors in LIB development. —YS

J. Am. Chem. Soc. 10.1021/jacs.1c05807 (2021).

MUSCULAR DYSTROPHY

A heartening advance

Duchenne muscular dystrophy is an X-linked genetic disorder that can be caused by diverse mutations in the dystrophin gene. These mutations cause abnormalities in both skeletal and cardiac muscles, with eventual respiratory failure and cardiomyopathy. As technology for respiratory support continues to improve, cardiomyopathy in these patients has become increasingly problematic. Atmanli *et al*. used CRISPR/Cas9 gene editing to correct the dystrophin gene abnormality in patient-derived induced pluripotent stem cells and then generated cardiomyocytes from these cells. This approach improved cardiomyocyte function in vitro and transcriptional patterns in a mouse model, suggesting potential therapeutic applications. —YN

Circ. Res. **129**, 602 (2021).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

GENOMICS

Next steps for the human genome reference

Twenty years since the publication of the first human genome sequences, Rood and Regev reflect on the legacy and next steps of the Human Genome Project in a Perspective. They discuss the impacts of the human reference and highlight the lasting issues that remain to be overcome, particularly in terms of diversity. Indeed, it remains unanswered whether a single human reference genome is suitable for applications today. How the human genome sequence can be used to offer mechanistic insight and improvement to human health are also key challenges that remain a central theme in research. Although much still needs to be done, it is clear the Human Genome Project has had and will continue to have enormous impact throughout science and society. —GKA

Science, abl5403, this issue p. 1442

ENZYME MECHANISM

Tracing a path to ethylene

Ethylene is produced industrially from fossil carbon sources, but plants and microbes produce small amounts through a handful of unusual enzymatic reactions. Copeland *et al.* studied a microbial ethylene-forming enzyme using oxygen isotope–tracing experiments and biochemical assays to test mechanistic proposals. This enzyme, which uses a non–heme iron center to activate oxygen, catalyzes two distinct oxidation reactions with different mechanisms. One reaction is entirely off-pathway and results in fragmentation of a co-substrate, arginine. The other reaction can completely fragment 2-oxoglutarate into ethylene, bicarbonate, and two molecules of carbon dioxide; however, it occasionally derails and yields an omega–hydroxy acid product. —MAF

Science, abj4290, this issue p. 1489

HUMAN GENOMICS

Repeats associated with phenotype

The degree to which repeated sequences within a genome affect human phenotypes has been difficult to establish. Mukamel *et al.* examined thousands of genomes in the UK Biobank and found that some of the largest effects of common genetic variants on human phenotypes, including those with clinical relevance, arise from protein-coding repeat polymorphisms (see the Perspective by Gymrek and Goren). Mapping the effects of the size and copy number of these repeated protein domains links genetic variation to human phenotypes, including lipoprotein(a) concentration, height, and male pattern balding. Furthermore, the alleles and frequencies of these repeated sequences differ between individuals of African and European descent, resulting in differences between the populations with clinical relevance for traits including lipoprotein(a) levels, a risk factor for coronary artery disease. —LMZ

Science, abg8289, this issue p. 1499; see also abl7794, p. 1440

SUPERCONDUCTIVITY

Restoring strangeness

Cuprate superconductors host a number of complex phases, and elucidating their relationship to each other may shed light on the nature of cuprate superconductivity. Two of these intriguing phases are the strange metal phase and the charge-density wave (CDW) order, which appear next to each other in the cuprate phase diagram. The hallmark of the strange metal phase, the linear dependence of electrical resistivity on temperature, disappears in the CDW phase. Wahlberg *et al.* strained thin films of the cuprate $YBa_2Cu_3O_{7-\delta}$ to suppress CDW order (see the Perspective by Le Tacon). Using resonant inelastic x-ray scattering and transport measurements, the researchers found that suppressing CDW restored the linear-in-temperature resistivity in the CDW portion of the phase diagram. —JS

Science, abc8372, this issue p. 1506; see also abi9685, p. 1438

QUANTUM OPTICS

Quantum trapping and shuffling

Programmable arrays of atoms or ions trapped in optical potentials have recently emerged as a leading platform for quantum simulation. Being able to interface into these arrays to access the quantum information being processed and pass it along to another module remains a challenge. Đorđević et *al.* developed a hybrid quantum system that combines atoms held in optical tweezers and a nanophotonic cavity to demonstrate full quantum control, efficient quantum nondestructive readout, and entanglement of atom pairs (see the Perspective by Kaufman). By combining atomic manipulation both inside and away from the cavity field and shuffling the atom qubits into and out of the cavity mode, the authors demonstrate a viable optical interface that could be scaled to larger systems. —ISO

Science, abi9917, this issue p. 1511; see also abk2617, p. 1436

HEART DEVELOPMENT

Pluripotency factor drives cardiogenesis

Research indicates that the adult mammalian heart does not contain cardiac stem cells, and the vast majority of cardiomyocytes do not divide. Heart regeneration is thus limited after injury. The postmitotic nature of cardiomyocytes blocks cardiac tumor formation but at the same time minimizes cardiomyocyte renewal. Chen *et al.* report that cell type–specific expression of pluripotency factors dedifferentiates adult cardiomyocytes to a state that resembles fetal cardiomyocytes, enabling adult cardiomyocytes to reenter mitosis (see the Perspective by Wang and Blau). Cardiomyocytes can be reprogrammed to a pluripotent state when expression of pluripotency factors is sustained over an extended period. If cardiomyocytes are only partially reprogrammed, then the heart regenerates without tumor formation. —BAP

Science, abg5159, this issue p. 1537 see also abl8679, p. 1439

PHYSIOLOGY

Degenerating with age with NF-κB

Aging is the most common cause of osteoarthritis. In a mouse model of spontaneous, age-related osteoarthritis of the knee, Catheline *et al*. found that nuclear factor κB (NF-κB) signaling in articular chondrocytes increased with age and correlated with the onset of joint degeneration. Activating NF-κB signaling in chondrocytes accelerated the onset of joint degeneration and promoted the expression of a proinflammatory, senescence-associated secretory phenotype, which was further enhanced by the NF-κB subunit p65. In contrast, the p50 subunit inhibited the expression of senescence-associated secretory phenotype (SASP) factors and cartilage degeneration. —AMV

Sci. Signal. **14**, eabf3535 (2021).

ANTIVIRAL IMMUNITY

Armor against adenovirus

Reactivation of human adenovirus (HAdV) is a major cause of mortality in children and immunocompromised individuals after allogeneic hematopoietic stem cell transplantation (allo-HSCT). Effective therapies for severe HAdV infections are lacking.

Using a three-dimensional intestinal organoid system, Jung *et al.* investigated how natural killer (NK) cells recognize and kill HAdV-infected intestinal epithelial cells. HAdV-infected intestinal epithelial cells strongly up-regulated a ligand called HLA-F that activates the NK cell receptor KIR3DS1 and enhances recognition and killing by NK cells. HAdV-infected pediatric allo-HSCT recipients who received KIR3DS1$^+$/HLA-Bw4$^+$ donor cells exhibited protection from severe HAdV infection and faster HAdV clearance. These findings suggest that the KIR3DS1/HLA-F axis is a promising target for the treatment of severe HAdV reactivation after allo-HSCT. —HMI

Sci. Immunol. **6**, eabe2942 (2021).

CORONAVIRUS

Stockpiling and control

A triumph that has emerged from the catastrophe of the severe acute respiratory syndrome coronavirus 2 pandemic has been the rapid development of several potent vaccines. However, 18 months into the pandemic and more than 6 months after vaccine approval, wealthy countries remain the major beneficiaries. Wagner *et al.* model the consequences of vaccine stockpiling in affluent countries on disease rates in lower- and middle-income countries and the consequences for the eruption of new variants that could jeopardize the early success of vaccines. For countries that can readily access vaccines, it would be better to share vaccines equitably to lower disease burdens in countries with less access, reduce the cost of having to be constantly vigilant for case imports, and minimize virus evolution. —CA

Science, abj7364, this issue p. 1488

RESEARCH ARTICLE SUMMARY

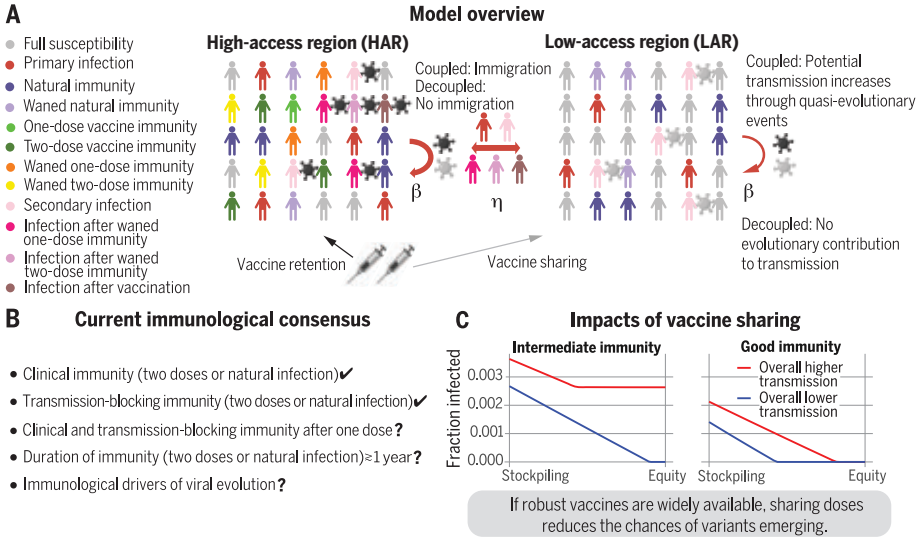
CORONAVIRUS

Vaccine nationalism and the dynamics and control of SARS-CoV-2

Caroline E. Wagner*†, Chadi M. Saad-Roy*†, Sinead E. Morris, Rachel E. Baker, Michael J. Mina, Jeremy Farrar, Edward C. Holmes, Oliver G. Pybus, Andrea L. Graham, Ezekiel J. Emanuel, Simon A. Levin, C. Jessica E. Metcalf, Bryan T. Grenfell*

INTRODUCTION: The COVID-19 pandemic has imposed massive public health burdens and economic costs, with more than 220 million confirmed cases and more than 4.5 million deaths globally so far. In some countries, the rapid deployment of safe and effective vaccines has reduced cases, hospitalizations, and deaths. However, many regions across the world have little access to doses, and infections are resurging. Furthermore, sustained transmission has led to the emergence of novel variants of the causative severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus with increased transmissibility, and against which vaccines may be less effective. Thus, understanding the effects of vaccine nationalism and allocation on the dynamics and control of SARS-CoV-2 is critical.

RATIONALE: We extend recent models of SARS-CoV-2 dynamics to consider the number of cases and potential for viral evolution in two hypothetical regions, one with high access and one with low access to vaccines. This modified framework allows for variation in the strength and duration of natural and vaccinal immunity, which are aspects of the virus and host response that remain uncertain, thus enabling general explorations of a number of potential scenarios. To model the sharing of vaccines, we begin by only coupling the two regions through their vaccination rates. For different immuno-epidemiological scenarios of SARS-CoV-2, we examine the medium- and long-term dynamics in both regions as a function of the fraction of vaccines shared. Then, we couple the regions through immigration and introduce potential transmission increases due to evolution. We use this coupled model to explore local and combined infection numbers and clinical burdens, as well as potential viral adaptation.



Effects of vaccine allocation on the epidemiological and evolutionary trajectories of two regions. (**A**) Each region is described by an immuno-epidemiological model, and the regions are potentially coupled through immigration and transmission increases driven by viral evolution. (**B**) Current consensus on host immune responses and clinical and transmission-blocking protection. (**C**) Epidemiological and evolutionary outcomes for different vaccine allocation schemes given the specific assumptions related to natural and vaccinal immunity described in (B). The terms "intermediate" and "good" immunity in (C) follow the descriptions provided in the main text. Additional scenarios can be explored using the online interactive application (https://grenfelllab.shinyapps.io/vaccine-nationalism/). Schematics of needles and viruses in (A) were created with BioRender.com.

RESULTS: In general, we find that stockpiling vaccines by countries with high availability leads to large increases in infections in countries with low vaccine availability, the magnitude of which depends on the strength and duration of natural and vaccinal immunity. A number of additional subtleties arise when the populations and transmission rates in each country differ depending on evolutionary assumptions and vaccine availability. Furthermore, the movement of infected individuals between countries combined with the possibility of evolutionary increases in viral transmissibility may greatly magnify local and combined infection numbers, suggesting that countries must invest in surveillance strategies to prevent case importation. Dose sharing is likely a high-return strategy because equitable allocation brings nonlinear benefits and also alleviates costs of surveillance (e.g., border testing, genomic surveillance) in settings where doses are sufficient to maintain cases at low numbers. Across a range of immunological scenarios, we find that vaccine sharing is also a powerful tool to decrease the potential for antigenic and transmission evolution of the virus, especially if infections after the waning of natural immunity contribute most to evolutionary potential.

CONCLUSION: Intuitively, our results indicate that unequal vaccine allocation will result in sustained transmission and increased case numbers in regions with low vaccine availability and thus to a higher associated clinical burden compared with a vaccinated population. Under certain scenarios, sustained local transmission could lead to an increased potential for antigenic evolution, which may result in the emergence of variants with novel antigenicity and/or transmissibility and affect epidemiological characteristics globally. Overall, our work underlines the importance of rapid, equitable vaccine deployment and the necessity to export vaccines to regions with low availability in parallel to their becoming available in regions with high access. Coordinated vaccination campaigns across the world, combined with improved surveillance and appropriate nonpharmaceutical interventions to prevent case importation, are imperative. ■

The list of author affiliations is available in the full article online.
***Corresponding author: caroline.wagner@mcgill.ca (C.E.W.); csaadroy@princeton.edu (C.M.S.-R.); grenfell@princeton.edu (B.T.G.)**
†These authors contributed equally to this work.
This is an open-access article distributed under the terms of the Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Cite this article as C. E. Wagner *et al.*, *Science* 373, eabj7364 (2021). DOI: 10.1126/science.abj7364

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abj7364

RESEARCH ARTICLE

CORONAVIRUS

Vaccine nationalism and the dynamics and control of SARS-CoV-2

Caroline E. Wagner[1]*†, Chadi M. Saad-Roy[2]*†, Sinead E. Morris[3], Rachel E. Baker[4,5], Michael J. Mina[6,7], Jeremy Farrar[8], Edward C. Holmes[9,10,11], Oliver G. Pybus[12], Andrea L. Graham[4], Ezekiel J. Emanuel[13], Simon A. Levin[4], C. Jessica E. Metcalf[4,14], Bryan T. Grenfell[4,14]*

Vaccines provide powerful tools to mitigate the enormous public health and economic costs that the ongoing severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic continues to exert globally, yet vaccine distribution remains unequal among countries. To examine the potential epidemiological and evolutionary impacts of "vaccine nationalism," we extend previous models to include simple scenarios of stockpiling between two regions. In general, when vaccines are widely available and the immunity they confer is robust, sharing doses minimizes total cases across regions. A number of subtleties arise when the populations and transmission rates in each region differ, depending on evolutionary assumptions and vaccine availability. When the waning of natural immunity contributes most to evolutionary potential, sustained transmission in low-access regions results in an increased potential for antigenic evolution, which may result in the emergence of novel variants that affect epidemiological characteristics globally. Overall, our results stress the importance of rapid, equitable vaccine distribution for global control of the pandemic.

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has led to more than 220 million infections and more than 4.5 million fatalities to date (*1*). Effective vaccines [e.g., (*2*–*4*)] have now been approved and are actively being deployed, but numerous important questions remain. Eventually, community immunity may be attained through the deployment of vaccines; however, if and when this occurs will be contingent on the characteristics of natural and vaccinal immunity (*5*–*7*). As illustrated by the rapid spread and high transmissibility of the Delta variant (*8*), the evolutionary potential of SARS-CoV-2 is a major potential obstacle for control (*9*).

[1]Department of Bioengineering, McGill University, Montreal, QC H3A 0C3, Canada. [2]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08540, USA. [3]Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY 10032, USA. [4]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08540, USA. [5]Princeton High Meadows Environmental Institute, Princeton University, Princeton, NJ 08540, USA. [6]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [7]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [8]The Wellcome Trust, London, UK. [9]Marie Bashir Institute for Infectious Diseases and Biosecurity, The University of Sydney, Sydney, NSW, Australia. [10]School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW, Australia. [11]School of Medical Sciences, The University of Sydney, Sydney, NSW, Australia. [12]Department of Zoology, University of Oxford, Oxford, UK. [13]Department of Medical Ethics and Health Policy, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. [14]Princeton School of Public and International Affairs, Princeton University, Princeton, NJ 08540, USA.
***Corresponding author: caroline.wagner@mcgill.ca (C.E.W.); csaadroy@princeton.edu (C.M.S.-R.); grenfell@princeton.edu (B.T.G.)**
†These authors contributed equally to this work.

Owing to strong public and political pressures and fear of waning immunity, some countries with high vaccine availability are currently resorting to "vaccine nationalism"—stockpiling vaccines to prioritize rapid access to their citizenry (*10*). Indeed, at the time of writing, 113 and 137 doses per 100 individuals have been administered in the United States and United Kingdom, respectively, whereas an average of 51 and 8.5 doses per 100 individuals have been administered in India and across Africa, respectively (*11*). Recently, the World Health Organization recognized that delayed access to vaccines in countries with low vaccine availability may lead to more evolutionary potential (*12*), which could result in immune escape or other phenotypic changes of interest (e.g., increases in transmission). The emergence of future variants capable of evading natural or vaccinal immune responses could threaten containment efforts globally. These concepts underlie the development of a number of policy tools, including the existing COVID-19 Vaccines Global Access (COVAX) initiative. Furthermore, to ensure that vaccine distribution is ethically sound and equitable, the "fair priority model" has been proposed (*13*–*15*) as a potential replacement to the currently planned proportional allocation (by population size) from COVAX.

Prior work has explored optimal prophylactic vaccine allocation for minimizing the final epidemic size of a fully immunizing infection [i.e., one that can be modeled using a susceptible-infected-recovered (SIR) framework] (*16*); when interaction between communities (or countries) is considered, equal vaccine distribution is increasingly advantageous in terms of minimizing case numbers (*16*). Modeling studies have also shown that coordinated influenza vaccine sharing would reduce the financial and infection burden of influenza outbreaks globally (*17*). Similar problems related to optimizing vaccine allocation have also been explored in networks with community structure (*18*), as well as in the face of economic constraints (*19*) and vaccination coalition formation (*20*), and for SARS-CoV-2 with age (and contact) heterogeneity (*21*, *22*).

We have recently shown that the strength and duration of immunity elicited after infection or receipt of one or two doses of a vaccine will have a crucial impact on the medium-term epidemiological and potential evolutionary outcomes (*5*, *6*). In this study, we extend these analyses to address potential epidemiological and evolutionary consequences of policies of vaccine nationalism or equitable access for a range of assumptions regarding the robustness of host immune responses. In reality, vaccine distribution is a public goods problem (*23*), and the optimal "global" allocation projected based on evolutionary and immunological uncertainties may differ from national optima because of the actual economic landscape of each country. In all cases, every nation, however, has a shared interest in reducing the potential for novel strains to arise, which is achievable by minimizing the global infection burden.

We consider a transnational extension of our model, comprising two countries with possibly different population sizes and seasonal transmission patterns. One country, the high-access region (HAR), chooses to allocate a fraction f of the total vaccine supply to the low-access region (LAR). The underlying immuno-epidemiological models for both countries account for both the duration of natural and vaccinal immunity and the residual decrease in host susceptibility to infection (relative to immunologically naïve individuals) after full natural or vaccinal immunity has waned; these models are described in detail in (*5*, *6*), where the more detailed structure (*6*) accounts for immunity after receipt of one or two vaccine doses. In the first "decoupled" framework (top panel of Fig. 1), we assume that the epidemiological dynamics of both countries are entirely independent, with the exception of their respective vaccination rates; we also compute a measure for the global potential for viral evolution of immune escape (*6*). In the second "coupled" framework (bottom panel of Fig. 1), we allow for immigration of infected individuals between the countries at rate η (*19*). Additionally, we approximate the stochastic occurrence of potential transmission increases (PTIs) in each country: Briefly, if the "potential net viral adaptation rate" [see supplementary materials, fig. S3, and (*6*)] exceeds a threshold, then

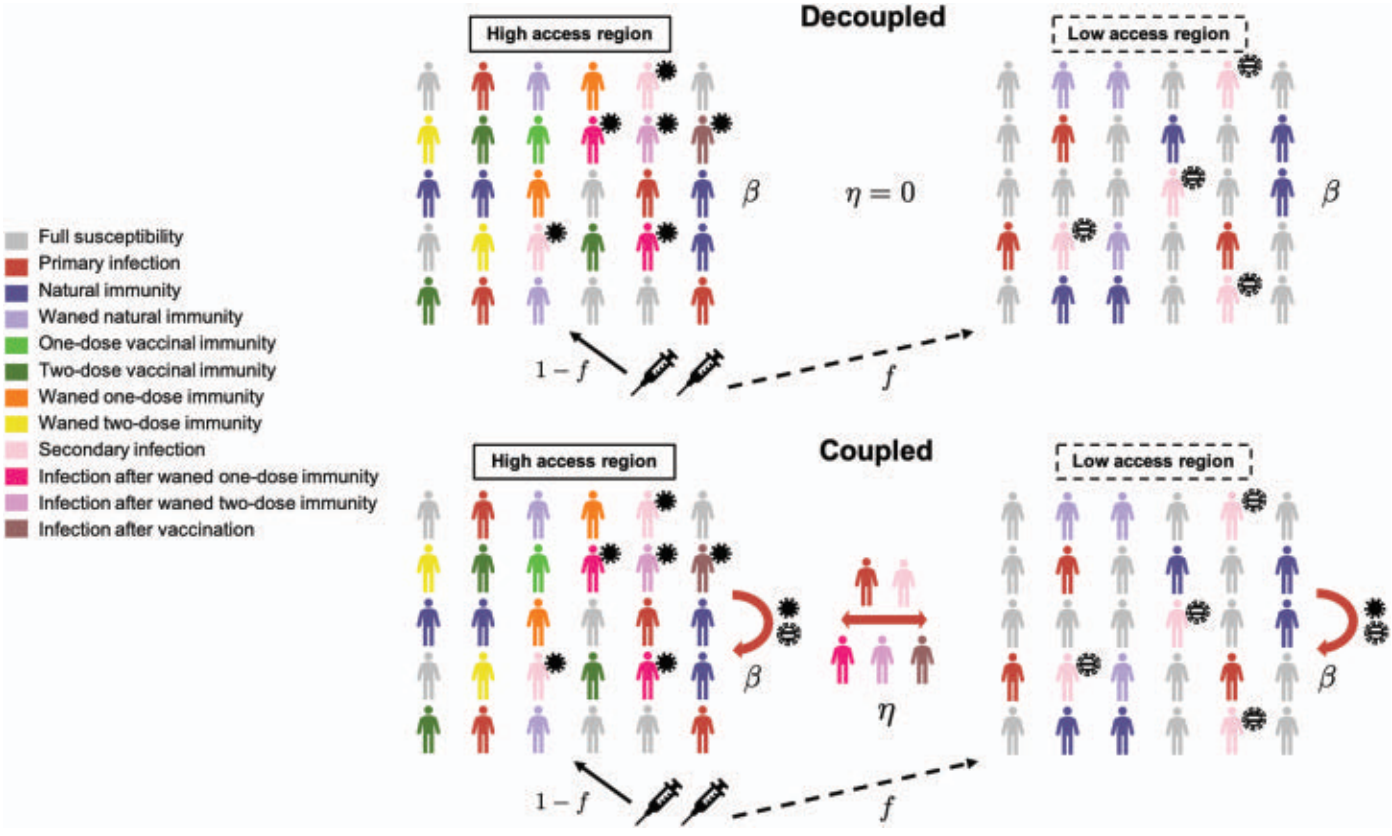


Fig. 1. Schematic depicting the two-country model. The underlying immuno-epidemiological models for each country are based on (*5*, *6*). Vaccines are allocated by the HAR to the LAR. In the coupled framework, immigration of infected individuals between the countries is considered, and the national transmission rate depends on PTIs in both countries, shown schematically as solid and striped virus particles in the HAR and LAR, respectively. In the decoupled framework, no immigration occurs, and the transmission rate is not influenced by PTIs. Full model details are provided in the supplementary materials.

there is a nonzero probability that the transmission rate in both countries increases. This follows evidence of enhanced binding of the SARS-CoV-2 spike protein receptor binding domain with the angiotensin-converting enzyme 2 (ACE2) receptor in more contagious SARS-CoV-2 variants, as well as potentially higher viral loads (*24*). Mutational changes to the spike protein furin cleavage site (e.g., at site 681) may also contribute to increased viral transmissibility (*25*). In this way, this assumption represents a pessimistic scenario where the evolution of pathogen immune escape is inevitably accompanied by increases in transmission, although we also compare our results with the more optimistic scenario where transmission increases do not occur. (Note that we are not modeling the complexities of variant dynamics and evolution explicitly.) In reality, transmission increases may plateau (*26*), and viral evolution may have more subtle effects on disease transmission, including modulating the susceptibility of partially immune hosts. The full mathematical details for both frameworks are described in the supplementary materials.

We begin with the decoupled framework and the simpler underlying "one-dose" vaccination model from (*5*) to compute the long-term equilibrium fraction of infections in both countries under a range of epidemiological and immunological scenarios. Then, with specific dosing regimes (*6*), we examine the short- and medium-term epidemiological dynamics and the global potential for evolution with the sharing of vaccines. Next, using the coupled framework, we compute national and combined case numbers in the medium term given different degrees of vaccine allocation from the HAR to the LAR for different immigration rates and average relative reproduction numbers. We do so for total as well as severe cases, with the expectation that the number of severe cases may be indicative of infections requiring hospitalization or the clinical burden of COVID-19, whereas the number of total cases reflects all infections regardless of severity. Finally, we compare the results of the coupled and decoupled frameworks for specific scenarios.

Decoupled framework

Equilibrium analyses

To obtain analytical intuition for the effect of vaccine nationalism, we first examine the simplest model of vaccine sharing, where both regions are only coupled through their vaccination rates. The dynamics of prophylactic vaccine distribution strategies are well understood when infections lead to recovery and lifelong immune protection (SIR) (*16*). However, natural and vaccinal immunity to SARS-CoV-2 is likely not lifelong, yet complete resusceptibility after the waning of immunity, as is assumed in susceptible-infected-recovered-susceptible (SIRS) frameworks, is also unlikely. Instead, we examine the role of both the strength and duration of immune responses with more general SIR(S) models (*5*, *6*). We first ignore the complexities of dosing regimes and extend the model in (*5*) to consider vaccine sharing in the decoupled framework (top panel of Fig. 1), assuming that a single immune category exists for vaccinated individuals and that vaccinal immunity may wane at a rate distinct from natural immunity. Because we assume that the infection dynamics in both countries are only coupled through their respective vaccination rates, a unique equilibrium of total infections exists [either disease-free or endemic; see supplementary materials and (*5*) for details]. To examine the long-term epidemiological effects of vaccine nationalism, we compute the total fraction of infections at equilibrium as the proportion of vaccines shared between countries is varied. In other words, for a fixed

global vaccination rate ν_{tot} (determined by the maximal rate of administration of the first dose $\nu_{0,tot}$ and the interdose period; see Materials and methods) and for a fraction f of vaccines allocated from the HAR to the LAR, the vaccination rates in the HAR and LAR are $(1-f)\nu_{tot}$ and $f\nu_{tot}$, respectively. We examine four immunity scenarios that range from poor to robust natural and vaccinal immune responses (Fig. 2).

When the characteristics of both countries are the same, sharing vaccines always decreases or maintains the total fraction of infections at the long-term equilibrium [see Keeling and Stattock (*16*) for the SIR extreme with a focus on two differently sized populations]. The intuition for this result is apparent from examining the underlying values for each country (Fig. 2C). The total fraction of infections is minimized whenever one of the countries does not vaccinate beyond the rate needed for herd immunity. Additionally, sharing does not have an appreciable impact on the total fraction of infections at equilibrium when vaccination rates are too low (Fig. 2A, top row) or overall transmission rates are higher and host immune responses are poorer (Fig. 2B, top row). Because of nonpharmaceutical interventions (NPIs) or intrinsic factors [e.g., demographics (*27*), population density (*28*), or vulnerabilities (*29*)], transmission rates in the two countries may be asymmetric. If there is less disease transmission in the HAR (modeled as a reduction in the transmission rate), then the "optimal" fraction of vaccines shared to minimize the combined equilibrium fraction of infections crucially depends on the magnitude of the vaccination rate. If vaccine supplies are low and immune responses poor, then sharing only a very small fraction of the vaccine supply is epidemiologically beneficial in terms of decreasing the overall burden. For stronger immune responses, augmenting vaccine sharing rates becomes increasingly beneficial from an epidemiological perspective, because the protective effects of the vaccine are maintained for longer within the population (compare the columns of the middle row of Fig. 2A). Similarly, as vaccine supplies increase (compare the colored curves in the middle row of Fig. 2A), the minimum value of infections occurs for increasingly large values of f, or fractions of vaccines shared. Eventually, when global vaccination rates are high, even for poor host immune responses, this minimum is attained when more than half of the vaccine supply is allocated to the LAR (first panel of the middle

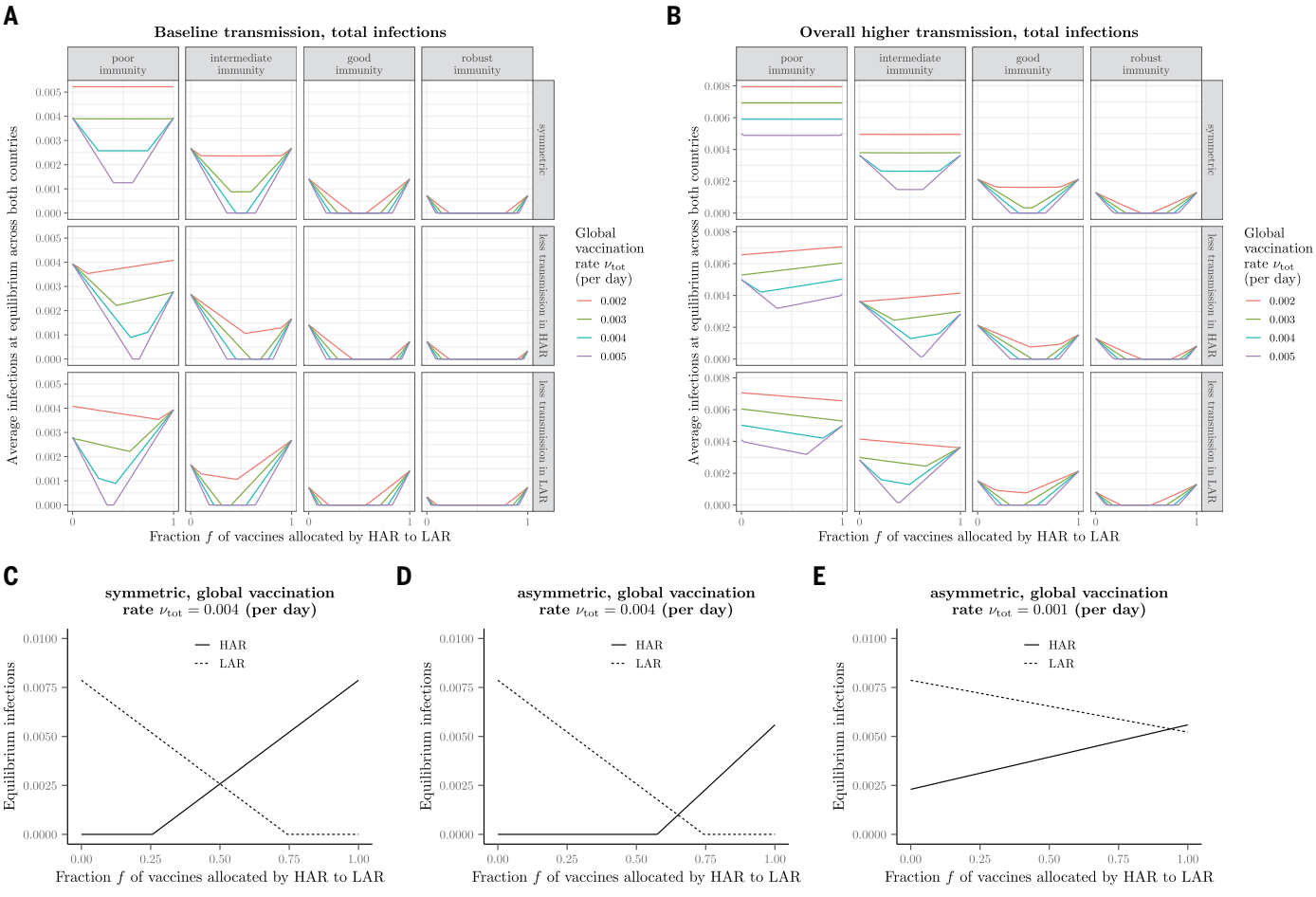


Fig. 2. Long-term equilibrium of the average fraction of infections. (**A** to **E**) Equilibrium infections as a function of the vaccine fraction allocated by the HAR to the LAR under different scenarios related to immunity, transmission, and vaccination rate. In all panels, immunity scenarios are as follows: poor immunity, $\frac{1}{\delta} = \frac{1}{\delta_{vax}} = 0.8$ years, $\epsilon = 0.8$; intermediate immunity, $\frac{1}{\delta} = \frac{1}{\delta_{vax}} = 1$ year, $\epsilon = 0.7$; good immunity, $\frac{1}{\delta} = \frac{1}{\delta_{vax}} = 1.5$ years, $\epsilon = 0.6$; and robust immunity, $\frac{1}{\delta} = \frac{1}{\delta_{vax}} = 2$ years, $\epsilon = 0.5$, where $\frac{1}{\delta}$ is the average duration of natural immunity, $\frac{1}{\delta_{vax}}$ is the average duration of vaccinal immunity, and epsilon is the relative susceptibility to infection after the waning of natural or vaccinal immunity. In the scenario with asymmetrical transmission rates between the two countries, the transmission rate in the country with lower transmission is taken to be 80% of the value in the symmetric case. In the scenarios with overall higher transmission rates (B), this same asymmetric assumption is made in addition to the baseline symmetric transmission rate being increased by 30% relative to the value in (A). In (C) to (E), illustrations of the equilibrium fraction of infections in each country with the intermediate immunity scenario are shown for (C) symmetric transmission with $\nu_{tot} = 0.004$, (D) asymmetric transmission (lower in HAR) with $\nu_{tot} = 0.004$, and (E) asymmetric transmission (lower in HAR) with $\nu_{tot} = 0.001$, with all other parameters as in (A). In all panels, the baseline transmission rate is $\beta = \frac{2.3}{5}$.

row of Fig. 2A). By symmetry, the opposite occurs if there is less transmission in the LAR. These trends are further magnified if overall transmission rates are increased (Fig. 2B). To further emphasize these effects, we present the long-term equilibrium of each country under representative scenarios in Fig. 2, C to E. In particular, a comparison of Fig. 2D and Fig. 2E illustrates the importance, and indirect benefit, of increasing vaccine supply. The relative sizes of the HAR and LAR populations can have important consequences for the fraction of vaccine allocation that minimizes the weighted fraction of infections (figs. S1 and S2). Overall, these results highlight the importance of continued NPIs that decrease transmission, such as rapid testing and physical distancing, in conjunction with ramping up vaccination and sharing vaccine supplies equitably to decrease overall burden.

Medium-term dynamics

Epidemiological considerations

To consider the near- and medium-term dynamic epidemiological effects of vaccine sharing, in Fig. 3 we explore the landscapes of immunity and infections across multiple scenarios for otherwise-symmetric countries (i.e., population size and seasonal transmission rates). In all scenarios, vaccine supply is assumed to be limited initially in the HAR [modeled as a one-dose policy, a lower maximal rate of administration of the first dose $\nu_{0,\mathrm{tot}}$, and no sharing ($f = 0$)] and then is assumed to increase. In conjunction with an increase in $\nu_{0,\mathrm{tot}}$, we allow for a transition to the recommended two-dose strategy ($\frac{1}{\omega} = 4$ weeks, third and fourth columns of Fig. 3, where $\frac{1}{\omega}$ is the average time between the first and second vaccine doses) and/or the initiation of equal sharing ($f = 0.5$) with the LAR (which is assumed to distribute vaccines using the same strategy as the HAR) (second and fourth columns of Fig. 3).

Intuitively, if one-dose immunity is robust (bottom panel of Fig. 3), then transitioning to a two-dose strategy leads to fewer individuals with robust vaccinal immunity, in turn giving rise to increases in infections in the short and medium term [compare the corresponding scenarios of the bottom panel of Fig. 3 and see also (*6*)]. In such a situation, "one-dose" strategies (i.e., either the first dose of a two-dose vaccine or the single dose of a one-dose vaccine) with equal sharing between countries suppress overall burden. Conversely, if one-dose immunity is poor, switching to the recommended two-dose regimen prevents the accumulation of individuals with waned one-dose immunity and thus potentially larger infection peaks in the longer term (top panel of Fig. 3). If poor one-dose immunity nevertheless reduces severity of infection after waning (unlike our pessimistic assumption), then the predicted clinical burden of severe cases would likely be lower. Finally, if one-dose immunity is poor and a one-dose policy is pursued, the first infection peak after ramping up vaccination in the HAR may be higher without sharing (top-left plot in the top panel of Fig. 3). This counter-intuitive finding arises because of the large accumulation of individuals with waned one-dose immunity who experience infection. This highlights the important role for population-level susceptibility (modulated by natural and vaccinal immune responses) and its dynamical interplay with transmission in determining the timing and burden of infections. The effects of different NPI scenarios, transmission patterns, and vaccination rates in either the HAR or LAR can be further explored with the online application (*30*).

Evolutionary considerations

The accumulation of individuals with various states of immunity (i.e., waned one-dose immunity or immunity after natural infection) may also lead to different evolutionary outcomes depending on the vaccine-sharing scheme pursued. Current evidence suggests that adaptive immune responses after natural infection with SARS-CoV-2 are fairly robust and long-lasting (*31*, *32*). This protection may be less certain in the context of subsequent infection with variant strains (*33*, *34*). Encouragingly, studies indicate that previously infected hosts are also largely protected (clinically and against breakthrough infections) against emerging variants after a single vaccine dose (*35*). However, recent work indicates poorer protection against the rapidly spreading Delta variant of concern, particularly after a single dose (*36*), although protection against severe disease still appears fairly robust (*37*–*39*). Interestingly, the protection conferred by a single vaccine dose in previously infected individuals may be even greater than that provided by two vaccine doses in naïve hosts (*40*). However, there are still many immunological uncertainties, for example, the duration and longer-term strength of this protection against existing strains and potential emerging variants remain unknown.

In fig. S3, we use the evolutionary framework from (*6*) to project the potential net viral adaptation rate [see supplementary materials for details and the interactive online application (*30*) for additional scenarios]. Overall, we find that uncertainties in evolutionary outcomes dominate our projections, echoing previous findings (*6*) [fig. S3 and the interactive online application (*30*)]. However, if the evolutionary potential for immune escape is highest among infections in hosts with natural immunity, the framework predicts that sharing vaccines always decreases global evolutionary potential (green curves of fig. S3) in the decoupled framework. Overall, when immunity after a single vaccine dose is robust, natural and vaccine-derived immunity will limit damaging pathogen evolution relative to the scenario with poor single-dose immunity (compare the top and bottom panels of fig. S3). However, if immunity is partial or waning, ongoing transmission might accelerate adaptation, supporting the need for continued monitoring of variants and their interaction with natural and vaccine-derived immunity.

Another intuitive result is that, in the LAR, sharing vaccines leads to increases in population immunity (for any dosing regime) and thus a decrease in infections and burden in the short term, even with poor one-dose immunity. In general, in the HAR, sharing decreases population immunity and increases infections in the short term. However, these changes are minimal and likely acceptable given the combined decrease in infections, illustrating the long-term benefits of vaccine sharing. In particular, although local cases in the HAR may increase in the short term, the longer-term disease risk would be lowered by sharing because of a decrease in the potential for the evolution of more transmissible or immune-escape viral variants.

Coupled framework

So far, we have assumed that the countries have decoupled disease dynamics. This simplification for tractability ignores infection importation as well as the possible emergence of variants from regions with more persistent infections. The issues that arise from the global circulation of SARS-CoV-2, particularly the variants of concern, are of considerable public health importance. Thus, we next explore these effects using the coupled framework presented in the bottom panel of Fig. 1.

Cumulative cases and pathogen evolution in the medium term

In Fig. 4, we plot the cumulative number of total and severe cases (see supplementary materials for details), assuming equal population sizes in both countries from the time of vaccine introduction until 5 years after the pandemic onset in the HAR, LAR, and combined, as well as the projected number of PTIs to have occurred in both regions by the end of the 5-year period. We do so for various vaccine allocation fractions between the HAR and LAR, as well as for a range of immigration rates, assuming symmetric transmission rates (Fig. 4, A and C), and relative mean reproduction numbers, assuming constant immigration rates (Fig. 4, B and D; see Materials and methods). In Fig. 4, A and B, we assume that infection after waned natural immunity contributes the most to viral adaptation, whereas in Fig. 4, C and D, we assume that infection after waned vaccinal immunity, and one-dose immunity in particular, contributes the most.

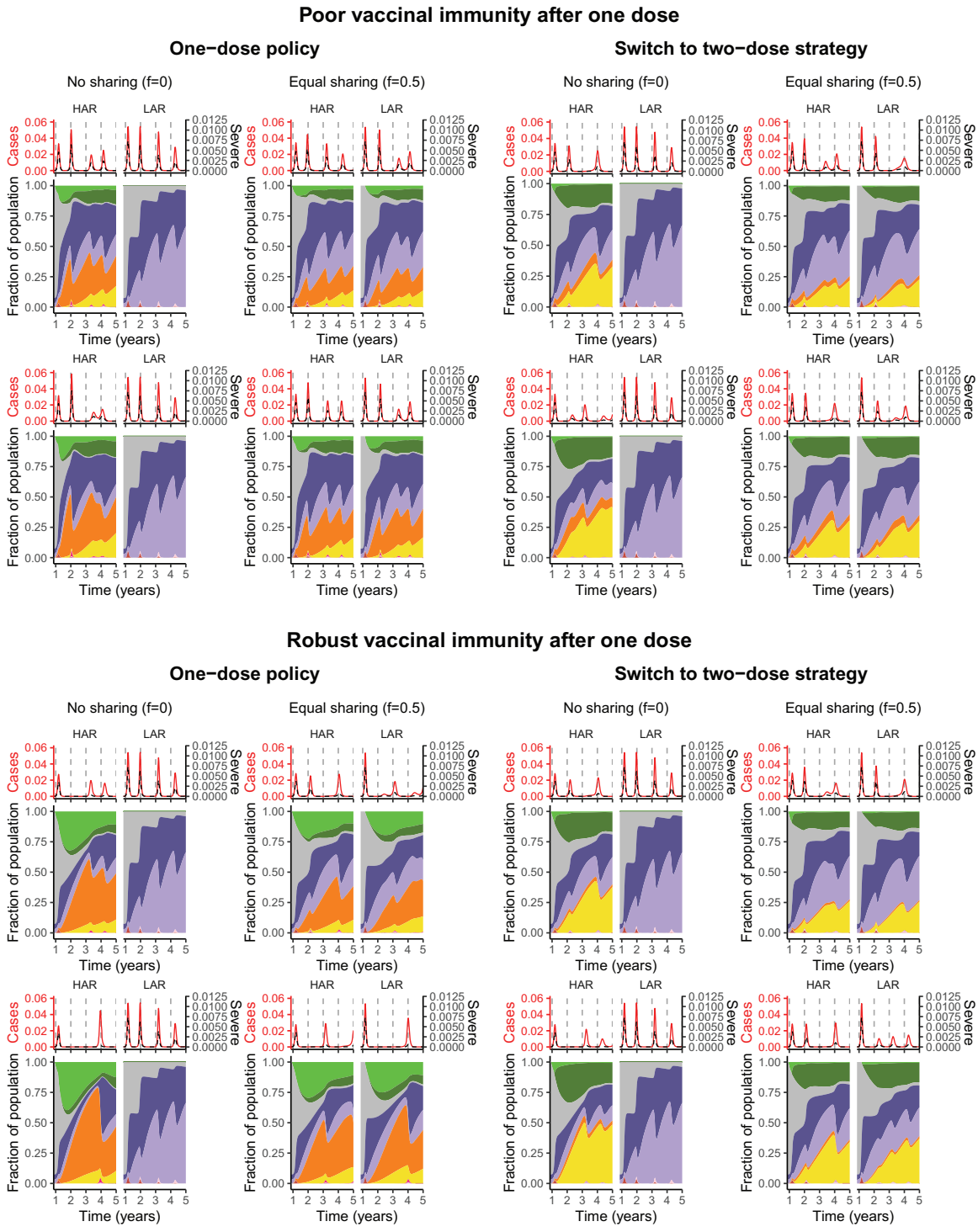


Fig. 3. Immune landscapes and infections in both countries under a range of vaccination strategies and assumptions related to the robustness of immune responses. Note that the color scheme is as in Fig. 1. In all panels, vaccination begins after week 48. Poor vaccinal immunity after one dose is represented by $\frac{1}{\rho_1} = 0.25$ year and $\epsilon_1 = 0.9$, where $\frac{1}{\rho_i}$ and is the average duration of vaccinal immunity following dose *i*, whereas robust vaccinal immunity after one dose means $\frac{1}{\rho_1} = 1$ year and $\epsilon_1 = 0.7$. Other immunity parameter values are as follows: $\frac{1}{\delta} = 1$ year, $\frac{1}{\rho_2} = 1$ year, $\epsilon = \epsilon_2 = 0.7$, $\epsilon_{V_1} = 0.1$, and $\epsilon_{V_2} = 0.05$. All other parameters, including the procedure for the calculation of severe cases, are described in the supplementary materials. In both the top and bottom panels, the top row depicts a switch from a maximum first-dose administration rate of 1 to 3% after week 60, whereas it is 1 to 5% for the bottom row (and concurrent with sharing, if it occurs).

Dependence on immigration rate

For equal population sizes and symmetric transmission rates, a weak dependence of total and severe cases as well as PTIs on the immigration rate η is observed, regardless of whether infection after waned natural or vaccinal immunity is assumed to contribute more to viral evolution (Fig. 4, A and C). Additionally, more equitable vaccine distribution minimizes

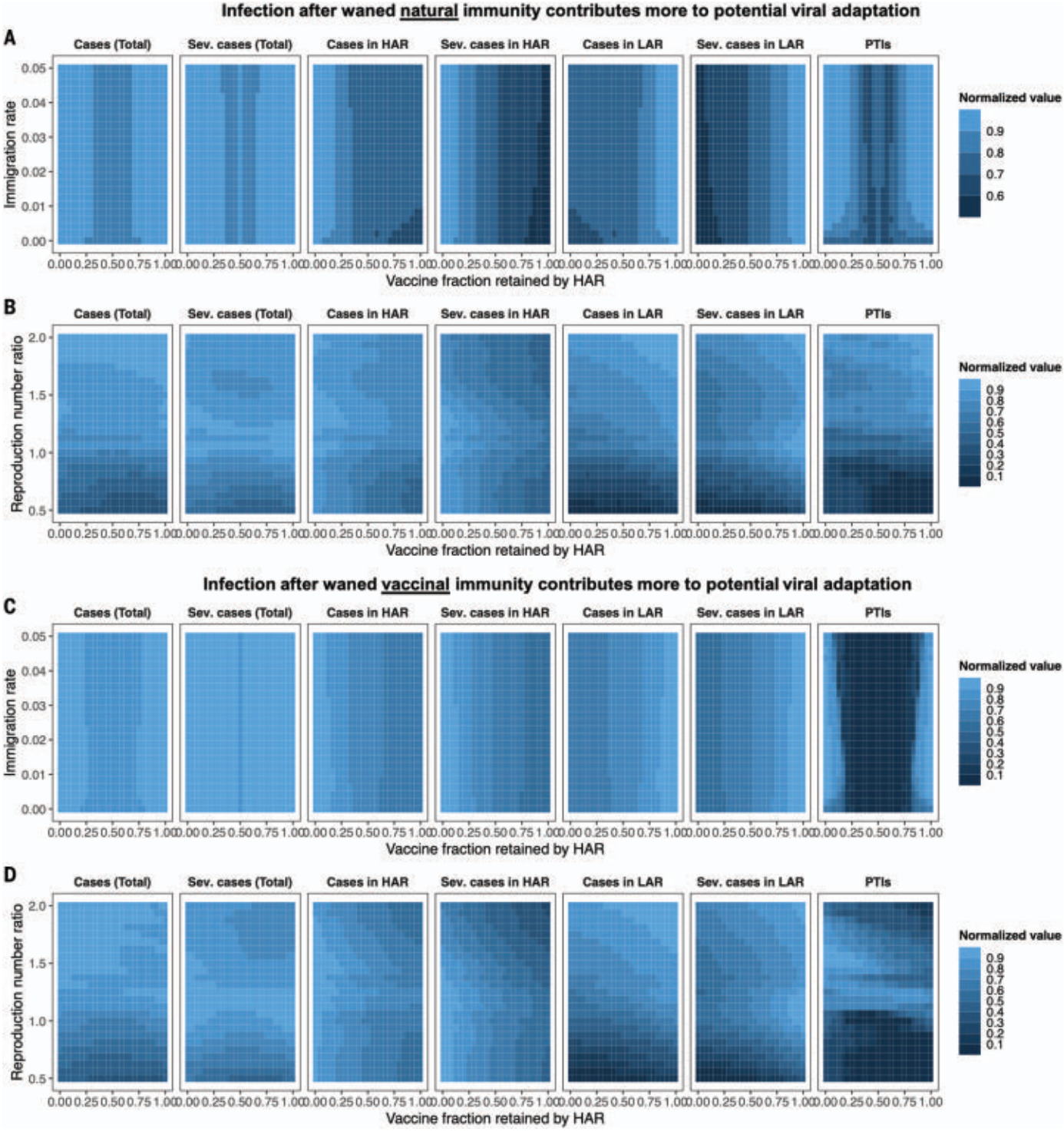


Fig. 4. Cumulative number of cases and PTIs in the medium term. (**A** to **D**) Heat maps depicting total and severe cases from the time of vaccine onset ($t_{vax} = 48$ weeks) through the end of the 5-year period for both countries (first and second columns from the left), the HAR (third and fourth columns from the left), and the LAR (fifth and sixth columns from the left), as well as the combined number of PTIs to have occurred in both countries at the end of 5 years (last column). Each grid point denotes the mean value of 100 simulations. The population of both countries is taken to be the same. Each area plot is internally normalized, such that the largest value in each plot is 1. The *x* axis indicates the fraction of vaccines retained by the HAR (i.e., $1 - f$); thus, the far right of a plot is the scenario where the HAR retains all vaccines ($f = 0$). In (A) and (C), both countries have the same average transmission rate ($\bar{R}_0$, see Materials and methods), and the immigration rate η is varied. In (B) and (D), the immigration rate is fixed at $\eta = 0.01$, and the relative mean transmission rate in the LAR, that is, $\bar{R}_{0,LAR}/\bar{R}_{0,HAR}$, is varied between 0.5 and 2. The seasonality of the transmission rates in both countries and periods of NPI adoption are identical and as described in the Materials and methods. In all simulations, we assume a two-dose strategy throughout, that is, $\frac{1}{\omega} = 4$ weeks, and take the maximal rate of administration of the first dose to be $\nu_{0,tot} = 2\%$. Assumed immunological parameters are $\frac{1}{\delta} = 1$ year, $\epsilon = 0.7$, $\epsilon_{V_1} = 0.1$, $\epsilon_{V_2} = 0.05$, $\epsilon_2 = 0.7$, and $\frac{1}{\rho_2} = 1$ year, and the one- to two-dose immune

response ratio is $x_e = 0.8$ (see Materials and methods). In (A) and (B), we assume that infection after waned natural immunity contributes more to potential viral adaptation and take $w_{IS} = 0.8$, $w_{IS1} = 0.2/x_e$, and $w_{IS2} = 0.2$, where w_i is the relative weight for evolutionary potential for each infection type i (see Materials and methods). In (C) and (D), we assume that infection after waned vaccinal immunity contributes more to potential viral adaptation and take $w_{IS} = 0.4$, $w_{IS1} = 0.8$, and $w_{IS2} = 0.8 \times x_e$ (see Materials and methods). Additional details related to the determination of severe cases are also provided in the supplementary materials.

total PTIs and combined cases in both scenarios. When natural infections contribute more to evolution (Fig. 4A) and η is low, the HAR must retain an increasing share of the vaccines to minimize local total cases as the immigration rate increases, but this is done at the expense of more cases in the LAR and PTIs. Notably, for $f \approx 0$, severe cases are minimized regardless of the immigration rate in the HAR and maximized in the LAR, which may have important clinical consequences. When infection after waned vaccinal immunity contributes more to viral evolution, this approach is no longer advantageous for the HAR, and the retained vaccine fraction sets the observed case numbers nearly independently of the assumed immigration rate. Further, large asymmetries in vaccine sharing (i.e., $f = 0$ or 1) result in much more marked relative numbers of PTIs in this scenario.

For the same total vaccine availability, the most realistic population asymmetry is that the LAR has a larger population. Under this condition (fig. S4) and when infection after waned natural immunity contributes more to evolution, total cases in the LAR are relatively insensitive to the amount of vaccine allocated, except for very large f; however, severe cases can be substantially reduced with vaccine sharing. Here, combined total cases and PTIs are minimized by minimizing cases in the HAR. When a greater number of total vaccines are available (i.e., a larger $\nu_{0,tot}$; fig. S5), total and severe cases in the more populous LAR decrease approximately monotonically with increasing vaccine allocation f, and more equitable vaccine allocation once again minimizes combined total and severe cases. When infection after waned vaccinal immunity contributes more to evolution, the trends are more similar to those for symmetric population sizes, and more equitable vaccine sharing is favored for both vaccination rates $\nu_{0,tot}$, given a larger population in the LAR (figs. S4 and S5).

When the LAR has a smaller population (fig. S6), a relatively weak dependence of total and severe case numbers and PTIs on η is still observed for both evolutionary scenarios, particularly for higher immigration rates. However, and particularly when infection after waned natural immunity contributes more to evolution, the minima in combined cases and PTIs are now observed for $f < 0.5$, that is, when the HAR retains more than half of the available vaccines. Importantly, we note that additional booster doses may further change these landscapes of immunity and, consequently, the projected burdens of total and severe cases.

Dependence on the relative transmission rate

The number of total and severe cases and PTIs show a greater sensitivity to the average reproduction number ratio $\bar{R}_{0,LAR}/\bar{R}_{0,HAR}$ between the two countries for a fixed immigration rate. Intuitively, for equal population sizes (Fig. 4) and when $\bar{R}_0$ in the LAR country is lower, having the HAR retain more than half of the vaccines ($f < 0.5$) is a good strategy for minimizing total PTIs and cases, regardless of the evolutionary scenario. Indeed, along with an increase in cases, the optimal vaccine allocation shifts closer and closer to equal sharing as the $\bar{R}_0$ values of both countries approach each other. These trends are similar when the LAR has a larger or smaller population than the HAR (figs. S4 and S6, respectively).

When $\bar{R}_0$ in the LAR is higher, trends are more complex. In general, regardless of the relative population size or evolutionary scenario, for a given $\bar{R}_{0,LAR}/\bar{R}_{0,HAR}$, cases in the HAR decrease with increasing vaccine retention (smaller f), whereas cases in the LAR increase (Fig. 4 and figs. S4 to S6). Severe cases in each region in particular are strongly reduced by increased vaccine availability. The increase in cases in the LAR is increasingly large at higher $\bar{R}_{0,LAR}/\bar{R}_{0,HAR}$. When infection after waned natural immunity contributes more to evolution and vaccine supply is sufficiently high to reduce case numbers in the LAR, PTIs are numerous when the HAR retains a large fraction of the vaccines (Fig. 4B and figs. S5B and S6B). This is due to sustained increased case numbers in unvaccinated individuals in the LAR. Conversely, when infection after waned vaccinal immunity contributes more to evolution, then having the HAR (with lower transmission rates) retain a larger fraction of the vaccines minimizes PTIs for any relative population size, because the high $\bar{R}_0$ in the LAR would result in large subsequent peaks containing individuals whose vaccinal immunity has waned with sharing (Fig. 4D and figs. S4D, S5D, and S6D). However, this strategy also leads to highly increased case numbers, including severe cases, in the LAR. Because LARs may also have more fragile health care systems, the increased clinical burden under this scenario may be particularly problematic.

Comparison of coupled and decoupled frameworks

We note that assumptions of large $\bar{R}_0$ in the LAR also result in very large initial infection peaks, which increase community immunity in the medium term. These initial waves are not reflected in the total case counts, however, because these values are summations from the time of vaccine initiation through the end of the 5-year period after the onset of the pandemic. Further, in Fig. 5 we illustrate the temporal effect of the coupled framework on the infection dynamics in the LAR and HAR relative to the first model with no immigration or explicit effect of PTIs on the transmission rate. When moderate asymmetry in $\bar{R}_0$ is assumed ($\bar{R}_{0,LAR}/\bar{R}_{0,HAR} = 1.2$), simulations using the decoupled framework suggest that a strategy where the HAR retains all vaccines would be highly beneficial for that country (Fig. 5, A to F; no PTIs are projected to occur, and low case numbers are observed throughout). However, this occurs at the expense of PTIs and infection burden in the LAR, which are both substantially higher. With the more realistic coupled framework, immigration and increases in transmission illustrate that this strategy is far less beneficial to the HAR than the decoupled framework would suggest, because substantially higher case numbers and pathogen evolution are predicted in this region. Although total cases in the HAR increase slightly when vaccines are equally distributed under the coupled framework (Fig. 5, G to L), substantial reductions in case numbers in the LAR result in fewer PTIs in that country, and total combined case numbers are also slightly lower. To untangle the effects of immigration and PTIs on dynamics in the coupled framework, we reproduce Figs. 4 and 5 allowing for immigration only in figs. S7 and S8, respectively. In other words, these figures represent a more optimistic evolutionary scenario in which the occurrence of a PTI does not increase transmission rates in either the HAR or LAR. Overall, we show that vaccines play an important role in minimizing cases (particularly severe cases) as well as potential viral adaptation in both regions. We also emphasize that imperfect vaccinal and natural immunity and asymmetries in population size and transmission rates add many nuances to this picture.

Caveats

A full list of caveats and future directions is presented in the supplementary materials; we briefly summarize them here. First, building on prior work and to focus on qualitative features, we ignore heterogeneities within countries, such as those due to age (*41*) or superspreading (*42*). We have also assumed simple scenarios for NPIs as in (*6*) and assumed that the seasonal transmission rates are similar in both countries. More granular, well-parameterized epidemiological models with these complexities

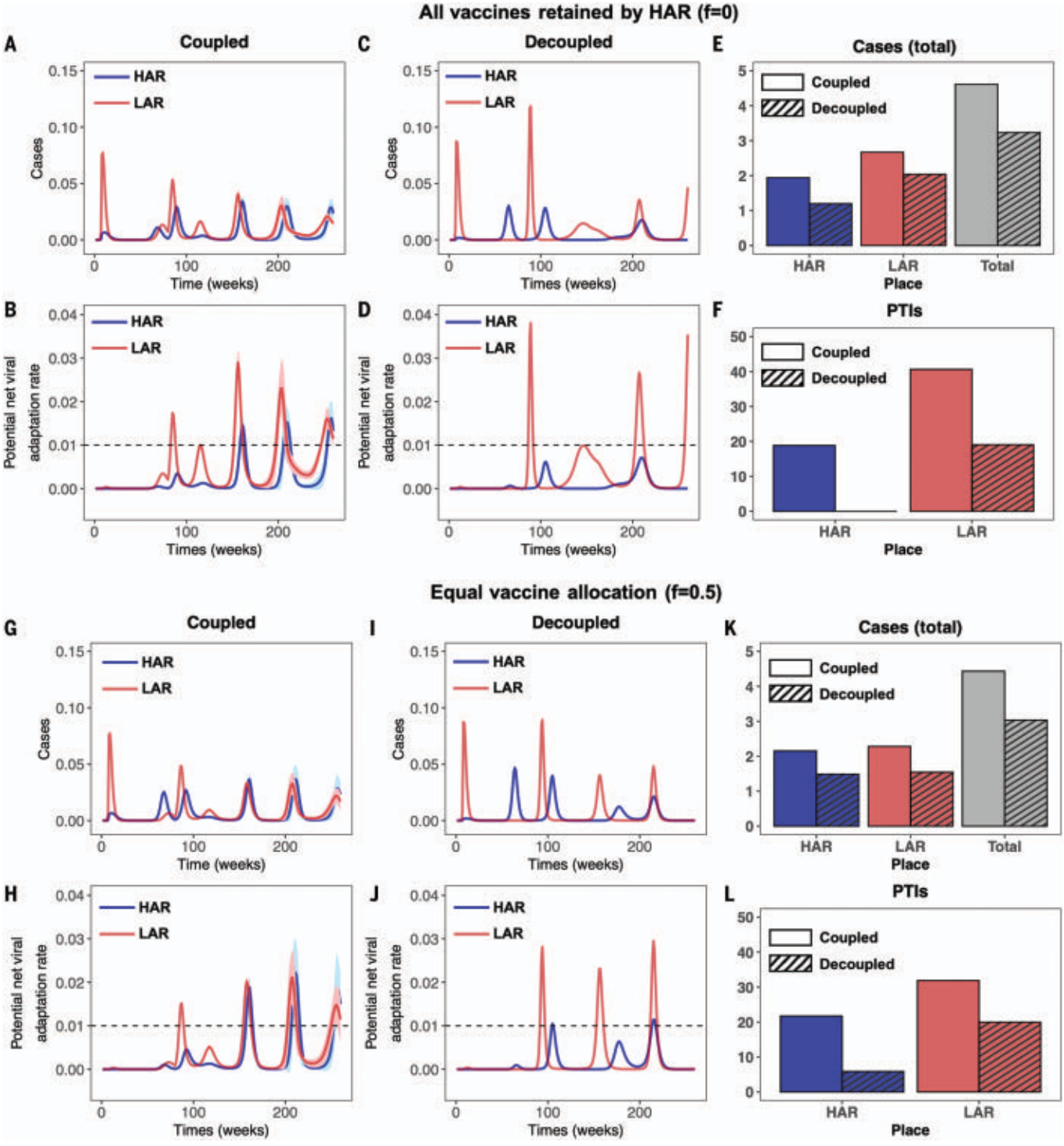


Fig. 5. Time series of cases and potential net viral adaptation rates. (**A** to **L**) Infections in the HAR (blue) and LAR (red) for the first 5 years after pandemic onset for the coupled [(A) and (G)] and decoupled [(C) and (I)] frameworks. Each simulation is run 100 times, with the average indicated by the solid line and the standard deviation shown with the corresponding ribbon. The average number of cumulative cases over all simulations from the time of vaccine onset $t_{vax} = 48$ weeks through the end of the 5-year period are shown in (E) for the HAR, LAR, and both countries combined (gray) for the coupled (solid) and decoupled (dashed) frameworks. Time series of the potential viral adaptation rate in both regions for the coupled [(B) and (H)] and decoupled [(D) and (J)] frameworks. The colors, averages, and standard deviations are as described above. The dashed horizontal line denotes $e_{cutoff} = 0.01$, the assumed threshold for the occurrence of a PTI (see Materials and methods). The average number of PTIs at the end of the 5-year period are shown in (F) and (L) for the HAR and LAR for the coupled (solid) and decoupled (dashed) frameworks. (A) to (F) correspond to the HAR retaining all vaccines ($f = 0$), whereas (G) to (L) correspond to equal vaccine sharing ($f = 0.5$). In all simulations, we take $\bar{R}_{0,LAR}/\bar{R}_{0,HAR} = 1.2$ and $\eta = 0.01$ and assume that infection after waned natural immunity contributes primarily to evolution (i.e., $w_{IS} = 0.8$, $w_{IS1} = 0.2/x_e$, and $w_{IS2} = 0.2$). All other parameters are identical to those in Fig. 4.

would lead to more accurate quantitative predictions. Additionally, underlying differences in susceptibility between the HAR and LAR, through factors such as variations in age structure or degrees of preexisting population-level immunity, may result in differential rates of severe cases and hospitalizations independent of vaccination. An explicit treatment of age structure and serostatus [as in (*21*)] in modeling vaccine rollout to estimate clinical burdens in HARs and LARs under different sharing schemes should be investigated further. We also note that our models focused on the situation where dose administration is contingent on supply, so that "sharing" vaccines from the HAR to the LAR decreases the vaccination rate in the HAR. However, certain HARs may have access to more doses than they can administer. In such a scenario, sharing would be even more beneficial: It would lead to no change in the HAR vaccination rate while increasing that of the LAR.

Moving forward, additional booster doses may be administered that could alter population-level immune landscapes, and including these in future models will be important for qualitative and quantitative predictions. Additionally, we omit vaccine hesitancy (*23*), though simple extensions of our previous models with hesitancy (*5*, *6*) could examine the resulting interplay with vaccine nationalism. Furthermore, our equilibrium analyses highlight the long-term importance of vaccine sharing regardless of when sharing is initiated. However, our scenarios that examine short- and medium-term implications strongly emphasize the benefits of close to "parallel" or concurrent vaccine sharing. The effects of delaying the initiation of sharing, for instance, until a certain proportion of individuals in the HAR are vaccinated, should be carefully examined.

Lastly, we have assumed the simplest evolutionary models, both for determining potential viral adaptation rates as well as for simulating potential increases in transmission rates. As more data become available, these should be refined (*43*–*46*), with possible directions including extending the model to explicitly track the transmission of different strains, and accounting for potential reductions in the strength of vaccinal immunity (i.e., the parameters $\epsilon_{V_1}, \epsilon_{V_2}, \epsilon_1$, and ϵ_2, where ϵ_{V_i} is the relative susceptibility to infection while in the i-dose vaccinated class and ϵ_i is the relative susceptibility to infection after the waning of i-dose immunity) due to the emergence of novel immune-escape variants (*47*). In particular, prolonged infections in immunosuppressed hosts may be disproportionately important for evolution (*48*). The online interactive application (*30*) allows for an in-depth exploration of the effect of different climate-driven seasonal transmission rates as well as a broad range of assumptions related to NPIs and immuno-epidemiological parameters.

Conclusion

Even as vaccine production increases, a number of countries are choosing to share little or no vaccines with countries that have very low vaccine availability. Vaccine nationalism, dosing regimes, and host immune responses have important interactive effects, and these will substantially shape epidemiological dynamics and evolutionary potential in the medium term. Additionally, unstable vaccine supply will also increase variability in the timing or availability of first and second doses.

Using extensions of our prior work (*5*, *6*), we incorporated vaccine sharing scenarios in two countries whose infection dynamics are either otherwise independent or coupled through immigration of infectious individuals and evolution-driven increases in transmission rates. When country profiles are symmetric, we find that sharing vaccines with countries that have low availability decreases overall infections and may also mitigate potential antigenic evolution. Asymmetries in population size or transmission rates introduce additional complexities, which are particularly marked when natural and vaccinal immunity is weak. Nevertheless, our models indicate that the prompt redistribution of vaccine surpluses is likely advantageous in terms of epidemiological and evolutionary outcomes in both countries and, by extension, globally. Ethical arguments also support this policy (*13*, *14*). Persistent increased disease transmission in countries with low vaccine availability also substantially undermines attempts at infection control through stockpiling in the country with high vaccine availability, which is not accounted for when disease transmission in both countries is assumed to be decoupled. Overall, our work highlights the importance of continued efforts in quantifying the robustness of immunity after vaccination. Furthermore, reevaluation of stockpiling policies as vaccine supplies increase is imperative, and ramping up global vaccination efforts is crucial.

Materials and methods

Determination of seasonal reproduction numbers

To reflect observed seasonal variation in transmission rates for respiratory infections arising from related coronaviruses (*49*), influenza (*49*), and respiratory syncytial virus (*50*), we base seasonal reproduction numbers in this work on those in (*5*), which were calculated in (*49*) based on the climate of New York City. Other seasonal patterns can be explored using the interactive online application (*30*). In all simulations, we modify these values to force a mean value for the basic reproduction number of $\bar{R}_0 = \langle R_0(t) \rangle = 2.3$, where t denotes each week in a year, by multiplying the climate-derived time series $R_{0,c}(t)$ by 2.3 and dividing by its average value, that is

$$R_0(t) = R_{0,c}(t)\frac{2.3}{\bar{R}_{0,c}}$$

When transmission rates are assumed to be asymmetrical, the mean value of the $R_0(t)$ time series is adjusted by the desired relative ratio. For example, for $\bar{R}_{0,\mathrm{LAR}}/\bar{R}_{0,\mathrm{HAR}} = 1.2$, the reproduction number time series in the HAR is as above, but in the LAR it is

$$\begin{aligned} R_{0,\mathrm{LAR}}(t) &= R_{0,c}(t)\frac{2.3}{\bar{R}_{0,c}}\frac{\bar{R}_{0,\mathrm{LAR}}}{\bar{R}_{0,\mathrm{HAR}}} \\ &= R_{0,c}(t)\frac{2.3}{\bar{R}_{0,c}} \times 1.2 \end{aligned}$$

Modeling of NPIs

In all simulations, we enforce periods of NPI adoption (arising from behaviors and policies such as lockdowns, mask wearing, and social distancing) in which the transmission rate is reduced from its seasonal value described in the previous section.

For the simulations of medium-term dynamics using the decoupled framework (i.e., Fig. 3 and fig. S3), we used the same NPI scenarios as in (*6*). For all simulations using the coupled framework (i.e., Figs. 4 and 5 and figs. S4 to S8), we assume that NPIs are adopted between weeks 8 and 44 after the pandemic onset, resulting in the transmission rate being reduced to 40% of its seasonal value. Between weeks 45 and 79, we assume that the transmission rate is 60% of its seasonal value—higher than during the previous period owing to either behavioral changes after the introduction of the vaccine or the emergence of more transmissible strains. Finally, we assume that NPIs are completely relaxed beyond week 80.

Simple vaccination model

The complete mathematical details relevant to the following components of the simple vaccination model are presented in the supplementary materials:

Model equations
Equilibrium calculations

Model with explicit dosing regimes

The complete mathematical details relevant to the following components of the model with explicit dosing regimes are presented in the supplementary materials:

Model equations
Calculation of cumulative case numbers
Linking vaccination rate to interdose period
Determination of potential net viral adaptation rate
Specification of one-dose vaccine efficacy from one- to two-dose immune response ratio
Calculation of the number of severe cases
Details for coupling between the countries

REFERENCES AND NOTES

1. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020). doi: 10.1016/S1473-3099(20)30120-1; pmid: 32087114
2. F. P. Polack *et al*., Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. *N. Engl. J. Med.* **383**, 2603–2615 (2020). doi: 10.1056/NEJMoa2034577; pmid: 33301246
3. M. Voysey *et al*., Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: An interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. *Lancet* **397**, 99–111 (2021). doi: 10.1016/S0140-6736(20)32661-1; pmid: 33306989
4. L. R. Baden *et al*., Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *N. Engl. J. Med.* **384**, 403–416 (2021). doi: 10.1056/NEJMoa2035389; pmid: 33378609
5. C. M. Saad-Roy *et al*., Immune life history, vaccination, and the dynamics of SARS-CoV-2 over the next 5 years. *Science* **370**, 811–818 (2020). doi: 10.1126/science.abd7343; pmid: 32958581
6. C. M. Saad-Roy *et al*., Epidemiological and evolutionary considerations of SARS-CoV-2 vaccine dosing regimes. *Science* **372**, 363–370 (2021). doi: 10.1126/science.abg8663; pmid: 33688062
7. C. M. Saad-Roy, S. A. Levin, C. J. E. Metcalf, B. T. Grenfell, Trajectory of individual immunity and vaccination required for SARS-CoV-2 community immunity: A conceptual investigation. *J. R. Soc. Interface* **18**, 20200683 (2021). doi: 10.1098/rsif.2020.0683; pmid: 33530857
8. Public Health England, "SARS-CoV-2 variants of concern and variants in England: Technical briefing 15" (Public Health England, 2021); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993879/Variants_ of_Concern_VOC_Technical_Briefing_15.pdf?fbclid= IwAR1yXHaljjsar1h6KQEAh66pFE4zAU_qID7BewdBw823FPtrSjMeSbQp-lE.
9. S. P. Otto *et al*., The origins and potential future of SARS-CoV-2 variants of concern in the evolving COVID-19 pandemic. *Curr. Biol.* **31**, R918–R929 (2021). doi: 10.1016/j.cub.2021.06.049; pmid: 34314723
10. D. P. Fidler, Vaccine nationalism's politics. *Science* **369**, 749–749 (2020). doi: 10.1126/science.abe2275; pmid: 32792371
11. J. Holder, Tracking coronavirus vaccinations around the world; www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html.
12. L. Eaton, Covid-19: WHO warns against "vaccine nationalism" or face further virus mutations. *BMJ* **372**, n292 (2021). pmid: 33526414
13. E. J. Emanuel *et al*., An ethical framework for global vaccine allocation. *Science* **369**, 1309–1312 (2020). doi: 10.1126/science.abe2803; pmid: 32883884
14. E. J. Emanuel, F. Luna, G. O. Schaefer, K. C. Tan, J. Wolff, Enhancing the WHO's proposed framework for distributing COVID-19 vaccines among countries. *Am. J. Public Health* **111**, 371–373 (2021). doi: 10.2105/AJPH.2020.306098; pmid: 33566663
15. L. M. Herzog, O. F. Norheim, E. J. Emanuel, M. S. McCoy, Covax must go beyond proportional allocation of covid vaccines to ensure fair and equitable access. *BMJ* **372**, m4853 (2021). pmid: 33402340
16. M. J. Keeling, A. Shattock, Optimal but unequitable prophylactic distribution of vaccine. *Epidemics* **4**, 78–85 (2012). doi: 10.1016/j.epidem.2012.03.001; pmid: 22664066
17. H. Mamani, S. E. Chick, D. Simchi-Levi, A game-theoretic model of international influenza vaccination coordination. *Manage. Sci.* **59**, 1650–1670 (2013). doi: 10.1287/mnsc.1120.1661
18. M. Salathé, J. H. Jones, Dynamics and control of diseases in networks with community structure. *PLOS Comput. Biol.* **6**, e1000736 (2010). doi: 10.1371/journal.pcbi.1000736; pmid: 20386735
19. P. Klepac, R. Laxminarayan, B. T. Grenfell, Synthesizing epidemiological and economic optima for control of immunizing infections. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14366–14370 (2011). doi: 10.1073/pnas.1101694108; pmid: 21825129
20. P. Klepac, I. Megiddo, B. T. Grenfell, R. Laxminarayan, Self-enforcing regional vaccination agreements. *J. R. Soc. Interface* **13**, 20150907 (2016). doi: 10.1098/rsif.2015.0907; pmid: 26790996
21. K. M. Bubar *et al*., Model-informed COVID-19 vaccine prioritization strategies by age and serostatus. *Science* **371**, 916–921 (2021). doi: 10.1126/science.abe6959; pmid: 33479118
22. N. Mulberry, P. Tupper, E. Kirwin, C. McCabe, C. Colijn, Vaccine rollout strategies: The case for vaccinating essential workers early. *medRxiv* 2021.02.23.21252309 [Preprint] (2021); doi: 10.1101/2021.02.23.21252309
23. C. E. Wagner *et al*., Economic and behavioral influencers of vaccination and antimicrobial use. *Front. Public Health* **8**, 614113 (2020). doi: 10.3389/fpubh.2020.614113; pmid: 33409264
24. J. Zahradník *et al*., SARS-CoV-2 RBD in vitro evolution follows contagious mutation spread, yet generates an able infection inhibitor. *bioRxiv* 2021.01.06.425392 [Preprint] (2021); doi: 10.1101/2021.01.06.425392
25. A. Rambaut *et al*., Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations (2020); https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563.
26. R. Burioni, E. J. Topol, Has SARS-CoV-2 reached peak fitness? *Nat. Med.* (2021). doi: 10.1038/s41591-021-01421-7; pmid: 34155413
27. N. G. Davies, P. Klepac, Y. Liu, K. Prem, M. Jit, CMMID COVID-19 working group, Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med.* **26**, 1205–1211 (2020). doi: 10.1038/s41591-020-0962-9; pmid: 32546824
28. H. M. Korevaar *et al*., Quantifying the impact of US state non-pharmaceutical interventions on COVID-19 transmission. *medRxiv* 2020.06.30.20142877 [Preprint] (2020); doi: 10.1101/2020.06.30.20142877
29. B. L. Rice *et al*., Variation in SARS-CoV-2 outbreaks across sub-Saharan Africa. *Nat. Med.* **27**, 447–453 (2021). doi: 10.1038/s41591-021-01234-8; pmid: 33531710
30. C. E. Wagner *et al*., Vaccine nationalism and the dynamics and control of SARS-CoV-2 interactive dashboard (2021); https://grenfelllab.princeton.edu/vaccine-nationalism.
31. J. M. Dan *et al*., Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* **371**, eabf4063 (2021). doi: 10.1126/science.abf4063; pmid: 33408181
32. S. P. Anand *et al*., Longitudinal analysis of humoral immunity against SARS-CoV-2 Spike in convalescent individuals up to 8 months post-symptom onset. *Cell Reports Medicine* **2**, 100290 (2021). doi: 10.1016/j.xcrm.2021.100290; pmid: 33969322
33. V. Shinde *et al*., Efficacy of NVX-CoV2373 Covid-19 vaccine against the B.1.351 variant. *N. Engl. J. Med.* **384**, 1899–1909 (2021). doi: 10.1056/NEJMoa2103055; pmid: 33951374
34. C. A. Prete Jr. *et al*., Reinfection by the SARS-CoV-2 P.1 variant in blood donors in Manaus, Brazil. *medRxiv* 2021.05.10.21256644 [Preprint] (2021); doi: 10.1101/2021.05.10.21256644
35. C. J. Reynolds *et al*., Prior SARS-CoV-2 infection rescues B and T cell responses to variants after first vaccine dose. *Science* **372**, eabh1282 (2021). pmid: 33931567
36. J. L. Bernal *et al*., Effectiveness of COVID-19 vaccines against the B.1.617.2 variant. *medRxiv* 2021.05.22.21257658 [Preprint] (2021); doi: 10.1101/2021.05.22.21257658
37. J. Stowe *et al*., Effectiveness of COVID-19 vaccines against hospital admission with the Delta (B.1.617.2) variant [Preprint] (2021); https://khub.net/documents/135939561/479607266/Effectiveness+of+COVID-19+vaccines+against+hospital+admission+with+the+Delta+%28B.1.617.2%29+variant.pdf/1c213463-3997-ed16-2a6f-14e5deb0b997?t=1623689315431.
38. S. Nasreen *et al*., Effectiveness of COVID-19 vaccines against variants of concern, Canada. *medRxiv* 2021.06.28.21259420 [Preprint] (2021); doi: 10.1101/2021.06.28.21259420
39. D. Lieber, "Pfizer vaccine less effective against delta infections but prevents severe illness, Israeli data show," *Wall Street Journal*, 6 July 2021; www.wsj.com/articles/pfizers-covid-19-vaccine-is-less-effective-against-delta-variant-israeli-data-show-11625572796?fbclid=IwAR1j5TqBVHtQH7LGxyc6GQUFhsuPvvVd2C7HsBd-7kvY6p-2Gk7-FipnKm4.
40. L. Stamatatos *et al*., mRNA vaccination boosts cross-variant neutralizing antibodies elicited by SARS-CoV-2 infection. *Science* **372**, 1413–1418 (2021). doi: 10.1126/science.abg9175; pmid: 33766944
41. J. S. Lavine, O. N. Bjornstad, R. Antia, Immunological characteristics govern the transition of COVID-19 to endemicity. *Science* **371**, 741–745 (2021). doi: 10.1126/science.abe6522; pmid: 33436525
42. R. Laxminarayan *et al*., Epidemiology and transmission dynamics of COVID-19 in two Indian states. *Science* **370**, 691–697 (2020). doi: 10.1126/science.abd7672; pmid: 33154136
43. C. M. Saad-Roy, A. B. McDermott, B. T. Grenfell, Dynamic perspectives on the search for a universal influenza vaccine. *J. Infect. Dis.* **219**, S46–S56 (2019). doi: 10.1093/infdis/jiz044; pmid: 30715467
44. K. Koelle, S. Cobey, B. Grenfell, M. Pascual, Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **314**, 1898–1903 (2006). doi: 10.1126/science.1132745; pmid: 17185596
45. E. M. Volz, S. L. Kosakovsky Pond, M. J. Ward, A. J. Leigh Brown, S. D. W. Frost, Phylodynamics of infectious disease epidemics. *Genetics* **183**, 1421–1430 (2009). doi: 10.1534/genetics.109.106021; pmid: 19797047
46. E. M. Volz, K. Koelle, T. Bedford, Viral phylodynamics. *PLOS Comput. Biol.* **9**, e1002947 (2013). doi: 10.1371/journal.pcbi.1002947; pmid: 23555203
47. T. Kustin *et al*., Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2 mRNA vaccinated individuals. *medRxiv* 2021.04.06.21254882 [Preprint] (2021); doi: 10.1101/2021.04.06.21254882
48. B. Choi *et al*., Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N. Engl. J. Med.* **383**, 2291–2293 (2020). doi: 10.1056/NEJMc2031364; pmid: 33176080
49. R. E. Baker, W. Yang, G. A. Vecchi, C. J. E. Metcalf, B. T. Grenfell, Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. *Science* **369**, 315–319 (2020). doi: 10.1126/science.abc2535; pmid: 32423996
50. R. E. Baker *et al*., Epidemic dynamics of respiratory syncytial virus in current and future climates. *Nat. Commun.* **10**, 5512 (2019). doi: 10.1038/s41467-019-13562-y; pmid: 31797866
51. C. E. Wagner *et al*., Code for Vaccine nationalism and the dynamics and control of SARS-CoV-2, Zenodo (2021); doi: 10.5281/zenodo.5119582

ACKNOWLEDGMENTS

Funding: This work was funded in part by Open Philanthropy; the Natural Sciences and Engineering Research Council of Canada through a Postgraduate-Doctoral Scholarship (C.M.S.-R.); the Cooperative Institute for Modelling the Earth System (CIMES) (R.E.B.); the James S. McDonnell Foundation 21st Century Science Initiative Collaborative Award in Understanding Dynamic and Multi-scale Systems (C.M.S.-R. and S.A.L.); the C3.ai Digital Transformation Institute and Microsoft Corporation (S.A.L.); Gift from Google, LLC (S.A.L.); the National Science Foundation (CNS-2027908 and CCF1917819) (S.A.L.); the US Centers for Disease Control and Prevention (B.T.G.); and the Flu Lab (B.T.G.). **Author contributions:** C.E.W., C.M.S.-R., and B.T.G. designed the study. C.E.W. and C.M.S.-R. performed the simulations and analyses and wrote the manuscript. S.E.M. developed the Shiny application. C.E.W., C.M.S.-R., S.E.M., R.E.B., M.J.M., J.F., E.C.H., O.G.P., A.L.G., E.J.E., S.A.L., C.E.J.M., and B.T.G. contributed to interpreting the results and editing the manuscript. **Competing interests:** O.G.P. has undertaken consulting for AstraZeneca relating to the genetic diversity and classification of SARS-CoV-2 lineages. The authors have no other competing interests. **Data and materials availability:** This manuscript contains no new data, and all referenced datasets are publicly available. Code for all analyses is available at Zenodo (*51*). 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

https://science.org/doi/10.1126/science.abj7364
Supplementary Text
Figs. S1 to S8
References (*52*–*67*)
MDAR Reproducibility Checklist

28 May 2021; accepted 9 August 2021
Published online 17 August 2021
10.1126/science.abj7364

RESEARCH ARTICLES

ENZYME MECHANISM

Hybrid radical-polar pathway for excision of ethylene from 2-oxoglutarate by an iron oxygenase

Rachelle A. Copeland[1]*, Shengbin Zhou[1]†, Irene Schaperdoth[1], Tokufu Kent C. Shoda[1]‡, J. Martin Bollinger Jr.[1,2]*, Carsten Krebs[1,2]*

Microbial ethylene-forming enzyme (EFE) converts the C3–C4 fragment of the ubiquitous primary metabolite 2-oxoglutarate (2OG) to its namesake alkene product. This reaction is very different from the simple decarboxylation of 2OG to succinate promoted by related enzymes and has inspired disparate mechanistic hypotheses. We show that EFE produces stereochemically random (equal cis and trans) 1,2-[$^2H_2$]-ethylene from (3*S*,4*R*)-[$^2H_2$]-2OG, appends an oxygen from O_2 on the C1-derived (bi)carbonate, and can be diverted to ω-hydroxylated monoacid products by modifications to 2OG or the enzyme. These results implicate an unusual radical-polar hybrid mechanism involving iron(II)-coordinated acylperoxycarbonate and alkylcarbonate intermediates. The mechanism explains how EFE accesses a high-energy carboxyl radical to initiate its fragmentation cascade, and it hints at capabilities of 2OG-dependent enzymes that may await discovery and exploitation.

Iron(II)- and 2-oxoglutarate–dependent (Fe/2OG) oxygenases catalyze hydroxylation, halogenation, epimerization, desaturation, cyclization, and ring-expansion reactions that are important for nutrient and energy acquisition, regulation, and biosynthesis in organisms ranging from microbes to humans (*1*–*3*). These divergent reactions arise from different fates of a common intermediate harboring a hydroxoiron(III) complex and carbon-centered radical, which is produced when an oxoiron (IV) (ferryl) intermediate abstracts hydrogen from the substrate (Fig. 1B, top) (*2*, *4*). The ferryl complex forms in a conserved pathway initiated by addition of O_2 to the Fe(II) cofactor (*2*, *4*, *5*). Evidence suggests that a formal replacement of the C1 carboxylate by the dioxygen unit yields an Fe(II)-peroxysuccinate complex (*5*). Heterolytic cleavage of the peroxide bond yields the common coproduct, succinate, as a ligand to the hydrogen-abstracting ferryl intermediate (*2*, *4*).

Microbial ethylene-forming enzyme (EFE) conforms to this functional paradigm only in its minor reaction. In ~30% of its turnovers, the canonical ferryl complex abstracts hydrogen from C5 of L-arginine (*6*) to initiate its oxidative fragmentation to guanidine and pyrroline-5-carboxylate (P5C) (Fig. 1A, bottom) (*7*–*9*). The ferryl complex that initiates L-arginine oxidation (RO) is not an intermediate in the major (~70% of turnovers) ethylene-forming (EF) reaction; the pathways branch from an earlier state (*6*). The EF reaction is a four-electron oxidation that cleaves three of the four carbon-carbon bonds in 2OG (C1–C2, C2–C3, and C4–C5) to yield three CO_2 equivalents and excise the C3–C4 unit intact. It spares, but nevertheless still requires, the L-Arg prime substrate of the minor RO reaction (Fig. 1A, top) (*6*–*9*). In the absence of L-Arg, EFE converts 2OG only to succinate and CO_2 (fig. S1A) (*10*, *11*).

Multiple polar-concerted mechanisms—essentially seven-center Eschenmoser/Grob-like multibond fragmentations—were initially proposed for the EF reaction (*9*, *12*). Each invoked an oxidized iron complex acting as nucleofuge to the electrofugal C5 carboxylate (fig. S1B). More recent work has invoked the Kolbe, Hunsdiecker-Borodin, and Barton decarboxylation reactions (among others) as precedents for the initial fragmentation in a multistep free-radical mechanism (Fig. 1B, bottom). In these reactions, oxidation of a carboxylic acid to a neutral radical initiates its fragmentation to CO_2 and a primary carbon-centered radical (*13*–*15*). For the EF reaction, the cognate sequence would have succin-1-yl ($R_1COO^•$ in Fig. 1B, bottom) fragmenting to propanoate-3-yl ($R_1^•$), from which multiple pathways to ethylene and CO_2 can be envisaged. Two recent computational studies proposed an unprecedented step to access the initiating carboxyl radical, in which O_2 formally inserts into the C1–C2 bond (Fig. 1B, bottom) rather than replacing C1 (top) (*16*, *17*). Cleavage of the O–O bond in the resultant Fe(II)-succinylperoxycarbonate complex was suggested to generate the succinyl radical. To date, no experimental data capable of distinguishing among these disparate polar and radical mechanisms have been reported, but the relative fragility of the EF pathway and robustness

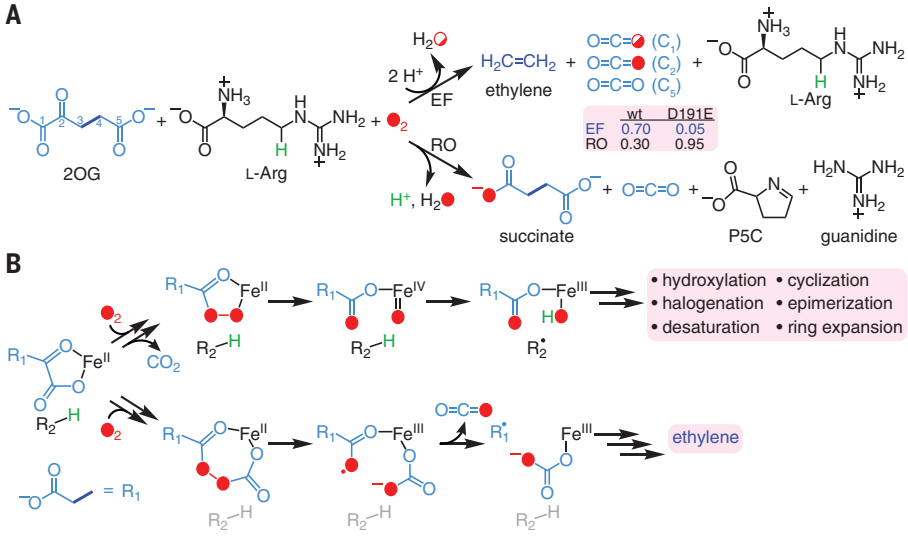


Fig. 1. Reactions catalyzed and possible pathways followed by EFE and other Fe/2OG oxygenases. (**A**) Ethylene-forming (EF, top) and L-arginine–oxidizing (RO, bottom) reactions catalyzed by EFE. The partition ratios of the wild-type (wt) and D191E variant protein are shown in the shaded box at the center. (**B**) Pathway to the prime substrate–derived radical ($R_2^•$) in a canonical Fe/2OG oxygenase (top) and pathway to the 2OG-derived propanoate-3-yl ($R_1^•$) in EFE, now experimentally validated in this work (bottom). The red shading traces the path of oxygen atoms from O_2 into the products.

[1]Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA. [2]Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA.
*Corresponding author. Email: rac346@psu.edu (R.A.C.); jmb21@psu.edu (J.M.B.); ckrebs@psu.edu (C.K.)
†Present address: School of Pharmaceutical Science and Technology, Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences (UCAS), Hangzhou 310024, China.
‡Present address: BioReference Laboratories, Elmwood Park, NJ 07407, USA.

of the RO pathway to substrate modifications and amino acid substitutions have emerged as potential clues. One recent study showed that substitution of cofactor ligand Asp^{191} with Glu (D191E) severely compromises ethylene formation but has minimal impact on L-arginine oxidation, thus shifting the partition ratio (EF:RO) from ~2.5 to ~0.05 (Fig. 1A) (*6*).

Ethylene stereochemistry implying a multistep mechanism

The two broad classes of mechanisms proposed to date evoke different predictions regarding the configuration of the ethylene product (Fig. 2A). The antiperiplanar disposition of electrofugal and nucleofugal groups required for the concerted mechanisms would fix the relative dispositions of the C3 and C4 methylene hydrogens, whereas progression through the propanoate-3-yl radical by β-scission of succin-1-yl would likely randomize the ethylene configuration. We accessed (3*S*,4*R*)-[$^2H_2$]-2OG by enzymatic deamination of the corresponding synthetic L-glutamic acid (see supplementary materials for the synthetic procedure) and evaluated the stereochemical course of ethylene formation by Fourier transform infrared (FTIR) absorption spectroscopy. We detected the 1,2-[$^2H_2$]-ethylene product in the gas above coupled reactions in which the chirally labeled L-Glu was first converted to 2OG by L-glutamate:pyruvate aminotransferase, and 2OG was then consumed by EFE (Fig. 2B). The peaks at 727 and 988 cm^{-1} arise from *trans*-1,2-[$^2H_2$]-ethylene, and the peak at 842 cm^{-1} from the cis isomer (*18*, *19*). Omission of EFE or use of unlabeled 2OG instead of (3*S*,4*R*)-[$^2H_2$]-2OG eliminated these FTIR absorption peaks, with the latter substitution producing a peak from unlabeled ethylene. The experimental spectrum from the (3*S*,4*R*)-[$^2H_2$]-2OG reaction matches most closely that of standard gas mixtures containing the two 1,2-[$^2H_2$]-ethylene isomers at approximately equal partial pressures (Fig. 2B and fig. S3), implying complete randomization of stereochemistry in the EFE reaction. We verified, using a published procedure based on the ability of isocitrate lyase to abstract only the pro-*S* hydrogens of succinate (*20*, *21*), that the coproduct emerging from the canonical RO pathway retained the expected (2*S*,3*R*) configuration (Fig. 2C), thus establishing that the random cis and trans stereochemistry observed in the ethylene product was introduced by EFE and not during the enzymatic conversion of the synthetic (3*S*,4*R*)-[$^2H_2$]-L-Glu to 2OG. This randomization implies the formation of an intermediate that can freely rotate around the C3–C4 bond (2OG numbering). Because this rotation would be restricted in the active site by an intact C2–C3 bond, the intermediate must have this bond cleaved; its most likely identity is the propanoate-3-yl radical.

Abortive products from a radical intermediate

Reactions of six 2OG analogs afforded additional evidence of the proposed radical intermediate in the form of partially fragmented products resulting from derailment of subsequent alkene formation. Predictable alternative decay pathways of this radical include coupling to either H$^\bullet$ (e.g., from the protein) or HO$^\bullet$ (a cofactor ligand) to produce a monoacid or ω-hydroxylated monoacid (ω-OH-acid) product, respectively (Fig. 3A). In reactions of EFE with 2OG analogs modified at C4 by mono- or difluorination, methylation, hydroxylation, or hydroxylation and methylation, the corresponding substituted ethylene products (or the carbonyl compounds to which they would convert) were either barely detectable or undetectable. In each case, liquid chromatography–mass spectrometry (LC-MS) chromatograms exhibited peaks at the mass-to-charge ratio (*m*/*z*) of the anticipated ω-OH-acid product (Fig. 3, B and D, and figs. S6 to S9). For the ω-OH-acid compounds that were commercially available (4-hydroxybutanoate, 2-methyl-3-hydroxypropanoate, and 2,3-dihydroxypropanoate), we verified that the alternative EFE reaction products eluted at the same times as the standards (Fig. 3, B and D, and fig. S9). For the C_6 2OG analog 2-oxoadipate (2OA), the presence of the additional methylene spacer between the ω-carbon radical and distal carboxylate would be expected to render it immune to a second fragmentation, and, indeed, oxidation of 2OA by EFE yielded primarily the corresponding ω-OH-acid product, 4-hydroxybutanoate (Fig. 3C and fig. S9), along with the simple decarboxylation product, glutarate, in a partition ratio of ~1:1. For each of the 2-oxodiacids, omission of the essential activator of the ethylene pathway, L-arginine, abolished production of the abortive

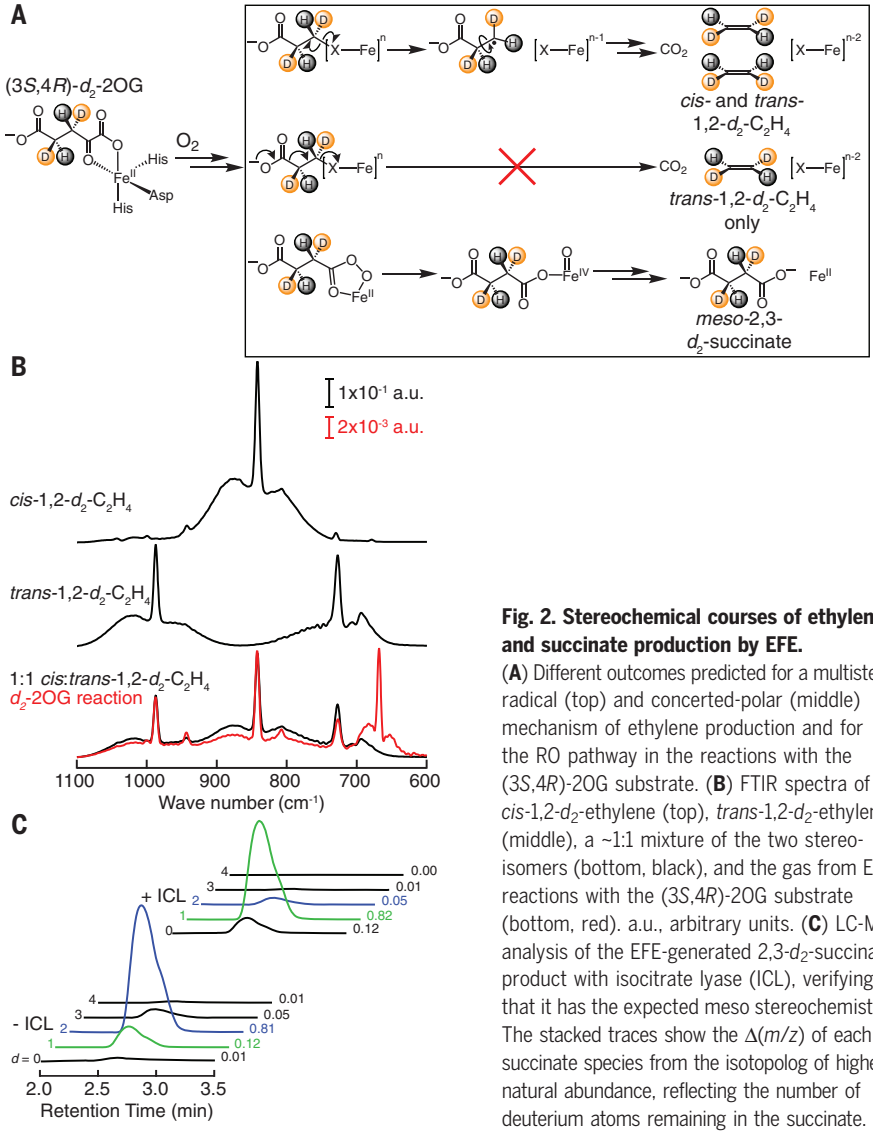


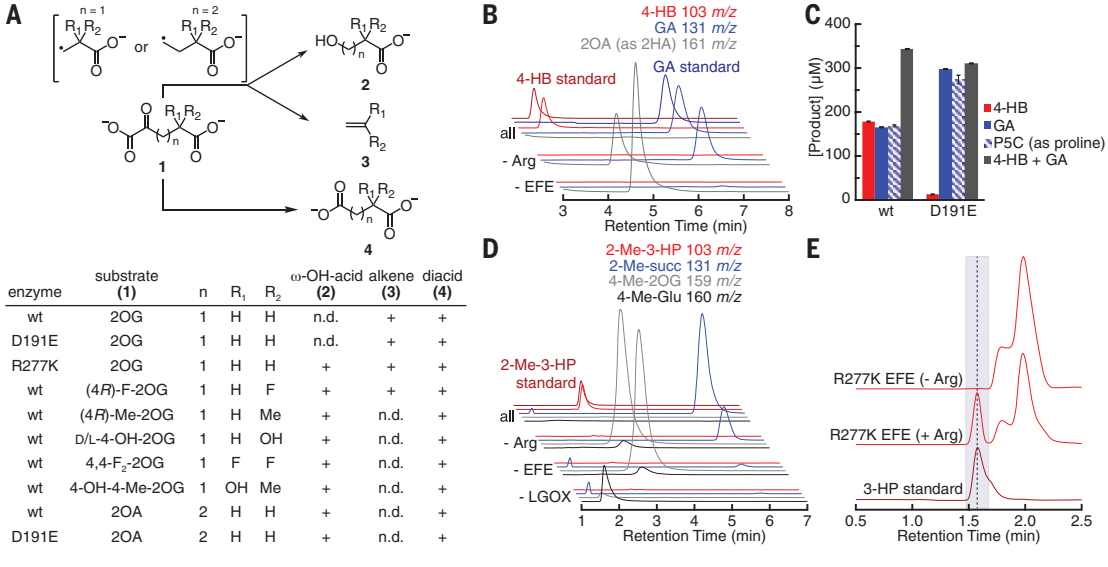
Fig. 2. Stereochemical courses of ethylene and succinate production by EFE. (**A**) Different outcomes predicted for a multistep-radical (top) and concerted-polar (middle) mechanism of ethylene production and for the RO pathway in the reactions with the (3*S*,4*R*)-2OG substrate. (**B**) FTIR spectra of *cis*-1,2-d_2-ethylene (top), *trans*-1,2-d_2-ethylene (middle), a ~1:1 mixture of the two stereoisomers (bottom, black), and the gas from EFE reactions with the (3*S*,4*R*)-2OG substrate (bottom, red). a.u., arbitrary units. (**C**) LC-MS analysis of the EFE-generated 2,3-d_2-succinate product with isocitrate lyase (ICL), verifying that it has the expected meso stereochemistry. The stacked traces show the Δ(*m*/*z*) of each succinate species from the isotopolog of highest natural abundance, reflecting the number of deuterium atoms remaining in the succinate.

ω-OH-acid product but not the ordinary C_{n-1} diacid coproduct (Fig. 3A). Similarly, the D191E variant protein, previously shown to produce only ~7% as much ethylene as wild-type EFE (*6*), converted 2OA almost exclusively to glutarate; 4-hydroxybutanoate was barely detectable (Fig. 3C). These observations associate the ω-OH-acid product specifically with the ethylene pathway. The diversion of the namesake pathway to ω-OH-acid products by six disparate 2OG modifications suggested that the corresponding reactivity might be elicited from 2OG itself by substitutions to amino acids in the EFE active site. Indeed, substitution of Arg^{277}, which interacts with the 2OG C5 carboxylate (*6*, *9*, *12*), by Lys (R277K) yielded a variant that produces both ethylene and 3-hydroxypropanoate (Fig. 3E and fig. S11), confirming that the ω-OH-acid can also be generated from the native substrate in a partition with the alkene.

Transfer of oxygen from O_2 to C1-derived (bi)carbonate

The most frequently proposed precursor to the ferryl intermediate, the Fe(II)-peroxysuccinate complex (Fig. 1B, top), could conceivably undergo O–O bond homolysis rather than the usual heterolysis to initiate the radical fragmentation; the resultant oxoiron(III)/succin-1-yl intermediate would represent, in essence, a charge-transfer excited state of the usual oxoiron(IV)/succinate intermediate. It is not obvious that the succin-1-yl radical formed by this route could persist long enough to initiate the C2–C3 fragmentation step. The recent computational studies suggesting C1–C2 insertion (*16*, *17*), rather than C1 replacement, by O_2 would provide a chemical rationale for accession of the radical manifold (Fig. 1B, bottom). The two possible products of O–O bond cleavage from this alternative intermediate would be the (bi)carbonatoiron(IV)/succinate and (bi)carbonatoiron(III)/succin-1-yl states, and, without the stabilizing effect of the strong bond to the oxo ligand, the high-valent state would be expected to have the higher energy. The predictions evoked by this mechanism are that (bi)carbonate, rather than CO_2, should be the initial product derived from C1 and that it should bear an oxygen atom derived from O_2. We used the enzymes phosphoenolpyruvate carboxylase and malate dehydrogenase (PEPC/MD) to trap the presumptive (bi)carbonate product as malate for LC-MS detection (*22*). For the reaction with 1-[^{13}C]-2OG, LC-MS chromatograms showed a prominent peak with a *m/z* of 134, equivalent to that predicted for malate with one atom of ^{13}C (Fig. 4A). This peak was nearly absent after identical treatment of EFE reactions containing unlabeled 2OG. In analysis of reactions carried out under $^{18}O_2$, a new peak at *m/z* 136 (+2 units) had about the same intensity as the peak at *m/z* 134. We could not detect ^{18}O-labeled malate in analysis of EFE reactions lacking L-Arg, suggesting that it resulted from incorporation (into malate by PEPC/MD) of a $^{13}C^{18}O^{16}O^-$ unit from the C1-derived (bi)carbonate product of the ethylene pathway.

We confirmed by high-resolution FTIR spectra of the gas from EFE reactions that the ethylene pathway appends an O_2-derived oxygen on C1 (Fig. 4B). Dehydration of the presumptive $C^{18}O^{16}O_2^{2-}$ product to CO_2 should result in retention of two-thirds of the ^{18}O (Fig. 1A, red shading). The presence of one ^{18}O atom shifts the energy of the CO_2 bending mode from 668 cm^{-1} to 663 cm^{-1} for $^{12}CO_2$ and from 648 cm^{-1} to 644 cm^{-1} for $^{13}CO_2$ (Fig. 4B, arrows) (*23*). The shifted features of the $^{16}O=^{12}C=^{18}O$ and $^{16}O=^{13}C=^{18}O$ isotopologs are visible in the spectra of gas from complete EFE reactions carried out under $^{18}O_2$ (Fig. 4B, colored traces) but not in the spectra of gas from reactions carried out under $^{16}O_2$ (black traces). The 663 and 644 cm^{-1} features are also absent from spectra of gas taken from control reactions lacking L-Arg (yellow) or containing the Fe/2OG L-arginine 3-hydroxylase VioC [from the biosynthetic pathway to viomycin (*24*)] in place of EFE (purple). Incorporation of ^{18}O from $^{18}O_2$ into CO_2 is thus strictly associated with ethylene production. Comparison of the spectra from reactions containing 2OG ^{13}C-labeled either only at C1 (blue) or at C1–C4 (green) further validates this conclusion. The EF pathway generates CO_2 equivalents from C1, C2, and C5 and should incorporate an O_2-derived oxygen into those derived from C1 and C2. In the spectra from reactions with 1-[^{13}C]-2OG (blue), features of both $^{16}O=^{13}C=^{18}O$ (from C1) and $^{16}O=^{12}C=^{18}O$



enzyme	substrate (1)	n	R_1	R_2	ω-OH-acid (2)	alkene (3)	diacid (4)
wt	2OG	1	H	H	n.d.	+	+
D191E	2OG	1	H	H	n.d.	+	+
R277K	2OG	1	H	H	+	+	+
wt	(4*R*)-F-2OG	1	H	F	+	+	+
wt	(4*R*)-Me-2OG	1	H	Me	+	n.d.	+
wt	D/L-4-OH-2OG	1	H	OH	+	n.d.	+
wt	4,4-F_2-2OG	1	F	F	+	n.d.	+
wt	4-OH-4-Me-2OG	1	OH	Me	+	n.d.	+
wt	2OA	2	H	H	+	n.d.	+
D191E	2OA	2	H	H	+	n.d.	+

Fig. 3. Detection of abortive ω-OH-acid products from 2OG analogs and the native 2-oxodiacid in wild-type and variant proteins. (**A**) Products generated by wild-type EFE and two variants from 2OG and C4-substituted 2OG analogs (*n* = 1) and from 2OA (*n* = 2). n.d., not detected. (**B**) LC-MS chromatograms demonstrating formation of 4-OH-butanoate (4-HB) from 2OA [detected after reduction to 2-hydroxyadipate (2HA)] and its dependence on the presence of L-Arg, which is required for ethylene production in the native reaction. (**C**) Quantification of the ω-OH-acid, diacid [glutaric acid (GA)], and P5C (after reduction to proline) products in the EFE reaction with 2OA. The values shown are means of two trials; the error bars show the standard errors of the means. (**D**) LC-MS chromatograms showing production of 2-Me-3-OH-propanoate (2-Me-3-HP) and 2-Me-succinate from 4-Me-2OG [generated by deamination of 4-Me-L-glutamate by L-glutamate oxidase (LGOX)] by EFE and its dependence on both enzymes and the L-Arg activator of the ethylene pathway. (**E**) LC-MS traces showing L-Arg–dependent production of 3-hydroxypropionic acid (3-HP) from 2OG by the R277K variant of EFE. Procedures are provided in the supplementary methods.

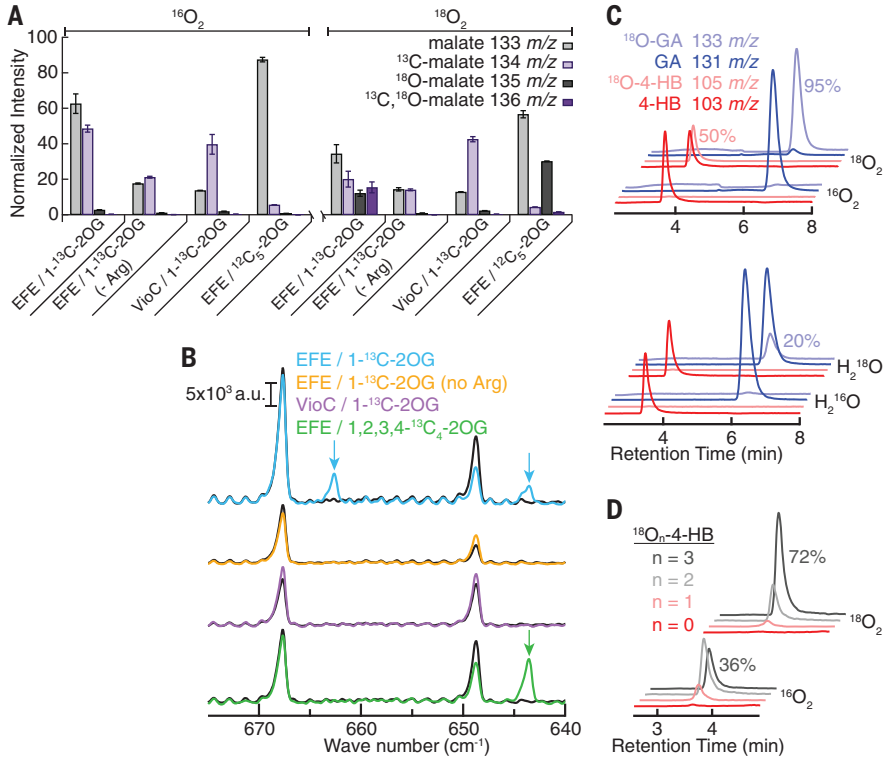


Fig. 4. Oxygen isotope tracer experiments showing transfer from O_2 to C1 of 2OG and from the resultant (bi)carbonate to the ω-carbon of the abortive product. (**A**) LC-MS quantification of malate isotopologs upon trapping of bicarbonate with phosphoenolpyruvate carboxylase and malate dehydrogenase (PEPC/MD). The fact that malate containing a $^{13}C^{18}O^{16}O^-$ unit (dark purple bars) is prominent only in the EFE reaction with 1-[^{13}C]-2OG, $^{18}O_2$, and L-Arg (first set of bars in right panel) associates this incorporation specifically with the ethylene pathway. The colored bars represent the means of two trials, and the error bars show the standard errors. (**B**) High-resolution FTIR spectra in the region of the CO_2 bending mode of the gas from enzyme reactions. The colored spectra are from reactions under $^{18}O_2$ and the black spectra from the otherwise identical reactions under $^{16}O_2$. The arrows indicate the peaks of the $^{16}O{=}^{12}C{=}^{18}O$ (left, blue) and $^{16}O{=}^{13}C{=}^{18}O$ (right, blue and green) mixed isotopologs generated specifically in the ethylene pathway of EFE. VioC is the L-arginine 3-hydroxylase from the viomycin biosynthetic pathway (*24*). (**C**) LC-MS chromatograms showing transfer of ^{18}O from $^{18}O_2$ to both the C_{n-1} diacid and the ω-OH-acid products generated by EFE from 2OA (top). The fact that ~50% of the 4-HB (the ω-OH-acid) product became labeled in the $^{18}O_2$ reaction implied a second source of the appended O atom. The reaction in ^{18}O-enriched solvent (86%, bottom) incorporated ^{18}O into the C_{n-1} diacid product glutarate (GA) but not the ω-OH-acid product (4-HB). (**D**) Prior exchange of the 2OA carboxylates by an extended incubation (with heating) in acidic $H_2{}^{18}O$ resulted in incorporation of ^{18}O into the 4-HB (bottom), showing that a carboxylate of the 2-oxodiacid substrate is the second source. Use of pre-exchanged 2OA and $^{18}O_2$ resulted in a maximum incorporation of 72% (top); the shortfall from unity is explained in the supplementary materials.

(from C2) are visible, and the peak of the ^{12}C mixed-oxygen isotopolog is slightly more intense, presumably because C2 is converted directly to CO_2 and need not lose an oxygen atom to enter the gas phase, as the C1-derived (bi)carbonate must do. In the corresponding spectra from the reaction with 1,2,3,4-[$^{13}C_4$]-2OG (green), only the peak for the $^{16}O{=}^{13}C{=}^{18}O$ mixed-oxygen isotopolog—now from both C1 and C2—is present, and it is correspondingly more intense than in the spectrum from the reaction with 1-[^{13}C]-2OG.

Evidence of radical-carbonate coupling and concerted-polar fragmentation

Observation of random cis and trans 1,2-[$^2H_2$]-ethylene stereochemistry, abortive ω-OH-acid products, and O_2-derived oxygen appended to C1 thus confirmed key expectations of the mechanism shown in Fig. 1B. We envisaged three possible paths for conversion of the propanoate-3-yl intermediate to ethylene. First, a second radical fragmentation could directly generate ethylene along with the one-electron reduced form of CO_2, which would reduce the Fe(III) cofactor to its resting Fe(II) state (fig. S13A, top). Second, a preceding electron transfer from the C3 radical to the cofactor, possibly coupled to ligand protonation, could generate a C3 carbocation that would readily undergo polar decarboxylation, yielding ethylene and the C5-derived CO_2 (fig. S13A, middle). Although these pathways would have precedent among mechanisms proposed for other enzymes, we disfavor them on the grounds that the first pathway ought not to be impeded by the C4 substitutions nor the R277K replacement, whereas the second might be compromised by electron-withdrawing substituents on C4 (e.g., 4-F-2OG, 4-F_2-2OG, and 4-OH-2OG) but not by either the electron-donating substituent (4-Me-2OG) nor the charge-conservative R277K substitution. Indeed, the observation of ω-OH-acid production from all five C4-substituted 2OG analogs and the variant protein suggests a stereoelectronic origin of the altered outcome.

The third pathway, involving (i) radical coupling between propanoate-3-yl and the Fe(III)-coordinated carbonate to generate 2-(1-carboxyethyl)carbonate followed by (ii) a five-center Eschenmoser/Grob-like fragmentation (Fig. 5, green box, and fig. S14A, bottom), would rationalize the observed sensitivity to C4 and active-site substitutions, because even modest steric perturbation of the antiperiplanar disposition of carboxylate electrofuge and carbonate nucleofuge required for such a polar-concerted reaction could slow it enough for elimination of the alcohol from the unstable carbonate monoester to compete. This mechanism would imply that the hydroxyl group installed on the ω-OH-acid products should be derived partly from O_2 (which donates one O atom to form the C1-derived carbonate) and partly from the C1 carboxylate of the 2-oxodiacid. The fraction deriving from each source would depend on the competition between carbonate-ligand dynamics and radical coupling. Stable coordination of one O atom through the reaction sequence would render only the two nonligand oxygen atoms, one derived from O_2 and one from the C1 carboxylate, subject to coupling to the substrate ω-radical. Under the reasonable assumption of rapid rotation of the carbonate ligand (as indicated by the curved arrow in Fig. 5), each source would contribute 50% of the incorporated oxygen. Alternatively, more facile ligand dynamics could allow for interchange of all three carbonate oxygens, possibly rendering them all equally susceptible to radical coupling. In this case, one-third of the incorporated O atoms would derive from O_2 and two-thirds from the C1 carboxylate.

The observed origins of the oxygen incorporated into the ω-OH-acid products derived

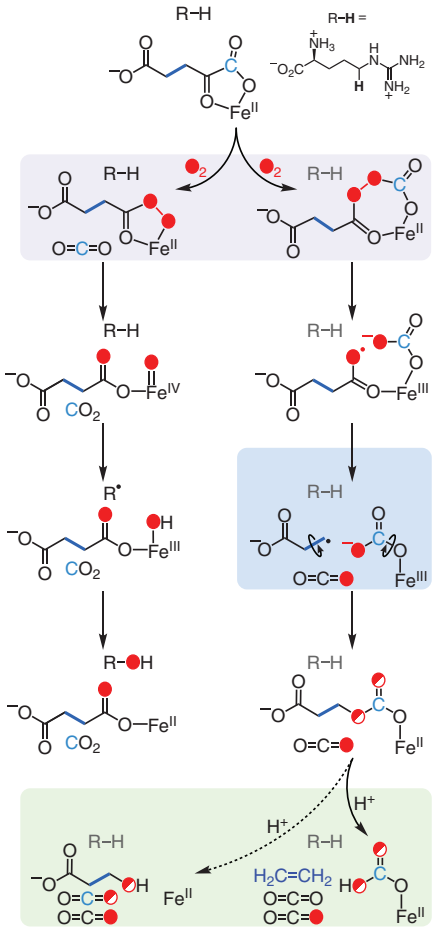


Fig. 5. Mechanisms of the EFE reactions implicated by this and prior studies. The isotope-tracer results presented here are summarized in red, and the steps supported by these experiments (gray and green), the detection of the abortive ω-OH-acid product (green), and the stereochemistry determination (blue) are indicated by the shaded boxes.

from 2OG, 2OA, and 4-OH-2OG provide evidence for radical-carbonate coupling and imply that the first case above—stable carbonate coordination with free rotation—is operant. About half of each ω-OH-acid product harbored ^{18}O when generated in reactions under $^{18}O_2$ in solvent of natural isotopic abundance (Fig. 4C and figs. S9 to S11). Use of solvent enriched in $H_2{}^{18}O$ and $^{16}O_2$ in the EFE reaction with 2OA resulted in formation of ^{18}O-glutarate (via exchange of the 2OA carbonyl) but not 4-^{18}OH-butanoate, implying that the other oxygen source is neither the solvent nor the substrate carbonyl (Fig. 4C). Use of 2OA that had been subjected to prior exchange of ^{18}O from solvent into its carboxylates (by incubation for several days in acidic H_2O containing ~90% ^{18}O) did result in incorporation of ^{18}O into the ω-OH-acid product (Fig. 4D), indicating that a carboxylate is the second source, as depicted in Fig. 5. The fractions of the 4-OH-butanoate having ^{18}O at all three positions from two paired reactions with ^{18}O label either only in the 2OA (36%) or in both 2OA and O_2 (72%) agree well with those predicted on the basis of the proposed mechanism and the isotopic compositions of the carboxylate-exchanged 2OA (90% ^{18}O) and dioxygen (0 or 90% ^{18}O) in the reactions (see calculations in the supplementary materials).

Complete hybrid radical-polar mechanism of ethylene production

The complete EFE mechanism shown in Fig. 5 is supported by both the random ethylene stereochemistry and the two O atom transfers—from O_2 to the C1-derived CO_2 [via (bi)carbonate] and from the C1-carboxylate to a hydroxylated product—that have not previously been seen in the reaction of an Fe/2OG oxygenase. These outcomes appear to flow from the recently proposed C1–C2 O_2-insertion step, which provides an entirely new way to envisage access to activated iron and fragmented 2OG intermediates and/or products in this class of enzymes (*16*, *17*). The mechanism of the net insertion [the two computational studies advanced different hypotheses (*16*, *17*)], precisely how this step is promoted by the EFE active site, and whether it is specific to EFE, occurs in other Fe/2OG systems, or may be elicited by enzyme engineering all remain to be established. In addition, although couplings of substrate carbon radicals to Fe(III)-coordinated alkylthiolate, halide, nitrite, and azide have been implicated in studies of related enzymes (*25*–*27*), we are not aware of any precedent for the proposed radical coupling to carbonate. The elucidation of these two unusual steps—one predicted computationally and the second wholly unanticipated—and the implication that ethylene forms from the unstable carbonate ester of 3-hydroxypropanoate, itself a target for biotechnological production (*28*), in a step that can readily be interdicted to preserve the ω-OH-acid raise the exciting possibility of synergistic development of EFE-enabled bioprocesses to both important feedstocks.

REFERENCES AND NOTES

1. R. P. Hausinger, *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68 (2004).
2. J. M. Bollinger Jr. *et al.*, in *2-Oxoglutarate-Dependent Oxygenases*, C. J. Schofield, R. P. Hausinger, Eds. (Royal Society of Chemistry, 2015), pp. 95–122.
3. S. Kal, L. Que Jr., *J. Biol. Inorg. Chem.* **22**, 339–365 (2017).
4. C. Krebs, D. Galonić Fujimori, C. T. Walsh, J. M. Bollinger Jr., *Acc. Chem. Res.* **40**, 484–492 (2007).
5. A. J. Mitchell *et al.*, *J. Am. Chem. Soc.* **139**, 13830–13836 (2017).
6. R. A. Copeland *et al.*, *J. Am. Chem. Soc.* **143**, 2293–2303 (2021).
7. H. Fukuda, H. Kitajima, T. Fujii, M. Tazaki, T. Ogawa, *FEMS Microbiol. Lett.* **59**, 1–5 (1989).
8. H. Fukuda *et al.*, *Biochem. Biophys. Res. Commun.* **188**, 483–489 (1992).
9. S. Martinez *et al.*, *J. Am. Chem. Soc.* **139**, 11980–11988 (2017).
10. K. Nagahama *et al.*, *J. Gen. Microbiol.* **137**, 2281–2286 (1991).
11. S. Martinez, R. P. Hausinger, *Biochemistry* **55**, 5989–5999 (2016).
12. Z. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4667–4672 (2017).
13. R. G. Johnson, R. K. Ingham, *Chem. Rev.* **56**, 219–269 (1956).
14. D. H. R. Barton, D. Crich, W. B. Motherwell, *J. Chem. Soc. Chem. Commun.* **1983**, 939–941 (1983).
15. S. H. Shi, Y. Liang, N. Jiao, *Chem. Rev.* **121**, 485–505 (2021).
16. J. Xue, J. Lu, W. Lai, *Phys. Chem. Chem. Phys.* **21**, 9957–9968 (2019).
17. S. S. Chaturvedi, R. Ramanan, J. Hu, R. P. Hausinger, C. Z. Christov, *ACS Catal.* **11**, 1578–1592 (2021).
18. M. M. Rochkind, *Anal. Chem.* **40**, 762–768 (1968).
19. D. Viglaska-Aflalo, M. Rey, A. Nikitin, T. Delahaye, *Phys. Chem. Chem. Phys.* **22**, 3204–3216 (2020).
20. T. T. Tchen, H. Vanmilligan, *J. Am. Chem. Soc.* **82**, 4115–4116 (1960).
21. C. Ducrocq, D. Fraisse, J. C. Tabet, R. Azerad, *Anal. Biochem.* **141**, 418–422 (1984).
22. M. D. Lane, H. Maruyama, R. L. Easterday, *Methods Enzymol.* **13**, 277–283 (1969).
23. I. E. Gordon *et al.*, *J. Quant. Spectrosc. Radiat. Transf.* **203**, 3–69 (2017).
24. J. Ju, S. G. Ozanick, B. Shen, M. G. Thomas, *ChemBioChem* **5**, 1281–1285 (2004).
25. D. P. Galonić, E. W. Barr, C. T. Walsh, J. M. Bollinger Jr., C. Krebs, *Nat. Chem. Biol.* **3**, 113–116 (2007).
26. M. L. Matthews *et al.*, *Nat. Chem. Biol.* **10**, 209–215 (2014).
27. E. Tamanaha *et al.*, *J. Am. Chem. Soc.* **138**, 8862–8874 (2016).
28. X. R. Jiang, X. Yan, L. P. Yu, X. Y. Liu, G. Q. Chen, *Nat. Commun.* **12**, 1513 (2021).

ACKNOWLEDGMENTS

We thank T. Zimudzi of the Penn State Materials Research Institute for assistance with FTIR spectroscopy, L. Collazo-Perez for assistance with enzyme purification, and A. Boal for helpful discussions and contributions to obtaining funding. **Funding:** This work was supported by the Office of Basic Energy Science within the US Department of Energy Office of Science (BES Award DE-SC0016255). **Author contributions:** R.A.C., J.M.B., and C.K. conceived of the study. R.A.C., S.Z., J.M.B., and C.K. designed the experiments and interpreted the results. R.A.C. performed biochemical assays and FTIR, LC-MS and gas chromatography–mass spectrometry data collection and analysis. S.Z. performed synthesis and nuclear magnetic resonance characterization of the L-glutamate isotopolog. R.A.C., I.S., and T.K.C.S. performed protein expression and purification. R.A.C. assembled data figures with input from J.M.B. and C.K. The manuscript was written by R.A.C., J.M.B., and C.K. and was edited by all authors. J.M.B. and C.K. supervised and acquired funding for the research. **Competing interests:** R.A.C., J.M.B., and C.K. are listed as inventors on provisional patent application 63/224,278, describing a new method to produce 3-hydroxypropanoate, which relies partly on this work and has been submitted by the Penn State Research Foundation. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abj4290
Materials and Methods
Figs. S1 to S14
Tables S1 and S2
References (*29*–*31*)

12 May 2021; accepted 30 July 2021
Published online 12 August 2021
10.1126/science.abj4290

BATTERIES

Carbon-free high-loading silicon anodes enabled by sulfide solid electrolytes

Darren H. S. Tan,[1] Yu-Ting Chen,[1] Hedi Yang,[1] Wurigumula Bao,[1] Bhagath Sreenarayanan,[1] Jean-Marie Doux,[1] Weikang Li,[1] Bingyu Lu,[1] So-Yeon Ham,[1] Baharak Sayahpour,[1] Jonathan Scharf,[1] Erik A. Wu,[1] Grayson Deysher,[1] Hyea Eun Han,[2] Hoe Jin Hah,[2] Hyeri Jeong,[2] Jeong Beom Lee,[2] Zheng Chen[1,3,4,5]*, Ying Shirley Meng[1,4,5]*

The development of silicon anodes for lithium-ion batteries has been largely impeded by poor interfacial stability against liquid electrolytes. Here, we enabled the stable operation of a 99.9 weight % microsilicon anode by using the interface passivating properties of sulfide solid electrolytes. Bulk and surface characterization, and quantification of interfacial components, showed that such an approach eliminates continuous interfacial growth and irreversible lithium losses. Microsilicon full cells were assembled and found to achieve high areal current density, wide operating temperature range, and high areal loadings for the different cells. The promising performance can be attributed to both the desirable interfacial property between microsilicon and sulfide electrolytes and the distinctive chemomechanical behavior of the lithium-silicon alloy.

Silicon (Si), which has a specific capacity exceeding 3500 mAh g^{-1}, has emerged as a promising alternative to graphite-based anodes (with a specific capacity of ~370 mAh g^{-1}) to increase the energy densities of lithium-ion batteries (LIBs), for various energy-storage applications such as electric vehicles and portable devices (*1*, *2*). In addition to being the second most abundant element in Earth's crust, Si is also environmentally benign and exhibits electrochemical potentials close to that of graphite (0.3 V versus Li/Li^+) (*2*). However, commercialization of Si anodes is hindered by its poor cycling and shelf life resulting from continuous solid electrolyte interphase (SEI) growth between the highly reactive Li-Si alloy and organic liquid electrolytes used in LIBs. These deficiencies are exacerbated by the large volumetric expansion (>300%) of Si during lithiation and by the loss of Li^+ inventory due to SEI growth and irreversibly trapped Li-Si alloy enclosed within (*3*).

Current efforts to mitigate capacity fade include the use of sophisticated Si nanostructures in combination with carbon composites and robust binder matrix to mitigate pulverization (*1*, *3*–*6*). Liquid electrolyte modifications, including the use of cyclic ethers, fluorinated additives, or other ionic liquids additives that stabilize the SEI, have also been explored (*5*, *7*). Improvements have been reported in numerous half-cell studies, but the varied amounts of lithium excess make it challenging to evaluate the effectiveness of each strategy. Among reports that demonstrated stable cycling in full cells (*5*, *8*–*20*), most are limited to 100 cycles, apart from a few that demonstrated longer cycle life using various prelithiation strategies to compensate for Li^+ inventory losses (*9*, *14*, *18*). Although prelithiation can be effective to extend cycle life, the ideal Si anode should be composed of pristine microsilicon (μSi) particles that do not require further treatment, reaping the benefits of low costs, ambient air stability, and environmentally benign properties. To realize this potential, two key challenges should be addressed: (i) stabilizing the Li-Si | electrolyte interface to prevent continuous SEI growth and trapped Li-Si accumulation, and (ii) mitigating growth of new interfaces induced by volume expansion that results in Li^+ consumption.

Si stability problems arise mainly from the interface with liquid electrolytes. The use of solid-state electrolytes (SSEs) in an all solid-state battery (ASSB) is a promising alternative approach, owing to its ability to form a stable and passivating SEI (*21*). Previous studies have reported the use of thin (submicrometer)–film type Si in ASSBs (*11*, *22*, *23*). Most ASSB reports have focused on the use of metallic Li, in an effort to maximize cell energy densities (*24*, *25*). However, small critical current densities of metallic Li anodes often dictate the need for operation under elevated temperatures, especially during cell charging. In this work, a μSi electrode consisting of 99.9 wt % μSi is used in μSi||SSE||lithium nickel cobalt manganese oxide (NCM811) cells to overcome both the interfacial stability challenges of μSi and the current density limitations of ASSBs.

Unlike conventional liquid-cell architectures, the SSE does not permeate through the porous μSi electrode (Fig. 1), and the interfacial contact area between the SSE and the μSi electrode is reduced to a two-dimensional (2D) plane. After lithiation of μSi, the 2D plane is retained despite volume expansion, preventing the generation of new interfaces. Bulk μSi exhibits an electronic conductivity of ~3×10^{-5} S cm^{-1} (fig. S1), comparable to that of most common cathode materials (~10^{-6} to

[1]Department of NanoEngineering, University of California San Diego, La Jolla, CA 92093, USA. [2]LG Energy Solution, Ltd., LG Science Park, Magokjungang 10-ro, Gangseo-gu, Seoul 07796, Korea. [3]Program of Chemical Engineering, University of California San Diego, La Jolla, CA 92093, USA. [4]Sustainable Power and Energy Center (SPEC), University of California San Diego, La Jolla, CA 92093, USA. [5]Program of Materials Science and Engineering, University of California San Diego, La Jolla, CA 92093, USA.
*Corresponding author. Email: shmeng@ucsd.edu (Y.S.M.); zhengchen@eng.ucsd.edu (Z.C.)

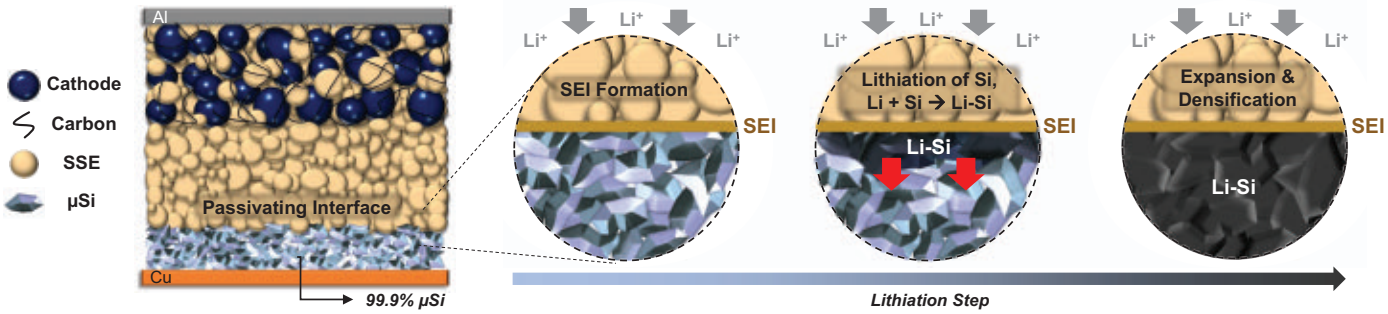


Fig. 1. Schematic of 99.9 wt % μSi electrode in an ASSB full cell. During lithiation, a passivating SEI is formed between the μSi and the SSE, followed by lithiation of μSi particles near the interface. The highly reactive Li-Si then reacts with Si particles within its vicinity. The reaction propagates throughout the electrode, forming a densified Li-Si layer.

10^{-4} S cm^{-1}) (*26*–*28*), so additional carbon additives are not necessary. Moreover, carbon is well known to be detrimental to the stability of sulfide SSEs, as it promotes SSE decomposition (*29*, *30*). Although certain types of carbon have been found to be compatible with anode-free ASSBs (*24*, *31*), eliminating carbon entirely in the μSi ASSB system is preferable. During lithiation of μSi, Li-Si formation can propagate throughout the electrode, benefiting from the direct ionic and electronic contact between Li-Si and μSi particles (Fig. 1). This process was found to be highly reversible without the need for any lithium excess. Separate full-cell experiments were performed, achieving current densities up to 5 mA cm^{-2}, operation between −20° and 80°C, and areal capacities of up to 11 mAh cm^{-2} (2890 mAh g^{-1}) for the different cells. Subsequent cycling of the μSi-NCM811 full cell at 5 mA cm^{-2} was found to deliver a capacity retention of 80% after 500 cycles, demonstrating the overall robustness of μSi enabled by ASSBs.

Results

Interface characterization

To demonstrate the importance of eliminating carbon in the anode, as well as the passivating nature of the Si-SSE interface, we characterized and quantified the SEI products from SSE decomposition with and without the presence of carbon additives. As most literature reports adopt Si composites containing between 20 and 40 wt % carbon additives (*5*, *8*–*20*), this was used as a basis for comparison against carbon-free μSi. Although Li metal is typically used as the counterelectrode in liquid electrolyte studies, its low critical current density in ASSBs make it unsuitable for studying our system (*32*, *33*). Likewise, limited kinetics in lithium indium alloys also make them inappropriate (fig. S2). Thus, NCM811, which is protected by a boron-based coating (*34*), was instead chosen as the counterelectrode, allowing direct evaluation of μSi in a full cell configuration without lithium excess. To prepare the samples, we used two composite Si-SSE anodes (with and without 20 wt % carbon additives). Composite anodes with SSEs were used to exaggerate any interfacial reactions for characterization. Figure 2A shows the voltage profiles of both cells during the first lithiation. The cell without carbon shows an initial voltage plateau around 3.5 V, typical of a μSi||NCM811 full cell. However, the cell with 20 wt % carbon shows a stark difference, with a lower initial plateau at 2.5 V, indicating extensive SSE electrochemical decomposition before reaching the lithiation potential above 3.5 V. Figure 2B compares the diffraction patterns of the pristine Si-SSE, lithiated Si-SSE, and lithiated Si-SSE-carbon composites. The lithiated Si-SSE sample retained the crystalline

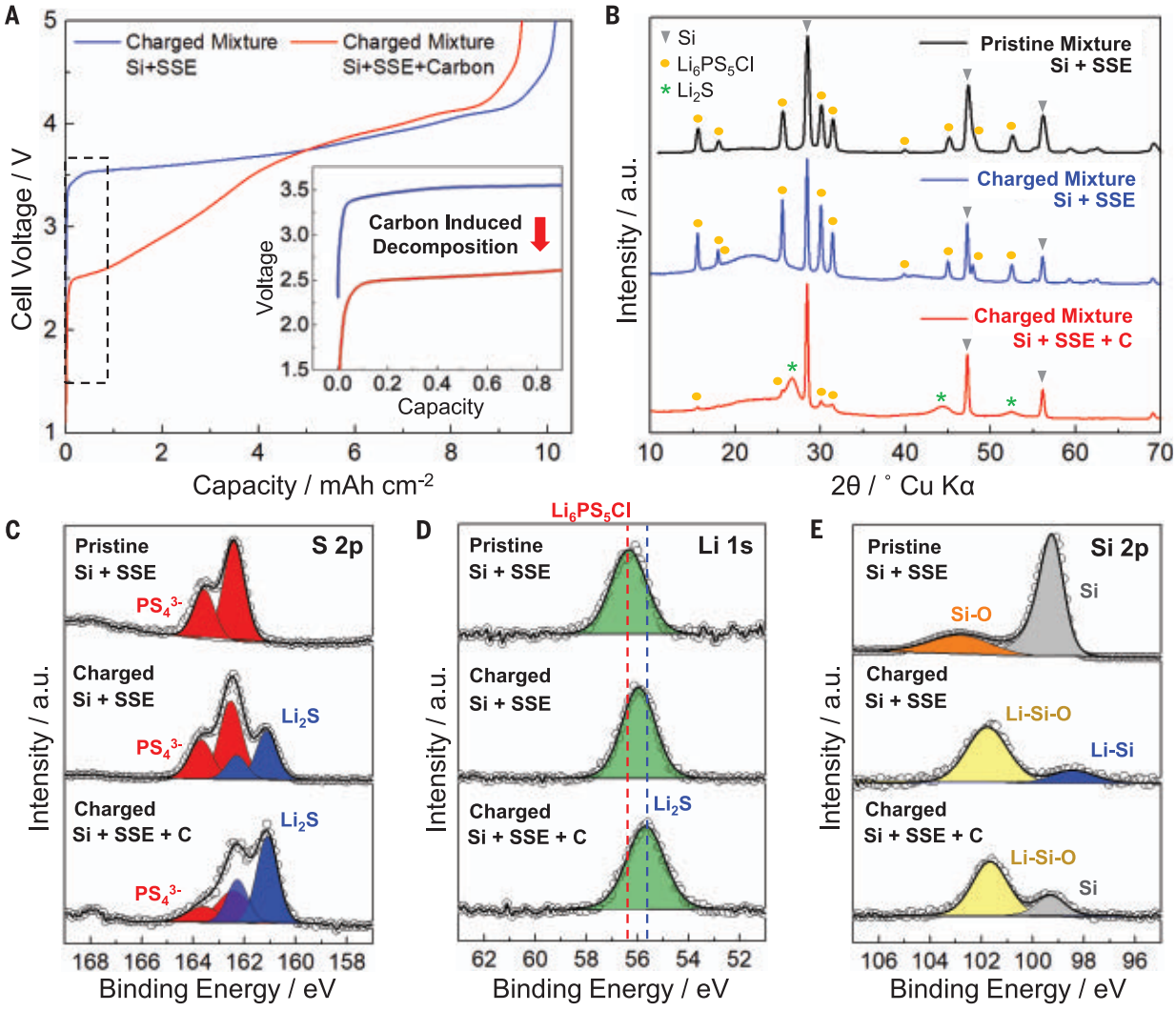


Fig. 2. Carbon effects on SSE decomposition. (**A**) Voltage profiles of μSi || SSE || NCM811 cells with and without carbon additives (20 wt %). Inset shows a lower initial plateau indicating SSE decomposition. (**B**) XRD patterns, and (**C** to **E**) XPS spectra of the (C) S 2p, (D) Li 1s, and (E) Si 2p core regions.

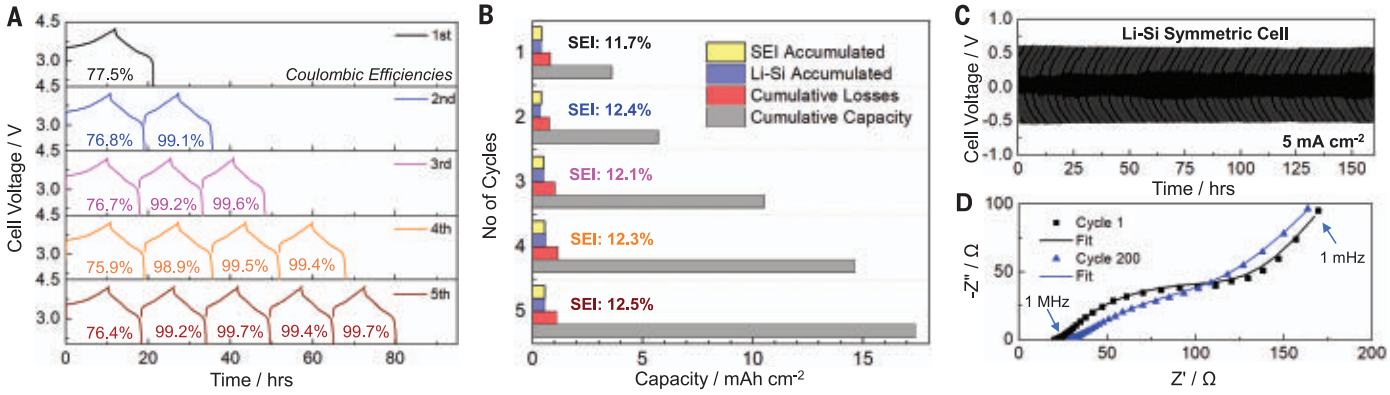


Fig. 3. Quantifying effects of SEI growth. (**A**) Voltage profiles of full cells used in titration gas chromatography. (**B**) Li-Si and SEI amounts relative to cell capacity. (**C**) Voltage profile of a Li-Si symmetric cell used for EIS, and (**D**) Nyquist plots.

structure of the SSE as well as that of the unreacted Si, with some signals of amorphous Li-Si manifesting as a hump at around 20°. Although some SEI is expected, the low amount formed at the interface is most likely not detectable with this bulk technique. However, in the cell where carbon is used, most of the pristine SSE's diffraction signals are no longer present, indicating severe electrolyte decomposition. During this process, nanocrystalline Li_2S forms as a major decomposition product and is observed as broad peaks appearing at 2θ angles of around 26°, 45°, and 52°.

These observations agree with the x-ray photoelectron spectroscopy (XPS) analysis in Fig. 2C, where the presence of carbon results in a greater extent of SSE decomposition, as seen by the formation of Li_2S (161 eV) in the S 2p region. Consequently, a larger decrease in peak intensities for the PS_4^{3-} thiophosphate unit signals is observed for the electrode containing carbon (Fig. 2C, bottom) compared to the electrode without carbon (Fig. 2C middle). Although the Li 1s region (Fig. 2D) is difficult to deconvolute owing to the presence of multiple Li^+ species, a shift toward lower binding energies is observed as a result of reduction of Li^+ from the pristine SSE. Although a smaller shift is observed in the sample without carbon (Fig. 2D middle), the peak position for the sample containing carbon is dominated by the Li_2S signals at about 55.6 eV (Fig. 2D bottom), reaffirming the previous observation using x-ray diffraction (XRD). In the Si 2p region, a native oxide layer is detected near the surface of the Si particles (Fig. 2E, top). Upon lithiation, this signal shifts to a lower binding energy. A peak with a binding energy consistent with that of Li-Si is found in the sample without carbon, whereas Si appears to remain unreacted in the sample with carbon. This is likely due to formation of the Li^+ consuming SEI products, severely limiting the lithiation of the μSi electrode itself. The vast disparities in the extent of SEI formation highlight the critical role of carbon in promoting decomposition as a result of its high specific area and electronic conductivity compared with Si. Previous studies also showed that the extent of SSE decomposition was highly dependent on the type of carbon material used, with the least decomposition observed when no carbon was used (*31*). For reduction to occur between solid-solid interfaces, sufficient contact with electronically conductive surfaces is needed. As such, carbon-free electrodes would substantially reduce SSE decomposition, resulting in improved first-cycle Coulombic efficiency (CE%) and rate capability compared with conventional carbon-containing electrodes (fig. S3, A and B).

Quantification of SEI components

Li_6PS_5Cl is reduced to form Li_2S, Li_3P, and LiCl (*29*), products that are highly passivating in nature (table S1). As such, SEI formation is expected to stabilize after the first cycle. Although capacity losses during cycling can be detected as a function of CE%, it is difficult to accurately deconvolute contributions from the SEI or trapped Li-Si, respectively. Titration gas chromatography (TGC) has been effectively used to quantify SEI and dead Li growth in Li metal batteries (*35*). In this study, TGC is similarly applied to quantify SEI growth and ascertain its passivating and stable nature. Five μSi||SSE||NCM811 full cells were assembled and cycled from one to five cycles, respectively (Fig. 3A). First-cycle CE% of ~76% was measured across all cells, which quickly rose to >99% from the second cycle. After cycling, the TGC method detailed in fig. S4 was applied to all five cells, where the difference between CE% losses and active Li^+ allows derivation of the SEI formed. The amounts of SEI accumulated, active Li^+ from Li-Si, sum of cumulative losses, and total cumulative capacities are plotted in Fig. 3B and summarized with error values in table S2. After the first cycle, the total amount of SEI formed was found to be 11.7% of the cell's capacity, and this amount increased slightly to 12.4% in the second cycle. In the subsequent cycles, both the accumulated SEI and the active Li^+ were found to remain stable and relatively unchanged, indicating interface passivation that prevents unwanted continuous reaction between Li-Si and the electrolyte. To evaluate the SEI stability over extended cycling, we fabricated and cycled a Li-Si symmetric cell at 5 mA cm^{-2} using a capacity of 2 mAh cm^{-2} per cycle (Fig. 3C). Electrochemical impedance spectroscopy (EIS) measurements found that the impedance remained stable over 200 cycles (Fig. 3D), confirming that the SEI is passivating in nature. Enlarged voltage profiles are shown in fig. S5. Conversely, impedance growth was observed in the NCM811 symmetric cell under similar conditions (fig. S5). Table S3 details the equivalent circuits' fitted values. The stable SEI formed at the Li-Si||SSE interface may also address self-discharge challenges faced by Si anodes, especially under increased temperature where unwanted parasitic reactions are aggravated. To illustrate this, we assembled and cycled solid and liquid electrolyte-based cells for five cycles at room temperature and at 55°C (fig. S6). Compared to the liquid cell, the solid cell was able to retain its charge (~99%), even at 55°C. Subsequent self-discharge tests, with 0 to 3 days of rest time at the charged state, also showed reduced loss of charge (3% versus 28%) in the solid cell, with limited voltage decay during resting (fig. S7).

Morphological evolution

As illustrated in Fig. 1, the μSi particles in the SSE cell remain in direct ionic (Li^+) and electronic (e^-) contact with each other. This allows for fast diffusion of Li^+ and transport of e^- throughout the electrode, unhindered by any electronically insulative components such as SEI or electrolyte. Galvanostatic intermittent titration technique experiments were performed, and average Li^+ diffusion coefficients

in the range of ~10^{-9} cm^2 s^{-1} were obtained (fig. S8). These are one to two orders higher than values reported on nano-Si or Si thin films in conventional liquid electrolytes (*36*–*38*), which have been limited by the interfacial stability challenges and extra impedance caused by thick SEI. To visualize the morphological evolution of Li-Si, we prepared cross-section scanning electron microscopy (SEM) images of three separate μSi electrodes by focused ion beam at the pristine, lithiated, and delithiated states.

At the pristine state (Fig. 4A), discreet μSi particles (2 to 5 μm) were observed, with an electrode porosity of 40% after calendering. After lithiation (Fig. 4B), the electrode becomes densified, with most pores disappearing between the pristine μSi particles. Moreover, the boundaries between separate μSi particles have entirely vanished. An enlarged view of the more porous region shows that the entire electrode has become an interconnected densified Li-Si alloy. After delithiation (Fig. 4C), the μSi electrode did not revert to its original discreet microparticle structure but instead formed large particles with voids between them. It is noted that a lower loading was used to image the entire void's depth at the delithiated state. Energy-dispersive x-ray (EDS) imaging confirms that the pores are indeed voids, with no evidence of SEI or SSE present between each delithiated particle. The morphological behavior observed is in stark contrast to morphological changes of μSi particles in liquid electrolyte systems (fig. S9, A and B), where lithiated μSi particles do not merge and remain separate as a result of SEI formed throughout the electrode. This chemomechanical behavior of the Li-Si alloy has been previously reported in studies on porous Si thin-film ASSBs using sulfide SSEs as well, where initial porosity incorporated into the pristine Si thin-film electrode was found to decrease during lithiation (*11*, *23*).

To quantify thickness growth and porosity changes during cycling, we prepared μSi electrodes with similar mass loadings of ~3.8 mg cm^{-2} and measured their thicknesses during charge and discharge states. At the pristine state, a thickness of ~27 μm was measured (fig. S10A), and after lithiation to $Li_{3.35}Si$, the thickness increased to ~55 μm (fig. S10B). This increase falls short of the expected >300% growth (*1*), indicating that a substantial decrease in initial 40% porosity must occur. To rationalize this, we calculated expected thicknesses versus porosities in table S4, which shows a low resulting porosity (<10%) of the ~55-μm μSi electrode after lithiation. This agrees with the qualitative observations made in Fig. 4, where considerable densification is observed compared to the pristine state. After delithiation (fig. S10C), a thickness of ~40 μm was measured, with a porosity of

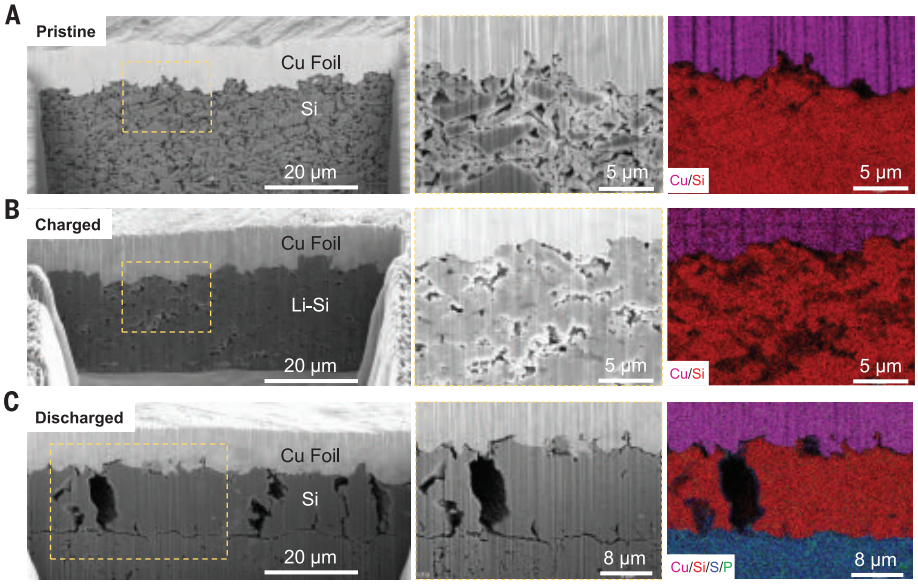


Fig. 4. Visualizing lithiation and delithiation of 99.9 wt % Si. (**A**) Pristine porous microstructure of μSi electrode. (**B**) Charged state with densified interconnected Li-Si structure. (**C**) Discharged state with void formation between large dense Si particles. Yellow dotted box represents enlarged porous regions of interest for each sample.

~30% calculated. The lower porosity at the delithiated state compared to the pristine 40% is expected, as some Li^+ remains in the anode (Fig. 3B). Despite the relatively large thickness and porosity changes, similar morphologies and thicknesses were observed after multiple cycles (fig. S11). This suggests that the mechanical properties of the Li-Si and SSE have a crucial role in maintaining the integrity of the interfaces as well as retaining contact with the anode along the 2D interfacial plane.

Although contact losses are less likely during lithiation, where volume expansion occurs, it is an important consideration during delithiation. However, good contact is still maintained between the SSE layer and the porous structure of the delithiated Li-Si (Fig. 4C). This indicates that some degree of Li-Si deformation occurred during cell cycling under a uniaxial applied stack pressure of 50 MPa with a homemade pressure rig (*39*). Although pristine μSi did not deform under calendering pressures of 370 MPa, existing reports found that hardness of Li-Si alloys decreases substantially as a function of lithiation, with values reaching as high as 10.0 to 11.6 GPa for pristine Si (*40*, *41*), to as low as 1.3 to 1.5 GPa for fully lithiated $Li_{3.75}Si$ (*40*, *42*). This trend agrees with our previous observations, where lithiated Li-Si with lower hardness could undergo sufficient deformation to form a dense alloy with low porosity (Fig. 4B), whereas delithiated Li-Si with higher hardness could not be fully deformed, evident from the large interparticle voids observed (Fig. 4C). Although the stack pressure of 50 MPa applied in this study is lower than the range of Li-Si hardness values reported in the literature, absolute conclusions on its deformability cannot be drawn from the hardness values alone. As an example, sulfide glasses, such as Li_2S-P_2S_5, were previously found to exhibit a hardness value of 1.9 GPa, yet they can be readily deformed at compaction pressures of 360 MPa (*43*). Likewise, crystalline Li_6PS_5Cl, which is expected to exhibit higher hardness than glassy Li_2S-P_2S_5, can also be deformed into a pellet for cell cycling at 50 MPa (*44*). Additionally, most mechanical studies conducted on Li-Si alloys have so far been limited to nano-indentation–based experiments on Si thin films, which can exhibit appreciably different mechanical properties compared with the μSi used in this work. As such, it is plausible for the lithiated μSi to undergo some degree of plastic deformation, especially at higher states of lithiation. To verify this, we lithiated a 99.9 wt % μSi electrode up to 0.01 V versus Li/Li^+ without pressure in a liquid electrolyte–based coin cell. Subsequently, we imaged the Li-Si electrode before and after pressing at 50 MPa. We observed that some degree of deformation in the lithiated Li-Si alloy was induced, along with a reduction in electrode porosity (fig. S9C). This deformation is necessary to maintain a good contact with the SSE, enabling high reversibility.

Electrochemical performance

To test the 99.9 wt % μSi in full cells, we prepared NCM811 cathode composites using a dry electrode process (fig. S12), with polytetrafluoroethylene (PTFE) as a binder to achieve thick electrodes. These electrodes were characterized by x-ray computed tomography (fig.

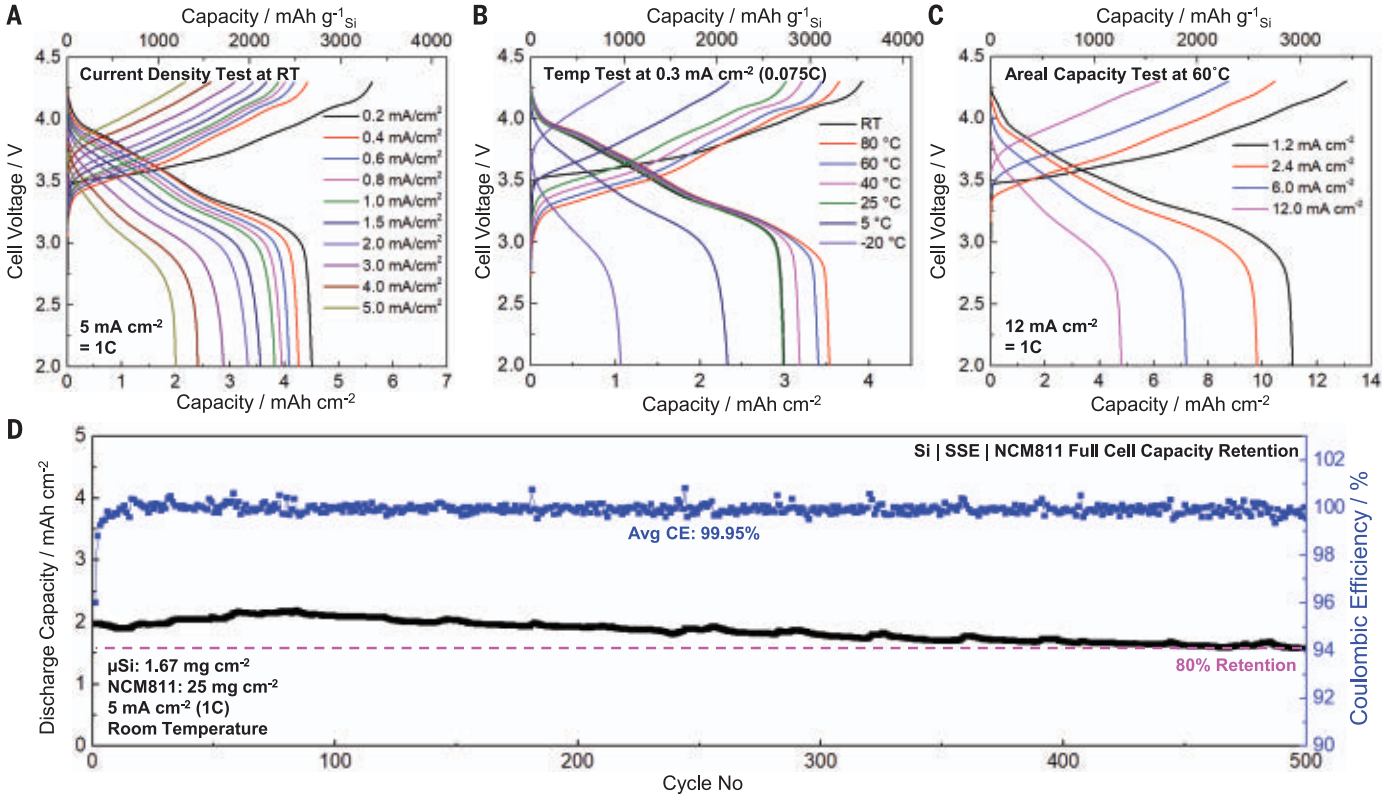


Fig. 5. μSi||SSE||NCM811 performance. (**A**) High current density test. (**B**) Wide temperature range test. (**C**) High areal capacity test. (**D**) Cycle life at room temperature. All cells were tested under similar charge and discharge conditions between 2.0 and 4.3 V. The first cycle voltage profile of each respective cell is plotted in black.

S13) and found to achieve improved electrode homogeneity and higher packing density compared with slurry casted electrodes. The dry processed cathode composites were then paired against the μSi anodes with an N/P ratio of 1.1 in full cells. Although higher N/P ratios reduce the likelihood of Li plating and cell shorting, they were found to deliver lower average CE% (fig. S14), likely as a result of higher states of charge reached by the NCM811 cathode for the same voltage cutoff, inducing undesirable impedance growth and contact losses.

Figure 5A shows the room-temperature galvanostatic cycling, where current is gradually increased from 0.2 mA cm^{-2} at the 1st cycle to 5 mA cm^{-2} at the 10th cycle. No evidence of a cell shorting occurs up to 5 mA cm^{-2} at room temperature, which is substantially higher than the typical room-temperature critical current density of Li metal ASSBs (*32*, *33*). In Fig. 5B, another full cell was cycled with a temperature range between −20° and 80°C using a moderate current density of 0.3 mA cm^{-2}. A lower current density is used to overcome the high bulk impedance of the SSE at low temperatures (*45*). Nonetheless, the cell does not exhibit shorting behavior despite charging at −20°C. To evaluate high areal loadings of μSi, we fabricated a full cell with a cathode sized to 12 mAh cm^{-2} (fig. S15). To overcome bulk impedance of the thick cathode electrode, we operated the full cell at 60°C to enhance Li^+ diffusion kinetics (fig. S16), a temperature commonly used in ASSB reports (*24*). By cycling at 1.2 mA cm^{-2}, the μSi anode was found to deliver reversible capacities of more than 11 mAh cm^{-2} (>2890 mAh g^{-1}) (Fig. 5C). Under continuous cycling at 12 mA cm^{-2}, the μSi anode delivers stable reversible capacity of more than 5 mAh cm^{-2} (>1250 mAh g^{-1}) (fig. S12). As room-temperature charge and discharge remains the ideal condition for ASSB operation, cycle life of the full cell was evaluated by maintaining a current density of 5 mA cm^{-2} at room temperature (Fig. 5D). The cell was found to achieve a capacity retention of 80% after 500 cycles and an average CE% of >99.9%. This capacity fade likely occurs as a result of cathode-SSE contact losses and cathode impedance growth (fig. S5). As the cell's capacity utilization at 5 mA cm^{-2} is arguably low, lower current cycling was also performed (fig. S17), where average CE% was found to also reach ~99.9%. Additionally, cell cycling at lower stack pressures is also presented in fig. S18. Battery performance described here still represents single-layered pelletized cells with thick SSE separators. The μSi anode's potential for high energy densities (>900 Wh $liter^{-1}$) can only be realized if SSE thickness can be reduced, in combination with multilayer form factors (table S5). Nonetheless, the electrochemical results shown above reaffirm the effectiveness of sulfide SSEs in enabling 99.9 wt % μSi anodes, capable of operating with high current densities, over a wide temperature range, and using high areal loadings, as well as achieving a long cycle and calender life. Overall, this approach presents substantial advantages to advance both the silicon anode and ASSB community, offering a pathway to address some of the fundamental interfacial and performance challenges of μSi anodes.

REFERENCES AND NOTES

1. Y. Jin, B. Zhu, Z. Lu, N. Liu, J. Zhu, *Adv. Energy Mater.* **7**, 1700715 (2017).
2. A. Franco Gonzalez, N.-H. Yang, R.-S. Liu, *J. Phys. Chem. C* **121**, 27775–27787 (2017).
3. Y. Yang *et al.*, *Sustain. Energy Fuels* **4**, 1577–1594 (2020).
4. C. Wang *et al.*, *Nat. Chem.* **5**, 1042–1048 (2013).
5. J. Chen *et al.*, *Nat. Energy* **5**, 386–397 (2020).
6. P. Parikh *et al.*, *Chem. Mater.* **31**, 2535–2544 (2019).
7. H. Shobukawa, J. Shin, J. Alvarado, C. S. Rustomji, Y. S. Meng, *J. Mater. Chem. A Mater. Energy Sustain.* **4**, 15117–15125 (2016).
8. H. J. Kwon *et al.*, *Nano Lett.* **20**, 625–635 (2020).
9. Q. Ma *et al.*, *ACS Appl. Energy Mater.* **3**, 268–278 (2019).
10. A. Baasner *et al.*, *J. Electrochem. Soc.* **167**, 020516 (2020).
11. S. Cangaz *et al.*, *Adv. Energy Mater.* **10**, 2001320 (2020).
12. G. Huang *et al.*, *ACS Nano* **14**, 4374–4382 (2020).
13. Y. Wang *et al.*, *Sci. Rep.* **10**, 3208 (2020).
14. W. An *et al.*, *Nat. Commun.* **10**, 1447 (2019).
15. Y. Li *et al.*, *Nat. Energy* **1**, 15029 (2016).
16. S. Choi, T. W. Kwon, A. Coskun, J. W. Choi, *Science* **357**, 279–283 (2017).

17. Z. Xu *et al.*, *Joule* **2**, 950–961 (2018).
18. X. Gao, F. Wang, C. Sam, S. Gollon, C. Yuan, *J. Electrochem. Energy Convers. Storage* **16**, 011009 (2019).
19. J.-G. Ren *et al.*, *Energy Technol. (Weinheim)* **1**, 77–84 (2013).
20. J.-K. Kim *et al.*, *Solid State Ion.* **268**, 256–260 (2014).
21. A. Banerjee, X. Wang, C. Fang, E. A. Wu, Y. S. Meng, *Chem. Rev.* **120**, 6878–6933 (2020).
22. W. Ping *et al.*, *Energy Storage Mater.* **21**, 246–252 (2019).
23. J. Sakabe, N. Ohta, T. Ohnishi, K. Mitsuishi, K. Takada, *Commun. Chem.* **1**, 24 (2018).
24. Y.-G. Lee *et al.*, *Nat. Energy* **5**, 299–308 (2020).
25. R. Xu *et al.*, *ACS Energy Lett.* **4**, 1073–1079 (2019).
26. S. Burkhardt *et al.*, *ACS Energy Lett.* **4**, 2117–2123 (2019).
27. X. Y. Qiu *et al.*, *Phys. Chem. Chem. Phys.* **14**, 2617–2630 (2012).
28. Z. Chen *et al.*, *Adv. Energy Mater.* **5**, 1401826 (2015).
29. D. H. S. Tan *et al.*, *ACS Energy Lett.* **4**, 2418–2427 (2019).
30. T. K. A. Schwietert *et al.*, *Nat. Mater.* **19**, 428–435 (2020).
31. W. Zhang *et al.*, *ACS Appl. Mater. Interfaces* **9**, 35888–35896 (2017).
32. S. Randau *et al.*, *Nat. Energy* **5**, 259–270 (2020).
33. J. Kasemchainan *et al.*, *Nat. Mater.* **18**, 1105–1111 (2019).
34. E. A. Wu *et al.*, *J. Electrochem. Soc.* **167**, 130516 (2020).
35. C. Fang *et al.*, *Nature* **572**, 511–515 (2019).
36. K. Pan, F. Zou, M. Canova, Y. Zhu, J.-H. Kim, *J. Power Sources* **413**, 20–28 (2019).
37. N. Ding *et al.*, *Solid State Ion.* **180**, 222–225 (2009).
38. K. Yoshimura, J. Suzuki, K. Sekine, T. Takamura, *J. Power Sources* **174**, 653–657 (2007).
39. J. M. Doux *et al.*, *Adv. Energy Mater.* **10**, 1903253 (2019).
40. L. A. Berla, S. W. Lee, Y. Cui, W. D. Nix, *J. Power Sources* **273**, 41–51 (2015).
41. V. Kulikovsky *et al.*, *Thin Solid Films* **516**, 5368–5375 (2008).
42. B. Hertzberg, J. Benson, G. Yushin, *Electrochem. Commun.* **13**, 818–821 (2011).
43. F. P. McGrogan *et al.*, *Adv. Energy Mater.* **7**, 1602011 (2017).
44. J.-M. Doux *et al.*, *J. Mater. Chem. A Mater. Energy Sustain.* **8**, 5049–5055 (2020).
45. C. Yu *et al.*, *ACS Appl. Mater. Interfaces* **10**, 33296–33306 (2018).

ACKNOWLEDGMENTS

We acknowledge the UCSD Crystallography Facility. This work was performed in part at the San Diego Nanotechnology Infrastructure (SDNI) of UCSD, a member of the National Nanotechnology Coordinated Infrastructure, which is supported by the National Science Foundation (grant ECCS-1542148). TGC measurement is made possible by the support of the Office of Vehicle Technologies of the US Department of Energy through Battery500 Consortium (grant DE-EE0007764). Characterization work was performed in part at the UC Irvine Materials Research Institute (IMRI) using instrumentation funded in part by the National Science Foundation Major Research Instrumentation Program under grant no. CHE-1338173. Y.S.M. thanks M. H. Kim and S. Bang from LG Energy Solutions for insightful discussions. **Funding:** This study was financially supported by the LG Energy Solution through the Battery Innovation Contest (BIC) program. Z.C. acknowledges funding from the start-up fund support from the Jacob School of Engineering at University of California San Diego. Y.S.M. acknowledges funding support from the Zable Endowed Chair Fund. **Author contributions:** D.H.S.T. and Y.S.M. conceived the ideas for the study. D.H.S.T. and H.Y. designed the experiments and cell configuration. Y.-T.C., W.B., B.S., W.L., B.L., S.-Y.H., B.S., and J.S., performed the XRD, XPS, TGC, FIB-SEM, and CT experiments. Z.C., J.-M.D., E.A.W., G.D., H.E.H., H.J.H., H.J., and J.B.L. participated in the scientific discussion and data analysis. D.H.S.T. wrote the manuscript. All authors discussed the results and commented on the manuscript. All authors have approved the final manuscript. **Competing interests:** A joint patent application on this work has been filed (US 63/157,012) between UC San Diego's Office of Innovation and Commercialization as well as LG Energy Solution, Ltd, and licensed by UNIGRID L.L.C. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abg7217
Materials and Methods
Figs. S1 to S18
Tables S1 to S5

23 January 2021; resubmitted 6 March 2021
Accepted 17 August 2021
10.1126/science.abg7217

HUMAN GENOMICS

Protein-coding repeat polymorphisms strongly shape diverse human phenotypes

Ronen E. Mukamel[1,2]*†, Robert E. Handsaker[2,3,4]*†, Maxwell A. Sherman[1,2,5], Alison R. Barton[1,2,6], Yiming Zheng[2,3], Steven A. McCarroll[2,3,4]*‡, Po-Ru Loh[1,2]*‡

Many human proteins contain domains that vary in size or copy number because of variable numbers of tandem repeats (VNTRs) in protein-coding exons. However, the relationships of VNTRs to most phenotypes are unknown because of difficulties in measuring such repetitive elements. We developed methods to estimate VNTR lengths from whole-exome sequencing data and impute VNTR alleles into single-nucleotide polymorphism haplotypes. Analyzing 118 protein-altering VNTRs in 415,280 UK Biobank participants for association with 786 phenotypes identified some of the strongest associations of common variants with human phenotypes, including height, hair morphology, and biomarkers of health. Accounting for large-effect VNTRs further enabled fine-mapping of associations to many more protein-coding mutations in the same genes. These results point to cryptic effects of highly polymorphic common structural variants that have eluded molecular analyses to date.

The human genome contains thousands of variable number of tandem repeat (VNTR) polymorphisms (*1*, *2*), but the effects of these polymorphisms on human phenotypes are largely unknown. VNTRs are multiallelic variants at which a nucleotide sequence, from seven to thousands of base pairs long, is repeated several to hundreds of times, with the number of repeats varying among individuals (fig. S1). Extreme alleles of VNTRs have been implicated in diseases including progressive myoclonus epilepsy (*3*) and facioscapulohumeral muscular dystrophy (*4*). However, because most VNTRs are invisible to single-nucleotide polymorphism (SNP) arrays and difficult to measure with short-read sequencing, they have not been considered in the genotype-phenotype association studies that have been central to recent work in human genetics.

We hypothesized that exome-sequencing data might contain unknown information about VNTR lengths and that VNTR alleles might segregate on specific SNP haplotypes, enabling statistical imputation (*5*) in SNP-phenotype datasets from hundreds of thousands of people, such as participants in the UK Biobank (UKB) (*6*).

[1]Division of Genetics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [2]Program in Medical and Population Genetics, Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard University, Boston, MA, USA. [3]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard University, Boston, MA, USA. [4]Department of Genetics, Harvard Medical School, Boston, MA, USA. [5]Computer Science and Artificial Intelligence Laboratory, MIT, Boston, MA, USA. [6]Bioinformatics and Integrative Genomics Program, Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.
***Corresponding author. Email: rmukamel@broadinstitute.org (R.E.M.); handsake@broadinstitute.org (R.E.H.); mccarroll@genetics.med.harvard.edu (S.A.M.); poruloh@broadinstitute.org (P.-R.L.)**
†These authors contributed equally to this work.
‡These authors co-supervised this work.

Exploring the phenotypic effects of coding VNTRs

We identified candidate VNTRs by scanning the human reference genome for tandemly repeated sequences (*7*). For each repeat, we estimated "diploid VNTR content," the sum of maternally and paternally derived allele lengths, in 49,959 exome-sequenced UKB participants (*8*) by measuring numbers of reads that aligned to the repeated sequence (*7*). We then used surrounding SNPs to identify haplotypes likely to have been coinherited from a recent common ancestor, enabling resolution of diploid measurements into allele-specific contributions and imputation of VNTR lengths into SNP-haplotypes of 437,612 additional UKB participants. We developed statistical algorithms to perform such analysis on extended SNP haplotypes for hundreds of thousands of individuals using sibling identical-by-descent information to benchmark accuracy and to optimize analysis parameters (*7*). We focused subsequent analysis on autosomal exon–overlapping repeats in 118 genes for which these measurements exhibited *cis* heritability in sibling pairs (table S1).

We applied this approach to identify relationships between coding VNTR alleles and 786 phenotypes (table S2) in up to 415,280 unrelated UKB participants (depending on phenotype) of European ancestry. This analysis found 185 statistically significant associations (table S3). To determine whether such associations were driven by VNTR length variation rather than by other variants with which the VNTRs were in linkage disequilibrium (LD), we performed fine-mapping analyses (*9*) considering nearby genotyped and imputed variants (*6*, *10*). Because variation at most VNTRs

Table 1. VNTRs within protein-coding sequences affect diverse human phenotypes. For each of five protein-altering VNTRs involved in phenotype associations that passed stringent fine-mapping criteria, *P* values [in linear mixed-model analyses of *N* = 415,280 unrelated UKB participants of European (EUR) ancestry] and estimated effect size ranges (across the longest and shortest alleles sufficiently common to be amenable to our computational analysis) are listed for the most strongly associated phenotype.

Gene	Cytoband	Repeat unit size	Repeat count (EUR)	Protein domain (effect)	Phenotype	Effect range ± SE	*P* value
LPA	6q25.3-q26	~5.6 kb (114 aa, two exons)	2–40	KIV (number)	Lipoprotein(a) concentration	5.1 ± 0.5 SD (= 233 ± 23 nmol/liter)	$4.4 \times 10^{-(25,121)}$
ACAN	15q26.1	57 bp (19 aa)	13–44	Chondroitin sulfate (size)	Height	0.49 ± 0.04 SD (= 3.2 ± 0.3 cm)	1.7×10^{-234}
TENT5A	6q14.1	15 bp (5 aa)	2–7	Unknown (size)	Height	0.09 ± 0.01 SD (= 0.6 ± 0.1 cm)	2.5×10^{-53}
MUC1	1q22	60 bp (20 aa)	20–125	Extracellular (size)	Serum urea	0.16 ± 0.01 SD (= 0.22 ± 0.01 mmol/liter)	2.7×10^{-163}
TCHH	1q21.3	18 bp (6 aa)	5–15	α-Helix rod (size)	Male pattern baldness score	−0.063 ± 0.006 SD	1.6×10^{-55}

arises from three or more alleles, VNTR variation was only partially correlated with individual SNPs, enabling this analysis to distinguish VNTR from SNP effects.

Nineteen phenotype associations involving five distinct VNTRs (Table 1, table S3, and fig. S1) exhibited evidence [FINEMAP (*9*) posterior probability >0.95] that VNTR length variation, rather than nearby SNPs, drove genotype-phenotype associations. For these five VNTRs, we improved genotyping accuracy by incorporating additional information from within-repeat variation or spanning reads to confirm the associations [figs. S2 and S3 (*7*)].

These associations appeared to explain some of the largest known GWAS signals for human phenotypes, including height, serum urea, and hair phenotypes, with some associations exhibiting strength comparable to or exceeding that of any single SNP in the genome.

Three VNTRs within exons of *TENT5A, MUC1,* and *TCHH* had not previously been implicated at these loci; a fourth (in *ACAN*) was recently reported in parallel work (*11*). Analysis also replicated an association between the length of the KIV-2 repeat in *LPA* and lipoprotein(a) concentration (*12*) [$P = 4.4 \times 10^{-(25,121)}$, BOLT-LMM (*13*)]. All five VNTRs were genotyped and imputed accurately (root mean square error ~1 repeat unit and/or $R^2 \geq 0.7$) according to benchmarks using cross-validation (fig. S4 and table S1) and the HGSVC2 long-read sequencing dataset (figs. S5 to S9) (*7*, *14*).

Fine-mapping of *LPA* variants influencing lipoprotein(a) concentration

Complex genetics involving VNTRs and SNPs at the same locus was revealed by analyzing lipoprotein(a) concentration [Lp(a)], elevated levels of which are a major risk factor for coronary artery disease (*15*). Lp(a) is almost completely heritable, with about half of its population variance explained by a VNTR-generated size polymorphism in the second kringle-IV (KIV) domain of apo(a) (*12*). Each KIV-2 repeat unit (~5.6 kb) spans two exons of *LPA*, which together encode a 114-aa copy of this domain. Longer alleles, those with more copies of the encoded kringle repeat, are known to associate with lower Lp(a) levels (*12*, *16*), reflecting retention of longer apo(a) isoforms in the endoplasmic reticulum (*17*). In the UKB, inheritance at the *LPA* locus explained most of the variance in Lp(a) measurements [$R = 0.93$ in sibling pairs sharing both *LPA* alleles, consistent with previous work (*18*)], with KIV-2 length explaining ~61% of this variance in a nonparametric model.

To identify additional *LPA* variants that might more completely explain Lp(a) variation and to explore their interactions with KIV-2 length, we used individuals heterozygous for either of two coding variants [combined minor allele frequency (MAF) = 0.05] that create null alleles that produce undetectable serum Lp(a) (*7*). This approach created an effective haploid model for Lp(a) and made it possible to systematically identify and measure the effects of Lp(a)–altering alleles (fig. S10). We performed stepwise conditional analysis to identify *LPA* sequence variants that associated with low Lp(a) despite occurring on short- or medium-length KIV-2 alleles that typically associate with higher Lp(a) levels (*7*).

These analyses identified associations with 17 protein-altering variants, each of which appeared to greatly reduce Lp(a) ($P < 1 \times 10^{-17}$ for each variant, Fisher's exact test or linear regression, table S4); 43% of European haplotypes were affected by at least one of these variants. Six variants predicted to partially or fully abolish constitutive splice sites and six missense variants achieved the strongest associations in 12 consecutive stages of stepwise analysis; five additional rare (MAF <1%) coding variants exhibited top or near-top associations in further conditional analyses (Fig. 1A, fig. S11, and table S4). The two variants with the largest impacts on Lp(a) variation in the European population (because of their high allele frequencies; MAF = 13 and 21%) were variants within the KIV-2 region that are computationally predicted to impair splicing (*19*) of KIV-2 exon 2. One of these splice variants has been experimentally validated (*20*). These variants reduced Lp(a) by 85 and 89%, respectively, when present within a single KIV-2 repeat unit; alleles carrying either variant on multiple repeat units within the VNTR produced nearly undetectable Lp(a) (fig. S12). Fine-mapping analyses identified three other common variants (MAF = 14 to 28%), two in the 5′ untranslated region (UTR) of *LPA*, which have both been observed to regulate translational activity (*21*, *22*), and one missense variant. All three variants associated with more modest effects on Lp(a) levels across a broad range of KIV-2 alleles (Fig. 1A and table S4).

The strong effects of the VNTR and SNPs at *LPA*, the large sample size of UKB, and the ability to chromosomally phase all of these variants accurately made it possible to identify nonlinear and cis-epistatic effects at *LPA*. Accounting for the effects of the 17 implicated coding variants at *LPA* showed that the inverse relationship between KIV-2 length and Lp(a) (*12*, *17*) breaks down for very short (high-protein-level) alleles (Fig. 1A). Throughout most of the KIV-2 length range (12 to 24 repeats), each one-repeat-unit decrease in KIV-2 length resulted in a 37% increase in Lp(a) (Fig. 1A). However, this effect was attenuated for alleles with fewer than 12 repeats and appeared to invert around eight repeats ($P = 9.4 \times 10^{-31}$, linear regression; Fig. 1A and fig. S13). Accounting for the nonlinear effect of KIV-2

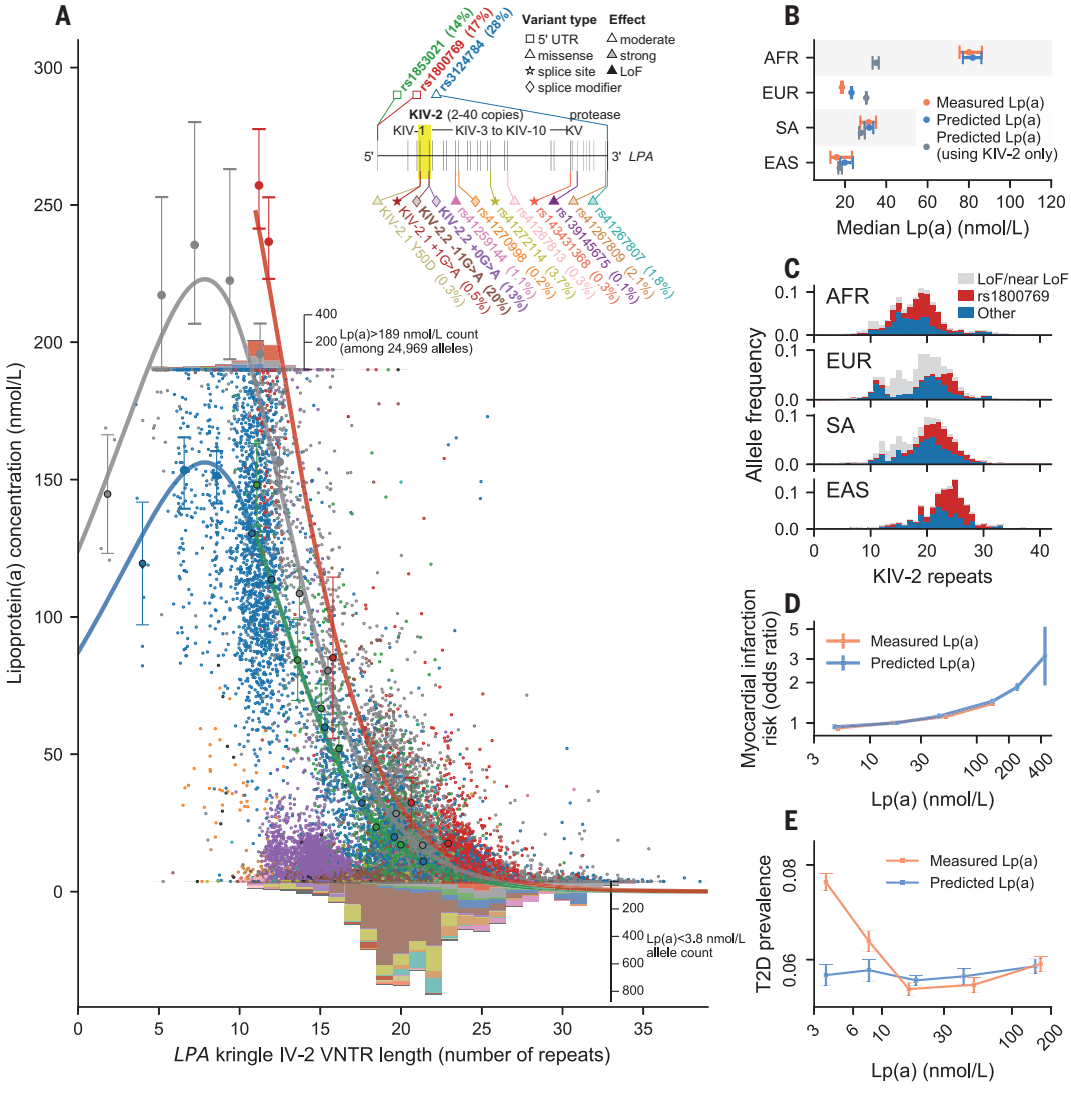


Fig. 1. Kringle IV-2 repeat length variation and 23 *LPA* SNPs together explain ~90% of lipoprotein(a) heritability. (**A**) Serum Lp(a) versus KIV-2 VNTR length in an effective-haploid model of Lp(a) involving 24,969 *LPA* alleles (in exome-sequenced UKB participants of European ancestry) for which the allele on the homologous chromosome was predicted to produce negligible Lp(a) (<4 nmol/liter). Colors indicate the 15 most common Lp(a)–modifying SNPs identified by fine-mapping analysis (full list in table S4). Curves indicate parametric fits of Lp(a) to KIV-2 length. Gray indicates alleles not carrying any Lp(a)–modifying SNPs; red, blue, and green are carriers of a single common Lp(a)–modifying SNP; large points are mean Lp(a) among such alleles in KIV-2 length bins. Error bars indicate 95% CIs. Histograms (top and bottom) show counts of Lp(a) measurements outside of the reportable range (<3.8 or >189 nmol/liter) colored by Lp(a)–modifying SNPs (*7*). (**B**) Observed and predicted median Lp(a) among individuals of African (AFR; *N* = 893), European (EUR; *N* = 42,162), South Asian (SA; *N* = 954), and East Asian (EAS; *N* = 156) ancestry. (**C**) *LPA* allele frequencies by ancestry. VNTR alleles in *cis* with a large-effect Lp(a)–reducing variant (respectively, the Lp(a)–increasing 5′ UTR variant rs1800769) are indicated in gray (respectively, red). (**D** and **E**) Myocardial infarction risk (D) and T2D prevalence (E) versus measured or genetically predicted Lp(a). Error bars indicate 95% CIs.

length and for phase-resolved *LPA* sequence variants explained 90% of the heritable variance (83% of total variance) in Lp(a) [versus ~60% of total variance in earlier work (*12*, *23*)].

Serum Lp(a) levels vary across populations (*12*), with median measurements fourfold higher among Africans than among Europeans, but the reason for this cross-population variation has been unclear. We found that this variation was largely explained by population differences in the allele frequencies of *LPA* sequence variants (Fig. 1B). Elevated Lp(a) in UKB participants of African ancestry (median 80.1 nmol/liter versus 18.5 nmol/liter in Europeans) was primarily explained by the paucity of alleles carrying variants that greatly reduced Lp(a) (~13% of African alleles versus ~43% of European alleles despite sufficient discovery power in both populations) and the higher frequency of the Lp(a)–increasing 5′ UTR variant among African alleles (MAF = 46% versus 17% in European alleles for rs1800769; Fig. 1C). These allele frequency differences also explained the apparent difference in shape of the Lp(a)–KIV-2 curve in different populations (fig. S14).

The accuracy of genetically predicted Lp(a) (R^2 = 0.83 in Europeans) enabled insights into epidemiological associations involving Lp(a). We observed that the myocardial infarction risk–increasing effect of higher Lp(a) (*15*, *24*) extends to extreme Lp(a) levels [odds ratio (OR) = 3.1, 95% confidence interval (CI) = 1.9 to 5.2 for individuals with genetically predicted Lp(a) >400 nmol/liter; Fig. 1D]. By contrast, lower genetically predicted Lp(a) did not associate with increased type 2 diabetes (T2D) risk, suggesting that the 17% (SE 1%) lower levels of Lp(a) observed in T2D patients represents reverse causation resulting from T2D itself, T2D-related liver comorbidities, or T2D medications (Fig. 1E, fig. S15, and table S5).

Human height is strongly affected by VNTRs in *ACAN* and *TENT5A*

Human height associates with hundreds of common alleles (*25*), generally with small effect sizes (<0.05 SDs). By contrast, size variation of a 57-bp (19-aa) repeat in the *ACAN* gene strongly associated with height ($P = 1.7 \times 10^{-234}$, BOLT-LMM), with an effect size differential of 0.49 SDs (SE 0.04), or 3.2 cm, between the longest and shortest European alleles (Fig. 2). This association, which appears to underlie one of the first reported genetic associations with height (*26*), was also observed in a parallel study using long-read sequencing in the deCODE cohort (*11*). Here, analysis in the larger, more diverse UKB cohort, which contains double the range of allelic variation,

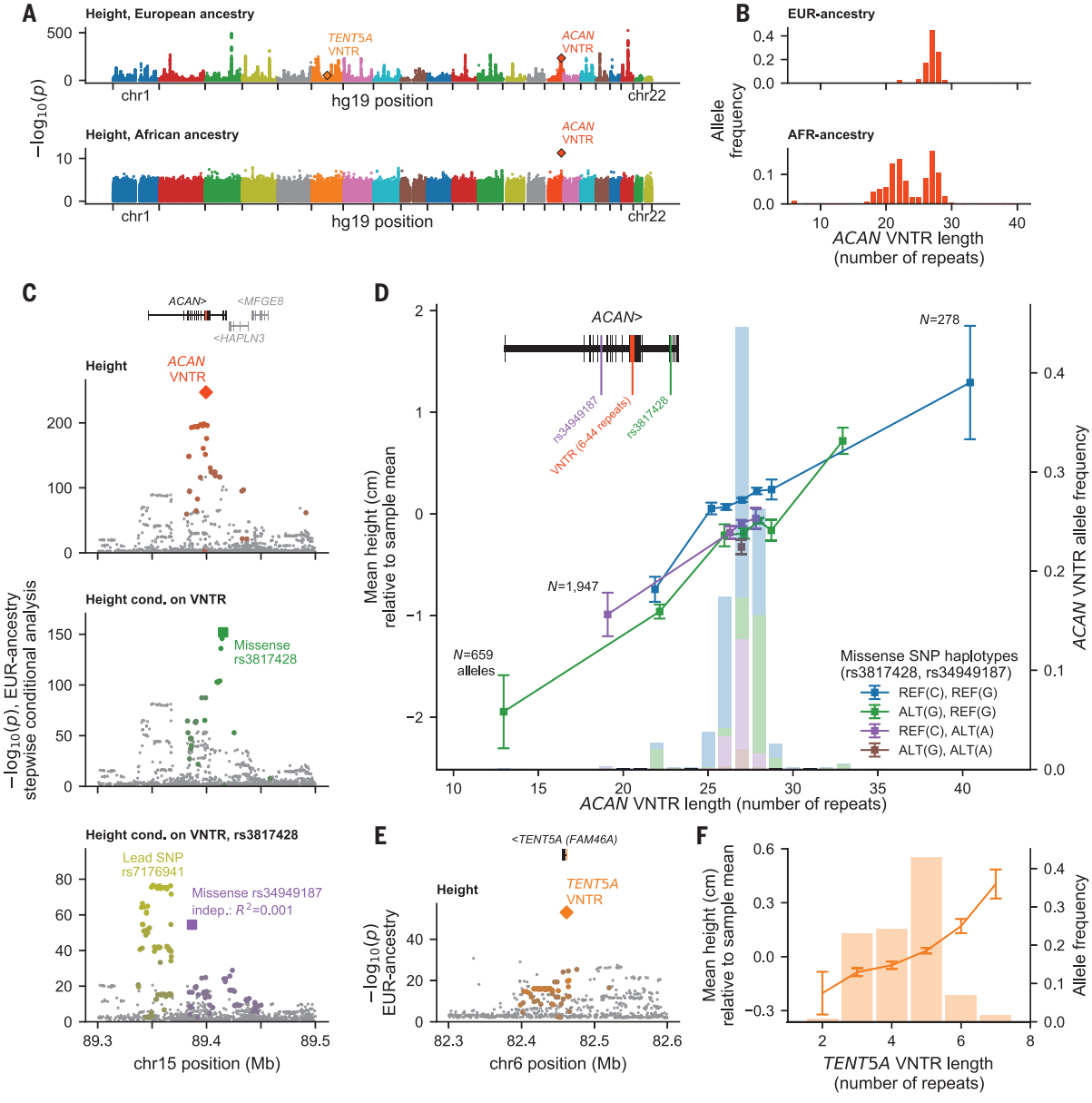


Fig. 2. Lengths of protein-coding repeat polymorphisms in *ACAN* and *TENT5A* associate with human height. (**A**) Genetic associations with height in UKB participants of EUR (top; N = 415,280) and AFR (bottom; N = 7543) ancestry. (**B**) *ACAN* VNTR allele length distributions. (**C**) Height association statistics at *ACAN* in three consecutive steps of stepwise conditional analysis (EUR N = 415,280). Large diamonds and squares indicate likely causal coding mutations; colored dots are variants in partial LD (R^2 > 0.1) with labeled variants. Height phenotypes were adjusted for genetic predictions computed using the rest of the genome (*7*). (**D**) Mean height of carriers (lines, left axis) and EUR allele frequencies (histograms, right axis) of *ACAN* alleles defined by VNTR length and missense SNP haplotype. Error bars indicate 95% CIs. Rare long alleles (40 to 42 repeats) were grouped into one bin. (**E**) Height associations at *TENT5A*. (**F**) Mean height and EUR allele frequencies for *TENT5A* VNTR alleles. Error bars indicate 95% CIs.

including a very short, six-repeat African allele and European alleles with up to ~44 repeats (Fig. 2, B and D), uncovered several additional insights.

Height exhibited an approximately linear relationship with length of the *ACAN* VNTR. Consistent increasing effects were observed across a series of at least nine distinct VNTR allele lengths, resulting in an association signal ($P = 1.7 \times 10^{-234}$, BOLT-LMM) stronger than that of any nearby variant, explaining 0.19% of height variance among European-ancestry UKB participants (Fig. 2, C and D). Moreover, among the 7543 UKB participants of African ancestry, the *ACAN* VNTR association was nearly 50% stronger than the association of any other variant in the genome ($P = 5.2 \times 10^{-12}$ for the VNTR versus $P = 1.4 \times 10^{-8}$ for the strongest SNP association) and explained a much larger 0.60% of height variance, primarily because of the greater VNTR length variation (SD = 3.7 repeats versus 1.5 repeats in Europeans; Fig. 2B). Imputation of the VNTR association into height association statistics from the African Ancestry Anthropometry Genetics Consortium cohort (*27*) replicated these results [with the VNTR explaining an estimated 0.42% of height variance; imputed $P = 5.8 \times 10^{-40}$ versus linear regression $P = 3.4 \times 10^{-20}$ for the strongest SNP association genome-wide; fig. S16 (*7*)].

Aggrecan, the protein encoded by *ACAN*, is a component of the extracellular matrix in growth

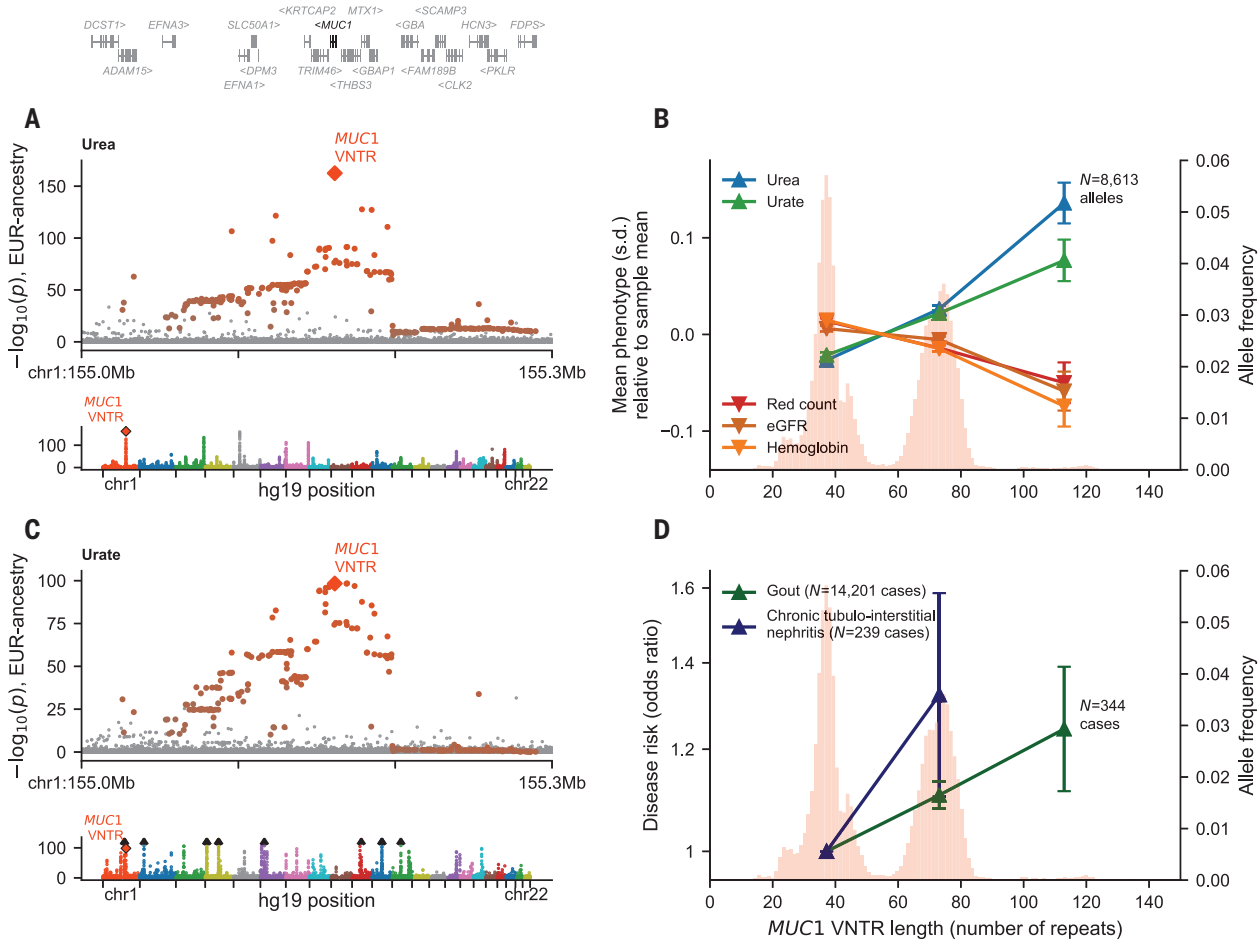


Fig. 3. *MUC1* VNTR length associates with multiple renal phenotypes. (**A** and **C**) Genetic associations with serum urea (A) and serum urate (C) at *MUC1* (top; orange dots indicate variants in LD with *MUC1* VNTR length (R^2 > 0.1) and genome-wide (bottom); *N* = 415,280 UKB EUR participants. (**B** and **D**) Mean phenotypes in carriers (B) or disease ORs (D) (lines, left axis) and allele frequencies (histograms, right axis) of *MUC1* VNTR alleles. VNTR alleles were stratified into three groups for phenotype analyses: short (<55 repeat units), long (55 to 95 repeat units), and very long (>95 repeat units). Error bars indicate 95% CIs. eGFR, estimated glomerular filtration rate.

plate cartilage and is required for normal growth plate cytoarchitecture (*28*). The VNTR generates 2.4-fold size variation in aggrecan's first chondroitin sulfate domain (CS1), in which amino acid residues are modified by long, charged polysaccharide chains that endow this extracellular matrix with key properties including the ability to hold large amounts of water (*29*).

As at *LPA*, incorporation of the *ACAN* VNTR into genetic association analysis (by stepwise conditional analysis) made it possible to identify additional genetic effects that are driven at *ACAN* by two common missense SNPs (Fig. 2C and table S6). These two missense SNPs, which affect ACAN globular domains, had two of the top three predicted deleteriousness scores (*30*) (CADD = 23.1 for rs3817428 and 27.6 for rs34949187) among common missense SNPs in *ACAN* and were corroborated by Bayesian fine-mapping (*9*) analysis (FINEMAP posterior probability of causality >0.99). A combined model including the VNTR and these SNPs explained 0.33% of height variance in Europeans.

Despite the strong effects of *ACAN* VNTR alleles on height, neither end of the allelic spectrum appeared to compromise ACAN function in any way detrimental to health. Whereas loss-of-function mutations in *ACAN* cause autosomal-dominant skeletal disorders (*31*), VNTR length variation did not associate at Bonferroni significance with any disease in the UKB ($P > 3 \times 10^{-4}$, logistic regression). A participant homozygous for the short, six-repeat allele (allele frequency = 1.2% among participants with African ancestry) had no reported musculoskeletal disease phenotypes.

A distinct coding VNTR in the *TENT5A* gene (previously named *FAM46A*) consisting of two to seven repeats of 15 bp also associated with height ($P = 2.5 \times 10^{-53}$, BOLT-LMM), with six VNTR alleles exhibiting monotonically increasing effects (Fig. 2, E and F). TENT5A, a poly(A) polymerase in which multiple coding variants have been linked to autosomal-recessive osteogenesis imperfecta (*32*), polyadenylates and increases expression in osteoblasts of the collagen genes *COL1A1* and *COL1A2* and other genes mutated in this disease (*33*).

Kidney function phenotypes shaped by a VNTR in *MUC1*

The *MUC1* gene encodes a secreted (cell surface–associated) protein (mucin 1) with cell-adhesive and anti-adhesive properties. *MUC1* harbors a VNTR that contains 20 to 125 repeats (*34*) of a 60-bp (20-aa) coding sequence that determines the length of a heavily glycosylated extracellular domain. Ultra-rare frameshift mutations within the *MUC1* VNTR cause autosomal-dominant tubulointerstitial kidney disease (*35*). In our analyses, length of the *MUC1* VNTR associated with several renal phenotypes (Fig. 3), including serum urea ($P = 2.7 \times 10^{-163}$, BOLT-LMM) and serum urate ($P = 4.7 \times 10^{-99}$,

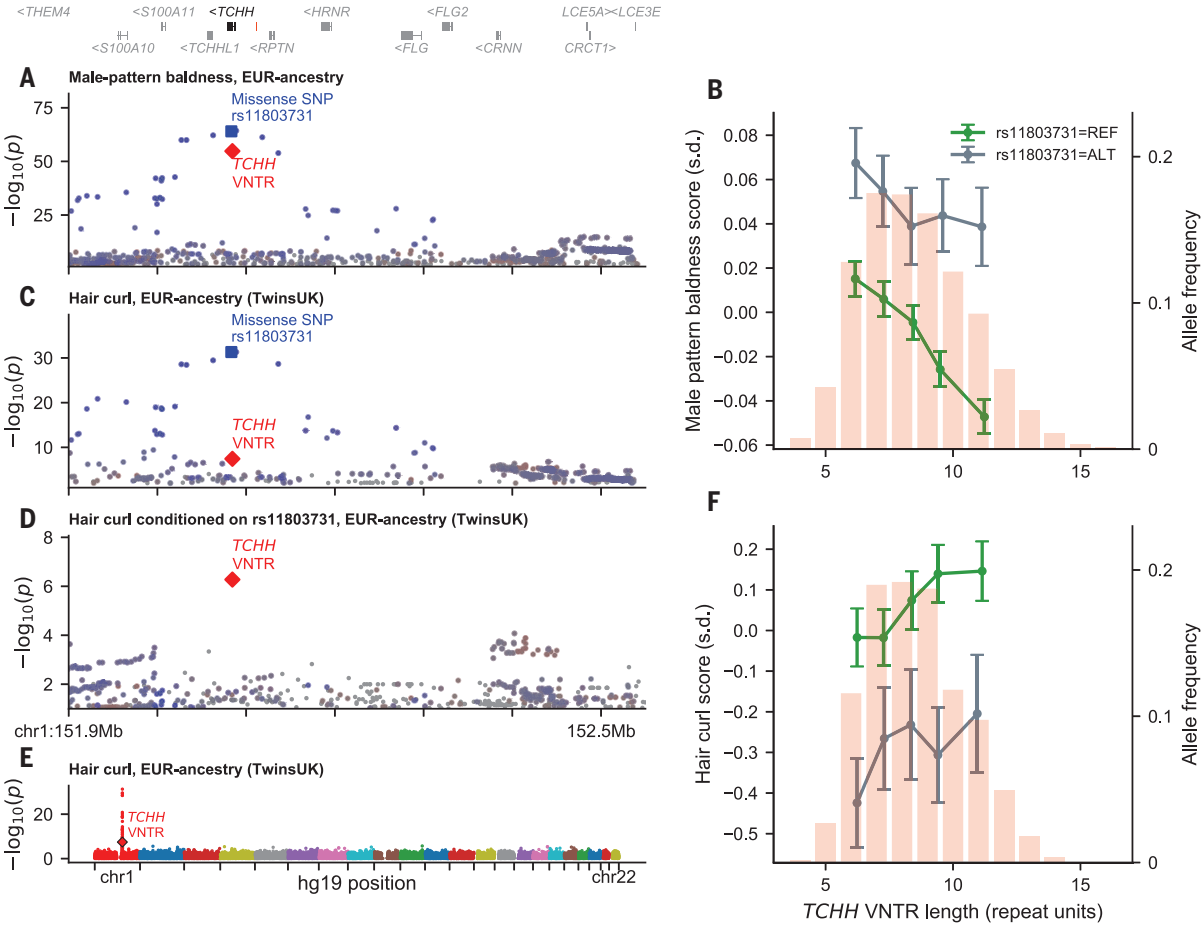


Fig. 4. *TCHH* VNTR length and missense SNP rs11803731 associate independently with hair phenotypes. (**A**) Genetic associations with male pattern baldness at *TCHH* (N = 189,537 male UKB EUR participants). Colors indicate partial LD (R > 0.1) with missense SNP rs11803731 (blue), the *TCHH* VNTR (red), or both rs11803731 and VNTR length (purple). (**B**) Mean baldness score in carriers (lines, left axis) and allele frequencies (histograms, right axis) of *TCHH* alleles. *TCHH* alleles were binned by VNTR length quintile and missense SNP rs11803731 status. (**C** and **D**) Genetic associations with hair curl at *TCHH* in N = 3334 TwinsUK participants [conditioned on rs11803731 in (D)]. (**E**) Genome-wide associations with hair curl in TwinsUK. (**F**) Relationship between *TCHH* allele length and hair curl [analogous to (B)].

BOLT-LMM). Longer VNTR alleles also associated with gout ($P = 3.6 \times 10^{-17}$, logistic regression), a disease caused by excessive uric acid crystallization in the joints.

The *MUC1* VNTR length polymorphism appeared to underlie some of the strongest, earliest reported SNP associations with serum urea and serum urate, two biomarkers of renal function that otherwise have somewhat independent heritability [genetic correlation = 0.25 (SE 0.01); Fig. 3, A and C]. For urea, the VNTR exhibited the strongest association genome-wide (matching that of a SNP on chromosome 5), explaining ~1% of heritable variance (~0.2% of total variance) in Europeans and accounting for nearly all of the association signal at the *MUC1* locus [previously reported as *MTX1-GBA* (*36*); Fig. 3A]. For urate, the VNTR also appeared to be the primary causal variant at a locus previously reported as *TRIM46* (*37*) (Fig. 3C). Longer *MUC1* alleles associated with increasing levels of both serum urea and urate across the VNTR length spectrum, with an incompletely dominant effect on urea ($P = 2.3 \times 10^{-20}$ for interaction, linear regression; fig. S17) but an additive effect on urate ($P = 0.56$ for interaction).

Associations with additional renal phenotypes indicated a complex relationship between *MUC1* VNTR length and kidney function (Fig. 3, B and D). Long *MUC1* alleles (>55 repeat units) increased the risk of gout (OR = 1.10; 95% CI = 1.08 to 1.13, $P = 1.2 \times 10^{-16}$, logistic regression) and chronic tubulointerstitial nephritis (OR = 1.31, 95% CI = 1.09 to 1.57, $P = 3.4 \times 10^{-3}$, logistic regression), which remained significant after correcting for 13 kidney diseases tested. However, *MUC1* VNTR allele length did not associate with chronic kidney disease (OR = 1.01, 95% CI = 0.99 to 1.04, $P = 0.33$, logistic regression) reported in 14,573 cases and only weakly influenced glomerular filtration rate as estimated from serum creatinine (beta = −0.19%, 95% CI = 0.11 to 0.28, for long versus short alleles). Long *MUC1* alleles associated with modest reductions in red blood cell counts (beta = −0.029 SD, SE = 0.002, $P = 1.5 \times 10^{-39}$, linear regression) and hemoglobin levels (beta = −0.031 SD, SE = 0.002, $P = 9.9 \times 10^{-44}$, linear regression), possibly reflecting an impact of reduced kidney function on erythropoietin production.

TCHH VNTR strongly associates with hair phenotypes

Repeat length variation in a coding VNTR in *TCHH* associated strongly with male pattern baldness ($P = 1.6 \times 10^{-55}$, BOLT-LMM). *TCHH* encodes trichohyalin, a protein that associates in regular arrays with keratin intermediate filaments and confers mechanical strength to the inner root sheath (*38*). The 18-bp VNTR encodes part of a highly stabilized alpha-helix

that forms an elongated rod structure (*39*). A rare nonsense mutation in *TCHH* has been implicated in uncombable hair syndrome (*40*), and a common haplotype containing the *TCHH* missense SNP rs11803731 (encoding a leucine to methionine substitution in TCHH) is by far the strongest genetic determinant of hair curl in individuals of European ancestry (*41*, *42*). In the UKB, the *TCHH* VNTR and rs11803731 exhibited independent associations with male pattern baldness (Fig. 4, A and B).

The *TCHH* VNTR appeared to be hypermutable and was poorly tagged by all nearby individual SNPs ($R^2 < 0.1$), leading us to wonder whether it might also contribute to hair curl in a way invisible to genome-wide association studies of this phenotype. Imputing *TCHH* VNTR alleles into the TwinsUK cohort (*43*) ($N = 3334$ genotyped individuals with hair curl phenotypes) revealed that the *TCHH* VNTR appeared to be the human genome's second-largest contributor to hair curl variation genome-wide (explaining ~1% of variance; $P = 3.6 \times 10^{-8}$, BOLT-LMM) after the missense SNP rs11803731 in *TCHH* (which explained ~4% of variance; Fig. 4, C to F). LD between the VNTR and rs11803731 further explained an association reported near *LCE3E* (450 kb upstream of *TCHH*) previously thought to be independent of *TCHH* (*42*) (Fig. 4, C and D).

Discussion

These results identify many strong effects of protein-coding VNTRs on human phenotypes. Most were among the strongest effects of all common variants identified for these phenotypes to date and resolved previously mysterious genetic associations for multiple traits. Incorporation of multiallelic VNTRs into fine-mapping analyses also helped to identify many more functional variants at the same loci, revealing the importance of incorporating allelic series of SNP and VNTR alleles into functional studies and epidemiological research.

These results are likely just the leading edge of a far larger set of VNTR-phenotype associations that future studies will reveal. In this work with exome-sequencing data, we were unable to analyze VNTRs that exist in noncoding sequences, are too short for depth-of-coverage to accurately measure length variation, or are too mutable to segregate well with SNP haplotypes. We anticipate that newer sequencing technologies applied to large, diverse cohorts will yield further insights into the mutational and evolutionary processes of VNTRs and their contribution to the "missing heritability" of human phenotypes.

A frustration in the study of human genetics has been that most reported genetic associations involve haplotypes of noncoding and missense SNPs with potential phenotypic contributions that are challenging to disentangle from one another and have first-order molecular effects that are opaque. VNTRs have several attributes that help to overcome these challenges. First, multiallelic VNTRs usually share only partial LD with nearby diallelic SNP and indel variants. Second, associations with protein-coding VNTRs implicate the size and copy number of specific protein domains, leading to specific, testable hypotheses about the effects of protein domains in biological systems. Third, the directions of coding VNTR associations have clear meaning, revealing whether risk is generated by having more or less of a domain. Finally, VNTRs generate natural allelic series of functionally distinct alleles that can be used for dose–response studies in human tissues and cellular models. We anticipate that these attributes will lead to new insights about the mechanisms by which gene and protein variation affect human biology.

REFERENCES AND NOTES

1. P. H. Sudmant *et al.*, *Nature* **526**, 75–81 (2015).
2. A. Sulovari *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23243–23253 (2019).
3. M. D. Lalioti *et al.*, *Nature* **386**, 847–851 (1997).
4. C. Wijmenga *et al.*, *Nat. Genet.* **2**, 26–30 (1992).
5. J. Marchini, B. Howie, S. Myers, G. McVean, P. Donnelly, *Nat. Genet.* **39**, 906–913 (2007).
6. C. Bycroft *et al.*, *Nature* **562**, 203–209 (2018).
7. Materials and methods are available as supplementary materials.
8. C. V. Van Hout *et al.*, *Nature* **586**, 749–756 (2020).
9. C. Benner *et al.*, *Bioinformatics* **32**, 1493–1501 (2016).
10. A. R. Barton, M. A. Sherman, R. E. Mukamel, P.-R. Loh, *Nat. Genet.* **53**, 1260–1269 (2021).
11. D. Beyter *et al.*, *Nat. Genet.* **53**, 779–786 (2021).
12. K. Schmidt, A. Noureen, F. Kronenberg, G. Utermann, *J. Lipid Res.* **57**, 1339–1359 (2016).
13. P.-R. Loh *et al.*, *Nat. Genet.* **47**, 284–290 (2015).
14. P. Ebert *et al.*, *Science* **372**, eabf7117 (2021).
15. R. Clarke *et al.*, *N. Engl. J. Med.* **361**, 2518–2528 (2009).
16. G. Utermann *et al.*, *J. Clin. Invest.* **80**, 458–465 (1987).
17. A. L. White, J. E. Hixson, D. L. Rainwater, R. E. Lanford, *J. Biol. Chem.* **269**, 9060–9066 (1994).
18. E. Boerwinkle *et al.*, *J. Clin. Invest.* **90**, 52–60 (1992).
19. K. Jaganathan *et al.*, *Cell* **176**, 535–548.e24 (2019).
20. S. Coassin *et al.*, *Eur. Heart J.* **38**, 1823–1831 (2017).
21. B. R. Zysow, G. E. Lindahl, D. P. Wade, B. L. Knight, R. M. Lawn, *Arterioscler. Thromb. Vasc. Biol.* **15**, 58–64 (1995).
22. K. Suzuki, M. Kuriyama, T. Saito, A. Ichinose, *J. Clin. Invest.* **99**, 1361–1366 (1997).
23. M. Trinder, M. M. Uddin, P. Finneran, K. G. Aragam, P. Natarajan, *JAMA Cardiol.* (2020).
24. D. F. Gudbjartsson *et al.*, *J. Am. Coll. Cardiol.* **74**, 2982–2994 (2019).
25. L. Yengo *et al.*, *Hum. Mol. Genet.* **27**, 3641–3649 (2018).
26. M. N. Weedon *et al.*, *Nat. Genet.* **40**, 575–583 (2008).
27. M. Graff *et al.*, *Am. J. Hum. Genet.* **108**, 564–582 (2021).
28. K. L. Lauing *et al.*, *Dev. Biol.* **396**, 224–236 (2014).
29. K. J. Doege, S. N. Coulter, L. M. Meek, K. Maslen, J. G. Wood, *J. Biol. Chem.* **272**, 13974–13979 (1997).
30. P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, M. Kircher, *Nucleic Acids Res.* **47** (D1), D886–D894 (2019).
31. L. Gleghorn, R. Ramesar, P. Beighton, G. Wallis, *Am. J. Hum. Genet.* **77**, 484–490 (2005).
32. M. Doyard *et al.*, *J. Med. Genet.* **55**, 278–284 (2018).
33. O. Gewartowska *et al.*, *Cell Rep.* **35**, 109015 (2021).
34. J. C. Fowler, A. S. Teixeira, L. E. Vinall, D. M. Swallow, *Hum. Genet.* **113**, 473–479 (2003).
35. A. Kirby *et al.*, *Nat. Genet.* **45**, 299–303 (2013).
36. Y. Okada *et al.*, *Nat. Genet.* **44**, 904–909 (2012).
37. A. Köttgen *et al.*, *Nat. Genet.* **45**, 145–154 (2013).
38. P. M. Steinert, D. A. D. Parry, L. N. Marekov, *J. Biol. Chem.* **278**, 41409–41419 (2003).
39. S. C. Lee *et al.*, *J. Biol. Chem.* **268**, 12164–12176 (1993).
40. F. B. Ü. Basmanav *et al.*, *Am. J. Hum. Genet.* **99**, 1292–1304 (2016).
41. S. E. Medland *et al.*, *Am. J. Hum. Genet.* **85**, 750–755 (2009).
42. F. Liu *et al.*, *Hum. Mol. Genet.* **27**, 559–575 (2018).
43. A. Moayyeri, C. J. Hammond, D. J. Hart, T. D. Spector, *Twin Res. Hum. Genet.* **16**, 144–149 (2013).
44. R. E. Mukamel, R. E. Handsaker, M. A. Sherman, A. R. Barton, Y. Zheng, S. A. McCarroll, P.-R. Loh, Codes and scripts for: Protein-coding repeat polymorphisms strongly shape diverse human phenotypes, Zenodo (2021); https://doi.org/10.5281/zenodo.4776804.

ACKNOWLEDGMENTS

We thank R. Gupta, J. Hirschhorn, M. Hujoel, S. Raychaudhuri, and M. Warman for helpful discussions. This research was conducted using the UKB resource under application no. 40709. Computational analyses were performed on the O2 High Performance Compute Cluster, which is supported by the Research Computing Group, at Harvard Medical School (http://rc.hms.harvard.edu). **Funding:** R.E.M. was supported by National Science Foundation (NSF) grant DMS-1939015 and National Institutes of Health (NIH) grant K25 HL150334. R.E.H. and S.A.M. were supported by NIH grant R01 HG006855. M.A.S. was supported by the MIT John W. Jarve (1978) Seed Fund for Science Innovation and by NIH fellowship F31 MH124393. A.R.B. was supported by NIH fellowship F31 HL154537 and training grant T32 HG 2295-16. P.-R.L. was supported by NIH grant DP2 ES030554, a Burroughs Wellcome Fund Career Award at the Scientific Interfaces, the Next Generation Fund at the Broad Institute of MIT and Harvard, and a Sloan Research Fellowship. TwinsUK is funded by the Wellcome Trust, Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd., and the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility, and Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust in partnership with King's College London. **Author contributions:** R.E.M., R.E.H., S.A.M., and P.-R.L. conceived and designed the study. R.E.M., R.E.H., and P.-R.L. designed and implemented the statistical methods and performed the computational analyses. Y.Z. helped to design and implement the VNTR partitioner algorithm. R.E.M., R.E.H, A.R.B., M.A.S., S.A.M., and P.-R.L. interpreted analytical results. All authors wrote and edited the manuscript. **Competing interests:** The authors declare no competing interests. **Data availability:** Access to the following data resources is available to all bona fide researchers by application: UKB (https://www.ukbiobank.ac.uk/), Twins UK (https://twinsuk.ac.uk/), the Haplotype Reference Consortium imputation panel (http://www.haplotype-reference-consortium.org/), and AAAGC height summary statistics (https://www.ebi.ac.uk/gwas/). Individual-level VNTR allele length estimates (resolved to phased SNP haplotypes) and genetically predicted Lp(a) values are available from UKB as a return from application no. 40709. **Code availability:** The following publicly available software resources were used to perform analyses in this work: Eagle2 (v2.3.5), https://data.broadinstitute.org/alkesgroup/Eagle/; Minimac4 (v1.0.1), https://genome.sph.umich.edu/wiki/Minimac4; BOLT-LMM (v2.3.5), https://data.broadinstitute.org/alkesgroup/BOLT-LMM/; FINEMAP (v1.3.1), http://www.christianbenner.com/; plink (v1.9 and v2.0), https://www.cog-genomics.org/plink2/; Tandem Repeats Finder (v4.09.1), tandem.bu.edu/trf/trf.html; the TOPMed Imputation Server, https://imputation.biodatacatalyst.nhlbi.nih.gov/; BLAT (v35), http://hgdownload.soe.ucsc.edu/admin/exe/; susieR (v0.10.1), https://stephenslab.github.io/susieR/; LDstore (v2.0), http://www.finemap.me; and ImpG (v1.0.1), https://github.com/huwenboshi/ImpG. Code and scripts used to perform analyses are available at Zenodo (*44*).

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abg8289
Materials and Methods
Supplementary Text
Figs. S1 to S17
Tables S1 to S8
References (*45*–*98*)
MDAR Reproducibility Checklist

29 January 2021; accepted 20 August 2021
10.1126/science.abg8289

SUPERCONDUCTIVITY

Restored strange metal phase through suppression of charge density waves in underdoped $YBa_2Cu_3O_{7-\delta}$

Eric Wahlberg[1], Riccardo Arpaia[1,2]*, Götz Seibold[3], Matteo Rossi[2]†, Roberto Fumagalli[2], Edoardo Trabaldo[1], Nicholas B. Brookes[4], Lucio Braicovich[2,4], Sergio Caprara[5,6], Ulf Gran[7], Giacomo Ghiringhelli[2,8], Thilo Bauch[1], Floriana Lombardi[1]*

The normal state of optimally doped cuprates is dominated by the "strange metal" phase that shows a linear temperature (T) dependence of the resistivity persisting down to the lowest T. For underdoped cuprates, this behavior is lost below the pseudogap temperature T^*, where charge density waves (CDWs), together with other intertwined local orders, characterize the ground state. We found that the T-linear resistivity of highly strained, ultrathin, underdoped $YBa_2Cu_3O_{7-\delta}$ films is restored when the CDW amplitude, detected by resonant inelastic x-ray scattering, is suppressed. This observation suggests an intimate connection between the onset of CDWs and the departure from T-linear resistivity in underdoped cuprates. Our results illustrate the potential of using strain control to manipulate the ground state of quantum materials.

Cuprate high-temperature superconductors belong to a class of materials where strong electron-electron correlations play a fundamental role, and whose unconventional properties might require abandoning traditional concepts of solid-state physics for a proper description (*1*). The "strange metal" phase of these superconductors is one of the most striking manifestations of the strong correlations. At optimal doping, this phase manifests as a linear temperature dependence of the resistivity that persists to the lowest T when superconductivity is suppressed (*1*–*4*). This behavior is fundamentally different from that observed in more conventional metals (*3*), where a T-linear dependence of the resistivity is found only at high temperatures where phonon scattering dominates the transport. T-linear resistivity is also found in other strongly correlated systems, including pnictides (*5*) and magic-angle bilayer graphene (*6*).

Recent developments have attempted to model this behavior by considering that the scattering time approaches the fundamental Planckian limit defined by $\tau = \hbar/k_B T$ (where $\hbar$ and k_B are the reduced Planck and Boltzmann constants) irrespective of the nature of the scattering process (*4*, *7*). Local interactions among quasiparticles suggest that ordinary metals cannot thermalize on such a short time scale; reaching the Planckian limit would require every particle to be entangled with every other particle in the system.

In cuprates, the T-linear resistivity is lost for doping above and below the optimal doping. In the overdoped regime, the recovery of an almost T^2 dependence of the resistivity, typical of a Fermi liquid, is a consequence of the increased screening of the electron-electron interactions caused by a higher charge carrier density. In the underdoped regime, the deviation from T-linear behavior happens at temperatures close to T^*, known as the pseudogap temperature, where states are missing at the Fermi energy (*8*). In the pseudogap region, the high-temperature superconductor phase diagram also hosts a plethora of intertwined electronic local ordering phenomena that break rotational/translational symmetry (*1*, *9*–*13*); charge density wave (CDW) order (*11*) is the most prominent one. The association between the departure from the T-linear resistivity and the occurrence of the pseudogap phenomenon has long been speculated. However, there is no consensus on the physics at play, nor on the causality hierarchy among pseudogap, local orders, and strange metal phenomenology (*14*, *15*). This is because the strange metal exhibits its most salient features in transport and its connection to spectral signatures is elusive (*16*). More specifically, the challenge is to disentangle the various possible mechanisms leading to the breakdown of the T-linear resistivity, such as the occurrence of the pseudogap and the appearance of local orders such as CDW. One way to address this challenge is to tune the local properties of underdoped high-temperature superconductors. In particular, the CDW can be strongly modified under pressure (*17*, *18*), strong magnetic fields (*19*), and strain in crystals (*20*) and thin films (*21*).

To tune the ground state in thin films of the cuprate material $YBa_2Cu_3O_{7-\delta}$ (YBCO), we use the geometric modification of its unit cell under the strong strain induced by the substrate. We show that the T-linear resistivity dependence is completely recovered (down to the superconducting critical temperature T_c) along the b axis when the CDW, detected by resonant inelastic x-ray scattering (RIXS), is suppressed along the a axis.

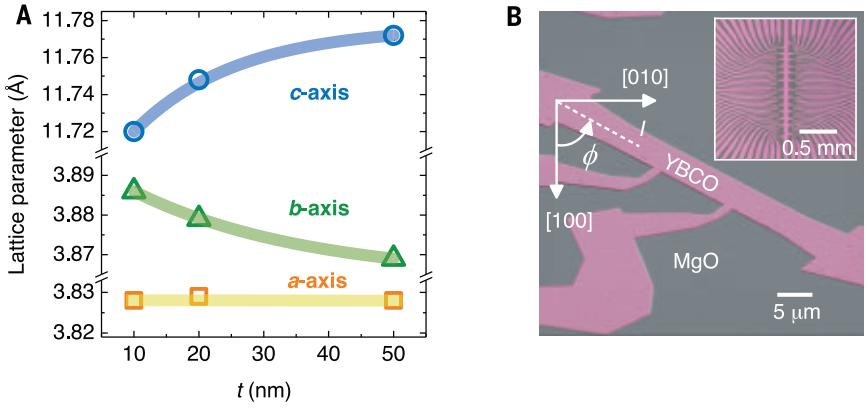


Fig. 1. Thickness dependence of the YBCO lattice parameters for films grown on MgO substrates. (**A**) Lattice parameters (*a*, *b*, and *c* axes are indicated by squares, triangles, and circles, respectively) of YBCO, measured by x-ray diffraction at 300 K, as function of the thickness of films with hole doping $p \approx 0.12$ grown on MgO. Thick lines are guides to the eye. (**B**) False-color scanning electron microscope image of a typical device used to measure the resistivity ρ, with the current I flowing at an angle ϕ with respect to the YBCO [100] direction (*a* axis). Inset: Overview of the patterned samples.

[1]Quantum Device Physics Laboratory, Department of Microtechnology and Nanoscience, Chalmers University of Technology, SE-41296 Göteborg, Sweden. [2]Dipartimento di Fisica, Politecnico di Milano, I-20133 Milano, Italy. [3]Institut für Physik, BTU Cottbus-Senftenberg, D-03013 Cottbus, Germany. [4]ESRF, European Synchrotron, F-38043 Grenoble, France. [5]Dipartimento di Fisica, Università di Roma "La Sapienza," I-00185 Roma, Italy. [6]CNR-ISC, I-00185 Roma, Italy. [7]Division of Subatomic, High-Energy and Plasma Physics, Chalmers University of Technology, SE-41296 Göteborg, Sweden. [8]CNR-SPIN, Dipartimento di Fisica, Politecnico di Milano, I-20133 Milano, Italy.
***Corresponding author. Email: floriana.lombardi@chalmers.se (F.L.); riccardo.arpaia@chalmers.se (R.A.)**
†Present address: Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA.

Thin-film devices

The films we use span a wide range of hole-doping p, going from the strongly underdoped ($p \approx 0.10$) up to the optimally doped ($p \approx 0.17$) regime (*22*, *23*). The strain was modified both by changing the substrate and by varying the film thickness t in a range from 50 nm down to 10 nm. We used (110)-oriented MgO and 1° vicinal angle (001) $SrTiO_3$ (STO) substrates to grow untwinned YBCO films (*24*). When the YBCO thickness is reduced to a few unit cells (t = 10 nm), films grown on MgO are characterized by a considerable elongation of the b axis and contraction of the c axis, with the total volume of the unit cell unchanged with respect to relaxed systems (Fig. 1A). For films grown on vicinal cut STO, the YBCO unit cell is instead almost thickness-independent (fig. S1).

Figure 1B shows a typical device used to measure the resistivity as a function of temperature. The devices are patterned at an angle ϕ with respect to the YBCO [100] direction by using a carbon mask in combination with electron beam lithography and Ar^+ ion milling (*22*, *23*, *25*).

Angular dependence of in-plane resistivity

Figure 2A shows the temperature dependence of the resistivity ρ, measured in two devices oriented along the YBCO a and b axes (ϕ = 0° and 90°, respectively) realized in an underdoped (p = 0.11) 50-nm-thick film grown on an MgO substrate. The resistivity anisotropy ratio at T = 290 K, defined by ρ_a(290 K)/ρ_b(290 K), is ~1.2, a value fully compatible with the level of hole doping (*26*). The temperature T_L, estimated as the temperature below which the resistivity normalized to 290 K deviates by 1% from the linear fit, is approximately the same for both devices, and it is very close to the pseudogap temperature T^* of single crystals at comparable doping measured using spectroscopic techniques (*8*, *27*). For this film thickness, we therefore have $T_L^a \approx T_L^b \approx T^*$.

Figure 2B shows analogous data for two devices patterned on a 10-nm-thick film on MgO (grown with the same deposition conditions as the 50-nm-thick film) with doping p comparable to Fig. 2A. The superconducting critical temperature T_c at this reduced thickness remains almost unchanged (*28*) relative to 50-nm-thick films. However, the overall $\rho(T)$ behavior is very different. We observe three striking features: (i) The in-plane anisotropy of the resistivity ρ_a(290 K)/ρ_b(290 K) is much larger relative to the 50-nm-thick film, (ii) the slopes $\gamma_{a,b}$ of the high-temperature linear resistivity along the a and b axes are very different, and (iii) the resistivity along the b axis $\rho_b(T)$ has a much broader temperature range of linearity, as is clearly highlighted by subtracting from $\rho_b(T)$ the linear fit $\rho_L(T) = \rho_0 + \gamma_b T$ of the high-temperature resistive behavior (Fig. 2B, inset). Here, ρ_0 and γ_b are respectively the intercept at T = 0 and the slope of the high-temperature linear dependence. At p = 0.14 and p = 0.147 (Fig. 2C and fig. S4C), for 10-nm-thick films we find that the T-linear behavior is completely recovered down to the superconducting transition. This finding is crucial because it indicates that the "strange metal" behavior is restored in ultrathin underdoped YBCO. What are the conditions for this to happen?

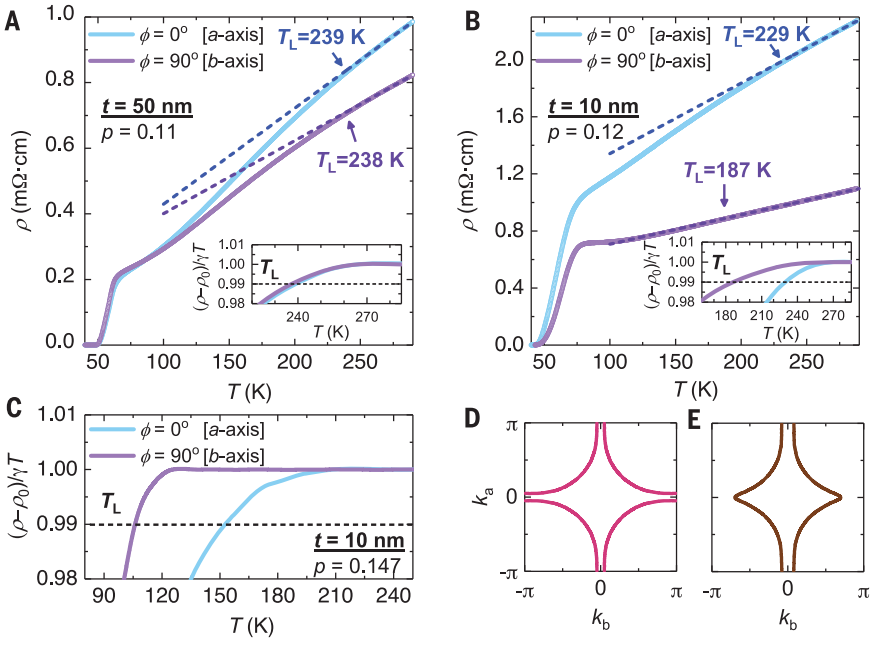


Fig. 2. Angular dependence of the YBCO in-plane resistivity as a function of the thickness of films on MgO under strain. (**A**) $\rho(T)$ of two devices, oriented along the a- and b-axis directions, patterned on a 50-nm-thick underdoped (p = 0.11) film. At T = 290 K, ρ_a/ρ_b = 1.2. The dashed lines are the linear fits of the curves for T > 260 K. The inset shows the determination of T_L, which is the temperature where the resistivity normalized to 290 K deviates by 1% from the linear fit (ρ_0 and γ are the coefficients of the fit). (**B**) $\rho(T)$ of two devices, oriented along the a- and b-axis directions, patterned on a 10-nm-thick underdoped (p = 0.12) film. At T = 290 K, ρ_a/ρ_b = 2.1. (**C**) Determination of T_L in two slightly underdoped devices, oriented along the a- and b-axis directions, patterned on a 10-nm-thick slightly underdoped (p = 0.147) film. $\rho(T)$ along the b axis is linear down to the onset of superconducting fluctuations around T = 100 K. (**D**) Simplified model of the typical Fermi surface of cuprates (here we disregard the influence of the CuO chains). For this kind of Fermi surface, the resistivity should be isotropic in the copper oxide planes. (**E**) Hypothetical anisotropic model Fermi surface that is compatible with the anisotropic transport of the 10-nm-thick devices.

The most prominent structural effect we encounter by reducing the thickness of the films is that the YBCO b axis expands, whereas the a axis is only slightly modified; at the same time, the total volume of the cell remains unaltered as a consequence of a c-axis contraction. One of the effects of the strain is therefore to increase the orthorhombicity of few-nm-thick films. For 10-nm-thick underdoped films ($p \approx$ 0.12), the values of the lattice parameters a and b are similar to those of YBCO with a much higher doping (*29*), close to the optimal doping. However, this effect cannot explain the anomalously high anisotropy ratio ρ_a(290 K)/ρ_b(290 K) that we observe (Fig. 2B), a value that in an optimally doped YBCO crystal is mainly related to the presence of CuO chains (*26*). Indeed, within a simple tight-binding description, an increase of ~0.02 Å in the b-axis lattice parameter—as we observe upon reducing the film thickness from 50 nm to 10 nm (Fig. 1A)—would reduce the corresponding hopping parameter between neighbor sites in the b direction by only ~1% (*30*). Moreover, the resulting (weak) modification of the electronic structure would induce the opposite anisotropy in the a- and b-axis resistivities, namely $\rho_b > \rho_a$ (because of the larger b-axis value), relative to what we experimentally observe. This rules out the increased orthorhombicity as a possible direct explanation of the anisotropy we observe.

The very different slopes of $\rho_a(T)$ and $\rho_b(T)$ give a hint of the physics at play in the 10-nm-thick films. Following Boltzmann transport

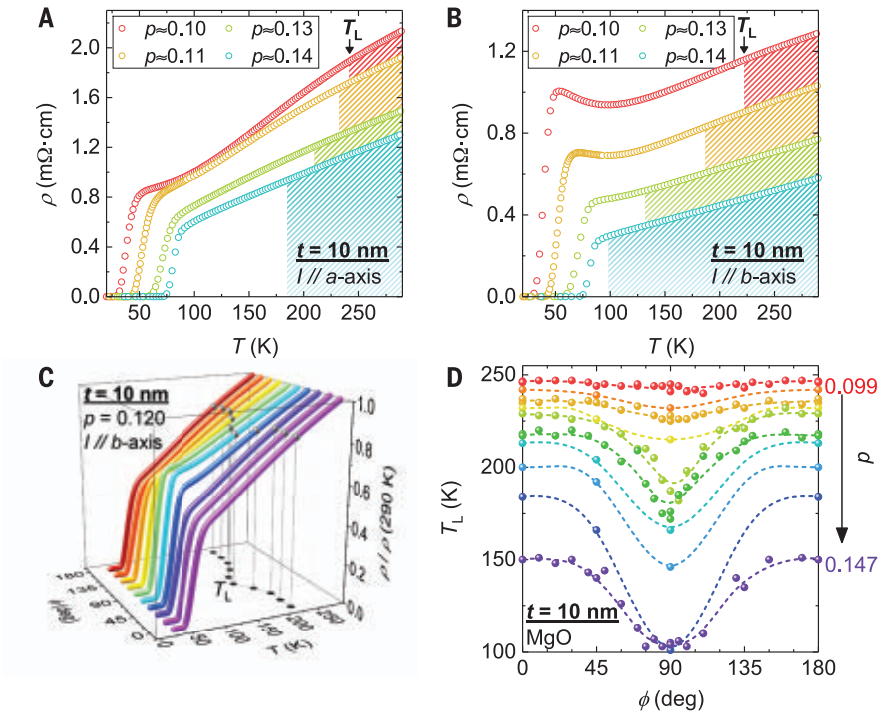


Fig. 3. Angular dependence of the in-plane resistivity of the 10-nm YBCO on MgO as a function of oxygen doping. (**A**) ρ(*T*) measured along the *a*-axis direction of 10-nm-thick films, with different doping levels, grown on MgO substrates. T_L is extracted as described in Fig. 1. The hatched regions represent the temperature range above T_L where at each *p* the *T*-linear resistivity regime occurs. (**B**) Same as (A), but with ρ(*T*) measured along the *b* axis. (**C**) ρ(*T*), normalized to its value at *T* = 290 K, for a 10-nm-thick film (*p* = 0.120) on MgO is shown as a function of ϕ (i.e., with respect to the *a*-axis direction). For each angle, the temperature T_L has been extracted (black spheres and dots). T_L is approximately constant as a function of ϕ, except around ϕ = 90°, where it exhibits a suppression. (**D**) T_L values as a function of ϕ for 10-nm-thick films with different doping levels (*p* ≈ 0.099, 0.103, 0.108, 0.117, 0.118, 0.120, 0.123, 0.134, 0.140, or 0.147).

theory, the conductivity is given by

$$\sigma_{a,b}(T) = 2e^2 \sum_k \frac{v^2_{F,a,b}}{\Gamma(k)}\{-n'_F\} \qquad (1)$$

where $v_{F,a,b}$ is the Fermi velocity, $\Gamma(k)$ is the k-dependent scattering rate, and n'_F is the derivative of the Fermi distribution. Consequently, the ratio $\Gamma(k)/v^2_{F,a,b}$ determines the slope $\gamma_{a,b}$ of the temperature-dependent film resistivity. From the experimental data, we have $\gamma_a \gg \gamma_b$ (at p = 0.11, γ_a = 4.9 μΩ·cm/K and γ_b = 2.1 μΩ·cm/K), which indicates that $v_{F,b} \gg v_{F,a}$ and therefore that the Fermi surface is already strongly anisotropic at room temperature. This is illustrated in Fig. 2, D and E, which show respectively the typical isotropic Fermi surface for the cuprates (where the contribution of the chains is neglected) and an anisotropic distorted Fermi surface, compatible with our experimental resistivity anisotropy. From Eq. 1, one may argue that anisotropic elastic scattering processes [e.g., due to small-angle scattering from impurities between CuO_2 planes (*31*)] can also account for the observed resistivity anisotropy. However, in this case $\Gamma(k)$ would be determined by the local density of states (*31*); that is, Γ would still be proportional to $1/v_F$. Therefore, only an anisotropy of the Fermi velocity, which breaks the C_4 symmetry of the crystal, can account for the observed resistivity anisotropy.

A Fermi surface of the type shown in Fig. 2E can be the consequence of a strong electronic nematicity in the system. Such a state has been extensively investigated from a theoretical point of view (*32*–*36*) and it has been experimentally found, by in-plane resistivity anisotropy measurements, in tetragonal $La_{2-x}Sr_xCuO_4$ films (*37*). We speculate that in the 10-nm thin films, the strain-induced distortion of the cell plays a fundamental role in stabilizing a nematic ground state already at room temperature.

The wider *T*-linear behavior of $\rho_b(T)$ in ultrathin films is not observed in YBCO on vicinal angle STO substrates. Here T_L does not change; it is the same along the *a* and *b* axes and going from 50-nm-thick to 10-nm-thick films (see fig. S3). This in agreement with the fact that the slopes $\gamma_{a,b}$ of $\rho_a(T)$ and $\rho_b(T)$ are comparable in 10- and 50-nm-thick films, so on STO substrates the Fermi surface is not substantially modified by strain at reduced thicknesses.

We have therefore arrived at the main result of our paper: A specific strain in underdoped 10-nm-thick YBCO films on a MgO substrate induces a nematic state that modifies the Fermi surface already at room temperature. But why would a distorted Fermi surface recover the *T*-linear resistivity behavior along the *b* axis? The answer to this question comes from the study of the ρ(*T*) dependence as a function of doping. Figure 3 shows $\rho_a(T)$ and $\rho_b(T)$ respectively as a function of the doping *p* (Fig. 3, A and B) and of the angle ϕ (Fig. 3, C and D) for 10-nm-thick YBCO films on MgO substrates (as a function of the thickness in fig. S4).

The $\rho_a(T)$ curves are rather conventional and in agreement with previous results (*22*, *26*). However, $\rho_b(T)$ (Fig. 3B) looks very different. As anticipated earlier, for $p \approx 0.14$ we have completely recovered the strange metal behavior: The *T*-linear dependence extends through the entire temperature range until superconductivity sets in. However, the situation changes at lower doping: For $p \approx 0.10$, the extracted T_L along *b* is close to the value along the *a* axis (Fig. 2, C and D), and $\rho_b(T)$ shows a pronounced upturn at low temperature before the superconducting transition. The upturn of the $\rho_b(T)$ in our 10-nm-thick YBCO films is observed at higher doping relative to the $\rho_a(T)$ (Fig. 3, A and B); indeed, it already appears at p = 0.13. This upturn in the resistivity has been attributed to the loss of high-mobility electron pockets because of the CDW order ending at $p \approx 0.08$ (*38*) and/or to the proximity to the antiferromagnetic Mott insulator through spin density waves (*39*–*41*). The doping dependence of $\rho_b(T)$ therefore points toward a strong involvement of the CDW order in the phenomenology we observe.

Resonant inelastic x-ray measurements of CDW

To characterize the CDW order in the YBCO films, we used RIXS at the Cu L_3 edge (~930 eV) (*23*). We investigated films with thickness of 10 and 50 nm grown on both MgO and STO at the doping $p \approx 0.125$ where the intensity of the CDW is strongest (*11*, *42*). To isolate the contribution of the CDW, we measured RIXS spectra at T = 70 K (i.e., close to T_c), where the CDW signal is maximized, and at T = 200 K, a temperature where the CDW contribution is negligible. The CDW peak has been explored along *a* and *b* with two orthogonal cuts along the *H* and *K* directions of reciprocal space, centered around the wave vector of the CDW $q_c^{CDW} = (H_{CDW}, K_{CDW})$. Shown in Fig. 4, A and B, is the quasi-elastic component of the RIXS spectra along the *a* and *b* axes for a 50-nm-thick film on MgO, as a function of *H*, at T = 70 K

and T = 200 K. Along both directions, at T = 200 K only a broad peak is present (Fig. 4, A and B, red regions); at T = 70 K, the signal is given by the sum of a broad peak (similar to that measured at high temperature) and of a narrow peak. This narrow, temperature-dependent peak is a signature of CDW; the broad, almost temperature-independent peak is instead a signature of short-range charge density fluctuations (CDFs) (*43*) (i.e., charge modulations, precursors of CDWs), which are present in the phase diagram at any temperature and in a very broad doping range, including the overdoped region. The same measurements as in Fig. 4, A and B, are reported in Fig. 4, C and D, for a 10-nm-thick film on MgO. Here, along the *b* axis (Fig. 4D), both the CDW and the CDF peaks are very similar to those measured in the thick film. Along the *a* axis (Fig. 4C), the situation is markedly different. The broad-in-$q_{//}$ CDF peak is unchanged with respect to the 50-nm-thick sample; in contrast, the narrow CDW peak, emerging at lower temperature, is almost negligible. This occurrence has been verified on the same sample, measuring along the *K* direction (fig. S5), and on other ultrathin films on MgO with different doping (fig. S6).

Note that our films are not perfectly untwinned (the untwinning degree is ~85% for films grown on MgO, ~90% for films grown on STO). The small temperature-dependent CDW signal measured along the *a* axis in the 10-nm films on MgO can be attributed to twinned domains. Indeed, once the twinning is taken into account, the actual CDW signal becomes effectively negligible in the *a*-axis direction. We conclude that in 10-nm-thick films on MgO, the CDW is unidirectional and directed along the *b* axis, whereas the CDW in YBCO grown on STO is thickness-independent (fig. S7). Strain-induced modifications of CDW have also been recently observed in YBCO single crystals, where the in-plane uniaxial compression along either *a* or *b* gives rise to an enhancement of the CDW in the orthogonal direction (*44*).

Discussion

Our results are in line with the theoretical predictions of unidirectional CDWs in the presence of nematicity (*34*, *35*), as a consequence of the modified Fermi surface. Our experiment additionally shows a clear correlation between the *T*-linear dependence of $\rho(T)$ and the charge order: The recovery of the *T*-linear resistivity along the *b* axis is associated with the suppression of the CDW along the YBCO *a* axis. In this scenario, only CDFs survive in the system at any temperature, which might be relevant both for Planckian metal (*45*) and marginal Fermi liquid theories (*46*) of the strange metal.

The strong correlation between the CDW and the breakdown of the *T*-linear behavior is further supported by the coincidence of the onset temperatures for the two phenomena. Figure S8 shows the temperature dependence of the CDW order for our 50- and 10-nm-thick films on MgO and STO substrates. Within experimental error, T_{CDW} (for p = 0.125) is the same as the temperature at which the resistivity departs from the linearity T_L^a and the same as the pseudogap temperature T^* [taken from literature (*8*, *27*)]. This is illustrated in Fig. 5, which shows a revised phase diagram for the 10-nm-thick films. Our RIXS measurements therefore require the revision of the common belief that the CDW is a low-temperature phenomenon happening at temperatures well below T_L and T^*, as indicated in early studies (*11*, *42*). The coincidence among T^*, T_L, and T_{CDW} is a common feature of doping above p = 1/8 where the CDW order is strong, as also supported by recent RIXS data (*43*), which are included for completeness in Fig. 5.

The quality of our samples allows us to exclude the possibility that structural changes of the CuO chains along the *b* axis in 10-nm films might have a role in the phenomenology we have observed. For overdoped films, where the conductivity of the chains causes an upward deviation from linearity in $\rho(T)$, we observe the same $\rho_b(T)$ dependence for 50-nm and 10-nm films and very similar values of T_L (fig. S2), which excludes CuO chain modification effects in our very thin films.

However, there is still an issue to be addressed: Why, in our films, does the disappearance of the CDW along the YBCO *a* axis

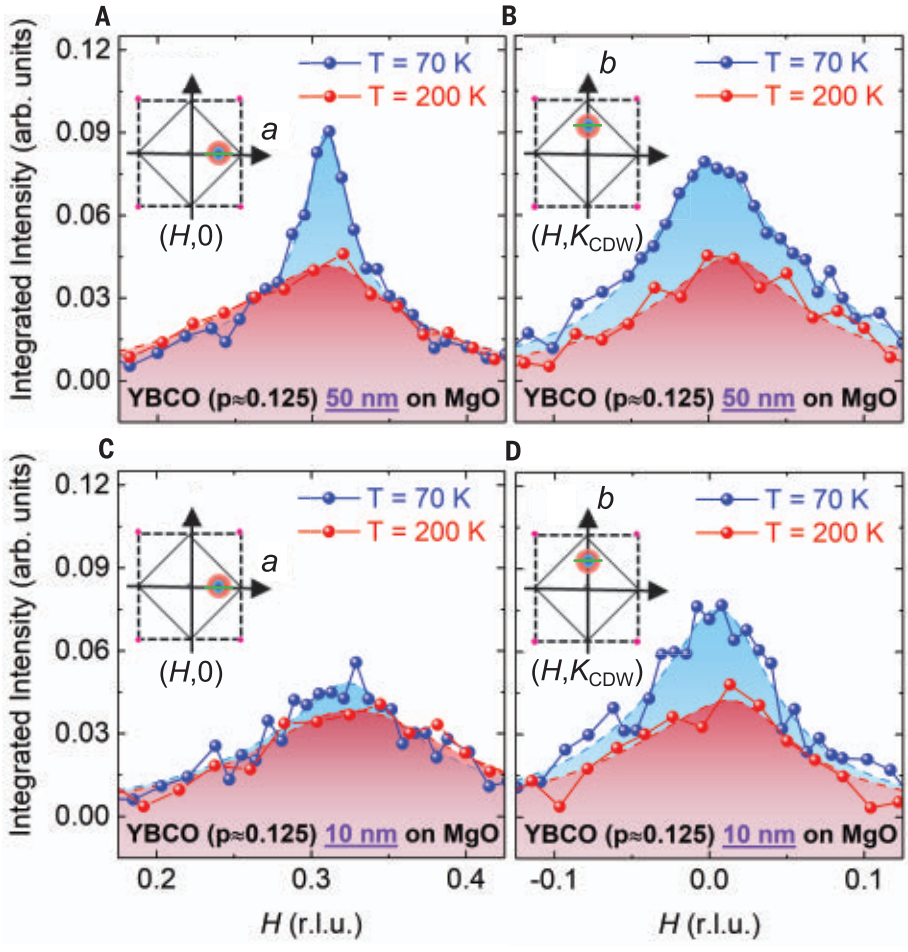


Fig. 4. Thickness dependence of the CDW in YBCO thin films. (**A** and **B**) Quasi-elastic scans measured at T = 70 K and T = 200 K on the 50-nm thin film on MgO along the (H, 0) direction (i.e., *a* axis) (A) and the (H, K_{CDW}) direction (i.e., *b* axis) (B). (**C** and **D**) Same as (A) and (B), but on a 10-nm-thick sample. The measurements were performed along the (H, 0) direction (i.e., *a* axis) (C) and the (H, K_{CDW}) direction (i.e., *b* axis) (D). In the 10-nm-thick sample, the CDW intensity along the *a* axis is almost negligible. If we take into account the percentage of twin domains (~15%) present in our films, we conclude that in films a few unit cells thick on MgO, the CDW is unidirectional along the *b* axis. The green line in the inset of each panel shows the direction of the scan relative to the CDW peak in reciprocal space.

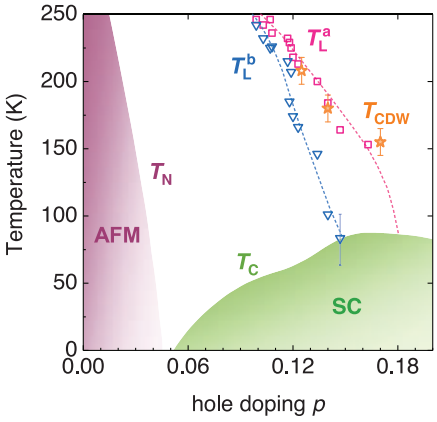


Fig. 5. Intertwined phases in the phase diagram of YBCO ultrathin films. The antiferromagnetic (AFM) and superconducting (SC) states are present below T_N and T_c, respectively [the T_c dome is derived from (*22*)]. T_L serves as a crossover line between the quantum-critical region of the phase diagram exhibiting strange metal properties (to the right of T_L) and the noncritical region, where the linearity of the resistivity versus temperature is lost (to the left of T_L). Along the *a* axis, $T_L = T_L^a$ (pink squares) is close to the pseudogap temperature T^* as determined by other techniques, whereas along the *b* axis (blue triangles), T_L is suppressed to lower temperatures. The error bar on T_L^b highlights the uncertainty of determining the correct value owing to proximity to T_c. The lines are guides to the eye. T_{CDW} (stars) is the onset temperature of the CDW, as determined by resonant inelastic x-ray scattering. The open star is related to the doping level investigated by RIXS in this manuscript (for 10-nm-thick films, revealed only along the *b* axis); the solid stars are from (*43*).

lead to a recovery of the *T*-linear dependence of the resistivity along the *b* axis (Figs. 2B and 4C)? It is not uncommon in other compounds that the occurrence of a unidirectional CDW phenomenon affects the transport properties in a direction orthogonal to the CDW *q*-vector (*47*). In our experiment, the reason can be linked to the nematic Fermi surface for very thin films. The generic Fermi surface (Fig. 2D) supports two scattering processes along the *b* axis associated with the quasi-nesting properties of the CDW *q*-vector: one within the Brillouin zone and the other connecting two adjacent Brillouin zones at (π, 0) (fig. S9A). For a conventional Fermi surface, the scattering rates of these two processes are comparable (*23*). However, for the distorted Fermi surface of Fig. 2E, the region around the (π, 0) point has a much higher density of states relative to the unperturbed one (fig. S9). As a consequence, the CDW scattering rate involving two adjacent Brillouin zones becomes larger than the scattering rate within the same Brillouin zone by a factor of >4 (*23*). Following the Boltzmann equation formalism, this strongly affects the resistivity along the *a* axis while leaving the resistivity along the *b* axis unaltered.

Finally, we consider the role of the pseudogap in the departure of the *T*-linear resistivity in underdoped cuprates. From the slopes of $\rho(T)$ along the *a* and *b* axes, $\gamma_{a,b}$, we infer that the Fermi surface is anisotropic at any doping (the ratio γ_a/γ_b is almost doping-independent). For $p \leq 0.1$, we observe that T_L, which at that doping level is much higher than T_{CDW} and comparable to T^* [taken from literature (*8*, *27*)], is isotropic ($T_L^a \approx T_L^b$). This hints at a pseudogap that is isotropic on a distorted Fermi surface. Such an isotropic pseudogap cannot explain the otherwise anisotropic T_L we observed at higher doping ($p > 0.1$). This allows us to conclude that the pseudogap is not the main effect responsible for the departure of the linearity when the CDW order is strong ($p > 0.11$), which is the range of doping in our experiment. However, one cannot exclude a major influence of the pseudogap on the transport at lower doping when T^* becomes much higher (close to room temperature and beyond) than T_{CDW}.

REFERENCES AND NOTES

1. B. Keimer, S. A. Kivelson, M. R. Norman, S. Uchida, J. Zaanen, *Nature* **518**, 179–186 (2015).
2. R. Daou *et al.*, *Nat. Phys.* **5**, 31–34 (2009).
3. J. A. N. Bruin, H. Sakai, R. S. Perry, A. P. Mackenzie, *Science* **339**, 804–807 (2013).
4. A. Legros *et al.*, *Nat. Phys.* **15**, 142–147 (2019).
5. S. Kasahara *et al.*, *Phys. Rev. B* **81**, 184519 (2010).
6. Y. Cao *et al.*, *Nature* **556**, 43–50 (2018).
7. J. Zaanen, *SciPost Phys.* **6**, 061 (2019).
8. T. Timusk, B. Statt, *Rep. Prog. Phys.* **62**, 61–122 (1999).
9. C. M. Varma, *Phys. Rev. B* **55**, 14554–14580 (1997).
10. Y. Sato *et al.*, *Nat. Phys.* **13**, 1074–1078 (2017).
11. G. Ghiringhelli *et al.*, *Science* **337**, 821–825 (2012).
12. J. M. Tranquada, B. J. Sternlieb, J. D. Axe, Y. Nakamura, B. S. Uchida, *Nature* **375**, 561–563 (1995).
13. B. Fauqué *et al.*, *Phys. Rev. Lett.* **96**, 197001 (2006).
14. E. Fradkin, S. A. Kivelson, J. M. Tranquada, *Rev. Mod. Phys.* **87**, 457–482 (2015).
15. T. Wu *et al.*, *Nat. Commun.* **6**, 6438 (2015).
16. A. Alexandradinata *et al.*, arXiv 2010.00584 [cond-mat.str-el] (1 October 2020).
17. O. Cyr-Choinière *et al.*, *Phys. Rev. B* **98**, 064513 (2018).
18. S. M. Souliou *et al.*, *Phys. Rev. B* **97**, 020503 (2018).
19. J. Chang *et al.*, *Nat. Phys.* **8**, 871–876 (2012).
20. H. H. Kim *et al.*, *Science* **362**, 1040–1044 (2018).
21. M. Bluschke *et al.*, *Nat. Commun.* **9**, 2978 (2018).
22. R. Arpaia, E. Andersson, E. Trabaldo, T. Bauch, F. Lombardi, *Phys. Rev. Mater.* **2**, 024804 (2018).
23. See supplementary materials.
24. R. Arpaia *et al.*, *Phys. Rev. Mater.* **3**, 114804 (2019).
25. S. Nawaz, R. Arpaia, F. Lombardi, T. Bauch, *Phys. Rev. Lett.* **110**, 167004 (2013).
26. Y. Ando, K. Segawa, S. Komiya, A. N. Lavrov, *Phys. Rev. Lett.* **88**, 137005 (2002).
27. I. M. Vishik *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18332–18337 (2012).
28. R. Arpaia *et al.*, *Phys. Rev. B* **96**, 064525 (2017).
29. J. D. Jorgensen *et al.*, *Phys. Rev. B* **41**, 1863–1877 (1990).
30. F. Liu, M. R. Press, S. N. Khanna, P. Jena, *Phys. Rev. B* **39**, 6914–6924 (1989).
31. E. Abrahams, C. M. Varma, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5714–5716 (2000).
32. S. A. Kivelson *et al.*, *Rev. Mod. Phys.* **75**, 1201–1241 (2003).
33. A. V. Maharaj, I. Esterlis, Y. Zhang, B. J. Ramshaw, S. A. Kivelson, *Phys. Rev. B* **96**, 045132 (2017).
34. Y. Gannot, B. J. Ramshaw, S. A. Kivelson, *Phys. Rev. B* **100**, 045128 (2019).
35. H. Yao, D. H. Lee, S. Kivelson, *Phys. Rev. B* **84**, 012507 (2011).
36. Y. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 68–72 (2020).
37. J. Wu, A. T. Bollinger, X. He, I. Božović, *Nature* **547**, 432–435 (2017).
38. D. Leboeuf *et al.*, *Phys. Rev. B* **83**, 054506 (2011).
39. P. Bourgeois-Hope *et al.*, arXiv 1910.08126 [cond-mat.supr-con] (17 October 2019).
40. S. Komiya, Y. Ando, *Phys. Rev. B* **70**, 060503 (2004).
41. S. Ono *et al.*, *Phys. Rev. Lett.* **85**, 638–641 (2000).
42. S. Blanco-Canosa *et al.*, *Phys. Rev. B* **90**, 054513 (2014).
43. R. Arpaia *et al.*, *Science* **365**, 906–910 (2019).
44. H.-H. Kim *et al.*, *Phys. Rev. Lett.* **126**, 037002 (2021).
45. L. Delacrétaz, B. Goutéraux, S. Hartnoll, A. Karlsson, *SciPost Phys.* **3**, 025 (2017).
46. G. Seibold *et al.*, *Commun. Phys.* **4**, 7 (2021).
47. A. A. Sinchenko, P. D. Grigoriev, P. Lejay, P. Monceau, *Phys. Rev. Lett.* **112**, 036601 (2014).
48. E. Wahlberg *et al.*, Restoring the strange metal phase via suppression of charge density waves in an underdoped cuprate superconductor; https://doi.org/10.5281/zenodo.4724109 (2021).

ACKNOWLEDGMENTS

We thank M. Grilli, C. Di Castro, and M. Moretti for enlightening discussions, D. Betto for initial tests at ID32 on ultrathin YBCO films, and L. Martinelli for helping with complementary RIXS measurements. This work was performed in part at Myfab Chalmers. The RIXS experimental data were collected at beamline ID32 of the European Synchrotron (ESRF) in Grenoble, France, using the ERIXS spectrometer designed jointly by the ESRF and Politecnico di Milano. **Funding:** Supported by the European Union's projects NANOCOHYBRI (Cost Action CA16218) and OXiNEMS (Horizon 2020, grant agreement 828784), the Knut and Alice Wallenberg foundation (project 2014.0102), and the Swedish Research Council (VR) under projects 2015-04368 (U.G.), 2018-04658 (F.L.), 2020-04945 (R.A.) and 2020-05184 (T.B.); Deutsche Forschungsgemeinschaft grant SE 806/19-1 (G.S.); Italian Ministry of University and Research PRIN project 2017Z8TS5B (G.G. and S.C.); and the University of Rome Sapienza through projects Ateneo 2018 (grant RM11816431DBA5AF), Ateneo 2019 (grant RM11916B56802AFE), and Ateneo 2020 (grant RM120172A8CC7CC7) (S.C.). **Author contributions:** F.L., E.W., R.A., and T.B. conceived and designed the experiments; E.W., R.A., and E.T. grew the samples; E.W. performed the transport measurements; R.A., E.W., F.L., T.B., M.R., R.F., G.G., and N.B.B. performed the RIXS measurements; R.A., G.G., and L.B. analyzed and interpreted the RIXS experimental data; F.L., E.W., R.A., and T.B. interpreted the transport measurements with the theoretical insights of U.G.; G.S. and S.C. performed the modeling of the experimental data and the theoretical calculations; and F.L., R.A., E.W., U.G., and T.B. wrote the manuscript with contributions from all authors. **Competing interests:** The authors declare no competing financial interests. G.G. is member of the Science Advisory Council of the ESRF. **Data and materials availability:** All experimental data shown in the main text and in the supplementary materials are accessible at the Zenodo repository (*48*).

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abc8372
Materials and Methods
Supplementary Text
Figs. S1 to S9
References (*49*–*64*)

21 May 2020; resubmitted 8 February 2021
Accepted 22 July 2021
10.1126/science.abc8372

REPORTS

QUANTUM OPTICS

Entanglement transport and a nanophotonic interface for atoms in optical tweezers

Tamara Đorđević[1]†, Polnop Samutpraphoot[1]†, Paloma L. Ocola[1]†, Hannes Bernien[2], Brandon Grinkemeyer[1], Ivana Dimitrova[1], Vladan Vuletić[3], Mikhail D. Lukin[1]*

The realization of an efficient quantum optical interface for multi-qubit systems is an outstanding challenge in science and engineering. Using two atoms in individually controlled optical tweezers coupled to a nanofabricated photonic crystal cavity, we demonstrate entanglement generation, fast nondestructive readout, and full quantum control of atomic qubits. The entangled state is verified in free space after being transported away from the cavity by encoding the qubits into long-lived states and using dynamical decoupling. Our approach bridges quantum operations at an optical link and in free space with a coherent one-way transport, potentially enabling an integrated optical interface for atomic quantum processors.

Systems of neutral atoms have recently emerged as a powerful platform for quantum simulations and quantum computation (*1*–*4*). Atom trapping, coherent control, and readout are enabled by optical beam arrays, whereas high-fidelity multi-qubit quantum gate operations are enabled either through tunneling (*5*) or coherent excitation into atomic Rydberg states (*6*). Many potential applications of these systems, from quantum networking to modular quantum computing, require a mechanism for transferring qubit states to optical photons without compromising the ability for multi-qubit control (*7*, *8*). Integration with a quantum optical interface, such as an optical cavity with switchable interactions, would enable the required atom-photon entanglement generation and distribution (*9*–*11*). A promising approach to achieve this is by spatially separating the cavity from the array and transporting selected atoms to and from the field as needed, preserving the coherence and scalability of atom array systems (Fig. 1A).

We demonstrate a system with a micrometer-size photonic crystal cavity (PCC) coupled to neutral atoms in movable optical tweezers that can perform high-fidelity single-qubit control and measurements as well as generate two-atom entanglement. Following the prior work that demonstrated coupling of atoms to photonic crystal structures (*12*–*14*), entangling atoms in macroscopic cavities (*15*, *16*), and coherently transporting atoms and ions in free space (*17*–*19*), we demonstrate PCC-mediated entanglement

[1]Department of Physics, Harvard University, Cambridge, MA 02138, USA. [2]Pritzker School of Molecular Engineering, University of Chicago, Chicago, IL 60637, USA. [3]Department of Physics and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author. Email: lukin@physics.harvard.edu
†These authors contributed equally to this work.

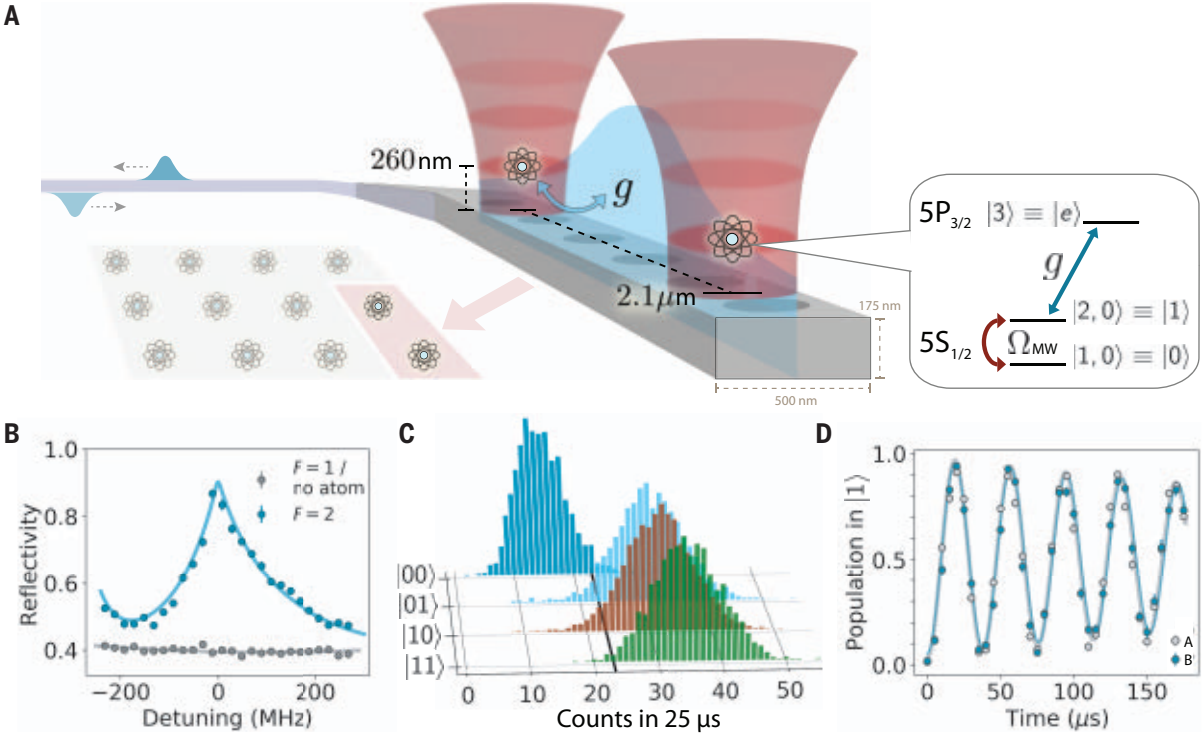


Fig. 1. Nanophotonic interface for atoms in tweezers. (**A**) A nanophotonic cavity could act as a photonic link for atom arrays serving as quantum information processors. Our system consists of two atoms in optical tweezers coupled to a fiber-linked cavity capable of a coherent one-way transport. (Inset) Atomic qubit states coupled by a microwave drive Ω_{MW}, with one coupled by the cavity field g to an excited state. (**B**) Cavity reflection spectrum for an atom in a strongly coupled state (blue) and an uncoupled state (gray), described by $(2g, \lambda, \kappa) = 2\pi \times (786, 6, 3800)$ MHz. (**C**) Readout of different two-qubit basis states by using the reflected photon counts. State $|00\rangle$ is distinguished from the others with fidelity $\mathcal{F} = 0.95(3)$. The atoms, cavity, and probe are on resonance. (**D**) Single-qubit Rabi oscillations in the cavity near-field showing preparation and coherent rotation with a π pulse fidelity $\mathcal{F}_\pi = 0.946(3)$.

generation that can be preserved and verified after transporting the atoms away from the cavity near-field. This is achieved by engineering a compact high-cooperativity photonic structure, establishing full quantum control in the near-field, encoding the qubits into field-insensitive states, and using dynamical decoupling sequences synchronized with the atomic transport.

In our experiments, two ^{87}Rb atoms were trapped in individually controlled optical tweezers and coupled to a PCC attached to a tapered optical fiber, enabling a low-loss photonic link out of the vacuum chamber (Fig. 1A) (*20*). The small cross-sectional area of the PCC (<0.1 μm^2) resulted in a small mode volume $V_m \sim 0.4\lambda^3$ that enabled high atom-photon cooperativity (*13*) and allowed for switching of the photon interactions by moving the atoms over ~1 μm in and out of the cavity near-field. Atoms at the PCC were trapped 260 nm above the surface by a standing wave created with the retro-reflection of the tweezer (*11*, *14*, *21*, *22*). Loading into the first lattice site of the standing wave was accompanied by a trapping potential deformation that resulted in an ~50% loading probability, which was increased to 80% by performing polarization gradient cooling during the transport to the PCC (*22*). [An efficiency above 90% without cooling during transport has been achieved with atoms initialized in the motional ground state (*21*).]

The cavity reflectivity strongly depends on the internal atomic state, with qubits encoded in $|0\rangle \equiv |F = 1, m_F = 0\rangle$ and $|1\rangle \equiv |F = 2, m_F = 0\rangle$. The PCC was tuned to be resonant with the $5S_{1/2}$, $F = 2 \rightarrow 5P_{3/2}$, $F' = 3$ optical transition and addresses all the m_F levels within these manifolds owing to the large cavity width κ. The reflection spectrum of a single atom is characterized by the extracted cooperativity $C = 4g^2/(\kappa\lambda) = 27(1)$ with a fit that accounts for thermal averaging (numbers in parentheses indicate 1 SD) (Fig. 1B) (*14*). When probed on resonance, the difference in reflectivity between an atom in the coupled $F = 2$ and uncoupled $F = 1$ manifold was used to perform a nondestructive single-shot readout. With two atoms, the uncoupled state $|00\rangle$ is distinguished from the other two-qubit states with high fidelity (Fig. 1C). This was used to determine the full two-atom state in all measurements at the PCC. For experiments in free space, readout was performed with state-selective push-out from the trap followed by fluorescence imaging. In both of these regimes, each pair of atoms was simultaneously prepared into $|0\rangle$ by using Raman-assisted state preparation (*22*) and addressed with a global microwave drive to perform single-qubit gates (Fig. 1D).

To coherently move the qubits from the near-field into free space, a millisecond coherence time was necessary. Choosing Zeeman-insensitive hyperfine sublevels as the qubit basis results in coherence times of $T'_2 = 2.1(1)$ ms and 9.7(8) ms in the near-field and free space, respectively, measured with a spin-echo pulse sequence (Fig. 2A). In the field-insensitive basis, the coherence is typically limited by the thermal motion that samples varying differential ac Stark shifts in the trap (*23*, *24*), and our measured T'_2 in free space is consistent with this model, given the trap parameters. The T'_2 at the PCC is likely further reduced by heating intrinsic to the near-field trapping geometry (*22*, *25*).

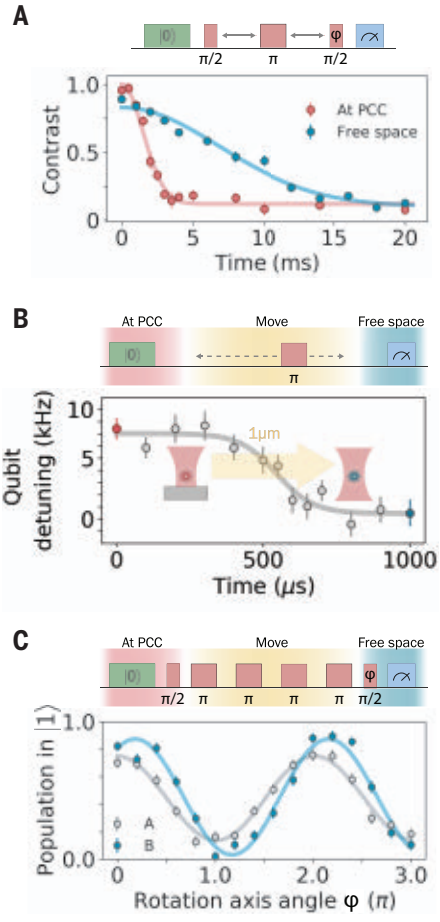


Fig. 2. Qubit coherence during the transport between the cavity and free space. (**A**) Measurement of the echo T'_2 for stationary traps at the PCC (red) and in free space (blue). (**B**) Changes in the trapping potential configuration during transport cause differences in the light shift and therefore qubit frequency, which can lead to additional dephasing. The gray line indicates fit to the programmed ramp. (**C**) Single-atom coherence for atoms in each trap (labeled "A" and "B") after transport measured in a Ramsey-type sequence. A four-pulse Carr-Purcell sequence symmetric with respect to the qubit frequency profile during transport was used to echo out the light shift variations.

Atomic coherence during transport is additionally affected by the change in differential ac Stark shift when moving between different trapping configurations. To observe this change, a detuned microwave π pulse was applied to single atoms at various times along the ~1-μm path from the PCC, where the measured population transfer is proportional to the instantaneous qubit frequency (*22*). From the resulting ramp profile in Fig. 2B, we obtained the minimal time of 650 μs to move from the PCC, limited by the mechanical response bandwidth of the scanning mirrors. Dephasing caused by fluctuations in the atom trajectory can be circumvented by using the Carr-Purcell decoupling pulse sequence (Fig. 2C) (*26*). This sequence was tested with the atom beginning in a superposition at the PCC and followed by $N \times \pi$ pulses along its transport, which performed better for an even N because of the symmetry of the ramp profile shown in Fig. 2B. The performance of the decoupling sequence was characterized separately in the two optical tweezers (trap A and trap B): For an optimal $N = 4$, an atom in trap B retains 0.88(5) of its initial coherence, limited by the move time and T'_2 (Fig. 2C). The lower contrast of 0.64(3) for an atom in trap A is ascribed to extra fluctuations in the scanning mirror, given that the traps are otherwise uniform. In addition to decoherence, each atom has an ~15% chance of being lost on the way out of the PCC, which is corrected for in the free space readout (*22*).

To generate an entangled state of atoms using the cavity, we performed a protocol known as "cavity carving" that projectively measures a classical superposition state while maintaining qubit indistinguishability (*27*, *28*), which can result in a maximally entangled Bell state (*16*, *29*). To implement the protocol, we initialized the atoms in $|00\rangle$, rotated by an angle θ to create the state

$$|\Psi_0\rangle = \cos^2\left(\frac{\theta}{2}\right)|00\rangle - \sin^2\left(\frac{\theta}{2}\right)|11\rangle - i\sin\left(\frac{\theta}{2}\right)\cos\left(\frac{\theta}{2}\right)(|01\rangle + |10\rangle) \quad (1)$$

We sent a weak coherent pulse into the cavity mode [$\bar{n} = 0.35$, chosen for negligible scattering error (*22*)] and conditioned on the detection of a reflected photon to project the state $|00\rangle$ out of the two-atom Hilbert space. The angle θ was chosen to maximize the overlap of the resulting state with $|\Psi^+\rangle = (|01\rangle + |10\rangle)/\sqrt{2}$, minimizing the population in $|11\rangle$ while accounting for the finite reflectivity of $|00\rangle$. For our system, the $|00\rangle$ reflectivity is already less than that of the other two-atom states (Fig. 1C). It was further reduced by an interferometer in the detection path and is ultimately limited by the interferometric stability (*22*). After the projection, an additional $\pi/2$ pulse was applied to

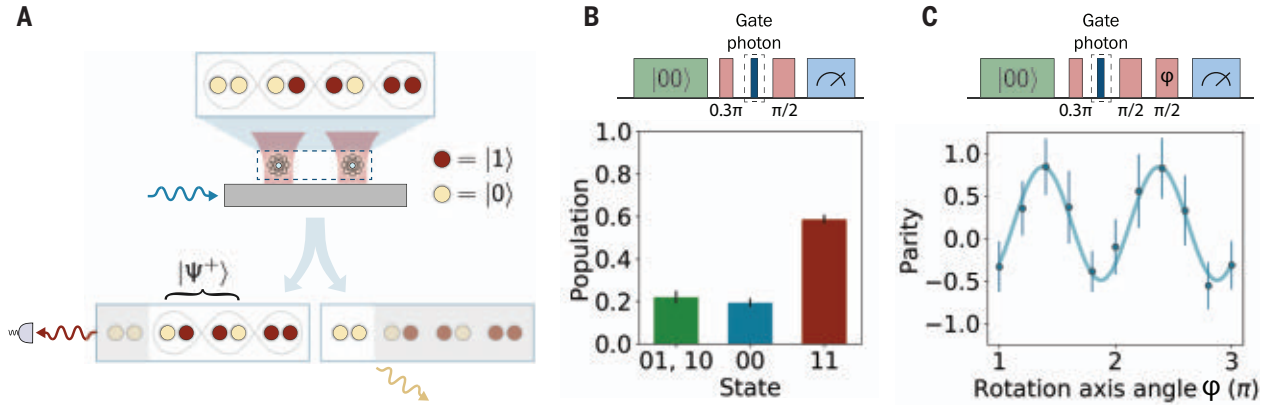


Fig. 3. Heralded entanglement at the cavity. (**A**) Protocol description. Postselection on a photon reflection from a superposition state carves out the $|00\rangle$ component, leaving the atoms predominantly in $|\Psi^+\rangle$. (**B**) Populations of the resulting state in different basis states. A π/2 pulse was applied before the measurement, converting the Bell state $|\Psi^+\rangle$ into $|\Phi^+\rangle$. {P_{11}, P_{00}} is extracted {with, without} an additional global π pulse before cavity reflection readout. P_{01} and P_{10} cannot be distinguished. (**C**) Parity oscillations of the $|\Phi^+\rangle$ state after rotating over a varied axis ϕ, showing nonclassical correlations between the qubits in all bases. The coherence $\rho_{11,00}$ is extracted from the amplitude. The combined coherence and populations give a Bell state fidelity of $\mathcal{F} = 0.72(3)$.

rotate the Bell state $|\Psi^+\rangle$ into $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ for ease of characterization.

The entangled state was first characterized in situ by using the cavity reflection readout shown in Fig. 1C. The $|\Phi^+\rangle$ fidelity $\mathcal{F} = \langle\Phi^+|\rho|\Phi^+\rangle = \frac{1}{2}(\rho_{00,00} + \rho_{11,11}) + \mathrm{Re}(\rho_{00,11})$ was determined by measuring populations and extracting coherence from the parity oscillation contrast upon another π/2 rotation over different axes (*16*, *22*). Populations are plotted in Fig. 3B with $\mathcal{P} = P_{00} + P_{11} = 0.78(3)$; the imbalance between $|00\rangle$ and $|11\rangle$ is discussed and modeled in the supplementary materials (*22*). The parity oscillation in Fig. 3C corresponds to a coherence of $\rho_{11,00} = 0.33(2)$. These measurements give a fidelity of $\mathcal{F} = 0.72(3)$ and a concurrence bound of $\mathcal{C} \geq 0.46(5)$ (*22*). The major part of the error (24%) is accounted for by the chosen initial angle θ for our protocol, with an additional 4% from the initialization imperfections. Although our approach uses a single gate photon, heralding on repeated gates can improve the entanglement fidelity at the expense of the success probability (*16*).

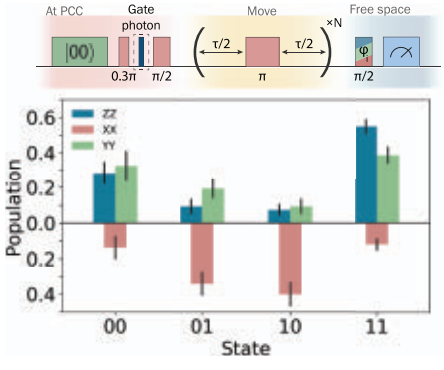


Fig. 4. Coherent transport of entangled atoms from the cavity to free space. After entangling the atoms at the cavity, the tweezers are moved away while applying dynamical decoupling. Free space readout in different bases after transport shows the remaining entanglement fidelity of $\mathcal{F} = 0.65(6)$, which is larger than the classical limit of 0.5.

Last, the entangled atoms were transported away from the cavity while applying decoupling pulses. The qubit state populations were measured in the *ZZ*, *XX*, and *YY* bases (Fig. 4), giving the population $\mathcal{P} = 0.78(6)$ and the coherence $\rho_{11,00} = 0.26(5)$, corrected for the atomic loss. This resulted in a fidelity of $\mathcal{F} = 0.65(6)$ and a lower bound to the concurrence of $\mathcal{C} \geq 0.35(11)$ [$\mathcal{F}_u = 0.52(5)$ and $\mathcal{C}_u \geq 0.04(10)$ without the readout correction]. Both $\mathcal{F} > 0.5$ and $\mathcal{C} > 0$ certify entanglement after transport at a 99% confidence level. The state fidelity is consistent with the decoherence observed for single atoms during transport (Fig. 2C), limited by atomic thermal motion and uncorrelated fluctuations in the programmed trajectories. The trajectory fluctuations can be reduced in the future by using an acousto-optic deflector (AOD) that would speed up the transport and enable atomic displacements of more than 100 μm, which is sufficient for high-fidelity Rydberg state control (*30*).

These results demonstrate that combining movable atoms in tweezers with nanophotonics constitutes a promising platform for the development of a quantum optical node for large-scale atomic quantum processors. If integrated with a free space atom array, our system could enable quantum teleportation of the atom array states to optical photons (*31*) or remote entanglement of arrays. This can be accomplished by using the demonstrated one-way transport, in which the atoms that are entangled with a photon at the PCC are then coherently moved to the free space array to interact through Rydberg gates. Extending our demonstration to a general-purpose quantum optical interface with a coherent two-way transport could enable feed-forward protocols such as quantum error correction by means of fast, high-fidelity quantum nondemolition cavity readout. Furthermore, with multiple atoms coupled to nanophotonic waveguides, entangled and arranged into arbitrary spatial configurations, important classes of many-particle quantum states can be created, such as atomic cluster or graph states (*32*), large-scale photonic Fock states (*33*), or Schrödinger cat states (*34*). Last, telecom transitions present in both alkaline-earth and alkali atoms (*35*, *36*) could extend atom-nanophotonic quantum networking to long distances, which is crucial for quantum communication and the scale-up of processing power (*8*).

REFERENCES AND NOTES

1. M. Saffman, T. G. Walker, K. Mølmer, *Rev. Mod. Phys.* **82**, 2313–2363 (2010).
2. S. Ebadi *et al.*, *Nature* **595**, 227–232 (2021).
3. P. Scholl *et al.*, *Nature* **595**, 233–238 (2021).
4. I. S. Madjarov *et al.*, *Nat. Phys.* **16**, 857–861 (2020).
5. A. M. Kaufman *et al.*, *Science* **345**, 306–309 (2014).
6. H. Levine *et al.*, *Phys. Rev. Lett.* **123**, 170503 (2019).
7. N. H. Nickerson, Y. Li, S. C. Benjamin, *Nat. Commun.* **4**, 1756 (2013).
8. C. Monroe *et al.*, *Phys. Rev. A* **89**, 022317 (2014).
9. T. Northup, R. Blatt, *Nat. Photonics* **8**, 356–363 (2014).
10. A. Reiserer, G. Rempe, *Rev. Mod. Phys.* **87**, 1379–1418 (2015).
11. T. G. Tiecke *et al.*, *Nature* **508**, 241–244 (2014).
12. J. D. Hood *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10507–10512 (2016).
13. D. E. Chang, J. S. Douglas, A. González-Tudela, C.-L. Hung, H. J. Kimble, *Rev. Mod. Phys.* **90**, 031002 (2018).
14. P. Samutpraphoot *et al.*, *Phys. Rev. Lett.* **124**, 063602 (2020).
15. S. Welte, B. Hacker, S. Daiss, S. Ritter, G. Rempe, *Phys. Rev. X* **8**, 011018 (2018).
16. S. Welte, B. Hacker, S. Daiss, S. Ritter, G. Rempe, *Phys. Rev. Lett.* **118**, 210503 (2017).
17. J. Beugnon *et al.*, *Nat. Phys.* **3**, 696–699 (2007).
18. A. Lengwenus, J. Kruse, M. Schlosser, S. Tichelmann, G. Birkl, *Phys. Rev. Lett.* **105**, 170502 (2010).
19. J. M. Pino *et al.*, *Nature* **592**, 209–213 (2021).
20. T. G. Tiecke *et al.*, *Optica* **2**, 70 (2015).
21. J. D. Thompson *et al.*, *Science* **340**, 1202–1205 (2013).

22. Materials and methods are available as supplementary materials.
23. S. Kuhr *et al.*, *Phys. Rev. A* **72**, 023406 (2005).
24. D. Reitz, C. Sayrin, R. Mitsch, P. Schneeweiss, A. Rauschenbeutel, *Phys. Rev. Lett.* **110**, 243603 (2013).
25. D. Hümmer, P. Schneeweiss, A. Rauschenbeutel, O. Romero-Isart, *Phys. Rev. X* **9**, 041034 (2019).
26. H. Y. Carr, E. M. Purcell, *Phys. Rev.* **94**, 630–638 (1954).
27. W. Chen *et al.*, *Phys. Rev. Lett.* **115**, 250502 (2015).
28. G. Barontini, L. Hohmann, F. Haas, J. Estève, J. Reichel, *Science* **349**, 1317–1321 (2015).
29. A. S. Sørensen, K. Mølmer, *Phys. Rev. Lett.* **90**, 127903 (2003).
30. H. Kübler, J. P. Shaffer, T. Baluktsian, R. Löw, T. Pfau, *Nat. Photonics* **4**, 112–116 (2010).
31. C. H. Bennett *et al.*, *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
32. H. Choi, M. Pant, S. Guha, D. Englund, *npj Quantum Inf.* **5**, 104 (2019).
33. V. Paulisch, M. Perarnau-Llobet, A. González-Tudela, J. I. Cirac, *Phys. Rev. A (Coll. Park)* **99**, 043807 (2019).
34. E. J. Davis, Z. Wang, A. H. Safavi-Naeini, M. H. Schleier-Smith, *Phys. Rev. Lett.* **121**, 123602 (2018).
35. J. P. Covey *et al.*, *Phys. Rev. Appl.* **11**, 034044 (2019).
36. S. G. Menon, K. Singh, J. Borregaard, H. Bernien, *New J. Phys.* **22**, 073033 (2020).

ACKNOWLEDGMENTS

We thank C. Senko, H. Levine, D. Bluvstein, R. Riedinger, M. Bhaskar, D. Levonian, J. Borregaard, F. Reiter, S. Schwartz, J. Thompson, and A. Zibrov for useful discussions and experimental contributions. **Funding:** This work was supported by the Center for Ultracold Atoms, the National Science Foundation, Air Force Office of Scientific Research Multidisciplinary University Research Initiative, Vannevar Bush Faculty Fellowship, US Department of Energy (DOE) QSA Center (contract 7568717), and Army Research Laboratory CDQI. B.G. acknowledges support from the US Department of Defense (DOD) NDSEG. The device was fabricated at the Harvard Center for Nanoscale Systems (CNS) (NSF ECCS-1541959). **Author contributions:** T.Đ., P.S., P.L.O., and H.B. built the experimental setup, performed the measurements, and analyzed data. B.G. and I.D. assisted with experiments. All work was supervised by V.V. and M.D.L. All authors discussed the results and contributed to the manuscript. **Competing interests:** V.V. and M.D.L. are cofounders and shareholders of QuEra Computing. The other authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and the supplementary materials.

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abi9917
Materials and Methods
Figs. S1 to S10
Table S1
References (*37*–*48*)

29 April 2021; accepted 29 July 2021
10.1126/science.abi9917

TOPOLOGICAL OPTICS

Topological insulator vertical-cavity laser array

Alex Dikopoltsev[1]†, Tristan H. Harder[2]†, Eran Lustig[1], Oleg A. Egorov[3], Johannes Beierlein[2], Adriana Wolf[2], Yaakov Lumer[1], Monika Emmerling[2], Christian Schneider[4], Sven Höfling[2], Mordechai Segev[1]*, Sebastian Klembt[2]*

Topological insulator lasers are arrays of semiconductor lasers that exploit fundamental features of topology to force all emitters to act as a single coherent laser. In this study, we demonstrate a topological insulator vertical-cavity surface-emitting laser (VCSEL) array. Each VCSEL emits vertically, but the in-plane coupling between emitters in the topological-crystalline platform facilitates coherent emission of the whole array. Our topological VCSEL array emits at a single frequency and displays interference, highlighting that the emitters are mutually coherent. Our experiments exemplify the power of topological transport of light: The light spends most of its time oscillating vertically, but the small in-plane coupling is sufficient to force the array of individual emitters to act as a single laser.

Topological insulators are systems that exhibit edge conductance that is protected against disorder and defects. Initially, this concept was related to electrons in solids, but this plentiful phenomenon was recently demonstrated in many wave systems such as photonics, cold atoms, and acoustics. Several years after establishing that topological properties can endow the transport of light with immunity against scattering from imperfections, the use of topological platforms to force injection locking of many semiconductor laser emitters was suggested. It was proposed (*1*, *2*) that the proper design of a multilaser array can lead to an edge mode flowing from one laser to another; force injection locking, despite some inevitable variations in the resonators; and eventually make all the elements lase as a single coherent source. This idea was demonstrated (*3*, *4*) in an array of asymmetrically coupled ring resonators (*5*), with standard semiconductor laser technology. Sequentially topological lasing using magneto-optic materials was demonstrated (*6*). Since then, topological insulator lasers were demonstrated in a variety of configurations (*7*, *11*), for example, a topological quantum cascade laser with valley edge modes (*7*), a topological bulk laser based on band inversion–induced reflection (*8*), a spin-momentum-locked topological laser (*9*), and a topological insulator laser with next-nearest-neighbor coupling (*10*). Moreover, recent theoretical work has shown the dramatically reinforced coherence of topological lasers compared with the corresponding nontopological devices (*12*). The topological insulator laser is thus a promising application of topological photonics, and many new ideas have emerged, for example, using topology in synthetic dimensions to force an array of semiconductor lasers to emit mode-locked pulses (*13*).

The main reasons why topological properties make this possible are as follows. First, the topological edge mode in all topological insulator systems has a nonzero group velocity; hence, it must flow between lasers. As such it forces injection locking of all emitters associated with the edge mode while suppressing

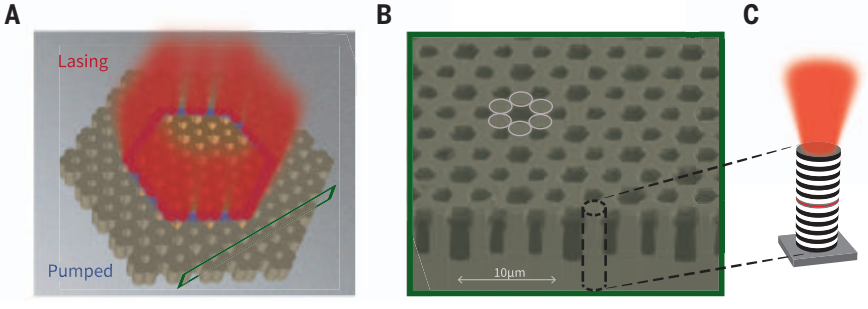


Fig. 1. Topological insulator VCSEL array. (**A**) Topological insulator VCSEL array, composed of two types of lattices: compressed and stretched honeycomb lattices, on either side of the topological interface. When the array is pumped on the interface (blue), the 30 VCSELs on the topological edge are injection-locked and lase vertically and coherently. (**B**) Scanning electron microscope image of the fabricated structure. Each pillar is a vertical microcavity, and the planar arrangement is based on the crystalline model with a six-pillar unit cell (white circles). (**C**) A single VCSEL: gain medium (red layer) sandwiched between two DBRs. Light emission is vertical.

[1]Physics Department, Technion, 32000 Haifa, Israel. [2]Technische Physik, Wilhelm-Conrad-Röntgen-Research Center for Complex Material Systems, and Würzburg-Dresden Cluster of Excellence ct.qmat, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany. [3]ITFO, Abbe Center of Photonics, Friedrich-Schiller-Universität Jena, D-07743 Jena, Germany. [4]Institute of Physics, University of Oldenburg, D-26129 Oldenburg, Germany.
***Corresponding author. Email: msegev@technion.ac.il (M.S.); sebastian.klembt@uni-wuerzburg.de (S.K.)**
†These authors contributed equally to this work.

independent lasing of localized groups of resonators. Second, the topological platform guarantees robustness against small differences between resonators that would reduce the efficiency of injection locking.

Forcing injection locking of a large array of lasers has been a challenge for decades. The first attempts to combine several semiconductor lasers used evanescently coupled one-dimensional (1D) arrays of edge emitters (*14*, *15*). However, these arrays emitted light with unstable modal structure, and their mutual coherence diminished when the number of emitters was larger than just a few lasers (*16*–*19*). A decade later, 2D arrays of evanescently coupled vertical-cavity surface-emitting lasers (VCSELs) were investigated (*20*, *21*). However, despite tour-de-force attempts to make coherent arrays of many VCSELs (*22*–*24*), large VCSEL arrays are now being used only in a mutually incoherent form. This means that such VCSEL arrays can be used, for example, to pump solid-state lasers but not for any application requiring coherence, interference, imaging, and the like. Obviously, finding a way to make large arrays of VCSELs act together as a single coherent source would enable applications that are beyond current technology.

Our topological insulator VCSEL array is designed to form a closed topological interface containing 30 emitters (Fig. 1). The structure is fabricated by molecular beam epitaxial growth of a quantum-dot gain medium (*25*) embedded in a single wavelength cavity layer sandwiched between two distributed Bragg reflectors (DBRs). The lattices are defined using electron-beam lithography and subsequent dry etching. Each VCSEL is a circular micropillar, a shape that, in principle, can support more than one polarization-dependent mode. However, as the diameter decreases in size, the spectral spacing between the modes increases, and the VCSEL tends to lase in a single polarization (*26*, *27*). We use the small diameter d = 2.5 μm and lattice constant $3a$ = 6.5 μm [for details, see (*28*)]. The VCSELs are ordered in the topological geometry known as the "crystalline" model (*29*–*33*) (Fig. 2A), where the topological nature arises from different geometries on both sides of an interface between "stretched" and "compressed" honeycomb lattices. Both lattices are achieved by stretching or compressing the distances inside a unit cell of six sites. In the compressed lattice, the six pillars in each unit cell are separated by radius r_1 from the center, where $r_1 < a$ (a is the distance in the honeycomb before the compression); in the stretched lattice, the distances are $r_2 > a$. The stretching and compression open bandgaps at the Dirac points of the underlying honeycomb lattice, and each lattice has a different topological invariant. The interface therefore supports topological edge states [see explanation and fig. S1 in (*28*)]. Figure 2B presents such a topological state that resides on the interface. To view the band structure of the two lattice bulks and of the topological interface, we obtain their Fourier space luminescence spectra along the K-Γ-K′ direction of the underlying honeycomb lattice (Fig. 2, C to E). When the bandgap of the "stretched" bulk array is aligned with the bandgap of the "compressed" bulk array, the interface exhibits a 553-μeV-wide topological gap. As Fig. 2D shows, new modes appear at the interface—the topological edge modes, which are known to display topological protection properties (*30*). These topological edge modes have group velocities whose direction depends on their intra–unit cell pseudo-spin, which in closed interface paths may carry orbital angular momentum (*33*). As shown in (*28*), we calculate the band structure of the system by using a continuous model for the envelope of the light field in the micropillar photonic cavities. This model entails rich physics; for example, it describes well all the photonic states of the array, including photonic bands composed of higher modes within each pillar. Comparing experiments with simulations, we find good agreement, highlighting that the modes in the gap are edge modes with nonzero group velocity.

Our VCSEL array lases in a topological edge mode of 30 VCSELs located on the topological interface (Fig. 3A). To drive lasing of the edge mode, we pump the interface with a pulsed hexagon-shaped beam using a spatial light modulator and measure the lasing pattern and spectrum [for more details, see (*28*)]. By imaging the out-of-plane optical output, we find that lasing occurs precisely at the topological interface composed of the 30 VCSELs on the edge (Fig. 3B). The lasing intensity structure is not uniform because of inevitable fabrication imperfections and the nonuniform position of the quantum dots in the gain layer within each pillar. However, despite the nonuniform lasing, the topological platform forces all the emitters on the topological interface to lock and lase as a single laser (rather than lasing in several independent localized groups), as we show next. Figure 3C presents the emission spectrum below and above the lasing threshold. When we pump below the lasing threshold, the broad emission spectrum implies that multiple modes emit amplified spontaneous emission. When the pumping is above the lasing threshold, we observe single-mode lasing exactly at the wavelength of the topologically protected modes. The lasing topological mode maintains a narrow emission spectrum at a stable energy with a 249-μeV linewidth, which is less than half the width of the topological bandgap in our system (553 μeV). Moreover, the lasing linewidth of a single individual VCSEL in our array is ~170 μeV. This means that our 30-emitter VCSEL array lases with a linewidth almost as narrow as that of a single emitter in the array. Figure 3D shows

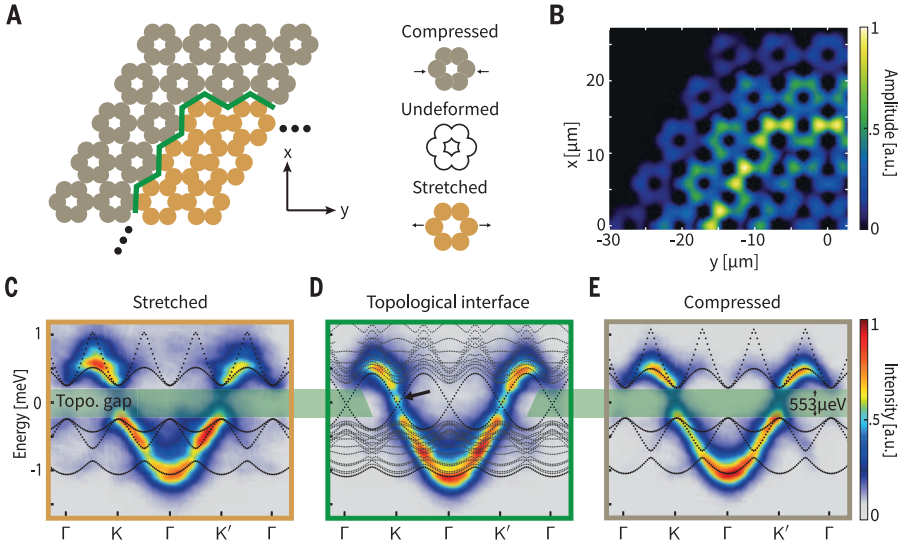


Fig. 2. Structure and photonic properties of the topological interface. (**A**) VCSEL array structure, composed of compressed (gray) and stretched (yellow) six-pillar unit cells of deformed honeycomb lattices, with uniform pillar-shaped VCSELs. The green line marks the topological interface. (**B**) Topologically protected mode centered on the interface between the two lattice types. a.u., arbitrary units. (**C** to **E**) Fourier-space photoluminescence spectra for the stretched bulk (C), topological interface (D), and the compressed bulk (E). The measured spectra are overlaid on the calculated band structure (black lines). The vertical axis is normalized so that 0 corresponds to 1.3003 eV, exactly the center of the topological gap. Note the measured photoluminescence peak from the photonic states in the topological gap [black arrow in (D)], exactly where the two bulk spectra [(C) and (E)] have bandgaps.

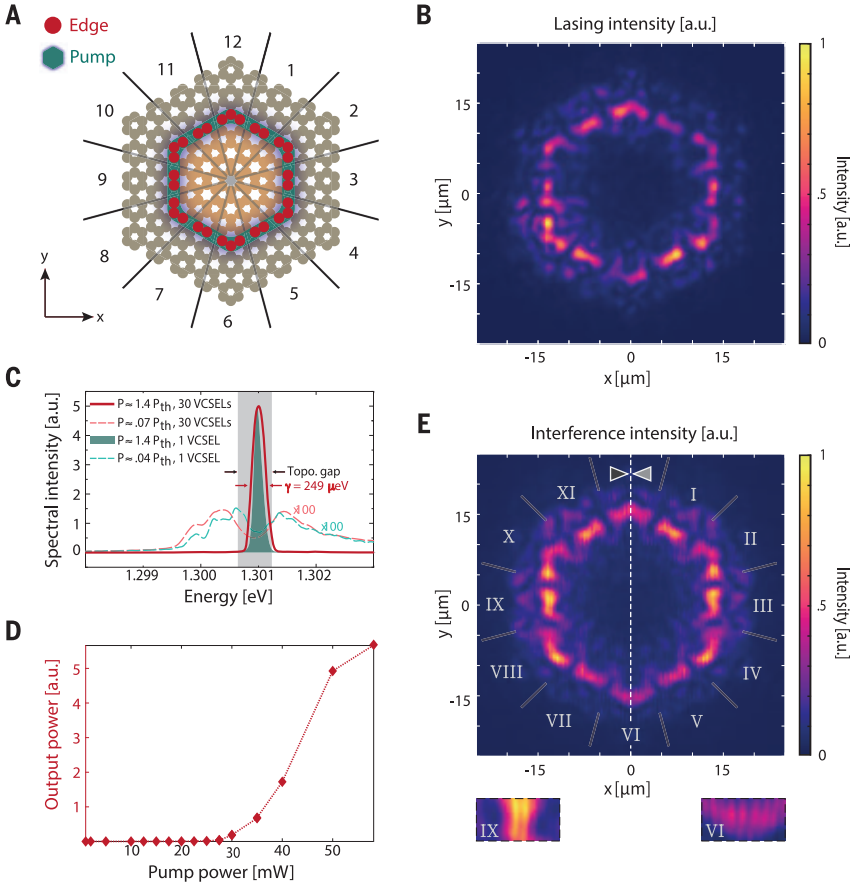


Fig. 3. Collective coherent lasing of the 30 VCSELs on the interface in the topological insulator array. (**A**) The topological VCSEL array. The inner bulk (yellow) (outer bulk, gray) is obtained by stretching (compressing) the honeycomb lattice. We pump the VCSELs on the topological interface (red) between the two bulks. (**B**) Lasing intensity of the topological edge mode. The shape includes corners and inevitable inhomogeneity, but the lasing is relatively uniform all along the topological edge. (**C**) Lasing spectrum of the topological array. The output spectra of a single VCSEL and the 30 VCSELs of the array are practically the same, below and above the lasing threshold. The measured lasing spectrum of the topological array (red line) is narrow (249 μeV) and fully contained within the topological gap (shaded region). (**D**) Output power versus pump power. (**E**) Interference between the lasing field and its mirror image, which measures coherence between lasers symmetrically located around $x = 0$. The fringes in region VI arise from interfering lasers in region 6 [of (A)] with themselves. The fringes in region IX show interference between the distinct lasers from regions 3 and 9 [of (A)]. Regions III and IX show interference fringes, even though emitters 3 and 9 are 13 VCSELs apart. The insets show zoom-ins of the fringes in regions VI and IX.

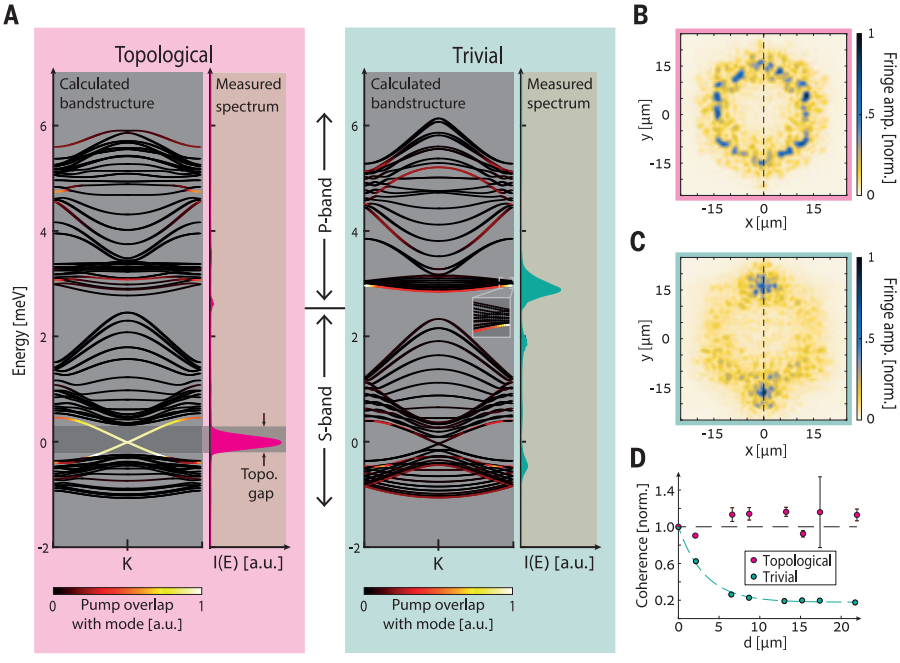


Fig. 4. Comparison between topological and trivial lasing in VCSEL arrays. (**A**) Calculated band structure of the photonic s- and p-bands of the topological and trivial arrays, along with the measured lasing spectra, I, as a function of energy E. The color of the bands indicates the overlap between the mode structure and the pump. Lasing in the topological array occurs exactly in the topological gap, where the group velocity is nonzero. This nonzero group velocity enables the injection locking of the VCSELs in the topological edge mode while suppressing lasing in localized groups of VCSELs. In the trivial array, lasing occurs in the lower edge of the p-band, where modes have group velocities approaching zero; hence, light does not circulate around the interface. Consequently, disorder causes lasing to occur in multiple independent groups of VCSELs, diminishing the mutual coherence. (**B** and **C**) Modulated part (fringe amplitude) of the mirror-interference measurements of the topological and trivial arrays, respectively. High modulation amplitude in each position is shown in (B), meaning that each of the 30 lasers of the topological interface is mutually coherent with its opposite partner. The graph in (C) displays coherence only near the mirror axis, which means that VCSELs that are far apart are not coherent with each other. The reason is lack of photonic transport in the trivial array, because the trivial edge modes have vanishing group velocity; hence, even weak disorder causes localized modes. (**D**) Modulation amplitude as a function of distance from the central VCSEL, averaged over all coherence paths and normalized to 1 at $d = 0$, extracted from (B) and (C). In the trivial case, coherence decays very rapidly, with a fit of ~3.5-μm decay length (dashed green). In the topological VCSEL array, high coherence is maintained around the whole interface. Error bars represent the standard deviation arising from local variations in fringe contrast.

the light output as a function of the input pump power, with a pumping threshold of $P_{th} \sim 25$ mW.

An important ingredient of coherence, especially when dealing with multi-emitters, is the interference between emitters located far apart from each other. To show the coherence of the lasing topological edge mode, we measure the interference of the output field with its mirror image around $x = 0$ (Fig. 3E) [setup described in (*28*)]. By doing this, we interfere the fields of every two VCSELs located on either side of the mirror axis. In this setting, the periodicity of the interference fringes arises mostly from the angle between the two arms of the interferometer. Specifically, the measured fringes in regions III and IX in Fig. 3E are the outcome of interference between VCSELs from region 3 and VCSELs from region 9 from Fig. 3A. The fringes in region VI in Fig. 3E are the result of the interference between VCSELs in region 6 and themselves. Importantly, the interference in regions III and IX is between VCSELs located 13 emitters and three corners apart from one another. This figure shows that every emitter interferes with its counterpart, showing that the coherence across the array is able to reestablish within each pumping cycle (even better results are expected with electrical pumping). The interference fringes are evidence that all 30 VCSELs act as a single coherent emitter.

Next, we compare the topological insulator VCSEL array with a trivial honeycomb VCSEL array (taking r_1, $r_2 \rightarrow a$) (Fig. 4) while exciting the trivial lattice with the same pump beam. Figure 4A shows the calculated band structure for the s- and p-bands, corresponding to the first and second modes of a single pillar, respectively. In both lattices, the measured spectrum corresponds to the modes with the maximum calculated overlap with the shaped pump beam. The measured spectrum of the topological array is localized in the topological gap, where the group velocity of the topological edge modes is high. By contrast, the spectrum of the trivial array is mainly located at the bottom of the p-band bulk, especially at the band edges where the group velocity goes to zero. The zero group velocity of the lasing modes in the trivial array means that localized groups of VCSELs tend to lase independently, in contrast to the topological array where the edge modes cannot be stationary. The wave in the topological mode has to circulate around the edge, and in doing that, it forces injection locking of all the VCSELs in that mode. It is therefore essential to examine the ability of VCSELs in the trivial array to interfere with one another, as we did for the topological array. Figure 4, B and C, shows the modulation amplitude extracted from the interference measurements, which is proportional to the mutual coherence of the array as a function of position. Comparing Fig. 4C with Fig. 4B, we notice a stark difference between the cases. The topological array shows coherence even for two lasers separated by half the interface length with three corners on its path, whereas the trivial honeycomb array results in coherence only around the autocorrelation points on the symmetry axis of the measurement, that is, only neighboring VCSELs show interference fringes. To compare the levels of coherence of the trivial and nontrivial arrays, we calculate the mean coherence as a function of the distance d from the VCSELs at $x = 0$ and normalize the coherence to 1 at $d = 0$ (where the central VCSEL interferes with itself) (Fig. 4D). The coherence of the topological insulator laser remains high along the whole path around the interface, whereas the trivial laser array shows the decay of coherence with increasing distance.

The topological VCSEL array separates between the vertical direction of lasing and the in-plane transport associated with the topological edge mode. This configuration endows the topological array with robustness for large variations of parameters. An immediate manifestation is shown in (*28*), where we examine the properties of the topological VCSEL array in a huge temperature range: between 4 and 200 K (fig. S7). The topological array also displays a narrow linewidth at 200 K, even though large temperature variations always change the refractive index and therefore the lasing wavelength. The linewidth of the topological VCSEL array remains narrow (~249 μeV) between 4 and 200 K (fig. S7C). This implies that the topological properties of this device are insensitive to changes in the wavelength. This happens because the topological VCSEL array scheme separates between the emission direction (where the DBRs and microcavity determine the wavelength) and the in-plane topological coupling, which are strongly coupled in all other schemes of topological insulator lasers.

The interference displayed in Fig. 3 confirms that our system is spatially coherent and that our topological 30-emitter VCSEL array acts as a single laser. The topologically protected transport, which is responsible for the injection locking of all the emitters, occurs in the plane of the chip—normal to the (vertical) direction of oscillation within each emitter. It is the topological design in which the VCSELs are positioned that forces all the emitters to act as a single laser. Research in topological photonics is now moving toward non-Hermitian and nonlinear topological phenomena, with the goal of using the properties of topological designs for new schemes of supreme semiconductor devices. The ability to separate between the direction of the topologically engineered platform and the oscillation direction within each emitter is a step toward this goal, offering the possibility to make large-scale coherent semiconductor laser arrays.

REFERENCES AND NOTES

1. G. Harari *et al*., in *Conference on Lasers and Electro-Optics (CLEO)* (OSA Technical Digest, Optical Society of America, 2016), paper FM3A.3.
2. G. Harari *et al*., *Science* **359**, eaar4003 (2018).
3. S. Wittek *et al*., in *Conference on Lasers and Electro-Optics (CLEO)* (OSA Technical Digest, Optical Society of America, 2017), paper FTh1D.3.
4. M. A. Bandres *et al*., *Science* **359**, eaar4005 (2018).
5. M. Hafezi, E. Demler, M. Lukin, J. Taylor, *Nat. Phys.* **7**, 907–912 (2011).
6. B. Bahari *et al*., *Science* **358**, 636–640 (2017).
7. Y. Zeng *et al*., *Nature* **578**, 246–250 (2020).
8. Z. K. Shao *et al*., *Nat. Nanotechnol.* **15**, 67–72 (2020).
9. Z.-Q. Yang, Z.-K. Shao, H.-Z. Chen, X.-R. Mao, R.-M. Ma, *Phys. Rev. Lett.* **125**, 13903 (2020).
10. Y. G. Liu, P. S. Jung, M. Parto, D. N. Christodoulides, M. Khajavikhan, *Nat. Phys.* **17**, 704–709 (2021).
11. J.-H. Choi *et al*., *Nat. Commun.* **12**, 3434 (2021).
12. I. Amelio, I. Carusotto, *Phys. Rev. X* **10**, 041060 (2020).
13. Z. Yang *et al*., *Phys. Rev. X* **10**, 011059 (2020).
14. D. E. Ackley, *Appl. Phys. Lett.* **42**, 152–154 (1983).
15. D. Botez, J. C. Connolly, *Appl. Phys. Lett.* **43**, 1096–1098 (1983).
16. S. S. Wang, H. G. Winful, *Appl. Phys. Lett.* **52**, 1774–1776 (1988).
17. H. G. Winful, S. S. Wang, *Appl. Phys. Lett.* **53**, 1894–1896 (1988).
18. R. K. DeFreez *et al*., *IEEE Photonics Technol. Lett.* **1**, 209–211 (1989).
19. H. G. Winful, *Phys. Rev. A* **46**, 6093–6094 (1992).
20. M. Orenstein *et al*., *Appl. Phys. Lett.* **58**, 804–806 (1991).
21. M. Orenstein, E. Kapon, J. P. Harbison, L. T. Florez, N. G. Stoffel, *Appl. Phys. Lett.* **60**, 1535–1537 (1992).
22. D. Francis, W. Yuen, H.-L. Chen, G. Li, C. Chang-Hasnain, *Electron. Lett.* **34**, 2132–2133 (1998).
23. F. Monti di Sopra *et al*., *Appl. Phys. Lett.* **77**, 2283–2285 (2000).
24. M. Miller, M. Grabherr, R. Jäger, K. J. Ebeling, *IEEE Photonics Technol. Lett.* **13**, 173–175 (2001).
25. H. Saito, K. Nishi, I. Ogura, S. Sugou, Y. Sugimoto, *Appl. Phys. Lett.* **69**, 3140–3142 (1996).
26. G. Edeltraud, O. Hess, *Spatio-Temporal Dynamics and Quantum Fluctuations in Semiconductor Lasers* (Springer, 2003).
27. P. Debernardi, G. P. Bava, C. Degen, I. Fischer, W. Elsasser, *IEEE J. Quantum Electron.* **38**, 73–84 (2002).
28. See supplementary materials.
29. L. Fu, *Phys. Rev. Lett.* **106**, 106802 (2011).
30. L. H. Wu, X. Hu, *Phys. Rev. Lett.* **114**, 223901 (2015).
31. G. Siroki, P. A. Huidobro, V. Giannini, *Phys. Rev. B* **96**, 041408 (2017).
32. Y. Yang *et al*., *Phys. Rev. Lett.* **120**, 217401 (2018).
33. X.-C. Sun, X. Hu, *Phys. Rev. B* **103**, 245305 (2021).

ACKNOWLEDGMENTS

M.S. would like to thank B. Shillman, the visionary founder of Congnex, for his continuing support for the Technion during crucial times. **Funding:** M.S. gratefully acknowledges an Advanced Grant from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 789339). The Würzburg group acknowledges financial support from the German Research Foundation (DFG) through the Würzburg-Dresden Cluster of Excellence on Complexity and Topology in Quantum Matter "ct.qmat" (EXC 2147, project ID 390858490). T.H.H. and S.H. acknowledge funding from the doctoral training program "Elitenetzwerk Bayern." T.H.H. acknowledges support from the German Academic Scholarship Foundation. The Würzburg team furthermore acknowledges support from the State of Bavaria. **Author contributions:** All authors contributed substantially to this work. **Competing interests:** Part of this work is included in a patent application filed by Technion and the University of Würzburg (international application no. PCT/IL2021/050489). **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and the supplementary materials.

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abj2232
Materials and Methods
Figs. S1 to S7
References (*34*–*37*)

28 April 2021; accepted 17 August 2021
10.1126/science.abj2232

CATALYSIS

Steam-created grain boundaries for methane C–H activation in palladium catalysts

Weixin Huang[1], Aaron C. Johnston-Peck[2], Trenton Wolter[3], Wei-Chang D. Yang[4], Lang Xu[3], Jinwon Oh[5], Benjamin A. Reeves[1], Chengshuang Zhou[1], Megan E. Holtz[2], Andrew A. Herzing[2], Aaron M. Lindenberg[5,6], Manos Mavrikakis[3], Matteo Cargnello[1,7]*

Defects may display high reactivity because the specific arrangement of atoms differs from crystalline surfaces. We demonstrate that high-temperature steam pretreatment of palladium catalysts provides a 12-fold increase in the mass-specific reaction rate for carbon-hydrogen (C–H) activation in methane oxidation compared with conventional pretreatments. Through a combination of experimental and theoretical methods, we demonstrate that an increase in the grain boundary density through crystal twinning is achieved during the steam pretreatment and oxidation and is responsible for the increased reactivity. The grain boundaries are highly stable during reaction and show specific rates at least two orders of magnitude higher than other sites on the palladium on alumina (Pd/Al_2O_3) catalysts. Theoretical calculations show that strain introduced by the defective structure can enhance C–H bond activation. Introduction of grain boundaries through laser ablation led to further rate increases.

The catalytic activities of supported metal nanoparticles (NPs) can depend on their surface structure and the exposed surface sites (*1*–*5*). Specific types of surface sites, such as terrace sites, steps, grain boundaries (GBs), and metal-support interface sites, can be manipulated to improve catalytic activity (*6*, *7*). For example, tetrahexahedral platinum (Pt) NPs with high-index facets exhibit enhanced catalytic activity in electrooxidation of formic acid and ethanol compared with Pt nanospheres (*8*). A silver catalyst with a high density of stacking faults showed superior activity and durability in the hydrogen evolution reaction (*9*).

Twin boundaries (TBs) and GBs are some of the most stable defects on metal surfaces and can serve as the active sites in certain electrocatalytic reactions (e.g., CO_2 electroreduction) (*10*–*12*). The improvement in performance results from the lattice strain induced by structural perturbations in the vicinity of the GBs at the catalyst surface (*13*), and this effect can lead to orders of magnitude higher catalytic rates (*14*). Although GBs have been recognized as promising defects for the activity of electrocatalysts, little is known about how they alter the catalytic properties in gas-phase heterogeneous reactions.

Here, we show that steam pretreatment of a palladium on alumina (Pd/Al_2O_3) catalyst enhanced its methane combustion activity, with the mass-specific reaction rate increasing by ≈12 times compared with the same sample treated in O_2. The extent of TB and GB formation was directly correlated to the activity of these Pd catalysts. Density functional theory (DFT) calculations showed that changes in reactivity can be explained by surface strain present in the immediate vicinity of GBs. The specific active sites exhibited a two orders of magnitude higher intrinsic rate.

Uniform colloidal Pd NPs were deposited onto a commercial Al_2O_3 support (average diameter of 15 nm; see Fig. 1A for size distribution and images). The catalytic performance of the Pd/Al_2O_3 catalyst was evaluated for methane combustion in the presence of steam (ramping the catalyst from 150°C to determine the light-off temperature) for the pristine catalysts and after several gas pretreatments (O_2, H_2, Ar, and steam) at temperatures of 300°C or higher (fig. S1A) (*15*). Negligible changes in light-off curves were observed relative to the pristine catalyst after conventional pretreatments (O_2, H_2, and Ar atmospheres) (fig. S1A) and had similar T_{50} values (temperatures needed to achieve 50% conversion of CH_4 to CO_2) of ≈423°C.

After a pretreatment in steam at 300°C, however, the catalyst showed a noticeable improvement in catalytic activity that led to the lowest T_{50} value (fig. S1A). The Pd/Al_2O_3 catalyst was then pretreated in either steam or O_2 at increasing temperatures (300°, 500°, 600°, and 700°C). No catalytic enhancement was found for the catalyst pretreated in O_2 (fig. S1B), but after pretreatment in steam, catalytic activity improved with increasing pretreatment temperature up to 600°C and showed no further change at 700°C (Fig. 1B). For this catalyst, a T_{50} of ≈373°C, which is 50°C lower than or the one treated in oxygen, was measured (Fig. 1D).

Cofeeding of steam in the reaction mixture usually has a detrimental effect on the methane combustion activity of Pd catalysts (*16*, *17*), so it was surprising that the steam pretreatment could increase the activity of the catalyst. The higher performance after steam pretreatment was stable for at least five cycles (5 hours spent on stream; fig. S2). We also considered the variability in steam concentration and processing time during the pretreatment. The improved activity was achieved at a water concentration as low as 0.8% (by volume) and did not change with increasing concentrations (fig. S3A). Changing the pretreatment duration from 0.5 hours to 2 hours did not change T_{50} (fig. S3B). These experiments suggest that the pretreatment temperature is more important than its duration and steam concentration.

Using high-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) analysis, we confirmed that there was no appreciable change in NP size distributions after any of the treatments (fig. S4), so the improved catalytic activity was not the result of particle sintering or redispersion. X-ray photoelectron spectroscopy (XPS) measurements demonstrated that in the O_2-pretreated catalyst, the Pd $3d_{5/2}$ peak was located at 336.6 eV (Fig. 1C and fig. S5), consistent with the PdO phase. The steam-pretreated catalyst showed a peak at 335.0 eV, attributable to metallic Pd, which was consistent with x-ray diffraction data (fig. S6). The catalysts were oxidized under reaction conditions (see below).

There is still a debate as to whether Pd, PdO, or a Pd/PdO mixed phase is the active phase for methane combustion (*18*). To fairly compare catalysts with the same oxidation state, a Pd/Al_2O_3 sample was prepared through sequential oxygen and hydrogen treatments [labeled as O_2-H_2–pretreated Pd/Al_2O_3 (*15*)] to convert the Pd oxide phase into metallic Pd, as confirmed by XPS (Fig. 1C). However, the activity for this sample was similar to that of the one pretreated exclusively in O_2 and much lower than that of the steam-pretreated catalyst (Fig. 1, D and E). We subjected the steam-pretreated sample to an oxygen treatment [labeled as steam-O_2–pretreated Pd/Al_2O_3 (*15*)] to convert the metallic Pd phase into PdO (as confirmed by XPS) (Fig. 1C) and test whether the oxidation would reduce its activity to that of the oxidized sample. No evident activation or deactivation in light-off performance was found as compared with the steam-pretreated sample (Fig. 1D). This demonstrated that the initial Pd oxidation state before catalysis did not correlate with the methane oxidation activity of the samples, and steam pretreatment improved

[1]Department of Chemical Engineering, Stanford University, Stanford, CA 94305, USA. [2]Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. [3]Department of Chemical and Biological Engineering, University of Wisconsin–Madison, Madison, WI 53706, USA. [4]Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA. [5]Department of Materials Science and Engineering, Stanford University, Stanford, CA 94305, USA. [6]Stanford Institute for Materials and Energy Sciences, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [7]SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA.
*Corresponding author. Email: mcargnello@stanford.edu

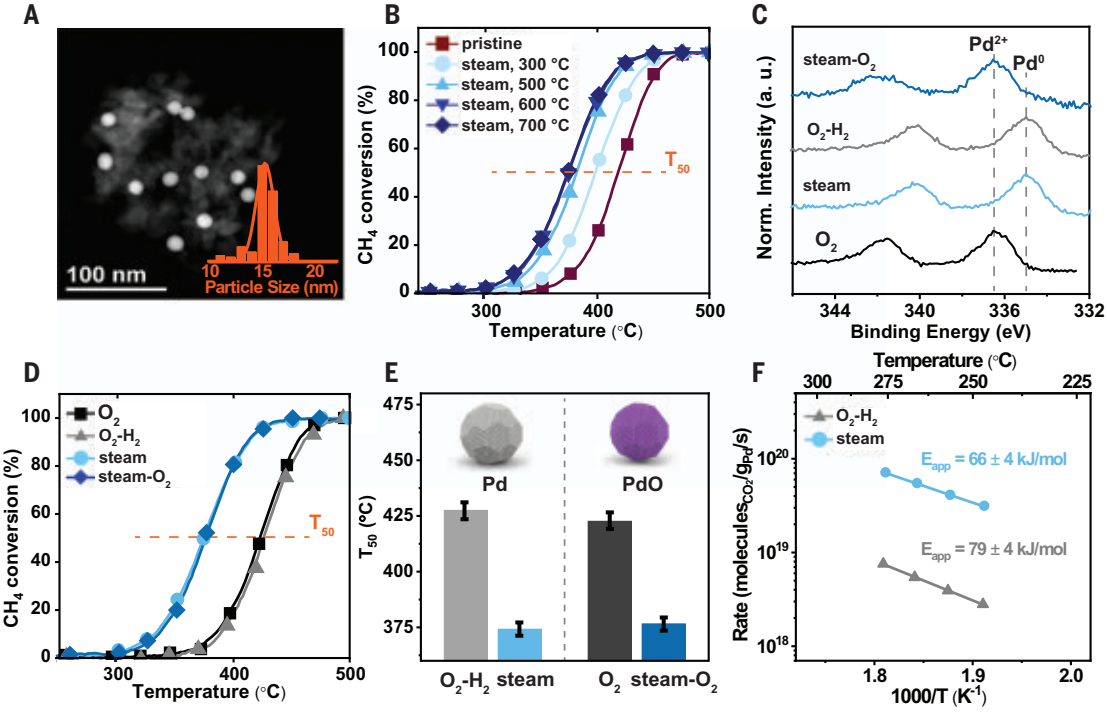


Fig. 1. Characterization and performance of Pd/Al_2O_3 catalysts upon different gas pretreatments. (**A**) Representative HAADF-STEM image of the pristine Pd/Al_2O_3. (Inset) Size distribution of Pd NPs. (**B**) Light-off curves for methane combustion (0.4% CH_4, 4.0% O_2, 4.2% H_2O, balance Ar) on pristine Pd/Al_2O_3 and after pretreatment in steam at increasing temperatures. (**C**) Pd 3*d* photoelectron spectra of Pd/Al_2O_3 after O_2 (600°C), O_2-H_2, steam (600°C), and steam-O_2 pretreatments. The spectral fittings are shown in fig. S5. a.u., arbitrary units. (**D**) Light-off curves and (**E**) T_{50} values of Pd/Al_2O_3 after O_2 (600°C), O_2-H_2, steam (600°C), and steam-O_2 pretreatments. Error bars represent the minimum and maximum measured values of at least three repeated experiments. (**F**) Arrhenius plots of methane combustion on Pd/Al_2O_3 after O_2-H_2 pretreatment and steam pretreatment (600°C).

the Pd activity regardless of its initial oxidation state.

Reaction rates were measured for these various Pd/Al_2O_3 catalysts. To avoid catalyst changes during kinetic experiments, steam was not added to the reaction mixture (fig. S7). The 600°C steam-pretreated catalyst had a ≈12 times higher mass-specific rate than the O_2-H_2–pretreated catalyst (Fig. 1F). Activating the first C–H bond is recognized as the rate-limiting step in methane oxidation, and Arrhenius plots showed an apparent activation energy for the O_2-H_2–pretreated catalyst of 79 ± 4 kJ·mol^{-1}, in agreement with previous studies (*19*). The steam-pretreated catalyst had lower activation energy (66 ± 4 kJ·mol^{-1}). The two samples showed similar prefactor values (≈3×10^{26} molecules of CO_2 per grams of Pd per second).

The effect of the PdO reducibility state was explored with methane temperature-programmed reduction (CH_4-TPR) (fig. S8A). Whereas methane oxidation was observed at ≈190°C for the steam-pretreated catalyst, the temperature shifted up to ≈215°C for the O_2-pretreated catalyst, consistent with the light-off curves (Fig. 1D). In temperature-programmed Pd oxidation (O_2-TPO) experiments, O_2 uptake was observed at ≈295°C for the steam-pretreated catalyst, versus ≈330°C for the O_2-pretreated one (fig. S8B), in line with DFT calculations (see below). Thus, the Pd phase in the steam-pretreated catalyst was more easily oxidized and reduced by O_2 and CH_4, respectively, and could be more active in methane oxidation.

Changes in support after steam treatment, including support hydroxylation, can promote reactivity of supported metal phases (*20*). We treated the alumina support in steam before depositing Pd NPs (*15*), but this sample showed an even higher T_{50} than that of a conventional Pd/Al_2O_3 with the same NP size (fig. S9). The decreased activity could have been caused by different metal-support interactions with hydroxylated alumina. Also, Pd/SiO_2 catalysts showed the same improvement in activity when treated in steam versus oxygen or oxygen-hydrogen atmospheres (fig. S10). Activity improvement could result from catalyst synthesis by-products being removed from the surface by the steam treatment. Phosphorus is an impurity in the initial Pd NPs, but it was mostly removed after steam or O_2-H_2 treatments, and there was no difference in its concentration between the samples (fig. S11 and table S2). These results suggest that the observed activity enhancement is related to structural changes in the Pd NPs.

We used HAADF-STEM to characterize the structure of the supported metallic Pd NPs (Fig. 2 and figs. S12 to S17). Oxidized Pd NPs underwent drastic electron beam–induced changes and could not be investigated in detail. The pristine Pd NPs on alumina mostly had an amorphous structure (fig. S12). However, after being treated in steam or O_2-H_2, the NPs crystallized and became highly faceted.

Different Pd exposed facets upon different gas pretreatments could account for the changes in reactivity (fig. S17). The distances from the particle center to the outermost surface planes were measured (table S3), and the corresponding three-dimensional crystal shape was derived using the Wulff construction. From the Wulff shape, the occurrence of different types of surface facets was extracted (fig. S18). Although samples showed different fractions of exposed facets, no trend correlated with the difference in catalyst activity (figs. S18 and S19), neither did the presence of voids in the NPs created by Kirkendall effects (figs. S12 to S15 and table S4). A similar procedure was used to analyze oxidized NPs, and although the beam sensitivity allowed measurements of only a few of them, comparable ratios of PdO {110} and {101} facets were observed (fig. S20 and table S5).

Both the steam- and O_2-H_2–pretreated samples presented TBs that lay parallel to the {111} planes (also referred to as Σ3 {111} TBs) (Fig. 2, A, B, and E, orange arrows). Separated by a coherent TB, the surface structure shows a symmetrical lattice arrangement with an ABC|CBA stacking sequence. The fast Fourier transform (FFT) diffractograms in Fig. 2J show two sets of patterns, in which the $(1\bar{1}1)$, (200) spots in grain G2 are mirrored, across the plane parallel to $(\bar{1}\bar{1}1)$, by $(\bar{1}\bar{1}\bar{1})$, $(00\bar{2})$ in grains G1 and G3, forming a typical coherent TB pattern.

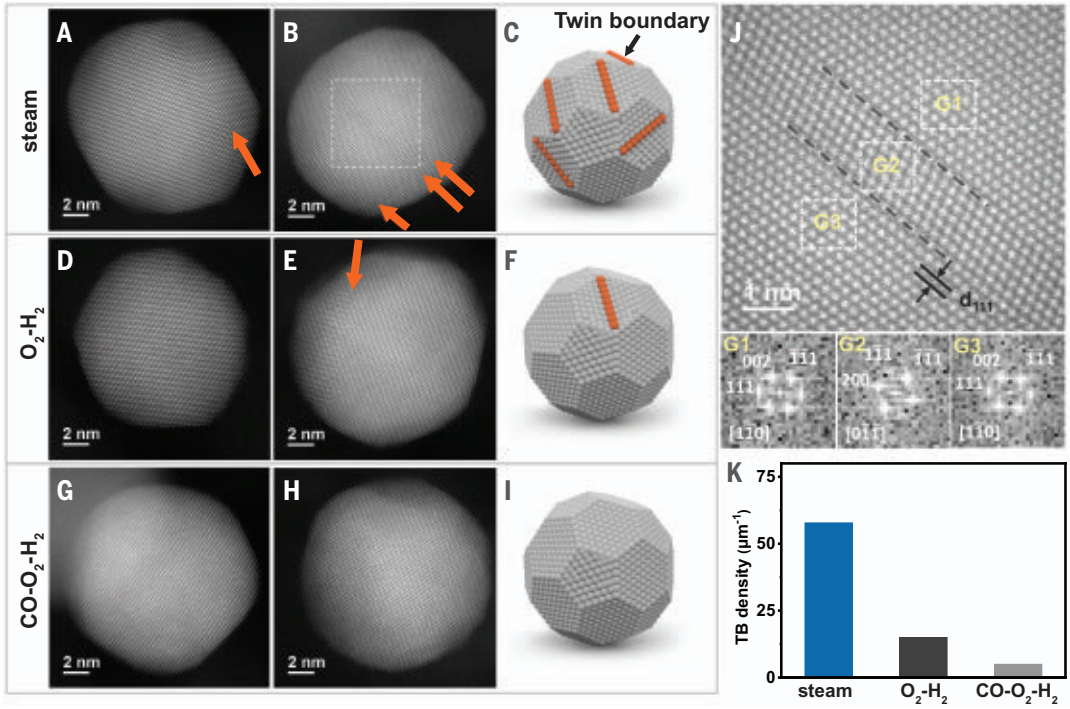


Fig. 2. Microstructure investigation. Atomic-resolution HAADF-STEM images of (**A** and **B**) steam-pretreated (600°C) Pd/Al_2O_3, (**D** and **E**) O_2-H_2–pretreated Pd/Al_2O_3, and (**G** and **H**) CO-O_2-H_2–pretreated Pd/Al_2O_3. (**C**, **F**, and **I**) Schematic of the TB density change in Pd NPs after different gas pretreatments. (**J**) The area marked by the box in (B), revealing the detailed arrangements of Pd atoms. The dashed lines highlight Σ3{111} TBs. Corresponding FFT images of grains (G1, G2, and G3) labeled in the top panel. (**K**) TB density statistical histogram of Pd/Al_2O_3 after steam (600°C), O_2-H_2, and CO-O_2-H_2 treatment.

To assess the relation between the presence of TBs and catalytic activity, we measured the TB surface density, which was taken to be the sum of the TB surface length over all measured NPs divided by the sum of the NP surface areas (*15*). We used experimental data on samples after different treatments to estimate the TB density to be ≈58 μm^{-1} for the steam-treated sample and ≈15 μm^{-1} for the O_2-H_2–pretreated sample (Fig. 2K and table S6). This result correlated well with the observed catalytic activity trend.

To confirm the role of TB density, a pristine Pd/Al_2O_3 sample was subjected to dilute CO treatment to cause Pd NPs to restructure into vicinal stepped surfaces (*4*) and decrease TB formation. The sample was further subjected to O_2 and H_2 treatments to remove the carbon coating induced by the CO treatment (fig. S21), reduce the Pd to a metallic state, and create a fully accessible and active Pd surface [labeled as CO-O_2-H_2–pretreated Pd/Al_2O_3, (*15*)]. No appreciable change in NP size was seen in the TEM images, and XPS confirmed the metallic state of Pd (fig. S22). The CO-O_2-H_2–pretreated sample had a low TB density of ≈4.9 μm^{-1}, as revealed by high-resolution HAADF-STEM images (Fig. 2, G to I, and fig. S15), and was also less active than both the steam-pretreated catalyst and the O_2-H_2 catalyst, with a higher T_{50} of ≈ 446°C (Fig. 3, A and B, and fig. S23).

These results were consistent with the observed correlation of the increased activity of the Pd/Al_2O_3 catalysts with TBs and the lack of TB formation under O_2, H_2, or CO atmospheres (Fig. 3B). We calculated an intrinsic reaction rate assuming that the atoms at the TB for the measured TB density in the steam-pretreated sample accounted for the increase in rate (*15*). This rate was ≈785 times greater than that on the in-plane Pd atoms.

Given the presence of the GB, enhanced catalytic activity associated with TBs could be related to strain effects. The exit wave power cepstrum transform was applied to scanning nanobeam electron diffraction data to explore the distribution of lattice strain in the steam- and CO-O_2-H_2–pretreated Pd/Al_2O_3 catalysts (*21*). The strain values are relative to a reference value (Lagrange strain), which was measured from the sum of all the diffraction patterns for individual NPs (Fig. 3C). Analysis revealed a radial lattice expansion near surfaces in both steam- and CO-O_2-H_2–pretreated samples regardless of the presence of a TB, and no correlation between the presence of a TB and a change in strain was identified in the Pd samples (figs. S24 and S25). A previous report analyzing atomic resolution images of supported Pt NPs reported a 1 to 2% expansion in the lattice planes immediately adjacent to the TB (*22*). The difference in results reported here could originate from the material system and synthesis, or from the lower spatial resolution of the strain routine implemented here lacking sensitivity. However, analysis of the prereaction NPs may not be sufficient given restructuring under reaction conditions and oxidation of the metallic surface.

Thus, we determined the thermal stability of the TBs under oxidizing conditions using environmental transmission electron microscopy (ETEM). Initially, the Pd/Al_2O_3 was exposed to an O_2 environment at a pressure of ≈0.87 Pa at room temperature (movie S1). The catalyst was heated at a rate of 100°C·s^{-1} and stabilized at ≈500°C (movie S2). We detected no apparent boundary segregation or disappearance as annealing temperature increased (Fig. 3, D and E). Instead, surface oxidation on the NP was observed during this process: The "cap" separated by the {111} TB was preferentially oxidized (Fig. 3, E and F), suggesting that the TB may promote oxygen dissociation and serve as the precursor structure to the formation of a GB between the Pd core and the surface PdO region. A slower heating rate experiment (200°C·min^{-1}) showed that the cap region was preferentially oxidized at ≈391°C in the same O_2 environment (fig. S26). We confirm the PdO phase formation on the basis of the FFT diffractogram (Fig. 3E, bottom inset) that matches well with tetragonal PdO and suggests that the oxidized cap region is oriented close to the $[\bar{1}11]$ zone axis. The PdO surface is bounded by the (110) and $(\bar{1}0\bar{1})$ facets (Fig. 3F). A time series of consecutive frames (each 2.5 s apart) selected from movie S2 shows the dynamic nature of the PdO facets in the oxidizing environment, with the PdO (110) surface replaced by the $(\bar{1}0\bar{1})$ surface (second frame) before reappearing (third frame) (as shown in fig. S27).

The ETEM identification explained the O_2-TPO data, in which the steam-pretreated catalyst oxidized at a lower temperature than the O_2-pretreated catalyst. We conclude that the original TB transformed into a general GB,

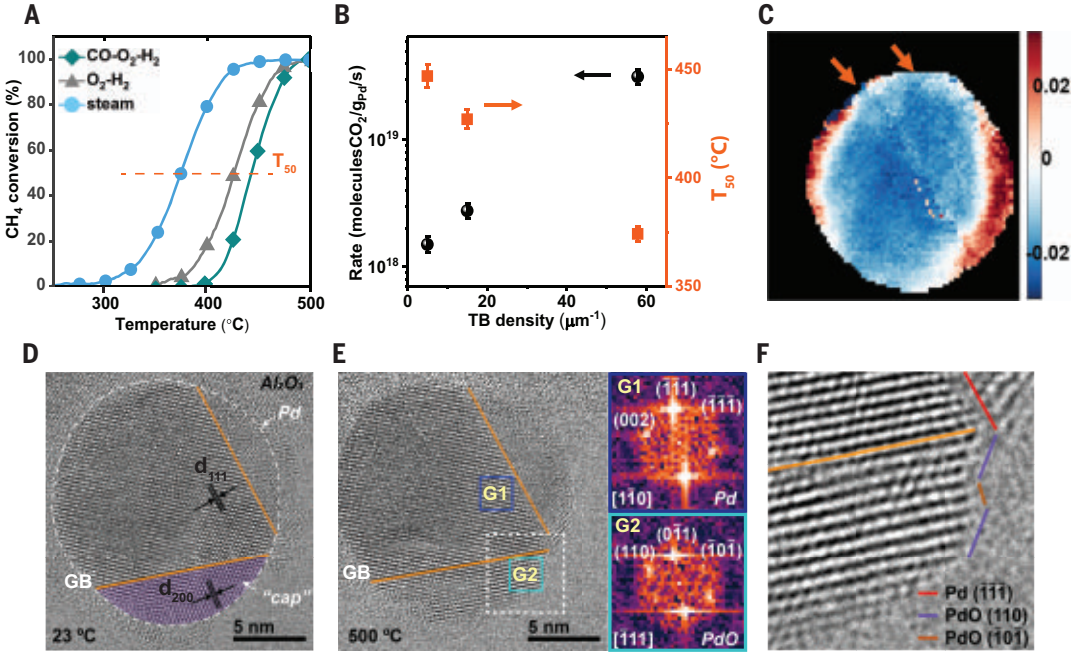


Fig. 3. Identifying TBs as the origin of increased catalytic activity and exploring their thermal stability. (**A**) Light-off curves of Pd/Al_2O_3 after steam (600°C), O_2-H_2, and CO-O_2-H_2 pretreatments. (**B**) Relationship between reaction rate and TB density and T_{50} and TB density. (**C**) Representative strain mapping for an individual Pd NP in the steam-pretreated catalyst relative to the reference values in the horizontal direction. The arrows denote the TBs. (**D** and **E**) Representative environmental TEM (ETEM) images of the same Pd nanoparticle in the steam-pretreated Pd/Al_2O_3 sample that was exposed to O_2 at 23° and 500°C, respectively, with the twin or grain boundaries highlighted by orange lines. FFT diffractograms [insets in (E)] indicating that the cap region separated by GB is preferentially oxidized at 500°C. (**F**) Zoom-in of the area marked by the white dashed box in (E) showing the exposed Pd and PdO facets in the vicinity of the GB (orange line).

indicating that these planar defects were maintained in an oxidizing condition at a high temperature of at least 500°C. The conversion of Pd to PdO transformed the crystalline lattice from cubic to tetragonal.

TBs might lead to an increased reaction rate for a variety of reasons. For example, they may impart structural irregularities that induce reconstruction of the PdO surface (*23*). Further, linear and/or point defects at the vicinity of the TB can lead to improved C–H bond activation (*24*). Alternatively, the oxidation of the metal surface may generate strain in the Pd/PdO heterostructure and in the fully oxidized PdO NP containing the GB. We explored this latter possibility in detail using DFT.

We performed plane wave DFT calculations for the dissociation of O_2 on Pd(111) and a TB model between two Pd(111) terraces (fig. S28). The binding energy of O_2 was calculated to be −1.07 and −1.19 eV on Pd(111) and the TB model, respectively. The O_2 dissociation barriers were calculated to be 0.66 and 0.49 eV on Pd(111) and the TB model, respectively. Overall, when referenced to gas-phase O_2, the O_2 dissociation transition state is 0.29 eV more stable on the TB model than on Pd(111) (which corresponds to a ~150-fold difference in rate at 400°C). These results indicate that the presence of twin boundaries will increase the O_2 population on the surface and facilitate O_2 dissociation relative to the Pd(111) facet, both suggesting that PdO formation is accelerated in the vicinity of twin boundaries. These results are in line with the CH_4-TPR and O2-TPO analysis reported above.

Plane wave DFT calculations were also performed to assess the effect of strain generated by the GBs on the activity of the catalyst for the first C–H bond-breaking step in methane combustion, known to be rate determining given the rate orders for methane and oxygen of ≈+1 and 0, respectively (fig. S29). PdO(110) and PdO(101) were chosen because the oxidized phase formed under reaction conditions (Fig. 3E) and preoxidation of the NPs did not hinder catalytic activity (Fig. 1D). These facets were specifically chosen because they were observed with ETEM (Fig. 3F). The catalyst surfaces were modeled as bulk PdO surfaces because the NPs that were placed under reaction conditions were measured to have PdO phases that were at least 4-nm thick, which corresponds to ~13 atomic layers. At this PdO thickness, there should be a sufficient number of PdO layers to resemble the properties of a bulk PdO surface, whereas the contribution from the supporting metallic Pd substrate is effectively screened.

Gibbs free energy diagrams and energy-optimized structures for adsorbed CH_4*, the transition states (TSs), and the final states (CH_3* + H*) were computed (Fig. 4, A to D). The calculation of the Gibbs free energy assumed a temperature of 400°C, at which the largest difference in methane conversion between steam-pretreated and O_2-pretreated catalysts was observed (Fig. 1D). Calculations were performed for each surface at −5% (i.e., 5% compressive), 0%, and +5% (i.e., 5% expansive) strain, so that the effects of both types of strain were explored and a trend could be established between the reactivity and the applied strain level.

The Gibbs free energy of adsorbed methane on PdO(101) with −5, 0, and 5% applied strain were −0.03, −0.07, and −0.25 eV, respectively (Fig. 4A). These values indicate that CH_4 binds more strongly to the surface with expansive strain, which is consistent with previously observed strain effects on metallic surfaces (*25*). The TS Gibbs free energy relative to gas-phase CH_4 was 0.89, 0.85, and 0.63 eV for −5, 0, and 5% applied strain, respectively. These energy values are effectively the Gibbs free energy differences between the C–H activation barrier and the desorption barrier of adsorbed CH_4. Expansive strain could increase the activity for C–H activation (as opposed to CH_4 desorption) on the PdO (101) surface.

To give a sense for the magnitude of this effect, a difference of ~0.13 eV in activation energy corresponds to one order of magnitude change in reaction rate constant at 400°C. Thus, imposing a 5% expansive strain on PdO(101) would yield an intrinsic reactivity enhancement corresponding to a 0.22 eV TS energy difference, and a coverage term enhancement due to stronger CH_4 binding by 0.18 eV. Combined, this leads to ≈3 orders of magnitude in intrinsic rate enhancement for the expansively strained PdO(101) surface, an estimate that is near the factor of 785 estimated from our experiments.

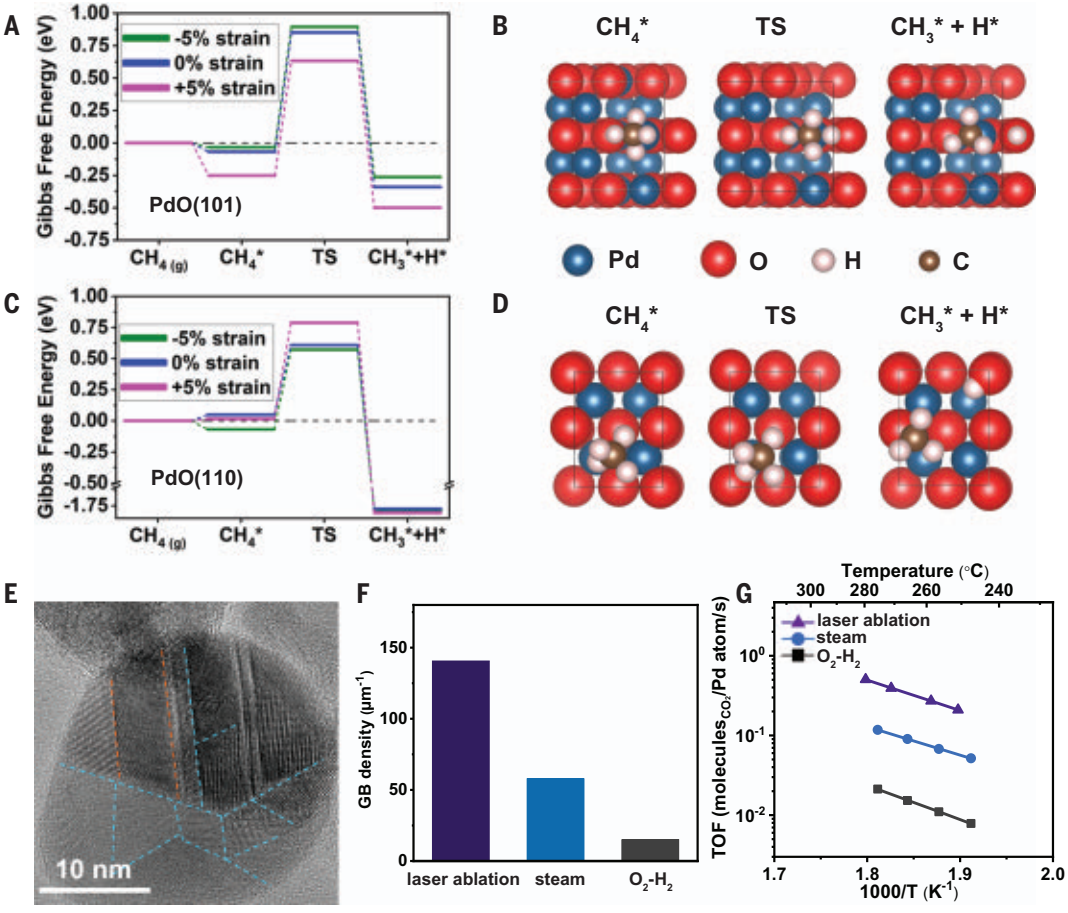


Fig. 4. Theoretical calculations for the effect of surface strain on the C–H breaking energetics and GB density effect investigation. (**A** and **C**) Calculated free energy diagrams for breaking the first C–H bond in CH_4 on PdO(101) and PdO(110), respectively. The Gibbs free energy values were calculated using a temperature of 400°C and referenced a clean surface and gas-phase CH_4 at infinite separation. (**B** and **D**) Top view of structures for adsorbed CH_4 (CH_4*), TS, and coadsorbed CH_3 and H (CH_3* + H*) on PdO(101) and PdO(110), respectively (side views are shown in fig. S32). Dashed rectangle denotes the unit cell. (**E**) High-resolution TEM image of laser ablation–generated Pd NPs. The orange dashed lines highlight Σ3{111} TBs, and the blue dash lines highlight GBs. (**F**) GB density statistical histogram of laser-generated Pd/Al_2O_3 and Pd/Al_2O_3 after steam (600°C) and O_2-H_2 pretreatments. (**G**) Arrhenius plots of methane combustion of laser-generated Pd/Al_2O_3 and Pd/Al_2O_3 after steam (600°C) and O_2-H_2 pretreatments. TOF, turnover frequency.

For comparison, the Gibbs free energy diagram for methane activation on the PdO(110) facet was also calculated (Fig. 4C). The reactivity of this surface was probed at its oxygen termination, which corresponds to a lower surface energy than its Pd termination under the experimental conditions used in this study (*26*). The application of expansive or compressive strain did not show a large effect on the binding energy of CH_4* or CH_3* + H* on this surface, where only narrow ranges of variances in the energies of CH_4* (0.05 eV) and CH_3* + H* (0.03 eV) were observed (Fig. 4C).

The TS energies on the PdO(110) surface showed a different trend than that observed on the PdO(101) surface as expansive or compressive strain was applied. The TS energies were 0.57, 0.61, and 0.79 eV for –5, 0, and 5% applied strain, respectively (Fig. 4C). These results suggest that a 5%-compressed PdO(110) could increase intrinsic reactivity corresponding to a total of ≈0.09 eV lower Gibbs free energy, which is less than an order or magnitude at 400°C, compared with 0% strain on that facet. Expansive strain would decrease reactivity on that facet.

The DFT results for PdO(101) and PdO(110) together demonstrate that a strained surface causes a higher C–H activation rate. As mentioned previously for the ETEM experiment, the PdO phase tended to form a cap region on one side of the GB, so the PdO phase could undergo compressive or expansive strain created by GBs, which, in turn, could increase reactivity. This idea is consistent with the experimental results showing higher activity for NPs that have a higher number of TBs.

Our DFT results suggest that the enhanced activity for the steam-pretreated Pd catalyst stems from a concerted effect of PdO formation and local strain caused by the presence of GBs. On the basis of the reactivity enhancement estimates calculated on the stretched Pd(101) and on the compressed PdO(110), and the respective comparisons with experimental rate measurements, we conclude that stretched Pd(101) is likely the dominant active site on the steam-pretreated catalysts. This assertion offers a potential avenue for improving Pd methane oxidation catalysts through better control of the facets and strain during PdO growth.

Activity enhancement by steam pretreatment was not exclusive of the Pd(15 nm)/Al_2O_3 catalyst. The same pretreatment-dependent activity profiles were observed in a Pd/Al_2O_3 catalyst prepared by conventional wet impregnation and in Pd/Al_2O_3 catalysts prepared from 8- and 12-nm colloidal Pd NPs (figs. S33 to S35). In all these cases, the steam-pretreated catalysts exhibited higher activity than the O_2-pretreated catalysts. The NPs with a larger size showed greater improvement in rates upon steam treatment (fig. S36). This observation may be explained by the more energetically favorable formation of TBs in larger NPs that can more easily accommodate defects. Despite a lower magnitude in rate increase likely caused by the smaller NP sizes and a less uniform surface structure in the wet-impregnated catalyst, these experiments confirm the generality of the effect of steam pretreatment on Pd catalysts.

The identification of GBs as highly active sites for methane combustion provides the opportunity to engineer Pd/Al_2O_3 catalysts for even more improved reactivity if the density of such defects can be increased. With this goal in mind, following previous reports (*9*), a laser ablation process was used to fabricate NPs rich in GBs (Fig. 4, E and F, and figs. S37 and S38) that were deposited on the same alumina used for the colloidal NPs. As predicted by our observations, the turnover frequency of the laser ablation-generated Pd/Al_2O_3 catalyst was four times higher than that of the steam-pretreated Pd(15 nm)/Al_2O_3 catalyst and hence

nearly 25 times higher than a conventional catalyst pretreated in just oxygen and hydrogen (Fig. 4G and table S7). These results may not only provide a general strategy for the rational design of the surface structure in heterogeneous nanocatalysts but can also help explain previous work on Pd-based catalysts for methane combustion that demonstrated improved reactivity upon aging or disparate thermal treatments (*27*).

REFERENCES AND NOTES

1. P. Nolte *et al.*, *Science* **321**, 1654–1658 (2008).
2. B. Hammer, *Top. Catal.* **37**, 3–16 (2006).
3. F. F. Tao *et al.*, *Nano Lett.* **16**, 5001–5009 (2016).
4. T. Avanesian *et al.*, *J. Am. Chem. Soc.* **139**, 4551–4558 (2017).
5. L. Wang *et al.*, *Science* **363**, 870–874 (2019).
6. P. P. Sharma *et al.*, *Angew. Chem. Int. Ed.* **54**, 13701–13705 (2015).
7. Y. Li *et al.*, *Nano Lett.* **17**, 1312–1317 (2017).
8. N. Tian, Z.-Y. Zhou, S.-G. Sun, Y. Ding, Z. L. Wang, *Science* **316**, 732–735 (2007).
9. Z. Li *et al.*, *Nat. Catal.* **2**, 1107–1114 (2019).
10. M. Zhao *et al.*, *ACS Catal.* **8**, 6948–6960 (2018).
11. X. Feng, K. Jiang, S. Fan, M. W. Kanan, *J. Am. Chem. Soc.* **137**, 4606–4609 (2015).
12. C. W. Li, J. Ciston, M. W. Kanan, *Nature* **508**, 504–507 (2014).
13. R. G. Mariano, K. McKelvey, H. S. White, M. W. Kanan, *Science* **358**, 1187–1192 (2017).
14. C. Tang *et al.*, *ACS Catal.* **10**, 2026–2032 (2020).
15. See the supplementary materials.
16. D. Ciuparu, M. R. Lyubovsky, E. Altman, L. D. Pfefferle, A. Datye, *Catal. Rev. Sci. Eng.* **44**, 593–649 (2002).
17. R. Gholami, M. Alyani, K. J. Smith, *Catalysts* **5**, 561–594 (2015).
18. Y. H. Chin, C. Buda, M. Neurock, E. Iglesia, *J. Am. Chem. Soc.* **135**, 15425–15442 (2013).
19. J. J. Willis *et al.*, *J. Am. Chem. Soc.* **139**, 11989–11997 (2017).
20. L. Nie *et al.*, *Science* **358**, 1419–1423 (2017).
21. E. Padgett *et al.*, *Ultramicroscopy* **214**, 112994 (2020).
22. T. Nilsson Pingel, M. Jørgensen, A. B. Yankovich, H. Grönbeck, E. Olsson, *Nat. Commun.* **9**, 2722 (2018).
23. N. M. Martin *et al.*, *ACS Catal.* **4**, 3330–3334 (2014).
24. H. H. Kan, J. F. Weaver, *Surf. Sci.* **603**, 2671–2682 (2009).
25. M. Mavrikakis, B. Hammer, J. K. Nørskov, *Phys. Rev. Lett.* **81**, 2819–2822 (1998).
26. J. Rogal, K. Reuter, M. Scheffler, *Phys. Rev. B* **69**, 075421 (2004).
27. O. Demoulin *et al.*, *J. Phys. Chem. B* **109**, 20454–20462 (2005).

ACKNOWLEDGMENTS

Funding: This work was mostly supported by NYSEARCH. Additional funding was provided by the Stanford Natural Gas Initiative. T.W., L.X., and M.M. acknowledge support for the computational modeling work by the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, Division of Chemical Sciences, Geosciences, and Biosciences (grant DE-FG02-05ER15731). J.O. was supported by the National Science Foundation under grant 1956300. B.A.R. was supported by the HydroGEN Advanced Water Splitting Consortium, established as part of the Energy Materials Network under the US Department of Energy, Office of Energy Efficiency and Renewable Energy, Hydrogen and Fuel Cell Technologies Office, under award DE-EE-0008084. Part of the experimental work was performed at the Stanford Nano Shared Facilities (SNSF), supported by the National Science Foundation under award ECCS-1542152. The computational work was performed in part by using supercomputing resources from the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility located at Lawrence Berkeley National Laboratory and operated under contract DE-AC02-05CH11231, and from the Center for Nanoscale Materials, a DOE Office of Science User Facility located at Argonne National Laboratory, supported by DOE contract DE-AC02-06CH11357. A.M.L. acknowledges support from the US Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, under contract DE-AC02-76SF00515.
Author contributions: W.H. and M.C. designed the project. W.H. performed catalyst synthesis, characterization, catalytic tests, and data analysis. A.C.J.-P. performed all high-resolution scanning transmission electron microscopy measurements and the related data analysis. T.W. and L.X. performed theoretical calculations supervised by M.M. W.-C.D.Y. conducted ETEM experiments and data analysis. J.O., B.A.R., and C.Z. conducted catalyst synthesis for laser ablation and the related catalytic tests supervised by A.M.L. M.E.H. and A.A.H. performed the surface strain data analysis. M.C. supervised the project. W.H. prepared a manuscript draft, and all authors contributed to the final manuscript. **Competing interests:** Stanford University has submitted a provisional patent application entitled "Promoting catalytic activity through steam treatment-induced nanoparticles restructuring" on which W.H. and M.C. are listed as inventors (US provisional patent number 63/084,847). The authors declare no competing financial interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abj5291
Materials and Methods
Supplementary Text
Figs. S1 to S38
Tables S1 to S7
References (*28*–*57*)
Movies S1 and S2

18 May 2021; accepted 18 August 2021
10.1126/science.abj5291

SYNTHETIC BIOLOGY

Cell-free chemoenzymatic starch synthesis from carbon dioxide

Tao Cai[1,2]†, Hongbing Sun[1,2]†, Jing Qiao[1,2]†, Leilei Zhu[2,3]†, Fan Zhang[1,2], Jie Zhang[2,3], Zijing Tang[2,3], Xinlei Wei[2,3], Jiangang Yang[2,3], Qianqian Yuan[2,4], Wangyin Wang[5], Xue Yang[2,4], Huanyu Chu[2,4], Qian Wang[2,4], Chun You[2,3], Hongwu Ma[2,4], Yuanxia Sun[2,3], Yin Li[1,2], Can Li[5], Huifeng Jiang[2,4], Qinhong Wang[1,2,4], Yanhe Ma[1,2,3]*

Starches, a storage form of carbohydrates, are a major source of calories in the human diet and a primary feedstock for bioindustry. We report a chemical-biochemical hybrid pathway for starch synthesis from carbon dioxide (CO_2) and hydrogen in a cell-free system. The artificial starch anabolic pathway (ASAP), consisting of 11 core reactions, was drafted by computational pathway design, established through modular assembly and substitution, and optimized by protein engineering of three bottleneck-associated enzymes. In a chemoenzymatic system with spatial and temporal segregation, ASAP, driven by hydrogen, converts CO_2 to starch at a rate of 22 nanomoles of CO_2 per minute per milligram of total catalyst, an ~8.5-fold higher rate than starch synthesis in maize. This approach opens the way toward future chemo-biohybrid starch synthesis from CO_2.

Starch is a main caloric component of food and animal feed, as well as an important industrial feedstock (*1*, *2*). Amylose and amylopectin polymers in starch granules consist of chains of glucosyl residues linearly linked by α-1,4-glycosidic bonds, interspersed by branching points of α-1,6-glycosidic bonds in the case of amylopectin (*3*). Starch synthesis in green plants involves about 60 steps and complex regulation (*4*, *5*). Although many efforts have been made to improve the production of starch in plants (*6*–*8*), the inefficiency of photosynthesis and the complexity of starch biosynthesis are obstacles (*9*). By contrast, advances in synthetic biology have enabled the design and construction of synthetic systems for more efficient CO_2 fixation (*10*–*14*) and chemical production (*15*, *16*). Inspired by the central principles of photosynthesis, extraordinary chemical catalysts have been developed to provide electrons (*17*) or hydrogen (*18*) more efficiently from solar energy and water for reducing CO_2 into simple chemicals (*19*, *20*). In this study, we used a chemical CO_2 reduction catalyst that produces reduced one-carbon (C1) units as an input to a chemoenzymatic pathway for cell-free starch synthesis.

To build this hybrid pathway, we first chose formic acid and methanol to serve as the candidate intermediates to bridge possible chemical catalysts and biological enzymes. We exploited formolase (fls) to design and construct the enzymatic part of the starch synthesis pathway from the candidate C1 intermediates (*21*). On the basis of a main set of 6568 reactions from the MetaCyc database (*22*) and ATLAS database (*23*) and two combinatorial sets of 15 formate and 8 methanol utilization reactions, we drafted two concise starch synthesis pathways from either formic acid or methanol by using the combination of combinatorial

[1]Department of Strategic and Integrative Research, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China. [2]National Center of Technology Innovation for Synthetic Biology, Tianjin 300308, China. [3]National Engineering Laboratory for Industrial Enzymes, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China. [4]CAS Key Laboratory of Systems Microbial Biotechnology, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China. [5]State Key Laboratory of Catalysis, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian 116023, China.
***Corresponding author. Email: ma_yh@tib.cas.cn**
†These authors contributed equally to this work.

algorithm and parsimonious flux balance analysis (comb-FBA) (*24*) and the COBRApy toolbox in Python (*25*) (fig. S1A and supplementary text). Starch could be synthesized, in principle, through only nine core reactions from CO_2 with formic acid or methanol as the C1 bridging intermediate (Fig. 1, inner circle).

In contrast to natural pathways that have evolved functionality and compatibility over hundreds of millions of years of selection, computationally designed pathways are often hindered by unpredictable and undesired interactions between enzymes from disparate biochemical contexts (*26*). To overcome these problems, we pursued a strategy of modular assembly and substitution. Two starch synthesis pathways were divided into more manageable modules (fig. S1A), including a C1 module (for formaldehyde production), a C3 module (for D-glyceraldehyde 3-phosphate production), a C6 module (for D-glucose-6-phospate production), and a Cn module (for starch synthesis). According to known enzymes in databases, five modules were initially constructed (C1a/C1b, C3a, C6a, and Cna). Although the C1a, C1b, and C3a modules displayed function individually (figs. S2, A, B, and G, and S3), the assembly of C1a or C1b with C3a did not result in detectable C3 compounds from formic acid or methanol (fig. S6A). We speculated that the marginal formaldehyde production from energy-efficient but thermodynamically unfavorable C1 modules may not be able to

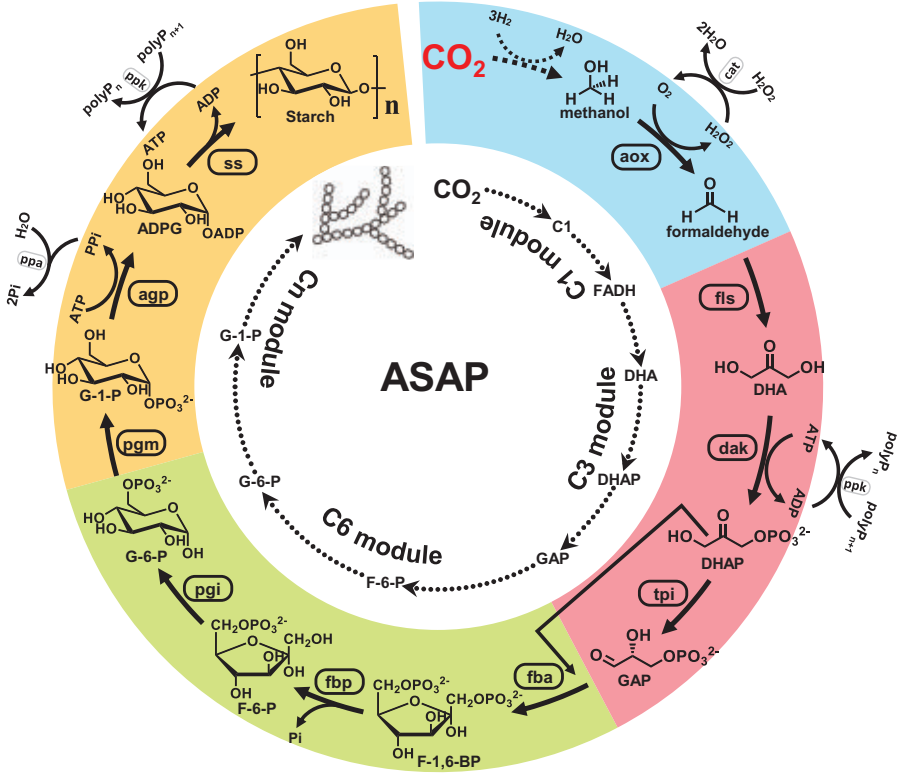


Fig. 1. Design and modular assembly of an artificial starch anabolic pathway. Inner circle: schematic of the artificial starch pathway drafted by computational pathway design with divided modules. C1 here indicates formic acid and methanol. Outer circle: schematic of artificial starch anabolic pathway (ASAP) 1.0, with individual modules colored. Auxiliary enzymes and chemicals are indicated. ADPG, ADP glucose; aox, alcohol oxidase; FADH, formaldehyde; F-1,6-BP, D-fructose-1,6-bisphosphate; F-6-P, D-fructose-6-phosphate; GAP, D-glyceraldehyde 3-phosphate; pgi, phosphoglucose isomerase; polyP, polyphosphate; pgm, phosphoglucomutase; ppa, pyrophosphatase; ppk, polyphosphate kinase; ss, starch synthase; tpi, triose-phosphate isomerase.

Table 1. Comparison of ASAP iterations with other natural and synthetic pathways. The average growth period of maize was assumed as 120 days. These numbers may vary depending on species, geographic location, and cultivation practices. The average molar weight of carbon unit in starch of maize was assumed as 27 g mol^{-1}. We note that starch synthesis in maize is considerably more complex than in our in vitro chemoenzymatic ASAP. For ASAP iterations, values are means, and error bars indicate SD (n = 3 replicates). NADPH, reduced form of nicotinamide adenine dinucleotide.

Pathway	Substrate	Energy input	Target product	Titer (mg $liter^{-1}$)	Time (hours)	Productivity (mg $liter^{-1}$ $hour^{-1}$)	Starch synthesis rate* (nmol C min^{-1} mg^{-1})
ASAP 1.0	Methanol	–	Amylose	30 ± 1	10	3.0 ± 0.1	0.15 ± 0.01
ASAP 2.0	Methanol	–	Amylose	230 ± 5	10	23 ± 1	2.0 ± 0.1
ASAP 3.0	CO_2	H_2	Amylose	1640 ± 86	4	410 ± 22	22 ± 1†
ASAP 3.1	CO_2	H_2	Amylopectin	1280 ± 6	4	320 ± 2	17.2 ± 0.1†
CETCH 5.4	$NaHCO_3$	NADPH	Glyoxylate	40.0	1.5	26.7	3.87‡
Calvin cycle (maize)	CO_2 and H_2O	Solar	Starch	–	120 days	–	2.58§

*Starch synthesis rate was calculated for the indicated substrates and shown as a nanomole of carbon converted to product per minute per milligram of total proteins. †The rate was calculated by using total amount of both catalyst and proteins (see supplementary materials). ‡The end product of CETCH is glyoxylate. Rate of CETCH was recalculated with 3.1 mg ml^{-1} of total proteins (*13*). §Calculation mainly based on reported data that starch is 26.1% of total biomass and total proteins, excluding storage protein in grain, are 2.17% of total biomass (*40*).

supply material for the key reaction of fls in the C3a module (supplementary text). We thus constructed the alternative C1 module with thermodynamically more favorable reaction cascades (fig. S2, C to E, C1c to e) (*21*, *27*). The most thermodynamically favorable C1e module was successfully assembled with the C3a module and achieved a substantially higher yield of C3 compounds from methanol (fig. S6A).

Assembling C1e + C3a with the C6a module (fig. S4A) produced negligible amounts of the target glucose-6-phosphate (G-6-P) (fig. S6B). We found the carbon flux was kinetically trapped at triose phosphates (fig. S8A) because of the unbalanced activity between dihydroxyacetone kinase (dak) in module C3a and fructose-6-phosphate aldolase (fsa) in module C6a (fig. S8, B and D). Furthermore, glycolaldehyde, which is the by-product of fls-catalyzed reaction in C3a, competitively inhibited the function of fsa (fig. S8, C and D, and supplementary text for more details of the incompatibilities between the C3a and C6a modules). Two alternative modules were constructed on the basis of different classes of aldolase (fig. S4, B and C, C6b and c). However, the extremely low activity of thermophilic fructose-1,6-bisphosphate aldolase/phosphatase at ambient temperature impeded the performance of C6c (table S1 and fig. S4E). For assembly of (C1e + C3a) + C6b, the function of *Escherichia coli* fructose-bisphosphatase (fbp) of module C6b was inhibited by adenosine 5′-triphosphate (ATP) and adenosine 5′-diphosphate (ADP), the essential cofactors of dak in the C3a module (fig. S9, A and B, and supplementary text). By coupling an ATP regeneration system with (C1e + C3a) + C6b [designated as (C1e + C3a) + C6b*], in which ATP was regenerated from ADP by consuming polyphosphate via polyphosphate kinase (table S1), we reduced ATP and ADP to a tolerable level and successfully produced G-6-P from methanol (figs. S9B and S6B).

Assembling C1e + C3a + C6b* with the Cna module (fig. S5A) failed to produce detectable amylose starch from methanol (fig. S6C). We observed that amylose synthesis via α-glucan phosphorylase of Cna was severely inhibited by the high inorganic phosphate (P_i)/α-D-glucose-1-phosphate (G-1-P) ratio, which could be formed from the upper part of the assembly (fig. S10, A and B, and supplementary text). Alternatively, we constructed an ATP-dependent Cnb module (fig. S5B), which is resistant to a high P_i/G-1-P ratio (fig. S10C). The assembly of (C1e + C3a + C6b*) + Cnb enabled 30 mg liter^{-1} amylose starch production from 20 mM methanol (Table 1 and fig. S6C).

With the assistance of computational pathway design and through assembling and substituting 11 modules constructed from a pool

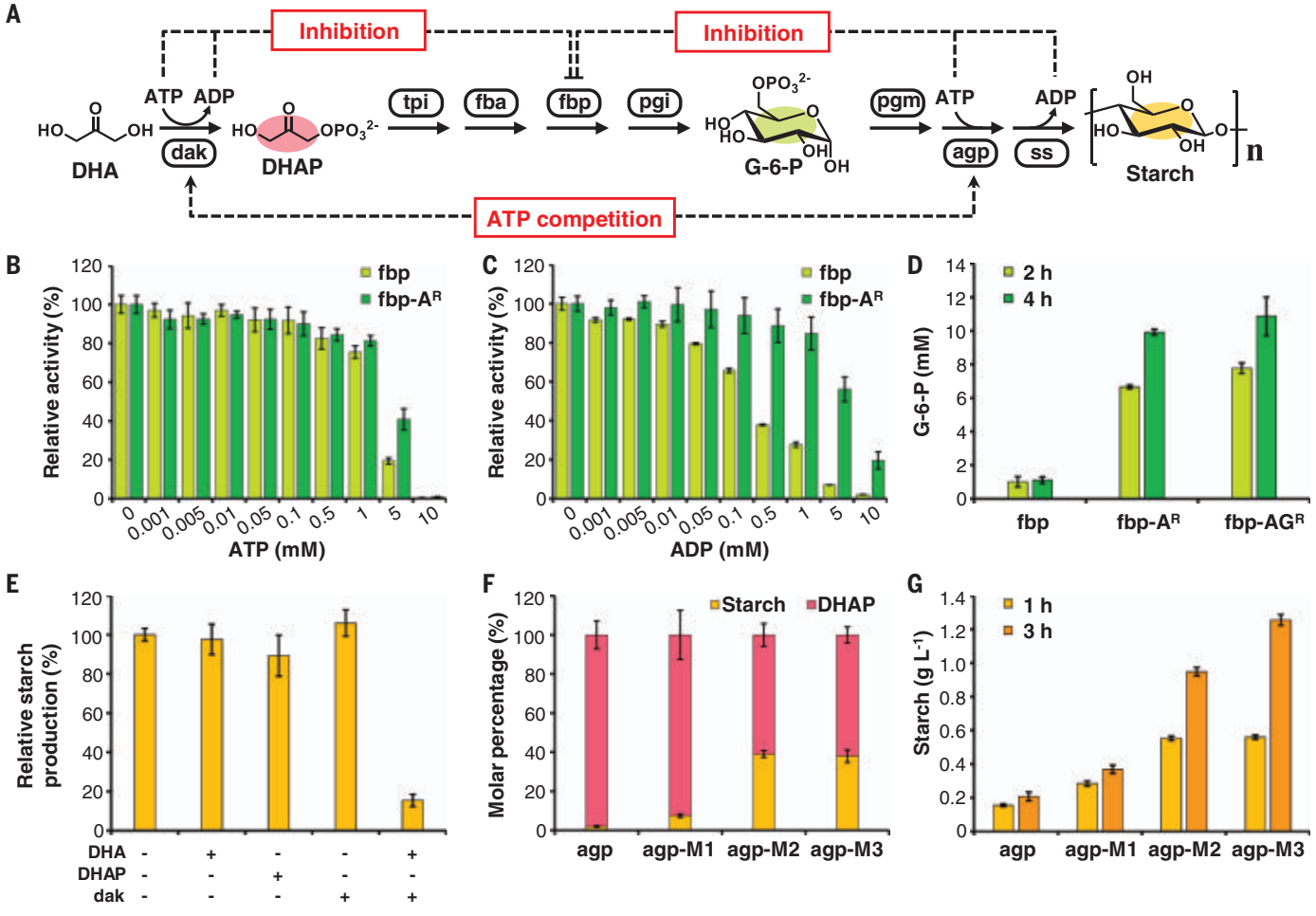


Fig. 2. Resolving main bottlenecks in ASAP. (**A**) Partial ASAP pathway from DHA to starch with bottlenecks indicated and key intermediates and product colored. (**B** and **C**) Inhibitory effect of ATP and ADP on fbp and fbp-A^R. (**D**) G-6-P production from 25 mM DHA via pathway with different fbp variants. (**E**) Inhibition by components from C3a model on the Cnb model. DHA, DHAP, and dak were present at concentrations of 25 mM, 10 mM, and 0.2 mg ml^{-1}, respectively. (**F**) Proportion of DHAP and starch (in glucose) produced from 25 mM DHA and 10 mM G-6-P in a competitive system, which includes dak and the Cnb module, with different agp variants. (**G**) Starch (in glucose) production from 25 mM DHA via partial ASAP as depicted in Fig. 2A. For (B) to (G), values are means, and error bars indicate SD (n = 3 replicates).

of 62 enzymes from 31 organisms (table S1), we established the artificial starch anabolic pathway (ASAP) 1.0 with 10 enzymatic reactions starting with methanol (Fig. 1, outer circle). The main intermediates and target product of ASAP 1.0 were detected by an isotopic ^{13}C-labeling experiment (fig. S11, A and C), validating its full function for starch synthesis from methanol.

After establishing ASAP 1.0, we sought to optimize this pathway by resolving potential bottlenecks. First, because of its low kinetic activity, the enzyme fls accounted for ~86% of the total protein dosage in ASAP 1.0 to sustain the metabolic flux and maintain toxic formaldehyde at a very low level (*28*, *29*). Directed evolution increased the fls catalytic activity, yielding the variant fls-M3 ($fls^{I28L/T90L/N283H}$), which showed an activity improvement of 4.7-fold toward 5 mM formaldehyde and a preference of dihydroxyacetone (DHA) as the main product (fig. S12 and supplementary text).

Even though they were maintained at a low level of 1 mM with the assistance of the regeneration system, ATP and ADP may still partially inhibit the function of *E. coli* fbp (Fig. 2, B and C), which is reported to be allosterically inhibited by adenosine 5′-monophosphate (AMP) (*30*). We found that the variant fbp-A^R ($fbp^{K104Q/R132I}$), which contains two mutations in the AMP allosteric site (*31*), alleviated ADP inhibition (Fig. 2, B and C) and substantially improved G-6-P production from DHA (Fig. 2D). Analysis of the inhibition pattern of the three kinds of nucleotide on fbp and fbp-A^R indicated that ATP or ADP was the determinant for inhibition in the system (table S5 and supplementary text). By integrating fbp-A^R with a reported variant resistant to G-6-P (*31*), a combined variant fbp-AG^R ($fbp^{K104Q/R132I/Y210F/K218Q}$) enabled a further improvement (Fig. 2D and supplementary text).

ATP competition between dak and ADP-glucose pyrophosphorylase (agp) was considered, as an increase in substrate DHA and its kinase dak resulted in an aberrantly lower starch production during the first 4 hours (Fig. 2A and fig. S10D). We confirmed that the coexistence of DHA and dak severely inhibited starch synthesis via Cnb (Fig. 2E) and output DHA phosphate (DHAP) as the dominant product over starch (Fig. 2F, first column), which validates that dak competitively consumed most of the ATP. Instead of reducing the dosage of dak, we tried to enhance the capacity of agp. Three high-activity agp variants were created in accordance with reported amino acid substitutions (*32*, *33*), and these variants displayed enhanced competition against dak (Fig. 2F). The best variant, agp-M3, successfully

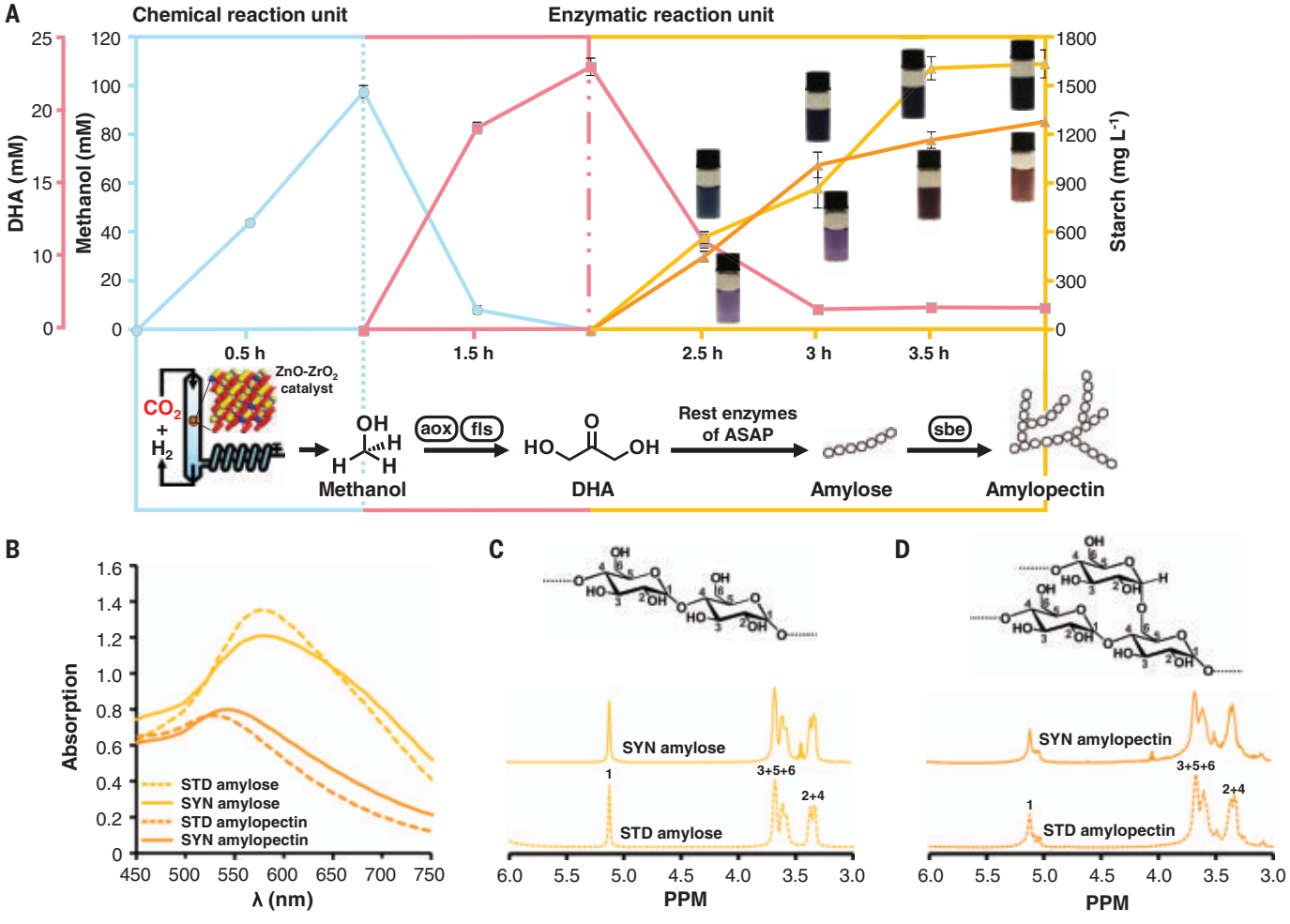


Fig. 3. Starch synthesis via ASAP from CO_2. (**A**) (Bottom) Time course of chemoenzymatic cascade reactions for starch synthesis from CO_2. sbe, starch branching enzyme. (Top) Alteration of key intermediates and starch (yellow curve indicating amylose and orange curve indicating amylopectin) during the time course. The production of amylose and amylopectin are also visualized in the vials by iodine dyeing at given time points. The reaction solution was diluted sixfold before iodine treatment. Values are means, and error bars indicate SD (n = 3 replicates). (**B**) Absorption spectra analysis of synthetic amylose and amylopectin after iodine treatment. The wavelength of maximum absorption (λ_{max}) of standard (STD) and synthetic (SYN) amylose is 577 nm and λ_{max} of standard and synthetic amylopectin is 529 and 543 nm, respectively. (**C** and **D**) ^1H nuclear magnetic resonance (NMR) spectrum of standard (STD) and synthetic (SYN) amylose and amylopectin.

increased starch synthesis from DHA by approximately sixfold (Fig. 2G).

Using these three engineered enzymes (fls-M3, fbp-AG^R, and agp-M3), we constructed ASAP 2.0, which produced ~230 mg liter^{-1} amylose starch in 10 hours from 20 mM methanol (Table 1). Compared with that of ASAP 1.0, the starch productivity of ASAP 2.0 was improved 7.6-fold. On the basis of ^{13}C-labeling liquid chromatography–mass spectrometry (LC-MS) analysis, ASAP 2.0 accumulated a lower concentration of intermediates than ASAP 1.0 (fig. S11, A and B), which indicates the effectiveness of our optimization strategies.

With the above success in ASAP 2.0, we proceeded to synthesize starch from CO_2 and hydrogen by coupling the enzymatic processes with CO_2 reduction by means of a previously developed inorganic catalyst, ZnO-ZrO_2 (*34*). Because of the unfavorable conditions of CO_2 hydrogenation, we developed a chemoenzymatic cascade system in ASAP 3.0 with a chemical reaction unit and an enzymatic reaction unit. To satisfy the demand of fls for a high concentration of formaldehyde and to avoid its toxicity to other enzymes (fig. S13), we further operated the enzymatic unit with two steps (Fig. 3A). In the chemical reaction unit, CO_2 was chemically hydrogenated to methanol at a rate of ~0.25 g hour^{-1} g^{-1} catalyst, and the produced methanol was constantly condensed and fed into the enzymatic unit to a final concentration of ~100 mM during the first hour. In the enzymatic unit, the methanol was first converted to ~22.5 mM C3 intermediate DHA for another 1 hour by supplementing two core enzymes and auxiliary catalase (cat) and then transformed to ~1.6 g liter^{-1} amylose starch in the subsequent 2 hours by supplementing the remaining eight core enzymes and auxiliary components (Fig. 3A). The synthetic amylose exhibited the same deep blue color and absorption maximum as standard amylose in the presence of iodine solution (Fig. 3B).

Natural starch contains ~20 to 30% amylose and 70 to 80% amylopectin (*3*). To synthesize amylopectin from CO_2, we introduced a starch branching enzyme (sbe) from *Vibrio vulnificus* (*35*) in ASAP 3.1. This setup produced ~1.3 g liter^{-1} amylopectin within 4 hours (Fig. 3A). The synthetic amylopectin had a reddish-brown color and a comparable absorption maximum as standard amylopectin after iodine treatment (Fig. 3B). Both the synthetic amylose and amylopectin exhibited the same one to six proton nuclear magnetic resonance signals as their standard counterparts (Fig. 3, C and D).

By using spatial and temporal segregation of steps, ASAP 3.0 achieved a high starch productivity of ~410 mg liter^{-1} hour^{-1} from CO_2. The starch synthesis rate of this chemoenzymatic pathway reached 22 nmol min^{-1} mg^{-1} of total catalyst and proteins, which is an 8.5-fold higher rate than that of starch synthesis via the Calvin cycle in maize (Table 1). This rate is also 5.7-fold higher than that of the synthetic crotonyl–coenzyme A (CoA)/ethylmalonyl-CoA/hydroxybutyryl-CoA (CETCH) cycle which has been recently extended into a platform to access different compounds directly from CO_2 (*12*, *13*, *36*). The theoretical hydrogen-to-methanol energy efficiency (η_{HME}) and methanol-to-starch energy efficiency (η_{MSE}) of ASAP is 85 and 61%, respectively, although these values do not consider energy consumption for processes such as enzyme production and maintenance of high temperature and pressure in the chemical step (supplementary text), which will compromise the energy efficiency of ASAP in practice. With an attainable solar-to-electricity efficiency (η_{SEE}) of 20% (*17*) and electricity-to-hydrogen efficiency (η_{EHE}) of 85% (*18*) in ideal photovoltaic and water-electrolysis devices, the theoretical maximal solar-to-starch efficiency ($\eta_{SSE} = \eta_{SEE} \times \eta_{EHE} \times \eta_{HME} \times \eta_{MSE}$) via ASAP will be 9%. With the estimated practical η_{HME}' of 68% considering the energy for temperature and pressure in the chemical step (*37*), the theoretical η_{SSE} is adjusted to 7%, which is comparable to the theoretical photosynthetic efficiency of solar energy to biomass for C3 (4.6%) and C4 (6%) plants (*38*) and is 3.5 fold of the estimated theoretical solar-to-starch efficiency for plants (2%) in a natural environment (*39*). Cell-free, chemoenzymatic, and efficient starch synthesis from CO_2 by ASAP provides an important starting point for applications such as industrial biomanufacturing of starch.

REFERENCES AND NOTES

1. P. L. Keeling, A. M. Myers, *Annu. Rev. Food Sci. Technol.* **1**, 271–303 (2010).
2. R. Höfer, "Sugar- and starch-based biorefineries" in *Industrial Biorefineries and White Biotechnology*, A. Pandey, R. Höfer, M. Taherzadeh, M. Nampoothiri, C. Larroche, P. L. Keeling, A. M. Myers, Eds. (Elsevier, 2015), chap. 4A, pp. 157–235.
3. R. L. Whistler, J. R. Daniel, "Molecular structure of starch" in *Starch: Chemistry and Technology*, R. L. Whistler, J. N. Bemiller, E. F. Paschall, Eds. (Academic Press, ed. 2, 1984), pp. 153–182.
4. T. Saithong, A. Meechai, S. Cheevadhanarak, S. Bhumiratana, *J. Comput. Sci. Syst. Biol.* **5**, 24–37 (2012).
5. M. R. Abt, S. C. Zeeman, *Curr. Opin. Plant Biol.* **55**, 109–117 (2020).
6. J. Li *et al.*, *Plant Cell Physiol.* **54**, 282–294 (2013).
7. N. Li, S. Zhang, Y. Zhao, B. Li, J. Zhang, *Planta* **233**, 241–250 (2011).
8. M. Hakata *et al.*, *Plant Biotechnol. J.* **10**, 1110–1117 (2012).
9. A. Bahaji *et al.*, *Biotechnol. Adv.* **32**, 87–106 (2014).
10. A. Bar-Even, E. Noor, N. E. Lewis, R. Milo, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 8889–8894 (2010).
11. A. Satanowski *et al.*, *Nat. Commun.* **11**, 5812 (2020).
12. M. Scheffen *et al.*, *Nat. Catal.* **4**, 105–115 (2021).
13. T. Schwander, L. Schada von Borzyskowski, S. Burgener, N. S. Cortina, T. J. Erb, *Science* **354**, 900–904 (2016).
14. T. E. Miller *et al.*, *Science* **368**, 649–654 (2020).
15. T. P. Korman, P. H. Opgenorth, J. U. Bowie, *Nat. Commun.* **8**, 15526 (2017).
16. S. Sherkhanov *et al.*, *Nat. Commun.* **11**, 4292 (2020).
17. M. Green *et al.*, *Prog. Photovolt. Res. Appl.* **29**, 3–15 (2021).
18. S. Shiva Kumar, V. Himabindu, *Mater. Sci. Energy Technol.* **2**, 442–454 (2019).
19. Y. Y. Birdja *et al.*, *Nat. Energy* **4**, 732–745 (2019).
20. R.-P. Ye *et al.*, *Nat. Commun.* **10**, 5698 (2019).
21. J. B. Siegel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3704–3709 (2015).
22. R. Caspi *et al.*, *Nucleic Acids Res.* **44**, D471–D480 (2016).
23. N. Hadadi, J. Hafner, A. Shajkofci, A. Zisaki, V. Hatzimanikatis, *ACS Synth. Biol.* **5**, 1155–1166 (2016).
24. X. Yang *et al.*, *Metab. Eng.* **56**, 142–153 (2019).
25. A. Ebrahim, J. A. Lerman, B. O. Palsson, D. R. Hyduke, *BMC Syst. Biol.* **7**, 74 (2013).
26. T. J. Erb, P. R. Jones, A. Bar-Even, *Curr. Opin. Chem. Biol.* **37**, 56–62 (2017).
27. A. Chou, J. M. Clomburg, S. Qian, R. Gonzalez, *Nat. Chem. Biol.* **15**, 900–906 (2019).
28. Y. S. Tai, K. Zhang, *Nat. Chem. Biol.* **11**, 384–386 (2015).
29. S. Poust *et al.*, *ChemBioChem* **16**, 1950–1954 (2015).
30. J. K. Hines, C. E. Kruesel, H. J. Fromm, R. B. Honzatko, *J. Biol. Chem.* **282**, 24697–24706 (2007).
31. J. S. Yang, S. W. Seo, S. Jang, G. Y. Jung, S. Kim, *PLOS Comput. Biol.* **8**, e1002612 (2012).
32. C. R. Meyer, J. A. Bork, S. Nadler, J. Yirsa, J. Preiss, *Arch. Biochem. Biophys.* **353**, 152–159 (1998).
33. C. R. Meyer, J. Yirsa, B. Gott, J. Preiss, *Arch. Biochem. Biophys.* **352**, 247–254 (1998).
34. J. Wang *et al.*, *Sci. Adv.* **3**, e1701290 (2017).
35. H. J. Jo, S. Park, H. G. Jeong, J. W. Kim, J. T. Park, *FEBS Lett.* **589**, 1089–1094 (2015).
36. S. Sundaram *et al.*, *Angew. Chem. Int. Ed.* **60**, 16420–16425 (2021).
37. E. S. Van-Dal, C. Bouallou, *J. Clean. Prod.* **57**, 38–45 (2013).
38. X. G. Zhu, S. P. Long, D. R. Ort, *Curr. Opin. Biotechnol.* **19**, 153–159 (2008).
39. Y. H. P. Zhang, C. You, H. Chen, R. Feng, "Surpassing photosynthesis: high-efficiency and scalable CO_2 utilization through artificial photosynthesis" in *Recent Advances in Post-Combustion CO_2 Capture Chemistry*. M. Attalla, Ed. (American Chemical Society, 2012), pp. 275–292.
40. C. S. Kuehn, J. G. Linn, D. G. Johnson, H. G. Jung, M. I. Endres, *J. Dairy Sci.* **82**, 2746–2755 (1999).

ACKNOWLEDGMENTS

We thank P. Kang from Tianjin University for introducing us to the fundamental principles of CO_2 electroreduction; Y. P. Zhang and J. Zhou (Institute of Microbiology) for granting the genome of *Clostridium acetobutylicum* and *Syn.* PCC 7942; H. Song for providing purified cat enzyme; P. Liu for molecular dynamic simulation; Z. D. Zhang and Q. C. Cao for LC-MS analysis; Y. Cai for NMR analysis; and Y. H. Yao for gas chromatography–mass spectrometry analysis. We thank G. P. Zhao, M. T. Reetz, and Z. T. Sun for advice on manuscript revision. We thank the core facility center at Tianjin Institution of Industrial Biotechnology, CAS, for instrument and technology support. We dedicate this work to the memory of Professor Arren Bar-Even, who devoted his life to synthetic pathway design and synthetic biology. **Funding:** This work was supported by the Tianjin Synthetic Biotechnology Innovation Capacity Improvement Action (grant TSBICIP-KJGG-008 to T.C.), the Key Research Program of the Chinese Academy of Sciences (grant ZDRW-ZS-2016-3 to Y.M. and T.C.), the Strategic Priority Research Program of the Chinese Academy of Sciences-Precision Seed Design and Breeding (grant XDA24020103-3 to H.J.), the Tianjin Outstanding Scholar Program (to Y.M.), Youth Innovation Promotion Association of CAS (grant 2016165 to T.C.), and the National Science Fund for Excellent Young Scholars (grant 31922047 to H.J.). **Author contributions:** Conceptualization: Y.M. and T.C.; Methodology: Y.M., T.C., H.J., H.M., L.Z., C.L., C.Y., and Y.S.; Investigation: T.C., H.S., J.Q., L.Z., F.Z., J.Z., Z.T., X.W., J.Y., Q.Y., W.W., X.Y., H.C., and Qian Wang; Visualization: T.C., Qinhong Wang, L.Z., and H.J.; Funding acquisition: Y.M., T.C., and H.J.; Project administration: Qinhong Wang and T.C.; Supervision: Y.M.; Writing—original draft: Y.M., T.C., Qinhong Wang, H.J., and L.Z.; Writing—review and editing: Y.M., T.C., Qinhong Wang, H.J., L.Z., C.L., and Y.L. **Competing interests:** This work is included in patent applications by the Tianjin Institute of Industrial Biotechnology, CAS: patent 202010858974.9 covering a method of starch synthesis from C1 feedstocks and patent 202010044853.0 covering the application of the mutants of fls. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abh4049
Materials and Methods
Supplementary Text
Figs. S1 to S13
Tables S1 to S7
References (*41*–*83*)

8 March 2021; accepted 12 August 2021
10.1126/science.abh4049

PALEOANTHROPOLOGY

Evidence of humans in North America during the Last Glacial Maximum

Matthew R. Bennett[1]*, David Bustos[2], Jeffrey S. Pigati[3], Kathleen B. Springer[3], Thomas M. Urban[4], Vance T. Holliday[5,6], Sally C. Reynolds[1], Marcin Budka[1], Jeffrey S. Honke[3], Adam M. Hudson[3], Brendan Fenerty[6], Clare Connelly[2], Patrick J. Martinez[2], Vincent L. Santucci[7], Daniel Odess[8]*

Archaeologists and researchers in allied fields have long sought to understand human colonization of North America. Questions remain about when and how people migrated, where they originated, and how their arrival affected the established fauna and landscape. Here, we present evidence from excavated surfaces in White Sands National Park (New Mexico, United States), where multiple in situ human footprints are stratigraphically constrained and bracketed by seed layers that yield calibrated radiocarbon ages between ~23 and 21 thousand years ago. These findings confirm the presence of humans in North America during the Last Glacial Maximum, adding evidence to the antiquity of human colonization of the Americas and providing a temporal range extension for the coexistence of early inhabitants and Pleistocene megafauna.

The late Pleistocene expansion of human populations into the Americas is the latest chapter in the "Out of Africa" migration of modern humans (*1*). The earliest evidence for dispersal and occupation remains controversial (*2*). Hypotheses range from those that favor Clovis as the earliest occupation [~13 thousand years ago (ka)] (*3*) to those that favor older ("pre-Clovis") sites dating to ~16.5 ka [i.e., post–Last Glacial Maximum (LGM); 26.5 to 19 or 20 ka (*4*)] or even older (*5*). Viable migration routes from Asia were dependent on timing and associated environmental conditions and could have occurred via an inland route through the Ice-Free Corridor (*6*, *7*), the Pacific Coastal Route (*8*, *9*), or both; however, these routes would have been closed or at least unlikely during the LGM (*10*).

Most early sites in North America, defined here as early post–LGM or older, are questioned because of either their dating (*3*) or uncertainty about the association of humans and purported artifacts (*11*). Fossil human footprints provide an alternative source of evidence for human presence when excavated from an in situ sedimentary sequence with good chronological control [e.g., (*12*–*14*)]. Unlike cultural artifacts, modified bones, or other more conventional fossils, footprints have a primary depositional context and are fixed on the imprinted surface.

[1]Institute for Studies in Landscapes and Human Evolution, Bournemouth University, Poole, BH12 5BB, UK. [2]National Park Service, White Sands National Park, P.O. Box 1086, Holloman AFB, NM 88330, USA. [3]US Geological Survey, Denver Federal Center, Box 25046, MS 980, Denver, CO 80225, USA. [4]Department of Classics and Cornell Tree Ring Laboratory, Cornell University, Ithaca, NY 14853, USA. [5]School of Anthropology, P.O. Box 210030, University of Arizona, Tucson, AZ 85721-0030, USA. [6]Department of Geosciences, University of Arizona, Tucson, AZ 85721-0030, USA. [7]National Park Service, Geologic Resources Division, 1849 C Street NW, Washington, DC 20240, USA. [8]National Park Service, Cultural Resources Directorate, Washington, DC 20240, USA.
*Corresponding author. Email: mbennett@bournemouth.ac.uk (M.R.B); daniel_odess@nps.gov (D.O.)

In this context, we report the discovery of a series of human footprints from White Sands National Park (WHSA Locality 2; Fig. 1A) in south-central New Mexico that occur on multiple stratigraphic horizons interbedded with seed layers. Our results indicate that humans were present on the landscape by at least ~23 ka, with evidence of occupation spanning approximately two millennia. These data provide definitive evidence of human occupation of North America south of the Laurentide Ice Sheet during the LGM.

WHSA lies within the Tularosa Basin, a topographically-closed basin that encompasses an area of ~14,000 km^2. The floor of the basin contains a gypsum playa (Alkali Flat; Fig. 1A and fig. S1) created by eolian deflation of late Pleistocene lake beds after the desiccation of pluvial Lake Otero. The precise history of Lake Otero is not currently known, but it was the largest of several perennial lakes that filled the basin between ~36 and 19 ka (*15*). Throughout this period, changes in the hydrological budget in the Tularosa Basin led to a range of water levels and therefore lake and wetland extents associated with fluctuations in precipitation, groundwater input, and inflow or outflow. Ichnofossils of extinct late Pleistocene fauna occur widely on the margins of the playa and include tracks of Proboscidea (mammoth), Folivora (ground sloth), Carnivora (canid and felid), and Cetartiodactyla (bovid and camelid), most of which are associated with human footprints (*16*–*18*).

WHSA Locality 2 is located on the eastern side of Alkali Flat on a shallow (<6 m) erosional scarp formed between the current playa and the White Sands gypsum dune field. A sedimentary sequence exposed by trenching consists of 1.25 m of lacustrine clays and silts intercalated with thinly bedded gypsiferous and siliciclastic sands, silts, and clays, which represent a transition from a lacustrine ecosystem to an alluvial regime in response to changing hydroclimate conditions during the late Pleistocene (Fig. 2 and fig. S2). Endocarps of ditch grass (*Ruppia cirrhosa*) occur as discontinuous layers throughout the sequence. Footprints (human, proboscidean, and canid) occur at all levels, both in cross sections and on excavated surfaces adjacent to the trench (Figs. 1, B to E, and 2 and figs. S4 to S10). A total of 61 human tracks have been identified and documented.

Human footprints reported from all surfaces have good anatomical definition (i.e., visible heel impressions, medial longitudinal arches, and toe pads), consistent with modern humans when compared to other Pleistocene footprint sites and the ichnotaxa *Hominipes modernus* (*19*, *20*). A geomorphometric comparison of a sample of WHSA tracks with a set of modern footprints (n = 356) and fossil footprints from Namibia (n = 78; supplementary text) reveals broad similarity (fig. S3). The WHSA tracks, similar to the fossil tracks from Namibia, are flatter-footed than the modern samples, similar to what is commonly reported for habitually unshod individuals [e.g., (*21*)]. The WHSA footprints also have longer toe pads that we suggest are associated with slippage of the foot during locomotion (*18*).

Seven footprint surfaces [track horizons 1 to 7 (TH1 to TH7)] and one multisurface, track-bearing package (TH8) are present. All footprint surfaces—including TH4 with 37 tracks— have been excavated, but we draw particular attention to TH1 and TH5 through TH7, which occur at depth in the base of the trenches. TH1, the stratigraphically lowest footprint horizon, is exposed in the north-south side trench where it exhibits several human tracks, including a two-step trackway (TH1; Fig. 2 and figs. S4 and S5). TH5 occurs as a surface at the western end of the site but can also be traced laterally 9.35 m onto the floor of the main east-west trench where two footprints occur with different orientations, separated vertically from one another by a few centimeters of sediment (TH5 and TH6; figs. S2 and S10). An additional footprint was also found in the base of the trench ~32 mm higher and 350 mm to the east (TH7; fig. S2). Inferences of stature, the age of the track-maker, and walking speeds are summarized in Fig. 2B, although caution is required as a result of the uncertain anthropometric affinity of the track-makers' population. Many tracks appear to be those of teenagers and children; large adult footprints are less frequent. One hypothesis for this is the division of labor, in which adults are involved in skilled tasks whereas "fetching and carrying" are delegated to teenagers. Children accompany the teenagers, and collectively they leave a higher number of footprints that are preferentially recorded in the fossil record. This pattern is common to all excavated surfaces.



Fig. 1. Study area and ancient footprints found at WHSA Locality 2. (**A**) Map showing approximate location of the study site. In accordance with the Archaeological Resources Protection Act of 1979, the National Parks Omnibus Management Act of 1998, and the Paleontological Resources Preservation Act of 2009, the precise location of the site is withheld. Interested parties may contact the National Park Service for this information, given a legitimate reason. (**B**) Human footprints on track horizon 4 (TH4). (**C**) Human footprint on TH5, located in the base of the main trench. (**D**) Surfaced 3D model of part of the main trench showing three human footprints on different surfaces on the trench floor. (**E**) Human trackways on TH4.

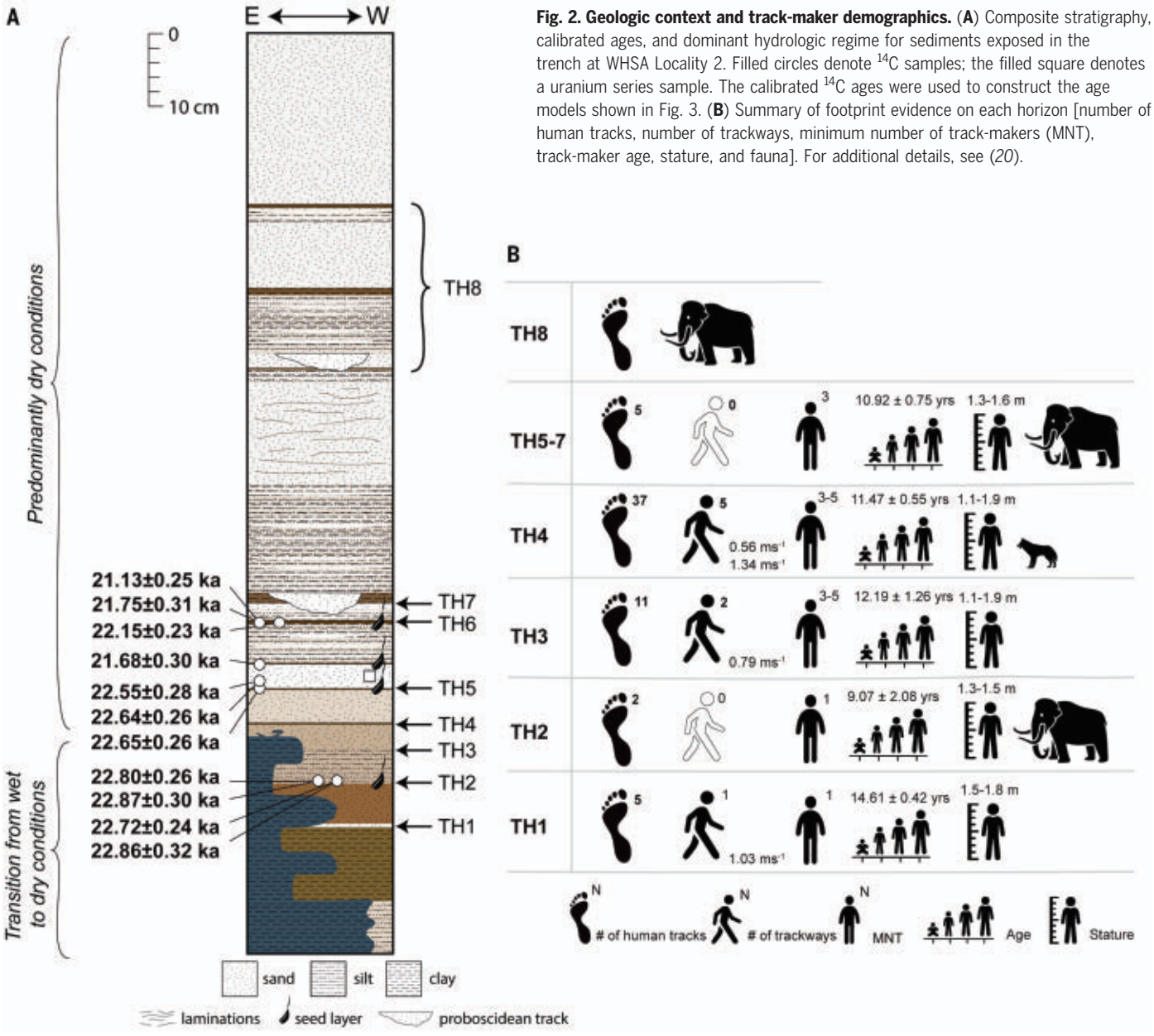


Fig. 2. Geologic context and track-maker demographics. (**A**) Composite stratigraphy, calibrated ages, and dominant hydrologic regime for sediments exposed in the trench at WHSA Locality 2. Filled circles denote ^{14}C samples; the filled square denotes a uranium series sample. The calibrated ^{14}C ages were used to construct the age models shown in Fig. 3. (**B**) Summary of footprint evidence on each horizon [number of human tracks, number of trackways, minimum number of track-makers (MNT), track-maker age, stature, and fauna]. For additional details, see (*20*).

We established chronological control for the footprint-bearing sediments by using radiocarbon dating of in situ layers of macroscopic seeds of the aquatic plant *Ruppia cirrhosa* (*20*). Seed layers were sampled from throughout the sedimentary sequence and yielded calibrated ages that range from 22.86 ± 0.32 to 21.13 ± 0.25 ka (n = 11) and maintain stratigraphic order within the uncertainties between TH2 and TH6 (Fig. 2). Radiocarbon dating of aquatic material, such as *R. cirrhosa*, may be subject to hard-water (or reservoir) effects that could conceivably make these ages too old, so we explored the potential for these effects at WHSA Locality 2 using three lines of evidence (*20*). First, the geologic and hydrologic settings along the shallow lake margin that hosts the human trackways make it unlikely that stands of emergent aquatic plants would harbor prominent hard-water effects. Second, our calibrated ^{14}C ages maintained stratigraphic order even when samples were separated by a few centimeters, which would not be the case if hard-water effects were large and variable. Finally, we evaluated independent chronologic data from the Tularosa Basin (*15*) and found that terrestrial and aquatic material yielded concordant ^{14}C ages between ~44 and 25 ka, which demonstrates that hard-water effects in pluvial Lake Otero persisted less than a few hundred years during this period (fig. S14). In our view, an improbable series of events would be required to introduce a large hard-water effect by ~23 ka when such effects were minimal for the previous ~20 thousand years. For these reasons, we conclude that our radiocarbon-based chronology is robust.

To determine the underlying distribution of the calibrated ^{14}C ages from WHSA Locality 2, we applied the kernel density estimation model (KDE_Model) function of OxCal4.4.2 r:5 (*22*, *23*). In addition, we used the OxCal boundary

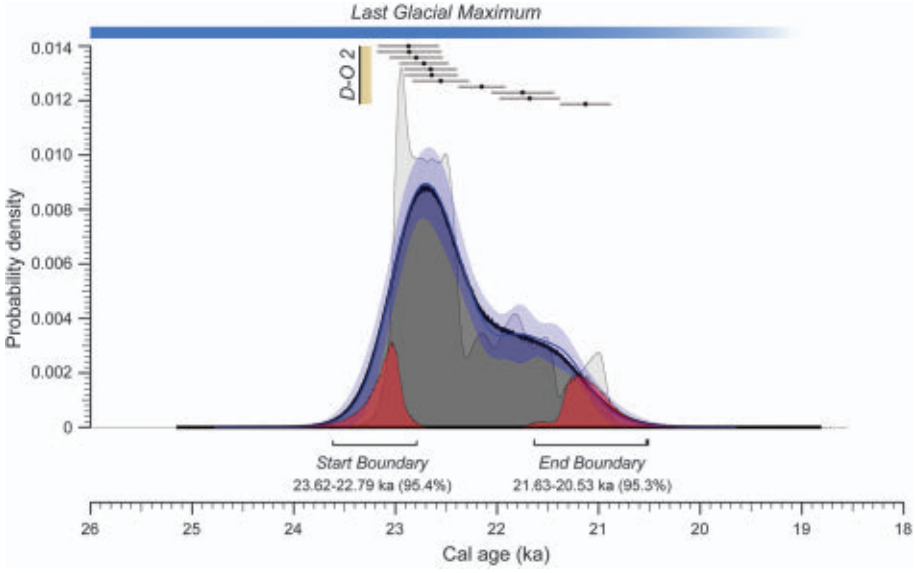


Fig. 3. Age dating and modeling results. KDE modeling results for the 11 calibrated ages corresponding to TH2 through TH6. The dark gray distribution outlined in black is the KDE distribution. The overlying blue line and lighter blue band show the mean and ±1σ for snapshots (n = 1000) of the KDE distribution, which indicate the importance of any features. The light gray distribution is the sum distribution for reference, and the red distributions are the start and end of the boundary distribution. This figure was generated using the KDE_Model and boundary functions in OxCal v.4.4.2 r:5 (*22*) and the IntCal20 calibration curve (*23*); accessed online 8 December 2020. Timing of the LGM (26.5 to 19 or 20 ka) (*4*) is shown by the horizontal blue bar; timing of Dansgaard-Oeschger event 2 (D-O 2, beginning at 23.3 ka) (*24*) is shown by the vertical tan bar.

function to determine the start and end date represented by the calibrated ages and the interval function to determine the amount of time represented by the track horizons (*20*). Our data and modeling results show that the ages of TH2 through TH6 span from ~23 to 21 ka (Fig. 3 and fig. S16), placing humans in southwestern North America for approximately two millennia during the LGM. The presence of proboscidean tracks in TH8 places an additional constraint on the upper age of the sequence, showing that it does not extend beyond the late Pleistocene.

The depositional sequence reported here exhibits a potential paleoclimate signal in which lacustrine conditions were succeeded by alluvial sedimentation that occurred in temporal correspondence with abrupt warming during Dansgaard-Oeschger event 2, beginning at ~23.3 ka (*24*). This drying event resembles sequences observed in other paleohydrological records in the southwestern US (*25*, *26*). In the Tularosa Basin, the lowered lake levels would have exposed a wide area along the lake margins where humans and megafauna traversed.

The evidence presented here confirms that humans were present in North America before the glacial advances of the LGM closed the Ice-Free Corridor (*9*, *27*) and the Pacific Coastal Route and prevented human migration from Asia (*7*). The overlap of humans and megafauna for at least two millennia during this time suggests that if people were hunting megafauna the practices were sustainable, at least initially. This also raises the possibility of a human role in poorly understood megafauna extinctions previously thought to predate their arrival (*28*, *29*) and makes "early" sites in the Americas appear more plausible. However, exactly when people first arrived in the Western Hemisphere and when continuous occupation was established are still both uncertain and contested. What we present here is evidence of a firm time and location when humans were present in North America.

REFERENCES AND NOTES

1. D. J. Meltzer, *The Great Paleolithic War: How Science Forged an Understanding of America's Ice Age Past* (Univ. of Chicago Press, 2015).
2. M. R. Waters, *Science* **365**, eaat5447 (2019).
3. S. J. Fiedel, B. A. Potter, J. E. Morrow, M. K. Faught, C. Vance Haynes Jr., *PaleoAmerica* **7**, 28–42 (2021).
4. P. U. Clark *et al*., *Science* **325**, 710–714 (2009).
5. L. Becerra-Valdivia, T. Higham, *Nature* **584**, 93–97 (2020).
6. R. J. Dawe, M. Kornfeld, *Quat. Int.* **444**, 56–71 (2017).
7. B. A. Potter *et al*., *Sci. Adv.* **4**, eaat5473 (2018).
8. A. J. Lesnek, J. P. Briner, C. Lindqvist, J. F. Baichtal, T. H. Heaton, *Sci. Adv.* **4**, eaar5040 (2018).
9. T. J. Braje *et al*., *Am. Antiq.* **85**, 1–21 (2019).
10. A. S. Dalton *et al*., *Quat. Sci. Rev.* **234**, 106223 (2020).
11. J. V. Ferraro *et al*., *Nature* **554**, E1–E2 (2018).
12. M. R. Bennett *et al*., *Science* **323**, 1197–1201 (2009).
13. N. Ashton *et al*., *PLOS ONE* **9**, e88329 (2014).
14. J. Duveau, G. Berillon, C. Verna, G. Laisné, D. Cliquet, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 19409–19414 (2019).
15. B. D. Allen, D. W. Love, R. G. Myers, *N. M. Geol.* **31**, 9–25 (2009).
16. D. Bustos *et al*., *Sci. Adv.* **4**, eaar7621 (2018).
17. T. M. Urban *et al*., *Sci. Rep.* **9**, 16470 (2019).
18. M. R. Bennett *et al*., *Quat. Sci. Rev.* **249**, 106610 (2020).
19. J. Y. Kim, K. S. Kim, M. G. Lockley, N. Matthews, *Ichnos* **15**, 126–139 (2008).
20. See supplementary materials for information on methods, materials, and additional text.
21. P. Francis, G. Schofield, *BMJ Open Sport Exerc. Med.* **6**, e000577 (2020).
22. C. Bronk Ramsey, OxCal, version 4.4.2, r:5 (2020); https://c14.arch.ox.ac.uk/oxcal.html.
23. P. J. Reimer *et al*., *Radiocarbon* **62**, 725–757 (2020).
24. A. Svensson *et al*., *Clim. Past* **4**, 47–57 (2008).
25. K. B. Springer, C. R. Manker, J. S. Pigati, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14522–14526 (2015).
26. K. M. Menking, V. J. Polyak, R. Y. Anderson, Y. Asmerom, *Quat. Sci. Rev.* **200**, 237–252 (2018).
27. P. D. Heintzman *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8057–8063 (2016).
28. J. M. Broughton, E. M. Weitzel, *Nat. Commun.* **9**, 5441 (2018).
29. D. J. Meltzer, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 28555–28563 (2020).
30. M. R. Bennett *et al*., Data release for Evidence for humans in North America during the Last Glacial Maximum, U.S. Geological Survey data release (2021); doi:10.5066/P9ABZEM9.
31. M. R. Bennett *et al*., Evidence of humans in North America during the Last Glacial Maximum: 3D Track Data, BORDaR (2021); doi:10.18746/bmth.data.00000186.

ACKNOWLEDGMENTS

Views and conclusions are those of the authors and do not necessarily reflect policies of the National Park Service (NPS). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the US government. Research at WHSA was conducted under a permit issued to D.O. (permit number 21-WHSA-01). **Funding:** M.R.B. acknowledges Natural Environment Research Council grants (NE/H004246/1 and NE/M021459/1). V.T.H. acknowledges the Argonaut Archaeological Research Fund (University of Arizona Foundation). Funding for J.S.P. and K.B.S. was provided by the US Geological Survey's Climate Research and Development Program and by the National Park Service through Interagency Agreement G14P20PG00274. T.M.U. acknowledges cooperative agreement P20AC00605 between Cornell University and the US Department of the Interior. D.B. and M.R.B. thank the Western National Parks Association for support. **Author contributions:** M.R.B., D.B., T.M.U., D.O., and S.C.R. conceptualized the project, were responsible for its management and resource acquisition, and did the fieldwork along with J.S.P. and K.B.S., who designed and led the stratigraphic analysis and radiocarbon dating. C.C., V.L.S., and P.J.M. provided field support, and J.S.H. and A.M.H. provided laboratory support. M.B. wrote several of the algorithms used in the footprint analysis. V.T.H. and B.F. provided supporting data. Analysis and data visualization were undertaken by M.R.B., K.B.S., J.S.P., and D.B. The original manuscript was drafted by M.R.B., with contributions from J.S.P., K.B.S., V.T.H., and D.O. Editing of the manuscript was undertaken by M.R.B., S.C.R., J.S.P., K.B.S., D.O., D.B., and V.T.H. **Competing interests:** The authors declare no competing or conflicting interests. **Data and materials availability:** Downloadable files of the data presented in the supplementary materials (tables S1 to S8) can be found at (*30*). The 3D footprints models are available at (*31*).

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abg7586
Supplementary Text
Figs. S1 to S19
Tables S1 to S8
References (*32*–*75*)

26 January 2021; accepted 25 August 2021
10.1126/science.abg7586

PLANT SCIENCE

A plant lipocalin promotes retinal-mediated oscillatory lateral root initiation

Alexandra J. Dickinson[1,2,3,4,5]*, Jingyuan Zhang[1], Michael Luciano[6], Guy Wachsman[1,4], Evan Sandoval[5], Martin Schnermann[6], José R. Dinneny[2,3], Philip N. Benfey[1,4]*

In *Arabidopsis*, de novo organogenesis of lateral roots is patterned by an oscillatory mechanism called the root clock, which is dependent on unidentified metabolites. To determine whether retinoids regulate the root clock, we used a chemical reporter for retinaldehyde (retinal)–binding proteins. We found that retinal binding precedes the root clock and predicts sites of lateral root organogenesis. Application of retinal increased root clock oscillations and promoted lateral root formation. Expression of an *Arabidopsis* protein with homology to vertebrate retinoid-binding proteins, TEMPERATURE INDUCED LIPOCALIN (TIL), oscillates in the region of retinal binding to the reporter, confers retinal-binding activity in a heterologous system, and, when mutated, decreases retinal sensitivity. These results demonstrate a role for retinal and its binding partner in lateral root organogenesis.

Plants continuously develop postembryonic lateral roots to forage for water and nutrients. The root clock, a temporal series of oscillating changes in gene expression, prepatterns sites for initiation of lateral root primordia (*1*, *2*). Blocking carotenoid metabolism genetically or with inhibitors is sufficient to dampen root clock oscillations and prevent lateral root initiation (*3*, *4*). Known apocarotenoid regulators of root development, including abscisic acid, strigolactone, anchorene, and β-cyclocitral, do not rescue lateral root initiation in plants with reduced carotenoid metabolism (*3*, *5*–*8*). In vertebrate development, β-carotene–derived retinoic acid has important functions in the oscillatory somitogenesis clock (*9*), neurogenesis, and vasculature development (*10*–*12*). Therefore, we hypothesized that retinoids may have similar developmental functions across different domains of life and that retinal signaling may play a role in the root clock.

To examine the potential for retinal signaling in the root, we deployed merocyanine aldehyde (MCA), a chemical reporter designed to measure vertebrate retinoid-binding protein activity (*13*). MCA has a cyanine-5 (Cy5) head group fused to a retinal-like moiety that fluoresces when bound to a lysine residue in retinoid-binding proteins (Fig. 1A). To test for the

[1]Department of Biology, Duke University, Durham, NC, USA. [2]Department of Plant Biology, Carnegie Institute of Science, Stanford, CA, USA. [3]Department of Biology, Stanford University, Palo Alto, CA, USA. [4]Department of Biology, Howard Hughes Medical Institute, Duke University, Durham, NC, USA. [5]Section of Cell and Developmental Biology, University of California San Diego, La Jolla, CA, USA. [6]Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD, USA.
*Corresponding author. Email: adickinson@ucsd.edu (A.J.D.); philip.benfey@duke.edu (P.N.B.)

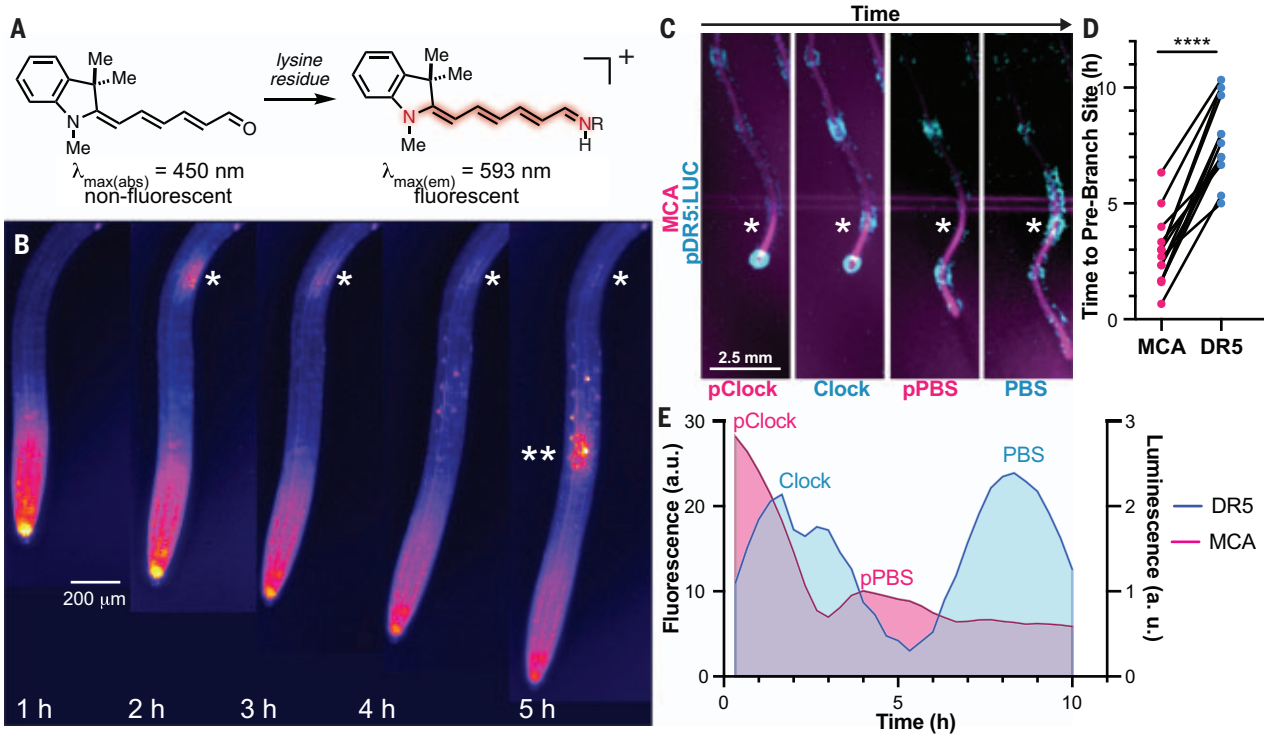


Fig. 1. A reporter for retinal-binding proteins predicts sites of lateral root organogenesis in *Arabidopsis*. (**A**) The chemical structure and fluorescence properties of merocyanine aldehyde (MCA) in the absence and presence of a lysine residue from a retinal-binding protein. (**B**) Time course of MCA fluorescence dynamics in *Arabidopsis* roots. The single and double asterisks indicate the site of fluorescent oscillations in the early differentiation zone of the root. (**C**) Representative images showing the spatial and temporal dynamics of MCA oscillations and the root clock (pDR5:LUC) in a growing root. (**D**) The pulse of MCA fluorescence indicates the presumptive prebranch site (PBS) before pDR5:LUC (time is measured from the start of the pDR5:LUC root clock oscillation). ****$P < 0.0001$. (**E**) A representative example of the temporal signal from MCA and pDR5:LUC in the region of the root that becomes the PBS [this graph was derived from the root shown in (C)]. a.u., arbitrary units.

presence of proteins that interact with retinal in the root, we treated *Arabidopsis* plants with 10 μM MCA. The majority of the mature root remained nonfluorescent with MCA treatment. However, MCA-induced fluorescence exhibited spatial and temporal patterns in developing regions of the root (movie S1 and Fig. 1B). Root meristems, regions of actively dividing cells, consistently displayed high levels of MCA fluorescence (Fig. 1B and fig. S1, A to C). In the early differentiation zone, pulses of MCA fluorescence were observed (Fig. 1B and fig. S1, A and B), which were reminiscent of the root clock. To determine whether MCA fluorescence pulses overlapped spatially or temporally with the root clock, we simultaneously tracked both processes using MCA treatment in roots expressing *pDR5:LUC*, an auxin-responsive reporter. The *pDR5:LUC* reporter line tracks the root clock and marks the prebranch site, the region of primary root tissue where lateral root organogenesis is initiated (*1*). MCA fluorescence was measured in the region of the root that oscillates (exhibits maximal luminescence of DR5 expression) and forms the prebranch site. We found that MCA signal preceded the root clock and marked a presumptive prebranch site by an average of 5 hours before the DR5 prebranch site (Fig. 1, C to E). Furthermore, like DR5 oscillations, MCA presumptive prebranch sites accurately predict the sites of lateral root primordia (fig. S2, A and B). These results indicate that *Arabidopsis* proteins capable of binding retinal are active in the specific region of the root that undergoes de novo organogenesis and that this activity precedes the peak of expression of auxin signaling during the root clock.

To further test the connection between putative retinal-binding protein dynamics and lateral root primordia development, we characterized MCA dynamics in roots in which DR5 oscillations are inhibited. D15 is a chemical inhibitor of carotenoid metabolism that reduces DR5 clock oscillations, blocks lateral root initiation, and decreases cell elongation (*3*, *4*). D15 had two effects on MCA spatial and temporal patterning. First, it decreased MCA fluorescence in the meristem (fig. S3, A and B). Second, it reduced the frequency of MCA pulses in the differentiation zone (fig. S3C). These results suggest that carotenoid metabolites are necessary for retinal binding in the root meristem and at sites of lateral root specification. D15 did not affect the amplitude of rare MCA pulses (fig. S3D), which suggests that the effect of D15 on MCA pulses is binary—pulses either occur at full amplitude or they do not occur at all. These findings are consistent with a model in which retinal biosynthesis is critical for retinal-binding protein activity. Previous work has demonstrated that apocarotenoid synthesis is necessary for initiating DR5 oscillations (*3*, *4*), further suggesting that retinal signaling precedes the root clock. These data are consistent with the hypothesis that retinal-binding proteins are involved in lateral root formation.

The correlation between MCA dynamics and lateral root development suggests a role for retinal in lateral root organogenesis (Fig. 2A and fig. S4A). To determine whether retinoids naturally occur in *Arabidopsis* and are affected by D15 treatment, we analyzed extracts from plants treated with D15 or a mock control using high-performance liquid chromatography–mass spectrometry (HPLC-MS). We found four compounds present in *Arabidopsis* that are decreased in D15-treated plants: retinal (apo16), 14'-apo-β-carotenal (apo14), 12'-apo-β-carotenal (apo12), and 10'-apo-β-carotenal (apo10) (fig. S4, A and B). To determine whether retinal, apo14, apo12, or apo10 could rescue D15 inhibition of the lateral root clock, we treated D15-inhibited

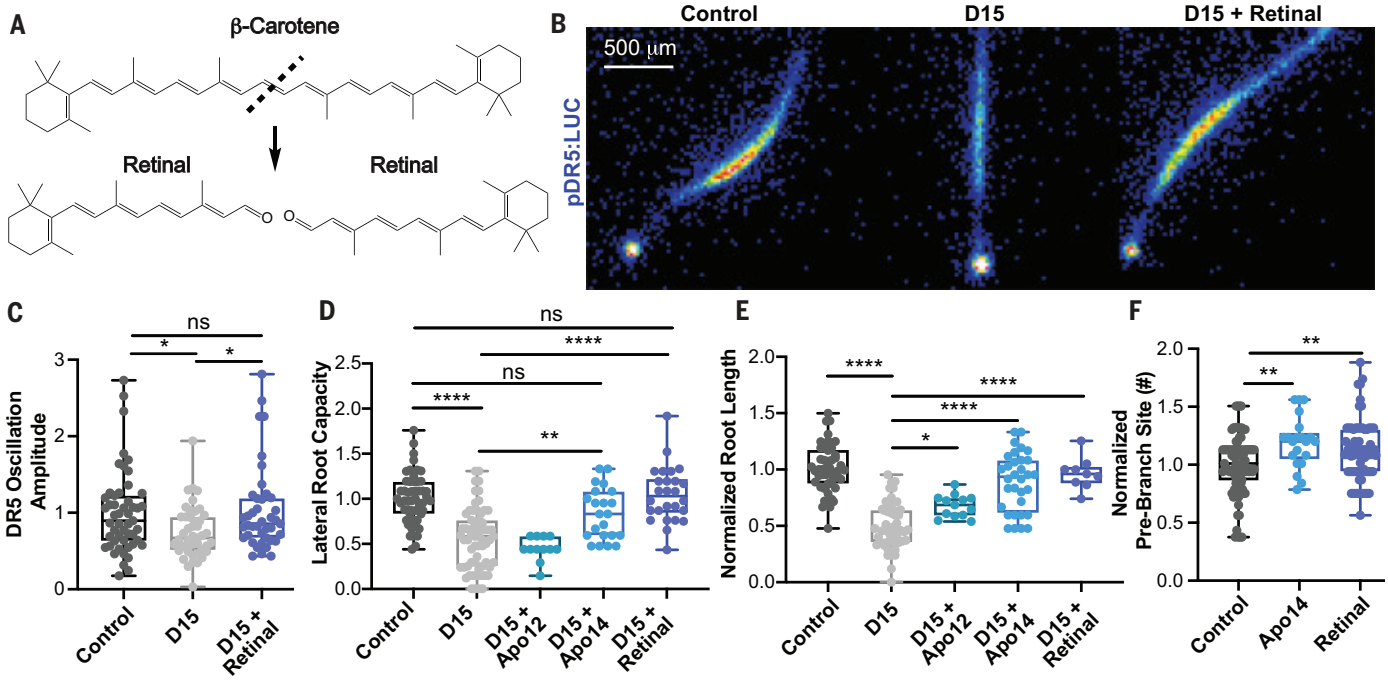


Fig. 2. Retinal rescues D15 inhibition of the root clock and lateral root organogenesis. (**A**) Chemical structures of retinal and its precursor, β-carotene. (**B**) Luminescent images of the maximum root clock oscillation in pDR5:LUC roots with the following treatments: vehicle control, D15, and D15 in combination with retinal. (**C**) Quantification of the maximum root clock oscillation intensity in pDR5:LUC roots. (**D**) Quantification of the capacity of plants to form lateral roots in D15-inhibited roots (normalized to the vehicle control). (**E**) Root length in plants exposed to D15 (normalized to the vehicle control). (**F**) The number of lateral root prebranch sites in pDR5:LUC roots treated with apo14 or retinal (normalized to the vehicle control). ns, not significant; *$P < 0.05$; **$P < 0.01$; ****$P < 0.0001$.

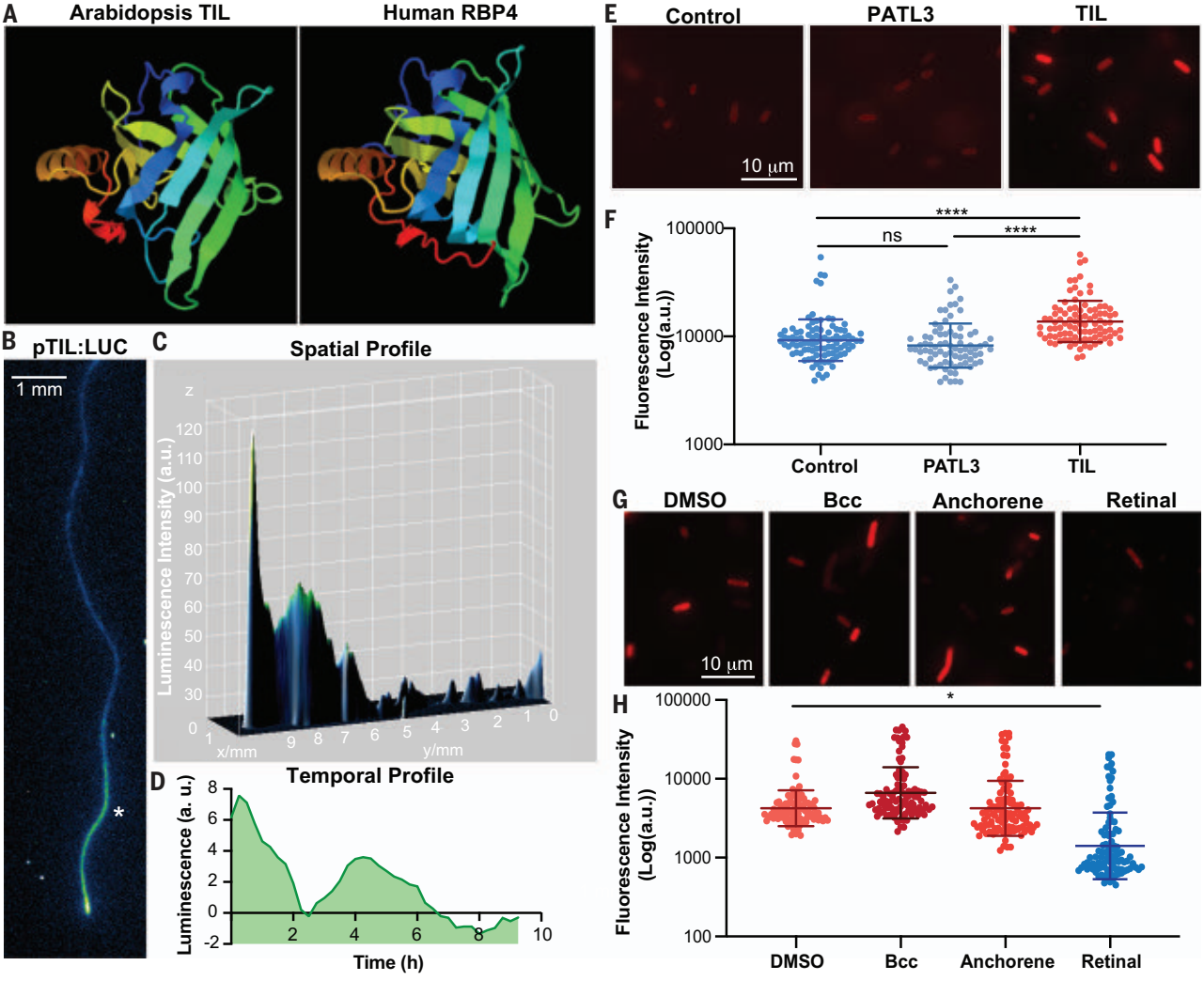


Fig. 3. TIL is a plant lipocalin that interacts with retinal. (**A**) Predicted protein structure of *Arabidopsis* TIL and human RBP4 using the protein homology–analogy recognition engine V 2.0 (PHYRE2) (*19*). (**B**) The TIL expression pattern in a representative pTIL:LUC root. The asterisk indicates the spatial region analyzed in (D). (**C**) A three-dimensional surface plot showing the spatial profile of the pTIL:LUC root shown in (B). (**D**) The luminescence intensity over time in the region of root highlighted by the asterisk in (B). (**E**) Representative images of *E. coli* cells transformed with *Arabidopsis* genes. (**F**) Fluorescence intensity of MCA-treated *E. coli* transformed with genes from *Arabidopsis*. The control cells were not transformed. (**G**) Representative images of *E. coli* cells expressing TIL that were pretreated with dimethyl sulfoxide (DMSO), β-cyclocitral (Bcc), anchorene, or retinal. (**H**) Fluorescence intensity of MCA-treated *E. coli* transformed with TIL and pretreated with DMSO, β-cyclocitral (Bcc), anchorene, or retinal. *$P < 0.05$; ****$P < 0.0001$.

roots with these compounds and quantified resulting changes in DR5 oscillations. We found that retinal, apo12, and apo14 significantly increased the amplitude of DR5 oscillations in D15-inhibited roots (fig. S4C). Just 1 μM of retinal added to D15-treated roots was sufficient to fully rescue the amplitude of root clock oscillations within 24 hours (Fig. 2, B and C, and movie S2). At this concentration, retinal also rescued cell elongation in D15-inhibited roots (fig. S4D). We next monitored the ability of retinal, apo12, and apo14 to rescue D15 inhibition of lateral root development. Only retinal and apo14 fully rescued D15 inhibition of primary root growth and lateral root capacity, which is defined as the ability of plants to form lateral roots after the tip of the primary root has been excised (Fig. 2, D and E). Retinal and apo14 significantly increased the number of lateral root primordia in untreated plants (Fig. 2F), which indicates that exogenous application of either retinal or apo14 is sufficient to induce ectopic organogenesis. Exogenous retinal treatment increased lateral root density (fig. S4E) as well as the total number of root primordia. These findings reveal that retinal and apo14 are sufficient to induce lateral root organogenesis. Enzymatic or nonenzymatic oxidation of the terminal double bond in apo14 generates retinal. Therefore, it is possible that retinal is the active signaling molecule during apo14 treatments.

To test whether retinal rescues D15 inhibition of MCA signal intensity, we examined MCA fluorescence in roots treated with D15 and retinal. Retinal rescued the MCA signal intensity in the meristem (fig. S5A), One interpretation is that retinal regulates activity of retinoid-binding proteins. An alternative hypothesis is that retinal induces the expression of a retinal-binding protein. To determine whether retinal biosynthesis occurs locally in the root, lateral root capacity was measured in roots where the entire plant was grown on D15 media except for the root tip (fig. S5B).

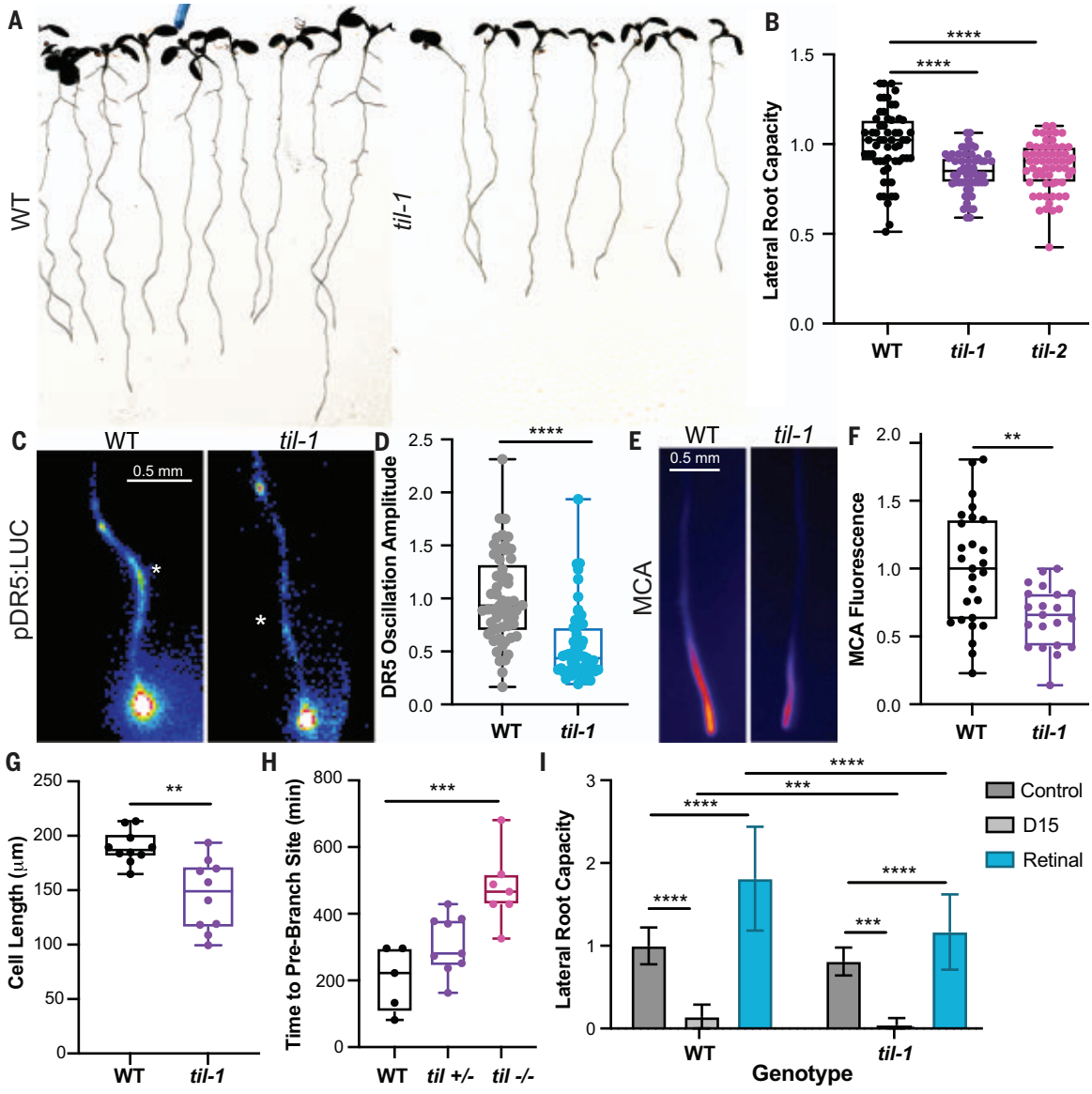


Fig. 4. TIL is essential for proper lateral root development. (**A**) Images of WT and *til-1* mutant seedlings. (**B**) Lateral root capacity (LRC) in WT and *til* mutants (normalized to WT roots). (**C**) Images showing the root clock oscillation maxima in WT and *til-1* mutant roots. (**D**) Quantification of the root clock oscillation amplitude in WT and *til-1* mutant roots. (**E**) Fluorescence images showing MCA signal intensity in WT and *til-1* mutant roots. (**F**) Quantification of the MCA fluorescence in the root tip (1-mm region) in WT and *til-1* mutant roots. (**G**) Average length of mature cortex cells in WT and *til-1* plants. (**H**) The time between the root clock maxima and the formation of a lateral root prebranch site, measured in pDR5:LUC roots. (**I**) Quantification of lateral root (LR) capacity in response to retinal and D15. Values are normalized with respect to the untreated roots for each genotype. Genotype by treatment interactions were significant for both treatments ($P \leq 0.0001$). For each sample, $n > 64$ plants. $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$.

Isolating the root tip from D15 treatment was sufficient to restore lateral root branching (fig. S5C), which indicates that retinal synthesis in immature root tissue regulates organogenesis. These data are consistent with the hypothesis that retinal biosynthesis, protein binding, and function in lateral root development are localized to the tip of the root.

To further understand the role of retinal in lateral root formation, we searched for *Arabidopsis* proteins with sequence homology to retinal-binding proteins in algae and vertebrates. We did not find significant homology between algae opsins and *Arabidopsis* proteins. However, we did identify homology between vertebrate and plant lipocalins, cross-kingdom protein transporters of small, hydrophobic molecules (*14*, *15*). TIL (AT5G58070), the primary lipocalin expressed in the *Arabidopsis* root, has sequence homology (E-value = 9×10^{-4}, percent identity = 25%) and predicted structural homology [template modeling score (TM-score) = 0.72] to RETINOL-BINDING PROTEIN 4 (RBP4), a vertebrate lipocalin (Fig. 3A and fig. S6). In plants, TIL has been implicated in preventing lipid peroxidation, particularly during light and heat stress (*16*). Analysis of an RNA sequencing (RNA-seq) expression map of the root during DR5 oscillations revealed that there is spatiotemporal overlap in TIL expression and the

root clock (*17*) (fig. S7). To investigate the TIL expression pattern dynamics more precisely, we generated *pTIL:LUC* lines and monitored expression in the root. We found that the TIL promoter confers expression that overlaps with the MCA fluorescence signal intensity (Fig. 3B). TIL expression is strongest in the meristem and oscillates in the early differentiation zone (Fig. 3, C and D, and movie S4). Together, these data suggest that TIL is a good candidate for a plant retinal-binding protein.

To test whether TIL binds retinal, we heterologously expressed TIL in *Escherichia coli* and measured the effect on MCA fluorescence. As a control, we expressed PATL3, a plant lipid-binding protein with a cellular retinaldehyde–binding protein (CRAL)–TRIO domain. PATL3 expression did not change MCA fluorescence, which indicates that the ability to bind lipids and homology to retinal-binding protein domains are not sufficient to induce MCA fluorescence (Fig. 3, E and F). By contrast, TIL significantly increased MCA fluorescence, indicating that it interacts with MCA (Fig. 3, E and F). Next, we hypothesized that if TIL binds to retinal, application of retinal should disrupt TIL-induced MCA fluorescence. Application of retinal to TIL-expressing *E. coli* significantly reduced MCA fluorescence, indicating that retinal and MCA compete for TIL binding (Fig. 3, G and H). By contrast, cells treated with anchorene or β-cyclocitral, carotenoid-derivatives with chemical similarities to retinal (fig. S8), did not diminish MCA fluorescence. These results suggest that TIL interacts with retinal but not with structurally similar apocarotenoids.

To determine the role of TIL in organogenesis, we examined root development in previously characterized TIL mutant alleles (*16*). Because both *til-1* and *til-2* mutant alleles had similar significant reductions in lateral root branching (Fig. 4, A and B), we focused further experimentation on the *til-1* allele. To determine whether TIL was important for establishing the root clock or for later stages of primordia development, we examined the expression of DR5 in *til-1*. We found that root clock amplitude is significantly inhibited in the *til-1* mutant, which suggests that TIL regulates the root clock (Fig. 4, C and D). In *til-1* mutants treated with MCA, fluorescence was significantly decreased (Fig. 4, E and F), indicating that meristematic and differentiation zone retinal binding is decreased in *til-1* roots. Furthermore, the *til-1* mutant has a defect in primary root growth because of a reduction in cell elongation, which phenocopies inhibition of carotenoid biosynthesis with D15 (Fig. 4G). Additionally, *til-1* mutants are slower to initiate lateral root primordia after the peak of the DR5 oscillation (Fig. 4H, fig. S9, and movie S3). This indicates that TIL contributes to the MCA signal in plants. These results support a model in which TIL regulates retinal signaling, root clock oscillations, and prebranch site formation.

Previous work identified TIL as being involved in heat stress tolerance (*15*). To test whether lateral root organogenesis is temperature dependent, we quantified the root clock oscillation before and after heat shock. The amplitude of the root clock significantly decreased after heat shock stress, indicating that organogenesis is a temperature-dependent process. To determine whether TIL has a role in this temperature dependency, root clock amplitude was measured in *til-1* pDR5:LUC plants. Temperature sensitivity of the root clock is suppressed in the *til* mutant [two-way analysis of variance (ANOVA) genotype by environment interaction $P = 0.008$], which suggests that TIL is important for proper response to heat stress during lateral root organogenesis (fig. S10).

To further explore the relationship between the plant lipocalins and the retinal response, we treated *til-1* roots with both retinal and D15 and quantified the effects on lateral root development. We found that *til-1* roots had significantly decreased sensitivity to retinal (two-way ANOVA genotype by treatment interaction $P < 0.0001$), which indicates that TIL is a key player in retinal-induced root organogenesis (Fig. 4I). By contrast, D15 treatment caused an increase in inhibition of lateral root capacity in *til-1* mutants (Fig. 4I). Mutants were significantly more sensitive to D15 than wild-type (WT) plants (two-way ANOVA genotype by treatment interaction $P < 0.0001$), which suggests that *til-1* is more sensitive to D15-induced retinal reduction compared with the wild type. This suggests that additional retinal perception pathways are still functional in *til-1* mutants. Overall, the changes in sensitivity in *til-1* provide genetic evidence that TIL is involved in retinal-mediated effects on lateral root development.

We identified retinal as an endogenous metabolite that is sufficient to induce post-embryonic root organogenesis. Reduction of retinal biosynthesis leads to inhibition of the root clock, the first known stage of lateral root organogenesis. Oscillations of activity of a chemical retinal-binding protein reporter predicted sites of lateral root organogenesis. We identified TIL, a plant lipocalin able to bind retinal, and discovered that it is essential for retinal perception and modulation of root clock oscillations. To date, at least three pathways have been identified that play a role in regulating the root clock. In addition to apocarotenals, these are pectin modification (*17*) and auxin (*18*). Whether these pathways act independently or in interlocking processes remains unknown. Our results show the convergent use of retinal-related compounds in the regulation of developmental clocks across the plant and animal kingdoms.

REFERENCES AND NOTES

1. M. A. Moreno-Risueno *et al*., *Science* **329**, 1306–1311 (2010).
2. W. Xuan, H. De Gernier, T. Beeckman, *Development* **147**, dev181446 (2020).
3. A. J. Dickinson *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **116**, 10563–10567 (2019).
4. J. M. Van Norman *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, E1300–E1309 (2014).
5. K. P. Jia *et al*., *Sci. Adv.* **5**, eaaw6787 (2019).
6. C. Ruyter-Spira *et al*., *Plant Physiol.* **155**, 721–734 (2011).
7. J. Agusti *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20242–20247 (2011).
8. I. De Smet *et al*., *Plant J.* **33**, 543–555 (2003).
9. M. B. Wahl, C. Deng, M. Lewandoski, O. Pourquié, *Development* **134**, 4033–4041 (2007).
10. J. Vermot, O. Pourquié, *Nature* **435**, 215–220 (2005).
11. M. Maden, *Nat. Rev. Neurosci.* **8**, 755–765 (2007).
12. S. C. Lin *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9234–9239 (2010).
13. I. Yapici *et al*., *J. Am. Chem. Soc.* **137**, 1073–1080 (2015).
14. J.-B. F. Charron *et al*., *Plant Physiol.* **139**, 2017–2028 (2005).
15. W. T. Chi, R. W. M. Fung, H. C. Liu, C. C. Hsu, Y. Y. Charng, *Plant Cell Environ.* **32**, 917–927 (2009).
16. S. Boca *et al*., *Plant Cell Environ.* **37**, 368–381 (2014).
17. G. Wachsman, J. Zhang, M. A. Moreno-Risueno, C. T. Anderson, P. N. Benfey, *Science* **370**, 819–823 (2020).
18. J. G. Dubrovsky *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8790–8794 (2008).
19. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. Sternberg, *Nat. Protoc.* **10**, 845–858 (2015).

ACKNOWLEDGMENTS

We thank D. Bergmann, M. Moreno-Risueno, L. Quadro, and I. Taylor for critical reading of the manuscript. We thank E. Sattely for use of her laboratory equipment, L. Quadro and E. Harrison for supplying apocarotenoid compounds, and M. Mijar for technical support. **Funding:** This work was supported by the Howard Hughes Medical Institute and the US National Institutes of Health under grant MIRA 1R35GM131725 (to P.N.B.) and by an Arnold and Mabel Beckman Postdoctoral Fellowship (to A.J.D.). The research of J.R.D. was supported in part by a Faculty Scholar grant from the Howard Hughes Medical Institute and the Simons Foundation. M.L. and M.S. are supported by the Intramural Research Program of the National Institutes of Health, the National Cancer Institute, and the Center for Cancer Research. **Author contributions:** A.J.D. and P.N.B. conceived of the work and drafted the manuscript. A.J.D., P.N.B., J.Z., M.L., M.S., G.W., E.S., and J.R.D. designed the experiments and interpreted the data. A.J.D., J.Z., M.L., G.W., and E.S. acquired the data. All authors revised the manuscript. **Competing interests:** The authors declare no competing interests. P.N.B. is the cofounder and chair of the Scientific Advisory Board of Hi Fidelity Genetics, Inc., a company that works on crop root growth. A.J.D. and P.N.B. are inventors on a patent application (63110088) submitted by Duke University that covers the use of retinal and related compounds in root growth. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abf7461
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*20*–*22*)
MDAR Reproducibility Checklist
Movies S1 to S4

17 November 2020; resubmitted 8 July 2021
Accepted 18 August 2021
Published online 26 August 2021
10.1126/science.abf7461

HEART DEVELOPMENT

Reversible reprogramming of cardiomyocytes to a fetal state drives heart regeneration in mice

Yanpu Chen[1], Felipe F. Lüttmann[1], Eric Schoger[2,3], Hans R. Schöler[4], Laura C. Zelarayán[2,3], Kee-Pyo Kim[4,5] Jody J. Haigh[6,7], Johnny Kim[1,8]*, Thomas Braun[1,8]*

Cardiomyocyte (CM) replacement is very slow in adult mammalian hearts, preventing regeneration of damaged myocardium. By contrast, fetal hearts display considerable regenerative potential owing to the presence of less mature CMs that still have the ability to proliferate. In this study, we demonstrate that heart-specific expression of Oct4, Sox2, Klf4, and c-Myc (OSKM) induces adult CMs to dedifferentiate, conferring regenerative capacity to adult hearts. Transient, CM-specific expression of OSKM extends the regenerative window for postnatal mouse hearts and induces a gene expression program in adult CMs that resembles that of fetal CMs. Extended expression of OSKM in CMs leads to cellular reprogramming and heart tumor formation. Short-term OSKM expression before and during myocardial infarction ameliorates myocardial damage and improves cardiac function, demonstrating that temporally controlled dedifferentiation and reprogramming enable cell cycle reentry of mammalian CMs and facilitate heart regeneration.

Adult mammalian cardiomyocytes (CMs) show little regenerative capability (*1*); therefore, it is beneficial to develop methods to stimulate adult CM proliferation after cardiac injury or disease (*2*, *3*). Induction of CM division is usually accompanied by loss of differentiated features, including sarcomere components; pronounced metabolic changes; and reexpression of embryonic genes. Although often coined "dedifferentiation," such a process might be more accurately called "partial reprogramming" because it resembles acquisition of a proliferation-competent, fetal-like state in which CMs have the ability to divide but have not lost all differentiated features (*4*). Forced expression of the reprogramming factors Oct4 (O), Sox2 (S), Klf4 (K), and c-Myc (M)—collectively known as OSKM or 4F (four factors)—induces somatic cells to dedifferentiate (*5*). Ubiquitous overexpression of OSKM in i4F mice (i4F, inducible four factors) results in cellular dedifferentiation, hyperplasia, and teratoma formation in various organs. However, teratomas do not form in the heart, thus reflecting the low proliferative rate of CMs (*6*). Because OSKM-mediated reprogramming is progressive and stochastic (*7*), we postulated that mitosis-refractory CMs may be pushed into a proliferative dedifferentiated state, when OSKM is expressed for an

[1]Department of Cardiac Development and Remodelling, Max Planck Institute for Heart and Lung Research, Bad Nauheim, Germany. [2]Institute of Pharmacology and Toxicology, University Medical Center Goettingen (UMG), Göttingen, Germany. [3]German Center for Cardiovascular Research (DZHK), Partner Site Göttingen, Göttingen, Germany. [4]Department of Cell and Developmental Biology, Max Planck Institute for Molecular Biomedicine, Münster, Germany. [5]Department of Medical Life Sciences, College of Medicine, The Catholic University of Korea, Seoul, Republic of Korea. [6]CancerCare Manitoba Research Institute, Department of Pharmacology and Therapeutics, Rady Faculty of Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada. [7]VIB, Flanders Institute for Biotechnology, Ghent University, Ghent, Belgium. [8]German Center for Cardiovascular Research (DZHK), Partner site Rhein/Main, Bad Nauheim, Germany.
***Corresponding author. Email: johnny.kim@mpi-bn.mpg.de (J.K.); thomas.braun@mpi-bn.mpg.de (T.B.)**

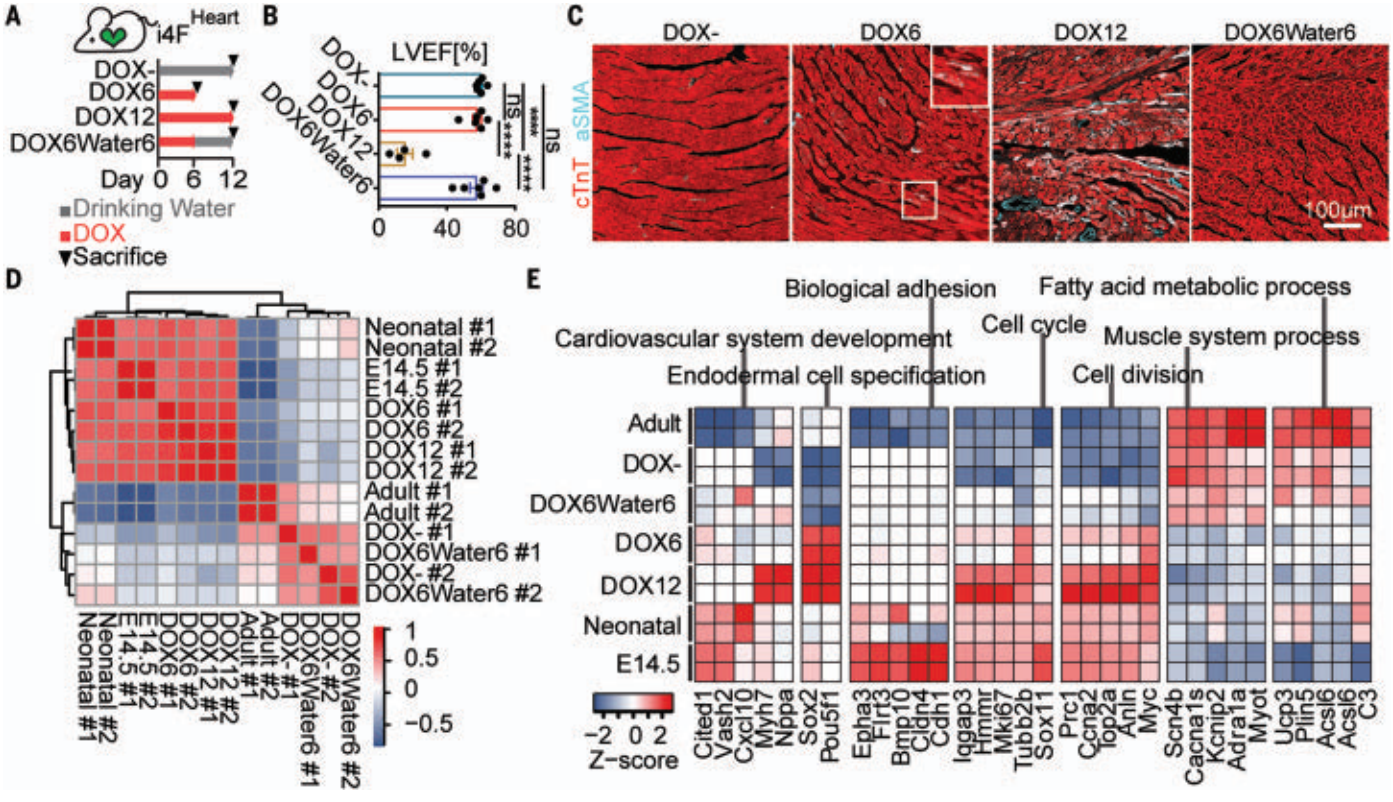


Fig. 1. Transient OSKM expression induces dedifferentiation of adult CMs. (**A**) Overview of Dox treatment of an i4F^{Heart} mouse. (**B**) LVEFs of Dox-treated i4F^{Heart} mice. Error bars indicate SEM. ns, not significant ($P > 0.05$); ****$P < 0.0001$. (**C**) Immunofluorescent staining of the cardiac dedifferentiation marker α-SMA (aSMA, cyan), and the CM marker cardiac troponin T (cTnT, red) in Dox-treated i4F^{Heart} mice. (**D**) Pearson correlation of transcriptomes from isolated CMs of Dox-treated i4F^{Heart} mice. (**E**) Heatmaps of top differentially expressed genes according to Gene Ontology classes.

extended period of time and expression in other cells is avoided.

To test this hypothesis, we overexpressed OSKM in purified juvenile mouse CMs that were fluorescently labeled with a nuclear H2BmCherry reporter (fig. S1A). Overexpression of OSKM substantially enhanced both EdU^+ and $PH3^+$ CM nuclei, indicative of G_2 or M phase entry (fig. S1, B to D). Consistent with previous reports (*8*), overexpression of c-Myc alone enhanced EdU incorporation. However, neither c-Myc alone nor OSK enabled G_2 or M phase entry, as indicated by the absence of $PH3^+$ CM nuclei (fig. S1, B to D). On the basis of these observations, we generated $i4F^{Heart}$ mice in which CM-specific expression of OSKM is enabled by administration of doxycycline (Dox) (Fig. 1A and fig. S2A). In these animals, Oct4 protein is expressed from a polycistronic $OSKM^{IRES\text{-}GFP}$ knock-in as early as 2 days after Dox supply (fig. S2, B and C). After 6 days of continuous Dox administration, $i4F^{Heart}$ mice appeared healthy and displayed normal left ventricular ejection fractions (LVEFs) and stroke volumes (SVs) (Fig. 1B and fig. S3, A and B). However, early signs of cardiac remodeling and CM dedifferentiation became apparent on day 6, as indicated by occasional expression of smooth muscle actin (SMA) in some CMs (*9*) (Fig. 1C). Prolonged expression of OSKM for an additional 6 days resulted in pronounced expression of SMA in CMs and substantial reduction of LVEFs and SVs (Fig. 1B and fig. S3, A and B). Stopping Dox administration in $i4F^{Heart}$ mice after 6 days terminated OSKM expression after 1 day (fig. S2, B and C), followed by rapid loss of SMA expression and restoration of LVEFs and SVs (Fig. 1, B and C, and fig. S3, A and B), clearly indicating that short-term OSKM-induced CM dedifferentiation is a reversible process. By contrast, $i4F^{Heart}$ mice treated with Dox for 12 days became increasingly sick and died despite Dox withdrawal (fig. S3, C and D).

Expression analyses of CMs from different developmental stages and of purified CMs from $i4F^{Heart}$ mice (fig. S4) revealed that continuous OSKM expression for 6 and 12 days induced adult $i4F^{Heart}$ CMs to adopt a gene expression profile that is very similar to those of neonatal and embryonic day 14.5 (E14.5) CMs (Fig. 1D and fig. S5, A and B). Consistent with the functional and morphological assessments, gene expression changes were almost completely reverted upon Dox withdrawal from animals treated for 6 days (Fig. 1, D and E, and fig. S6). The most profound gene expression changes were related to metabolism, cell proliferation, and division. Gene expression of key glycolytic enzymes was up-regulated, whereas genes associated with fatty acid metabolism and beta-oxidation became down-regulated (Fig. 1E and figs. S5B and S6, E to H). Virtually all gene expression changes were unidirectional and progressive, whereas expression of OSKM remained at the same level (Fig. 1E); this suggests a stochastic process that is highly dependent on the length of Dox administration. Likewise, $i4F^{Heart}$ mice treated with the 6-day Dox regimen displayed only a moderate increase of *Myh7* and *Nppa*, whereas animals treated for 12 days were characterized by strongly increased *Myh7* and *Nppa* expression (Fig. 1E), a substantial increase of heart size and heart weight–to–body weight ratio, a pronounced decrease of CM size, and an increase of total CM numbers (fig. S7, A to F).

Notably, the $i4F^{Heart}$ mice developed epithelial-like, E-cadherin$^+$/Nanog$^+$/$OSKM^{IRES\text{-}GFP+}$ neoplasms within the myocardium 21 days after sustained Dox treatment (Fig. 2, A and B). Nanog-expressing induced pluripotent stem cell (iPSC)–like clones obtained from these neoplasms could be maintained even in the absence of Dox (Fig. 2, A and B). These observations, which were fully recapitulated by Dox administration to tamoxifen-treated $i4F^{\alpha MHC\text{-}MCM}$ mice (fig. S8, A to F), demonstrate that CM-restricted expression of OSKM dedifferentiates and reprograms adult CMs as dividing pluripotent cells in vivo. The CM origin of the neoplasms was further confirmed by introduction of a $Rosa26^{H2BmCherry}$ reporter (hereafter referred to as $i4F^{Heart/mCherry}$). Notably, these mice carry only one LoxP-flanked $OSKM^{IRES\text{-}GFP}$ allele (not two as in $i4F^{Heart}$ animals) and thereby express lower levels of OSKM upon Dox treatment (fig. S9, A to C). The comparatively low levels of OSKM expression in $i4F^{Heart/mCherry}$ hearts induced only moderate CM dedifferentiation and cardiac remodeling 12 days after initiation of Dox administration (fig. S9, C to F) and did not increase heart size (fig. S9, G and H) or CM numbers (fig. S9I). Nevertheless, $i4F^{Heart/mCherry}$ mice developed lineage-traced, Nanog$^+$/H2BmCherry$^+$ cardiac neoplasms within the myocardium by week 7 of sustained Dox treatment (Fig. 2B). Nanog$^+$ cells isolated from independent heart tumors (fig. S10A) stably expanded in culture (Fig. 2, Aiii and Biii), had intact karyotypes (fig. S10B), and formed a mouse chimera after injection into blastocysts (fig. S10, C and D). These data demonstrate that the degree of CM dedifferentiation and proliferation depends on the expression levels of OSKM as well as the duration of expression.

Rod-shaped CMs purified from $i4F^{Heart/mCherry}$ and $i4F^{Heart}$ hearts underwent cell cycle reentry and complete division in the presence of Dox in culture (Fig. 3A and movies S1 to S6). In contrast to previous reports, binucleated CMs, which represent 85% of CMs in murine hearts (*10*), underwent mitosis and cytokinesis, similar to mononucleated CMs. CMs went through only one round of division in vitro and did not form iPS cells despite continuous Dox supply for more than 3 months. However, consistent with the lentivirus-mediated overexpression of OSKM (fig. S1, A to D), continuous Dox supply resulted in substantially increased proliferation of neonatal $i4F^{Heart/mCherry}$ CMs

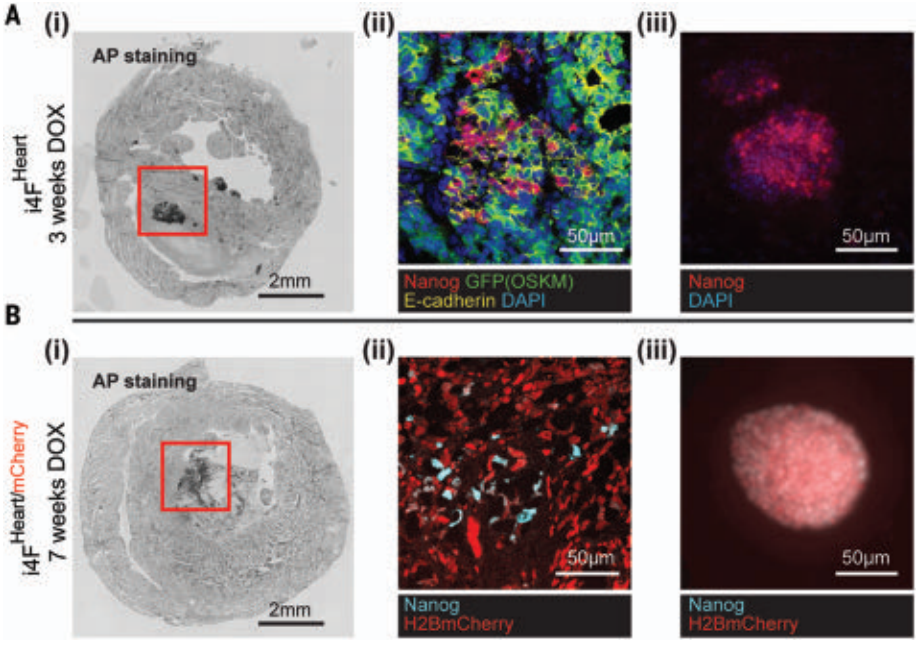


Fig. 2. Extended OSKM expression induces formation of heart tumors derived from adult CMs. (**A**) Three weeks of Dox treatment in $i4F^{Heart}$ mice. (**B**) Seven weeks of Dox treatment in $i4F^{Heart/mCherry}$ mice. (i) Alkaline phosphatase staining of CM-derived neoplasms. (ii) Immunofluorescent staining of cardiac neoplasms from sections outlined in red in (i). (iii) Immunofluorescent images of Nanog-expressing clones derived from $i4F^{Heart}$ and $i4F^{Heart/mCherry}$ neoplasms. DAPI (4′,6-diamidino-2-phenylindole) marks DNA blue. GFP, green fluorescent protein.

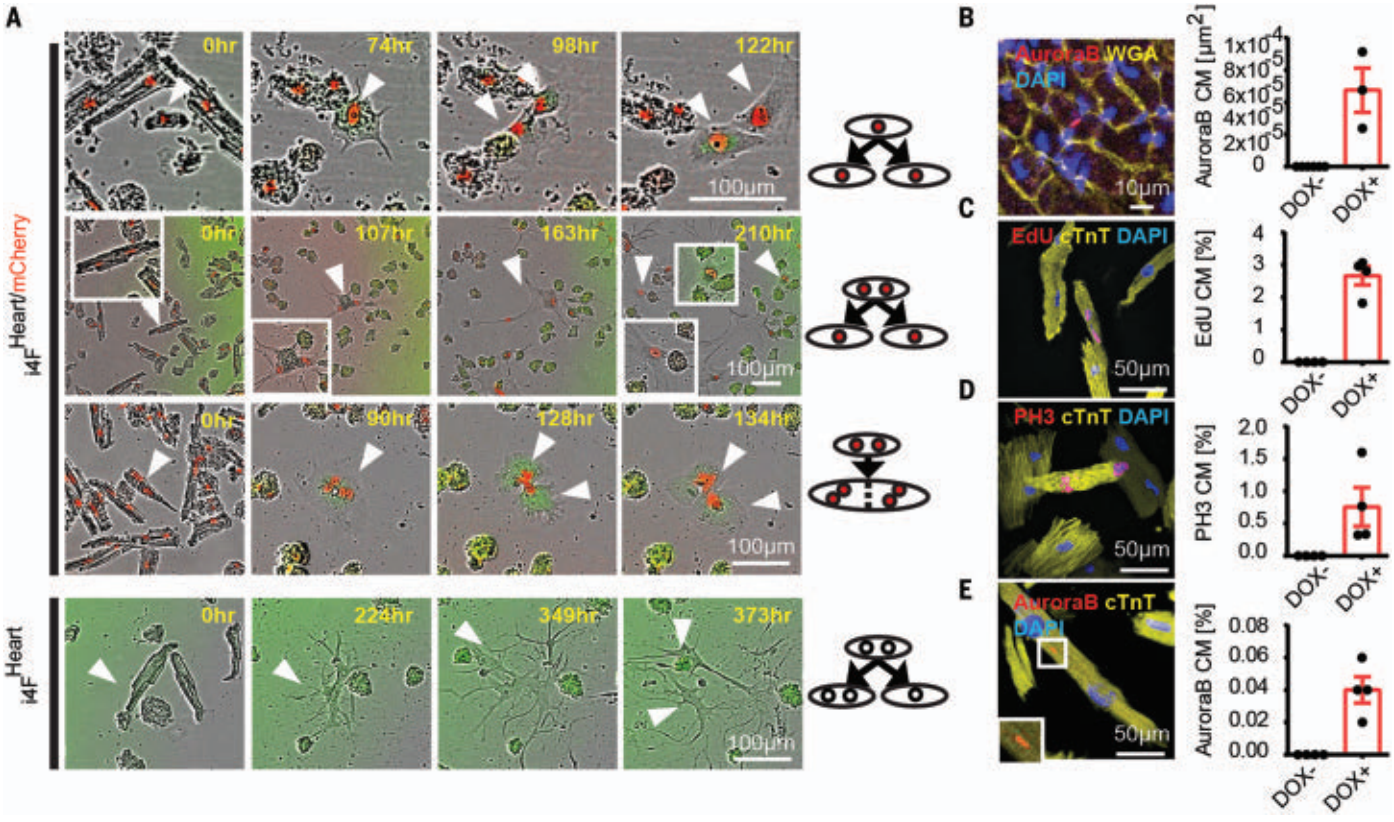


Fig. 3. OSKM expression drives division of mono- and binucleated adult CMs. (**A**) Time-lapse imaging of purified adult CMs from i4F^{Heart} and i4F$^{Heart/mCherry}$ mice exposed to Dox. Arrowheads point to dividing CMs. White boxes indicate related CMs marked with arrowheads at higher magnification. (**B** to **E**) Immunostaining of cross-sectioned hearts for the cytokinesis marker AuroraB kinase localized at the midbody (red) and wheat germ agglutinin (WGA, yellow) (B), as well as of EdU (C), the mitosis marker PH3 (D), and AuroraB kinase (E) in freshly isolated adult CMs [(C) to (E)] from i4F^{Heart} mice treated with Dox for 12 days. Error bars indicate SEM.

(fig. S11, A and B, and movies S7 and S8). Neonatal i4F$^{Heart/mCherry}$ CMs completely converted into Nanog-expressing iPS cells within 5 days (fig. S11, C and D) and formed a mouse chimera after injection into blastocysts (fig. S11E). Continuous Dox administration to neonatal [postnatal day 1 (P1)] and juvenile (P7) i4F$^{Heart/mCherry}$ mice led to substantial heart growth within 1 week (fig. S12, A and B), prominent cardiac dedifferentiation (fig. S12C), and increased CM proliferation rates relative to mock-treated controls (fig. S12, D to H). Notably, all of these effects were less pronounced in juvenile i4F$^{Heart/mCherry}$ mice compared with newborn i4F$^{Heart/mCherry}$ mice (fig. S12, A to F). These data suggest that environmental cues are present in adult hearts, facilitating survival and/or growth of dedifferentiating adult CMs, whereas less mature neonatal and/or juvenile CMs do not depend on these factors for reprogramming. Alternatively, a rare subpopulation of slowly dividing CMs that is more susceptible to full OSKM-dependent reprogramming might exist in vivo (*11*).

Although fully mature CMs are relatively resistant to OSKM treatment, as indicated by their inability to form iPSCs in culture (Fig. 3A), treatment of adult i4F^{Heart} mice with Dox for 12 days resulted in formation of AuroraB$^+$ CMs within the myocardium (Fig. 3B and fig. S13, A to C) and a significant increase in CM numbers (fig. S7F). Cell cycle reentry of adult CMs was unambiguously confirmed by analysis of purified cardiac troponin T–expressing CMs that were positive for EdU$^+$ and PH3$^+$ and harbored AuroraB at the midline between CM nuclei (Fig. 3, C to E, and fig. S13, D and E). The data show that OSKM-dependent reprogramming and dedifferentiation of CMs are determined by their developmental stage: (i) in neonatal CMs, OSKM expression sustains an immature, proliferative state; (ii) in juvenile CMs, it prevents maturation and mitotic exit; and (iii) in mitosis-refractory, fully differentiated adult CMs, it allows acquisition of a mitosis-competent state.

Because transient expression of OSKM in CMs stimulates proliferation without loss of CM identity, we wanted to know whether this also enables cardiac repair. Short-term OSKM expression led to repair of the amputated cardiac apex and minimal fibrotic scaring in juvenile i4F$^{Heart/mCherry}$ mice at P7, a developmental time point when CMs normally have ceased to proliferate and heart regenerative capacity is completely lost (*12*) (fig. S14, A to C). EdU incorporation rates were significantly enhanced in CM nuclei throughout the heart but were highest adjacent to the initial injury site (fig. S14, D and E), indicating that post-injury changes to the microenvironment promote cell cycle entry of partially reprogrammed CMs. These data show that transient OSKM expression extends the regenerative window of juvenile murine hearts.

To test whether transient reprogramming contributes to regeneration of adult hearts, we subjected i4F^{Heart} mice to myocardial infarction (MI) by ligation of the left anterior descending artery (LAD). Mice were treated with a 6-day Dox regimen 6 days before MI [pretreatment (PT)], 1 day after MI [acute treatment (AT)], or 6 days after MI [therapeutic treatment (TT)] (Fig. 4 and fig. S15). Infarcted i4F^{Heart} hearts displayed a significant decrease in scar size compared with untreated controls under all tested conditions (Fig. 4, A and B, and figs. S15 and S16, A to C). EdU-incorporating PCM1$^+$ CMs were significantly increased in the infarct border zone in pretreated and acutely treated i4F^{Heart} hearts (Fig. 4, C and D, and fig. S16, D and E). None of these effects were observed in Dox-treated wild-type mice (fig. S17, A to D). Serum cardiac troponin I levels (fig. S18, A and

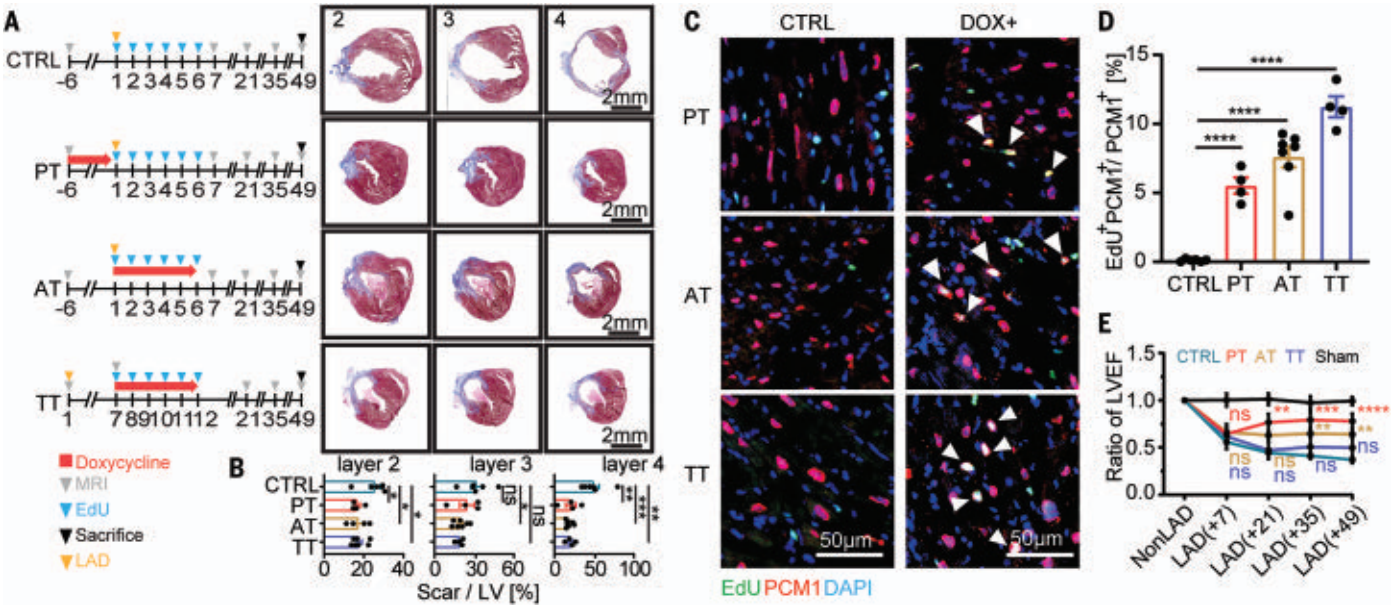


Fig. 4. Heart regeneration by partial CM-specific reprogramming. (**A**) (Left) Schematic outline of Dox treatments for LAD-ligated i4F^{Heart} mice. CTRL, control; PT, pretreatment; AT, acute treatment; TT, therapeutic treatment. (Right) Masson's trichrome staining of serial transverse sections, showing fibrotic scar tissue in blue and healthy myocardium in red. (**B**) Quantification of scar sizes in left ventricles (LVs) from (A). (**C** and **D**) Images (C) and quantification (D) of EdU incorporation (green) in PCM1$^+$ CMs (red, marked by arrows) at the border zone of LAD-ligated i4F^{Heart} hearts. (**E**) Serial fMRI measurements showing relative changes of cardiac function over time. Error bars indicate SEM. ns, not significant ($P > 0.05$); *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$.

B) or rates of MI-induced CM death (fig. S18C) were not changed between different groups 24 hours after LAD ligation, ruling out the possibility that OSKM expression enhanced CM survival and protected the myocardium from damage. Similarly, immunostaining against CD45 and CD31 in heart cross sections did not reveal differences between Dox-treated i4F^{Heart} mice and controls (fig. S18, D and E), suggesting that overexpression of OSKM in CMs does not alter inflammatory or vascular responses upon infarction. Collectively, these data indicate that transient expression of OSKM in adult CMs enables adult heart repair through proliferation of preexisting CMs.

Serial functional magnetic resonance imaging (fMRI) over a period of 2 months revealed the absence of cardiac tumor formation under all tested conditions (Fig. 4E and fig. S19). Cardiac function was improved in i4F^{Heart} mice treated with Dox before and during MI, in contrast to cardiac function in control mice. Improvement of cardiac function was most prominent in pretreated i4F^{Heart} mice. Mice treated with Dox 6 days after MI did not display significant improvement of cardiac function (Fig. 4E and fig. S19) despite a substantial reduction in scar size. These data suggest that newly formed CMs might not be fully functional or might be suboptimally arranged within the myocardium to exert full contractile force. Improving function of newly formed CMs may further narrow the gap between the condition of infarcted and sham-treated, non-infarcted hearts. In conclusion, our study shows that forced transient dedifferentiation and reprogramming of adult CMs improves left ventricle systolic function after MI, particularly when reprogramming is initiated as early as possible. Beyond heart regeneration, our study demonstrates that reprogramming and transformation of nonmitotic cells is possible but is associated with the risk of cellular transformation when reprogramming is not controlled tightly. Our study suggests therapeutic possibilities for achieving functional recovery of organs that fail to regenerate under normal conditions. Future tests will be needed to determine whether controlled CM-specific expression of OSKM by exogenous means elicits cardiac repair (*13*).

REFERENCES AND NOTES

1. O. Bergmann *et al.*, *Cell* **161**, 1566–1575 (2015).
2. H. Sadek, E. N. Olson, *Cell Stem Cell* **26**, 7–16 (2020).
3. N. U. N. Nguyen *et al.*, *Nature* **582**, 271–276 (2020).
4. Y. Zhu, V. D. Do, A. M. Richards, R. Foo, *J. Mol. Cell. Cardiol.* **152**, 80–91 (2021).
5. K. Takahashi, S. Yamanaka, *Cell* **126**, 663–676 (2006).
6. M. Abad *et al.*, *Nature* **502**, 340–345 (2013).
7. J. Hanna *et al.*, *Nature* **462**, 595–601 (2009).
8. M. J. Bywater *et al.*, *Nat. Commun.* **11**, 1827 (2020).
9. T. Kubin *et al.*, *Cell Stem Cell* **9**, 420–432 (2011).
10. M. Patterson *et al.*, *Nat. Genet.* **49**, 1346–1353 (2017).
11. M. Cui *et al.*, *Dev. Cell* **53**, 102–116.e8 (2020).
12. E. R. Porrello *et al.*, *Science* **331**, 1078–1080 (2011).
13. T. Kisby *et al.*, *Adv. Therap.* **4**, 2000141 (2021).

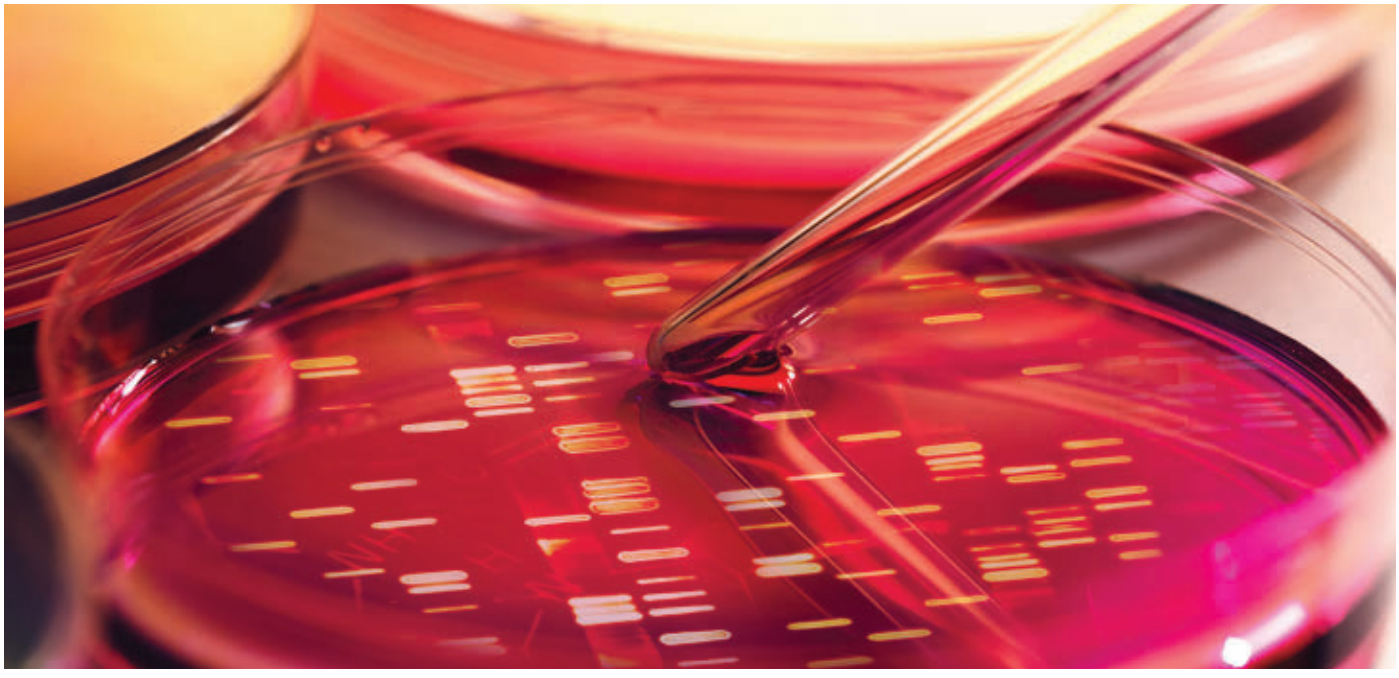
ACKNOWLEDGMENTS

We thank the core facilities in MPI-BN headed by A. Wietelmann and S. Günther. **Funding:** This work was supported by the DFG [Excellence Cluster Cardio-Pulmonary Institute (CPI), SFB TRR81 TP A02 and SFB 1213, TP A02 and B02, SFB 1002 C07/INF] and the German Center for Cardiovascular Research. **Author contributions:** Y.C. designed and performed most experiments, interpreted results, analyzed the data, and wrote the manuscript. F.F.L., E.S., and L.C.Z. performed experiments and interpreted results. K.-P.K. and H.R.S. interpreted results and performed karyotyping. J.J.H. interpreted results. T.B. interpreted results and wrote the manuscript. J.K. designed experiments, conceived the project, interpreted results, analyzed the data, and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** RNA sequencing data are deposited in Gene Expression Omnibus (GEO) under accession no. GSE160918. The ROSA26-OSKM mouse model is available from J.J.H., T.B., or J.K. under a material transfer agreement with the Vlaams Instituut voor Biotechnologie (VIB).

SUPPLEMENTARY MATERIALS

https://science.org/doi/10.1126/science.abg5159
Materials and Methods
Figs. S1 to S19
References (*14–23*)
MDAR Reproducibility Checklist
Movies S1 to S8

11 January 2021; resubmitted 18 May 2021
Accepted 6 August 2021
10.1126/science.abg5159



Multitasking at its best: Researchers can now drill down to the single-cell level and simultaneously extract gene and protein data.

Combine and conquer: Single-cell omics multiplies

Analyzing the genome, transcriptome, or proteome of a single cell is old hat now–it's time to do them all at once.

By Alan Dove

Ever since 17th century autodidact Antonie van Leeuwenhoek gazed through his handcrafted lenses to discover a menagerie of "animalcules," scientists have been straining for a closer look. With the advent of molecular biology, the study of cells has focused on the diversity of progressively smaller structures, and in recent years on progressively smaller samples as well.

Shrinking sample sizes are important for understanding a central paradox of multicellular biology: Each cell acts as an individual but must coordinate every activity with its neighbors. What biologists really want, then, is a way to study every molecule in every cell individually, then correlate the results.

It's a tall order, but recent advances in biochemistry, microfluidics, bioinformatics, and cell culture have already gone a long way toward filling it. Investigators can now sequence genomic DNA or messenger RNA, catalog proteins, or characterize chromatin structures in individual cells, and repeat the process hundreds or thousands of times in an experiment, producing extraordinarily detailed pictures of both individual and collective cell activity. As these single-cell techniques continue to develop, scientists are taking the next logical step, combining analyses to look at two or more parameters in each cell at a time.

Separate is not equal

Initially, researchers tried to synthesize single-cell data from separate cells, for example, by performing single-cell messenger RNA sequencing on one sample, and chromatin accessibility analysis on a duplicate sample. They would then link the patterns computationally, assuming that each sample had exactly equivalent cells in it. "However, there are computational challenges in trying to link different modalities across cells, even if the cells came from the same original sample," says Michael Schnall-Levin, founding scientist at **10X Genomics** in Pleasanton, California.

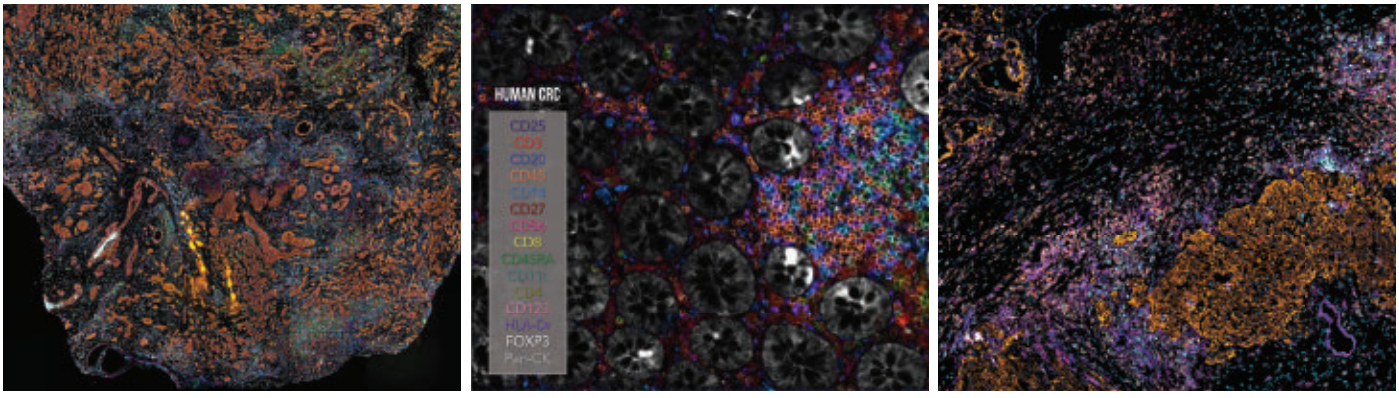
More recently, cell biologists and technology companies have developed a new generation of methods, now variously called single-cell multiomics, next-generation multiomics, or multimodal omics. Regardless of terminology, the idea is to perform multiple analyses at the single-cell level on a single sample. Instead of synthesizing correlations computationally, single-cell multiomics looks directly at each cell in multiple ways. "Some of the earliest methods enabled simultaneous detection of RNA and proteins, such as CITE-seq and REAP-seq, or of RNA and CRISPR guide RNAs, such as Perturb-seq, from the same cell," says Schnall-Levin. While the initial proof-of-concept experiments focused on immune cells, single-cell multiomics has now gained traction in cancer research as well.

Regardless of the parameters they want to track, investigators new to single-cell analysis should spend some time thinking about sample preparation. "This is especially important when measuring analytes with different stabilities … I would say that sample prep and experimental methods are a balance to make sure different analytes can be optimally detected in unison," says Schnall-Levin.

Once the laboratory portion of the experiment is done, single-cell multiomics presents unique bioinformatics challenges. Every additional analyte is another layer of data to process, multiplying the total amount of information researchers need to handle. However, there are some compensating advantages to the multiomics approach. "Having joint measurement of '-omes' eliminates the need for computational strategies to infer relationships between different single-cell datasets," explains Schnall-Levin.

As a major provider of single-cell sequencing technologies, 10X has adapted its product line to cater to the mushrooming multiomics trend. The company cont.>

IMAGE: © ISAK55\SHUTTERSTOCK.COM



A 13-plex immune profiling assay on a human non-small cell lung cancer sample (**left**); a 15-plex assay in a human colorectal cancer sample with an overlay of biomarkers (**middle**); and a 20-plex assay on a human breast cancer sample (**right**).

now sells equipment and reagents to analyze several different combinations of targets. Schnall-Levin expects those offerings to continue to expand as the field develops. In particular, he anticipates investigators combining more types of analyses into single experiments. Indeed, some immunologists are already measuring as many as five parameters at once at the single-cell level, and that number seems likely to grow.

Getting your drops in a row

As researchers add more dimensions of analysis to their single-cell experiments, they're encountering some tricky problems. "The typical path people are taking into multiomics is adding on other easy molecules on top of RNA, like protein or epigenetic markers … which are easy to get at in a whole-genome and whole-transcriptome type of level," says Anup Parikh, vice president of software and informatics at **Mission Bio** in South San Francisco, California.

While correlating the transcriptome and expressed proteome in a single cell provides a useful cross-check, researchers increasingly want to include measurements of entirely orthogonal characteristics, such as the cell's chromosome copy number or genome sequence. Because DNA sequencing entails disrupting the cell, though, it's not compatible with simultaneous transcriptome or proteome analyses that require intact cells.

To address that, Mission Bio offers a microfluidic system that can send single cells through a series of analyses, all while tracking each cell on its journey. The system separates individual cells from a suspension, isolating each cell into its own droplet with a molecular barcode. These barcoded droplets then pass through microfluidic channels on the Mission Bio device, like packages on a conveyor belt, visiting separate reaction chambers that carry out each analysis.

"We never lose anything from the process, but we can separate out a reaction where we do some protease treatment to break open the cell, independently from the reaction where we put polymerases to amplify the DNA or [analyze] protein or anything else out of that cell," says Parikh. By placing nondestructive steps ahead of destructive ones, the system can obtain proteomic, transcriptomic, genomic, and epigenomic data from each cell.

With so many options, researchers must be careful not to bite off more than they can chew. "When you approach these types of projects, it gets complex really quickly, and so you want to ensure that you've got a really well-grounded scientific question, because you can … generate immense datasets," says Aaron Llanso, Mission Bio's director of translational applications and market development.

That kind of focus also informs the company's approach to data processing. "We've taken a very methodical approach to building our informatics solutions specific to the applications that help our scientists and our pharmaceutical companies; so it's less of a shotgun approach where we have general tools, and more 'can I help answer a question about clonal evolution of cancer in a specific cancer type?'" says Parikh.

Over the rainbows

Given the power of microfluidics to provide fine-scale control over cells and reagents, it's no surprise that other companies also offer microfluidic solutions for specific types of single-cell analysis. For **Canopy Biosciences** in St. Louis, Missouri, that meant updating the venerable technique of immunofluorescence microscopy. Traditionally, researchers have used antibodies labeled with fluorescent dyes of different colors to visualize the expression of a few different cellular proteins at a time, but that approach doesn't scale beyond a handful of epitopes. Cells express many more proteins than there are fluorescent dye colors.

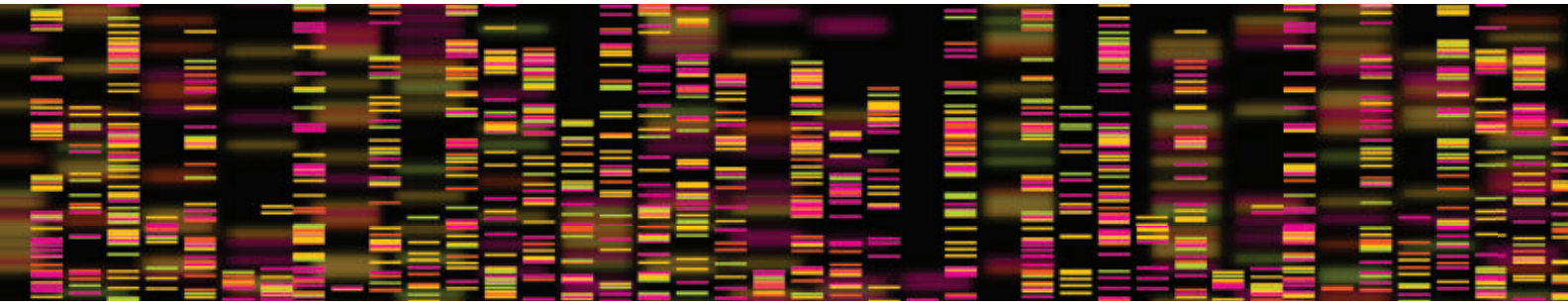
Canopy's microfluidic system automates the fluorescent labeling and imaging processes, then repeats them. "We do this staining in an iterative fashion to build really high-plex assays, so you stain with a cocktail of up to five antibodies [on a microfluidic chip], collect image data on those five antibodies, and then erase away the fluorescent signal, and you can stain again," says Thomas Campbell, product manager at Canopy. The system then stacks the image data to map the expression of all the proteins across each cell in the sample. Campbell adds that while the capacity of the chip doesn't have any fixed limit, so far it's generally been used for analyzing sets of 20 to 50 proteins per experiment.

The chips also preserve the cells. "You can store your sample in the fridge for up to 2 years, and then take it back out and pick up right where you left off, and build on top of that assay, to kind of reinterrogate the sample at a later date," says Campbell. That could be an especially useful feature for investigators working with precious clinical samples. It also means that as the technology continues to develop, users will be able to apply new techniques to previous specimens.

In principle, the Canopy system could find use in any field that's traditionally relied on immunofluorescence. "One of the nice things about the platform is it's quite flexible with regard to sample type; we do tissue sections, both fresh, frozen, and [formalin-fixed paraffin-embedded], and we can also do a wide array of different cell suspensions," Campbell says. He adds that most current customers are working in immunology or oncology, but the company has also supported COVID-19 immunoprofiling studies and neuroscience applications.

cont.>

IMAGE: COURTESY CANOPY BIOSCIENCES



Featured participants

10X Genomics
10xgenomics.com

BD Biosciences
bd.com

Canopy Biosciences
canopybiosciences.com

Genewiz
genewiz.com

Mission Bio
missionbio.com

The sort squeeze

Microfluidics isn't the only way to approach single-cell multiomics. Indeed, for immunologists, isolating and analyzing single cells is hardly a new idea; they've been using flow cytometers to separate and study fluorescently tagged cells in numerous ways for years. One major manufacturer of flow cytometry equipment, **BD Biosciences** in Franklin Lakes, New Jersey, now also sells a complete set of add-on components tailored to single-cell multiomics.

Indeed, BD now sells a range of equipment, consumables, and reagents covering almost the entire pipeline of a single-cell multiomics experiment, from blood collection tubes and cell dissociation buffers to flow cytometers.

"Once the [flow cytometry enrichment] happens, then we go into the BD Rhapsody system, which is a very unique system for single-cell partitioning," says Ranga Partha, senior director of marketing at BD. The Rhapsody microwells are designed to preserve fragile cells, such as granulocytes, neutrophils, and stem cells, and the system also includes a scanner for visual quality control. Accessory equipment and reagents for protein and RNA profiling, either targeted or tracking the entire transcriptome, round out the system. The company is also developing a variable, diversity, and joining [V(D)J] sequencing system that will characterize antibody gene rearrangements in single cells.

Though the current Rhapsody assays clearly cater to immunologists, that field's overlap with cancer research has not escaped the company's notice. "I believe the future of the technology is going to be [focusing on] biomarker discovery in the field of oncology," says Partha.

Even for scientists with extensive flow cytometry experience, though, single-cell multiomics presents unique challenges. Partha cites sample preparation, cell viability checks, and bioinformatics as the three biggest hurdles for newcomers to the field. "Having skilled and trained researchers [on staff] to do all three of them is challenging," he says.

To address that issue, he urges researchers to talk to colleagues who've done similar experiments and contact the companies that sell the reagents and equipment they intend to use. "I think having a conversation with a technical application specialist from the company that provides the assay would be a good step … but I think it's sometimes missed by the researcher, and they continue the experiment, and then toward the end the data complexity becomes very high," says Partha. Like other experts in the field, he points to bioinformatics as an evolving challenge that's likely to become even more important as researchers add more assays to their single-cell experiments.

Get professional help

Despite the difficulties, the popularity of single-cell multiomics is growing almost as fast as the complexity of the experiments. "It's definitely a hot, emerging area," says Haythem Latif, director of clinical development services at **Genewiz** in South Plainfield, New Jersey. Not everyone who's interested in the field is equipped to dive into it, though. Fortunately, research companies such as Genewiz are happy to help investigators who would rather hire out some or all of the experimental work.

When picking a service company, scientists should expect to spend some time discussing their research, the specific questions they want to address, and what kinds of technologies the company can bring to bear on the problem. "We're working quite collaboratively … making sure those types of technical parameters are matched with the biological question that's at hand," says Latif. These early conversations can head off trouble later. "If there are some feasibility issues in what you're trying to do biologically, those can get potentially called out or worked around in a different way," Latif says.

Service providers should always customize their approach to each project, but in many cases they'll converge on well-trodden paths. Recently, for example, one popular strategy has been to apply 10X Genomics' transcriptomic profiling technology with surface protein tagging for corresponding proteomic information, all at the single-cell level. "That's a nice clean workflow. It's been developed with a lot of customers … to kind of get those two pieces of information together," says Latif. He adds that researchers are also starting to add spatial analyses in order to localize proteins on cell surfaces.

While oncologists and immunologists have been the biggest adopters of single-cell multiomics so far, Genewiz representatives point out that other fields are also looking into the technology. "[In] stem cell differentiation and pluripotency there's a lot of interest … to help characterize the stem cells and their differentiation process, and that's also being done in neuroscience," says Dave Corney, the company's principal scientist.

Regardless of the specific strategies they're pursuing, experts agree that single-cell multiomics is just starting to take off, with immense potential for future discoveries. "The human body has 40 trillion cells that are changing all the time … the ability to have tools for researchers to make these discoveries will result in us knowing more about what causes cancer or Alzheimer's or heart disease–and that's what we want to do," says Schnall-Levin.

Alan Dove is a science writer and editor based in Massachusetts.

IMAGE: © ZITA\SHUTTERSTOCK.COM



Cell-Sorting Platform

Sphere Fluidics' Cyto-Mine is an automated platform that integrates single-cell screening, sorting, dispensing, imaging, and clone verification to streamline workflows, reduce costs, improve throughput, and enable high-value clones to be captured in a single run. Thanks to its simplified internal design, it is now faster and easier to manufacture, commission, and service. The new electronics design improves system performance, reliability, and stability, while the updated software features provide a better user experience, reduce user–instrument interaction, and allow for more robust data analysis. The instrument's enhanced capability and performance enable users to access more functionality, and the improved software offers a more advanced, reliable system to further simplify antibody discovery, cell-line development, cell therapy, and synthetic biology workflows.

Sphere Fluidics
For info: +44-(0)-1223-628890
www.spherefluidics.com

Fully Characterized FFPE Tissue Blocks

AMS Biotechnology has established an unmatched commercial biorepository and procurement network that includes a large selection of fully characterized and annotated formalin-fixed paraffin-embedded (FFPE) tissue blocks to help streamline your research. We offer an extensive collection of frozen and FFPE human and animal tissue samples from normal and many diseased tissue types. All FFPE tissue blocks are processed using standard clinical procedures, under Institutional Review Board approved protocols. Each tissue sample is pathologically verified and quality assessed by a board-licensed pathologist before being shipped with donor demographics and clinical details. Characterized sample sets are now available and include many clinically relevant mutations and biomarkers, such as EGFR, PD-L1, KRAS, NRAS, MSI, MMR, and BRAF. In addition, comprehensively NGS-characterized FFPE solid tumor tissues suitable for assay development, genotyping, and identification of new mutational hotspots are available. FFPE tissue blocks can be supplied ready to use or can be tailored to suit your workflow and provided as slides, scrolls, or array formats.

AMS Biotechnology
For info: +44-(0)-1235-828200
www.amsbio.com

RNA Extraction Kits

Porvair Sciences has launched kits for rapid extraction of total RNA from a wide range of biological samples. These proprietary kits enable extraction of up to 1,000 µg of application-ready pure RNA in as little as 10 min. Extraction and purification of RNA is required for numerous downstream molecular biology applications, such as real-time PCR, quantitative PCR, Western blotting, and RNA amplification. The Chromatrap RNA Extraction kit from Porvair Sciences follows the simple procedure of sample lysis, RNA capture, washing, and finally elution of purified RNA. Porvair's extraction kits employ a uniquely designed spin column and proprietary buffer chemistry as well as a reliable protocol to isolate superior-quality RNA from a variety of sample sources. Using the Chromatrap kit, scientists can now process up to 10^7 cultured animal cells, 50 mg of animal tissue, 0.5 mL of blood, 250 mg of plant material, 5 x 10^7 yeast cells, or 5 x 10^8 bacteria in a single run.

Porvair Sciences
For info: +44-(0)-1978-666222
www.chromatrap.com/chromatrap-rna-extraction-kit

Multiplex IHC Kits

Akoya Biosciences' Opal Multiplex IHC kits make multiplex results accessible to anyone who works with standard immunohistochemistry (IHC). With Opal, you can select antibodies for simultaneous IHC detection based on performance rather than species. Opal kits are optimized for reliable spectral unmixing and simultaneous measurement of three to eight IHC targets, plus a nuclear stain. Opal enables you to (1) use the best primary antibodies, regardless of species, with no cross-reactivity; (2) identify multiple cell phenotypes while retaining spatial and morphological context that is lost with bulk measurements and flow cytometry; and (3) get more information from precious and scarce samples.

Akoya Biosciences
For info: +1-571-524-4181
www.akoyabio.com/phenoptics/opal-kits-reagents

Viral Nucleic Acid Extraction Kit

BioChain has successfully provided enough reagents to test over 3 million people across the United States for COVID-19 infection, due to the recent launch of their vamPure Viral Nucleic Acid Extraction kit, which allows for fast and efficient total nucleic acid extraction from human swab samples. The final extract is ideal for subsequent detection of various viral species, including the novel coronavirus responsible for COVID-19. BioChain's new kit has been hailed as a high-quality, cost-effective, sensitive, and reliable tool by clinical labs that need to process tens of thousands of test samples per day. The vamPure kit is easy to use and ideal for application on automation platforms, which greatly facilitates acquisition of test results. The kit can be purchased in small quantities or bulk packaging, depending on the customer's needs.

BioChain Institute
For info: +1-888-762-2568
www.biochain.com

Electronically submit your new product description or product literature information! Go to www.science.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

RUTGERS
School of Arts and Sciences

DEPARTMENT OF MOLECULAR BIOLOGY AND BIOCHEMISTRY

The Department of Molecular Biology and Biochemistry at Rutgers, The State University of New Jersey, New Brunswick (Busch Campus), invites applicants for a tenure-track **Assistant Professor** position. The position requires a PhD degree in molecular biology or a related field by the appointment date (**9/1/2022**). The successful candidate will be responsible for teaching undergraduate and graduate students, and leading a creative research program in the general areas of Biochemistry or Molecular Biology. Applicants who study metabolism in health and disease using biochemical and metabolomics approaches are especially encouraged to apply. Outstanding researchers in other areas will also be considered. The research program of the new faculty member will bolster the Department's existing strengths in the study of cancer, aging and gene regulation.

The MBB Department is part of an expanding program within the Division of Life Sciences on the Busch Campus, where the Center for Advanced Biotechnology and Medicine, the Waksman Institute, and the Robert Wood Johnson Medical School are also located. Faculty members also work closely with the Rutgers Cancer Institute of New Jersey and Rutgers Child Health Institute. We have strong consolidated, interdepartmental graduate programs in Molecular Biosciences. The campus is located within one hour of New York and Philadelphia in a highly desirable area.

Rutgers University has one of the most diverse student bodies among institutions of higher learning in the country, and boasts several programs established to nurture equity and inclusion that enhances the achievement of excellence by faculty members and students. We encourage applications from women and members of communities who are underrepresented in the sciences, and will evaluate the potential of the applicant to mentor and empower minoritized students.

The position is highly competitive with regard to start-up funds, laboratory space and salary. Interested individuals are encouraged to apply by submitting the following materials:

- Curriculum vitae
- 2-3 page statement of research plans
- Contact information for three scientists who can provide letters of reference
- 1-page statement discussing your past and/or potential contributions to diversity in scientific research and training Interested individuals must apply online at **https://jobs.rutgers.edu/postings/141827**

Applications should be submitted as soon as possible but not later than **October 31, 2021**; modestly late applications will likely be considered.

Rutgers University is an equal opportunity/affirmative action employer.

Cleveland Clinic
Lerner Research Institute

The Department of Cancer Biology at the Lerner Research Institute, home to laboratory, translational, and clinical research at the Cleveland Clinic, is seeking to recruit **faculty at all academic levels** with a particular interest in breast, lung, genitourinary (prostate, kidney, bladder), and pancreatic cancers. The overarching goal of the Department of Cancer Biology is to define the underlying molecular mechanisms and clinical behavior of cancer to identify new therapeutic strategies. The position presents an exceptional opportunity to partner with outstanding clinical oncology programs to translate basic discoveries to the clinic. In addition, Department Faculty are members of the NCI-designated Case Comprehensive Cancer Center and the Cleveland Clinic Lerner College of Medicine at Case Western Reserve University.

Links to the Department and Lerner Research Institute are here:
Cancer Biology: **https://www.lerner.ccf.org/cancerbio/**
LRI: **http://intranet.lerner.ccf.org/lri/**.

Applicants must have a Ph.D., M.D., or both, and an exceptional track record of accomplishment. The successful applicant will be supported by generous start-up funds.

Candidates should submit a CV, short narrative of research interests and plans (2 page maximum), and three names for references, via e-mail to Alex Almasan, Ph.D., c/o Christina Stuart, **stuartc2@ccf.org**.

Cleveland Clinic is a not-for-profit, multi-specialty academic medical center that integrates clinical and hospital care with research and education. It is rated by U.S. News and World Report as the #2 hospital in the country and #5 in Cancer.

Cleveland Clinic is an Equal Opportunity/Affirmative Action Employer: Women/Minorities/Veterans/Individuals with Disabilities. Smoke/Drug-free Environment.

THE UNIVERSITY OF NEW MEXICO

FRESHWATER ECOSYSTEM ECOLOGY ASSISTANT PROFESSOR DEPARTMENT OF BIOLOGY, UNIVERSITY OF NEW MEXICO

We invite applications for a full-time, tenure-track 9-month appointment at the **Assistant Professor** level in the area of **Freshwater Ecosystem Ecology**. We are interested in applicants with demonstrated commitments to **(1)** pursuing a vigorous, externally funded research program, with an emphasis in arid and/or semi-arid ecosystems; **(2)** engaging in interdisciplinary research and activities associated with the Grand Challenge in Sustainable Water (**https://grandchallenges.unm.edu/three-grand-challenges/sustainable-water-resources/**); **(3)** achieving excellence in teaching and mentoring at undergraduate and graduate levels; and **(4)** diversity, equity, inclusion, student access, and community outreach; **(5)** and postdoctoral experience is preferred by the start of the appointment. UNM is both a research-intensive university (Carnegie R1) and a minority-majority institution. It is located in Albuquerque, a city of diverse cultures in a biodiverse, scenic landscape. The Department of Biology is family-friendly, with 32 tenure-track faculty members (**https://biology.unm.edu/core-faculty.shtml**) and ~1300 undergraduate majors, a thriving graduate training program, and vibrant portfolio of extramurally funded research. Strong campus-wide connections to interdisciplinary programs are related to water, including the Bosque Ecosystem Monitoring Program, the Sevilleta Long Term Ecological Research program, the Water Resources Program, Museum of Southwestern Biology, Center for Stable Isotopes, Center for Water and the Environment, and the Analytical Geochemistry Laboratory in Earth and Planetary Sciences.

A Ph.D. in the life sciences is required at the time of application. For complete job requirements, see Assistant Professor, Freshwater Ecosystem Ecology at **https://unmjobs.unm.edu, Req15596**.

Applicants should submit a cover letter including a paragraph addressing potential connections to the UNM Grand Challenge in Sustainable Water, curriculum vitae, three representative reprints, names of three references, and three two-page statements addressing past, present and future efforts in **(1)** research, **(2)** teaching, and **(3)** diversity and inclusion. For best consideration, all application materials must be submitted through UNMjobs, **https://unmjobs.unm.edu** by **October 15, 2021**. Questions related to this posting may be directed to Dr. Marcy Litvak, **mlitvak@unm.edu**.

The University of New Mexico is an Equal Opportunity/Affirmative Action Employer and Educator. Women and underrepresented minorities are strongly encouraged to apply.

VANDERBILT UNIVERSITY

The **Department of Biochemistry** at Vanderbilt University School of Medicine Basic Sciences invites applications for a tenure-track faculty appointment at the rank of **Assistant Professor** as part of the **Destination Vanderbilt** faculty hiring initiative. We seek candidates who are using innovative approaches to address important and challenging problems in all aspects of biology at the molecular and mechanistic level.

The candidate will be expected to develop an independent research program that compliments current departmental and institution strengths, which include genome maintenance, cancer biology, structural biology, mass spectrometry, enzymology, signal transduction, toxicology, drug discovery, host-microbe interactions, and neurobiology.

The successful candidate will join a vibrant and diverse research community. The Biochemistry department is consistently positioned in the top tier for NIH funding among all US medical schools. Our exceptional faculty have received considerable recognition, including receipt of major University awards, election to the National Academy of Sciences, and a Nobel Prize in Physiology or Medicine (Stanley Cohen, 1986). We are proud of our history and strive towards continued excellence in research, training, and service in the field of biochemistry.

Incoming faculty will have a supportive peer group and extensive mentoring. Biochemistry is part of the Basic Sciences within the School of Medicine at Vanderbilt University. This culture is enhanced by the diverse activities of several trans-institutional centers and institutes (https://tinyurl.com/yyqdq3t3; https://tinyurl.com/y2cwu27g). The appointee will have access to state-of-the-art core facilities (https://tinyurl.com/y3s24nds; https://tinyurl.com/y96ky46j) and will be eligible for internal funding mechanisms that foster team-science approaches. Trainees in the appointee's lab also may be eligible for stipend support from over 25 NIH-supported T32 training programs.

Applicants should apply to Vanderbilt through Interfolio at: **apply.interfolio.com/94568**. Evaluation of applicants will begin **November 1, 2021**.

Please submit: a curriculum vitae; two-page research plan; one-page statement of contributions to equity, diversity and inclusion; one-page mentoring/teaching statement; reprints of up to three recent publications; and arrange for three recommendation letters to be sent via Interfolio. Additional information is available at **https://medschool.vanderbilt.edu/biochemistry**.

Vanderbilt Biochemistry is committed to promoting diversity, equity, and inclusion in our scientific community (https://medschool.vanderbilt.edu/biochemistry/dei/). Vanderbilt University is an Equal Opportunity/Affirmative Action employer committed to recruiting and retaining an academically and culturally diverse community of faculty. Vanderbilt has a strong institutional commitment to diversity in all areas and particularly encourages applications from minorities, women, individuals with disabilities, and members of other underrepresented groups.

ScienceCareers online @sciencecareers.org

By **Laura Rossi**

What my coach gave me

First I burst into uncontrollable tears. Then I stood and began to frantically shake my arms. Thank goodness I was working from home; if my colleagues had seen me, who knows what they would have thought. I wasn't sure what I thought myself. It was my first session with my personal coach, whom I had hired to try to address the challenges I was experiencing as a new tenure-track professor. I went into the online session wanting—and expecting—immediate, concrete advice. Instead, we were talking about a toxic former mentor and practicing physical exercises. Could this possibly be the answer to my problems?

I had heard about the idea of a personal coach a few months earlier from a more senior academic who was interviewing for a position at my institution. We connected over shared experiences, so I felt comfortable telling her I was a bit overwhelmed balancing life with a newborn and the demands of a tenure-track position. For one thing, I wanted—and thought I needed—to be perpetually available to answer emails, which kept me attached to my inbox as though my life depended on it. I also felt I needed to say yes to every request that came my way, and soon found myself not having time for the work I really needed to do to build my lab and advance my career.



"Working with her makes me feel … capable and strong."

This colleague told me that earlier in her career, her department chair had suggested she hire a coach. It really helped her, she said, and she gave me the name of hers. I was skeptical. I had already taken leadership and time management courses that did not deliver on their promises, and I was not keen on more advice from somebody not currently working in academia. Besides, everyone around me appeared to be doing well without a coach. A trusted senior colleague even commented that hiring a coach might look bad to the department head.

Nonetheless, I knew I wouldn't last much longer if I didn't get some help. I had money from my startup package for leadership training that I was going to lose otherwise. So, I decided to give coaching a try.

Despite my misgivings about that first session, I stuck with it. After another session or two, I began to see that my coach was helping me identify the underlying issues that were driving my behavior—which gave me the knowledge and confidence to change. She helped me understand that my compulsions to check email and agree to requests were subconscious attempts to compensate for the lack of caring I felt from that toxic mentor. This knowledge helped me succeed at something I had attempted many times before but failed at miserably: creating strict no-email times. This time, I reminded myself of the ways I already care for my colleagues—being kind, helping when I can, responding to emails and requests thoughtfully at times that work for me. That helped me set boundaries and reclaim focused time for myself and my research.

After that breakthrough, my coach and I moved on to other challenges. She helped me learn to deal with rejection more objectively, for example by highlighting negative feedback in one color and positive in another, which helps me quiet my inner critic by reminding myself that I also receive praise. She taught me a scheduling tool that keeps me productive and on track, both at work and in my personal life. As important as the concrete advice is how working with her makes me feel: capable and strong.

After the money from my startup package ran out, I wanted to continue with my coach. So I did something I never would have before: I asked the department head for funding. She said yes, and noted that it was a good investment not only for me, but also for the department: If I'm happy and successful, the department benefits, too.

I'm not saying that now my life is perfect and I have no stress at all. I don't know whether I will get tenure. What I do know is that if I had not reached out for help, I might be writing a resignation letter instead of this essay. ■

Laura Rossi is an assistant professor at the Delft University of Technology. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

The UNIVERSITY of OKLAHOMA Health Sciences Center

ASSISTANT/ASSOCIATE PROFESSOR
FACULTY POSITION IN CELL BIOLOGY

The Department of Cell Biology at the University of Oklahoma Health Sciences Center (OUHSC) invites exceptional scientists to apply for a tenure-track eligible, **assistant** or **associate professor** position. Candidates with research interests addressing fundamental questions relevant to cilia/ciliopathies, centrosomes/centrosomopathies, microtubules and pathologies related to microtubule defects are encouraged to apply. The successful candidate is expected to develop/maintain an independent and dynamic research program. Their program will develop within the context of a growing cohort of basic scientists focused on mechanistic studies of primary cilia. The Oklahoma Medical Research Foundation (OMRF) is adjacent to OUHSC and maintains a collaborative partnership with OUHSC. Scientist at OMRF are also a source of collaborators for studies in this field. All OMRF core facilities are available to OUHSC investigators.

The Department of Cell Biology is one of five basic science departments within the College of Medicine at OUHSC, providing a highly interactive cohort of scientists with interests in cilia, cell signaling, cancer, vision, neuroscience and cardiovascular biology. For further information about the department go to **https://basicsciences.ouhsc.edu/cb/**.

OUHSC is a comprehensive health center that includes Colleges of Pharmacy, Dentistry, Nursing, Allied Health, Public Health, Medicine, and the Graduate College. It is home to two hospitals and an NCI-designated Cancer Center. OUHSC Core Facilities include the state-of-the-art Laboratory for Molecular Biology and Cytometry Research, which provides services specializing in genomics, proteomics, bioinformatics, cytometry, and imaging.

OUHSC is located in downtown Oklahoma City. The greater Oklahoma City area has a rich cultural base with modern, vibrant business and entertainment districts. The cost of living is enviably low relative to other university communities across the country.

Successful applicants will receive a competitive salary, generous, multi-year start-up package, and be assigned newly renovated laboratory space.

Job requirements include evidence of outstanding research productivity and a PhD, MD or equivalent. Candidates should submit a curriculum vitae, a description of current research interests and future research plans as one pdf file. Letters of recommendation will be requested at a later stage. Please address the application to Dr. Marie Hanigan, Chair of the Cell Biology Search Committee and send the application packet to Paula Slaughter at **paula-slaughter@ouhsc.edu**.

The University has a commitment to diversity, equality and inclusion. All qualified applicants will receive consideration. The University of Oklahoma Health Sciences Center is an equal employment opportunity and affirmative action institution http://www.ou.edu/eoo. Individuals with disabilities and protected veterans are encouraged to apply.

HEALTH
UNIVERSITY OF UTAH

In anticipation of substantial growth over the next five years, the Department of Human Genetics at the University of Utah School of Medicine (**www.genetics.utah.edu**) seeks outstanding applicants for one or more tenure-track positions at the level of **Assistant** or **Associate Professor**. We seek highly creative scientists who use genetics to investigate fundamental biological problems. We encourage applicants whose research focuses on human and medical genetics; evolutionary and functional genetics and genomics; computational genomics; genome technology development; and research programs using established or novel model organisms.

As part of a vibrant community of faculty with a strong track record of collaborative mentorship, research, and funding, the Department of Human Genetics lies at the interface between basic and clinical sciences. This creates ample opportunities for interdisciplinary research (e.g., our Center for Genetic Discovery, Transformative Excellence Program in Evolutionary Genetics and Genomics, and Center for Genomic Medicine). As a department, we value diversity and equity, and believe that the best science is done when researchers of diverse backgrounds are integrated and supported in an inclusive manner. We seek faculty who share these values. Our institution is set in a unique geographical landscape that attracts a heterogeneous and productive scientific community. Successful candidates will receive a generous startup package and enjoy a stimulating research environment that places a strong emphasis on innovation and interaction.

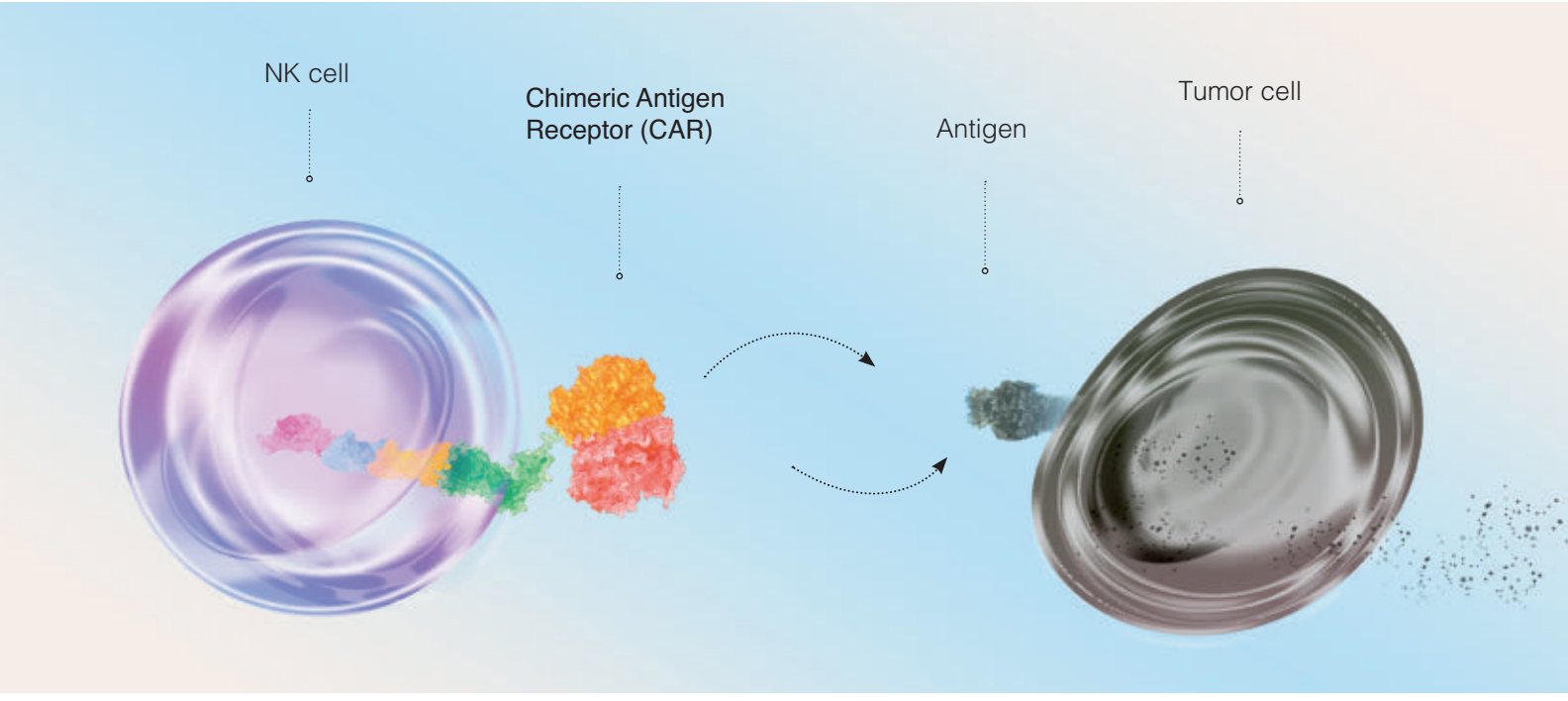
For further information, please contact **Kandacel@genetics.utah.edu**.
https://utah.peopleadmin.com/postings/121711

The University of Utah Health (U of U Health) is a patient-focused center distinguished by collaboration, excellence, leadership, and respect. The U of U Health values candidates who are committed to fostering and furthering the culture of compassion, collaboration, innovation, accountability, diversity, integrity, quality, and trust that is integral to our mission.

Application Deadline: **November 1, 2021**

Custom CAR-NK Development

Engineering NK Cells for Cell Therapy R&D



Custom CAR-NK Development

CAR-T cells are powerful tools for cell therapy. Engineering other immune cells such as NK cells can offer novel therapeutic options that address some of the disadvantages of T cell therapy.

A few advantages of CAR-NK cell therapy include:

- NK cells are part of the body's innate immune system and are the body's normal cancer-fighting reserve
- NK cells have different target recognition mechanisms as an innate lymphoid cell
- CAR-NK cell-based approaches can be generated from allogeneic donors
- The absence of TCR strongly reduces the risk of GvHD

In conclusion, CAR-NK cell therapy could offer the advantage of an intrinsic tumor-killing capacity possessed by NK cells in addition to their CAR-dependent killing mechanism, impeding tumor immune escape mechanisms.

ProMab offers completely personalized CAR-NK cell development service, offering many different options to accelerate cell therapy research. Learn more about how ProMab can introduce CAR-NK development into your work.

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