Using optical fiber networks to eavesdrop on the planet p. 1312

Deep learning enhances density functional theory pp. 1322 & 1385

Nuclear membrane pores dilate under strain p. 1341

Science

$15
10 DECEMBER 2021
science.org

AAAS

LEAF EVOLUTION

Similar genes but different expressions form grass and eudicot leaves p. 1377

Gordon Research Conferences

New Conferences in 2022

We are excited to be back and looking forward to welcoming you in 2022!

- Additive Manufacturing of Soft Materials
- Biology of Winter
- Calcium Signaling
- Cell-Cell Fusion
- Computational Materials Science and Engineering
- Electrochemical Interfaces in Energy Conversion and Storage
- Fluids in Disease Transmission and Contamination
- Frontal Cortex
- Genetic Biocontrol
- Graft Preservation in Heart Transplantation
- Hypothalamus
- Immunoengineering
- Immunometabolism in Health and Disease
- In Vivo Ultrasound Imaging
- Lipidomics
- Multiscale Mechanochemistry and Mechanobiology
- Neural Mechanisms of Acoustic Communication
- Radionuclide Theranostics for the Management of Cancer
- Single-Cell Cancer Biology
- Single-Cell Genomics
- Systems Aging
- Unconventional Semiconductors and Their Applications

Apply today to one of the 225 GRCs and their associated Gordon Research Seminars in 2022! Find your conference community at **www.grc.org.**

Esophageal cancer prevention through early diagnosis and treatment: Lessons from Hebei Province, China

In the 1970s, esophageal cancer was the leading cause of cancer-related death in Hebei Province, China, with one in five people dying from this disease in some high-risk counties. In the late-stage disease, patients cannot eat or drink, and often die within 6 months without proper treatment. The 5-year survival rate is under 10%, with deaths causing considerable economic and emotional hardship.

Oncologists at the Fourth Hospital of Hebei Medical University tackled this problem using a three-pronged approach. First, they set up a provincial cancer registry that tracked the cancer epidemiology and burden, laying a foundation for comprehensive cancer prevention and control. Second, they focused on determining the etiology of esophageal cancers in Hebei. Researchers determined that contributing factors included contamination of the local water supply with toxic nitrogen–containing compounds, mold contamination of food resulting from long-term storage, trace-element deficiencies in the soil, and poor nutrition among the local residents. The genetics of the local population was also a factor, as the cancer appeared repeatedly within certain families. And third, an etiology prevention strategy was put in place that included improvements to the quality of the drinking water, better administration of storage for agricultural produce, and an educational campaign with recommendations for a healthy lifestyle.

But these measures proved insufficient. Oncologists at the Fourth Hospital therefore set out to determine if early detection and treatment of esophageal cancer was possible.

A simple, inexpensive test was developed based on the detection of glycogen. Iodine, when sprayed on esophageal mucosal cells, changes their color if the cells are normal. No color change indicates potentially cancerous cells. Researchers at the hospital also participated in a multicenter clinical research program to develop additional screening technologies for esophageal cancer. They achieved good success in early detection, according to Baoen Shan, director of the Fourth Hospital and the Provincial Institute of Oncology. Furthermore, the hospital was the first in China to develop a technique for endoscopic resection of esophageal cancer as a treatment modality. Over 1 million people in 18 provinces have now been trained in the procedure. Over 230,000 esophageal cancer screenings have been performed in Hebei in the past 15 years and 4,000 early-stage cases have been detected, with a 5-year survival rate of 92.7%. 11,429 cases of precancerous esophageal cancer lesions have been identified.

Studies have been performed to find treatments that might block the formation of precancerous lesions to prevent or treat early-stage esophageal cancer, including the extraction and purification of compounds from traditional Chinese medicines, such as *p*-hydroxylcinnamaldehyde isolated from the *Cochinchina momordica* seed and lupeal acetate from *Cortex Periplocae* root. Traditional Chinese medicine prescriptions were also tested in an attempt to prevent or impede precancerous lesions in the esophagus, mitigating the need for surgery, radiotherapy, or chemotherapy.

Thanks to 45 years of comprehensive prevention and control, the esophageal cancer mortality rate in Hebei decreased 82.4% from the 1970s to 2018, a markedly faster decline than the average across China of 56.6% during the same period. The 5-year survival rate increased from ~8.0% in the 1990s to 34.4% in 2013, higher than both the Chinese average (30.3%) and that found in the United States (21.1%). This reduction in mortality and morbidity saves patients in Hebei an estimated RMB 532 million (USD 83.4 million) in direct medical expenses and RMB 2.1 billion (USD 329.3 million) in indirect expenses annually. The work being done in Hebei has considerably reduced patient medical expenses, saved national medical resources, and significantly advanced the prevention and control of upper gastrointestinal cancers in China.

"The money aside, what we are doing is life-saving," Shan says, "and lives are priceless."

Sponsored by

IMAGES PROVIDED BY THE BEIJING HEALTH PROMOTION ASSOCIATION

Leveraging life science innovation

By encouraging increased innovation and entrepreneurial aspirations in academic laboratories, the BioInnovation Institute (BII) spawns dozens of startups and nurtures their development. The new BII & Science *Prize for Innovation will inspire even more advances.*

In Copenhagen, Denmark, nearly 70 startup companies began at BII. This nonprofit institute funded by the Novo Nordisk Foundation "supports life-science innovation," says BII CEO Jens Nielsen. "We are particularly interested in taking excellent science, typically in academia but in some cases at hospitals and other research institutions, and helping to translate it into a business idea with a life of its own, specifically through market investment to build a company." In short, BII builds connections. Nielsen describes that work as "being a bridge between academia and early-stage companies in order to help them attract financing."

This year, BII played a partnering role in a new connection, the BII & *Science* Prize for Innovation. "We started this prize with *Science* to increase the emphasis on the need for innovation in the academic setting," Nielsen explains.

With more than 20 years of experience in academia, Nielsen developed an interest in translating university research into catalysts for new companies. "We need to empower that science and introduce it into society in order to address numerous large societal challenges." He adds, "Our goal is to emphasize this, and I cannot think of a better partner than *Science*, which is a symbol of excellence."

Lessons from life science

BII focuses specifically on the life sciences, in which Nielsen says, "We have seen a lot of innovation." New medicines are a critical example, with those innovations depending in large part on research initiated in universities. "Today, many of the big pharmaceutical companies are sourcing new drug leads through the acquisition of biotech startups that, in many cases, can trace their roots back to academic research," Nielsen says. "This has resulted in the establishment of a very nice 'food chain.' For example, oncology treatment has been revolutionized by expanding on ideas that started in universities and hospitals and promoting their growth and application in the market."

Today's achievements in the life sciences represent only the beginning of what lies ahead. "Life-science innovation allows skilled scientists to not only express their passion within their field of work academically but also bring their ideas to market for the benefit of people and societies around the globe," says Robert Langer, Institute Professor at the Massachusetts Institute of Technology and holder of more than 1,400 patents being used in over 400 companies.

As the chair of BII's scientific advisory board, Langer knows the innovation environment in Denmark. "The innovation potential there looks promising, based on a long academic tradition and many new patent filings every year. I know from experience in the U.S. that raising the entrepreneurial capacity in society takes a strong collaboration between dedicated organizations in an ecosystem."

The collaboration between BII and *Science* promises even more innovation. "The interest and passion for life-science innovation are currently gaining momentum among scientists, and the new BII *& Science* Prize for Innovation supports this progress," Langer says.

The description of this prize states: "We seek scientists who can show that they have reached across field boundaries."

IMAGE PROVIDED BY BIOINNOVATION INSTITUTE

That leads to multidisciplinary thinking, which is fundamental to Langer's own success. In 2011, he and Nobel laureate Phillip Sharp described going even further, pointing out the "increasing need to merge expertise that goes beyond the interdisciplinary intersection of fields to the emergence of new disciplines" (*1*). Thinking back on that concept, Langer says, "I think that convergence is becoming more and more useful for innovation and for startup companies."

Research to reduce pregnancy loss

Multidisciplinary research also plays a critical role in reducing one of life's cruelest moments: the loss of a pregnancy. According to the Mayo Clinic (*2*), approximately 10% to 20% of all known pregnancies end in a miscarriage. But the actual number is probably higher, as many miscarriages occur early in pregnancy before the mother is aware she is pregnant.

Henriette Svarre Nielsen, professor of medicine at the University of Copenhagen and head of the Recurrent Pregnancy Loss Unit at Amager and Hvidovre Hospital, has spent 20 years trying to find the causes of so many pregnancy losses. "From long-term research, I now understand that pregnancy loss is not just a binary event in the life of people who want a child," she says. "Yet, when the loss happens, parents don't receive an explanation capable of easing their grief and sorrow and enabling relevant action to avoid recurrence." She wants to change that.

Working with BII, Svarre Nielsen is leading a project on the early diagnosis of pregnancy loss. The study's description notes that it focuses on developing "a diagnostic test that can inform couples after a pregnancy loss to guide relevant actions [with the potential of increasing] live birth rates" (*3*).

Svarre Nielsen and her colleagues started enrolling patients 1 year ago by selecting couples who experienced a pregnancy loss to obtain various biosamples for analysis. It is well described that approximately 60% of pregnancy losses can be explained by a chromosomal abnormality. "The intriguing perspective of this study is to explore what triggers the rejection of euploid pregnancies—those in which the fetus has the normal number of chromosomes," Svarre Nielsen explains.

To determine the reason, scientists perform genome sequencing to "look for a genetic cause of death and characterize losses where the fetus is apparently healthy," Svarre Nielsen says. She aims to do two things: first, to provide a diagnosis that gives the parents some closure, and second, to identify a way to avoid a future pregnancy loss, perhaps through treatment.

Svarre Nielsen adds that women who have lost pregnancies are at a higher risk of other health problems later in life, including diabetes, heart disease, and mental-health disorders. She states, "If the precursors of these diseases are treated, we may be able to reduce pregnancy loss. A reduction of just 5% would significantly increase the number of wanted children, given the high frequency of pregnancy losses." Beyond saving pregnancies, Svarre Nielsen asks: "Can we actually use pregnancy loss and a few tests to identify women at risk of disease 10 or 15 years later?"

> "I know from experience in the U.S. that raising the entrepreneurial capacity in society takes a strong collaboration between dedicated organizations in an ecosystem."
>
> — Robert Langer

Over the horizon

Many more areas of life science will benefit from innovative scientists. Nielsen mentions "the green transformation" and the "need to establish a more sustainable society." For example, innovations in life science could potentially eliminate carbon dioxide produced from agriculture.

"Science is often driven by curiosity, which emphasizes the need to broaden one's research focus in order to drive innovation and develop new solutions," Langer says. "We need to celebrate those who dare to think differently and follow their passion."

References

1. P. A. Sharp, R. Langer, *Science* **333**, 527, doi: 10.1126/science.1205008.
2. Mayo Clinic, "Miscarriage"; available at https://www.mayoclinic.org/diseases-conditions/pregnancy-loss-miscarriage/symptoms-causes/syc-20354298.
3. https://bii.dk/faculty/professor-henriette-svarre-nielsen-2.

Sponsored by

BII BioInnovation Institute

CONTENTS

10 DECEMBER 2021
VOLUME 374 • ISSUE 6573

1312

Researchers unspool fiber optic cables used to eavesdrop on the rumblings beneath Mount Meager, a volcano in British Columbia, Canada.

NEWS

IN BRIEF

1302 News at a glance

IN DEPTH

1304 How bad is Omicron? Some clues are emerging
New variant appears to evade immunity and shows signs of spreading more rapidly
By K. Kupferschmidt and G. Vogel

1305 Archaeologist reinstated as Max Planck director
After allegations of bullying, Nicole Boivin returns to supervisory role *By A. Curry*

1306 Trial of Harvard chemist poses test for U.S.'s China Initiative
Prosecution of Charles Lieber "misguided," critics say *By J. Mervis*

1307 Gene editing produces single-sex litters in mice
CRISPR approach may curb culling of lab animals, chicks *By E. Pennisi*

1308 How pandemic pressure is re-engineering clinical trials
COVID-19 spurred nimbler trials that demand less of patients—and many changes may stick *By J. Couzin-Frankel*

1310 X-ray telescope to study how magnetic objects sculpt light
NASA mission could reveal engines driving black hole jets *By D. Clery*

1311 Key cancer results failed to be reproduced
Project to replicate high-impact preclinical cancer studies delivers sobering verdict
By J. Kaiser

FEATURES

1312 World wide web
Researchers are weaving optical fibers into a low-cost network of seismic sensors that can probe Earth's hidden dynamics
By P. Voosen
PODCAST

INSIGHTS

PERSPECTIVES

1318 A lipid primes the final cut in dividing cells
A newly described pathway activates separation of lens cells at the end of cytokinesis *By J. A. Brill and A. Wilde*
RESEARCH ARTICLE p. 1339

1319 Interactomes in the era of deep learning
Deep learning provides an atomic snapshot of the yeast protein interactome *By J. Pereira and T. Schwede*
RESEARCH ARTICLE p. 1340

1321 Keep it simple and switch to pure tellurium
Tellurium switch operates memories through crystalline-liquid-crystalline phase changes
By R. Calarco and F. Arciprete
REPORT p. 1390

1322 Artificial intelligence "sees" split electrons
Machine-learning creates a density functional that accounts for fractional charge and spin *By J. P. Perdew*
REPORT p. 1385

1323 Martian water escape and internal waves
Lower atmospheric processes are vital for assessing water loss from Mars
By E. Yiğit

1325 Oncolytic virotherapy as immunotherapy
Recognizing immune responses to oncolytic virotherapy opens the way for new combinations *By A. Melcher* et al.

PHOTO: JACQUELYN SMALE

POLICY FORUM

1327 Filling gaps in trustworthy development of AI
Incident sharing, auditing, and other concrete mechanisms could help verify the trustworthiness of actors
By S. Avin et al.

BOOKS *ET AL.*

1330 Interpreting Jerusalem
Science, politics, identity, and conflict converge in investigations of the city of peace
By T. E. Levy

1331 Humanity versus the world
A biologist confronts the evolutionary arms race at play in human-remade landscapes
By M. E. Hannibal

LETTERS

1332 Wastewater: China's next water source
By Y. Zhang et al.

1332 Noise pollution disrupts freshwater cetaceans
By Z.-T. Wang et al.

1333 STEM disparities we must measure
By J. B. Freeman

1334 Technical Comment abstracts

1334 Errata

RESEARCH

IN BRIEF

1335 From *Science* and other journals

REVIEW

1338 Immunology
Early-life imprinting of unconventional T cells and tissue homeostasis
M. G. Constantinides and Y. Belkaid
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABF0095

RESEARCH ARTICLES

1339 Cell biology
PI(3,4)P2-mediated cytokinetic abscission prevents early senescence and cataract formation *F. Gulluni* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABK0410
PERSPECTIVE p. 1318

1340 Structure prediction
Computed structures of core eukaryotic protein complexes *I. R. Humphreys* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABM4805
PERSPECTIVE p. 1319

1341 Cell biology
Nuclear pores dilate and constrict in cellulo *C. E. Zimmerli* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABD9776

1342 Neurodevelopment
Mouse and human share conserved transcriptional programs for interneuron development *Y. Shi* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABJ6641

1343 Coronavirus
Protection against SARS-CoV-2 Beta variant in mRNA-1273 vaccine–boosted nonhuman primates *K. S. Corbett* et al.

1353 Coronavirus
Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant *J. Zhang* et al.

1360 Catalysis
Sabatier principle of metal-support interaction for design of ultrastable metal nanocatalysts *S. Hu and W.-X. Li*

1366 Hydrophobicity
Charge transfer across C–H···O hydrogen bonds stabilizes oil droplets in water
S. Pullanchery et al.
PODCAST

1370 Forest ecology
Multidimensional tropical forest recovery
L. Poorter et al.

REPORTS

1377 Plant science
Evolution of the grass leaf by primordium extension and petiole-lamina remodeling
A. E. Richardson et al.

1381 Superconductivity
Discovery of segmented Fermi surface induced by Cooper pair momentum
Z. Zhu et al.

1385 Quantum chemistry
Pushing the frontiers of density functionals by solving the fractional electron problem
J. Kirkpatrick et al.
PERSPECTIVE p. 1322

1390 Phase-change memory
Elemental electrical switch enabling phase segregation–free operation *J. Shen* et al.
PERSPECTIVE p. 1321

1395 Optical imaging
Time-of-flight 3D imaging through multimode optical fibers *D. Stellinga* et al.

1399 Graphene
Detection of graphene's divergent orbital diamagnetism at the Dirac point
J. Vallejo Bustamante et al.

DEPARTMENTS

1300 Editorial
Academic freedom under fire
By Jonathan R. Cole

1301 Editorial
Universities are not political prizes
By H. Holden Thorp

1406 Working Life
Sciencing while brown
By Christopher J. Hernandez

ON THE COVER

A stem of the grass *Andropogon gerardii* (also known as "big bluestem") emerging from a leaf sheath. Grasses, like other monocots, have tall, cylindrical leaves that visually differ from the wide, flat leaves of eudicot plants. By means of computational modeling and developmental genetic analysis, researchers have uncovered similarities in the leaf development mechanisms of grasses and eudicot plants. See page 1377. *Photo: Bruce Leander/Lady Bird Johnson Wildflower Center*

New Products.....1404
Science Careers.....1405

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.

ILLUSTRATION: ROBERT NEUBECKER

EDITORIAL

Academic freedom under fire

Jonathan R. Cole is the John Mitchell Mason University Professor at Columbia University, New York, NY, USA, where he was also the Provost and Dean of Faculties from 1989 to 2003. jrc5@columbia.edu

Academic freedom is under fire across the world—from Hong Kong, where the Chinese government's crackdown on open discussion has spurred an exodus of scholars from universities, to the United States, where there is a rising tide of anti-intellectualism and assaults on free inquiry. The effort to undermine this cornerstone of American democracy has been driven by former President Trump and his administration, members of Congress, state governors, and legislators. Alas, it also comes from faculty and students on the ideological right and left, and even from some presidents of the country's research universities. These are powerful forces that must be overcome.

The defense of academic freedom is never easy because the university is not a cloister and is no more a safe space from criticism or conflict than the society at large. Thus, there will be unpopular opinions expressed in scholarly research and in the classroom. But it is often unpopular opinions that shift paradigms, fell orthodoxies, and advance knowledge and inquiry to new levels of achievement. As the Kalven Committee stated in its 1967 report on the university's role in political and social action in the United States, "a good university, like Socrates, will be upsetting." De facto speech codes, the idea of "privileged knowledge," and weaponizing misunderstood concepts such as "critical race theory" all limit research on political grounds and threaten academic inquiry from within its own environment. To cancel speeches, to sanction students or faculty for their views, to protect oneself from the upsetting ideas and language that are encountered in great texts and novels may "purify" universities but only by diminishing them. American academician Robert Hutchins, perhaps the greatest defender of academic freedom, said that the problem with witch hunts was "not how many faculty would be fired for their beliefs, but how many who think they might be…"

Academic freedom can slip away slowly, and if ignored, could leave society in a climate of censorship, mistrust, and even fear. Now is the time for bold and courageous leadership. Presidents and provosts of universities in the United States need to act collectively by, for example, issuing a public statement that defends academic freedom. Boards of trustees or regents should sign on as supporting this statement of principle. Moreover, the political and overly commercial influence on governing boards must end. Members of these boards too often consist of political appointments or elected representatives with a political point of view. Thus, the tendency is for special interests to cloud their notions of how the university should be governed—from developing the institution's mission and strategic goals to establishing policies. Rather than reinforcing university values—foundational academic, intellectual, and artistic freedoms—university governance is vulnerable to economic and political forces. People across this country, no matter what side of the political divide, want excellent education for their families. Universities must make the case that excellence is threatened by political interference and either restructure their boards or more closely scrutinize member influence.

> **"Universities must make the case that excellence is threatened by political interference…"**

In 1915, and again in 1940, the American Association of University Professors released a statement on the principles of academic freedom. It is time for these principles to be examined and fortified. US courts have defended academic freedom as falling in the penumbra of the Constitution's First Amendment, which enforces the freedom of speech, among other rights. An update to these principles that ensures the free search for truth and its free exposition in the name of the common good would fortify academic freedom and potentially be codified into federal law. This effort could involve a presidential blue ribbon panel of experts that is open to new thinking about the principles. It should include discussion of free expression by students as well as faculty and protect remarks made outside their formal capacity as faculty members.

The university has a central role in the growth of knowledge, the exposure of students to diverse thoughts and differing views, and the preparation of a next generation for civic life. If that role is being undermined from outside or even within the university, its leaders should rise to its defense, even if it means jeopardizing their own jobs, or removing those leaders who are a threat. It is a public principle worth defending at a high personal cost. When called for, such courageous action will educate the public and legislators on the importance of academic freedom.

–Jonathan R. Cole

10.1126/science.abn5447

PHOTO: CHESTER HIGGINS. © THE NEW YORK TIMES

Universities are not political prizes

Institutions of higher education are valued by democracies because of their civic goal—to foster in the next generation of citizens a regard for individual freedoms and rights, principled debates, and tolerance for opposing opinions. The leaders of these institutions are expected to demonstrate their commitment to these values by supporting academic freedom—the ability of faculty, staff, and students to challenge wisdom, explore new ideas, and advance knowledge through free inquiry. But lately, some university administrators have been responding to the ever more polarized political climate by giving lip service to academic freedom while playing politics—either ignoring or playing both sides of conflicts that threaten to undermine the very tradition of free and diverse thinking and discourse.

Given the polarized political environment in the United States, tensions over the content of university classes, research, and speakers have only escalated. These problems exist because of, or are made worse by, university administrators who pretend that all is well while they are actually being attacked under the radar. Take the recent incident involving the University of Florida (UF), which blocked liberal professors from testifying as experts in a lawsuit that challenged the state's voting laws. The university rationalized that as a public institution, it must remain neutral when the real reason was that it didn't want to anger political overlords. Last month, UF reversed its decision after a public outcry by its stakeholders, particularly when it was discovered that other professors had been quietly muzzled. It would have made more sense for UF to acknowledge the potential pressure to comply with the state's conservative administration up-front while making clear that its professors are free to "speak truth to power" when called upon, sticking to its commitment to academic freedom. It is a mistake for scientists to ascribe tensions over academic freedom to the world of the humanities and qualitative social sciences, because limiting academic freedom in one discipline has implications for all of academia. Administrators at UF recently insisted that a new degree not include the word "critical," a word that is just as important to humanists as "evolution" is to scientists.

> **"...some university administrators have been...giving lip service to academic freedom while playing politics..."**

As a former university chancellor and provost, I know that university administrators must constantly try to please two different audiences—a generally liberal faculty and generally conservative alumni. When they speak to alumni, they tell heart-tugging stories about students who were transformed by higher education, celebrate athletic success, and laud life-saving advances developed on their campus. When they talk to the faculty, they emphasize how the generosity of the alumni and positive performance of the endowment are bringing new resources to scholarship and teaching. Rarely does it come up that one audience skews conservative and the other skews liberal. When a contentious issue arises, such as a conflict over an invited speaker, presidents tend to grit their teeth and adjust their responses to match the mindset of their audience. Through this "code-switching," they tell alumni concerned with the so-called "woke left" about the few conservative faculty and speakers who successfully gave their talks. They tell the faculty about the importance of academic freedom. Both audiences can be mollified for a while, but when an incident occurs that brings the conflict into the open, the tension is exacerbated by having been swept under the rug earlier.

University administrators should stop playing both sides. It's antithetical to the freedom of thought and constructive dialogues that should percolate throughout the campus. A more forthright approach would allow college presidents to say, for example, "I disagree with this person but believe the university should be a place where they can speak." The stammering and decision reversals, like what we just witnessed by UF, just breeds suspicion from both sides.

Academic freedom is fragile. We need university leaders to stop taking steps to avoid offending alumni, trustees, and political figures because it undermines their own institutions. Some truth serum and a more direct approach won't solve all the problems for college presidents—and may even lead to some losing their jobs—but more transparency about their views is a step toward preserving universities as highly trusted institutions by all communities. Universities are not cults or political prizes. It's time for their leaders to explain that in simple and clear terms.

–H. Holden Thorp

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

PHOTO: CAMERON DAVIDSON

10.1126/science.abn4724

NEWS

> “In essence, this campaign was a hall of mirrors, endlessly reflecting a single fake persona.”
>
> **From a report by Meta (formerly Facebook)** finding that Chinese entities created a phony Swiss biologist on social media and amplified his claims that the United States was intimidating the World Health Organization on COVID-19’s origin.

IN BRIEF

Edited by **John Travis**

ENVIRONMENT

Call for U.S. to address plastic waste

The United States is the world’s largest generator of plastic waste and among the top dozen sources of plastic pollution in the ocean, yet it lacks comprehensive strategies to study or reduce the problem, reports a panel convened by the U.S. National Academies of Sciences, Engineering, and Medicine. The country produced 42 million tons of plastic waste in 2016, with as much as 2 million tons of that winding up in the ocean, the report released last week notes. That’s roughly one-quarter of the estimated global total dumped into the ocean each year. Currently, recycling in the United States is “grossly insufficient” to tackle the problem, the panel warns. The group calls for developing a nationwide research strategy by the end of 2022 to better understand the size and drivers of the plastics problem. It also urged greater U.S. measures to cut plastic waste, pointing to initiatives in other countries, such as restricting free plastic bags and requiring manufacturers to take back packaging.

Bringing equity to AI

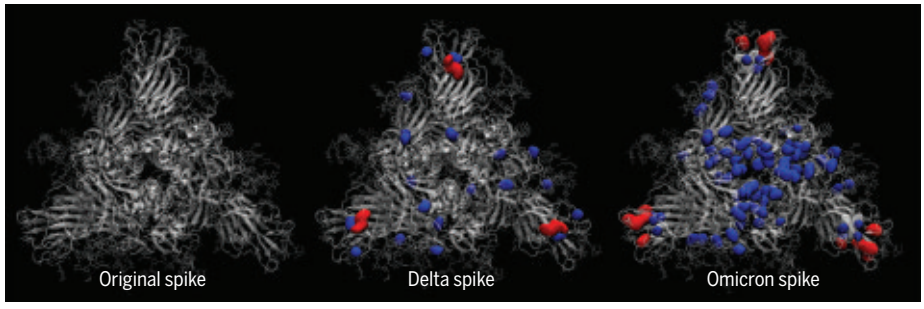
DIVERSITY | Computer scientist Timnit Gebru, whose research showed facial recognition software has a bias against women and people of color, announced last week she’s launching the Distributed Artificial Intelligence Research Institute (DAIR). The nonprofit institute will develop artificial intelligence software serving marginalized groups and aims “to counter Big Tech’s pervasive influence on the research, development and deployment of AI,” according to a statement. In an early project, DAIR researchers will create a public data set of aerial images to study how apartheid shaped land use in South Africa. Gebru co-led Google’s Ethical AI group until last year, when she clashed with the Silicon Valley giant over a paper she co-authored raising concerns about potential harm from language-processing technology used by Google and other companies. (Gebru says she was fired, although Google called her departure a resignation.) DAIR will kick off with $3.7 million in funding from multiple foundations and other groups.

An antibody for Omicron?

COVID-19 | A monoclonal antibody (mAb) authorized this year for COVID-19 neutralizes pseudoviruses equipped with all 37 of the mutations in the spike surface protein of the Omicron variant, according to results from Vir Biotechnology. The company, which developed the drug with GlaxoSmithKline, this week added the latest data to a 1 December preprint reporting that the antibody maintains potent neutralizing activity against pseudoviruses with individual Omicron spike mutations. The drug, called sotrovimab, was authorized by the United States in May and by the United Kingdom last week. EU regulators are still evaluating it. The antibody, delivered by intravenous infusion, is for nonhospitalized people with mild to moderate COVID-19 symptoms. Regeneron said last week its COVID-19 mAb cocktail (casirivimab and imdevimab) may lose some potency against the Omicron variant, and it is conducting lab studies to confirm that. Separately this week, the World Health Organization advised against use of antibody-filled plasma from recovered COVID-19 patients as a therapy, saying it does not improve survival.

Abortion case hears little science

LAW | Science and medicine barely played a role in last week’s oral arguments for a potentially landmark U.S. Supreme Court case, although fetal viability was mentioned often and several groups had filed amicus briefs summarizing relevant research results. The justices are reviewing the constitutionality of a Mississippi law that bans most abortions after 15 weeks of pregnancy. Previous Supreme Court cases had established that abortions could not be banned if the fetus couldn’t yet survive outside the womb, a threshold



A visualization of mutations (blue, red, and white) that distinguish the Omicron spike protein from those of the original SARS-CoV-2 pandemic coronavirus and the Delta variant.

IMAGE: MIA ROSENFELD AND FIONA KEARNS/AMARO LAB/UNIVERSITY OF CALIFORNIA, SAN DIEGO

ECOLOGY

Hippo gut microbes make a splash

A study of hippo pools in Africa suggests the microscopic gut residents of these massive animals have an unexpected influence outside their bodies. Like almost all animals, hippos (*Hippopotamus amphibius*) host beneficial gut microbes, such as bacteria that aid digestion. But when hippos defecate into the large pools they wade and swim in along the Serengeti's Mara River, they release living members of their intestinal microbiome and effectively create a poolwide "metagut," University of Florida ecologist Christopher Dutton and colleagues reported last week in *Scientific Reports*. These microbes can change the water chemistry, which could help explain the massive fish kills that take place during seasonal floods, when pool water is washed downstream, the team proposes. The animals also share bacteria with each other as they drink the water, possibly boosting the digestive capabilities and immune defenses of all the hippos in the pond. "Hippos shape the world around them through microbes associated with their feces," says Robert Naiman, an emeritus ecologist at the University of Washington, Seattle. Other creatures that live in stagnant pools or slow-moving streams, such as alligators, may do something similar, he says.

that is about 23 or 24 weeks of gestation given current medical care. Justice Sonia Sotomayor noted that a pregnant person's risk of death is 14 times higher in childbirth than in abortion, and that few scientists think fetuses feel pain before 24 weeks. But the court largely focused on nonscience questions, such as whether to overturn precedents and whether viability should be the basis of any policy. A decision is expected by next summer.

Biomedical funder expands

PHILANTHROPY | The Chan Zuckerberg Initiative (CZI), which 5 years ago set out to devote $3 billion over 10 years to biomedical research, now plans to more than double its spending and develop tools for studying biological processes. Funded by Facebook co-founder Mark Zuckerberg and pediatrician Priscilla Chan, who are married, CZI's science program launched with a wildly ambitious goal: to prevent, manage, or cure all disease by 2100. It funds research in areas such as neurodegeneration and rare diseases, as well the CZ Biohub, which supports studies of cell biology and infectious disease at three Northern California universities. In the next phase, the organization will create a network of biohubs and two free-standing institutes—one focused on advanced imaging and the second to study natural and artificial intelligence. CZI's total commitment will extend to $6.4 billion over 20 years.

BY THE NUMBERS

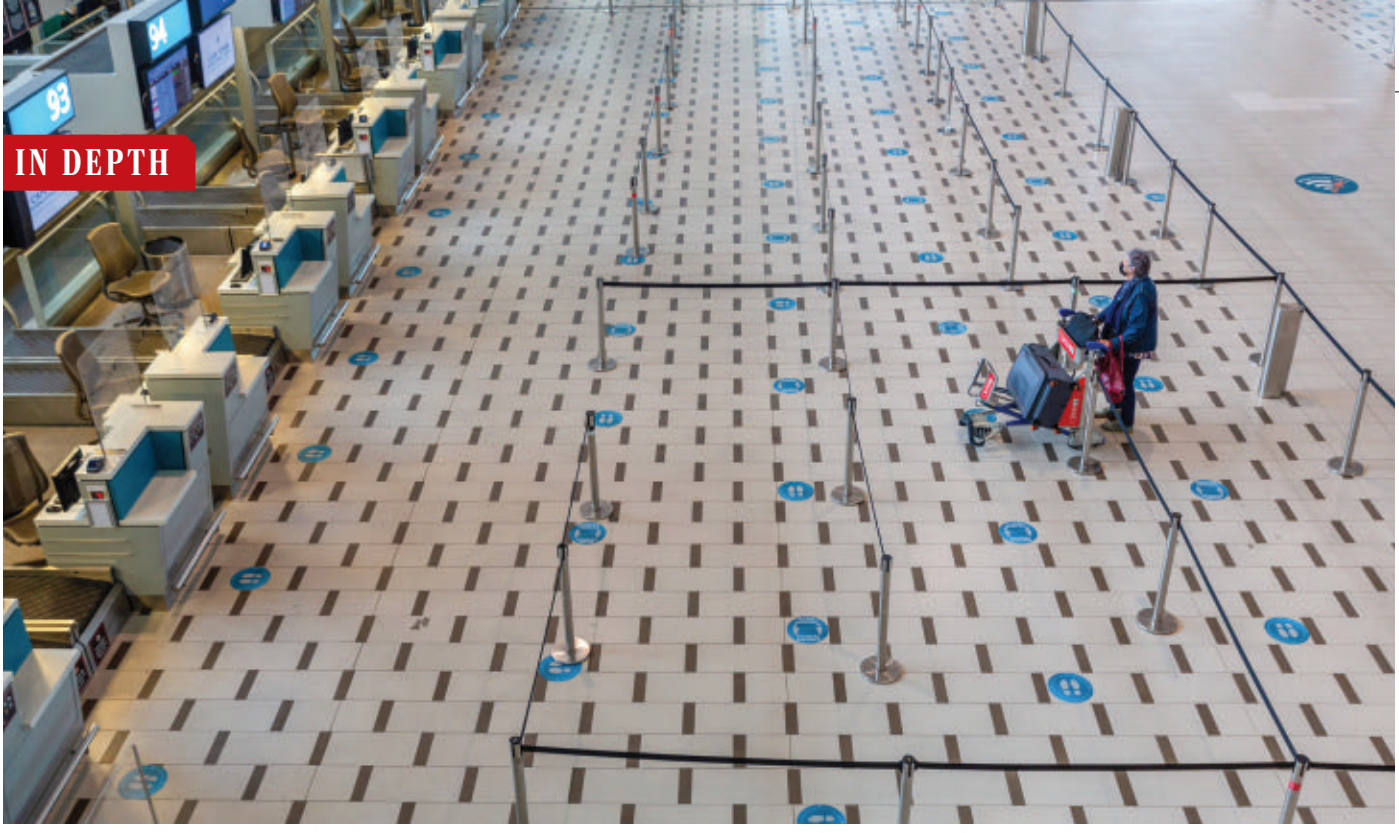
69,000

Estimated increase in global malaria deaths for 2020 compared with 2019, 70% (47,000) of them attributed to pandemic disruptions and mostly children under 5 years old. (World Health Organization's annual malaria report)

Prizes for defending science

AWARDS | An independent science integrity consultant and a racial health disparities researcher who have both faced scientific or public pushback have won this year's John Maddox Prize for standing up for science. Microbiologist Elisabeth Bik received the £3000 award for her data sleuthing, notably uncovering duplicated or modified images that can signal fraud and other problems. Physician-scientist Mohammad Sharif Razai of St George's, University of London won the early career prize for work that included addressing vaccine hesitancy among ethnic minorities and health disparities that result from systemic racism. The award is sponsored by the London-based charity Sense about Science and *Nature*, where John Maddox was editor until 1995.

PHOTO: TONY HEALD/MINDEN PICTURES



The departure terminal at Cape Town International Airport on 3 December. Many countries have halted travel from southern Africa in a bid to slow Omicron's spread.

PHOTO: DWAYNE SENIOR/BLOOMBERG/GETTY IMAGES

COVID-19

How bad is Omicron? Some clues are emerging

New variant appears to evade immunity and shows signs of spreading more rapidly

By **Kai Kupferschmidt** *and* **Gretchen Vogel**

Testing stations and hospital wards in Gauteng, South Africa's most populous province. A company's Christmas party in Oslo, Norway, that became a superspreading event. Infection patterns in the United Kingdom. Scientists are scouring patchy evidence from around the world to better understand Omicron, the new SARS-CoV-2 variant, and what it might mean for the next phase of the pandemic. Three weeks after Omicron was discovered, there are still mostly questions, but a few hints have emerged—some worrisome, others more encouraging.

Researchers are focusing on three key questions: Can Omicron evade immunity from vaccines or previous infections? How transmissible is it? And how much severe disease will it cause?

The most solid clues so far pertain to the first question—and they are not reassuring. The genome alone—with more than 30 mutations in the all-important spike protein—suggested the variant might well be the best yet at dodging our immune defenses. Early data from South Africa seem to confirm that worry. That country has seen three massive COVID-19 waves so far, one with the original SARS-CoV-2, one with the Beta variant (which never made much headway outside the country and has now disappeared), and one with Delta. Past studies found a previous infection offered imperfect, but significant protection against Beta and Delta, and many had hoped South Africa had built up enough population immunity to dampen further spread of SARS-CoV-2.

But a new study, published as a preprint on 2 December, suggests that won't work with Omicron. Juliet Pulliam, an infectious disease epidemiologist at the South African Centre of Excellence in Epidemiological Modelling and Analysis, and her colleagues analyzed 35,670 reinfections among nearly 2.8 million positive tests carried out through late November. The proportion of reinfections rose significantly as Omicron started to spread, suggesting an earlier infection only offers half as much protection against the new variant as it does against Delta. That's a sign Omicron is able to escape at least some of the immune system's defenses. "This does not bode well for vaccine-induced immunity," says virologist Florian Krammer of the Icahn School of Medicine at Mount Sinai.

"The exact numbers are fraught with issues," says William Hanage, an epidemiologist at Harvard University. But, he says, "It is a first pass that provides a good enough comparison to show us that, as we might have expected, reinfections are a big deal with Omicron." How big a problem that will become depends on whether vaccinations and previous infections still protect against severe disease, says Justin Lessler, an epidemiologist at the University of North Carolina, Chapel Hill. "We knew there were going to be variants and there was going to be immune escape. The hope is that all these reinfections are mild and it's not such a big deal," he says. "That's been my index of when the pandemic is 'over'—when we have big waves of infection but not of serious disease."

Whether Omicron is more transmissible than its predecessors—as both Alpha and Delta were—is harder to judge. Omicron cases in South Africa have risen steeply in the past few weeks, but that could be explained in part by chance or the variant's ability to infect those who are vaccinated or had a previous infection.

But Jeremy Farrar, head of the Wellcome Trust, sees cause for concern. "The evidence

that this is more transmissible is getting stronger every day," he says. In the United Kingdom, the number of positive polymerase chain reaction tests in which the gene encoding the spike protein cannot be detected (a sign of a likely Omicron infection) is increasing rapidly. In Oslo, a company Christmas party at a restaurant became a superspreading event, with at least 120 people testing positive; 19 cases so far have been confirmed as Omicron. (All attendees were vaccinated and had tested negative before the event.) In Denmark, 53 of 150 high school students who attended a party went on to test positive for Omicron.

"None of this alone tells us that this is more transmissible," says Kristian Andersen, an infectious disease researcher at Scripps Research. Superspreading events, for instance, have been a hallmark of SARS-CoV-2 from the start. "But Omicron is really rare still, so the fact that we see early cases being associated with superspreading events is quite concerning," Andersen says.

Early signs that Omicron causes less severe symptoms than previous variants offer some reassurance. Doctors in South Africa are reportedly seeing a larger proportion of mild COVID-19 cases in the hospital than at the start of earlier waves. The number of hospital patients infected with SARS-CoV-2 has been rising rapidly, but that includes "incidental" cases—patients seeking care for other reasons who test positive for the virus as well. Data through 6 December indicate the number who needed oxygen support was lower than in previous waves, suggesting fewer patients are suffering the serious lung damage from COVID-19 that has put so many in the hospital during the pandemic.

But it's too early to tell whether Omicron is really more benign. Many early cases in South Africa have been linked to a university outbreak and occurred in young people, who are less susceptible to severe disease. Previous infections could also be providing some protection, as could the steadily climbing vaccination rate in South Africa. Or it might simply be too early to see many serious cases, which can take weeks to develop and always make up a small proportion of the total number. "I haven't seen anything yet that tells me whether this is as severe or less severe, or more severe," Farrar says. "At the moment, my working assumption is that the clinical syndrome of illness is the same as previous variants."

If that assumption holds, but the virus spreads more rapidly than Delta, more people would get severely sick in a short time period, which could mean a huge extra burden on health care systems that are already stretched thin—especially in places with low vaccine uptake and low levels of infection-induced immunity.

Even if Omicron causes milder disease, rapid spread could still quickly overwhelm hospitals in many places. "A small percentage of a large number is still a large number," says genomicist Mads Albertsen of Aalborg University, who serves on a panel advising the Danish government on SARS-CoV-2 variants. And it's not just about deaths and hospitalizations, says Mary Bushman, an epidemiologist at Harvard. "Part of what we need to think about is whether it's causing Long Covid," Bushman says.

More data from countries with different vaccination patterns will soon give a better picture of the threat Omicron poses. Scientists are particularly interested to see whether people who have had a booster shot are better protected.

> **"The evidence that this is more transmissible is getting stronger every day."**
> **Jeremy Farrar**, Wellcome Trust

In the meantime countries are scrambling to slow the variant's spread, with few signs of success. Bans against travelers from southern Africa are quickly losing their justification now that the virus seems entrenched in dozens of countries. Denmark, which has identified 183 Omicron cases so far, is trying to contain spread by broadening quarantine rules—asking not just people infected with the new variant and their close contacts to isolate, but also the close contacts of close contacts. But the rapid spread already makes that strategy impractical, Albertsen says.

That means it's down to the standard defenses such as wearing masks, social distancing, vaccination, testing, and isolation for those who test positive. "It's doing the basics well that matters, whatever the variant is called," Farrar says. Maria Van Kerkhove, an epidemiologist at the World Health Organization, says countries should pay extra attention to getting all of their vulnerable people fully vaccinated, including the elderly and those with conditions that can worsen COVID-19. "These are the people that governments should be targeting right now," she says.

Van Kerkhove is exasperated that with Omicron on their doorstep, many countries in the Northern Hemisphere haven't done enough to control big winter outbreaks of Delta. "We're not even out of the middle of this pandemic yet," she says, "and we're moving in the wrong direction." ■

SCIENTIFIC COMMUNITY

Archaeologist reinstated as Max Planck director

After allegations of bullying, Nicole Boivin returns to supervisory role

By **Andrew Curry**

A Berlin court this week reinstated embattled archaeologist Nicole Boivin as a director of the Max Planck Institute for the Science of Human History (MPI-SHH), a world-leading institute in the study of prehistory.

Boivin was removed from her directorship in October after an internal investigation by the Max Planck Society (MPG) reportedly found she bullied junior staffers and took credit for other researchers' work, among other charges.

Boivin initially agreed to remain at MPI-SHH as a researcher. Then, in November, she filed for an injunction with the High Civil Court in Berlin, seeking to block the demotion while she challenges it in a planned lawsuit. Judges granted the temporary injunction on 6 December, effectively reinstating Boivin to her position, with supervisory authority and a multimillion-euro budget.

In an email to staffers later that day, Boivin celebrated the decision. "I am very happy to be back in place," she wrote. She called her removal "a massive miscarriage of justice I am confident will be rectified in the courts."

MPG has pledged to continue to press for her dismissal. "This is disappointing insofar as this ruling does not take into account the welfare of the employees at the Institute," MPG Secretary General Rüdiger Willems said in a statement. "[MPG] will appeal the ruling." The society planned to meet this week to decide how to "adequately protect the employees," the statement said.

This week's ruling turned in part on MPG's decision to remove Boivin quickly, before the society's Senate had time to approve her demotion. Boivin's lawyer argued that the move violated the society's own bylaws. The court agreed. ■

Chemist Charles Lieber says he is innocent of charges that he failed to fully disclose financial ties to China.

SCIENCE AND SECURITY

Trial of Harvard chemist poses test for U.S.'s China Initiative

Prosecution of Charles Lieber "misguided," critics say

By **Jeffrey Mervis**

The arrest of Charles Lieber on 28 January 2020 drew worldwide attention to the Department of Justice's China Initiative, launched in November 2018 amid concerns that China was improperly benefiting from U.S.-funded research. Lieber, who was the chair of Harvard University's chemistry department at the time of his arrest and celebrated for his research on nanomaterials, is one of 23 U.S. academics, most of them of Chinese descent, who have been targeted under the initiative.

Like most of the other defendants, Lieber is not accused of espionage or theft of intellectual property—two concerns driving the China Initiative. Instead, when his trial begins on 14 December in a Boston federal courtroom, prosecutors will try to persuade a jury that Lieber lied about his ties to a Chinese university and failed to report income to tax and banking authorities.

The trial comes amid calls from academics, Democratic lawmakers, and others for U.S. Attorney General Merrick Garland to end or at least curtail the China Initiative. Critics say the government shouldn't be seeking criminal convictions of scientists who may have only failed to properly document all of their research support—errors that might typically draw civil or administrative penalties. "Lacking evidence of more serious wrongdoing, [Lieber] has become the target of a tragically misguided government campaign [that has] criminalized them," wrote some 40 of Lieber's colleagues, including seven Nobel laureates, in a March letter to Garland.

The government's track record in such cases is mixed. In the past 2 years, eight academic scientists have received prison sentences of up to 37 months after pleading guilty to various charges, mostly involving failure to disclose financial ties to Chinese institutions. Only two, a husband and wife at Nationwide Children's Hospital in Columbus, Ohio, were charged with stealing trade secrets. Last summer, the government dropped its prosecution of seven scientists, six of whom faced charges of violating U.S. immigration law, and its case against one scientist who returned to China before his indictment is unlikely to move forward. (Federal investigators have also brought charges against dozens of scientists working in industry.)

Prosecutors lost the first case against an academic that went to trial, that of mechanical engineer Anming Hu. (Hu is weighing a reinstatement offer from the University of Tennessee, Knoxville, which had fired him.) Now, court watchers are waiting to see how prosecutors fare against six other researchers who have elected to go to trial (see chart, below), with Lieber's case the next in line.

The six charges against Lieber are linked to interactions with the Wuhan University of Technology (WUT). In late 2011, according to the indictment, Lieber agreed to become a "strategic scientist" at WUT and to lead a new "WUT-Harvard joint nano key laboratory" there. Several months later, according to the government, Lieber agreed to join a Chinese foreign talent recruitment program. Together, prosecutors say, the new positions provided him with a monthly salary of up to $50,000, $150,000 in living expenses, and $1.74 million over 3 years in research funding.

In most instances, it is not illegal for U.S. academics to receive foreign support. But the government alleges Lieber violated federal requirements to report such arrangements to

The China Initiative heads to court

Six academic researchers are preparing to contest an array of federal charges, including defrauding federal research agencies and making false statements. Two have been fired and two suspended without pay.

NAME AND INSTITUTION*	CHARGES	ARRESTED	TRIAL DATE	UNIVERSITY STATUS AFTER ARREST
Charles Lieber (Harvard University)	False statements to Department of Defense (DOD), National Institutes of Health; false reporting	January 2020	14 Dec. 2021	Suspended with pay
Simon Ang (University of Arkansas)	Wire fraud on NASA, DOD grants; false statements	May 2020	7 Feb. 2022	Terminated
Zhengdong Cheng (Texas A&M University)	Defrauding NASA, false statements	August 2020	4 Apr. 2022	Terminated
Franklin Tao (University of Kansas)	Wire fraud on Department of Energy (DOE), National Science Foundation (NSF) grants; false statements	August 2019	18 Apr. 2022	Suspended without pay
Gang Chen (Massachusetts Institute of Technology)	Wire fraud on DOE grant, false reporting	January 2021	Not scheduled	On leave with pay
Mingqing Xiao (Southern Illinois University)	Wire fraud on NSF grant, false reporting	May 2021	Not scheduled	Suspended without pay

*at time of arrest

CREDITS: (PHOTO) AP PHOTO/CHARLES KRUPA; (DATA) U.S. DEPARTMENT OF JUSTICE

two of his funding agencies, the Department of Defense (DOD) and the National Institutes of Health (NIH). Harvard didn't learn about the deals until 2015, the government says.

In April 2018, according to prosecutors, Lieber told DOD that "he had never been asked to participate" in the foreign talents program. Several months later, prosecutors say Harvard told NIH that Lieber "has represented that he is not and has never been a participant" in a Chinese talents program. The government says both assertions were false.

Lieber has pleaded not guilty to all the charges, and his lawyers declined to comment. "We will do our talking in the courtroom," says Marc Mukasey of Mukasey Frenchman LLP, a New York City law firm.

So far, Lieber's attorneys have been largely unsuccessful in attacking the government's case. U.S. District Court Judge Rya Zobel, who will preside, denied a request to suppress statements from Lieber's 3-hour interrogation after his arrest, ruling that he had not been coerced. A different judge threw out Lieber's request to clarify what it means to "participate" in a foreign talents program, part of his assertion that federal disclosure rules are ambiguous and confusing.

For Lieber to be found guilty of lying to DOD and NIH, the prosecution must prove he knew his statements were false and that the information was germane to his grants from those agencies. (In Hu's case, U.S. District Judge Thomas Varlan found that Hu did not intentionally mislead NASA, that university officials had assured Hu that NASA rules restricting work with China didn't apply to his research project, and that Hu had done the work he promised to NASA.)

The tax and banking charges against Lieber—which were added several months after his arrest—are less open to interpretation. "You either check the box or you don't" on the relevant paperwork, said one lawyer familiar with the case who requested anonymity. The lawyer speculates that prosecutors added the charges to induce Lieber to make a plea deal. "You don't always criminalize [those violations]," the attorney says, "but it increases the pressure … to plead guilty to something."

One common thread in the pending cases is their staggering cost. Chen and Lieber are the only two defendants still receiving a salary from their universities. Several have launched fundraising campaigns to cover legal fees that could exceed $1 million, with varying success. Just one defendant, Massachusetts Institute of Technology engineering professor Gang Chen, has an employer willing to pay for his defense. Harvard has refused, saying Lieber violated his contract by not fully disclosing his WUT ties. ■

GENETICS

Gene editing produces single-sex litters in mice

CRISPR approach may curb culling of lab animals, chicks

By **Elizabeth Pennisi**

In some farmers' ideal world, cows would birth only females, sows would bear no boars, and chicks would all grow up to be hens. There would be no need to kill millions of male animals, which don't produce eggs or milk.

Now, scientists are a step closer to this reality. Researchers have harnessed the gene editor CRISPR to produce litters of mice of all one sex. That's a potential boon to agriculture and may offer a more immediate advantage in research. "The paper shows a state-of-the-art solution to producing single-sex species" with "impressive results," says Ehud Qimron, a CRISPR expert at Tel Aviv University who was not involved with the work.

The impact for lab animals may be huge. "In the past 5 years around 25,000 papers were published using mice in sex-specific research studies," says co-author James Turner, a molecular geneticist at the Francis Crick Institute. "If we could prevent the generation of the unstudied sex, the number [saved] would be in the hundreds of thousands."

Other methods exist to skew the male/female ratio of newborn animals. Scientists can sort sperm by the weight of the sex chromosome (the Y is lighter than the X), or cause embryos of one sex to die before birth. In a study published 2 years ago, researchers using CRISPR managed to produce altered mice in which four of five litters were all female.

In the current work, that efficiency jumps to 100%, a welcome gain. The authors also target a gene that is conserved in many animals, so the technique may prove useful for more than just mice. "The approach seems generalizable to other animal species," including birds and fish, says Michael Wiles, a molecular geneticist at the Jackson Laboratory who was not involved with the study. It could even help with the recovery of endangered species, Wiles says, depending on which sex was in short supply.

Many kinds of research studies require animals of a particular sex. Scientists engineering new lines of genetically modified mice prefer males, which can father new litters every few days, whereas females require 6 to 8 weeks to produce one. Studies of reproductive tissues only need males or females, as do certain hormone and cancer studies. "Developing ways to have same-sex litters is of utmost importance," Qimron says.

In the new study, Turner and Crick molecular geneticist Charlotte Douglas and Peter Ellis from the University of Kent began by identifying a molecular target that would effectively eliminate embryos. It had to be "expressed at a high enough level and at the right time during development to ensure 100% elimination," Turner says. The team picked a gene called *Topoisomerase 1* (*TOP1*), which is key for cell division; disabling it should lead to the quick demise of a very early embryo.

The team then separated the two parts of CRISPR: the enzymatic complex that physically disrupts a target gene, and the "guide RNA" that recognizes the gene and guides the complex to the right spot in the genome. The researchers put the guide RNA targeting *TOP1* into the mother mouse's genome and attached the DNA encoding CRISPR's cutting complex to the father's Y chromosome. The enzyme and guide RNA were reunited only when sperm with the Y chromosome fertilized the female's eggs, creating the X/Y combination that defines a male.

When the developing male embryo was only a few dozen cells, the effects of gene editing kicked in, killing that embryo before it got a chance to implant in the mother's womb. And indeed, no male pups were born, the team reported last week in *Nature Communications*. The reverse was true when researchers attached the CRISPR complex to the male's X chromosome; in that case, CRISPR kicked in when the male and female X chromosomes were united, and all the female embryos failed to implant.

The modification also made things easier on the mouse mothers. When targeted embryos die before implantation, the mother has fewer embryos to sustain, making it more likely that the other embryos will thrive. Perhaps that's why the disappearance of the undesirable embryos resulted in larger than expected litters of the desired sex. (The litters should have been half the usual size, but they were just 30% to 40%

smaller.) *TOP1* is present in many animals, so the approach should work in other species, the team says.

That could alleviate some ethical dilemmas, some researchers say. "The choice is made before the animal is born," says Tak Mak, a geneticist at the Princess Margaret Cancer Centre. For his breast cancer studies, he only uses female mice and so has had to sacrifice all males born. This new technology "will eliminate this unpleasant and inefficient reality," he says.

Although Wiles expects same-sex litters may become routine in labs, he's not optimistic about agriculture. "Sadly, it's 'dead in the water' for food production while the world fears [genetically modified organisms]."

Sue Leary, president of the Alternatives Research & Development Foundation, a nonprofit seeking alternatives to animals in research, says she sees little benefit to the new approach. "You can't solve an ethical problem with another ethical problem, which is genetic engineering," she says. The method is "not pragmatic" for culling farm animals, she argues, saying researchers should adopt alternatives to using mice.

Few seem concerned that the method will be tried in humans anytime soon. It works best with species that have big litters and short gestations, Qimron says. And it requires genetic modifications not feasible in humans.

These mice have been genetically modified to parent single-sex litters.

The team is making mice capable of producing these litters freely available to academic scientists. Tel Aviv University immunologist Mordechay Gerlic, whose lab is the one that managed to produce some all-female mice litters with CRISPR a couple years ago, says he looks forward to the next steps. "I was very happy to see that they used a similar approach, and target the same region [in the Y chromosome]," Gerlic says. In the future, as the approach expands to other animals, he says, "it will be important to define other regions in the Y chromosome." ■

THE NEW NORMAL

How pandemic pressure is re-engineering clinical trials

COVID-19 spurred nimbler trials that demand less of patients—and many changes may stick

By **Jennifer Couzin-Frankel**

The pandemic's shadow has Tami Minnier questioning everything about medical practice—including how clinical trials scrutinize it. "We have all these traditions that we haven't really challenged," says Minnier, a nurse at the University of Pittsburgh Medical Center (UPMC) and its chief quality officer.

Like many others who operate the global machinery of trials, which test everything from cancer drugs to the benefits of exercise, Minnier has spent years lamenting their shortcomings, including their sluggish pace, their shortage and lack of diversity of volunteers, and that they're often too small to produce conclusive results. But change has been slow. COVID-19 upended that, compelling trials to adapt to new constraints, such as limits on in-person participation. It also led to a proliferation of sometimes creatively designed trials testing ways to fight the virus.

Clinical trialists hope many reforms will stick. "I'm really interested in the extent to which the pandemic has served as a forcing mechanism to change the conduct of trials for the better," says Joseph Unger, a health services researcher and biostatistician at the Fred Hutchinson Cancer Research Center. "It's like a natural experiment."

UPMC, with hundreds of active clinical trials, is one of many places around the world where the experiment is playing out. Prepandemic, clinical research "could be methodical and careful and take years," Minnier says, but when patients seriously ill with a new disease flooded emergency rooms, health care providers were awash in questions. Should they give steroids? Would flipping someone onto their stomach improve breathing? When should doctors resort to a ventilator? Queries quickly became research studies. This "questioning of, 'Why are we doing what we're doing'" in patient care was, Minnier says, "the driving force for me personally" to start to rethink so much across medicine.

A 10-minute walk away, at UPMC's 600-bed main hospital, physician Derek Angus treats patients in the intensive care unit (ICU) and runs clinical trials for sepsis, a bloodstream infection. Although sepsis is the leading cause of death in hospitals, sepsis trials have, historically, accomplished little. Patients and families often decline to participate because the trials test one drug at a time, often with a 50% chance of getting a placebo. "Meanwhile, you're really sick, in septic shock," says Angus, who has tried for years to design what he calls "smarter, more efficient trials."

After swiveling to COVID-19, Minnier and Angus bonded over their desire for trials that offer quick, clear answers. Part of the solution, they believe, is to fuse clinical research with patient care, which can both speed recruitment and lighten the burden on participants. For example, if a clinical trial requires the same blood test a patient's doctor is seeking, Angus says, it could tap into those results rather than requiring a separate blood draw. Once patients are enrolled, he wants trials to feel seamless, a model he compares to the iPhone his elderly mother operates with ease. The phone is "designed to look effortless," he says, "even though the engineering is more complicated now than it's ever been."

Behind the scenes, re-engineered data infrastructure at UPMC is smoothing the process. In the years preceding COVID-19, Oscar Marroquin, a cardiologist who leads the institution's health care data and analytics activities, took steps to meld millions of electronic medical records with ongoing or potential trials, making it easier to flag patients who might be eligible or track them if enrolled. That system became more quickly integrated into research when COVID-19 struck, and Marroquin is now applying it beyond the coronavirus, for example to studies of cancer and back pain.

Minnier and Angus, who was recently named UPMC's chief innovation officer, are also helping other UPMC researchers design statistically powerful studies that yield rapid results. One way to accomplish this is with a pragmatic trial, which assesses the power of interventions in patients already getting them. Another uses Bayesian statistics, which allow patients to be randomized in different ways as the trial progresses, depending on how a treatment performs.

UPMC physicians recently launched a COVID-19 trial that combines those innovations to study monoclonal antibodies, a treat-

PHOTO: FRANCIS CRICK INSTITUTE

ILLUSTRATION: KATTY HUERTAS

ment for recently diagnosed people. Anyone in the UPMC system and the nearby region who tests positive for COVID-19 and is eligible for the antibodies can be referred to get the therapy, and a UPMC physician explains that they will receive one of several options. Those who agree are automatically enrolled in the trial and their care—a single antibody infusion—doesn't change. But behind the scenes, volunteers are randomly assigned to one of three antibody therapies, unless the doctor or patient requests a specific one. "No one knows if they all work equally well," Angus says—an even more pressing question now, given the rise of the Omicron variant.

About 60% of those approached, nearly 10,000 people so far, have consented to get monoclonal antibodies and ended up in the study; typically, 10% to 20% of those eligible for a trial agree to participate. "The sample size is staggering," Angus says, and both he and Minnier plan to stretch these trial designs beyond COVID-19. Angus recently met with pharmacists to discuss comparing different antibiotics in certain hospitalized patients. Minnier wants to do the same for blood pressure medications given in the ICU, and diabetes drugs for outpatients. Pragmatic and Bayesian trials aren't new, but Angus and Minnier say COVID-19 underscored their appeal and feasibility.

"We may have underestimated our ability to pivot and change rapidly," says Peter Merkel, a rheumatologist at the University of Pennsylvania who leads an international network called the Vasculitis Clinical Research Consortium. Several years earlier, Merkel received a grant to experiment with remote informed consent and patient visits for trials, but the project struggled because patients and providers weren't always accepting of the technology, and the computer infrastructure was difficult to navigate.

Pandemic pressure made all the difference. This time, "We were able to turn what were pilot ideas to the norm," Merkel says. COVID-19 forced his network to collect far fewer blood samples from volunteers, and the team is reconsidering the frequency of future sample collection and physical exams. "Are there things you can omit without doing major harm" to the trial, he wonders, such as imaging studies, tissue samples, and frequent in-person checkups?

"There's really no good reason to go back" to the pre–COVID-19 trial model, agrees Yana Najjar, an oncologist at UPMC. Like Merkel, she reoriented her trial testing a drug combination for melanoma, allowing participants to stretch out the time between intravenous doses from 3 weeks to 6 and shipping medications to sites near her patients. These changes, Najjar and Merkel say, not only make life easier for existing participants, but may help diversify trials, by attracting people who live in rural areas, lack transportation, or struggle to find time and resources to visit a trial site regularly. "We want our trials to represent that population with the disease, not just the patients who live" near study sites, Merkel says. "This democratizes trials."

The shift comes with hurdles. "There's much more burden on the data science team," says Manisha Desai, a biostatistician at Stanford University. A trial she's involved in, on pulmonary arterial hypertension, assesses how far participants can walk in 6 minutes at home, while researchers gauge fatigue and shortness of breath via a video link—a COVID-19–driven change. Desai notes the data are "so much noisier than the gold standard," which is measured on an in-person test at a clinic. There's less uniformity among walking paths, for example. Unger is studying to what degree, if any, shunting parts of studies outside cancer centers has compromised quality.

At UPMC, the new normal is a work in progress. Marroquin doesn't know yet how many researchers will be willing to link up their trials with his vast, new infrastructure. And he acknowledges that the system can't help power trials beyond UPMC. Still, like others, he sees change enduring. A key goal, Angus says, is to spur enthusiasm among more patients by offering easier to navigate trials that minimize risk and demand less sacrifice. "The design should basically ask for as little altruism" as possible, he says—while tackling as many questions as it can to speed science along. ■

IXPE's three telescopes sit 4 meters from their detectors on a mast that extends after launch.

IMAGE: NASA/BALL AEROSPACE

ASTRONOMY

X-ray telescope to study how magnetic objects sculpt light

NASA mission could reveal engines driving black hole jets

By **Daniel Clery**

This week, NASA will launch a small mission with an ambitious task: to peer into violent cosmic objects for clues to how they work. By detecting how intense magnetic fields in collapsed stars and black holes align, or polarize, the x-rays they emit, the Imaging X-ray Polarimetry Explorer (IXPE) could reveal how those objects spew out radiation in the first place. "We're just trying to find out how do you produce these x-rays anyway," says principal investigator Martin Weisskopf of NASA's Marshall Space Flight Center.

"The community has been waiting for this for a long time," says x-ray astronomer Feryal Özel of the University of Arizona, who is not part of the project. NASA's Orbiting Solar Observatory 8, launched in 1975, detected a smattering of polarized x-rays from a single source, the Crab nebula, Weisskopf says. Since then, the questions such x-rays might answer have piled up, but no mission has set out to measure them—in part because the 1975 mission suggested there might be too few to be worth the effort. Weisskopf says he proposed dedicated x-ray polarimetry missions several times before succeeding with IXPE, a $190 million mission that is set to launch on 9 December from the Kennedy Space Center in Florida on a SpaceX Falcon 9 rocket.

X-rays are emitted when gas is heated into a 100-million-degree plasma, a roiling soup of electrons and ions. Photons have magnetic and electric fields that oscillate perpendicularly to their path, typically in random directions. However, magnetic conditions at their birth or interactions on their journey can polarize the photons by forcing those oscillations into the same plane.

IXPE will spend at least 2 years scrutinizing x-ray sources with three identical telescopes—cheaper than one big one and a hedge against failures. Each telescope is a cylinder containing two dozen concentric shells that focus x-rays through grazing reflections. (X-rays penetrate standard telescope mirrors.) Sensors built by the Italian Space Agency detect the x-rays and their polarization in a layer of dimethyl ether gas. An x-ray hitting a gas atom knocks out an electron that tends to shoot off in the direction of polarization, leaving a visible trail. Imaging many trails and their spread tells observers how polarized the light is, and in which direction.

Pulsars are a prime target for IXPE. These spinning, city-size remnants of dead stars emit beams of radio waves, x-rays, and other radiation that sweep past Earth like a lighthouse beacon. Weisskopf says rival theories of beam generation suggest the x-rays originate in different locations on the pulsar: all across its surface, at its poles, or in its atmosphere. Each theory predicts the x-ray polarization signal should vary with the timing of the pulses in a way IXPE should be able to distinguish. "We will get a result, as long as we launch," he predicts.

Also in the mission's sights are magnetars, stellar remnants like pulsars, but with even more powerful magnetic fields, 100 million times stronger than any magnet made on Earth. The magnetic field lines force fast-moving electrons into helical paths, causing them to spray out polarized x-rays known as synchrotron radiation. By measuring how the polarization of the x-rays changes as the magnetar spins, observers will be able to map the geometry of the field across the entire globe—and watch for tangles that lead to eruptive outbursts. "With IXPE we may really see the magnetic field direction, the twisted magnetosphere," says theorist Matthew Baring of Rice University.

Polarization could also reveal how hungry the supermassive black hole at the center of our Milky Way has been in recent history. Gas and dust swirling around an active, accreting black hole emit x-rays as they are heated by gravitational forces close to the event horizon. The x-rays now arriving directly from the Milky Way's black hole are dim, suggesting it is quiescent. But x-rays emitted longer ago could also be arriving, having followed dogleg paths after being scattered by distant gas clouds. That scattering should imprint a polarization on the x-rays, a sign that they came from the galactic center and not the cloud itself. "If we look at clouds and see polarization, that would be a smoking gun," says astrophysicist Philip Kaaret of the University of Iowa, who worked on a rival polarimeter proposal that lost out to IXPE. Their brightness might show whether the black hole was more active tens of thousands of years ago.

IXPE's biggest coup might be in helping understand the mechanics of powerful jets launched by supermassive black holes in distant galaxies. These jets blast material up to 10 million light-years out into space—that's 100 times the diameter of the Milky Way. Researchers believe the spin of the black hole funnels material and field lines from churning charged particles near the accretion disk into the jets, which spew from both poles.

Other telescopes have seen x-rays emitted close to the base of the jet, and researchers believe synchrotron radiation is responsible. But what powers the electrons to near–light-speeds as they spiral around the field lines? One possibility is shock waves created by fast-moving plasma colliding with slower moving material. Another is magnetic reconnection, when stressed field lines snap and reconnect. Both processes release energy that can accelerate electrons.

Polarization measurements by IXPE could offer insight into how big and chaotic this emission region is and which process accelerates the electrons. Both might be at work, Baring says. "That makes nature richer," he says, "and keeps theorists busy." ■

BIOMEDICINE

Key cancer results failed to be reproduced

Project to replicate high-impact preclinical cancer studies delivers sobering verdict

By **Jocelyn Kaiser**

An ambitious project that set out 8 years ago to replicate findings from top cancer labs has drawn to a discouraging close. The Reproducibility Project: Cancer Biology (RP:CB) reported this week that when it attempted to repeat experiments drawn from 23 high-impact papers published about 10 years ago, fewer than half yielded similar results.

The findings pose "challenges for the credibility of preclinical cancer biology," says psychologist Brian Nosek, executive director of the Center for Open Science (COS), a co-organizer of the effort. The project also points to a need for authors to share more details of their experiments so others can try to reproduce them, he and others involved argue. Indeed, vague protocols and uncooperative authors, among other problems, ultimately prevented RP:CB from completing replications for 30 of the 53 papers it had initially flagged, the team reports in two capstone papers in *eLife*.

"It is useful to have this level of objective data about how challenging it can be to measure reproducibility," says Charles Sawyers of Memorial Sloan Kettering Cancer Center, who reviewed the designs and results for some of the early replication studies. But, he adds, "It's hard to know whether anything will change as a consequence."

Nosek's center and the company Science Exchange set up RP:CB in 2013, after two drug companies reported they could not reproduce many published preclinical cancer studies. The goal was to replicate key work published by journals such as *Science*, *Nature*, and *Cell* from 2010 to 2012. With foundation funding, the organizers designed replication studies that were peer reviewed by *eLife* to ensure they would faithfully mimic the original experiments.

The project's staff soon ran into problems because all the original papers lacked details such as underlying data, protocols, statistical code, and reagent sources. When authors were contacted for this information, many spent months tracking down details. But only 41% of authors were very helpful; about one-third declined or did not respond. Other problems surfaced when labs began experiments, such as cells that did not behave as expected in a baseline study.

The project ended up paring the initial list of 53 papers, comprising 193 key experiments, to just 23 papers with 50 experiments. They completed all replications for 18 of those papers and some experiments for the rest; starting in 2017, the results from each one have been published, mostly as individual papers in *eLife*. All told, the experimental work cost $1.5 million.

Results from only five papers could be fully reproduced. Other replications yielded mixed results, and some were negative or inconclusive. Overall, only 46% of 112 reported experimental effects met at least three of five criteria for replication, such as a change in the same direction—increased cancer cell growth or tumor shrinkage, for example. But even when the effects reappeared, their magnitude was usually much more modest, on average just 15% of the original effect. "That has huge implications for the success of these things moving up the pipeline into the clinic. [Drug companies] want them to be big, strong, robust effects," says Tim Errington, project leader at the COS.

The findings are "incredibly important," says Michael Lauer, deputy director for extramural research at the National Institutes of Health (NIH). At the same time, Lauer notes the lower effect sizes are not surprising because they are "consistent with … publication bias"—that is, the fact that the most dramatic and positive effects are the most likely to be published. And the findings don't mean "all science is untrustworthy," Lauer says.

Indeed, labs have reported findings that support most of the papers, including some that failed in RP:CB. And two animal studies that weren't replicated by RP:CB have led to promising early clinical results—for an immunotherapy drug and a peptide designed to help drugs enter tumors.

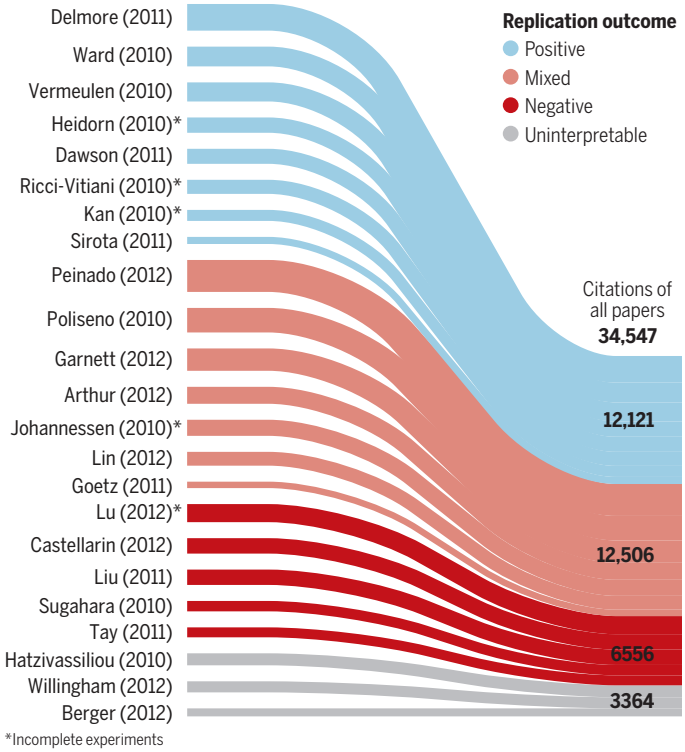
Still, the findings underscore how elusive reliable results can be in some areas. Johns Hopkins University infectious diseases physician-scientist Cynthia Sears, who reviewed papers on links between gut bacteria and colon cancer that were not fully replicated, says simple differences, such as the local bacteria in a lab animal quarters, can sway results. RP:CB, she adds, was "an instructive experience."

If there's one key message, it's that funders and journals need to beef up requirements that authors share methods and materials, RP:CB leaders say. NIH data sharing rules starting in January 2023 may help, Lauer notes.

Such rules may help but not fully solve the problem, Sawyers says. "In the end, reproducibility will likely be determined by results that stand the test of time, with confirmation and extension of the key findings by other labs." ■

Disappointing numbers

Out of 53 prominent preclinical cancer papers, only 23 could be put to the test, and many did not have clearly reproducible results.



*Incomplete experiments

CREDITS: (GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) REPRODUCIBILITY PROJECT: CANCER BIOLOGY

FEATURES

WORLD WIDE WEB

Researchers are weaving optical fibers into a low-cost network of seismic sensors that can probe Earth's hidden dynamics

By **Paul Voosen**

In August, as Iceland's long summer days began to wane, Sölvi Thrastarson made his 10th visit to a volcano so young it lacks a name. Since it began to erupt in March on a peninsula a short drive south of Reykjavík, the volcano has drawn flocks of tourists to its fiery but relatively tame outbursts. Thrastarson, a geophysics graduate student at ETH Zürich, joined a throng of sightseers on a ridge of weathered basalt 2 kilometers from the summit, and watched as globs of lava leaped out of the caldera. Camera shutters clicked as the volcano surged. "This really is the perfect tourist eruption," Thrastarson says.

The swelling crowds have been a boon for Thrastarson and his science. In April, Neyðarlínan, the Icelandic emergency telecom company, extended a fiber optic line to the volcano, providing internet access to a region lacking cell service. For the tourists, the fiber is a digital lifeline. For the researchers, it is a way to take the volcano's very pulse.

PHOTO: WOJCIECH GAJEK

In 2019, researchers in the Swiss Alps brought fiber to the foot of the Rhône Glacier (in the valley behind the foreground ridge) and detected motions that were traced to icequakes.

For several weeks in late spring, from a small utility shack in an impromptu parking lot plowed out of an old lava field, Thrastarson and colleagues tapped the fiber with an "interrogator"—a box that fired laser pulses along the cable and recorded the response. Every few days, a team member drove in from Reykjavík to retrieve the data set. From the comfort of their cozy homes, the researchers sifted through the data for clues to the interior fluctuations driving the restive volcano. It was about as easy as volcanology can get. The hardest part, Thrastarson says with a laugh, was fending off tourists who thought their shack was a bathroom.

The demands of the internet era have draped the world in a web of fiber. Fiber now connects neighborhoods and utilities, traffic cameras and mass transit. For geoscientists, fiber had always been a means to an end—a way to hook up a weather station or an undersea pressure sensor. But that has changed in the past few years. Now, the fiber itself is the sensor, says Jonathan Ajo-Franklin, a geophysicist at Rice University. "We're seeing data where we've never seen data before."

The principles of fiber optic sensing are relatively simple. The cables are bundles of glass fibers, each no thicker than a human hair, that carry information encoded in light. Small, randomly oriented defects within the fibers act like tiny mirrors, scattering light. The interrogators—or boxes, as most researchers call them—work much like radars. They fire a laser pulse into an unused fiber and record the pattern of reflections coming back from defects along the length of the cable. When an external pressure wave crosses a section of the fiber—be it from an earthquake or a footfall—it stretches and squeezes the defects. The reflections in that section are displaced by nanometers, leading to a phase shift in the rebounding light. By firing thousands of pulses per second, researchers can build up a picture of a passing seismic wave, at a distance of up to 100 kilometers or more along the fiber.

Unlike traditional seismometers, which are spaced many kilometers apart, fiber offers the equivalent of a seismometer every meter or two along the cable. This density, combined with the low cost and ruggedness of fiber, has prompted researchers to lay cables on glaciers, volcanoes, permafrost, and earthquake fault zones—any place the earth might crack or crunch, grind or grate. They have also tapped into unused "dark" fibers in existing telecom cables to pick up vibrations from sources as faint as pedestrians and cars. "This fiber

At a new, unnamed volcano near Reykjavík, Iceland (left), fiber has revealed noise that may correspond to eruptive bursts. Along the rim of Iceland's ice-capped Grímsvötn volcano (right), researchers buried fiber with a weighted sled—and heard a mysterious hum that could be magma-fueled tremors ringing the ice cap like a bell.

is really everywhere, and that is a game changer," says Biondo Biondi, a geophysicist at Stanford University. "We can have seismic sensors everywhere on the cheap."

For the past few years, many fiber measurements were mere proofs of principle. But the field is now maturing, says Andreas Fichtner, a seismologist at ETH Zürich and Thrastarson's adviser. Fiber is revealing previously unknown earthquake faults, the hidden mechanics of glaciers and avalanches, and volcanic gurglings that could aid in predicting eruptions. "It's getting beyond the hype, to where people start doing science with it," Fichtner says.

LIKE MANY BREAKTHROUGH scientific techniques, fiber sensing has its origins in U.S. military research. Beginning in the 1980s, the Navy towed fiber optic cables behind ships to sense the sounds of enemy submarines—one reason the technique is still sometimes called distributed acoustic sensing (DAS). By the late 2000s, the oil and gas industry was lining its pipelines and boreholes with fiber. Technicians used backscattered laser light to look for sharp temperature changes—a sign of a ruptured well or pipe—or to detect artificial seismic waves from air guns on the surface, in order to probe the structure of the surrounding rock. Today, the fracking industry uses borehole fibers to monitor rock fracturing and the microearthquakes caused by the high-pressure injection of water.

By the mid-2010s, academic scientists were adopting the technique. A pioneering effort came in 2015, when a team of scientists at the German Research Centre for Geosciences (GFZ Potsdam) took advantage of unused fibers on a 15-kilometer cable connecting two geothermal power plants in Iceland. Philippe Jousset and his colleagues were able to not only detect distant earthquakes, but also locate the rupture sources by measuring differences in the arrival times of earthquake waves on either side of small bends in the cable.

In California, meanwhile, Eileen Martin, then a student with Biondi, was running a 2.5-kilometer loop of fiber in utility tunnels under Stanford. In that hushed milieu, the array picked up not only earthquakes, but also the vibrations of traffic, footfalls—even waves on the ocean. It only went offline once, Martin adds, when someone jostled the fiber in the computer room.

The nascent field's biggest splash came when Ajo-Franklin and colleagues tapped a fiber the Monterey Bay Aquarium Research Institute runs off the California coast to undersea instruments. They deployed their interrogator box in 2018, during a 4-day maintenance shutdown of the undersea instruments, and detected a small earthquake that struck California at the time. More interesting was what the arrival times of the earthquake waves revealed in the surrounding rock: a previously unknown fault zone under the cable, just 10 kilometers off the coast. The finding showed that fiber could detect unknown earthquake threats. For fiber, says Martin, now at the Colorado School of Mines, "that was the single most convincing case anybody has made."

Fichtner had specialized in using data collected by others to image the planet's deep interior. But in DAS, he saw a potential paradigm shift in the making. The promise of the technique turned him into a field seismologist. He decided to target remote, frozen environments where traditional seismometers are especially expensive and difficult to deploy.

Those efforts started close to home with the Rhône Glacier in the Swiss Alps, where 15 square kilometers of ice is retreating in the face of global warming. In 2019, near the front of the glacier, Fichtner's group laid a kilometer-long cable in the shape of triangle, covered it with snow, and left it in place for 1 week. Despite the short deployment, the fiber captured a set of fast motions that were traced to occasional icequakes—previously seen only on ice sheets, not smaller glaciers. The fiber triangle traced the source of the quakes to the same slip patch—suggesting some parts of the glacier bed behave almost like earthquake fault zones. "This changes our understanding of how glaciers move significantly," Fichtner says.

In January, the researchers studied a far faster flow of snow and ice: avalanches. They attached their box to a cable running up a snow-draped mountain in southwestern Switzerland at the Sionne Valley avalanche test site, where researchers use helicopter-dropped explosives to trigger artificial avalanches. As the avalanches thun-

PHOTOS: (LEFT TO RIGHT) KRISTÍN JÓNSDÓTTIR; HILDUR JÓNSDÓTTIR

dered down the slope, the fiber detected internal structures in the flows, including "roll waves" produced by instabilities in their dense flowing core. "You see waves that had only been predicted theoretically," Fichtner says. "You see them immediately."

Buoyed by these results, the team shifted its sights to one of Iceland's largest and most dangerous volcanoes, Grímsvötn, which is buried beneath an ice cap but monitored only by a lone seismic station. It last erupted for 4 days in 2011, blowing a hole in the ice cap and lofting ash into the stratosphere that grounded hundreds of transatlantic flights. As global warming accelerates melting of the ice cap, it could uncork Grímsvötn; the declining weight could allow pressurized magma to more easily fracture its rock jailhouse and escape. The researchers wanted to know whether Grímsvötn was growing more restless.

That led Fichtner, Thrastarson, and others to make the 2-day trek to the caldera in April. Using a sled weighed down by an oil barrel and towed by a snowcat, they plowed a half-meter trench in the ice and snow, burying a 12-kilometer-long cable that circumscribed the caldera before bending into the center. They left the instrument box running for several months, beaming its data back. Fichtner was astonished at the upheaval they detected: some 1800 small earthquakes within 10 kilometers of the caldera, more than 10 times the number seen by the seismometer.

They also picked up a distinct hum, smoother than any known kind of volcanic tremor. Their best guess is that it is a resonance, generated as tremors ring the 300-meter-thick ice cap like a bell. If so, changes in the hum could signal changes in ice thickness, or provide warning of increased volcanic activity, Fichtner says.

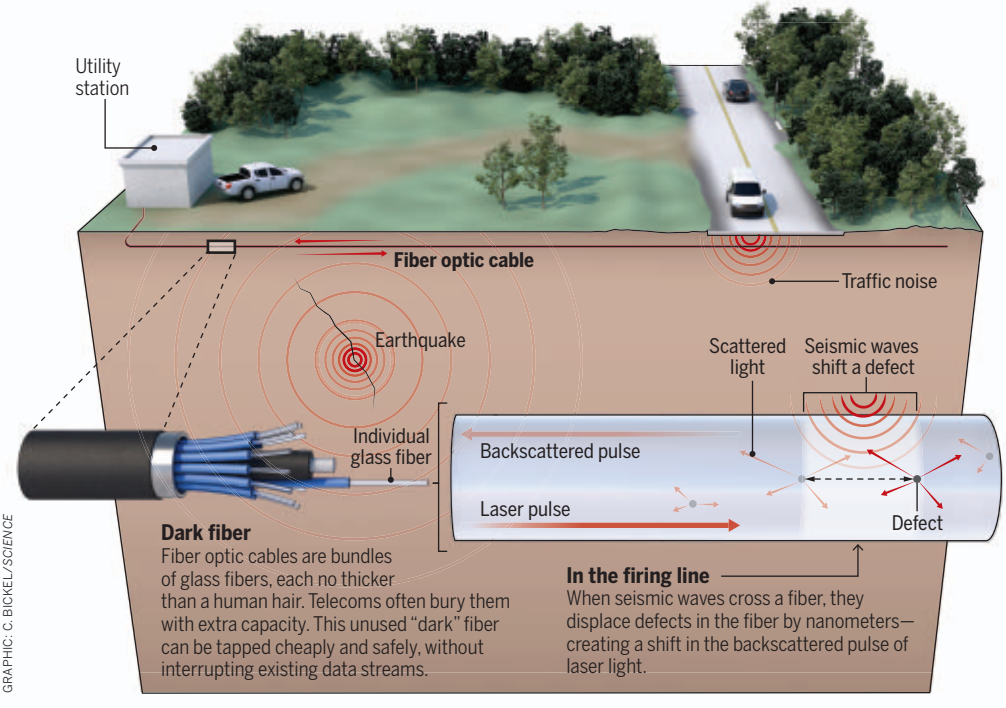
Fichtner isn't alone in his focus on frozen worlds. In 2016, Ajo-Franklin led a team to Fairbanks, Alaska, to see whether fibers could monitor permafrost, the subsurface layer of frozen soil, chock full of preserved organic matter, that is threatened by climate change across much of the Arctic. They laid a 4-kilometer-long fiber cable in a crisscrossing array at a military research site. More than 100 small borehole heaters warmed the surrounding soil, and an instrument that wobbles like a badly loaded washing machine set off vibrations at the surface. The resulting data—hundreds of terabytes, flown back in hard drives in students' luggage—showed the fibers could indeed detect thawing: Meltwater significantly slowed the speed of the seismic waves. The study, to be published this year, raises the possibility of using dark fiber and ambient earthquake waves to track permafrost thaw—which could inform projections of how much carbon these thawing soils will release with continued Arctic warming.

TWO YEARS AGO, a 7.1 magnitude earthquake struck Ridgecrest, California, 180 kilometers north of Los Angeles—the most powerful quake to hit California in 20 years. Zhongwen Zhan, a seismologist at the California Institute of Technology, was elated. He knew Ridgecrest would ring for days with aftershocks; by deploying a fiber array to detect and analyze them, he could learn about other earthquake hazards. The local telecom helped his team hook up four boxes on unused fibers, including one stretching 8 kilometers under the town. In just a few days, the team had installed the equivalent of thousands of seismic sensors. And then they watched while the ground continued to shake.

On a traditional seismic hazard map, Ridgecrest was a single pixel, the whole area lumped together with the same risk. The aftershocks caught by the fiber array revealed the drastic variations within that pixel, with one side of the town shaking three times more strongly than the other. Using a different set of vibrations—mini–seismic waves from traffic—to image the shallowest parts of subsurface, Zhan and colleagues found that the shaky side of town sat on far looser sediments, a risk unknown to residents. The array also identi-

Good vibrations

Fiber optic cables don't just carry data from sensors; now, they are the sensors. By offering the equivalent of a seismometer every meter or so along their length, fibers portend a cheaper way to study motions on Earth.



GRAPHIC: C. BICKEL/*SCIENCE*

A world of noise

Existing or newly deployed fibers have much to listen to.

Earthquakes
Fibers not only locate distant quakes, but also reveal the composition and structure of surrounding rock "illuminated" by seismic waves.

Volcanoes
In regions that are barely monitored by traditional seismometers, fibers can be used to detect the small tremors that often foreshadow eruptions.

Cryosphere
Fibers have revealed hidden motions in glaciers and avalanches. They could also be used to detect the timing and extent of thawing permafrost.

Oceans
Offshore cables have detected storms, ship traffic, and acoustic waves that probe temperatures in the deep ocean—a way to assess its capacity to buffer climate change.

City monitoring
Shifts in traffic and footfalls have been correlated to pandemic lockdowns. The data could also be used to track commercial activity.

fied faults that were misplaced or missing in geological maps. "We were able to see it all," Zhan says.

In the future, Zhan hopes such a rapid response won't be necessary, because DAS arrays exploiting unused telecom fibers will be integrated into the permanent seismic networks that continuously monitor earthquake-prone regions and generate early warnings. Those telecom fibers will be especially useful in remote regions and offshore, where traditional seismic sensors are expensive to establish and maintain. "DAS really shows strong potential for detecting earthquakes earlier than conventional sensors," he says.

Offshore fibers could detect not just earthquakes, but also shifting pressures from tides and currents. Add those capabilities to the standard instruments that are usually attached to ocean cables, and you have a new standard for ocean observation, says Charlotte Krawczyk, geophysicist at GFZ Potsdam. "This will be a new kind of global monitoring," she says.

Its reach may extend far from shore. Last year, researchers set a new record for DAS length with a next-generation interrogator that detected vibrations at a distance of 120 kilometers on a fiber cable running off Svalbard, the world's northernmost inhabited island. The team, led by Martin Landrø at the Norwegian University of Science and Technology, sensed earthquakes from the Mid-Atlantic Ridge, storms, ocean swells, ship traffic, and the lonesome calls of blue and fin whales, raising hopes for an unobtrusive way of monitoring the mammals, Landrø says. "These are numbers we'd like to get a better grip on."

And in 2020, a seafloor cable deployed from the Canary Islands picked up waterborne sound waves from an earthquake that shook the ocean floor. Zhan was part of a team that used such ocean sound waves, which travel faster in warmer water, to measure the ocean's temperature change over time. If fibers start to pick up these waves regularly, it would be "amazing," he says—a way to identify where the deep ocean is heating up, and whether its capacity to absorb 90% of the heat from global warming is diminishing.

Fiber has made inroads on ice, land, and water. Tieyuan Zhu, a geophysicist at Pennsylvania State University (Penn State), University Park, thinks ground-based fibers can even be used to study the air. With a small array at Penn State, he has shown how they can capture and locate thunder while singling out lightning strikes, which shake the ground in a way that's distinct from thunder. They can also distinguish pelting rain from gusty winds. In June at the nearby Shale Hills Critical Zone Observatory, he and his colleagues deployed a fiber array to probe how storms, growing more severe with global warming, deepen the ground's "weathering layer"—the tens of meters or so of soil and rock that sit above bedrock. Zhu says the fiber easily heard the sounds of rain soaking into the weathering layer from the remnants of Hurricane Ida, which passed through the eastern United States in September. "We saw a beautiful signal variation."

A ship off Chile's coast lays a cable that connects to California. Offshore fiber has revealed storms and swells.

FIBER IS POISED to break out in science, but its potential in the applied world may be even greater, says Nathaniel Lindsey, a geophysicist at FiberSense, a startup in California targeting city-sensing applications. Already, small university arrays, sensitive to car traffic and human footsteps, have witnessed the stark falloff in activity because of the coronavirus pandemic, and its resumption as lockdowns are lifted. They can detect fires, landslides, and all sorts of natural threats. Fibers laid into bridges and buildings can detect when infrastructure is close to failure, and they could replace the motion sensors and cameras used to monitor borders and fences. Unlike cameras, DAS "doesn't care if it's pouring or there's fog," Biondi says. The next step will be getting a city to allow a large, sustained network of DAS arrays. In San Jose, California, Biondi is targeting just that, pitching the city on opening up unused fibers in its telecom network for monitoring.

The downside to fiber is the staggering amount of data it produces. In the past, seismologists could gather data from their isolated, widely spaced instruments and crunch it later, on a laptop. The thousands of distinct "sensors" along a single fiber, in contrast, can gather hundreds of terabytes of data in a matter of days. Long-term arrays will produce petabytes, requiring processing in real time, Fichtner says. "Then we can't store our data anymore," he says. "We will have to learn how to do science on the fly." Many labs are developing artificial intelligence algorithms to efficiently sift through the data and look for patterns.

Fichtner's lab already has a data problem—it has done so many deployments that it doesn't have enough researchers to work through the collected data. They've just started to tease apart data captured from the unnamed Icelandic volcano. One mystery they have yet to unravel: low-frequency pulses of activity every few seconds, which could correspond to small eruptive bursts or the collapse of bubbles in the magma.

Despite the data deluge, Fichtner's team isn't stopping. The researchers just wrapped up a monthlong deployment Greece, where they tapped 30 kilometers of cable running through the north of Athens, overnight adding 10,000 seismometers to the 20 in the region, hoping to capture hidden faults. In October, they interrogated a submarine cable that connects the islands of Santorini and Ios. It runs right past Kolumbo, a large, mysterious underwater volcano that last erupted in 1650 but has shown worrying signs of activity.

With the fibers watching, Fichtner says, Kolumbo is putting on a show. The vibrations just keep coming. ■

PHOTO: GOOGLE

ScienceImmunology.org

READY TO PUT THE SPOTLIGHT ON YOUR RESEARCH?

Submit your research:
cts.ScienceMag.org

Science Immunology
AAAS

Twitter: @SciImmunology
Facebook: @ScienceImmunology

CELL BIOLOGY

A lipid primes the final cut in dividing cells

A newly described pathway activates separation of lens cells at the end of cytokinesis

By **Julie A. Brill**[1,2] and **Andrew Wilde**[2,3]

Defects in cytokinesis, the process of cell separation at the end of the cell division cycle, can lead to aneuploidy, genome instability, and ultimately, cancer (*1*). Accumulating data also implicate cytokinesis defects in various noncancer syndromes, many of which present with neurological or visual abnormalities. The recent discovery that patients with oculoskeletodental syndrome, caused by inactivating *PIK3C2A* (phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2α) mutations, develop early-onset cataracts and secondary glaucoma suggested that the encoded protein might normally function in cytokinesis (*2*). On page 1339 of this issue, Gulluni *et al.* (*3*) report that loss of *PIK3C2A* inhibits abscission, the final stage of cytokinesis, and triggers lens cell senescence and cataract formation in zebrafish, mice, and humans.

PI3K-C2α, which is encoded by *PIK3C2A*, is a phosphatidylinositol (PI) kinase that modifies membrane lipids through phosphorylation. Individual PI phosphates (PIPs) occur within the inner leaflet of cellular membranes, where they form distinct microdomains that recruit specific proteins and thereby regulate diverse cellular processes. The best characterized PIP is phosphatidylinositol 4,5-bisphosphate [PI(4,5)P_2], which, together with phosphatidylinositol 4-phosphate [PI(4)P], establishes plasma membrane identity; PIPs also bind and facilitate the localization and activity of transmembrane and peripheral proteins at the plasma membrane (*4*). Dynamic remodeling of PIPs is a key regulatory mechanism, and multiple studies confirm the importance of PIP remodeling during cytokinesis, notably in the assembly and constriction of the contractile ring that is required before abscission (*5*–*7*).

Inhibition of cytokinesis leads to G_1 arrest in the next cell cycle, followed by senescence (*8*). Gulluni *et al.* found that senescent multinucleated cells (resulting from defective

[1]Cell Biology Program, The Hospital for Sick Children, Toronto, ON, Canada. [2]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada. [3]Department of Biochemistry, University of Toronto, Toronto, ON, Canada. Email: julie.brill@sickkids.ca; andrew.wilde@utoronto.ca

PHOTO: PAUL WHITTEN/SCIENCE SOURCE

Cataracts are characterized by opacity forming in the lens, which can be caused by disrupted cytokinesis that changes the structure of the lens.

cytokinesis) accumulate and intercellular bridges (ICBs), the final membranous connection between two daughter cells, persist in cells from *PIK3C2A*-null patients. These phenotypes suggest a role for PI3K-C2α in late stages of cell division. Indeed, the authors found that PI3K-C2α localizes to ICBs. Direct binding to both $PI(4,5)P_2$ in the plasma membrane and γ-tubulin restricts PI3K-C2α to the ICB central bulge, a structure called the Flemming body. Furthermore, catalytic activity and proper localization of PI3K-C2α are needed for abscission. Gulluni *et al*. determined that PI3K-C2α at the ICB uses PI(4)P as a substrate to produce a localized pool of $PI(3,4)P_2$, a lipid not previously known to act in cytokinesis.

The final stage of cytokinesis involves recruitment of the endosomal sorting complexes required for transport III (ESCRT-III) family of charged multivesicular body protein 4 (CHMP4) proteins to the Flemming body (*9*). Studies in tissue culture cells have revealed that CHMP4 is recruited to the Flemming body through centrosomal protein of 55 kDa (CEP55) and ALG-2–interacting protein X (ALIX). CHMP4B is then released from the Flemming body owing to its polymerization into helical filaments that constrict the membrane of the ICB, leading to membrane fission and physical separation of the daughter cells (*10*). Recruitment and subsequent release of CHMP4B from the Flemming body must be tightly regulated to ensure that abscission occurs at the correct time and not, for example, while lagging chromosomes persist in the ICB, which would create genomic instability (*11*). Notably, heterozygous mutations in *CHMP4B* are linked with early-onset cataracts. However, mice that lack CEP55 develop normally except within the brain. CEP55 is thus not essential for cell division in all cells but is specifically required in neurons, raising the possibility of parallel, cell type–specific pathways to abscission.

Gulluni *et al*. find that lens cells express low amounts of ALIX but can still recruit CHMP4B to the ICB, suggesting that these cells use a variation of the ESCRT-III abscission pathway. In human tissue culture cells, the authors discovered that the requirement for ALIX can be bypassed by the ESCRT-II protein vacuolar protein-sorting–associated protein 36 (VPS36), which binds $PI(3,4)P_2$ in the Flemming body. Depletion of PI3K-C2α results in a corresponding loss of $PI(3,4)P_2$ from this structure, along with VPS36 and CHMP4B. These observations suggest that a specific lipid microdomain is needed to bypass ALIX function during cytokinesis.

The redundancy of the ALIX and $PI(3,4)P_2$ pathways raises the question of why certain cell types might be particularly susceptible to abscission defects. Answers may lie in either the frequency of divisions or the tissue context in which the cell divisions occur. Adherent cells grown in culture and depleted of abscission factors, including fibroblasts lacking CEP55, display only low levels of cytokinesis failure. By contrast, mice that lack CEP55 and ALIX exhibit more severe cell division defects in neurons (*12*, *13*). One possibility could relate to forces within the tissue. Adherent cells in culture can migrate away from each other upon mitotic exit and consequently undergo cytofission simply by pulling themselves apart (*14*). Within a tissue such as the brain or the lens, geometric constraints to motility may render cells more susceptible to abscission defects.

Differences in cell type– and tissue-specific cell division pathways could also stem from the need to respond to different regulatory pathways. Abscission is thought to be subject to a checkpoint mechanism, delaying the final step of daughter cell separation until all the chromosomes have left the intercellular bridge. Recent findings suggest that activation of abscission is also susceptible to the state of nuclear envelope reassembly and membrane tension across the ICB (*9*). Could it be that each of these events feeds into a different arm of the abscission pathway? PI3K-C2α has been found to generate PI(3)P in response to shear stress and increased membrane tension (*15*). By using independent pathways or those with overlapping function, abscission may have evolved to respond to a variety of events to ensure its correct timing and the maintenance of genome stability. ■

REFERENCES AND NOTES

1. T. Fujiwara *et al*., *Nature* **437**, 1043 (2005).
2. D. Tiosano *et al*., *PLOS Genet*. **15**, e1008088 (2019).
3. F. Gulluni *et al*., *Science* **374**, eabk0410 (2021).
4. G. R. V. Hammond *et al*., *Science* **337**, 727 (2012).
5. F. Gulluni, M. Martini, E. Hirsch, *J. Cell. Biochem*. **118**, 3561 (2017).
6. E. M. Storck, C. Özbalci, U. S. Eggert, *Annu. Rev. Biochem*. **87**, 839 (2018).
7. S. Frémont, A. Echard, *Curr. Biol*. **28**, R458 (2018).
8. A. Panopoulos *et al*., *Mol. Biol. Cell* **25**, 3105 (2014).
9. M. Vietri, M. Radulovic, H. Stenmark, *Nat. Rev. Mol. Cell Biol*. **21**, 25 (2020).
10. N. Elia, R. Sougrat, T. A. Spurlin, J. H. Hurley, J. Lippincott-Schwartz, *Proc. Natl. Acad. Sci. U.S.A*. **108**, 4846 (2011).
11. N. T. Umbreit *et al*., *Science* **368**, eaba0712 (2020).
12. J. N. Little *et al*., *J. Neurosci*. **41**, 3344 (2021).
13. M. H. Laporte *et al*., *Sci. Rep*. **7**, 44767 (2017).
14. C. L. Dix *et al*., *Dev. Cell* **45**, 132 (2018).
15. A. Boukhalfa *et al*., *Nat. Commun*. **11**, 294 (2020).

ACKNOWLEDGMENTS

The authors are grateful to J. Ashkenas for help with preparing the manuscript. The authors are funded by Natural Science and Engineering Research Council discovery grants (RGPIN-2016-06775 to J.A.B. and RGPIN-2019-05782 to A.W.).

10.1126/science.abm7949

BIOCHEMISTRY

Interactomes in the era of deep learning

Deep learning provides an atomic snapshot of the yeast protein interactome

By **Joana Pereira**[1,2] *and* **Torsten Schwede**[1,2]

Characterizing macromolecular interactions allows better understanding of the inner workings of a cell. However, all methods available today have limitations: Some tell us whether two macromolecules interact, others provide atomic detail about the interaction partners or, at best, the structures of isolated assemblies without cellular context. On page 1340 of this issue, Humphreys *et al*. (*1*) describe a new computational approach, founded on the ongoing deep learning revolution in structure bioinformatics (*2*, *3*), to predict the composition and model the three-dimensional (3D) structure of protein-protein interactions at the same time. They apply their approach to a eukaryotic system—baker's yeast (*Saccharomyces cerevisiae*)—and predict and accurately model more than 1500 protein-protein interactions, 106 of which were not seen before, paving the way to high-throughput, high-accuracy modeling of entire cells.

Determining the 3D structures of macromolecules and their interactions provides important information about macromolecular mechanisms, which can be used, for example, in drug development or exploited in biotechnology. Experimental structural biology methods such as macromolecular crystallography (MX) and high-resolution cryo–electron microscopy (cryo-EM) provide atomic-level detail of macromolecular structures and their assemblies (*4*). Such experiments are laborious and require purification of the macromolecules from their cellular context. Although techniques such as yeast two-hybrid (Y2H) and cross-linking mass spectrometry (XL-MS) allow for large-scale detection of interaction partners, methods such as site-directed mutagenesis or Förster energy resonance transfer (FRET) experiments characterize individual interactions and interfaces. This information can be used to guide the modeling of assemblies by, for example, macromolecular docking in integra-

tive (also known as hybrid) approaches that combine a variety of data types from low-resolution experiments with computational modeling to generate 3D representations of macromolecular assemblies (*5*).

In recent years, structural biology has seen its horizons drastically expanded by computational techniques for structure prediction (see the figure), fueled by the evolution of machine learning algorithms (*6*) as well as a rapid increase of experimental information in open databases such as the Protein Data Bank, which celebrates its 50th anniversary this year. The Critical Assessment of Structure Prediction (CASP) experiment has, since 1994, provided a platform for testing protein structure prediction methods and, during its history, has lived through (and stimulated) several revolutions (*7*). For example, the development of sensitive methods for the detection of remote homologous relationships boosted homology-based modeling, and the use of coevolution information further improved the modeling of proteins without homologs of known structure. This latter method is based on the idea that residues close in space are evolutionarily coupled, and that coupling signals extracted from multiple sequence alignments can be used to predict close contacts in 3D. This not only proved to be useful for the prediction of protein 3D structures, but also readily expanded to the realm of intermolecular interactions, acting as a fast and accurate method to screen and predict protein-interacting pairs in, for example, the proteome of a bacterium (*Escherichia coli*) (*8*, *9*).

This year, a new breakthrough occurred and a new era in structural bioinformatics started (*2*, *3*): DeepMind's AlphaFold2 algorithm (*6*) became the first computational method to reach close-to-experimental atomic accuracy for individual protein structures in CASP (*10*). The basis of this success was the combined use of state-of-the-art deep learning methods with massive amounts of computing power and the vast structure and sequence data accumulated over the past five decades. This promoted a quick and intense activity in the community, with RoseTTAFold rising shortly as a close academic competitor of AlphaFold2 (*11*). Both methods make use of state-of-the-art deep learning approaches but differ in their core architecture. Still, an important part of both is the use of evolutionary couplings from multiple sequence alignments, which are efficiently handled within their underlying networks to predict interatomic contacts and accurately compute 3D coordinates for the atoms in a target protein from its amino acid sequence. Given the previous success of such signals for the identification of protein-protein interactions (*8*, *9*), it makes sense to explore such methods to improve the prediction and modeling of protein-protein interactions and their assemblies at the atomic level.

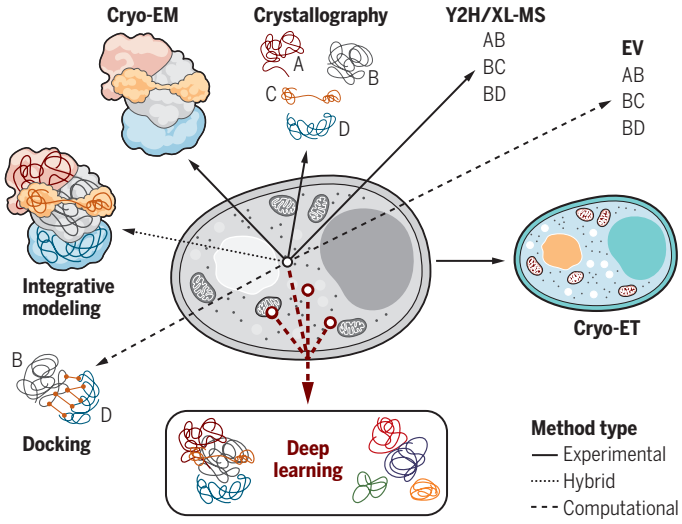
Although most efforts focused on adapting the AlphaFold2 and RoseTTAFold workflows to model protein complexes of known composition and stoichiometry (*12*), Humphreys *et al*. combined the speed of RoseTTAFold's contact prediction algorithm with the high accuracy of AlphaFold2's folding engine and suggest a new method to accurately predict and model at the same time protein pairs across the baker's yeast proteome, the first eukaryote to have its interactome modeled in such a high-throughput fashion. Scanning through ~8 million putative protein pairs, Humphreys *et al*. predicted those more likely to interact on the basis of strong coevolutionary signals and replaced macromolecular docking by protein structure prediction of the joint pair to model the 3D structure of the assembly. The method was able to accurately predict the composition and model the structure of more than 1500 interacting pairs spanning almost all key eukaryotic cell processes, including 106 undescribed assemblies that may highlight previously unknown processes, as well as more than 600 previously known interacting pairs (according to low-resolution biophysical data).

The work by Humphreys *et al*. is a step closer to the modeling of entire cells at high resolution and has already inspired further studies into the interactome of the human mitochondrion (*13*). Currently, methods such as MX and electron microscopy (EM) provide high-resolution atomic representations of macromolecular machines in isolation. Cellular cryo–electron tomography (cryo-ET) has the potential to provide a detailed snapshot of the network of macromolecular interactions, but so far only subnanometer resolution can be obtained (*14*). Artificial intelligence (AI)–based highly accurate proteome-wide modeling of interactions may be able to compensate that resolution gap in a timely manner, especially for more complex organisms. Notwithstanding, methods such as AlphaFold2 and RoseTTAFold provide a static model; incorporating the transient and dynamic nature of macromolecular assemblies will need to be addressed in the future.

This work also highlights the success of open science and community-based method development. AlphaFold2, developed by a commercial company, was made openly available to the entire community, including its source code. This promoted the quick development of different AI-based bioinformatic methods for various goals, such as the Humphreys *et al*. study. AI-based methods are clearly promoting a shift in the way life sciences research will be carried out in the future, where 3D computational models will routinely inspire new experimentally testable hypotheses. ■

Methods to study macromolecular machines

Deep learning–based methods complement experimental techniques, allowing the proteome-wide prediction and modeling of protein assemblies.



Cryo-EM, cryo–electron microscopy; cryo-ET, cryo–electron tomography; EV, evolutionary couplings; Y2H, yeast two-hybrid; XL-MS, cross-linking mass spectrometry.

[1]Biozentrum, University of Basel, Basel, Switzerland. [2]SIB Swiss Institute of Bioinformatics, Biozentrum, University of Basel, Basel, Switzerland. Email: torsten.schwede@unibas.ch

REFERENCES AND NOTES

1. I. R. Humphreys *et al*., *Science* **374**, eabm4805 (2021).
2. A. N. Lupas *et al*., *Biochem. J.* **478**, 1885 (2021).
3. S. M. Kandathil, J. G. Greener, D. T. Jones, *Proteins* **87**, 1179 (2019).
4. T. Nakane *et al*., *Nature* **587**, 152 (2020).
5. A. Sali, *J. Biol. Chem.* **296**, 100743 (2021).
6. J. Jumper *et al*., *Nature* **596**, 583 (2021).
7. A. Kryshtafovych *et al*., *Proteins* **89**, 1607 (2021).
8. Q. Cong *et al*., *Science* **365**, 185 (2019).
9. A. G. Green *et al*., *Nat. Commun.* **12**, 1396 (2021).
10. J. Pereira *et al*., *Proteins* **89**, 1687 (2021).
11. M. Baek *et al*., *Science* **373**, 871 (2021).
12. R. Evans *et al*., bioRxiv 10.1101/2021.10.04.463034 (2021).
13. J. Pei *et al*., bioRxiv 10.1101/2021.09.14.460228 (2021).
14. M. Turk, W. Baumeister, *FEBS Lett.* **594**, 3243 (2020).

ACKNOWLEDGMENTS

We thank G. Studer, J. Durairaj, and X. Robin for helpful discussions.

10.1126/science.abm8295

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

Keep it simple and switch to pure tellurium

Tellurium switch operates memories through crystalline-liquid-crystalline phase changes

By **Raffaella Calarco**[1] *and* **Fabrizio Arciprete**[2]

Researchers, inspired by neurobiological functions and architecture of the human brain, have been developing neuromorphic computing by using artificial neurons and synapses to perform processing and storage in the same physical place. However, no conceptual technology can become reality without hardware. The class of memory devices known as phase change (PC) memories is expected to enable more efficient learning algorithms in neuromorphic computers, owing to their capability to also work as a processing unit (*1*). On page 1390 of this issue, Shen *et al.* (*2*) use pure elemental tellurium to build an electrical switch with a large drive current and rapid switching speed that can be used to efficiently operate PC cells in the cross-point memory architecture (see the figure). By creating the logic gate out of a single element, the design gets around challenges presented by material stoichiometry and selective elemental migration. The single-element design represents a step toward high-density, fast, and nonvolatile PC memories.

[1]Institute for Microelectronics and Microsystems (IMM), National Research Council (CNR), Via del Fosso del Cavaliere 100, 00133 Roma, Italy. [2]Dipartimento di Fisica, Università di Roma "Tor Vergata," Via della Ricerca Scientifica 1, 00133 Rome, Italy.
Email: raffaella.calarco@artov.imm.cnr.it

Chalcogenide alloys are the material of choice because they have the ability to reversibly change their electrical resistivity upon the application of proper electrical stimuli and are considered to be an ideal candidate for making memristive devices to emulate neuronal synaptic behavior. In these alloys—many of them tellurium-based—the resistive switching mechanism is thermally induced through resistive heating (*3*). In some materials, the change in resistivity is tied to a structural phase change from crystalline to amorphous, whereas in others, the change in resistivity occurs without any structural transformation. The former kind of material is good for creating nonvolatile PC memory, with nonvolatile referring to the ability to retain memory without a power supply. The amorphous phase has a high resistivity and represents the "0" logic state. The crystalline phase has a low resistivity and represents the "1" logic state. By contrast, the latter kind of material is good for creating ovonic threshold switches (OTSs), which can connect and interrupt the electrical path on demand. When a threshold voltage is reached, both the PC memory and the OTS would let the current flow pass. However, once the voltage bias is removed, the PC memory would retain its highly conductive crystalline phase, whereas the OTS would return to its initial high-resistivity state.

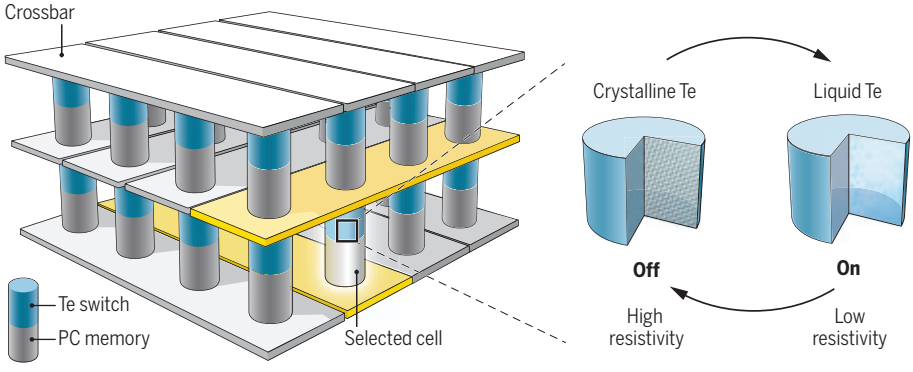
Although this difference in OTS and PC memories has been known at least since the 1960s (*4*, *5*), there has been a renewed research interest because of their potential application in a computer architecture known as the crossbar array, or more specifically, the 3D XPoint (*6*, *7*). In this architecture, a single PC memory cell is selected for operation by applying a proper bias to two crossing electrode bars (see the figure). Neighboring cells, placed along the same bars, are usually biased at half the voltage and act as outlets for unwanted currents to spill over. The switches—that is, the OTSs—connected to each PC memory cell allow for current flow only to the cell where full bias is applied. Thus, a single XPoint element is composed of a PC memory cell and a switch.

To successfully build hardware using this architecture, a switch must possess the following properties: a high on-current density of at least 10 MA/cm^2, enough to drive the selected PC memory cell; a high endurance of up to 10^8 cycles, on par with that of PC cells; and good thermal stability to avoid heat-related failures (*8*). Researchers have built OTSs using tellurium-based complex alloys that can fulfill some of the requirements, but most of them suffer from alloy composition deviation under switching cycles (*8*).

Instead of a classical OTS, Shen *et al.* designed a pure tellurium switch that not only satisfies the aforementioned requirements but also shows a switching speed of around 10 ns, an on-off current ratio of 10^3, and the ability to retain its threshold-switching behavior up to 400°C for 30 min. The authors attribute the impressive thermal stability of the pure tellurium switch to its crystalline phase and not to the amorphous one, as is typical of OTSs. The switching mechanism is based on the presence of a potential energy barrier at the interface between the electrode and the elemental-crystalline semiconducting tellurium. The rectifying characteristics of the energy barrier enable the flow of a negligible leakage current in the off state. Once in the on state and above 2.5 V, the tellurium film is melted and becomes highly conductive. After removing the applied voltage, the recrystallization of liquid tellurium occurs spontaneously on the nanosecond time scale. As evidence for the melting process, the authors observed that the tellurium crystal was oriented along

Switching between crystalline and liquid phases

Shown is a schematic of the crossbar-array memory architecture 3D XPoint. The individual cells are connected by the crossbars (electrodes), with each cell containing a tellurium (Te) switch and a phase change (PC) memory unit.



GRAPHIC: N. CARY/*SCIENCE*

a different direction than before, albeit with the identical trigonal structure. The switching phenomena observed by Shen *et al.* might be peculiar for a material with a low crystallization temperature, fast crystallization speed, and large resistivity difference between the on and off states. At variance with a classic OTS, which is an electric field–driven switch, the tellurium switch gets its resistivity contrast from the metallic-like and semiconducting characteristics of its liquid and crystal phases, respectively.

The past half-century has seen major advances in communications and computing, and this trend is expected to continue at a high pace. Such expansion produces an exponential growth of big data, analyzed through artificial intelligence approaches on conventional computers. The demand for more computational power also comes with proportionally higher operational energetic costs. An increase in computational efficiency, offered by better devices, may be able to curb the carbon footprint that is generated, if the growth in efficiency can outpace demand. Looking at the potential breakthrough for applications, the integration of two single-element devices—that is, a single-element PC memory with a single-element switch—might be of interest. The single-element components will minimize element migration and enhance the robustness of the architecture. Shen *et al.* have proved tellurium to be the material of choice for the switch, whereas other researchers have suggested antimony as a promising candidate for a single-element PC memory (*9*). Furthermore, a scalability down to 60 nm has been demonstrated by the authors for a single XPoint element, with the on-current density increasing quadratically with decreasing device size while maintaining the same switching time, pointing at further scalability. What has been achieved by Shen *et al.* is unprecedented and provides a robust method to realize crystalline elemental switches that bear new perspectives for 3D XPoint architectures. ■

REFERENCES AND NOTES

1. G. W. Burr *et al.*, *IEEE Trans. Electron Dev.* **62**, 3498 (2015).
2. J. Shen *et al.*, *Science* **374**, 1390 (2021).
3. M. Wuttig, N. Yamada, *Nat. Mater.* **6**, 824 (2007).
4. W. R. Noverthover, A. D. Pearson, US Patent 3117013 (1964).
5. S. R. Ovshinsky, US Patent 3271591 (1966).
6. Intel, Intel Optane Memory H10 with Solid State Storage (2018); www.intel.com/content/www/us/en/products/m.
7. D. Kau *et al.*, in *Proceedings of the 2009 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2009), pp. 1–4; https://ieeexplore.ieee.org/document/5424263.
8. M. Zhu, K. Ren, Z. Song, *MRS Bull.* **44**, 715 (2019).
9. M. Salinga *et al.*, *Nat. Mater.* **17**, 681 (2018).

10.1126/science.abm7316

QUANTUM CHEMISTRY

Artificial intelligence "sees" split electrons

Machine-learning creates a density functional that accounts for fractional charge and spin

By **John P. Perdew**

Chemical bonds between atoms are stabilized by the exchange-correlation (xc) energy, a quantum-mechanical effect in which "social distancing" by electrons lowers their electrostatic repulsion energy. Kohn-Sham density functional theory (DFT) (*1*) states that the electron density determines this xc energy, but the density functional must be approximated. This is usually done by satisfying exact constraints of the exact functional (making the approximation predictive), by fitting to data (making it interpolative), or both. Two exact constraints—the ensemble-based piecewise linear variation of the total energy with respect to fractional electron number (*2*) and fractional electron *z*-component of spin (*3*)—require hard-to-control nonlocality. On page 1385 of this issue, Kirkpatrick *et al.* (*4*) have taken a big step toward more accurate predictions for chemistry through the machine learning of molecular data plus the fractional charge and spin constraints, expressed as data that a machine can learn.

Efficient computer prediction of what molecules and materials can exist, and with what properties, can be enabled with DFT, through the self-consistent solution of effective one-electron time-independent Schrödinger equations. However, in cases in which exact constraints are important for the proper sharing of the electrons and their spins among the atoms, their neglect can lead to some of the worst qualitative failures of standard density functionals (*5*). This problem can be illustrated by the case of a sodium (Na) atom well separated from a chlorine (Cl) atom (see the figure) (*6*). The exact energy minimizes at zero electron transfer between neutral atoms because the exact energy contribution from each atom is a linkage of straight-line segments that connect with sharp corners at integer electron numbers. However, simple density functionals that round off these corners minimize with nonzero electron transfer. A Cl atom that should have 17 electrons might wrongly be assigned on average 17.4 electrons because its electron number fluctuates between 17 with 60% probability and 18 with 40% probability.

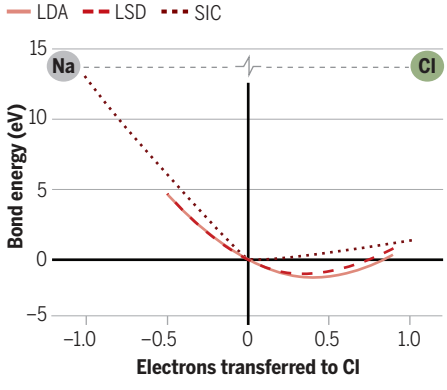
Human beings have developed approximations to the exact density functional by positing equations and determining how well they agree with exact solutions and experimental data. Artificial intelligence (machine learning with the use of deep neural networks) could learn it by recognizing features and patterns in the density after training on similar data (*7*). This process is analogous to the way that both human beings and machines can recognize faces.

The DeepMind 2021 (DM21) functional developed by Kirkpatrick *et al.* is technically a local hybrid (nonlocal) functional. It has features that can be constructed from ingredients such as the local spin densities, their gradients in space, the noninteracting kinetic energy densities, and the exact unscreened and screened exchange energy densities. DM21, which can be used with standard Kohn-Sham electronic structure

Departments of Physics and Chemistry, Temple University, Philadelphia, PA 19122. Email: perdew@temple.edu

Not staying neutral

In the limit of an infinite bond length, sodium chloride (NaCl) should be neutral atoms, not ions. The total energy should follow the straight-line segments of the exact Kohn-Sham density functional theory, as imitated by the Perdew-Zunger self-interaction correction (SIC) functional. However, a spurious transfer of 0.4 electrons to Cl is predicted by the spin-unpolarized local density approximation (LDA) and by the spin-polarized local spin density (LSD) approximations.



codes, was fit to some data sets for main-group molecules and then tested on others. On a large and diverse suite of 55 data sets for different thermochemical molecular properties (*8*), the weighted absolute error of DM21 is 4 kcal/mol.

This very small error compared with that for most functionals results from the large number of well-chosen ingredients and from the fitted molecular data. The error is essentially the same whether the fractional charge and spin data are included (DM21) or not (DM21m). However, inclusion of those data improved the performance of DM21 for charge-transfer and strong-correlation problems not included in the test suite, such as binding energy curves for H_2^+ and H_2, charge transfer in a DNA base pair, and a compressed hydrogen chain. DM21 impressively captures strong correlation without symmetry breaking.

A comparison can be made to the strongly constrained and appropriately normed (SCAN) functional (*9*), which was created by positing equations that satisfy 17 exact constraints but not the fractional charge and spin constraints. Unlike DM21, SCAN is not fitted to any bonded system and does not use expensive exact exchange. On the same suite of 55 test sets, without any exact exchange ingredient but with a standard dispersion correction, SCAN makes an 8 kcal/mol error. However, when SCAN is density-corrected (evaluated on the Hartree-Fock density instead of its own self-consistent density), that error is reduced to 6 kcal/mol (*10*). This density correction also eliminates charge-transfer errors of SCAN.

For the chemistry of main-group elements, DM21 is very good, although it may be less accurate for transition metal chemistry, a more challenging problem to which it was not fitted. Solids and liquids could also be described unsatisfactorily for several reasons: because they are not included in the DM21 fitting sets (although a variant of DM21 was also fit to the electron gas of uniform density) and because atoms and small molecules can be well described by using full exact exchange at long range, whereas extended systems cannot.

The fractional spin constraint in DM21, although exact in principle, might suppress spin symmetry breaking (as it does for the binding energy curve of H_2) that can be both real and revealing in extended systems. As noted by Anderson (*11*), time-dependent fluctuations of the electron density or spin density persist over time scales that can grow large as the size of the system grows to that of a crystal lattice. These fluctuations break the symmetries that are found in an exact symmetry-preserving ground-state wave function, which may only predict the densities or spin densities averaged over an infinite time interval (*12*). For example, the net local spin density of an antiferromagnetic solid such as nickel oxide (NiO) can be zero over an infinite time interval but can display localized spin moments that alternate in direction from one transition metal atom to another in an ordered and fixed array and persist for years.

These local spin moments can be predicted by use of standard spin density functionals, including SCAN. Supercell calculations (with larger than minimal unit cells) (*13*) find that those moments (and the electrical insulation they produce) persist even above the antiferromagnetic ordering temperature. Because it can break symmetries, SCAN accounts for and reveals the strong correlations that occur, for example, in the cuprate high-temperature superconducting materials, explaining both the insulator-to-metal transition that occurs with doping and the spin-and-charge stripes (*14*). Symmetry-breaking SCAN also works well for the hydrogen chain (*13*).

The importance of DM21 developed by Kirkpatrick *et al.* is not that it yields the ultimate density functional but that an artificial intelligence (AI) approach addressed the fractional electron and spin problem that has resisted a direct analytical solution to creating the functional. Their work and that of (*15*) suggest that more predictively accurate density functionals can be designed by combining constraint satisfaction with AI fitting to large and diverse data sets. ■

REFERENCES AND NOTES

1. W. Kohn, L. J. Sham, *Phys. Rev.* **140** (4A), A1133 (1965).
2. J. P. Perdew, R. G. Parr, M. Levy, J. L. Balduz Jr., *Phys. Rev. Lett.* **49**, 1691 (1982).
3. A. J. Cohen, P. Mori-Sánchez, W. Yang, *J. Chem. Phys.* **129**, 121104 (2008).
4. J. Kirkpatrick *et al.*, *Science* **374**, 1385 (2021).
5. A. J. Cohen, P. Mori-Sánchez, W. Yang, *Science* **321**, 792 (2008).
6. J. P. Perdew, in *Density Functional Methods in Physics*, vol. 123, R. M. Dreizler, J. da Providencia, Eds. (Plenum Press, 1985), p. 265.
7. J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, K. Burke, *Phys. Rev. Lett.* **108**, 253002 (2012).
8. L. Goerigk *et al.*, *Phys. Chem. Chem. Phys.* **19**, 32184 (2017).
9. J. Sun, A. Ruzsinszky, J. P. Perdew, *Phys. Rev. Lett.* **115**, 036402 (2015).
10. G. Santra, J. M. L. Martin, *Chem. Theory Comput.* **17**, 1368 (2021).
11. P. W. Anderson, *Science* **177**, 393 (1972).
12. J. P. Perdew, A. Ruzsinszky, J. Sun, N. K. Nepal, A. D. Kaplan, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2017850118 (2021).
13. Y. Zhang *et al.*, *Phys. Rev. B* **102**, 045112 (2020).
14. Y. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 68 (2020).
15. S. Dick, M. Fernandez-Serra, *Phys. Rev. B* **104**, L161109 (2021).

ACKNOWLEDGMENTS

This work was supported by NSF DMR-1930528 and US Department of Energy DE-SC0018331.

10.1126/science.abm2445

GRAPHIC: ADAPTED FROM (6) BY K. FRANKLIN/*SCIENCE*

PLANETARY SCIENCE

Martian water escape and internal waves

Lower atmospheric processes are vital for assessing water loss from Mars

By **Erdal Yiğit**[1,2]

Evidence has accumulated in recent years that suggests that Mars used to have more habitable conditions (*1*). Determining the processes that led to its current cold and dry state are important for understanding habitability more generally. Observations and modeling efforts show that atmospheric escape has adversely affected Mars' habitability primarily by irreversibly diminishing the atmospheric water reservoir through loss of atomic oxygen and hydrogen into space (*2*). Understanding this requires connecting dust storms, atmospheric waves, and atmospheric water.

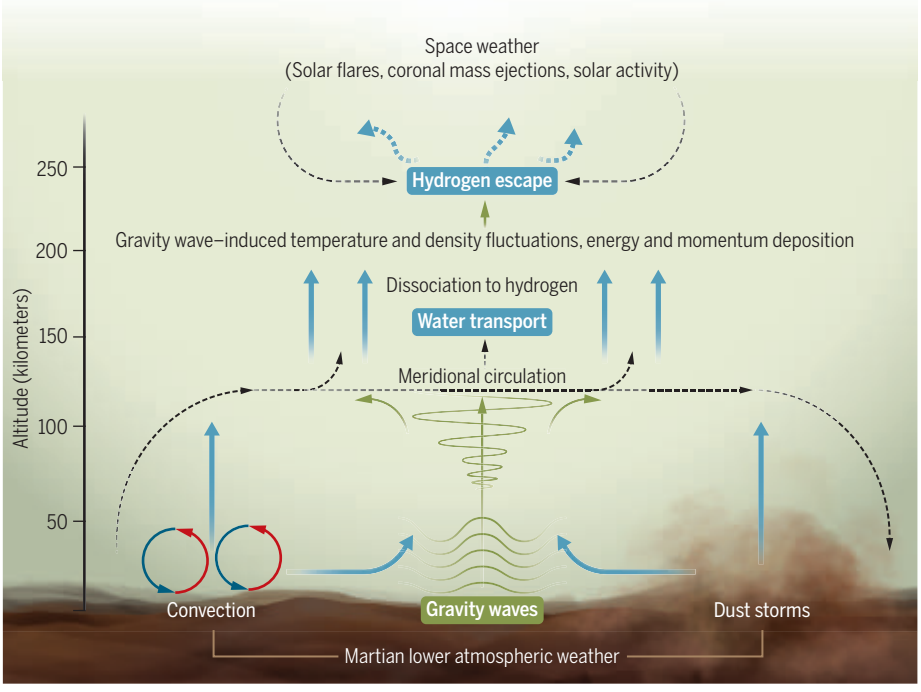
Geological considerations require that water was more abundant on Mars early in its history, along with Mars being wetter and warmer (*1*). The thermal escape of atomic hydrogen (H) to space is the primary mechanism for long-term loss of water (see the figure). Molecular hydrogen (H_2) forms in the middle atmosphere by photochemical processes (*3*) and is transported upward, where it is decomposed to H, which can more easily escape to space (*4*). Recent general circulation modeling studies (*5*), as well as observations by the MAVEN (Mars Atmosphere and Volatile Evolution) spacecraft (*6*) and the ExoMars Trace Gas Orbiter (TGO) (*7*), demonstrated that water can be directly transported during the perihelion season to the thermosphere, where it is dissociated. However, the role of the lower atmospheric weather and the associated wave-induced vertical coupling processes are insufficiently explored.

Processes that led to the loss of water on Mars are complex and require a whole-

[1]Department of Physics and Astronomy, George Mason University, Fairfax, VA, USA. [2]Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen, Germany. Email: eyigit@gmu.edu

Getting water off of Mars

Atmospheric coupling processes that play a major role in the direct transport of water to the thermosphere are shown. Atmospheric gravity (buoyancy) waves may have a key role in strengthening the meridional circulation responsible for the upward water transport and in enhancing hydrogen escape to space.



atmosphere perspective. Lower atmospheric dust storms are a key component of the martian weather, with wide-reaching implications for the martian whole-atmosphere system. They often occur during perihelion when Mars is closest to the Sun, which happens during martian southern summer. Major dust storms are global phenomena and can last for several months (*8*). Atmospheric (gravity) waves, generated primarily in the lower atmosphere, are ubiquitous features of all planetary atmospheres (*9*) and can directly propagate to the thermosphere, coupling the different atmospheric layers and producing substantial density fluctuations (*10*).

NASA's MAVEN mission has shown that observable seasonal changes exist in the escape of hydrogen, with the highest escape flux occurring during martian perihelion (*11*). Also, lower atmospheric deep convection can facilitate increased transport of water vapor to the middle atmosphere during martian dust storms, which can ultimately enhance hydrogen loss to space (*12*). Martian global-scale climate models can, within limitations, help diagnose the processes that influence the transport of water. Whereas some global models show that water can penetrate to higher altitudes only during the perihelion season, when the meridional circulation cell is sufficiently strong (*5*), other models underestimate the amount of water vapor transported to high altitudes (*13*).

Space weather in the form of solar flares and coronal mass ejections can lead to increased hydrogen escape from the martian upper atmosphere and should be included in the extrapolation of water loss (*14*). However, these processes occur on relatively shorter time scales (hours to days) compared to dust storm effects, which can last for several months and can substantially alter the martian whole atmosphere. Because Mars lacks an intrinsic global magnetic field, the role of space weather is important in the present time and could also have played a crucial role in atmospheric loss over the history of Mars (*4*). All these results indicate strongly that the loss of water to space in the form of hydrogen escape cannot be fully understood without considering lower atmospheric processes. However, all these studies miss an important mechanism of vertical coupling, i.e., upward-propagating lower atmospheric waves, especially gravity (buoyancy) waves. During global dust storms, gravity waves encounter favorable propagation conditions from the lower atmosphere to the thermosphere (*15*), where they can potentially control Jeans escape of hydrogen to space through wave-induced fluctuations of temperature and density. Gravity waves are also known to drive a mesospheric meridional circulation on Earth and Mars; thus, the waves are likely to contribute to the strengthening of the dust-storm–time middle atmospheric meridional circulation that enhances water transport to the thermosphere.

> **"...hydrogen escape cannot be fully understood without considering lower atmospheric processes."**

To better understand the global transport of water and loss to space, coordinated observational and modeling efforts are required to characterize the large-scale and small-scale variability and waves in the martian whole atmosphere, especially during transient events such as dust storms. Future three-dimensional climate modeling studies, accounting for the generation, propagation, and dissipation of a broad spectrum of atmospheric waves, will be necessary to elucidate the role of the internal gravity waves in the (direct) transport of water to thermospheric and exospheric altitudes and its loss to space. These whole-atmosphere modeling studies would have to simultaneously account for the dynamical, thermal, and compositional effects produced by gravity waves and their modulation by larger-scale waves, such as tides.

Whereas wave propagation can produce local density and temperature fluctuations, gravity-wave momentum and energy deposition can substantially change the background atmospheric temperature, which is a key parameter in thermal escape. Future coincident coordinated observations are required to constrain models and wave activity, and to help characterize the whole-atmosphere distribution of water and its constituents. Using several current observational capabilities such as ExoMars TGO, MAVEN, and Mars Reconnaissance Orbiter in complementary ways may help accomplish this lofty goal. ■

REFERENCES AND NOTES

1. B. M. Jakosky, R. J. Phillips, *Nature* **412**, 237 (2001).
2. M. S. Chaffin, J. Deighan, N. M. Schneider, A. I. F. Stewart, *Nat. Geosci.* **10**, 174 (2017).
3. V. A. Krasnopolsky, *Icarus* **321**, 62 (2019).
4. B. Jakosky *et al.*, *Icarus* **315**, 146 (2018).
5. D. S. Shaposhnikov, A. S. Medvedev, A. V. Rodin, P. Hartogh, *Geophys. Res. Lett.* **46**, 4161 (2019).
6. S. W. Stone *et al.*, *Science* **370**, 824 (2020).
7. A. A. Fedorova *et al.*, *Science* **367**, 297 (2020).
8. S. K. Jain *et al.*, *Geophys. Res. Lett.* **47**, e2019GL085302 (2020).
9. A. S. Medvedev, E. Yiğit, *Atmosphere* **10**, 531 (2019).
10. E. Yiğit *et al.*, *Geophys. Res. Lett.* **42**, 8993 (2015).
11. M. S. Chaffin *et al.*, *Geophys. Res. Lett.* **41**, 314 (2014).
12. N. G. Heavens *et al.*, *Nat. Astron.* **2**, 126 (2018).
13. J. Y. Chaufray *et al.*, *Icarus* **353**, 113498 (2021).
14. M. Mayyasi *et al.*, *Geophys. Res. Lett.* **45**, 8844 (2018).
15. T. Kuroda, A. S. Medvedev, E. Yiğit, *J. Geophys. Res. Planets* **125**, e2020JE006556 (2020).

10.1126/science.abg5893

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

CANCER

Oncolytic virotherapy as immunotherapy

Recognizing immune responses to oncolytic virotherapy opens the way for new combinations

By **Alan Melcher**[1], **Kevin Harrington**[1], **Richard Vile**[2]

Oncolytic virus (OV) therapy is now something of a misnomer. The initial paradigm was that naturally occurring or genetically engineered viruses would preferentially infect, replicate in, and lyse cancer cells relative to normal cells, leading to selective direct cytotoxicity. The immune system used to be considered detrimental, given its potential to recognize and respond to the virus, prevent its replication, and limit viral spread and tumor cell killing (*1*). This view has now been largely replaced by a model in which infection of the tumor by the virus makes it more visible to the immune system for recognition and attack. Tumor cell killing by the virus and/or immune effector cells effectively generates an in situ vaccine, with tumor antigens released into a microenvironment modified by viral infection to reverse tumor-induced immune suppression.

By 2020, there were ~100 reported clinical trials using OVs, treating more than 3000 cancer patients. Approvals were granted in different countries for picorna-, adeno-, and herpes viruses, although only one agent was approved by the US Food and Drug Administration (FDA). The growing recognition of the immune basis of oncolytic virotherapy coincided with the success of immunotherapy in the clinic, particularly immune checkpoint inhibitors (ICIs), which block negative regulatory elements that control immune tolerance, specifically the programmed cell death protein 1 (PD-1) axis and cytotoxic T lymphocyte–associated antigen 4 (CTLA-4). These drugs revealed that the immune system can and does respond to cancer, but tumors evolve to escape immune detection through immunoediting (*2*). However, if the immune system can be reenergized and reactivated, immune control of the tumor can be reestablished.

The number of OVs being tested in preclinical models, with a particular focus on their immunogenicity, has since increased. However, in some ways the field has been struggling with an embarrassment of (potential) riches. It has been unclear how to prioritize the most promising preclinical agents to take forward to clinical testing. Additionally, how to accurately measure the immunogenicity of tumor cell death after OV infection—particularly for viruses that do not readily infect mouse cells (e.g., coxsackie virus, adenovirus, and measles virus) and so cannot be tested reliably in preclinical models—has been a challenge. The variety of OVs covers DNA, single-stranded RNA (ssRNA), and double-stranded RNA (dsRNA) viruses, and the immunobiological consequences after infection of tumor and/or other cells varies, including the signaling pathways activated, such as retinoic acid–inducible gene I (RIG-I)–like receptors for viral RNA and the cyclic GMP–AMP synthase (cGAS)–stimulator of interferon genes (STING) pathway for cytosolic DNA. The degree to which these pathways remain intact within cancer cells, and the relative contribution of the response to infection of nonmalignant cells in the tumor microenvironment (such as cancer-associated fibroblasts, endothelial cells, macrophages, and other immune cell subtypes), adds further complexity to understanding how tumor immunogenicity is generated by OVs.

Much current work is focused on the ability to genetically modify OVs to encode transgenes, such as cytokines, ICIs, tumor-associated antigens, bispecific T cell engagers, and microRNAs, as payloads to boost the antitumor immune response. However, the choice of these encoded transgenes has often been based on limited mechanistic understanding. For example, granulocyte-macrophage colony-stimulating factor (GM-CSF) is expressed in many OVs being clinically tested, even though the evidence supporting the benefit of virally encoded GM-CSF is limited. A further challenge is the inevitable antiviral (alongside antitumor) immune response activated by treatment. Even if viruses to which patients have not been previously exposed are used, neutralizing antiviral antibodies (NAbs) will increase with treatment, potentially compromising repeat administrations, particularly if the route of administration is systemic. Even though intravenously delivered OVs (e.g., adenovirus and vaccinia virus) can infect tumors in patients (*3*, *4*), despite the presence of NAbs in the case of reovirus (*5*), concerns about sufficient delivery to tumors have led to the administration of many OVs—including talimogene laherparepvec (T-Vec), a clinically approved OV for the treatment of advanced melanoma—being restricted to direct intratumoral injection.

The rapid expansion of ICIs into the clinic has affected the development of OVs and other immunotherapeutics in unexpected ways. Usually, advances in the clinic lag years behind the laboratory discoveries that underpin them, but this has reversed. To improve on the benefit from ICIs, which remains limited to an overall response rate of only 10 to 15% of all cancer patients (although this is higher in certain tumor types and some patient groups), combination strategies have been empirically tested without sufficient understanding of the immunobiology of why they might (or might not) work. Trials of ICI combinations with chemotherapy or radiotherapy are relatively easy to design because the ICI is added to current standard-of-care regimens, and such trials have met with considerable success (*6*, *7*). However, immunotherapy-OV combinations have so far been disappointing. In almost all such studies, the collection and analysis of patient samples has been inadequate and/or unreported, so it remains generally unknown why combinations, which showed promise in preclinical models, have failed to translate into patient benefit. This lack of correlative, translational data particularly applies to negative trials (which are at least as important, in terms of mechanistic understanding, as positive studies), leaving the immune changes that occur in patients after treatment unknown.

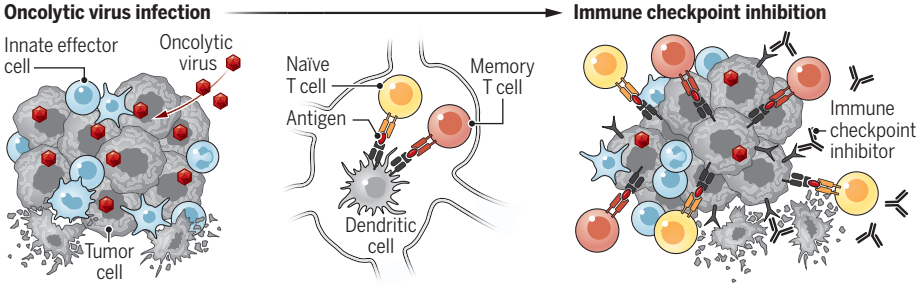
There have been few large, randomized controlled trials including OVs so far, and those that have been reported have been discouraging. A study of vocimagene amiretrorepvec (TOCA 511), a replicating retrovirus encoding cytosine deaminase, which converts the prodrug 5-fluorocytosine into 5-fluorouracil, failed in a phase 2/3 study of direct injection into the surgical cavity on first or second resection of high-grade glioma, randomized against standard of care (*8*). A second phase 2b study involving pexastimogene devacirepvec (Pexa-Vec), a vaccinia virus encoding GM-CSF, was also unsuccessful when tested in sorafenib-refractory hepatocellular carcinoma (HCC) (*9*). Both studies have plausible explanations for their lack of success. In the TOCA 511 study, patients had fewer treatment cycles than in earlier phase trials, and for Pexa-Vec, the poor responses of advanced HCC are notoriously difficult to reverse.

[1]The Institute of Cancer Research, London, UK. [2]Mayo Clinic, Rochester, MN, USA. Email: alan.melcher@icr.ac.uk; kevin.harrington@icr.ac.uk; vile.richard@mayo.edu

However, the trial that best illustrates both the promise and challenges of immunovirotherapy is the testing of T-Vec with the anti-PD1 ICI pembrolizumab (MASTERKEY-265, NCT02263508). T-Vec is a genetically modified type I herpes simplex OV with viral *ICP34.5* and *ICP47* genes deleted, to enhance tumor tropism and reduce neurovirulence, and encoding GM-CSF. Despite T-Vec being clinically approved, the phase 3 study in advanced melanoma on which its approval was based pre-dated widespread use of ICIs (and BRAF inhibitors) in this disease and had subcutaneous GM-CSF as its control arm. Nevertheless, a randomized study of T-Vec with ipilimumab (an anti-CTLA4 ICI) was promising (*10*), and clinical analysis of the phase 1b lead-in stage of the randomized pembrolizumab ±T-Vec study was also encouraging, with impressive clinical responses (*11*).

Oncolytic viruses are multifaceted tumor killers

Oncolytic viruses delivered to the tumor not only kill tumor cells directly, but also potentiate antitumor immune responses by releasing antigens and activating inflammatory responses. This may allow immunologically "cold" tumors to become "hot."



Oncolytic viral infection kills tumor cells, releasing antigens and recruiting innate effector cells.

At the **lymph node**, viral, self, and tumor antigens are presented to naïve and memory T cells, which, primed and expanded, travel to the tumor.

Activated T cells infiltrate the tumor, where they recognize antigens and kill tumor cells. This antitumor response can be boosted by immune checkpoint inhibitors.

Alongside clinical data from this early stage of MASTERKEY-265, translational readouts from 21 patients addressed the hypothesis suggested by preclinical data (of this and other ICI-OV combinations) that the virus would turn the immunologically "cold" tumor "hot" and thus prime for ICI efficacy. "Heat" illustrates the variable level of immune activation within a tumor and is described using parameters such as PD-L1 expression, T cell infiltration, and interferon gene signature. However, heat can denote different things, and there remains huge uncertainty around its meaning and importance. Heat could be of the wrong sort, reflecting immune activation directed against the virus rather than immunogenically weaker tumor antigens. Heat also needs to be generated at the right time as well as in the right place. Measuring heat and how its different immunological forms influence antiviral and antitumor immune effectors is a critical area of research that offers the potential to unleash the real value of OV as an immunotherapy (see the figure). However, as a biomarker, immunological heat is currently of limited practical use, with PD-L1 expression being the only manifestation routinely tested and used for clinical decision-making.

In the early stages of the MASTERKEY-265 trial, detailed patient sample analysis showed that OV injection turned cold tumors hotter (*11*). Unfortunately, however, the later randomized phase 3 trial was stopped because of clinical futility. Given the encouraging early clinical data, this setback requires explanation. It may, again, be down to the clinical context, in that a larger group of melanoma patients, with limited metastatic disease, did too well with single-agent pembrolizumab (with an expected 5-year survival rate approaching 50%) for the addition of T-Vec to make a significant difference. More provocatively, there were subtle differences in the protocols between the phase 1b and 3 stages of the trial, with T-Vec injections starting 5 weeks earlier in phase 1b, before pembrolizumab, whereas the two treatments began at the same time in phase 3. Hence, in the phase 3 component, the virus may not have had time to heat up the tumor before ICI treatment began. Whatever the reason, this study illustrates, for OVs in particular and immunotherapy more generally, how informative immune analysis can be and that the testing of a sufficient number of patients within a consistent, clinically appropriate trial setting is crucial.

Looking at immunotherapy more broadly, the mechanistic basis of successful treatments is incompletely understood and targets only a small fraction of the pathways that are increasingly studied. For example, stimulatory antibodies targeting positive immune checkpoints, such as CD40 and 4-1BB, have yet to show substantial clinical activity despite preclinical promise, and this is also true for small molecules targeting other inhibitory immune targets, such as indoleamine 2,3-dioxygenase (IDO) (*12*). Hence, OVs are not alone in struggling to make the transition from early- to late-stage trial success and clinical adoption, although encouraging single-agent phase 1 studies of OVs have recently been reported (*13*, *14*). These trials involved clinically challenging methods of delivery, namely of a herpes simplex OV through intratumoral catheters (*13*) and neural stem cell delivery of an adenovirus injected into the surgical resection cavity (*14*), both in high-grade glioma. Limited, correlative immune data were included in these studies, but the challenge of therapy and immune analyses in larger trials (in terms of practicality and outcome) remains.

The progression of a drug from promising early-phase clinical and translational data to more widespread application in patients is always challenging. For immunotherapy, this has been achieved for ICI and chimeric antigen receptor (CAR)–T cells but not yet for OVs, where the need for combination strategies to improve outcomes brings complications. What is required is an imaginative and collaborative approach to early-then-late–phase trials, perhaps focusing on less immunotherapy-responsive diseases than melanoma, such as microsatellite stable colorectal and ovarian cancer, or brain tumors. Studies should maximize translational immune analysis to improve scientific insights, which can then support larger trials to determine true efficacy as well as reverse translation back to the laboratory for truly iterative basic and clinical research. Multiarm trials, which combine experimental arms in the same study, may be one way to cover more ground faster, to fully realize the potential of OVs as powerful and adaptable candidates to turn cold tumors hot, thus improving the benefits of current immunotherapy for more cancer patients. ■

REFERENCES AND NOTES

1. A. C. Filley, M. Dey, *Front. Oncol.* **7**, 106 (2017).
2. G. P. Dunn *et al.*, *Annu. Rev. Immunol.* **22**, 329 (2004).
3. R. Garcia-Carbonero *et al.*, *J. Immunother. Cancer* **5**, 71 (2017).
4. C. J. Breitbach *et al.*, *Nature* **477**, 99 (2011).
5. A. Samson *et al.*, *Sci. Transl. Med.* **10**, 422 (2018).
6. L. Paz-Ares *et al.* *N. Engl. J. Med.* **379**, 2040 (2018).
7. S. J. Antonia *et al.*, *N. Engl. J. Med.* **379**, 2342 (2018).
8. T. F. Cloughesy *et al.*, *JAMA Oncol.* **6**, 1939 (2020).
9. M. Moehler *et al.*, *OncoImmunology* **8**, 1615817 (2019).
10. J. Chesney *et al.*, *J. Clin. Oncol.* **36**, 1658 (2018).
11. A. Ribas *et al.*, *Cell* **170**, 1109 (2017).
12. G. V. Long *et al.*, *Lancet Oncol.* **20**, 1083 (2019).
13. G. K. Friedman *et al.*, *N. Engl. J. Med.* **384**, 1613 (2021).
14. J. Fares *et al.*, *Lancet Oncol.* **22**, 1103 (2021).

10.1126/science.abk3436

GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

POLICY FORUM

ARTIFICIAL INTELLIGENCE

Filling gaps in trustworthy development of AI

Incident sharing, auditing, and other concrete mechanisms could help verify the trustworthiness of actors

By **Shahar Avin[1], Haydn Belfield[1,2], Miles Brundage[3], Gretchen Krueger[3], Jasmine Wang[4], Adrian Weller[2,5,6], Markus Anderljung[7], Igor Krawczuk[8], David Krueger[5,9], Jonathan Lebensold[4,9], Tegan Maharaj[9,10], Noa Zilberman[11]**

The range of application of artificial intelligence (AI) is vast, as is the potential for harm. Growing awareness of potential risks from AI systems has spurred action to address those risks while eroding confidence in AI systems and the organizations that develop them. A 2019 study (*1*) found more than 80 organizations that have published and adopted "AI ethics principles," and more have joined since. But the principles often leave a gap between the "what" and the "how" of trustworthy AI development. Such gaps have enabled questionable or ethically dubious behavior, which casts doubts on the trustworthiness of specific organizations, and the field more broadly. There is thus an urgent need for concrete methods that both enable AI developers to prevent harm and allow them to demonstrate their trustworthiness through verifiable behavior. Below, we explore mechanisms [drawn from (*2*)] for creating an ecosystem where AI developers can earn trust—if they are trustworthy (see the figure). Better assessment of developer trustworthiness could inform user choice, employee actions, investment decisions, legal recourse, and emerging governance regimes.

Common themes in statements of AI ethics principles include (i) assurance of safety and security of AI systems throughout their life cycles, especially in safety-critical domains; (ii) prevention of misuse; (iii) protection of user privacy and source data; (iv) ensuring that systems are fair and minimize bias, especially when such biases amplify existing discrimination and inequality; (v) ensuring the decisions made by AI systems, as well as any failures, are interpretable, explainable, reproducible and allow challenge or remedy; and (vi) identifying individuals or institutions who can be held accountable for the behaviors of AI systems. These principles address concerns that include accidents in robotic systems; erroneous judgments from AI systems used by physicians or in court; misuse of AI in surveillance, manipulation, or warfare; and risks to privacy and concerns about systemic bias (*3*).

In the study of trust in technology, a common approach differentiates trust in people (individuals and institutions) and trust in technology artifacts (*4*). Whereas trust in artifacts mainly relies on competence and reliability, trust in people also relies on motives and integrity. Trust can be earned by providing reliable evidence that AI systems, and the processes used to develop and deploy them, address potential harms. This evidence carries further weight in an ecosystem where principles for preventing harms are well established, and where failure to adhere to principles carries meaningful consequences; a failure to establish an ecosystem that links trust to trustworthiness could spread into a general loss of trust in AI systems, compounding the harm from specific systems with the harm of foregone benefits. Concerns regarding motives, although crucial to some aspects of trust, are mostly outside the scope of the proposed mechanisms.

Trustworthy AI development presents considerable challenges. Technical standards that assure that an AI system adheres to the ethical principles mentioned are often lacking. Thus, experts need to evaluate specific AI systems in the contexts where they are developed and deployed. Experts may not be incentivized to address potential harms from their own organizations, and cooperation across organizations can raise antitrust law concerns. The proposed mechanisms we describe help address these challenges by sharing relevant information and incentivizing expert evaluation, which together can inform public assessments of AI developers' trustworthiness.

Beyond AI development, we recognize that the broader sociotechnical context, including but not limited to AI procurement, deployment, social context, and use, will require additional engagement and measures. Although our mechanisms focus on AI systems at or close to deployment, where requirements and context are clearer, they also extend to earlier stages of development. Separately, we note the need for AI developers to earn trust by consistently displaying trustworthy behavior more generally, including healthy, equitable, and diverse work environments, clear antiretaliation policies to protect whistleblowers, and broad environmental, ethical, and social responsibilities.

MECHANISMS

Red team exercises

To address concerns of misuse and new vulnerabilities, a growing number of AI developers are turning to "red teams": professionals who consider a system from the perspective of an adversary to identify exploitable vulnerabilities, which can then be mitigated. To date, AI red teams exist mostly within large industry and government labs, though experts also engage in "red team" activity in academia and through consultancy. AI red-teaming could form a natural extension of the cybersecurity red team community, though the data-driven and increasingly general nature of AI systems requires new domains of expertise.

We see space for the formation of a community of AI red team experts that shares experience across organizations and domains. Such exchange is not currently commonplace, though there has been a growing trend to publish threat modeling of AI systems (*5*). For example, there are public technical discussions on the feasibility of criminals using adversarial attacks on machine learning (ML) models or on the possibility of misusing large-scale language models for online disinformation. Red-teaming could be carried out by an independent third party to address antitrust concerns (*6*).

[1]Centre for the Study of Existential Risk, University of Cambridge, Cambridge, UK. [2]Leverhulme Centre for the Future of Intelligence, University of Cambridge, Cambridge, UK. [3]OpenAI, San Francisco, CA, USA. [4]School of Computer Science, McGill University, Montreal, QC, Canada. [5]Department of Engineering, University of Cambridge, Cambridge, UK. [6]The Alan Turing Institute, London, UK. [7]Centre for the Governance of AI, Oxford, UK. [8]École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland. [9]Mila, Montreal, QC, Canada. [10]Faculty of Information, University of Toronto, Toronto, ON, Canada. [11]Department of Engineering Science, University of Oxford, Oxford, UK. Email: sa478@cam.ac.uk

Audit trails

External audits, whether mandated by regulation or undertaken voluntarily, would form an important piece of the AI trustworthiness ecosystem (see below). To enable audits, AI developers would need to adopt best practices for documenting their development process and systems' makeup and activities. Clear standards for retaining information during development and operations exist in other domains (*7*). Standards and logging mechanisms have yet to be created to cover the range of AI applications, though there are some ongoing domain-specific efforts, for example, for automated vehicles (*8*). Although industry standards can raise antitrust concerns, lessons from other safety-critical industries suggest that such concerns can be addressed if standards are mandated by governments; if they are voluntary, developed in an open and participatory manner; or if they are accessible on fair, reasonable, and nondiscriminatory terms (*6*).

Early progress can be seen in ethics frameworks that formalize questions to ask during the development process (e.g., Rolls-Royce Aletheia Framework or the Machine Intelligence Garage Ethics Framework), in emerging guidelines for documenting certain features of AI models [e.g., Annotation and Benchmarking on Understanding and Transparency of Machine Learning Lifecycles (ABOUT ML) or Model Cards], and in proposals for continuous monitoring and logging. Developers are expected, for example, to record the provenance of all data that are used to train models and to record outcomes of benefit and risk assessments conducted before deployment. Further progress requires collective effort to develop widely accessible and free standards for audit trails.

Interpretability and explainability

Assuring the safety, accountability, and fairness of AI systems is often challenged by their "black box" nature. Many researchers aim to address this challenge, either by restricting AI systems to human-readable, rules-based behaviors or by explaining systems' outputs, for example, by highlighting salient inputs. Research in this area has highlighted the importance of specific principles: (i) methods should provide sufficient insight for the end-user to understand how a model produced an output; (ii) the interpretable explanation should be faithful to the model, accurately reflecting its underlying behavior; and (iii) techniques should not be misused by AI developers to provide misleading explanations for system behavior. The challenge remains to translate these principles into verifiable practice.

Growth in interpretability and explainability research, for example, in attribution methods that reliably explain specific predictions of computer vision models, is welcome. However, more research effort should be directed toward combining techniques to address ethical and safety concerns. For example, greater effort could be focused on increasing the explainability of accidents and systemic biases. Further, we should (i) develop common standards for domain-specific interpretability criteria and objectives, for example, that can faithfully explain possible differences in predictions for different populations; (ii) develop tests that measure compliance with such standards; and (iii) identify domains where the context of use requires a high level of interpretability and develop performant interpretable models for these domains.

Complementary to interpretability and explainability, which help interrogate the outputs of AI systems, reproducibility allows external teams to recreate an AI system and interrogate it, verifying claims made by developers. Initiatives like the Association for Computing Machinery (ACM) artifact review and badging and the ML reproducibility challenge incentivize reproducibility in research settings.

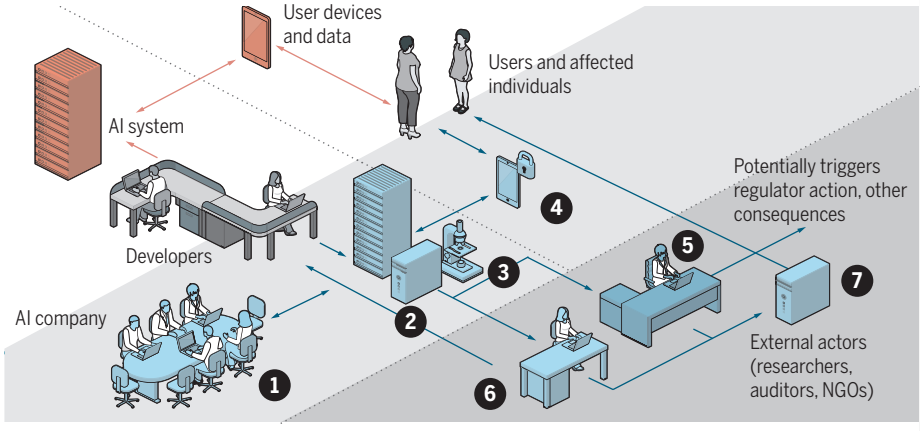
Privacy-preserving machine learning

Concerns regarding privacy in ML include unauthorized access to the data that are used to train models, privacy violations and targeting of individuals and communities through inferring sensitive information from trained models, and unauthorized access to the trained model itself. Research on privacy-preserving ML (PPML) has developed complementary techniques to address these concerns. Federated learning techniques allow the centralized training of a model with decentralized data, without raw data ever leaving the source device (*9*). Differential privacy techniques modify the development process such that trained models retain meaningful statistical patterns at the population level but reduce the risk of inferring information about individuals (*10*). Encrypted computation allows data owners and model developers to train models without either side gaining access to the information of the other. Together, these techniques can help mitigate privacy concerns, though each involves trade-offs, for example, regarding ease of development or training efficiency.

PPML techniques lack standard software libraries and awareness among AI developers. However, a growing community of open-source implementation projects has enabled progress toward wider adoption. Projects that extend PPML to existing ML frameworks, such as PyTorch's Opacus and Tensorflow Privacy, have seen particular growth. Projects like Federated AI (FedAI), PySyft, Flower, and OpenFL provide added value by considering privacy beyond the

Promoting and verifying trustworthiness

Existing relations (●) between organizations that develop artificial intelligence (AI) and users often leave gaps that make users unable to verify the trustworthiness of these organizations. The proposed relations (●) are supported by the mechanisms described in the text, which either (i) help developers adopt best practices in their internal processes and handling of user data (1 to 4, light-gray background) or (ii) incentivize external actors to evaluate the trustworthiness of developers and systems (5 to 7, dark-gray background) and share that information with users. Together, these mechanisms promote a flow of information about trustworthiness from developers, through external actors [researchers, auditors, nongovernmental organizations (NGOs)], to users.



1 Internal red teams **2** Audit trails **3** Interpretability tools **4** Privacy-preserving machine learning **5** Third-party audits **6** Bias and safety bounty programs **7** Public incident database

GRAPHIC: N. DESAI/*SCIENCE*

narrow context of model training; they illustrate to AI developers how to reason about where data is stored and what trade-offs exist in preserving privacy. To further accelerate adoption, we recommend establishing reliable support for active PPML projects, open standards, algorithm benchmarks, and educational resources.

Third-party auditing

For best practices to engender trust, AI developers must follow them and be seen to follow them. This is complicated by limitations on the information that can be shared publicly by AI developers, for example, private user data. A solution adopted in several other industries is third-party auditing, where an auditor gains access to restricted information and in turn either testifies to the veracity of claims made or releases information in an anonymized or aggregated manner. Third-party auditing sidesteps several antitrust concerns (*6*); it is also well positioned to leverage technical solutions, such as secure multiparty computation, that allow verification of claims without requiring direct access to sensitive information.

Auditing can take many forms, involving varying mixes of government and private actors and a range of funding models and information-sharing practices. A recent concrete proposal for independent auditing of AI systems highlights three key ingredients: (i) proactive independent risk assessment, (ii) reliance on standardized audit trails, and (iii) independent assessment of adherence to guidelines and regulations (*11*).

Auditing can only contribute to trust if auditors themselves are trusted and if failures to pass audits carry meaningful consequences; it is therefore essential that auditors have strong incentives to report their findings accurately and have protections for raising concerns when necessary. Reputational mechanisms, as well as government and civil society backing, would help provide such incentives.

Bias and safety bounties

The complexity of AI systems means that it is possible for some vulnerabilities and risks to evade detection before release. For similar reasons, in the field of cybersecurity, experts research flaws and vulnerabilities in published software and publicly available hardware. What began as an antagonistic relationship between vendors and external security researchers led to the development of "bug bounties" and responsible disclosure: mechanisms by which security experts can carry out their research and be financially rewarded for their findings while companies benefit from the discoveries and have a period in which to address them before they are publicly revealed. These mechanisms have helped align incentives in cybersecurity, have led to more secure systems, have and helped increase trust in companies that meaningfully and continuously engage in bug-bounty programs.

A similar approach could be adopted to reward external parties who research bias and safety vulnerabilities in released AI systems. At present, much of our knowledge about harms from AI comes from academic researchers and investigative journalists, who have limited access to the AI systems they investigate and often experience antagonistic relationships with the developers whose harms they uncover. The Community Reporting of Algorithmic System Harms project from the Algorithmic Justice League explores the potential for bounty programs that cover a broad range of harms from AI systems, including unfair bias. This July, Twitter offered bounties to researchers who could identify biases in their image-cropping algorithm. Note that such bounty systems do not shift the burden from AI developers—more resources should also be invested in surfacing and addressing vulnerabilities and biases before product release.

Sharing of AI incidents

As AI systems move from labs to the world, theoretical risks materialize in actual harms. Collecting and sharing evidence about such incidents can inform research and development as well as regulatory mechanisms and user trust. However, any AI developer is disincentivized to share incidents in their own systems, owing to reputational harms, especially if they cannot trust competitors to share similar incidents—a classic collective-action problem (*12*). Mechanisms are needed that enable coordination and incentivize sharing.

Incident sharing could become a regulatory requirement. Until then, voluntary sharing can be incentivized, for example, by allowing anonymous disclosure to a trusted third party. Such a third party would need to have transparent processes for collecting and anonymizing information and operate an accessible and secure portal. The Partnership on AI is experimenting with such a platform through its AI Incident Database, where information about AI incidents is compiled from both public sources and reporting from developers (*13*). Recently, the Center for Security and Emerging Technology developed a taxonomy of three categories (specification, robustness, and assurance) based on the incidents reported, with more than 100 incidents exemplifying each category (*14*).

CONCLUSION

The mechanisms outlined here provide concrete next steps that can help the public assess the trustworthiness of AI developers. Although we stress the need for a broader ecosystem that considers stages both before and after development, and that can enforce meaningful consequences, we see the verification of trustworthy behavior by developers as an important part of the maturation of the field of AI. These mechanisms enable targeted and effective regulation; for example, the European Union has proposed AI regulation that includes incident sharing, audit trails, and third-party auditors (*15*). We invite greater engagement with this urgent challenge at the interface of interdisciplinary research and policy. ■

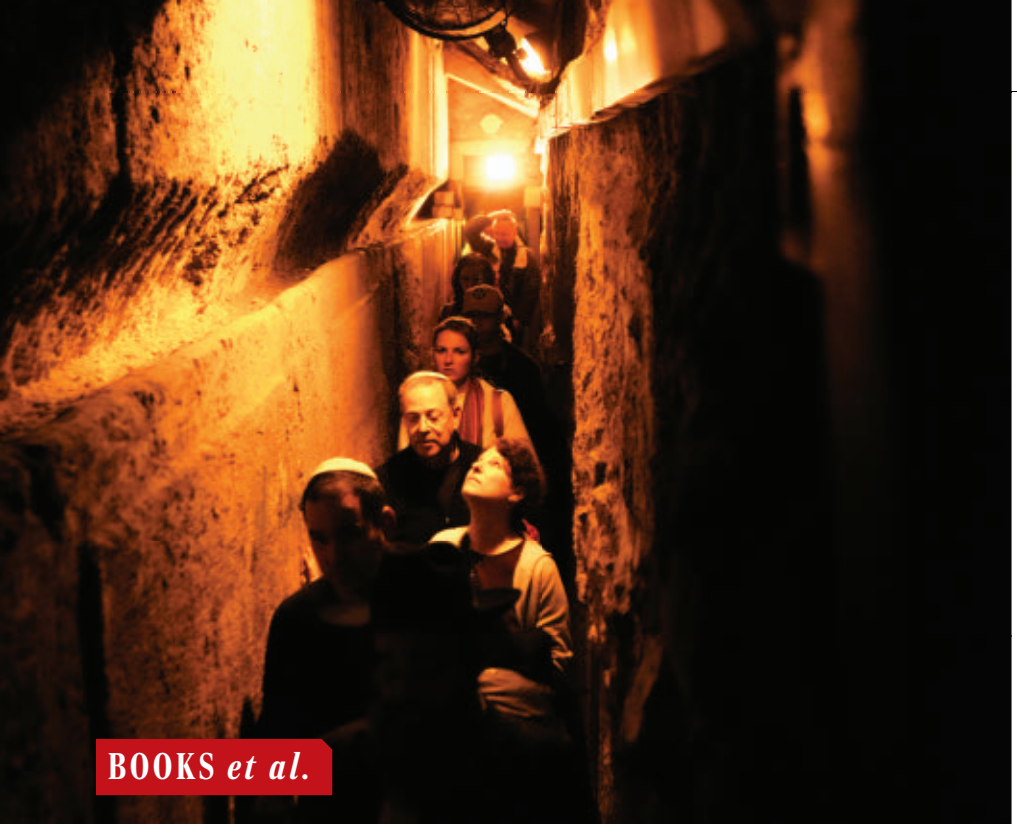
REFERENCES AND NOTES

1. A. Jobin, M. Ienca, E. Vayena, *Nat. Mach. Intell.* **1**, 389 (2019).
2. M. Brundage *et al.*, arXiv:2004.07213 (2020).
3. M. Whittaker *et al.*, "AI Now report 2018" (AI Now Institute, 2018); https://ainowinstitute.org/AI_Now_2018_Report.pdf.
4. S. Thiebes, S. Lins, A. Sunyaev, *Electron. Mark.* **31**, 447 (2021).
5. P. Xiong *et al.*, arXiv:2101.03042 (2021).
6. S. Hua, H. Belfield, *Yale J. Law Technol.* **23**, 415 (2021).
7. R. Bell, *ACM Int. Conf. Proceed. Ser.* **162**, 3 (2006).
8. International Organization for Standardization (ISO), "Report on standardisation prospective for automated vehicles (RoSPAV)" (ISO/TC 22 Road Vehicles, ISO, 2021; https://isotc.iso.org/livelink/livelink/fetch/-8856347/8856365/8857493/ISO_TC22_RoSPAV.pdf.
9. P. Kairouz *et al.*, arXiv:1912.04977 (2019).
10. A. Dwork *et al.*, in *Theory of Cryptography Conference* (Springer, 2006), pp. 265–284.
11. G. Falco *et al.*, *Nat. Mach. Intell.* **3**, 566 (2021).
12. A. Askell, M. Brundage, G. Hadfield, arXiv:1907.04534 (2019).
13. S. McGregor, arXiv:2011.08512 (2020).
14. S. McGregor, The first taxonomy of AI incidents (Partnership on AI, 2021); https://incidentdatabase.ai/blog/the-first-taxonomy-of-ai-incidents.
15. European Commission, "Proposal for a regulation of the European parliament and of the council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts" (COM/2021/206 final, European Commission, 2021); https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

ACKNOWLEDGMENTS

The report from which this paper draws was composed by 59 authors and benefitted from further feedback from colleagues and at workshops [see (*2*) for details]. After the corresponding author, the next five authors and the last six authors each form sets; within each set, authors contributed equally and are ordered alphabetically. We thank S. Bhatnagar for graphic design assistance and two anonymous reviewers for insightful suggestions. S.A. received funding from the Isaac Newton Trust and from Templeton World Charity Foundation, Inc. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of Templeton World Charity Foundation, Inc. H.B. received funding from the Casey and Family Foundation. M.B. and G.K. received funding from OpenAI. A.W. received funding from a Turing AI Fellowship under grant EP/V025379/1, the Alan Turing Institute, and the Leverhulme Trust through the Centre for the Future of Intelligence. M.A. received funding from Open Philanthropy and support from the Future of Humanity Institute. I.K. acknowledges support from the European Research Council Horizon 2020 research and innovation program under grant 725594 (time-data). D.K. received funding from Open Philanthropy.

10.1126/science.abi7176



BOOKS *et al.*

Visitors travel through the Western Wall tunnels under the Old City in Jerusalem in December 2009.

ARCHAEOLOGY

Interpreting Jerusalem

Science, politics, identity, and conflict converge in investigations of the city of peace

By **Thomas E. Levy**

Shortly after the 1967 Six-Day War between Israel and a coalition of its Arab neighbors, the Israeli poet Yehuda Amichai visited the newly unified city of Jerusalem, later describing it as "a port city on the shore of eternity" (*1*). The poem pays tribute to the metaphorical role played by the landlocked city, referencing its deep-time history and its intersectionality with the God of the three monotheistic faiths—Judaism, Christianity, and Islam.

Jerusalem and its history are closely linked to how people—ranging from scientists to religious practitioners to atheists—have imagined this mountain settlement over time. In his new book, *Under Jerusalem*, journalist Andrew Lawler seeks to unpack how the story of the city of Jerusalem, with its amazing underground world of ancient tunnels, roads, buildings, sewage works, and buried artifacts, is told in relation to the people who have populated it for longer than 5000 years. This ambitious book traces more than 150 years of continuous archaeological investigation of the world's most excavated city.

In Lawler's view, the social identity of Jerusalem's storytellers—that is, whether they see themselves primarily as archaeologists, historians, theologians, university professors, activists, or politicians—largely determines how they confront the city's buried past. He therefore pays careful attention to the nationality, ethnicity, political leanings, and socioeconomic standing of the individuals who appear in his account.

Lawler begins his story in 1863 with the earliest systematic investigation of the city, conducted by French senator Louis-Félicien Joseph Caignart de Saulcy. While excavating a large burial complex known as the Tomb of the Kings, the amateur archaeologist mistakenly concluded that he had found the final burial site of the Israelite King David and his son Solomon, setting the stage for a series of competing digs conducted by researchers from France, Germany, and the United Kingdom. These endeavors, often firmly rooted in Christian perspectives of the Holy Land, brought renewed interest to Palestine, a forgotten corner of the dying Ottoman Empire.

Contemporary standard practices for exploring archaeological sites focus on careful surface-down excavations. However, in the 1860s, in an effort to locate the remains of King David's palace, two young British military men, Charles Wilson and Charles Warren, used tunneling methods to explore Jerusalem's labyrinth of underground passages, channels, and buildings. A number of subsequent excavators, most of them Israeli, eventually followed suit.

Lawler interrogates the complex ideological motivations that compel the region's two Indigenous peoples, Jews and Arabs, to explore and alter the city. Jewish investigators are often motivated by a desire to explore clues to their heritage that lurk underground. Their efforts are frequently deeply interwoven with ideas about Jewish sovereignty in Jerusalem today. For Arab interpreters, it is the aboveground Islamic monuments that most often resonate with their identity and spur their investigations.

Under Jerusalem highlights the numerous archaeological atrocities that have been carried out in this contested city. These include the Jordanian destruction of the Jewish Quarter and its 27 synagogues in the 1948 war, the Israeli bulldozing of the Mugrabi Quarter adjacent to the Western Wall of the Temple Mount shortly after the 1967 war, and the currently occupied houses that have been undermined by archaeological tunneling. Some of these incidents reflect the evolution of archaeological practice, others the religion and politics of the actors—but all reflect a desire to use Jerusalem for the perceived good of the groups to which the investigators belong.

Under Jerusalem: The Buried History of the World's Most Contested City
Andrew Lawler
Doubleday, 2021. 464 pp.

Lawler hints that a commitment to science-based archaeology by a new generation of Israeli researchers, coupled with innovative underground excavation techniques, may herald a more objective telling of the history of the world's most contested city. This offers a distinctive opportunity for archaeologists to write a new narrative of the "city of peace" from the ground up. ■

REFERENCES AND NOTES

1. Y. Amichai, "Yerushalayim ir namal al sefat ha-neẓaḥ in Yerushalayim" (1967).

10.1126/science.abn0362

The reviewer is at the Department of Anthropology, University of California, San Diego, La Jolla, CA 92093, USA. Email: tlevy@ucsd.edu

PHOTO: MICHAL FATTAL/BACKYARD/REDUX

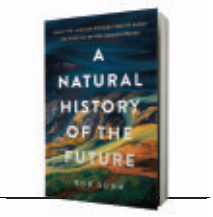
ANTHROPOCENE

Humanity versus the world

A biologist confronts the evolutionary arms race at play in human-remade landscapes

By **Mary Ellen Hannibal**

A Natural History of the Future: What the Laws of Biology Tell Us About the Destiny of the Human Species
Rob Dunn
Basic Books, 2021. 320 pp.



While global concern is focused on how to avoid climate cataclysm, more attention is needed on the myriad other ways humanity is affecting the living world. Even if we could halt fossil fuel emissions tomorrow, we would still need to make some big changes. Evolutionary biologist Rob Dunn's timely new book, *A Natural History of the Future: What the Laws of Biology Tell Us About the Destiny of the Human Species*, is a guide to this complex problem and offers palatable solutions.

Homo sapiens are just one species among millions, and yet we exert outsized control over conditions on Earth. With undeniable ingenuity and brute power, we have wrestled prairies into factory farms and valleys into gleaming towers of cement and steel. We have stopped up rivers and redirected them with dams and levees, and we have hijacked nature's timetables to suit our expedience.

All of these interventions are too clever by half. We have failed to take sufficient notice of nature's implacable power, both invisible in the machinations of natural selection and overwhelmingly present in drowned shorelines and skies blackened with wildfire smoke. As Dunn puts it, "We are like a driver who somehow gets down the road, despite being too short to see out the window, a little drunk, and very fond of acceleration."

Of all human-made habitats, our monocropped farms are the largest. According to Dunn, the collective area of corn planted on Earth is the size of France. Wheat, barley, rice, sugarcane, cotton, and tobacco fields make up a human-jiggered geography. These vast areas produce food for us—humanity eats more than half the net primary production of photosynthesis—but they also destroy habitats that for millennia hosted birds, native plants, mammals, butterflies, bees, and other diverse life-forms.

It may surprise the reader that in addition to displacing native species, agricultural fields host "hundreds, perhaps thousands, of pest and parasite species that live nowhere else." Dunn estimates that more new species have evolved in our crops than on the Galápagos Islands.

Cities likewise shape-shift not only their skylines but the composition of species living within them. In geological time, cities have arisen so abruptly that Dunn likens them to volcanic eruptions. Again, we crowd many species out but create space for others, including rats, pigeons, and bedbugs.

Dunn estimates that "no fewer than a thousand animal species now live indoors." Meanwhile, cities and industrial agriculture operations are connected by roads and other means of transport that effectively act as wildlife corridors for entities including deadly viruses. "[W]e are investing in an infrastructure that ensures their survival," notes Dunn.

One way humanity arrived atop the food chain was by eliminating predators. "Our ancestors," Dunn observes, "were often eaten." But the calculus changed when hominins began to hunt. We have been killing those that would kill us—from big-toothed predators to insect pests and microscopic pathogens—ever since. Our time at the top is, however, ephemeral: "Eventually, our enemies catch up," notes Dunn.

Natural selection works to sustain something like an arms race between *Homo sapiens* and our perceived enemies. Dunn describes an experiment undertaken by Michael Baym and colleagues, in which a giant petri plate was divided into a series of columns filled with increasing concentrations of antibiotics. Bacteria with no genetic resistance to the antibiotics were introduced into the outside (lowest-concentration) column, "defenseless as sheep." The bacteria were knocked back by the antibiotics, but some were able to reproduce, and genetic variation emerged in subsequent generations.

After a brief pause, "one and then many lineages evolved resistance to the lowest concentration in the antibiotic...." Again and again, the bacteria slowed then surged, generating mutations that could survive the next round. If we paid more attention to the rate of resistance, Dunn argues, we could manage "the river of life" with surgical strikes rather than by bombing everything in sight.

A Natural History of the Future is a clear and important read with one flaw. Dunn not only overuses the term "law," he never defines it. While concepts such as "corridors" and "niche" do help govern the diversity of life and are applicable whether the species in question are hummingbirds or cockroaches, they are not immutable laws. His excellent explication of this could have been achieved without fudging semantics. ■

The reviewer is the author of *Citizen Scientist: Searching for Heroes and Hope in an Age of Extinction* (The Experiment, 2016). Email: maryellenhannibal@gmail.com

PHOTO: JACOB MILLER

Students weed an experimental plot in East Bethel, MN, where scientists are studying biodiversity and resilience.

10.1126/science.abl9604

An aerial photo shows a sewage treatment plant in Hebei Province, China. China plans to address water scarcity by scaling up its wastewater treatment capacity.

Edited by **Jennifer Sills**

Wastewater: China's next water source

In the face of growing water scarcity, China has tried to cope by setting strict regulations to control total water consumption (*1*), limiting agricultural and industrial water consumption (*2*), and moving water around the country through the huge South-to-North Water Diversion Project (*3*, *4*). However, increasing water demand and decreasing water quality are forcing China to look for additional solutions. Recognizing the potential of wastewater as a water source, China has set ambitious goals for the amount of wastewater it can reclaim for use: more than 35% of wastewater discharge in the Beijing-Tianjin-Hebei region and more than 25% in prefecture-level water-scarce cities by 2025 (*5*).

In 2019, China's urban wastewater discharge was about 75 billion cubic meters (*2*), and the amount of reclaimed water was less than 10 billion cubic meters, less than 15% of urban wastewater discharge. However, China's cities are well positioned to increase their reclaimed water production. Newer cities are already equipped with modern water management facilities, which handle water supply, drainage, and sanitation as components of a centralized, integrated system (*6*). In addition to increasing the capacity of current wastewater treatment systems, the primary obstacle in cities is the high cost of water treatment, which must be absorbed by the water producers or distributed among potentially unsupportive stakeholders.

In contrast, in China's vast rural areas, water reclamation is in its infancy. The small population density and unstructured layout of residential buildings have led to varied wastewater discharge methods. In towns trying to recycle wastewater, the construction of small-scale wastewater collection pipeline networks will be the first step. Instead of the large, centralized treatment facilities in cities, rural areas will require small, decentralized treatment systems serving each local area.

To meet its wastewater reclamation goals, China must increase and restructure its investments. The national government must fund new and upgraded wastewater infrastructure. Public-private partnerships, a government financing model used to encourage companies to participate in the construction of public infrastructure, have decades-long payback periods for wastewater treatment, a heavy capital burden for companies. To correct this, the government has proposed alternative methods of project financing, such as real estate investment trusts, national green development funds, and central banks' environmental low-cost funding support. Successful wastewater reuse will require the coordination of central and local governments, the supervision of the environmental protection bureaus, the participation of the scientific community, and the support of society. It will also depend on flexible market-based approaches to incentivize market players, such as tax incentives and low interest rate funding for companies investing in the technology and water price adjustments to manage higher production costs.

Yanxia Zhang, Jiazhen Huo, Xiaojin Zheng*
School of Economics and Management, Tongji University, Shanghai 200092, China.
*Corresponding author.
Email: xjzheng@tongji.edu.cn

REFERENCES AND NOTES

1. The State Council the People's Republic of China, "The Thirteenth Five-Year period to implement the strictest water resources management system assessment results released" (2021); www.gov.cn/xinwen/2021-09/29/content_5639965.htm [in Chinese].
2. Ministry of Housing and Urban-Rural Development of the People's Republic of China, "2019 China urban and rural construction statistical yearbook" (2020) [in Chinese].
3. M. Webber *et al.*, *WIREs Wat.* **2**, e1556 (2021).
4. Ministry of Water Resources of the People's Republic of China, "Project overview" (2021); http://nsbd.mwr.gov.cn/zw/gcgk/ [in Chinese].
5. National Development and Reform Commission People's Republic of China, "Guidance on promoting the resource utilization of wastewater" (2021); www.ndrc.gov.cn/xwdt/tzgg/202101/t20210111_1264795_ext.html [in Chinese].
6. Ministry of Housing and Urban-Rural Development of the People's Republic of China, "GB/T 34173-2017: Urban drainage and sewage treatment service" (2017); http://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=5998B3DBF3314A7A76A9945750FB7D21 [in Chinese]

10.1126/science.abm6738

Noise pollution disrupts freshwater cetaceans

Anthropogenic noise pollution is an emerging stressor for aquatic animals, especially whales, dolphins, and porpoises, which rely heavily on sound production, transmission,

PHOTO: COSTFOTO/BARCROFT MEDIA/GETTY IMAGES

and reception (*1*, *2*). Yet ocean soundscapes have aroused great concern (*3*), whereas little attention has been paid to anthropogenic noise pollution in riverine habitats.

Unlike the marine environment, riverine ecosystems are constricted to areas that often face inevitable competition with humans. All cetaceans that live in rivers, including five dolphin species and a porpoise, have declined substantially both in range and numbers, especially in Asia (*4*). The Yangtze finless porpoise (*Neophocaena asiaeorientalis asiaeorientalis*), the sole freshwater porpoise in the world, which is only found in two Chinese lakes and some parts of the Yangtze River, is listed as Critically Endangered by the International Union for Conservation of Nature (*5*).

Noise pollution affects freshwater cetaceans across regions. The Yangtze has been the world's busiest inland river in terms of shipping since 2005 (*6*), and its daily average vessel flow and annual cargo movement have increased over the past decade (*7*). Noise levels at approximately 88% of the river's porpoise distribution area are sufficient to disrupt the behavior of the finless porpoise, and 40% of the river has noise levels high enough to cause reduction in the porpoise's hearing sensitivity (*8*, *9*). Underwater noise exposure from vessels in the Ganges River in India elevated metabolic stress in Ganges river dolphins (*Platanista gangetica gangetica*) (*10*). Busy vessel traffic in the Indus River in Northern Pakistan threatens the Indus River dolphin (*P. g. minor*) (*11*). Compared with coastal Irrawaddy dolphins (*Orcaella brevirostris*), freshwater Irrawaddy dolphins in Indonesia's Mahakam River showed a hypersensitivity to intensive boat traffic (*12*).

Addressing noise pollution will require improvements to shipping lanes, ports, and ships. Shipping lanes should be regulated to protect channel resources. Optimal navigation speeds and restrictions of ship speeds in ecologically important areas would help mitigate noise pollution in cetacean habitats. The government should identify and eliminate illegal ports, especially those that have been built along the Yangtze in China. Ships that are docked should be required to use power from land rather than continuing to run their engines. The construction and renovation of environmentally friendly and economical ships should be developed and put to use. Subsidies should be provided to companies using outdated vessels to help them replace the ships with newer, quieter technology, such as more efficient propellers. Finally, governments should require all river traffic to meet operation and management standards that reduce noise pollution.

Zhi-Tao Wang, Peng-Xiang Duan, Ke-Xiong Wang*, Ding Wang*
The Key Laboratory of Aquatic Biodiversity and Conservation of the Chinese Academy of Sciences, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China.
*Corresponding author. Email: wangk@ihb.ac.cn (K.-X.W.); wangd@ihb.ac.cn (D.W.)

REFERENCES AND NOTES

1. S. L. King, V. M. Janik, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13216 (2013).
2. S. D. Simpson *et al.*, *Nat. Commun.* **7**, 10544 (2016).
3. C. Duarte *et al.*, *Science* **371**, eaba4658 (2021).
4. Marine Mammal Commission, "Freshwater dolphins and porpoises" (2021); www.mmc.gov/priority-topics/species-of-concern/freshwater-dolphins-and-porpoises/.
5. D. Wang, S. T. Turvey, X. Zhao, Z. Mei, "*Neophocaena asiaeorientalis* ssp. *Asiaeorientalis*" (The IUCN Red List of Threatened Species, 2013); www.iucnredlist.org/species/43205774/45893487.
6. X. Y. Zhu, "The annual cargo throughput of the Yangtze River trunk line has ranked first in the world for many years," *China News* (2019); www.chinanews.com/cj/2019/07-05/8885178.shtml [in Chinese].
7. Editorial Committee of the Yangtze River Yearbook, *Yangtze River Yearbook in 2020* (Yangtze River Yearbook Press, Wuhan, 2020) [in Chinese].
8. Z. T. Wang *et al.*, *Environ. Pollut.* **262**, 114310 (2020).
9. Z. T. Wang *et al.*, *Ecotoxicol. Environ. Safe.* **226**, 112860 (2021).
10. M. Dey *et al.*, *Sci. Rep.* **9**, 15426 (2019).
11. F. Perveen, S. A. Mehmood, S. Ahmed, Z. U. Reman, *J. Life Sci.* **5**, 39 (2011).
12. D. Kreb, K. D. Rahadi, *Aquat. Mamm.* **30**, 363 (2004).

10.1126/science.abf0222

STEM disparities we must measure

The National Science Foundation's (NSF's) Survey of Earned Doctorates, an annual census of all newly minted US PhDs since 1957, has been a cornerstone in science, technology, engineering, and mathematics (STEM) diversity efforts (*1*). Along with NSF's other national workforce surveys, the Survey of Earned Doctorates provides critical data needed to document underrepresented groups in STEM and determine their eligibility for funding and federal resources. However, NSF's surveys do not include sexual orientation and gender identity (SOGI) questions, despite the alarming disparities that lesbian, gay, bisexual, transgender, and queer (LGBTQ) people face in US STEM fields (*2*).

LGBTQ people are estimated to be 17 to 21% less represented in STEM than statistically expected (*3*, *4*) and are less likely to major in STEM, persist in STEM, earn STEM degrees, and be in STEM occupations (*2*, *5*, *6*). Harmful biases and unsupportive STEM environments appear to be partly at fault. For instance, LGBTQ scientists experience more career barriers and workplace harassment than non-LGBTQ scientists, even when controlling for other demographic and career-related factors (*7*).

In April, after years of delays, NSF began piloting SOGI questions for its surveys, and the agency presented its first results on 2680 respondents in November (*8*). Although NSF had longstanding concerns that the questions would be perceived as too sensitive, their results showed instead that break-off and nonresponse rates (which indicate discomfort with survey questions) were virtually 0%. Moreover, LGBTQ and non-LGBTQ respondents alike overwhelmingly reported feeling comfortable providing SOGI data to a federal agency like NSF. However, NSF stopped short of any indication that it would consider adding SOGI questions to its surveys now, suggesting instead that more research is needed.

Any incremental knowledge NSF will gain by additional research is outweighed by the urgent need for nationwide SOGI data on the STEM workforce. Agencies from the Department of Education to the Department of Labor have already been collecting SOGI data on hundreds of thousands of respondents for years (*9*). In 2016, the Department of Education began collecting SOGI data on its national surveys of high-schoolers and college students (*10*). Five years later, NSF is increasingly out of sync with its federal counterparts.

Demographic questions are never perfect. NSF's race and ethnicity questions reflect federal-wide standards established in 1977 and later revised in 1997 (*11*). The questions have been criticized for being insufficiently nuanced (*12*), but NSF has never hesitated to include them, because race and ethnicity data are critically needed to resolve urgent STEM opportunity gaps. The same logic applies to SOGI data. NSF should always work to improve how it collects demographic information, but striving for more perfect survey methods should never trump its congressionally mandated responsibility to address the challenges faced by underrepresented groups in STEM. LGBTQ people in STEM should be no exception.

Jonathan B. Freeman
New York University, New York, NY 10003, USA.
Email: jon.freeman@nyu.edu

REFERENCE AND NOTES

1. NSF, Survey of Earned Doctorates (2021); www.nsf.gov/statistics/srvydoctorates/.
2. J. B. Freeman, *Pol. Insights Behav. Brain Sci.* **7**, 141 (2020).
3. J. B. Freeman, *Nature* **559**, 27 (2018).
4. E. Cech, M. Pham, *Soc. Sci.* **6**, 12 (2017).
5. B. E. Hughes, *Sci. Adv.* **4**, eaao6373 (2018).
6. D. Sansone, C. S. Carpenter, *PLOS ONE* **15**, e0241596 (2020).
7. E. Cech, T. Waidzunas, Sci. Adv. **7**, eabe0933 (2021).
8. R. L. Morrison, J. Chandler, F. Lan, K. Jamrick, "Session B-1: Measuring sexual orientation & gender identity:

Measuring SOGI among college graduates: Results from a methodological experiment using a non-probability sample," Federal Committee on Statistical Methodology 2021 Research and Policy Conference (2021); https://copafs.org/fcsm-2021-conference-program/.
9. R. Ellis *et al*., "Assessing the feasibility of asking about sexual orientation and gender identity in the current population survey: Results from cognitive interviews," (Working Paper Number RSM2018-06, US Census, 2018); www.census.gov/library/working-papers/2018/adrm/rsm2018-06.html.
10. E. M. Christopher, L. Burns, "Session 4B: Questionnaire evaluation: Using paradata and metadata to assess effects of addition of sensitive items to an ongoing longitudinal survey," 2021 FedCASIC Conference (2021); www.census.gov/fedcasic/fc2021/archive.php.
11. Office of Management and Budget, "Revisions to the standards for the classification of federal data on race and ethnicity" (1997); https://obamawhitehouse.archives.gov/omb/fedreg_1997standards.
12. C. M. Snipp, *Annu. Rev. Soc.* **29**, 563 (2003).

10.1126/science.abn1103

TECHNICAL COMMENT ABSTRACTS

Comment on "An early Miocene extinction in pelagic sharks"

Gavin J. P. Naylor, Arthur de Lima, José I. Castro, Gordon Hubbell, Mario C. C. de Pinna

Sibert and Rubin (Reports, 4 June 2021, p. 1105) report an early Miocene extinction in pelagic sharks based on the loss of shark denticle diversity in two widely separated deep-sea sediment cores. We assert that the pattern observed is not a consequence of extinction but of shifting species ranges induced by global current reorganization.
Full text: dx.doi.org/10.1126/science.abj8723

Response to Comment on "An early Miocene extinction in pelagic sharks"

Elizabeth C. Sibert and Leah D. Rubin

Naylor *et al*. argue that the existence of multiple denticle types within a single species precludes the use of this metric as a measure of the decline of multiple shark species. We show that species-level shark diversity would have to decrease by >90% to account for the observed >70% denticle extinction, implying that the early Miocene shark extinction was larger than previously recognized.
Full text: dx.doi.org/10.1126/science.abj9522

Comment on "An early Miocene extinction in pelagic sharks"

I. Feichtinger, S. Adnet, G. Cuny, G. Guinot, J. Kriwet, T. A. Neubauer, J. Pollerspöck, K. Shimada, N. Straube, C. Underwood, R. Vullo, M. Harzhauser

Sibert and Rubin (Reports, 4 June 2021, p. 1105) claim to have identified a previously unidentified, major extinction event of open-ocean sharks in the early Miocene. We argue that their interpretations are based on an experimental design that does not account for a considerable rise in the sedimentation rate coinciding with the proposed event, nor for intraspecific variation in denticle morphology.
Full text: dx.doi.org/10.1126/science.abk0632

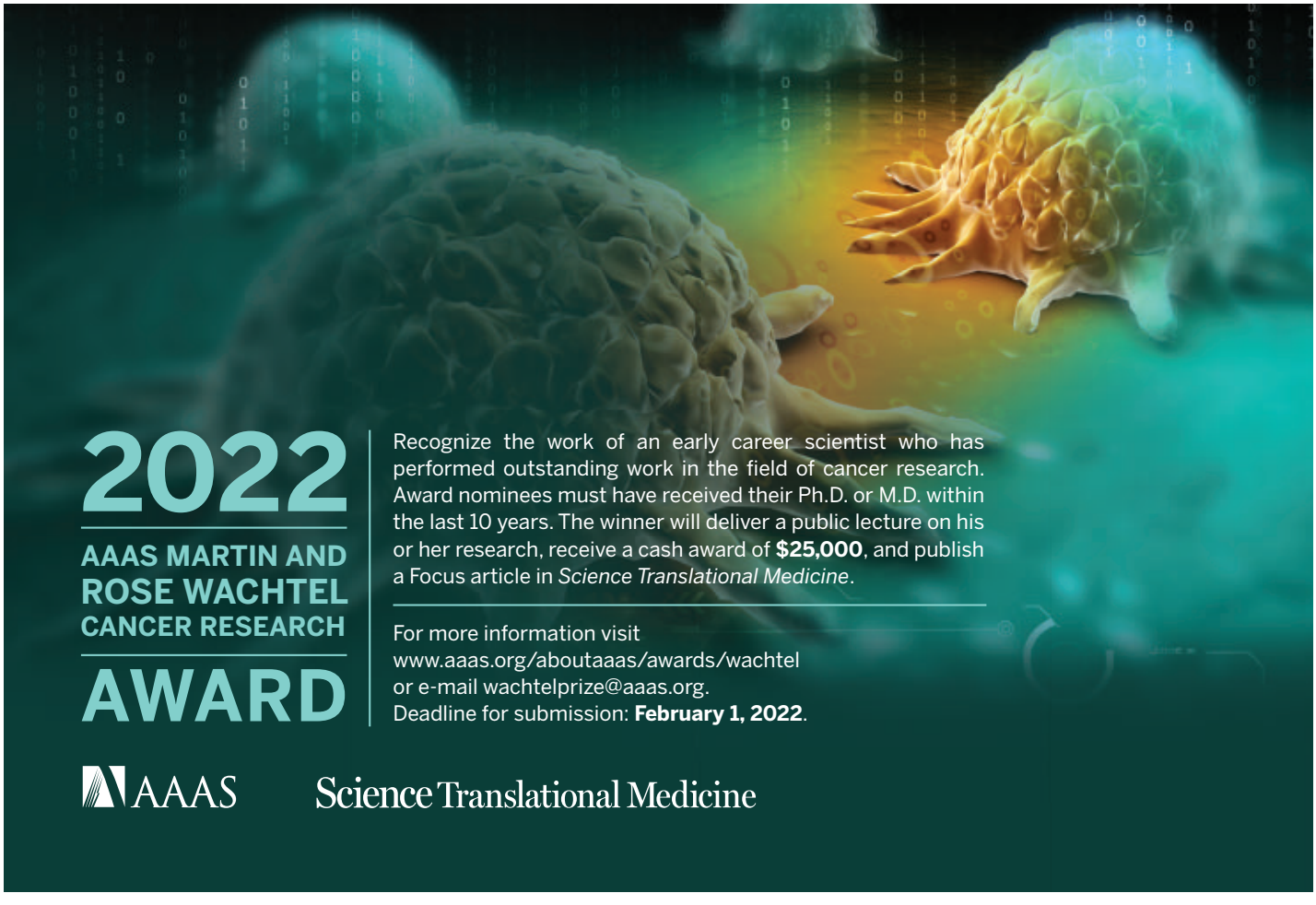
Response to Comment on "An early Miocene extinction in pelagic sharks"

Elizabeth C. Sibert and Leah D. Rubin

Feichtinger *et al*. assert that the reduction in denticle abundance and diversity we found are incorrect, claiming that we failed to consider changes in sedimentation rate. However, we used standard methods that explicitly account for changes in sedimentation rate and density. We maintain that our initial dataset and conclusions are robust and provide evidence for a major early Miocene extinction in pelagic sharks.
Full text: dx.doi.org/10.1126/science.abk1733

ERRATA

Erratum for the Research Article "Cauliflower fractal forms arise from perturbations of floral gene networks," by E. Azpeitia *et al*., *Science* **374**, eabn5313 (2021). Published online 10 December 2021; 10.1126/science.abn5313



RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

HYDROPHOBICITY

Why oil and water do not mix

It is well known that oil forms stable droplets that carry a negative electrophoretic mobility (and negative charge) upon dispersing in water. However, the underlying mechanism is a long-debated topic. Using vibrational sum-frequency scattering spectroscopy, Pullanchery *et al*. recorded the interfacial vibrational spectrum in the oxygen–deuterium and carbon–hydrogen stretching regions of a hexadecane–water interface. Their spectral analysis accompanied by molecular dynamics simulations showed that water molecules form "improper" interfacial hydrogen bonds with alkyl hydrogens, resulting in the water-to-oil charge transfer that stabilizes oil droplets. This work demonstrates that sum-frequency scattering spectroscopy is a powerful technique that can improve our understanding of hydrophobicity in water-mediated chemical and biological systems. —YS

Science, abj3007, this issue p. 1366

Oil and water mixtures form stable droplets in part due to "improper" hydrogen bonds between water molecules and the alkyl chains of oil.

PHOTO: TURTLE ROCK SCIENTIFIC/SCIENCE SOURCE

PLANT SCIENCE

Shared systems in leaf development

The long, narrow leaves of grasses look rather different from the often shorter, flatter leaves of eudicot plants. Richardson *et al*. combined developmental genetics and computational modeling to reveal that these two types of leaves, which are widely separated by evolution, have more in common than expected. Expression of similar patterning genes in the primordial zone is confined to a wedge for the eudicot leaf but expanded to concentric domains in the grass leaf, driving development of the cylindrical, encircling sheath characteristic of these leaves. Addition or removal of gene expression in a marginal zone contributes to the development of the broader leaf characteristic of eudicots. Thus, grass and eudicot leaves are diversified elaborations of shared toolkits. —PJH

Science, abf9407, this issue p. 1377

GRAPHENE

Detecting orbital magnetism

Graphene's electronic structure has been predicted to lead to an unusual orbital response to magnetic fields. However, detecting this orbital magnetism is difficult because it is usually masked by the signal stemming from spins. Vallejo Bustamante *et al*. managed to capture this response by placing two giant magnetoresistance detectors below a sample of graphene sandwiched by layers of hexagonal boron nitride. These detectors picked up a strong diamagnetic response from an undoped sample, consistent with theoretical predictions. The technique may be useful in the investigation of other two-dimensional materials. —JS

Science, abf9396, this issue p. 1399

SUPERCONDUCTIVITY

Imaging a peculiar Fermi surface

Running a current through a superconductor can cause the superconducting gap to close on a section of the Fermi surface. However, observing this segmented Fermi surface directly is tricky. To do so, Zhu *et al*. worked with a thin film of the topological insulator bismuth telluride placed on top of superconducting niobium diselenide. A small applied magnetic field caused a screening current, which in turn led to a segmented Fermi surface in the topological insulator layer. Using a scanning tunneling microscope, the researchers were able to map out this Fermi surface. —JS

Science, abf1077, this issue p. 1381

OPTICAL IMAGING

A view through a fiber

Reconstructing a three-dimensional (3D) image of a scene typically involves sending out pulses of light and timing their return. For endoscope applications in bioimaging or imaging inside difficult-to-reach places inside machines, the typical approach using bulk optics may not be viable. Stellinga *et al*. found that 3D imaging can be achieved using multimode optic fibers. After characterizing the transmission matrix of the fiber, optical pulses can be used to reconstruct 3D images of a number of scenes. Because this approach can use fibers the width of a human hair, the results are promising

for miniaturized and minimally invasive 3D endoscopic imaging applications. —ISO

Science, abl3771, this issue p. 1395

CANCER

Chimeric costimulation

Improving chimeric antigen receptor (CAR) T cell therapy is essential for avoiding relapses in patients with hematological cancers. To address common mediators of relapse, Katsarou *et al*. developed chimeric costimulatory receptors, which bind to a second tumor-associated antigen and provide intracellular signaling to support CAR T cell activation. The authors observed increased tumor cell killing in vitro and in vivo, even against tumor cells with very low antigen density, and this was associated with improved CAR T cell persistence. —CSM

Sci. Transl. Med. **13**, eabh1962 (2021).

FOREST ECOLOGY

Resilient secondary tropical forests?

Although deforestation is rampant across the tropics, forest has a strong capacity to regrow on abandoned lands. These "secondary" forests may increasingly play important roles in biodiversity conservation, climate change mitigation, and landscape restoration. Poorter *et al*. analyzed the patterns of recovery in forest attributes (related to soil, plant functioning, structure, and diversity) in 77 secondary forest sites in the Americas and West Africa. They found that different attributes recovered at different rates, with soil recovering in less than a decade and species diversity and biomass recovering in little more than a century. The authors discuss how these findings can be applied in efforts to promote forest restoration. —AMS

Science, abh3629, this issue p. 1370

NEUROSCIENCE

Insight into brain repair

One approach to enhancing cellular regeneration in the brain is to exploit the molecular mechanisms used by the juvenile brain when regeneration potential is high. Bayin *et al.* found that irradiation injury in the neonatal cerebellum causes a progenitor cell that usually makes glial cells to instead switch its potential and make neurons. Single-cell sequencing implicated the transcription factor Ascl1 as the factor that controls this switch in potential, and the switch did not occur in the absence of Ascl1. This study provides new insight into the plasticity potential of the juvenile brain and potential targets to enhance regeneration after injury. —NJA

Sci. Adv. 10.1126/sciadv.abj1598 (2021).

Forests such as this secondary tropical forest along the Amazon River near Iquitos, Peru, can recover some, but not all, old-growth attributes quickly.

IN OTHER JOURNALS

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

CELL BIOLOGY

Taking a sip from the lumen

The gut continuously samples its contents for potentially harmful substances. Intestinal goblet cells help to maintain the gut epithelium by secreting mucus. These cells also help to monitor the gut lumen by forming so-called goblet cell–associated antigen passages. In mice, Gustafsson *et al.* used high-resolution imaging approaches to show that gut luminal contents were transcytosed across goblet cells. Endocytosed material was also delivered in a more conventional fashion to endosomes and lysosomes. Acetylcholine triggered both endocytosis and goblet cell mucus secretion; however, both processes could be independently modulated to balance the needs of intestinal barrier maintenance and immune sampling of the gut contents. —SMH

eLife **10**, e67292 (2021).

FOREST ECOLOGY

Drying reduces tropical tree growth

Long-term climate changes and climate extremes can reduce tree growth and limit potential carbon sequestration by forests. Bauman *et al.* used demographic data from 23 sites over 49 years to disentangle the effects of mean and anomalous climate conditions on tree growth in Australia's tropical forests. They found that trees in warmer (low-elevation) sites

PHOTOS: (LEFT TO RIGHT) AMAZON-IMAGES/ALAMY STOCK PHOTO; IMAGEBROKER/ALAMY STOCK PHOTO

PLANT ECOLOGY

Survival traits for the treetops

Epiphytes, plants that grow on the stems and branches of other plants, account for about 10% of all vascular plants. However, knowledge of their biology has lagged that of ground-rooted plants. To address this lack, Hietz *et al.* assembled data on the functional traits of almost 3000 epiphytic species, mainly from the neotropics. The authors found that epiphytes typically have lower photosynthetic rates than trees and ground-rooted herbs and lower leaf nitrogen content. These traits mean that epiphyte leaves tend to be thick and retain more water, reflecting the constraints of the nutrient and water supply in the forest canopy. Although this work draws together much-needed data on epiphyte leaf traits, much remains to be learned about the root physiology and reproductive habits of these plants. —AMS

J. Ecol. 10.1111/1365-2745.13802 (2021).

Epiphytic plants tend to have thicker leaves to help conserve water and nutrients.

grew faster, but anomalies of high solar radiation and atmospheric water demand reduced tree growth. Fast-growing trees and drier forests were also more sensitive to dry periods, which are increasing over time. This study shows the potential for climate change to limit carbon storage in tropical forests. —BEL

Glob. Change Biol. 10.1111/gcb.15982 (2021).

METABOLOMICS

Panning for metal-bound small molecules

Because of their important role as cofactors and their high reactivity, microbes devote considerable energy and biosynthetic resources to producing small molecules that aid in solubilization, transport, and chaperoning of transition metal ions. Aron *et al.* developed a native mass spectrometry method that enables the identification of these compounds. By adjusting the pH and adding metal salts after chromatographic separation but before ionization, the authors were able to identify metal-bound complexes in culture extracts and complex samples. Iron-binding siderophores were easily observed in various contexts, as were several known zinc- and copper-binding molecules. Metal sequestration is a common antimicrobial strategy, and this approach may provide insights into how microbes acquire scarce metal ions. —MAF

Nat. Chem. 10.1038/s41557-021-00803-1 (2021).

PHYSICS

Dots simulate magnetism

Engineered quantum systems can be used to simulate the complex many-body physics that is difficult to probe in "real" materials or to calculate using classical computers. Atomic systems, including cold atoms in optical lattices and trapped ions, have been at the forefront of this quantum simulation effort. In a complementary development, van Diepen *et al.* used a solid-state system, a linear array of quantum dots in a GaAs–AlGaAs heterostructure, to simulate quantum magnetism. By controlling the potentials of the four quantum dots in the array, the researchers were able to tune the exchange interactions between the single electrons occupying each dot. The method may be generalizable to longer arrays and different lattice geometries. —JS

Phys. Rev. X **11**, 041025 (2021).

CANCER

A gene target for bile duct cancer

Cholangiocarcinomas are rare and highly aggressive cancers that occur within bile ducts. Activation of the fibroblast growth factor receptor 2 (FGFR2) pathway is frequently observed for those bile duct tumors that occur within the liver (intrahepatic cholangiocarcinoma) and show mixed therapeutic responses. Cleary *et al.* performed genomic analysis of several hundred biliary tract cancers and discovered a new class of recurrent mutations in the *FGFR2* gene. *FGFR2* extracellular domain in-frame deletions were found to be an alternative mechanism of FGFR2 activation and oncogenic transformation that was experimentally targeted by treatment with FGFR inhibitors. —PNK

Cancer Discov. **11**, 2488 (2021).

ORGANIC CHEMISTRY

Aldehyde deuteration

Replacing hydrogen in organic molecules with its heavier deuterium isotope is a widely used strategy for studying drug metabolism and chemical mechanisms more broadly. Gadekar *et al.* sought to optimize deuterium incorporation at the carbonyl site of benzaldehyde derivatives, a challenging compound class for this reaction. By correlating rates with a variety of parameters associated with both the substrates and *N*-heterocyclic carbene catalysts, the authors were able to maximize deuteration efficiency while minimizing a competing coupling reaction between the aldehydes. —JSY

ACS Catal. **11**, 14561 (2021).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CANCER

Progress with oncolytic viruses

Oncolytic virotherapy was first thought to work by viruses selectively infecting and then lysing tumor cells. These viruses can be engineered to selectively target tumor cells and also to deliver various types of payloads to improve their efficacy. However, after much clinical testing, it has become clear that oncolytic viruses are a form of immunotherapy whereby viral infection drives an antiviral and antitumor immune response. In a Perspective, Melcher *et al.* discuss the current challenges of developing effective oncolytic virotherapy that can induce an antitumor response, often in combination with other immunotherapeutics. Although some encouraging phase 1 trials have recently been reported, there have been disappointments in converting early promise to long-term uptake and approval, highlighting the need to incorporate tumor immunology analyses in clinical trials and to consider how combination trials can be improved. —GKA

Science, abk3436, this issue p. 1325

IMMUNOLOGY

Atypical T cells take a different path

Conventional T cells recognize peptides that are presented by polymorphic major histocompatibility complexes (MHCs). By contrast, many tissue-resident T cells, such as mucosal-associated invariant T cells, invariant natural killer T cells, and γδ T cells, respond to modified peptides and small molecules presented by conserved MHC-like molecules. Unconventional T cells are important for host defense and tissue repair and seed tissues during critical early-life windows of development. Constantinides and Belkaid review recent advances describing how unconventional T cell subsets compete for early-life signals, including those from the microbiota, which instruct their development and have enduring consequences for the health of the host. —STS

Science, abf0095, this issue p. 1338

CELL BIOLOGY

ESCRTing lenses away from senescence

ESCRT proteins control membrane fusion in various key cellular processes, but the mechanisms involved are still incompletely understood. Gulluni *et al.* report that ESCRT recruitment at the cytokinetic bridge is mediated by the binding of an ESCRT-II subunit to the signaling lipid phosphatidylinositol 3,4-bisphosphate (see the Perspective by Brill and Wilde). This pathway acts in parallel to a known cascade driven by a protein called ALIX, but its failure is sufficient to lead to premature senescence in the lens of fish, mouse, and human eyes, where ALIX is expressed at lower levels. These results point to an evolutionarily conserved pathway for the cell-specific control of cytokinesis that serves to protect from senescence and the early onset of cataracts. —SMH

Science, abk0410, this issue p. 1339; see also abm7949, p. 1318

STRUCTURE PREDICTION

Deep learning for protein interactions

The use of deep learning has revolutionized the field of protein modeling. Humphreys *et al.* combined this approach with proteome-wide, coevolution-guided protein interaction identification to conduct a large-scale screen of protein-protein interactions in yeast (see the Perspective by Pereira and Schwede). The authors generated predicted interactions and accurate structures for complexes spanning key biological processes in *Saccharomyces cerevisiae*. The complexes include larger protein assemblies such as trimers, tetramers, and pentamers and provide insights into biological function. —VV

Science, abm4805, this issue p. 1340; see also abm8295, p. 1319

CELL BIOLOGY

Mechanosensitive nuclear pores

The nucleus of eukaryotic cells is enclosed by the nuclear envelope, a double membrane punctuated with nuclear pore complexes (NPCs). These giant channels in the nuclear envelope mediate nucleocytoplasmic exchange. Zimmerli *et al.* show that the mechanical status of the nuclear membranes controls their nuclear pore diameter. Pulling forces imposed through nuclear membranes lead to stretching of NPCs and dilation of their diameter, whereas relief of such forces causes NPC constriction. Thus, the control of nuclear size and shape is functionally linked with NPC conformation and nucleocytoplasmic transport activity. —SMH

Science, abd9776, this issue p. 1341

NEURODEVELOPMENT

Surveying brain interneuron development

As transient structures in early brain development, the ganglionic eminences generate dozens of different types of interneurons that go on to migrate throughout and weave together the developing brain. Shi *et al.* analyzed human fetal ganglionic eminences. Single-cell transcriptomics revealed unexpected diversity in the types of progenitor cells involved. The human ganglionic eminence depends more heavily on intermediate progenitor cells as workhorses than does the developing neocortex, with its greater reliance on radial glial cells. —PJH

Science, abj6641, this issue p. 1342

CORONAVIRUS

Delta's spike

Understanding the molecular mechanisms of the increased transmissibility and immune evasion of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants is critical to guiding current and future intervention strategies. Zhang *et al.* determined cryo–electron microscopy structures of the full-length spike protein trimers of the Delta, Kappa, and Gamma variants of SARS-CoV-2 and studied their function and antigenic properties. The Delta spike protein fused membranes more efficiently at low levels of the cellular receptor ACE2, and its pseudotyped viruses infected target cells substantially more rapidly than all other variants tested, possibly at least partly accounting for its heightened transmissibility. Mutations of each variant rearranged the antigenic surface of the N-terminal domain of the spike protein but only caused local changes in the receptor-binding domain, consistent with greater resistance to neutralizing antibodies. These findings elucidate the molecular events that have led these viruses to adapt in human communities and to evade host immunity. —VV

Science, abl9463, this issue p. 1353

CATALYSIS

Rational design of stable nanocatalysts

Sintering of nanoparticles is one of the main causes of their catalytic deactivation. Rational design of nanocatalysts that are stable against sintering is a grand challenge in heterogeneous catalysis. Hu *et al.* present kinetic theories for two competing sintering mechanisms, Ostwald ripening and particle

migration, which relate the rates of both processes to fundamental interaction energies in metal nanoparticle-support combinations. Using kinetic simulations for hundreds of such pairs, the authors show a universal volcano dependence of the sintering kinetics on the metal-support binding energy that can serve as a single descriptor to predict nanoparticle growth rates. The revealed scaling relations are a good start in the development of high-throughput screening computational approaches to drive discovery of sintering-resistant nanocatalysts. —YS

Science, abl9828, this issue p. 1360

VACCINES

Boosting adjuvant activity

Adjuvants can enhance immune responses against poorly immunogenic antigens, making them an important component of many vaccines despite the limited number currently approved for clinical use. Silva *et al.* developed a new adjuvant formed by the Toll-like receptor 4 agonist monophosphoryl lipid A (MPLA) self-assembled into nanoparticle structures with detergent-like saponin. Compared with a panel of tested adjuvants, including other saponin-based formulations, saponin–MPLA nanoparticles (SMNPs) induced superior antibody and germinal center responses after immunization with a poorly immunogenic HIV antigen in both mice and rhesus macaques. SMNPs increased lymphatic flow in a mast cell–dependent manner and increased antigen delivery to draining lymph nodes, suggesting mechanisms by which saponin adjuvants act to enhance humoral immune responses. —CO

Sci. Immunol. **6**, eabf1152 (2021).

CANCER

Growing tumor blood vessels, STAT

The co-transcriptional activators YAP/TAZ induce changes in gene expression in endothelial cells that promote tumor angiogenesis. Shen *et al.* sought to understand how YAP/TAZ are activated in colorectal cancers and melanomas. Cytokines abundant in the tumor microenvironment promoted the interaction of YAP/TAZ with the transcription factor STAT3. This interaction enabled the nuclear translocation of YAP/TAZ. Pharmacological inhibition or genetic ablation of YAP/TAZ or STAT3 in endothelial cells reduced the growth of colorectal cancers and melanomas in mice. —WW

Sci. Signal. **14**, eabj8393 (2021).

PHASE-CHANGE MEMORY

Single element switch

Phase-change materials are attractive for computer memory and switching, in part due to their small size and fast switching speeds. However, competitive materials frequently have many elements, which decreases the switching reliability. Shen *et al*. built a pure tellurium device that is capable of fast switching through a phase transformation (see the Perspective by Calarco and Arciprete). Unlike many other phase-change materials, the change in resistance happens because the tellurium melts during the switching process. The resulting device can be switched 100 million times before failure and is an appealing route for avoiding the issues from multi-element phase-change materials. —BG

Science, abi6332, this issue p. 1390; see also abm7316, p. 1321

CORONAVIRUS

A look at variant-specific boosters

The evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants of concern (VOCs) has raised the question of whether current COVID-19 vaccines protect against VOCs and if a variant specific vaccine may be needed. Of the currently identified VOCs, the Delta variant is believed to be the most transmissible, whereas the Beta variant appears to be the most vaccine resistant. Corbett *et al*. looked at the effect of vaccine boosting using either the original WA-1 strain vaccine or a Beta variant–specific booster. Around 6 months after the primary two-dose vaccine series, a third boost vaccination resulted in higher neutralizing antibody levels against VOCs in nonhuman primates. Regardless of whether the boost was from the original vaccine or the Beta-specific version, similar increases in neutralizing antibody levels were observed and resulted in enhanced viral protection. —PNK

Science, abl8912, this issue p. 1343

QUANTUM CHEMISTRY

Improving DFT with deep learning

In the past 30 years, density functional theory (DFT) has emerged as the most widely used electronic structure method to predict the properties of various systems in chemistry, biology, and materials science. Despite a long history of successes, state-of-the-art DFT functionals have crucial limitations. In particular, significant systematic errors are observed for charge densities involving mobile charges and spins. Kirkpatrick *et al*. developed a framework to train a deep neural network on accurate chemical data and fractional electron constraints (see the Perspective by Perdew). The resulting functional outperforms traditional functionals on thorough benchmarks for main-group atoms and molecules. The present work offers a solution to a long-standing critical problem in DFT and demonstrates the success of combining DFT with the modern machine-learning methodology. —YS

Science, abj6511, this issue p. 1385; see also abm2455, p. 1322

REVIEW SUMMARY

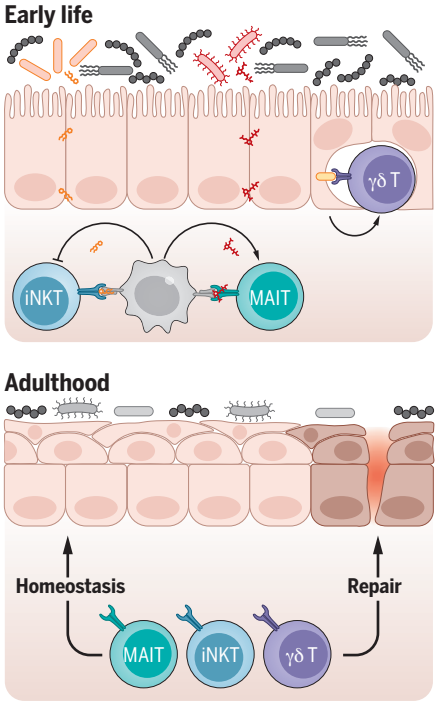
IMMUNOLOGY

Early-life imprinting of unconventional T cells and tissue homeostasis

Michael G. Constantinides* and Yasmine Belkaid*

BACKGROUND: In addition to providing antimicrobial immunity, the immune system governs numerous aspects of host physiology, in part through the coordinated action of tissue-resident lymphocytes. Although conventional T cells recognize peptides presented by polymorphic major histocompatibility complexes (MHCs), a large proportion of T cells within tissues are specific for modified peptides and small molecules. These unconventional T cells are restricted by monomorphic MHC molecules, many of which exhibit a high level of conservation across species, suggesting an essential role for these cells throughout evolution. Some unconventional T cells express semi-invariant T cell receptors (TCRs) that limit their antigenic range analogously to innate immune receptors. Termed innate-like T cells, these populations developmentally acquire effector characteristics prior to thymic egress, including rapid cytokine release and expression of chemokine receptors and integrins. Without the necessity of antigen-mediated activation within secondary lymphoid organs, innate-like T cells accumulate within tissues earlier than conventional effector T cells. Consequently, recent work has shown that these unconventional T cells are particularly responsive to early-life signals and promote the maintenance of tissue homeostasis as a coordinated network.

ADVANCES: As a consequence of their emergence in early life, unconventional T cells contribute to the initial dialog between the host and the microbiota, on which some subsets depend for their development. Exposure to commensal microbes during the first weeks of life imprints the abundance of mucosal-associated invariant T (MAIT) cells and invariant natural killer T (iNKT) cells in tissues, due to both TCR-mediated activation and modulation of tissue-specific chemokines. Epithelial cell expression of host molecules in early life is also necessary for the accumulation of a defined subset of γδ T cells. Absence of these early-life signals cannot be compensated for later in life and has long-lasting consequences for host physiology, rendering adult animals more susceptible to inflammation. Although the antimicrobial functions of unconventional T cells have been established, recent work has demonstrated their contributions to host metabolism, tissue repair, and the maintenance of homeostasis. In addition to mediating epithelial repair directly via the expression of growth factors, γδ T cells can also promote tissue homeostasis by inducing the release of alarmins, which recruit regulatory T cells and reparative type 2 innate lymphoid cells, as well as promoting thermoregulation within adipose tissue. Alarmins also enable a subset of unconventional $CD8^+$ T cells to release type 2 cytokines that promote wound healing. Additionally, MAIT cells express transcripts associated with tissue repair and contribute to healing, with commensal microbes and their metabolites enhancing these reparative functions. However, the response of MAIT cells and other unconventional T cells to microbial exposure is context-dependent, with the proinflammatory cytokines released during an infection inducing a distinct transcriptional profile characterized by effector functions, accompanied by diminished expression of tissue repair genes. Although unconventional T cells exhibit overlapping roles, their disparate stratification within epithelial tissues suggests that each population may serve a unique purpose in maintaining tissue homeostasis. Animal models deficient in subsets of unconventional T cells exhibit compensatory increases of other unconventional T cells, indicating that these populations constitute a network of redundant cellular mechanisms that contribute to tissue resilience. However, inflammation can abrogate the expression of epithelial molecules that maintain a subset γδ T cells, resulting in a compensatory increase in more inflammatory γδ T cells that predispose the tissue to chronic inflammation.



Early-life imprinting of unconventional T cells determines tissue physiology in adulthood. Unconventional T cells, including invariant natural killer T (iNKT) cells, mucosal-associated invariant T (MAIT) cells, and γδ T cells, develop in response to microbial and epithelial signals in early life and subsequently promote tissue repair and homeostasis as a coordinated network.

OUTLOOK: Although the causes of competition between unconventional T cells remain to be determined, the demonstrated contributions of these populations to barrier immunity and tissue homeostasis provide opportunities for novel therapeutics. Indeed, the abundance of unconventional T cells from infancy and their recognition of conserved microbial products could be exploited to combat early-life infections, even with the increasing prevalence of antibiotic resistance. Although the microbial responsiveness of unconventional T cells may enable the development of pre/probiotics that modulate their function and use for the prevention of dysbiosis, the dependence of some subsets on early-life microbial interactions suggests that antibiotic use and malnutrition during this period may impart lifelong consequences for immunity. Furthermore, although the ability of unconventional T cells to promote epithelial growth and angiogenesis is beneficial in the context of tissue repair, these functions can also exacerbate tumorigenesis. Therefore, the development of safe and effective treatments requires an understanding of the unique intricacies of unconventional T cells. We discuss the identification of signals responsible for the development of unconventional T cells and their accumulation within tissues. Additionally, we examine how alterations in early-life signals or inflammatory challenges can imprint the unconventional T cell network and affect tissue homeostasis for the long term. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: constantinides@scripps.edu (M.G.C.); ybelkaid@niaid.nih.gov (Y.B.)
Cite this article as M. G. Constantinides and Y. Belkaid, *Science* 374, eabf0095 (2021). DOI: 10.1126/science.abf0095

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abf0095

REVIEW

IMMUNOLOGY

Early-life imprinting of unconventional T cells and tissue homeostasis

Michael G. Constantinides[1]*† and Yasmine Belkaid[1,2]*

Unconventional T cells—including invariant natural killer T (iNKT) cells, mucosal-associated invariant T (MAIT) cells, and defined subsets of γδ T cells—are restricted by monomorphic major histocompatibility complex class Ib (MHC-Ib) molecules and seed tissues during development. Early-life instructive signals, including those derived from the microbiota, establish homeostatic set points for unconventional T cells, a phenomenon that has lifelong consequences for the regulation of tissue immunity, inflammation, and repair. Unconventional T cells compete for niches within tissues, and recent evidence supports the idea that the fundamental role of these cells in tissue physiology may result from their action as a network with overlapping and potentially synergistic functions, rather than as individual subsets.

The immune system, previously appreciated for its role in antimicrobial protection, is now known to also govern numerous aspects of host physiology (*1*). Part of this control occurs through the coordinated action of tissue-resident lymphocytes, including unconventional T cells that accumulate in response to specific developmental cues and exposure to defined antigens. Although conventional T cells recognize peptides presented by polymorphic major histocompatibility complex (MHC) proteins, a large proportion of T cells within tissues are specific for modified peptides and small molecules, including defined lipids and metabolites produced by microbes (*2*). These unconventional T cells are restricted by monomorphic MHC class Ib (MHC-Ib) molecules; some populations, such as invariant natural killer T (iNKT) cells, mucosal-associated invariant T (MAIT) cells, and subsets of γδ T cells, express a semi-invariant T cell receptor (TCR) that limits their antigenic range analogously to an innate immune receptor (*2*) (Table 1). Termed innate-like T cells, these populations developmentally acquire effector characteristics prior to thymic egress, including rapid cytokine release and expression of chemokine receptors and integrins, which results in their accumulation in tissues earlier than conventional effector T cells, during the second trimester in humans and within 3 weeks of birth in mice (*3*–*8*). This dependence on early-life signals raises fundamental questions regarding the extent to which the early emergence of unconventional T cells may provide critical functions before the establishment of classical immunologic memory. The evolutionary conservation of many MHC-Ib molecules suggests that unconventional T cells may serve a critical role in host physiology, particularly in mediating the interaction with the microbiota and maintaining tissue homeostasis (Fig. 1). Recent evidence also supports the idea that unconventional T cells compete for niches within tissues and may exert overlapping functions associated with the preservation of tissue homeostasis. We discuss the identification of signals responsible for the development of unconventional T cells and their accumulation within tissues, and how alterations in early-life signals or inflammatory challenges can imprint the unconventional T cell network and affect tissue homeostasis for the long term.

[1]Metaorganism Immunity Section, Laboratory of Immune System Biology, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA. [2]NIAID Microbiome Program, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA.
***Corresponding author. Email: constantinides@scripps.edu (M.G.C.); ybelkaid@niaid.nih.gov (Y.B.)**
†Present address: Department of Immunology and Microbiology, Scripps Research, La Jolla, CA 92037, USA.

Dependence on early-life signals: Microbiota and epithelially derived signals

Barrier tissues are the primary route of exposure to the microbiota and are enriched in unconventional T cells (Fig. 2). The ability of MHC-Ib molecules to present antigens with specific chemical moieties or amino acid sequences, which can be derived from a large constituency of the microbiota, places immune cells that recognize them as ideal candidates for the constitutive sensing and recognition of microbial antigens. Indeed, recent evidence suggests that unconventional T cells, in addition to using microbiota-derived metabolites for their development and function, may also contribute to the beneficial dialog between the microbiota and its host.

Innate-like T cell subsets acquire effector characteristics prior to thymic egress, which results in their accumulation in tissues during early life (*3*–*8*). Studies of iNKT cells provided the first demonstration that early-life exposure to the microbiota controls the development and tissue accumulation of unconventional T cells. iNKT cells express a semi-invariant TCR α-chain that enables them to recognize both endogenous and microbial lipids presented by CD1d (*9*). Although iNKT cells can be selected on self-ligands (*10*), they are hyporesponsive in the absence of commensals, and germ-free (GF) mice exhibit fewer iNKT cells in the thymus, spleen, and liver (*11*, *12*). However, iNKT cells accumulate within the colonic lamina propria and lungs of GF mice as a consequence of elevated expression of the chemokine CXCL16 by epithelial cells, exacerbating colonic inflammation and allergic airway responses (*13*). Although the *Cxcl16* gene is hypermethylated in the absence of commensals, colonization of neonatal GF mice diminishes DNA methylation, abrogating CXCL16 production and preventing the subsequent recruitment of iNKT cells (*13*). Conversely, microbial

Table 1. Unconventional T cells. Populations of unconventional T cells present in humans (h) and mice (m) are listed, with innate-like T cell subsets in boldface. DETC, dendritic epidermal T cell; IEL, intraepithelial lymphocyte; GEM, germline-encoded mycolyl-reactive.

Population	MHC-Ib	Antigen	TCR repertoire
γδ17 T	**Unknown**	**Unknown**	**Vγ4 or Vγ6 (m)**
DETC (m)	**Unknown**	**Unknown**	**Vγ5 Vδ1 (m)**
γδ IEL	**Unknown**	**Butyrophilin-like molecules**	**Vγ4 (h), Vγ7 (m)**
MAIT	**MR1**	**Microbial metabolites**	**Vα7.2 Jα33/20/12 (h)** **Vα19 Jα33 (m)**
iNKT	**CD1d**	**Glycolipids**	**Vα24 Jα18 (h)** **Vα14 Jα18 (m)**
GEM (h)	**CD1b**	**Glycolipids**	**Vα7.2 Jα9 (h)**
T_C1	Qa1 (m)	Peptides	Polyclonal
Type II NKT	CD1d	Glycolipids	Polyclonal
$T_C1/17$	H2-M3 (m)	*N*-formylated peptides	Polyclonal

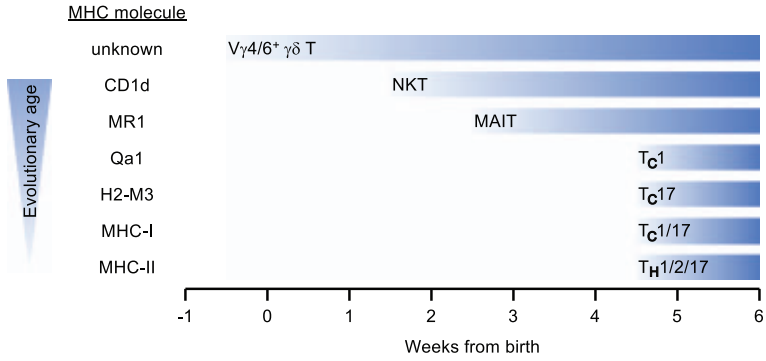


Fig. 1. Emergence of unconventional T cells. The emergence of unconventional T cells during murine development correlates with the evolutionary age of the MHC molecules they recognize.

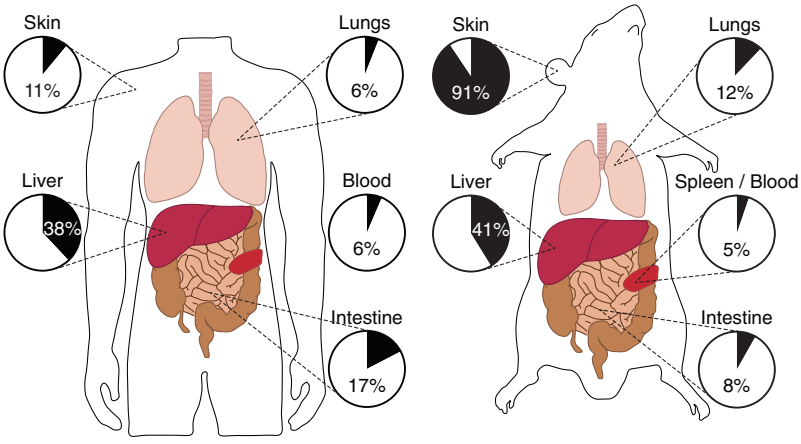


Fig. 2. Frequency of unconventional T cells. Unconventional T cells are enriched in barrier tissues and organs that drain these sites, such as the liver. The cumulative frequency of MAIT, iNKT, and γδ T cells among all T cells is shown for the indicated tissues of humans and mice.

exposure 5 weeks after birth is insufficient to alter the methylation of *Cxcl16* and minimize the accumulation of iNKT cells (*13*). Thus, epithelial cells are permissive to epigenetic modifications in early life that consequently have long-lasting effects on the abundance of iNKT cells in barrier tissues and predispose animals to immune pathologies.

In addition to imprinting the frequency of iNKT cells through the regulation of chemokines, the microbiota also provides antigens that can directly modulate the development of iNKT cells during the first 2 weeks of life. The intestinal symbiont *Bacteroides fragilis* expresses sphingolipids that inhibit iNKT activation by preventing CD1d-mediated presentation of agonistic lipids (*3*), including endogenous ligands that mediate iNKT selection in the absence of commensals. As such, early-life exposure to *B. fragilis* prevents the accumulation of iNKT cells within the colon, but not the lungs, of the progeny, indicating that the inhibitory effects of the sphingolipids are localized to the predominant site of *B. fragilis* colonization (*3*). Although these animals are protected from experimentally induced colitis once they become adults, colonization of GF mice at 2 weeks of age is insufficient to inhibit iNKT cell proliferation, resulting in a higher frequency of these proinflammatory T cells and rendering the adult animals more susceptible to inflammation (*3*).

As well as inhibiting the accumulation of iNKT cells in the colon via sphingolipids, *B. fragilis* also regulates the abundance of other unconventional T cells through another microbial molecule, polysaccharide A (PSA). Using the transcription factor PLZF to identify innate and innate-like lymphocytes (*14*, *15*), *B. fragilis* colonization was found to increase the frequency and number of $PLZF^+$ innate-like αβ T cells and $PLZF^+$ innate lymphocytes ($TCR\beta^-TCR\gamma\delta^-$) within the neonatal thymus in a PSA-dependent manner (*16*). Although antibiotic treatment of adult mice did not alter the abundance of $PLZF^+$ innate-like cells within the thymus, mice treated with antibiotics until weaning exhibited a reduced number of these thymic lymphocytes as adults and were more susceptible to experimentally induced colitis (*16*). Thus, microbial colonization during the neonatal period imprints the abundance of iNKT cells in barrier tissues, which can have enduring effects on host physiology.

In contrast to mice, where iNKT cells are abundant, the predominant unconventional αβ T cells in humans are MAIT cells, which recognize intermediates of vitamin B2 (riboflavin) synthesis presented by MR1 (*17*). Because riboflavin synthesis is broadly conserved among bacteria and fungi, MAIT cells respond to a diverse array of microbes, including known commensals (*4*, *18*). Genetically identical mice housed in distinct cages exhibit a striking variability in MAIT cell abundance (*4*), which is suggestive of a dependence on the microbiota and parallels the variability in MAIT cell frequency between human individuals (*19*, *20*). In further support of a role for the microbiota, mice devoid of live microbes have fewer thymic $CD24^+CD44^-$ MAIT cells (*21*). Additionally, these immature MAIT cells express reduced levels of Nur77, a transcription factor that correlates with TCR signal strength, indicating that microbial antigens promote thymic selection of MAIT cells (*21*). Consequently, MAIT cells are nearly absent from peripheral tissues in animals that are either GF or colonized with commensals that do not synthesize riboflavin (*4*, *21*–*23*). Among all lymphocyte subsets, MAIT cells are the most affected by the absence of the microbiota (*4*). In conventionally housed mice, MAIT cells accumulate in barrier tissues before weaning, and neonatal colonization of GF animals with riboflavin-synthesizing intestinal commensals results in a comparable number of MAIT cells at these sites (*4*, *21*). However, although colonization of adult GF mice promotes the maturation of MAIT cells in the thymus, microbial exposure later in life is insufficient to induce MAIT cells in peripheral tissues (*4*, *21*), indicating that MAIT cell abundance in tissues is imprinted by the microbiota in early life. Topical application of purified riboflavin intermediates on neonatal, but not adult, GF mice also results in the maturation of MAIT cells in the thymus and their accumulation in tissues (*4*, *21*). Thus, these ligands are sufficient to induce MAIT cells in the absence of additional microbial antigens and can derive from commensals at other barrier surfaces, including the skin. The systemic effects of riboflavin derivatives on MAIT cell development parallel the distal impact of other microbial metabolites, including short-chain fatty acids, aryl hydrocarbon receptor ligands, polyamines, and PSA (*24*). In utero transfer of these metabolites from the maternal microbiota would explain the seemingly paradoxical presence of functionally mature MAIT cells in human fetal tissue before microbial colonization (*5*). Enterobacteriaceae, a family of Gram-negative bacteria

that are enriched for riboflavin synthesis genes (*4*), are enriched during the last trimester of pregnancy in humans and colonize both human and mouse neonates from birth (*25*–*28*), leading to an abundance of MAIT cell antigens during early-life development.

Additional microbial metabolites may also regulate the development of MAIT cells because MR1 has also been shown to present antigens that do not derive from riboflavin biosynthesis (*29*, *30*). Because MAIT cells have recently been shown to contribute to both tissue repair and antimicrobial immunity (*4*, *31*–*34*), the microbial imprinting of MAIT cell frequency in early life would be expected to have long-term implications for tissue homeostasis.

Although the vast majority of unconventional T cells are tissue-resident, some subsets can acquire migratory capabilities under defined settings—a property that may enable these cells to mediate the distal and long-term impact of the microbiota. For instance, broad-spectrum antibiotic treatment during the neonatal period has been shown to increase the severity of experimentally induced psoriasis in adult mice due to persistent intestinal dysbiosis (*35*). The resulting loss of microbial diversity corresponds with an increase in the frequency of interleukin-22 (IL-22)–producing $V\gamma4^+$ γδ T cells within the skin [Heilig and Tonegawa nomenclature (*36*)], which is necessary for disease pathology (*35*). Conversely, antibiotic treatment of adult animals ameliorates psoriasis and reduces the frequency of IL-$22^+V\gamma4^+$ γδ T cells (*35*), supporting the idea that the early-life microbiota is uniquely susceptible to external perturbations that can imprint long-term microbial dysbiosis and promote proinflammatory unconventional T cells. More generally, both experimental and clinical observations support the idea that early-life alterations in microbiota can have a detrimental impact on the host (*37*). However, the extent to which maladapted early interactions between unconventional T cells and the microbiota predispose to disease remains an open question.

In addition to early-life microbial signals that regulate unconventional T cells, expression of host molecules during defined developmental periods can shape local T cell populations. Within barrier tissues in particular, epithelial cells have been shown to provide early instructive signals for defined subsets of unconventional T cells. For instance, epithelial cells display butyrophilin-like (Btnl) molecules such as Btnl1 that induce TCR-dependent stimulation of intraepithelial lymphocyte (IEL) γδ T cells (*38*). Although the induction of Btnl1 on enterocytes is sufficient to expand the intestinal $V\gamma7^+$ IEL population, expression of Btnl1 after 7 weeks of age does not alter their abundance (*38*), demonstrating an early-life window for epithelial cell imprinting of unconventional T cells. Despite the abundance of unconventional T cells within barrier tissues, much remains to be understood about the role of epithelial and stromal cells in providing early-life instructive signals.

Competition between unconventional T cells: A cellular network

Unconventional T cells have been often explored one subset at a time, but recent evidence supports the idea that these cells should be assessed as a network of cells with potentially overlapping and/or redundant functions. Indeed, although unconventional T cells respond to disparate antigens, they express remarkably similar transcriptional programs (*39*), resulting in an inherent competition that results in the maintenance of a relatively stable level of effector cells within tissues.

The ability of unconventional T cells to compete for defined niches has been shown in various tissues. For instance, mice deficient in microRNAs miR-181a and miR-181b exhibit a severe reduction of iNKT cells and a corresponding increase of $NK1.1^+$ γδ T cells in the liver (*40*), where iNKT cells represent a substantial population. Although iNKT cells are nearly absent from all organs in these animals (*41*), the compensatory expansion of γδ T cells does not occur in the thymus or lymph nodes (*40*), which suggests that this interplay between unconventional T cells may only occur in tissues where they are abundant. Additionally, hepatic $NK1.1^+$ γδ T cells express type 1 markers, including the IL-12 receptor and interferon-γ (IFN-γ) (*42*), whereas the predominant iNKT subset within the liver is the type 1 iNKT1 population (*9*). Thus, these populations likely compete for the same trophic or survival factors, leading to an inverse correlation between them. In the dermis, where γδ T cells are the predominant type 17 effector cells in mice, both MAIT cells and iNKT cells are increased in *Tcrd*$^{-/-}$ animals (*4*). However, cutaneous MAIT cells and γδ T cells are unaffected by the lack of iNKT cells in CD1d-deficient mice (*4*), likely because of the comparatively low number of iNKT cells in the skin. By contrast, in the thymus, spleen, and liver, where iNKT cells account for a large proportion of effector T cells, *Cd1d*$^{-/-}$ mice show a compensatory increase of MAIT cells (*22*). The thymic MAIT cells that expand are at the mature $CD44^+$ developmental stage (*22*), which coincides with the expression either RORγt or T-bet and the associated effector molecules and receptors (*43*); these findings support the hypothesis that the competition between unconventional T cells may be mediated by cytokines. Recently, the intrinsic ability of these populations to compete was confirmed in a patient harboring a homozygous point mutation in *MR1* associated with an absence of MAIT cells and concomitant expansion in γδ T cells (*44*). Experimentally, cytokines such as IL-23 and IL-1 have been proposed to contribute to this phenomenon, but for the most part, the nature of the trophic or survival factors responsible for the tissue-specific composition and frequency of unconventional T cells remains poorly understood (Fig. 3).

Although competition exists among unconventional T cells, each subset may serve distinct roles in promoting tissue physiology as a result of inherent differences in localization. For example, cutaneous $V\gamma5^+V\delta1^+$ dendritic epidermal T cells (DETCs) are restricted to the epidermis, whereas $V\gamma4^+$ and $V\gamma6^+$ γδ T cells reside in the dermis with conventional αβ T cells (*45*) and MAIT cells are positioned below the dermal-epidermal interface (*4*), termed the basement membrane. However, the signals that regulate the stratification of unconventional T cells within tissues have not been established, nor is it known whether this localization is necessary for their function.

The extent to which the competition among unconventional T cells contributes to disease susceptibility has not been established. Indeed, infection or chronic infection may displace tissue-resident unconventional T cells, opening the door to an expansion of inappropriate competitors. For example, a recent study revealed that subsets of γδ T cells can also compete with one another under inflammatory conditions. The healthy human intestine contains $V\gamma4^+V\delta1^+$ IELs that recognize the butyrophilin-like molecules BTNL3 and BTNL8 and promote tissue homeostasis though their cytotoxic and tissue repair programs (*38*, *46*). However, the chronic inflammation caused by celiac disease abrogates BTNL8 expression on epithelial cells, resulting in loss of $V\gamma4^+V\delta1^+$ IELs and a corresponding expansion of inflammatory $V\delta1^+$ IELs bearing an H-J1 CDR3γ motif and producing more IFN-γ (*47*). Although the removal of dietary gluten restores BTNL8 expression, it is insufficient to reconstitute the $V\gamma4^+V\delta1^+$ IEL population, indicating that chronic inflammation can permanently alter the tissue milieu. To what extent alterations in the repertoire and frequencies of unconventional T cells cause predisposition to inflammatory states remains to be determined.

The ability of unconventional T cells to compete within tissues may have, in some settings, hindered our ability to understand the precise role of individual subsets. This phenomenon also raises an intriguing hypothesis that the ability of unconventional T cells to compensate for each other contributes to tissue resilience by preserving essential tissue functions.

Tissue repair and maintenance of homeostasis

Although the antimicrobial functions of unconventional T cells have been established (*2*), emerging evidence indicates that one of the

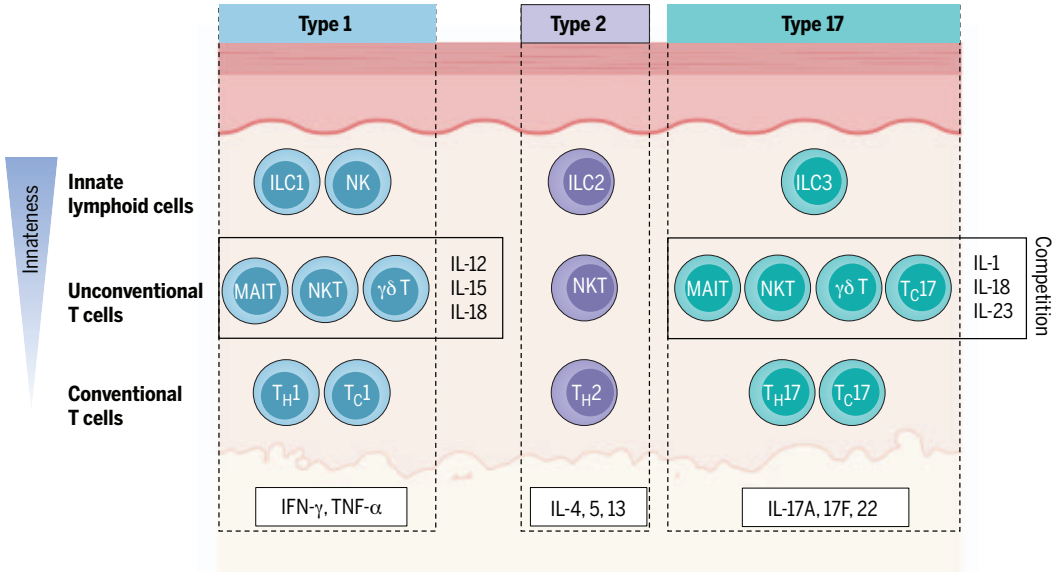


Fig. 3. Unconventional T cells within tissues. Shown are the different populations of effector lymphocytes present in epithelial tissues, organized by their effector molecules (columns) and the extent of their innate characteristics (rows). As a consequence of their similar transcriptional programs, unconventional T cells express overlapping cytokine receptors, resulting in competition for interleukins. Animal models deficient in subsets of unconventional T cells exhibit compensatory increases of other unconventional T cells, indicating that these populations constitute a network of redundant cellular mechanisms that contribute to tissue resilience.

dominant functions of these cells may be associated with the preservation or restoration of tissue homeostasis. Such a property can be, at least in part, explained by the ability of unconventional T cells to rapidly respond not only to defined ligands, but also to alarmins such as IL-18 or IL-33 that are produced in the context of tissue damage (*48*).

Intestinal γδ IELs and DETCs produce fibroblast growth factors, which promote the proliferation and differentiation of epithelial cells, returning these tissues to homeostasis after damage (*49*, *50*). Upon injury, epidermal keratinocytes at the wound edge up-regulate the expression of butyrophilin-like molecules encoded by *Skint* genes (*51*), which are recognized by DETCs in a TCR-dependent manner (*52*). Expression of *Skint* genes is diminished in the skin of aged mice, resulting in fewer DETCs at the site of the wound relative to young mice and, consequently, delayed wound closure (*51*). After activation, DETCs in close proximity to the wound lose their characteristic dendritic morphology and express insulin-like growth factor 1 (IGF-1), which reduces keratinocyte apoptosis (*50*, *53*). Although both αβ and γδ T cells from acutely wounded human tissue express IGF-1, T cells from chronic wounds do not and are refractory to TCR stimulation (*54*). This suggests that chronic wounds may persist in part because of the unresponsiveness of cutaneous T cells.

In addition to mediating epithelial repair directly via the expression of growth factors, γδ T cells can also promote tissue homeostasis through the recruitment of regulatory T (T_{reg}) cells and type 2 innate lymphoid cells (ILC2s). Influenza infection of neonatal mice results in the accumulation of $CD27^{-}$ γδ T cells that produce IL-17, which is sufficient to induce IL-33 expression by lung epithelial cells (*55*). Deficiency in either γδ T cells or IL-33 abrogates the accumulation of ILC2s and T_{reg} cells and their production of amphiregulin, resulting in decreased survival after infection (*55*). Although *Tcrd*$^{-/-}$ mice exhibit increased inflammation, their viral titers are comparable to wild-type controls (*55*), supporting the idea that γδ T cells may provide protection to influenza by promoting tissue repair rather than enhancing antiviral immunity. Additionally, $PLZF^{+}V\gamma6^{+}$ γδ T cells within adipose tissue promote the accumulation of $ST2^{+}$ T_{reg} cells by producing IL-17 and tumor necrosis factor, which induce IL-33 release from adipose stromal cells (*56*). Because IL-33 promotes adaptive thermogenesis by inducing expression of uncoupling protein 1 (UCP1) in brown and beige adipocytes (*57*), γδ T cells also contribute to thermoregulation. The frequency of γδ T cells within adipose tissue increases after cold exposure, and in the absence of γδ T cells or IL-17A, mice exhibit decreased UCP1 expression and a greater loss of body temperature (*56*), indicating that γδ T cells respond to temperature fluctuations and maintain homeostasis. More recently, γδ T cells have been shown to contribute to the adaptation of gut epithelial cells to nutrient sensing (*58*). Notch signaling was found to contribute to the activation of γδ T cells in this context, but it is important to highlight that in most settings, the molecular mechanisms by which unconventional T cells are activated and contribute to host physiological responses remain largely unknown.

Cutaneous MAIT cells are positioned below the basement membrane, which serves as a scaffold for the migration and attachment of progenitor cells during wound healing (*59*). MAIT cells are also found along the basement membrane in the human colon (*32*), which suggests that this may be a characteristic of MAIT cells within epithelial tissues. In addition to their unique localization, MAIT cells express a tissue repair transcriptional program, which is enhanced by TCR-mediated stimulation alone (*4*, *31*, *32*). However, when activated in the presence of stimulatory cytokines, MAIT cells acquire a distinct transcriptional profile characterized by effector functions, with diminished expression of tissue repair genes (*32*). Similarly, in response to topically applied *Staphylococcus epidermidis*, a skin commensal that does not induce inflammation in this manner (*60*), cutaneous MAIT cells up-regulate the expression of genes associated with angiogenesis and tissue repair, including angiogenin and the growth factors IGF-1 and HGF (*4*). In accordance with the notion that TCR activation promotes the tissue repair program of MAIT cells, topical application of the stimulatory riboflavin derivative 5-OP-RU is sufficient to increase re-epithelialization of cutaneous excision wounds in wild-type mice. By contrast, MAIT cell–deficient animals exhibit less wound healing, indicating that MAIT cells contribute to tissue repair in vivo (*4*).

In iNKT cell–deficient *Traj18*$^{-/-}$ mice, cutaneous wound healing is delayed relative to wild-type controls, in part due to decreased collagen deposition and increased neutrophil infiltration (*61*, *62*). However, the *Traj18*$^{-/-}$ mice used in these studies were recently found to lack MAIT cells in addition to iNKT cells because the null allele abrogates expression of other Jα genes, including the invariant *Traj33* used by the murine MAIT cell TCR (*63*, *64*), making it impossible to distinguish the contributions of these unconventional T cell populations in this strain. Antibody blockade of CD1d decreases iNKT cell frequency within cutaneous wounds and promotes wound

closure, which was interpreted to indicate that iNKT cells inhibit tissue repair (*65*). However, the recently described competition among unconventional T cells could mean that prevention of CD1d-mediated iNKT activation and accumulation after injury may increase the abundance of other unconventional T cells that have been shown to promote wound healing, including MAIT cells and γδ T cells. Thus, it remains to be determined whether iNKT cells also participate in tissue repair.

Although T cells that recognize *N*-formylated peptides presented by the MHC-Ib molecule H2-M3 display a diverse TCR repertoire (*66*), they exhibit innate-like effector characteristics and have recently been shown to promote tissue repair (*67*–*69*). After topical application of *S. epidermidis*, H2-M3–restricted RORγt$^+$ CD8$^+$ T cells preferentially accumulate in the epidermis, where they express immunoregulatory and tissue repair transcriptional signatures (*60*, *68*). The transcriptional program of the H2-M3–restricted RORγt$^+$ CD8$^+$ T cells closely resembles that of MAIT cells after the resolution of a *Legionella* infection, which also display tissue repair genes (*31*). Conversely, after an inflammatory intradermal inoculation of *S. epidermidis*, H2-M3–restricted T-bet$^+$ CD8$^+$ T cells result (*60*, *69*), which lack the immunoregulatory and tissue repair characteristics of the RORγt$^+$ cells (*68*). Because homeostatic interactions with the microbiota are necessary for the induction of a tissue repair signature in H2-M3–restricted T cells, robust cytokine-mediated activation likely induces effector characteristics analogously to MAIT cells (*32*). After a cutaneous injury, H2-M3–restricted RORγt$^+$ CD8$^+$ T cells accumulate at the wound edge, where damage-associated alarmins, and in particular IL-18, promote their production of type 2 cytokines, which contribute to the re-epithelialization of the wound and thus restore tissue homeostasis (*68*, *69*).

Growing evidence indicates that the role of unconventional T cells is highly dependent on their tissue localization, host developmental stage, and disease context. Indeed, although the ability of unconventional T cells to promote epithelial growth and angiogenesis is beneficial in the context of tissue repair, these functions can also exacerbate tumorigenesis. For instance, the production of IL-17 by lung-resident Vγ6$^+$Vδ1$^+$ T cells in response to the local microbiota promotes neutrophil infiltration into lung adenocarcinomas and increases tumor development (*70*). Conversely, γδ T cells and other unconventional T cells also exhibit antitumor responses (*71*). MAIT cells were recently shown to target myeloma cells pulsed with the riboflavin derivative 5-OP-RU, and an MR1-restricted T cell clone from human peripheral blood was found to kill a broad range of cancer cells (*72*, *73*). Microbiota-mediated conversion of bile acids regulates the accumulation of iNKT cells in the liver, which inhibit hepatic tumor growth (*74*). Furthermore, unconventional T cells make attractive targets for chimeric antigen receptor T cell therapy because their restriction by monomorphic MHC-Ib molecules avoids the possibility of graft-versus-host disease (*71*).

Because of the complexity and interrelated functions of unconventional T cells, this evolutionarily ancient arm of the immune system is expected to contribute extensively to the control of both host physiology and disease states. As such, leveraging the ability of these tissue-resident cells to rapidly respond to canonical antigens, alarmins, and/or common survival factors may represent an important and highly physiological therapeutic strategy.

REFERENCES AND NOTES

1. L. C. Rankin, D. Artis, Beyond Host Defense: Emerging Functions of the Immune System in Regulating Complex Tissue Physiology. *Cell* **173**, 554–567 (2018). doi: 10.1016/j.cell.2018.03.013; pmid: 29677509
2. D. I. Godfrey, A. P. Uldrich, J. McCluskey, J. Rossjohn, D. B. Moody, The burgeoning family of unconventional T cells. *Nat. Immunol.* **16**, 1114–1123 (2015). doi: 10.1038/ni.3298; pmid: 26482978
3. D. An *et al*., Sphingolipids from a symbiotic microbe regulate homeostasis of host intestinal natural killer T cells. *Cell* **156**, 123–133 (2014). doi: 10.1016/j.cell.2013.11.042; pmid: 24439373
4. M. G. Constantinides *et al*., MAIT cells are imprinted by the microbiota in early life and promote tissue repair. *Science* **366**, eaax6624 (2019). doi: 10.1126/science.aax6624; pmid: 31649166
5. E. Leeansyah, L. Loh, D. F. Nixon, J. K. Sandberg, Acquisition of innate-like microbial reactivity in mucosal tissues during human fetal MAIT-cell development. *Nat. Commun.* **5**, 3143 (2014). doi: 10.1038/ncomms4143; pmid: 24452018
6. J. D. Haas *et al*., Development of interleukin-17-producing γδ T cells is restricted to a functional embryonic wave. *Immunity* **37**, 48–59 (2012). doi: 10.1016/j.immuni.2012.06.003; pmid: 22770884
7. K. Narayan *et al*., Intrathymic programming of effector fates in three molecularly distinct γδ T cell subtypes. *Nat. Immunol.* **13**, 511–518 (2012). doi: 10.1038/ni.2247; pmid: 22473038
8. P. Tieppo *et al*., The human fetal thymus generates invariant effector γδ T cells. *J. Exp. Med.* **217**, e20190580 (2020). doi: 10.1084/jem.20190580; pmid: 31816633
9. M. G. Constantinides, A. Bendelac, Transcriptional regulation of the NKT cell lineage. *Curr. Opin. Immunol.* **25**, 161–167 (2013). doi: 10.1016/j.coi.2013.01.003; pmid: 23402834
10. A. Bendelac, P. B. Savage, L. Teyton, The biology of NKT cells. *Annu. Rev. Immunol.* **25**, 297–336 (2007). doi: 10.1146/annurev.immunol.25.022106.141711; pmid: 17150027
11. B. Wei *et al*., Commensal microbiota and CD8+ T cells shape the formation of invariant NKT cells. *J. Immunol.* **184**, 1218–1226 (2010). doi: 10.4049/jimmunol.0902620; pmid: 20048124
12. G. Wingender *et al*., Intestinal microbes affect phenotypes and functions of invariant natural killer T cells in mice. *Gastroenterology* **143**, 418–428 (2012). doi: 10.1053/j.gastro.2012.04.017; pmid: 22522092
13. T. Olszak *et al*., Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* **336**, 489–493 (2012). doi: 10.1126/science.1219328; pmid: 22442383
14. M. G. Constantinides, B. D. McDonald, P. A. Verhoef, A. Bendelac, A committed precursor to innate lymphoid cells. *Nature* **508**, 397–401 (2014). doi: 10.1038/nature13047; pmid: 24509713
15. A. K. Savage *et al*., The transcription factor PLZF directs the effector program of the NKT cell lineage. *Immunity* **29**, 391–403 (2008). doi: 10.1016/j.immuni.2008.07.011; pmid: 18703361
16. M. Ennamorati *et al*., Intestinal microbes influence development of thymic lymphocytes in early life. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 2570–2578 (2020). doi: 10.1073/pnas.1915047117; pmid: 31964813
17. M. G. Constantinides, Interactions between the microbiota and innate and innate-like lymphocytes. *J. Leukoc. Biol.* **103**, 409–419 (2018). doi: 10.1002/JLB.3RI0917-378R; pmid: 29345366
18. L. Le Bourhis *et al*., Antimicrobial activity of mucosal-associated invariant T cells. *Nat. Immunol.* **11**, 701–708 (2010). doi: 10.1038/ni.1890; pmid: 20581831
19. G. Ben Youssef *et al*., Ontogeny of human mucosal-associated invariant T cells and related T cell subsets. *J. Exp. Med.* **215**, 459–479 (2018). doi: 10.1084/jem.20171739; pmid: 29339446
20. P. Chen *et al*., Circulating Mucosal-Associated Invariant T Cells in a Large Cohort of Healthy Chinese Individuals From Newborn to Elderly. *Front. Immunol.* **10**, 260 (2019). doi: 10.3389/fimmu.2019.00260; pmid: 30838000
21. F. Legoux *et al*., Microbial metabolites control the thymic development of mucosal-associated invariant T cells. *Science* **366**, 494–499 (2019). doi: 10.1126/science.aaw2719; pmid: 31467190
22. H. F. Koay *et al*., A three-stage intrathymic development pathway for the mucosal-associated invariant T cell lineage. *Nat. Immunol.* **17**, 1300–1311 (2016). doi: 10.1038/ni.3565; pmid: 27668799
23. E. Treiner *et al*., Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. *Nature* **422**, 164–169 (2003). doi: 10.1038/nature01433; pmid: 12634786
24. M. G. Rooks, W. S. Garrett, Gut microbiota, metabolites and host immunity. *Nat. Rev. Immunol.* **16**, 341–352 (2016). doi: 10.1038/nri.2016.42; pmid: 27231050
25. M. C. Arrieta, L. T. Stiemsma, N. Amenyogbe, E. M. Brown, B. Finlay, The intestinal microbiome in early life: Health and disease. *Front. Immunol.* **5**, 427 (2014). doi: 10.3389/fimmu.2014.00427; pmid: 25250028
26. J. L. Combellick *et al*., Differences in the fecal microbiota of neonates born at home or in the hospital. *Sci. Rep.* **8**, 15660 (2018). doi: 10.1038/s41598-018-33995-7; pmid: 30353125
27. D. M. Chu *et al*., Maturation of the infant microbiome community structure and function across multiple body sites and in relation to mode of delivery. *Nat. Med.* **23**, 314–326 (2017). doi: 10.1038/nm.4272; pmid: 28112736
28. O. Koren *et al*., Host remodeling of the gut microbiome and metabolic changes during pregnancy. *Cell* **150**, 470–480 (2012). doi: 10.1016/j.cell.2012.07.008; pmid: 22863002
29. M. J. Harriff *et al*., MR1 displays the microbial metabolome driving selective MR1-restricted T cell receptor usage. *Sci. Immunol.* **3**, eaao2556 (2018). doi: 10.1126/sciimmunol.aao2556; pmid: 30006464
30. A. N. Keller *et al*., Drugs and drug-like molecules can modulate the function of mucosal-associated invariant T cells. *Nat. Immunol.* **18**, 402–411 (2017). doi: 10.1038/ni.3679; pmid: 28166217
31. T. S. C. Hinks *et al*., Activation and In Vivo Evolution of the MAIT Cell Transcriptome in Mice and Humans Reveals Tissue Repair Functionality. *Cell Rep.* **28**, 3249–3262.e5 (2019). doi: 10.1016/j.celrep.2019.07.039; pmid: 31533045
32. T. Leng *et al*., TCR and Inflammatory Signals Tune Human MAIT Cells to Exert Specific Tissue Repair and Effector Functions. *Cell Rep.* **28**, 3077–3091.e5 (2019). doi: 10.1016/j.celrep.2019.08.050; pmid: 31533032
33. H. Wang *et al*., MAIT cells protect against pulmonary Legionella longbeachae infection. *Nat. Commun.* **9**, 3350 (2018). doi: 10.1038/s41467-018-05202-8; pmid: 30135490
34. B. van Wilgenburg *et al*., MAIT cells contribute to protection against lethal influenza infection in vivo. *Nat. Commun.* **9**, 4706 (2018). doi: 10.1038/s41467-018-07207-9; pmid: 30413689
35. P. Zanvit *et al*., Antibiotics in neonatal life increase murine susceptibility to experimental psoriasis. *Nat. Commun.* **6**, 8424 (2015). doi: 10.1038/ncomms9424; pmid: 26416167
36. J. S. Heilig, S. Tonegawa, Diversity of murine γ genes and expression in fetal and adult T lymphocytes. *Nature* **322**, 836–840 (1986). doi: 10.1038/322836a0; pmid: 2943999
37. S. Tamburini, N. Shen, H. C. Wu, J. C. Clemente, The microbiome in early life: Implications for health outcomes. *Nat. Med.* **22**, 713–722 (2016). doi: 10.1038/nm.4142; pmid: 27387886
38. R. Di Marco Barros *et al*., Epithelia Use Butyrophilin-like Molecules to Shape Organ-Specific γδ T Cell Compartments. *Cell* **167**, 203–218.e17 (2016). doi: 10.1016/j.cell.2016.08.030; pmid: 27641500
39. M. Gutierrez-Arcelus *et al*., Lymphocyte innateness defined by transcriptional states reflects a balance between proliferation and effector functions. *Nat. Commun.* **10**, 687 (2019). doi: 10.1038/s41467-019-08604-4; pmid: 30737409

40. I. Sandrock *et al*., MicroRNA-181a/b-1 Is Not Required for Innate γδ NKT Effector Cell Development. *PLOS ONE* **10**, e0145010 (2015). doi: 10.1371/journal.pone.0145010; pmid: 26673421
41. J. Henao-Mejia *et al*., The microRNA miR-181 is a critical cellular metabolic rheostat essential for NKT cell ontogenesis and lymphocyte development and homeostasis. *Immunity* **38**, 984–997 (2013). doi: 10.1016/j.immuni.2013.02.021; pmid: 23623381
42. S. Hatano, T. Murakami, N. Noguchi, H. Yamada, Y. Yoshikai, CD5$^-$NK1.1$^+$ γδ T Cells that Develop in a Bcl11b-Independent Manner Participate in Early Protection against Infection. *Cell Rep.* **21**, 1191–1202 (2017). doi: 10.1016/j.celrep.2017.10.007; pmid: 29091759
43. H. F. Koay *et al*., A divergent transcriptional landscape underpins the development and functional branching of MAIT cells. *Sci. Immunol.* **4**, eaay6039 (2019). doi: 10.1126/sciimmunol.aay6039; pmid: 31757835
44. L. J. Howson *et al*., Absence of mucosal-associated invariant T cells in a person with a homozygous point mutation in *MR1*. *Sci. Immunol.* **5**, eabc9492 (2020). doi: 10.1126/sciimmunol.abc9492; pmid: 32709702
45. Y. Cai *et al*., Pivotal role of dermal IL-17-producing γδ T cells in skin inflammation. *Immunity* **35**, 596–610 (2011). doi: 10.1016/j.immuni.2011.08.001; pmid: 21982596
46. D. Melandri *et al*., The γδTCR combines innate immunity with adaptive immunity by utilizing spatially distinct regions for agonist selection and antigen responsiveness. *Nat. Immunol.* **19**, 1352–1365 (2018). doi: 10.1038/s41590-018-0253-5; pmid: 30420626
47. T. Mayassi *et al*., Chronic Inflammation Permanently Reshapes Tissue-Resident Immunity in Celiac Disease. *Cell* **176**, 967–981. e19 (2019). doi: 10.1016/j.cell.2018.12.039; pmid: 30739797
48. P. Rider, E. Voronov, C. A. Dinarello, R. N. Apte, I. Cohen, Alarmins: Feel the Stress. *J. Immunol.* **198**, 1395–1402 (2017). doi: 10.4049/jimmunol.1601342; pmid: 28167650
49. Y. Chen, K. Chou, E. Fuchs, W. L. Havran, R. Boismenu, Protection of the intestinal mucosa by intraepithelial γδ T cells. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14338–14343 (2002). doi: 10.1073/pnas.212290499; pmid: 12376619
50. J. Jameson *et al*., A role for skin γδ T cells in wound repair. *Science* **296**, 747–749 (2002). doi: 10.1126/science.1069639; pmid: 11976459
51. B. E. Keyes *et al*., Impaired Epidermal to Dendritic T Cell Signaling Slows Wound Repair in Aged Skin. *Cell* **167**, 1323–1338.e14 (2016). doi: 10.1016/j.cell.2016.10.052; pmid: 27863246
52. L. M. Boyden *et al*., Skint1, the prototype of a newly identified immunoglobulin superfamily gene cluster, positively selects epidermal γδ T cells. *Nat. Genet.* **40**, 656–662 (2008). doi: 10.1038/ng.108; pmid: 18408721
53. L. L. Sharp, J. M. Jameson, G. Cauvi, W. L. Havran, Dendritic epidermal T cells regulate skin homeostasis through local production of insulin-like growth factor 1. *Nat. Immunol.* **6**, 73–79 (2005). doi: 10.1038/ni1152; pmid: 15592472
54. A. Toulon *et al*., A role for human skin-resident T cells in wound healing. *J. Exp. Med.* **206**, 743–750 (2009). doi: 10.1084/jem.20081787; pmid: 19307328
55. X. J. Guo *et al*., Lung γδ T Cells Mediate Protective Responses during Neonatal Influenza Infection that Are Associated with Type 2 Immunity. *Immunity* **49**, 531–544.e6 (2018). doi: 10.1016/j.immuni.2018.07.011; pmid: 30170813
56. A. C. Kohlgruber *et al*., γδ T cells producing interleukin-17A regulate adipose regulatory T cell homeostasis and thermogenesis. *Nat. Immunol.* **19**, 464–474 (2018). doi: 10.1038/s41590-018-0094-2; pmid: 29670241
57. J. I. Odegaard *et al*., Perinatal Licensing of Thermogenesis by IL-33 and ST2. *Cell* **166**, 841–854 (2016). doi: 10.1016/j.cell.2016.06.040; pmid: 27453471
58. Z. A. Sullivan *et al*., γδ T cells regulate the intestinal response to nutrient sensing. *Science* **371**, eaba8310 (2021). doi: 10.1126/science.aba8310; pmid: 33737460
59. E. Rognoni, F. M. Watt, Skin Cell Heterogeneity in Development, Wound Healing, and Cancer. *Trends Cell Biol.* **28**, 709–722 (2018). doi: 10.1016/j.tcb.2018.05.002; pmid: 29807713
60. S. Naik *et al*., Commensal-dendritic-cell interaction specifies a unique protective skin immune signature. *Nature* **520**, 104–108 (2015). doi: 10.1038/nature14052; pmid: 25539086
61. H. Tanno *et al*., Invariant NKT cells promote skin wound healing by preventing a prolonged neutrophilic inflammatory response. *Wound Repair Regen.* **25**, 805–815 (2017). doi: 10.1111/wrr.12588; pmid: 28940971
62. H. Tanno *et al*., Contribution of Invariant Natural Killer T Cells to Skin Wound Healing. *Am. J. Pathol.* **185**, 3248–3257 (2015). doi: 10.1016/j.ajpath.2015.08.012; pmid: 26468976
63. J. Xie *et al*., Deficiency of Mucosal-Associated Invariant T Cells in TCRJα18 Germline Knockout Mice. *Immunohorizons* **3**, 203–207 (2019). doi: 10.4049/immunohorizons.1900035; pmid: 31356166
64. R. Bedel *et al*., Lower TCR repertoire diversity in Traj18-deficient mice. *Nat. Immunol.* **13**, 705–706 (2012). doi: 10.1038/ni.2347; pmid: 22814339
65. D. F. Schneider *et al*., Prevention of NKT cell activation accelerates cutaneous wound closure and alters local inflammatory signals. *J. Surg. Res.* **171**, 361–373 (2011). doi: 10.1016/j.jss.2010.03.030; pmid: 21067780
66. T. Chun *et al*., Induction of M3-restricted cytotoxic T lymphocyte responses by N-formylated peptides derived from Mycobacterium tuberculosis. *J. Exp. Med.* **193**, 1213–1220 (2001). doi: 10.1084/jem.193.10.1213; pmid: 11369792
67. Y. Bediako *et al*., SAP is required for the development of innate phenotype in H2-M3–Restricted Cd8$^+$ T cells. *J. Immunol.* **189**, 4787–4796 (2012). doi: 10.4049/jimmunol.1200579; pmid: 23041566
68. J. L. Linehan *et al*., Non-classical Immunity Controls Microbiota Impact on Skin Immunity and Tissue Repair. *Cell* **172**, 784–796. e18 (2018). doi: 10.1016/j.cell.2017.12.033; pmid: 29358051
69. O. J. Harrison *et al*., Commensal-specific T cell plasticity promotes rapid tissue adaptation to injury. *Science* **363**, eaat6280 (2019). doi: 10.1126/science.aat6280; pmid: 30523076
70. C. Jin *et al*., Commensal Microbiota Promote Lung Cancer Development via γδ T Cells. *Cell* **176**, 998–1013.e16 (2019). doi: 10.1016/j.cell.2018.12.040; pmid: 30712876
71. D. I. Godfrey, J. Le Nours, D. M. Andrews, A. P. Uldrich, J. Rossjohn, Unconventional T Cell Targets for Cancer Immunotherapy. *Immunity* **48**, 453–473 (2018). doi: 10.1016/j.immuni.2018.03.009; pmid: 29562195
72. M. D. Crowther *et al*., Genome-wide CRISPR-Cas9 screening reveals ubiquitous T cell cancer targeting via the monomorphic MHC class I-related protein MR1. *Nat. Immunol.* **21**, 178–185 (2020). doi: 10.1038/s41590-019-0578-8; pmid: 31959982
73. N. A. Gherardin *et al*., Enumeration, functional responses and cytotoxic capacity of MAIT cells in newly diagnosed and relapsed multiple myeloma. *Sci. Rep.* **8**, 4159 (2018). doi: 10.1038/s41598-018-22130-1; pmid: 29515123
74. C. Ma *et al*., Gut microbiome-mediated bile acid metabolism regulates liver cancer via NKT cells. *Science* **360**, eaan5931 (2018). doi: 10.1126/science.aan5931; pmid: 29798856

ACKNOWLEDGMENTS

Funding: Supported by NIAID Division of Intramural Research grants ZIA-AI001115 and ZIA-AI001132 (Y.B.), the Cancer Research Institute Irvington Postdoctoral Fellowship Program (M.G.C.), and NIAID Division of Extramural Activities grant K22-AI146217 (M.G.C.). **Author contributions:** M.G.C. and Y.B. conceptualized the article content, wrote and edited the manuscript, and designed the figures. **Competing interests:** The authors declare no competing interests.

10.1126/science.abf0095

RESEARCH ARTICLE SUMMARY

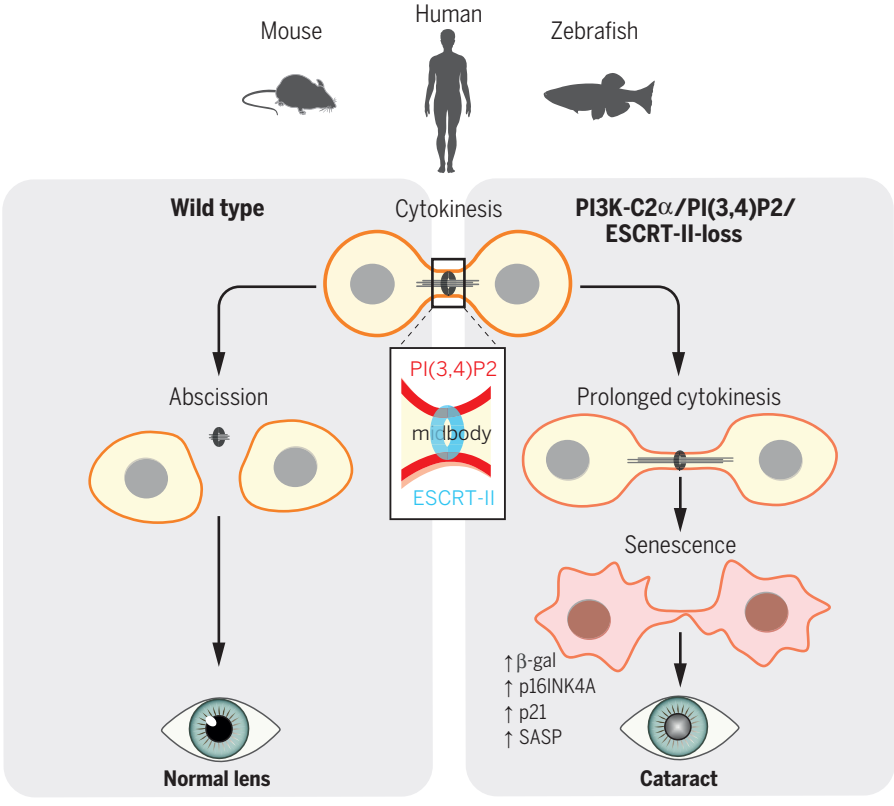
CELL BIOLOGY

PI(3,4)P2-mediated cytokinetic abscission prevents early senescence and cataract formation

Federico Gulluni*, Lorenzo Prever, Huayi Li, Petra Krafcikova, Ilaria Corrado, Wen-Ting Lo, Jean Piero Margaria, Anlu Chen, Maria Chiara De Santis, Sophie J. Cnudde, Joseph Fogerty, Alex Yuan, Alberto Massarotti, Nasrin Torabi Sarijalo, Oscar Vadas, Roger L. Williams, Marcus Thelen, David R. Powell, Markus Schueler, Michael S. Wiesener, Tamas Balla, Hagit N. Baris, Dov Tiosano, Brian M. McDermott Jr., Brian D. Perkins, Alessandra Ghigo, Miriam Martini, Volker Haucke, Evzen Boura, Giorgio Roberto Merlo, David A. Buchner, Emilio Hirsch*

INTRODUCTION: In response to aging, senescence plays a major role in the onset of various degenerative diseases, including the slow development of cataracts, a clouding of the eye lens that troubles the vision of millions of people worldwide. Transparent cells that form the eye lens divide throughout lifetime, but in the elderly, this process is less efficient, and senescent cells trigger the decline in transparency. Although pathogenesis is largely unclear, patients with genetic predisposition to early-onset cataract can help to define molecular mechanisms.

RATIONALE: Lack of the lipid kinase PI3K-C2α (phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2α) causes premature cataracts in fish, mice, and humans. This indicates that PI3K-C2α is involved in the maintenance of an evolutionarily conserved mechanism that protects the lens from the loss of transparency. On the basis of this, we aimed to define the role of PI3K-C2α in the lens, starting from the notion that this kinase is involved in mitosis and that correct cell division protects from senescence and eventually cataract development.



During cytokinesis, PI3K-C2α produced PI(3,4)P_2 (red) at the midbody, triggering the recruitment of the ESCRT-II subunit VPS36 (light blue). When PI3K-C2α, PI(3,4)P_2, or VPS36 is lost, the secondary ingression, where abscission occurs, does not form properly, resulting in prolonged cytokinesis. This leads to aneuploidy and early onset of senescence, which is particularly evident in the lens epithelium, resulting in cataract.

RESULTS: We found that PI3K-C2α played a specific role in the last step of cell division, called cytokinetic abscission. At the end of the mitotic process, the two nascent daughter cells form a contractile ring that constricts the membrane and produces an intercellular bridge that is eventually cleaved through biochemical reactions orchestrated by a specific organelle, the midbody. PI3K-C2α was found to be enriched at the midbody, where it produces a specific lipid moiety known as phosphoinositide (3,4) bisphosphate [PI(3,4)P_2]. PI(3,4)P_2 appeared to be required for efficient organization of the membrane-severing machinery, a multiprotein aggregate in which a cascade of three serially acting, ESCRT proteins–containing complexes eventually cleave the intercellular bridge. At the midbody, the endosomal sorting complexes required for transport I (ESCRT I) complex triggers the enucleation of the ESCRT II protein aggregate and then that of the membrane-severing ESCRT III machinery. A key component of the ESCRT II complex, named VPS36, anchored on locally preformed PI(3,4)P_2 to properly stabilize the ESCRT II protein aggregate at the midbody. This was found in fish, mice, and humans, indicating that this process has been conserved throughout vertebrate evolution. Without PI3K-C2α and its lipid product PI(3,4)P_2, the ESCRT machinery loses efficiency, and cytokinesis is delayed. Prolonged cytokinesis triggers the expression of p16INK4A and the activation of the transcriptional program, leading to senescence, likely protecting against the genomic instability that ensues cell division after refusion and tetraploidization. However, cytokinesis evolved as a resilient cellular process, and redundant pathways are common in the organization of the various ESCRT complexes. For example, an alternative pathway to ESCRT III organization that bypasses the need for an ESCRT II aggregate relies on the function of the linker protein ALIX. In the lens, ALIX was poorly expressed, and the PI3K-C2α, PI(3,4)P_2, and VPS36 pathway was specifically required to protect lens cells from cytokinetic delay, senescence, and ultimately, cataract formation.

CONCLUSION: This study provides evidence that a pathway based on PI3K-C2α, PI(3,4)P_2, and VPS36 positioning at the midbody improves robustness of cytokinetic abscission and is nonredundant in the eye lens. Failure of this process defines a pathogenetic mechanism of cataract development, ultimately delineating a path linking ESCRT dysfunction to senescence. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: federico.gulluni@unito.it (F.G.); emilio.hirsch@unito.it (E.H.)
Cite this article as F. Gulluni *et al.*, *Science* **374**, eabk0410 (2021). DOI: 10.1126/science.abk0410

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abk0410

RESEARCH ARTICLE

CELL BIOLOGY

PI(3,4)P2-mediated cytokinetic abscission prevents early senescence and cataract formation

Federico Gulluni[1]*, Lorenzo Prever[1], Huayi Li[1], Petra Krafcikova[2], Ilaria Corrado[1], Wen-Ting Lo[3], Jean Piero Margaria[1], Anlu Chen[4], Maria Chiara De Santis[1], Sophie J. Cnudde[1], Joseph Fogerty[5], Alex Yuan[5], Alberto Massarotti[6], Nasrin Torabi Sarijalo[7], Oscar Vadas[8,9], Roger L. Williams[10], Marcus Thelen[11], David R. Powell[12], Markus Schueler[13], Michael S. Wiesener[7], Tamas Balla[14], Hagit N. Baris[15,16,17], Dov Tiosano[15,17], Brian M. McDermott Jr.[18,19], Brian D. Perkins[5], Alessandra Ghigo[1], Miriam Martini[1], Volker Haucke[3,19], Evzen Boura[2], Giorgio Roberto Merlo[1], David A. Buchner[4,20], Emilio Hirsch[1]*

Cytokinetic membrane abscission is a spatially and temporally regulated process that requires ESCRT (endosomal sorting complexes required for transport)–dependent control of membrane remodeling at the midbody, a subcellular organelle that defines the cleavage site. Alteration of ESCRT function can lead to cataract, but the underlying mechanism and its relation to cytokinesis are unclear. We found a lens-specific cytokinetic process that required PI3K-C2α (phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2α), its lipid product PI(3,4)P_2 (phosphatidylinositol 3,4-bisphosphate), and the PI(3,4)P_2–binding ESCRT-II subunit VPS36 (vacuolar protein-sorting-associated protein 36). Loss of each of these components led to impaired cytokinesis, triggering premature senescence in the lens of fish, mice, and humans. Thus, an evolutionarily conserved pathway underlies the cell type–specific control of cytokinesis that helps to prevent early onset cataract by protecting from senescence.

Cytokinesis is the final step of mitosis, driving the physical separation of the two daughter cells, in which an organelle called midbody orchestrates the cleavage of the intercellular bridge. At the midbody, several cytokinesis-associated events take place—including cytoskeleton rearrangements, cell cycle regulation, membrane traffic, and plasma membrane remodeling—that progressively constrict the intercellular bridge until cleavage, in a process called abscission (*1*).

The ESCRT (endosomal sorting complexes required for transport) machinery directs a conserved membrane cleavage reaction that is important in multiple cellular processes, such as multivesicular endosome (MVE) formation, virus budding, plasma membrane repair, and importantly, cytokinetic abscission. All of these processes share a similar topology of budding away from the cytosol, starting with the recruitment of early ESCRT components (ESCRT-0, -I and -II) to the designated membrane (*2*). The final step of membrane fission relates to the self-polymerization and remodeling of the ESCRT-III subunits into helical filaments on the inner side of the membrane (*3*). Cytokinetic abscission has been thought to require only ESCRT-I, ESCRT-III, and the ESCRT-associated ALG2-interacting protein X (ALIX) that serves to recruit ESCRT-III in abscission. Nonetheless, mice that lack ALIX show reduced brain size without affecting other tissues, indicating the presence of redundant pathways that enable cytokinesis in non-neuronal cells (*4*–*7*). Consistent with this view, increasing evidence suggests that recruitment of the ESCRT-III component charged multivesicular body protein 4B (CHMP4B) to the midbody involves not only ALIX but also other parallel and additional pathways controlled by the ESCRT-I and -II cascade (*8*–*11*). However, the precise mechanism underlying ESCRT-II recruitment to the midbody and the resulting ESCRT-III accumulation remains incompletely understood (*1*).

During cargo sorting and budding at endosomes, ESCRTs are targeted to the membrane through multiple low-affinity interactions with 3-phosphoinositides (*12*–*14*). Phosphoinositides are produced on intracellular membranes and at the cytosolic face of the plasma membrane, where they serve as docking sites for proteins that govern diverse processes, including endocytosis, intracellular signaling, and vesicular trafficking (*1*). Whether a similar phosphoinositide-based mechanism directs ESCRT recruitment to the abscission site during cytokinesis is unknown.

Cytokinetic defects can end up in binucleation, and the ensuing tetraploidization can lead to senescence. Abnormal development of the lens as well as late-onset cataracts of the elderly are linked to premature cellular senescence (*15*), and cytokinetic defects may trigger these abnormalities (*16*). In addition, loss of members of the ESCRT machinery, such as CHMP4B and VPS4, determines early cataract onset in vivo (*17*–*19*), thus suggesting that perturbation of ESCRT-mediated cytokinesis can disturb the appropriate organization of lens epithelial cells. Thus, a connection between ESCRT-mediated abscission, senescence, and cataract development can be envisioned, but molecular details of this cascade of events are still poorly defined.

Senescence and premature aging in patients carrying null mutations of *PIK3C2A*

PIK3C2A (phosphatidylinositol-4-phosphate 3-kinase catalytic subunit type 2A)–null patients have congenital syndromic features reminiscent of premature aging, including early onset of cataract and secondary glaucoma (*20*). To investigate a possible role of *PIK3C2A* in premature senescence, we analyzed markers such as senescence-associated β-galactosidase (SA-β-gal) and p16INK4A (*21*) in *PIK3C2A*-null fibroblasts from patients and their controls. Whereas less than 10% of wild-type cells scored positive, more than 50% of *PIK3C2A*-null fibroblasts from different families displayed pronounced SA-β-gal activity (Fig. 1A and fig. S1A). Similar results were observed in *Pik3c2a*$^{-/-}$ primary mouse embryonic fibroblasts (MEFs) (fig. S1B) and in human lens epithelial cells in which *PIK3C2A* was suppressed (HLE-B3) (fig. S1C). After 2 weeks of culture, expression of senescence markers [p16INK4A, p21, BCL2/BAX ratio, and senescence-associated secretory phenotype (SASP)] were significantly higher in *PIK3C2A*-null patients' fibroblasts than in wild-type controls (Fig. 1B and fig. S1D) (3.02-fold induction ± 0.34, n = 5 replicates,

[1]Department of Molecular Biotechnology and Health Sciences, University of Turin, Turin 10126, Italy. [2]Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Praha, Czech Republic. [3]Leibniz-Forschungsinstitut für Molekulare Pharmakologie (FMP), 13125 Berlin, Germany. [4]Department of Biochemistry, Case Western Reserve University, Cleveland, OH 44106, USA. [5]Department of Ophthalmic Research, Cole Eye Institute, Cleveland Clinic, Cleveland, OH 44106, USA. [6]Dipartimento di Scienze del Farmaco, Università degli Studi del Piemonte Orientale, "A. Avogadro", Largo Donegani 2, 28100 Novara, Italy. [7]Department of Nephrology and Hypertension, Friedrich-Alexander University Erlangen Nürnberg, Erlangen, Germany. [8]Section des Sciences Pharmaceutiques, University of Geneva, 1211 Geneva, Switzerland. [9]Department of Microbiology and Molecular Medicine, University of Geneva, 1211 Geneva, Switzerland. [10]Medical Research Council (MRC) Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK. [11]Institute for Research in Biomedicine, Università della Svizzera Italiana, Bellinzona, Switzerland. [12]Pharmaceutical Biology, Lexicon Pharmaceuticals, The Woodlands, TX 77381, USA. [13]Division of Nephrology and Internal Intensive Care Medicine, Charite University, Berlin, Germany. [14]Section on Molecular Signal Transduction, Program for Developmental Neuroscience, Eunice Kennedy Shriver NICHD, National Institutes of Health, Bethesda, MD 20892, USA. [15]Division of Pediatric Endocrinology, Ruth Children's Hospital, Rambam Medical Center, Haifa 30196, Israel. [16]The Genetics Institute, Rambam Health Care Campus, Haifa, Israel. [17]Rappaport Family Faculty of Medicine, Technion–Israel Institute of Technology, Haifa 30196, Israel. [18]Department of Otolaryngology–Head and Neck Surgery, Case Western Reserve University School of Medicine, Cleveland, OH 44106, USA. [19]Freie Universität Berlin, Faculty of Biology, Chemistry and Pharmacy, 14195 Berlin, Germany. [20]Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH 44106, USA.
*Corresponding author. Email: federico.gulluni@unito.it (F.G.); emilio.hirsch@unito.it (E.H.)

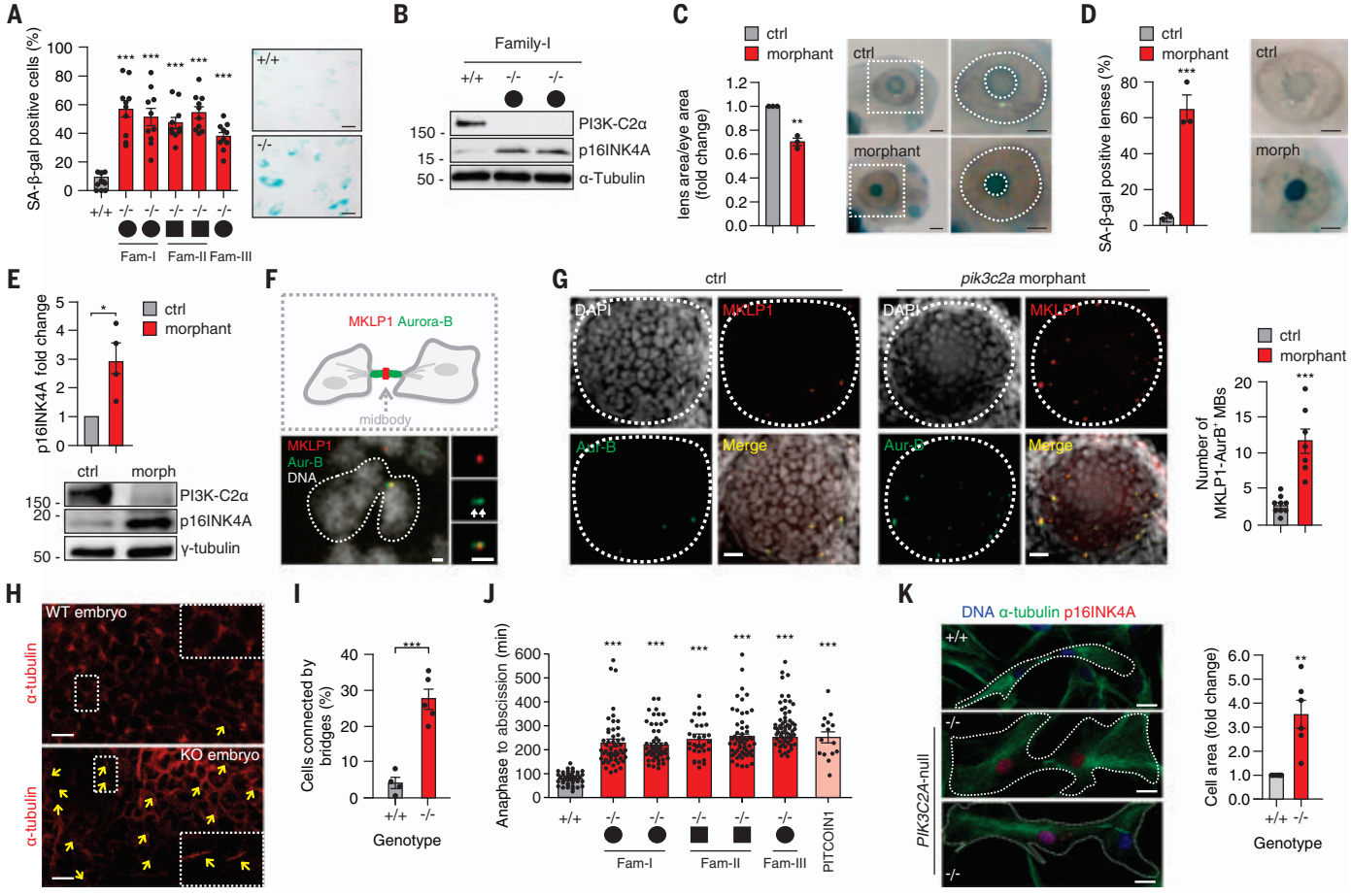


Fig. 1. Loss of PI3K-C2α induces defective abscission and early senescence. (**A**) Quantification and representative images of SA-β-gal–positive fibroblasts derived from Family I, II, and III after 2 weeks in culture. +/+ genotypes are shown in the chart as pulled together. (**B**) Immunoblot analysis of p16INK4A protein level in fibroblasts derived from Family I after 2 weeks in culture. (**C**) Quantification and representative images of the ratio between lens and eye size in control and *pik3c2a* morphant 72 hours post-fertilization (hpf) zebrafish embryos. (**D**) Quantification and representative images of SA-β-gal intensity on the lens of control and *pik3c2a* morphant 72-hpf embryos. (**E**) Immunoblot analysis of p16INK4A and PI3K-C2α in control and *pik3c2a* morphant embryos (n = 4 pools of 15 embryos each). (**F** and **G**) Confocal images of whole-mount immunofluorescence performed on 72-hpf embryos lens using MKLP1 (red) and Aur-B (green) antibodies to stain midbody and TO-PRO-3 (gray) to stain nuclei. (**H**) Immunofluorescence of wild-type and *Pik3c2a*$^{-/-}$ embryo sections by using α-tubulin to mark intercellular bridges connecting cells in cytokinesis. (**I**) Quantification of the number of cells connected by bridges (%). n = 6 fields in at least four independent experiments. (**J**) Time-lapse analysis of the time required to progress from anaphase to cytokinesis in fibroblasts derived from patients with homozygous deletion of PI3K-C2α or in control fibroblasts treated with PITCOIN1. (**K**) (Left) Immunofluorescence of p16INK4A (red), DNA (blue), and α-tubulin (green) in wild-type and *PIK3C2A*-null fibroblast. (Right) Quantification of cell area in control and *PIK3C2A*-null fibroblasts. If not previously specified, all results are shown as mean or representative picture of at least three independent experiments ± SEM. *P < 0.05; **P < 0.01; ***P < 0.001.

P = 0.0013; fig. S1, E to G). Similar results were observed in PI3K-C2α–depleted HLE-B3 cells (fig. S1, E, F, and H) (3.31-fold induction ± 0.23, n = 3 replicates, P < 0.0001; fig. S1I). Next, a selective inhibitor for PI3K-C2α (PITCOIN1) was tested on human fibroblasts and found to significantly induce p16INK4A expression (fig. S1J), confirming that loss of the lipid products of this kinase could induce senescence.

Loss of PI3K-C2α induces early senescence, defective lens development, and cataracts in vivo

To test whether PI3K-C2α depletion causes premature senescence in the eye lens, we focused on zebrafish in which *pik3c2a* was suppressed (fig. S2A) that develop to term and recapitulate the human phenotype more faithfully than do *Pik3c2a*$^{-/-}$ mice, which die in utero (22). The lens size was 25% smaller in *pik3c2a* morphants than in controls (Fig. 1C). We also found a significantly higher number of SA-β-gal–positive lenses in *pik3c2a*-suppressed fish than in controls (65.5 versus 5.1%, respectively) (Fig. 1D and fig. S2B). Levels of other senescence markers (p16INK4A, p21, BCL2/BAX ratio, and SASP) were also increased in *pik3c2a* morphants (Fig. 1E and fig. S2, C to E). Similarly, p16INK4A was elevated in wild-type embryos treated with PITCOIN1 (fig. S2F).

To further investigate whether *pik3c2a* loss in adult fish caused premature lens cell senescence and cataracts akin to *PIK3C2A*-null patients, we generated and examined compound heterozygotes from two zebrafish strains carrying distinct null mutations in the *pik3c2a* gene. The frequency of genotypes in the offspring from *pik3c2a*$^{+/-}$ intercrosses followed the expected Mendelian ratio (table S1), and the gross morphology of the mutants was indistinguishable from that of controls. All compound heterozygotes carrying the two distinct null alleles (n = 118) displayed lenticular abnormalities (P < 0.005, two-tailed Fisher's exact test), consisting in circular cataract and posterior lenticonus (fig. S3A and movie S1), like what is observed in *PIK3C2A*-null patients (fig. S3B) (20).

To challenge these findings in a different model organism, we analyzed the eye of a mouse homozygous for a *Pik3c2a* hypomorphic allele (*Pik3c2a*$^{hypo/hypo}$) (*23*) and found reduced eye size (fig. S3C). *Pik3c2a*$^{hypo/hypo}$ lenses displayed reduced numbers of KI-67–positive nuclei and increased p16INK4A expression compared with those of controls (fig. S3, D and E). To analyze whether this could occur independently of a developmental defect, loss of PI3K-C2α in the adult eye from Cre-mediate gene deletion induced a significant up-regulation of p16INK4A expression and other senescence markers (p21, BCL2/BAX ratio, and SASP) in the lens (fig. S4, A to D), leading to overt cataracts within 3 months (fig. S4E). This was consistent with cataract patients, in which *PIK3C2A* lens expression is significantly lower in individuals with age-related cataracts than in age-matched controls (36.7% ± 5.2 reduction; $P = 0.0001$) (*24*). Thus, loss of PI3K-C2α in vertebrates can induce early senescence, defective lens development, and cataracts.

Cytokinesis defects in cells lacking PI3K-C2α

Abnormal cell division and subsequent tetraploidization can lead, in the lens epithelium, to senescence and cataract formation (*15*, *16*). We reverted to zebrafish morphants as well as PITCOIN1-treated embryos and analyzed tetraploidy by means of flow cytometry in dissociated cells of the developing lens (fig. S5, A to C). The number of cells in the G_2/M phase of the cell cycle was substantially increased (twofold difference between controls and either *pik3c2a* morphants or PITCOIN1-treated embryos), indicating increased number of cells stalled in late cytokinesis (*25*). In agreement with this hypothesis, immunofluorescence analysis revealed a threefold enrichment of lens epithelial cells connected by a cytokinetic bridge positive for markers of late cytokinesis, such as Aurora B and MKLP1 (Fig. 1, F and G, and fig. S5, D and E). Cells connected by intercellular bridges were also observed in embryonic day 11.5 (E11.5) *Pik3c2a*$^{-/-}$ mouse embryos (Fig. 1, H and I). Furthermore, time-lapse analysis of the progression from anaphase onset to abscission showed that *Pik3c2a*$^{-/-}$ MEFs were delayed in the completion of cytokinesis (fig. S5, F and G, and movie S2). Whereas post-anaphase furrowing and intercellular bridge establishment proceeded normally (fig. S5, F and G, and movie S2), cytokinesis failure and refusion appeared more frequently (fig. S5, H and I).

Like fish and mice, *PIK3C2A*-null patients' cells (*20*) showed a significantly delayed abscission (Fig. 1J) and binucleation (fig. S5J). Mutant cells still connected with the intercellular bridge resulted positive for the senescence marker p16INK4A (Fig. 1K and fig. S6, A and C). In addition, mutant daughter cells that were stalled in cytokinesis exhibited p16INK4A positivity in only one of the two nuclei, indicating that the increase in p16INK4A expression did not occur after mitosis but appeared in cells enduring delayed abscission (Fig. 1K and fig. S6A, bottom right). In agreement, cell flattening that is typical of senescent cells occurred in p16INK4A-positive *PIK3C2A*-null cells that did not complete cytokinesis (Fig. 1K, right). This phenotype was recapitulated in *PIK3C2A*-depleted HLE-B3 cells (fig. S6, B and D) as well as in cells treated with an inhibitor of cytokinesis [Paprotrain (*26*)] (fig. S6E). Thus, data from humans, mice, and cultured cells indicate that senescence in cells or tissues that lack PI3K-C2α function could be a direct consequence of defective and/or delayed cytokinesis.

PI3K-C2α localizes at the midbody and directs abscission

To dissect the molecular mechanism underlying the requirement for PI3K-C2α to complete cytokinesis and to protect cells from senescence, we used HeLa cells, in which depletion of PI3K-C2α delayed abscission (fig. S7, A to E). Localization of endogenous PI3K-C2α during cytokinesis was analyzed by means of immunofluorescence and found in a characteristic ring-like arrangement at the midbody. PI3K-C2α appeared to be embraced by intercellular-bridge microtubules, indicating that PI3K-C2α was associated with the midbody ring (Fig. 2A and figs. S7, F, top, and G, and S8A). Similar results were seen in cells transfected with a green fluorescent protein (GFP)–tagged PI3K-C2α (figs. S7, F, bottom, and S8, B and C, and movie S3). Both transfected GFP–PI3K-C2α and the endogenous protein were absent from the cleavage furrow during early cytokinesis but became enriched at the midbody during late stages of cytokinesis (fig. S7F). Thus, PI3K-C2α localizes to the midbody and controls late cytokinetic abscission.

Coincidental binding of γ-tubulin and PI(4,5)P_2 localizes PI3K-C2α at the midbody

During the furrowing process, PI(4,5)P_2 progressively accumulates at the membrane surrounding the furrow (*27*). The PX domain of PI3K-C2α specifically binds PI(4,5)P_2 with high affinity and is required for the activation of PI3K-C2α catalytic activity (*28*). We thus analyzed the localization in cytokinesis of a PI3K-C2α PX-domain mutant as well as a kinase inactive (KD) form, a clathrin-binding domain mutant (Δ1-380), and a GFP-tagged wild-type protein and found that only the PX-binding mutant reduced PI3K-C2α localization (Fig. 2B and fig. S9A).

Previous proximity biotinylation experiments identified PI3K-C2α as a putative γ-tubulin interactor (*29*). In agreement, we found γ-tubulin to be enriched in an unbiased proteomic analysis of PI3K-C2α interactors (table S2). Given the specific localization of γ-tubulin at the midbody (*30*), we tested its possible function in recruiting PI3K-C2α during abscission by analyzing the localization of PI3K-C2α and γ-tubulin. As expected, both proteins were enriched at the midbody where they colocalized (Fig. 2C). Immunoprecipitation from cells synchronized in cytokinesis of myc-PI3K-C2α together with γ-tubulin revealed that the two proteins interact during abscission (Fig. 2D). A docking model of the PI3K-C2α and γ-tubulin complex showed a region comprising the C-terminal part of the helical domain of PI3K-C2α fitting to the γ-tubulin crystal structure (fig. S9, B to D). We created an enhance GFP (EGFP) fusion protein between the C-terminal part of the PI3K-C2α helical domain [referred to hereafter as γ-tubulin binding domain (GBD), amino acids 981 to 1203] and analyzed its intracellular localization by means of immunofluorescence. EGFP-GBD displayed complete colocalization with γ-tubulin in HeLa cells during interphase (fig. S9, E and F). Pull-down assays by using GST-GBD showed that the GBD is sufficient to precipitate γ-tubulin from cells synchronized in cytokinesis (Fig. 2E). Last, mutation of three residues within the γ-tubulin interaction site (Q1022A, T1025A, and S1081A) inhibited γ-tubulin binding (Fig. 2F) as well as colocalization at the centrosome and midbody (Fig. 2G). These data demonstrate that coincidental binding of PI(4,5)P_2 and γ-tubulin drives PI3K-C2α localization. (Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. In the mutants, other amino acids were substituted at certain locations; for example, Q1022A indicates that glutamine at position 1022 was replaced by alanine.)

PI3K-C2α produces PI(3,4)P_2 at the midbody

To characterize the role of PI3K-C2α at the midbody, we explored the function of its kinase activity during cytokinesis. Delayed abscission in HeLa cells was rescued through reexpression of a small interfering RNA (siRNA)–resistant wild-type PI3K-C2α, whereas add-back of the KD form had no effect (Fig. 3A and fig. S10A), demonstrating that cytokinesis completion requires PI3K-C2α kinase activity at the midbody.

The scaffold function of PI3K-C2α is involved in spindle organization at metaphase, and lack of PI3K-C2α causes aberrant chromosome congression (*31*). In line with these findings, loss of PI3K-C2α increased the amount of LAP-2–positive lagging chromosomes (fig. S10B), which when entrapped at the bridge cause cytokinetic delay. To exclude a role of the scaffold function, we reexpressed

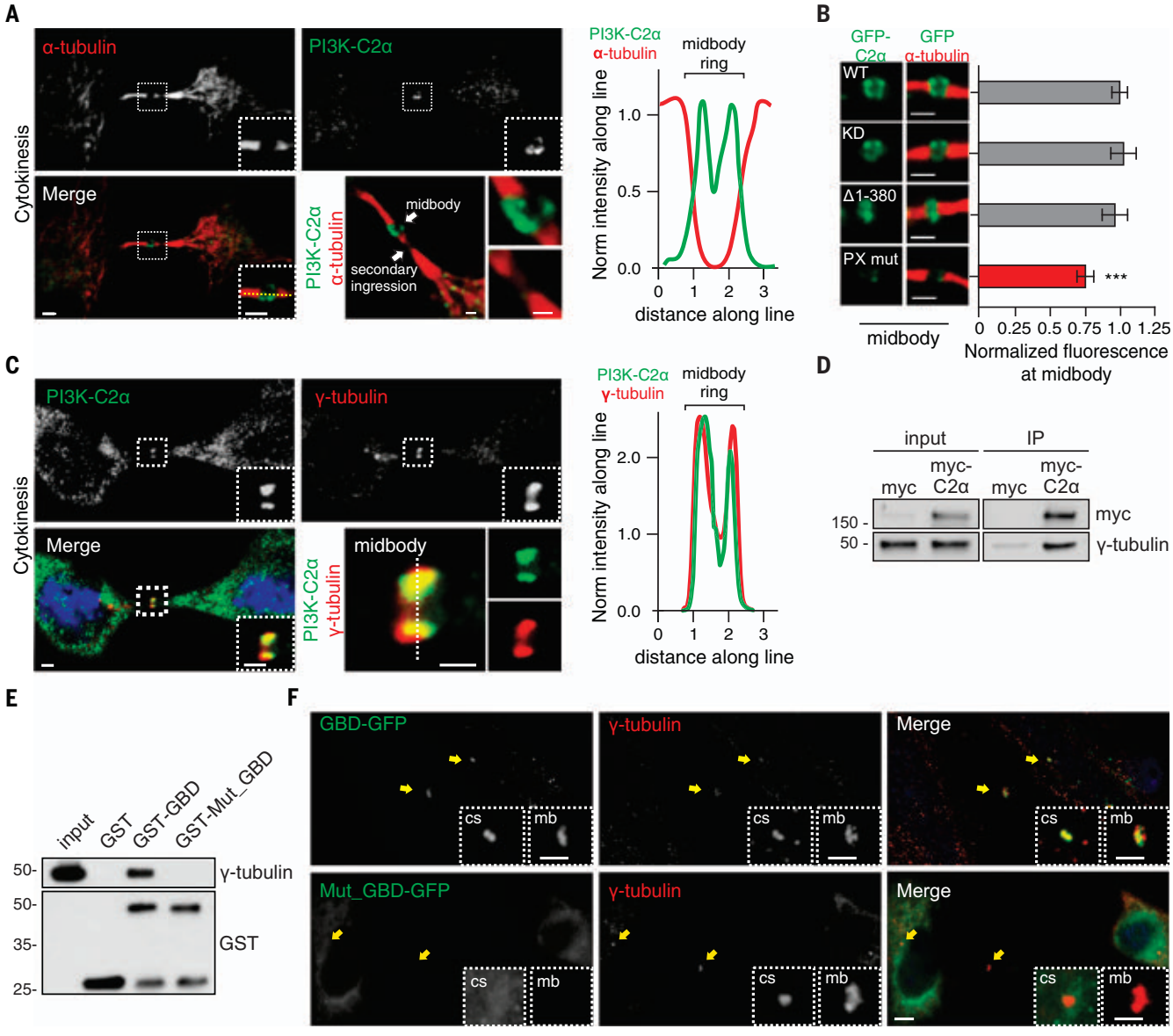


Fig. 2. PI3K-C2α localizes to midbody through the PX- and the γ-tubulin–binding domain. (**A**) Confocal images of HeLa cells stained for PI3K-C2α (green) and α-tubulin (red) (left) during cytokinesis and (right) fluorescence intensity along the line showing localization of PI3K-C2α at midbody. (**B**) Hela cells transfected with wild-type, kinase inactive (KD), clathrin-binding deletion (Δ1-380), and PX-binding mutant PI3K-C2α GFP-tagged. Immunofluorescence staining by using antibody to GFP showing (left) the enrichment of the different PI3K-C2α constructs and (right) their quantification at the midbody. (**C**) Confocal images of (left) HeLa cells stained for PI3K-C2α (green), 4′,6-diamidino-2-phenylindole (DAPI) (blue), and γ-tubulin (red) and (right) fluorescence intensity along the line showing colocalization between PI3K-C2α and γ-tubulin. (**D**) Immunoprecipitation of myc-PI3K-C2α and immunoblot anti- either myc or γ-tubulin. (**E**) Pull-down experiment by using GST-GBD and GST-GBD$^{Q1022A\text{-}T1025A\text{-}S1081A}$ from cells synchronized in cytokinesis. (**F**) Immunofluorescence staining of GFP-GBD and GFP-GBD$^{Q1022A\text{-}T1025A\text{-}S1081A}$ (green) with γ-tubulin (red) during cytokinesis. Enlarged section shows colocalization with centrosome and midbody. If not previously specified, all results are shown as mean or representative picture of at least three independent experiments ± SEM. ***$P < 0.001$.

a kinase-dead version of PI3K-C2α that allows proper chromosome congression but fails to produce PI(3,4)P_2 (*31*). The kinase-dead mutant was able to abolish LAP2-positive bridges but not delayed cytokinesis, thus excluding that defective abscission relies on the role of PI3K-C2α in metaphase (fig. S10B).

PI3K-C2α has been reported to produce phosphatidylinositol 3-phosphate [PI(3)P] on endosomes and PI(3,4)P_2 at the plasma membrane (*22*). To address the identity of the PI3K-C2α lipid product at the midbody, we analyzed the localization of 3-phosphoinositides during cytokinesis. PI(3,4,5)P_3, mainly produced by class I PI3Ks, was almost undetectable at the midbody (fig. S11A). As expected (*32*), PI(3)P was enriched at the midbody (fig. S11B), and no significant changes in its levels were observed upon PI3K-C2α depletion (Fig. 3B). PI(3)P mainly appeared inside the intercellular bridge (Fig. 3B and fig. S11C), differently from the ring-like distribution of PI3K-C2α. By contrast, PI3K-C2α localization matched that of PI(3,4)P_2 (Fig. 3C and fig. S11, D and E), and suppression of PI3K-C2α resulted in a 73% reduction in PI(3,4)P_2 abundance, whereas levels of PI(3)P remained unchanged (Fig. 3, B and C).

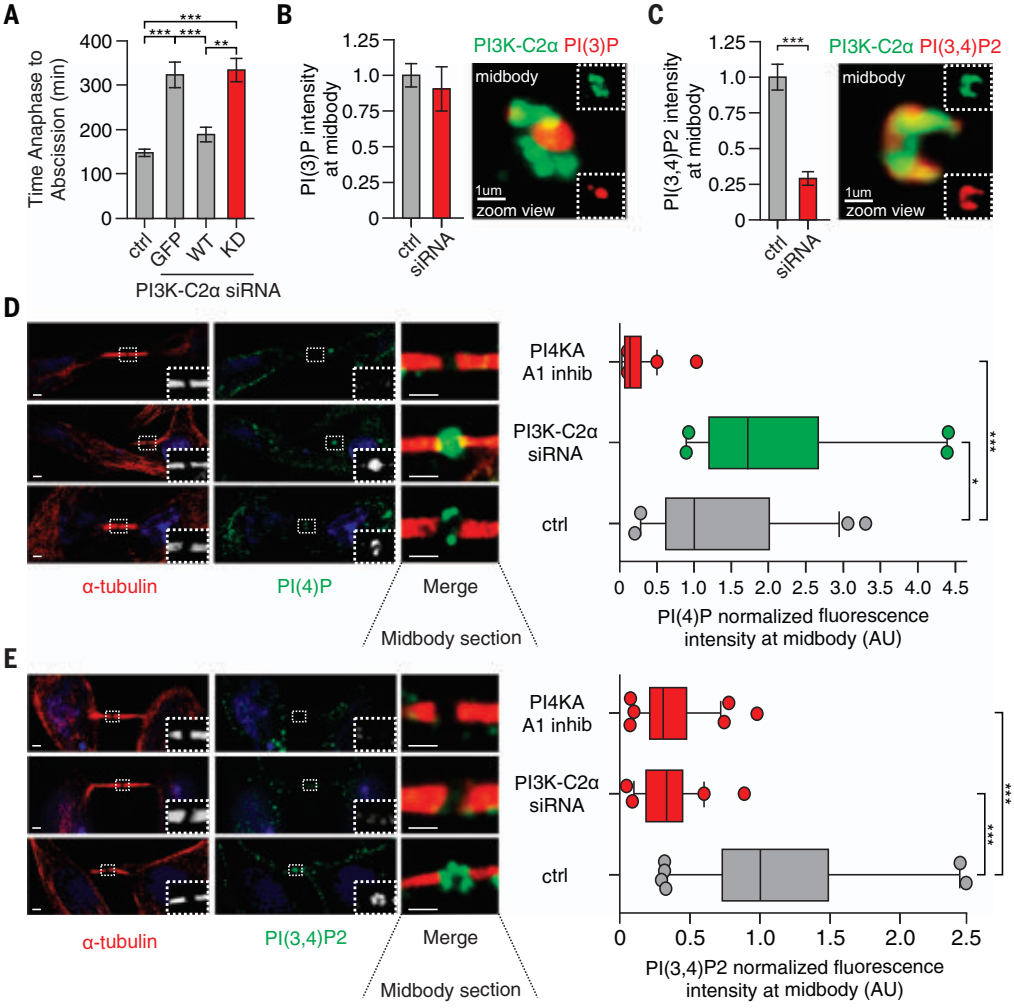


Fig. 3. PI3K-C2α produces PI(3,4)P_2 at the midbody ring. (**A**) Time-lapse microscopy measuring the time required to progress from anaphase to abscission in HeLa cells treated with siRNA for PI3K-C2α alone or expressing GFP and siRNA-resistant GFP–wild-type (WT) or GFP-kinase inactive (KD) form of PI3K-C2α. $n \geq 25$ cells imaged. (**B**) (Left) Quantification of PI(3)P levels at midbody by means of immunofluorescence staining ($n \geq 100$ cells) and (right) confocal image of 3D reconstructed midbody stained for PI(3)P (red) and PI3K-C2α (green). (**C**) (Left) Quantification of PI(3,4)P_2 levels at midbody by immunofluorescence staining ($n \geq 100$ cells) and (right) confocal image of 3D reconstructed midbody stained for PI(3,4)P_2 (red) and PI3K-C2α (green). (**D**) Quantification of PI(4)P levels at midbody in control, A1 inhibitor, and siRNA-treated cells. $n \geq 100$ cells, mean ± SD. (**E**) Quantification of PI(3,4)P_2 levels at midbody in control, A1 inhibitor, and siRNA-treated cells. $n \geq 100$ cells, mean ± SD. If not previously specified, all results are shown as mean or representative picture of at least three independent experiments ± SEM. *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

PI3K-C2α generates PI(3,4)P_2 starting from PI(4)P (*33*), suggesting that PI4K may synthetize PI(4)P at the midbody. Because PI4KA is well known to be active at the plasma membrane (*34*), we treated cells with the PI4KA-specific inhibitor A1 and found it to abolish PI(4)P at the midbody (Fig. 3D). Conversely, depletion of PI3K-C2α significantly increased PI(4)P at the midbody (Fig. 3D) and reduced PI(3,4)P_2 to the same extent of PI4KA inhibition (Fig. 3E), thus indicating that PI3K-C2α uses PI(4)P as a substrate.

PI(3,4)P2 locally produced by PI3K-C2α controls the recruitment of ESCRT-II/VPS36 to the midbody

Cytokinetic abscission relies on a cascade of interactions in which ESCRT-I binds to ALIX, which eventually recruits ESCRT-III (*4*, *6*). Unexpectedly and in contrast to this canonical model, cataracts have been observed in humans and mice harboring inactivating mutations of the ESCRT-III component CHMP4B (*17*) but not in ALIX-null mice (*7*). We hypothesized that PI3K-C2α might be a key component of an ALIX-independent pathway that leads to cytokinetic abscission that is of importance in the lens. The lens epithelium of both fish and mice, although expressing high levels of PI3K-C2α, appeared nearly devoid of ALIX (Fig. 4, A and B, and fig. S12A).

These data led us to hypothesize that in cells expressing low or negligible amounts of ALIX, cytokinetic abscission depends on PI3K-C2α. We further reasoned that even in cells that express ALIX, such as HeLa cells, the PI3K-C2α pathway may act synergistically with ALIX to drive completion of cytokinesis. Thus, we tested whether PI3K-C2α–derived PI(3,4)P_2 could recruit effector proteins linking ESCRT-III to the cytokinetic bridge. Among the known ESCRT components, VPS36, a core subunit of ESCRT-II, binds PI(3,4)P_2 (*13*). Immunofluorescence staining of endogenous VPS36 showed that during cytokinesis, this protein was enriched at the midbody, where it colocalized with PI(3,4)P_2 in ring-like structures (Fig. 4C). Like PI3K-C2α, VPS36 was absent from the cleavage furrow but was enriched at the midbody during late cytokinesis (fig. S12, B to D). Time-lapse imaging of living cells confirmed the enrichment of GFP-VPS36 at the midbody during abscission (Fig. 4D and movie S4). VPS36 suppression caused delayed cytokinesis (fig. S13A) and refusion during abscission like that observed in cells lacking PI3K-C2α (fig. S13B).

To dissect the order of events during cytokinetic abscission, we analyzed the effects of the depletion of either PI3K-C2α or VPS36 on the localization of the various components of the ESCRT machinery. Suppression of either PI3K-C2α or VPS36 had no effect on TSG101 (ESCRT-I) localization at the midbody (fig. S13C). Similarly, down-modulation of TSG101 did not lead to changes in either PI3K-C2α or PI(3,4)P_2 amounts. Thus, PI3K-C2α acts downstream of TSG101 in an ESCRT-I–independent manner (fig. S13, D and E). Conversely, suppression of PI3K-C2α, but not class III PI3K inhibition, resulted in a significantly reduced recruitment of VPS36 to the midbody (Fig. 4E and fig. S13F, respectively). These findings were confirmed in *PIK3C2A*-null human fibroblasts that showed a >50% reduction in VPS36 and

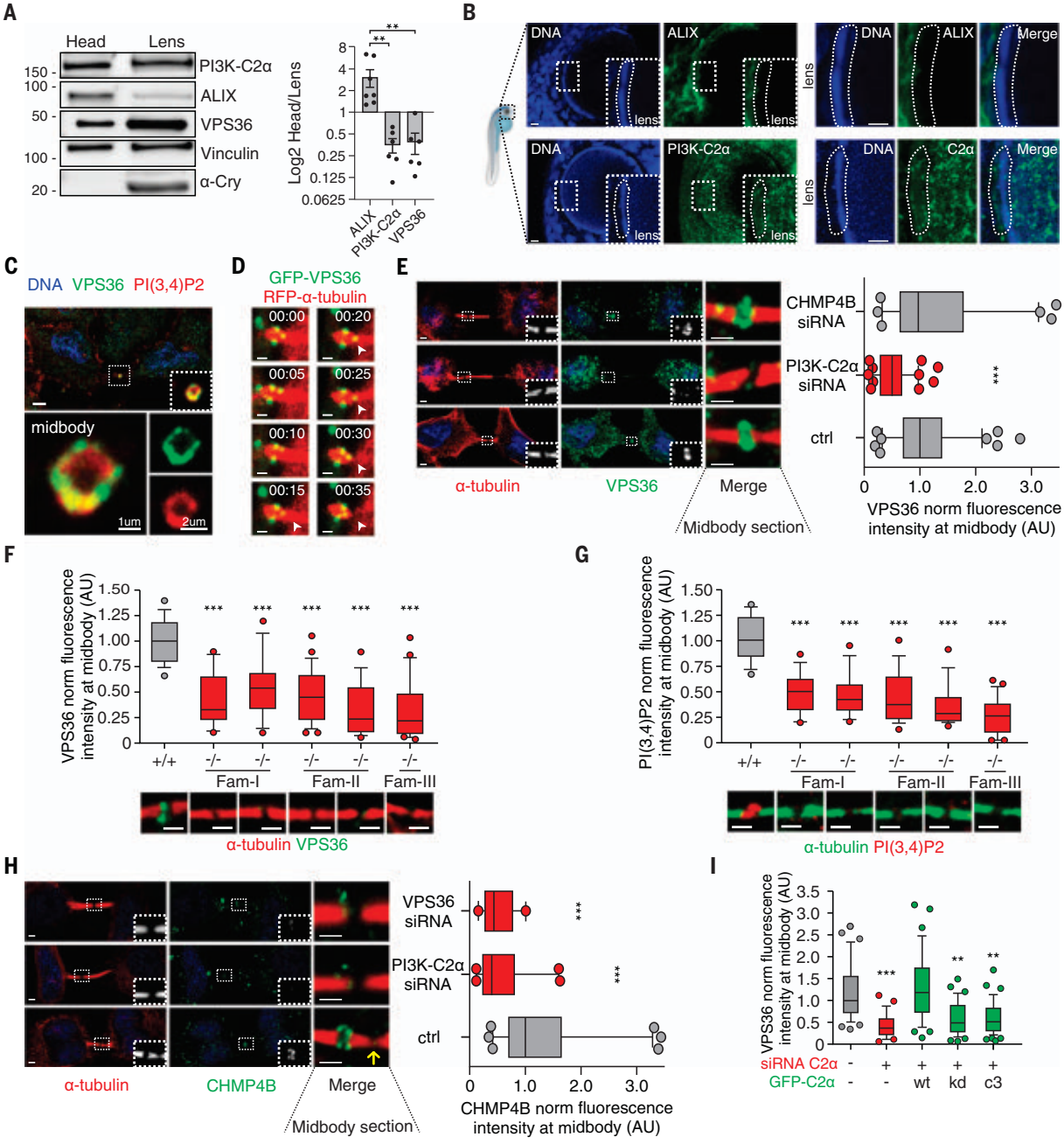


Fig. 4. PI3K-C2α/VPS36 axis is an ALIX alternative pathway to recruit CHMP4B to the midbody in lens. (**A**) Immunoblot analysis (left) and quantification plot (right) showing Log2 of head/lens protein level for PI3K-C2α, ALIX, and VPS36 in adult zebrafish. (**B**) Immunofluorescence showing (top) ALIX or (bottom) PI3K-C2α expression in 72-hpf zebrafish lens. (**C**) Immunofluorescence staining showing colocalization between endogenous VPS36 and PI(3,4)P_2 at the midbody ring. (**D**) Snapshots taken from time-lapse imaging of GFP-VPS36 during cytokinesis in HeLa cells stably expressing red fluorescent protein (RFP)–α-tubulin. Arrowheads indicate enrichment of VPS36 at the abscission site. (**E**) Quantification of VPS36 levels at midbody in control and siRNA-treated cells. $n \geq 200$ cells, mean ± SD. (**F** and **G**) Immunofluorescence staining and quantification of endogenous (F) VPS36 and (G) PI(3,4)P_2 levels at the midbody in fibroblasts derived from patients with homozygous deletion of PI3K-C2α. +/+ genotypes were pulled together. $n \geq 50$ cells, mean ± SD. (**H**) Quantification of CHMP4B levels at midbody in control and siRNA-treated HeLa cells. $n \geq 120$ cells, mean ± SD. (**I**) Quantification of VPS36 levels at midbody in control and siRNA-treated cells upon transfection of siRNA-resistant WT, KD, and PI(3)P–producing (C3) forms of PI3K-C2α. $n \geq 30$ cells, mean ± SD. If not previously specified, all results are shown as mean or representative picture of at least three independent experiments ± SEM. $^{**}P < 0.01$; $^{***}P < 0.001$.

PI(3,4)P_2 levels at the midbody as well as elongated intercellular bridges (Fig. 4, F and G, and fig. S14A). Removal of TSG101 together with PI3K-C2α synergized and almost completely abolished VPS36 recruitment to the midbody (fig. S13G). Thus, PI3K-C2α–derived PI(3,4)P_2 facilitates VPS36 recruitment and ESCRT-III assembly. In agreement, PI4KA inhibition reduced the ESCRT-III subunit CHMP4B at the midbody (fig. S14B), but suppression of CHMP4B had no effect on VPS36 localization (Fig. 4E), confirming that VPS36 is recruited before ESCRT-III during abscission. At the midbody, depletion of either PI3K-C2α or VPS36 reduced CHMP4B enrichment

(Fig. 4H), but overexpression of VPS36-mCherry increased CHMP4B-GFP localization (fig. S15A). Thus, VPS36/ESCRT-II participates in the hierarchical recruitment of ESCRT-III during abscission. Midbody fractionation showed that PI3K-C2α loss led to significantly reduced levels of VPS36 and CHMP4B (fig. S15, B and C) but not of the ESCRT-I subunit TSG101 and Citron kinase (fig. S15, B to D).

Last, VPS36 localization at the midbody in PI3K-C2α–suppressed cells was completely restored by a wild-type PI3K-C2α but neither by the KD nor by the C3 mutant that produces PI(3)P only (Fig. 4I and fig. S15E) (*33*, *35*). Thus, $PI(3,4)P_2$ produced by PI3K-C2α controls VPS36 localization, contributing to the midbody enrichment of CHMP4B driving cytokinesis.

VPS36-dependent CHMP4B recruitment to the midbody through an ALIX-independent pathway

Our results pointed to two mechanisms for recruitment of ESCRT-III at the midbody: one depending on ALIX (*12*, *25*) and the other on PI3K-C2α/ESCRT-II (fig. S15F). In agreement with this hypothesis, we observed a 16, 49, and 76% reduction in CHMP4B levels at the midbody upon suppression of either ALIX or PI3K-C2α or their concomitant depletion, respectively (Fig. 5A). No changes in ALIX localization at the midbody could be observed after loss of either PI3K-C2α or VPS36 (Fig. 5B), suggesting that ALIX and PI3K-C2α/VPS36 operate independently of each other (fig. S15F). Thus, CHMP4B can associate with the midbody through either ALIX (ALIX pathway) or PI3K-C2α/ESCRT-II (fig. S15F). Furthermore, the complex of VPS36 and VPS22, anchored at the midbody membrane, can recruit CHMP6, eventually promoting the addition of CHMP4B (fig. S15F) (*8*). Consistent with this model, suppression of CHMP6 resulted in a nearly 50% reduction in CHMP4B levels at the midbody (Fig. 5C). Furthermore, CHMP6 recruitment was decreased after suppression of either PI3K-C2α or VPS36 but not after suppression of ALIX (Fig. 5D). Conversely, removal of CHMP6 did not impair VPS36 recruitment (Fig. 5E). Thus, a pathway anchoring CHMP4B/ ESCRT III to the midbody plasma membrane acts in parallel to ALIX and depends on PI3K-C2α/PI(3,4)2, VPS36, and CHMP6 (fig. S15F).

Binding of VPS36 to $PI(3,4)P_2$ is required for VPS36 localization to the midbody

VPS36 function is highly conserved during evolution, but its structure substantially diverged from yeast to mammals, with multiple changes in phosphoinositide binding selectivity (*13*, *14*, *36*, *37*). The yeast GRAM-like ubiquitin binding in EAP45 (GLUE) domain in VPS36 shows specificity toward PI(3)P, whereas the mammalian GLUE domain binds strongly but

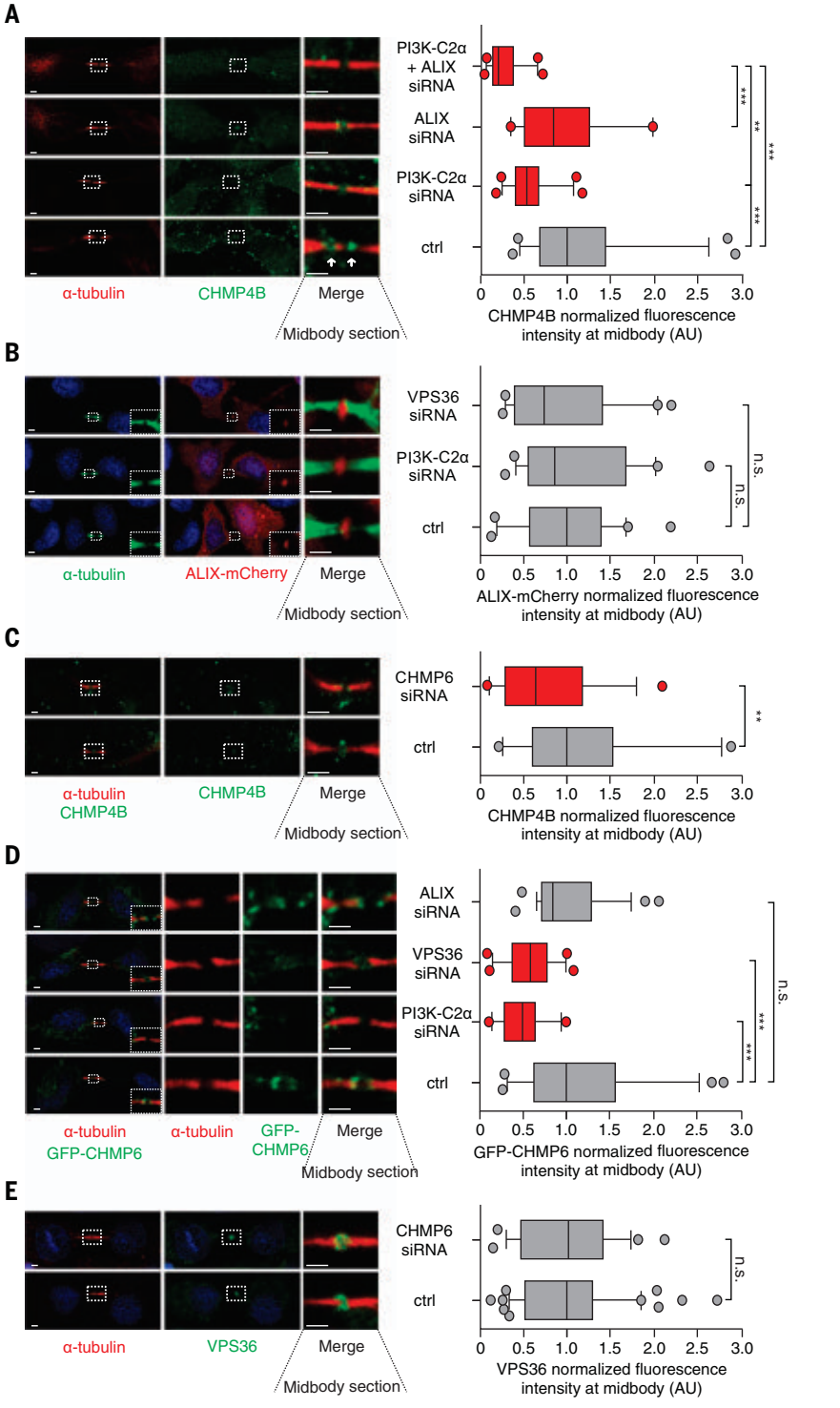


Fig. 5. VPS36 recruits CHMP4B to the midbody through ALIX parallel pathway. (**A**) Quantification of CHMP4B levels at midbody in HeLa cells treated with control siRNA or siRNA targeting PI3K-C2α, ALIX, or PI3K-C2α and ALIX. $n \geq 50$ cells, mean ± SD. (**B**) Quantification of ALIX levels at midbody in HeLa cells transfected with ALIX-mCherry and treated with control or siRNA targeting PI3K-C2α or VPS36. $n \geq 60$ cells, mean ± SD. (**C**) Quantification of CHMP4B levels at midbody in HeLa cells treated with control or siRNA targeting CHMP6. $n \geq 45$ cells, mean ± SD. (**D**) Quantification of GFP-CHMP6 levels at midbody in HeLa cells transfected with GFP-CHMP6 and treated with control or siRNA targeting PI3K-C2α, VPS36, or ALIX. $n \geq 50$ cells, mean ± SD. (**E**) Quantification of VPS36 levels at midbody in HeLa cells treated with control or siRNA targeting CHMP6. $n \geq 40$ cells. If not previously specified, all results are shown as mean or representative picture of at least three independent experiments (n.s. = not significant; $^{**}P < 0.01$; $^{***}P < 0.001$).

relatively promiscuously to phosphoinositides, including PI(3,4)P_2 (*13*, *14*). The molecular explanation for this evolutionary change is the presence of additional charged residues in mammalian VPS36 flanking the charged residues responsible for PI(3)P binding in yeast (*36*). We mutated all three of the additional basic residues from the mammalian GLUE domain, partially reestablishing the yeast PI-binding sequence (fig. S16A). Mutant mammalian VPS36 (yMut-VPS36) showed a punctate staining in cells during interphase that resembled that of wild-type VPS36 localization (fig. S16B). Similar to yeast VPS36, yMut-VPS36 showed a threefold increase in colocalization with PI(3)P-positive vesicles (Fig. 6A). Conversely, reduced colocalization of yMut-VPS36 with PI(3,4)P_2 was observed (Fig. 6B). In agreement, liposome binding assays showed reduced binding of yMut-VPS36 to PI(3,4)P_2 and increased association with PI(3)P (Fig. 6C and fig. S16C). Furthermore, a pull-down assay by using PI(3,4)P_2–coated beads revealed a reduced ability of yMut-VPS36 to bind to PI(3,4)P_2 (Fig. 6D). The previously reported H0m mutant of VPS36 (*14*) that lacks binding to ESCRT-I retained its interaction with PI(3,4)P_2 (Fig. 6D).

Considering that suppression of PI3K-C2α decreased PI(3,4)P_2 and VPS36 enrichment at the midbody, we tested whether the reduced binding observed for yMut-VPS36 to PI(3,4)P_2 might be sufficient to prevent its localization to the abscission site. Live cell imaging of GFP-H0m-VPS36 and GFP-yMut-VPS36 showed a 24 and 72% reduction of their localization at the midbody, respectively (Fig. 6E, fig. S16D, and movies S5 to S7). Thus, PI(3,4)P_2 provides a spatial cue that enhances VPS36 localization.

Loss of either PI3K-C2α or VPS36 impairs secondary ingression formation

Defective ESCRT assembly during cytokinesis often results in impaired secondary ingression formation, which is the final cause of delayed abscission (*38*, *39*). Suppression of either PI3K-C2α or VPS36 led to stable and elongated intercellular bridges and a reduced percentage of cells displaying a secondary ingression at one or both sides, respectively (fig. S17, A and B). Although control HeLa cells formed the secondary ingression at an expected distance from the midbody of 0.77 μm (fig. S17C) (*38*), loss of either PI3K-C2α or VPS36 led to an average of 0.90 and 0.88 μm, respectively (fig. S17C). Thus, loss of PI3K-C2α impairs secondary ingression formation.

Recruitment of ESCRT machinery to the midbody causes deformation of the secondary ingression site into wavy patterns of the membrane at the intercellular bridge (*40*). To support the role of PI3K-C2α and VPS36 in the assembly of the ESCRT machinery at the midbody, a kymograph analysis of the sequential wavy constrictions occurring at the membrane flanking the midbody was analyzed by use of live cell imaging. Depletion of either PI3K-C2α or VPS36 caused a significant elongation of the intercellular bridge, likely because of the stretching movement of the two dividing cells (fig. S17D). These intercellular bridges were very stable, with significantly fewer constriction events compared with that of control cells (fig. S17, E and F). No differences in localization

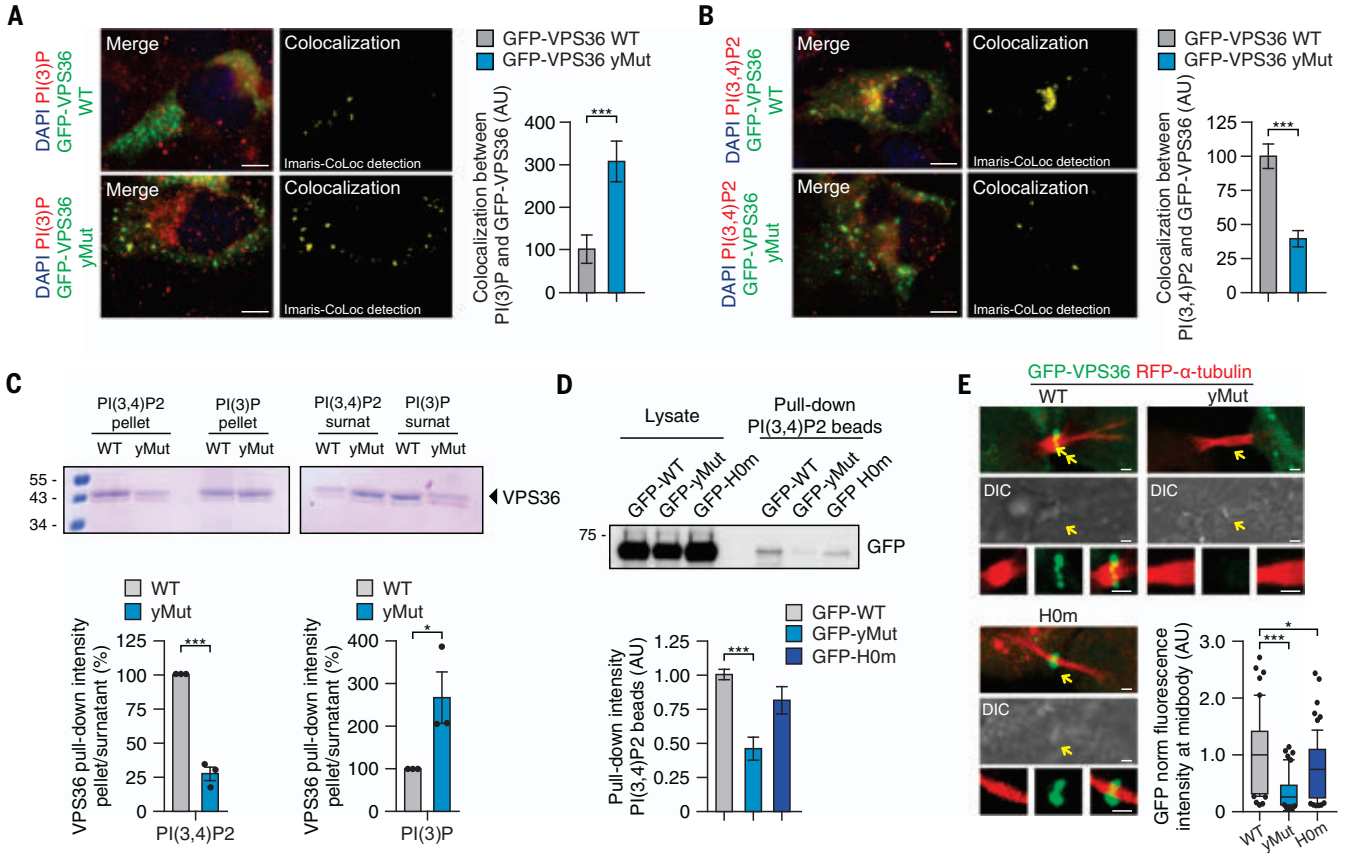


Fig. 6. PI(3,4)P_2 is required for VPS36 recruitment to the midbody. (**A** and **B**) (Left) Immunofluorescence and (right) quantification by using Imaris software of the colocalization between GFP-VPS36 WT or GFP-VPS36-yMut with either (A) PI(3)P or (B) PI(3,4)P_2. (**C**) Lipid sedimentation assay showing binding of WT and yMut VPS36 to PI(3,4)P_2 and PI(3)P. (**D**) Pull-down experiment by using PI(3,4)P_2–coated beads and lysates from cells expressing WT, yMut, or H0m VPS36. (**E**) Time-lapse of HeLa cells stably expressing RFP–α-tubulin, showing localization of WT, yMut, or H0m VPS36 enrichment at the midbody during abscission. Quantification reports the average fluorescence intensity observed in $n \geq 60$ cells and normalized on total GFP fluorescence inside each cell (mean ± SD). If not previously specified, all results are shown as mean or representative picture of at least three independent experiments ± SEM. *$P < 0.05$; ***$P < 0.001$.

at the intercellular bridge of other secondary ingression site effectors, such as FIP3 and F-actin, were observed (fig. S17, G and H).

Loss of VPS36 induces cytokinesis defects and early senescence in the lens

Our data suggest that loss of VPS36 should phenocopy the lack of PI3K-C2α. Analysis of cytokinesis in the lens of zebrafish embryos in which *vps36* was suppressed (fig. S18A) showed that MKLP1-positive midbodies, double stained with either Aurora B (Fig. 7A) or α-tubulin (fig. S5, D and E), were threefold more abundant than in controls (Fig. 7B and fig. S5E). Like with *pik3c2a* suppression, this was not due to increased proliferation because the lens area in *vps36* morphants was 31% smaller than in wild-type controls (Fig. 7C). Consistent with this idea, a reduced number of phospho-H3–positive cells was detected in the lens of *vps36* or *pik3c2a* morphants (fig. S18B), and reduced EdU incorporation was found in HLE-B3–suppressed cells as well as after treatment with the cytokinesis inhibitor Paprotrain (fig. S18C). Furthermore, expression of senescence markers (p16INK4A, SA-β-Gal, p21, BCL2/BAX ratio, and SASP) was more abundant in the eyes of *vps36* morphants than in wild-type controls (Fig. 7, D and E, and fig. S19, A to C). Similar results were observed in VPS36-deficient HLE-B3 cells, where p16INK4A (fig. S19D) and SA-β-Gal (fig. S19E) were increased, with p16INK4A appearing in the nuclei of cells blocked in cytokinesis (fig. S19F). Appearance of senescence markers (SA-β-gal, p16INK4A, and p21), previously observed in *pik3c2a* morphants (Fig. 1, C to E, and fig. S2, C to E), was prevented by a morpholino-resistant wild-type *pik3c2a* mRNA but not by the KD or by the C3 *pik3c2a* mutants (fig. S19, G to I) (*35*). Thus, PI3K-C2α kinase activity specifically linked to cytokinesis is involved in protection from senescence. Last, we tested whether cytokinesis-interfering inhibitors targeting Aurora-B (ZM447439), Myosin-II (blebbistatin), and MKLP2 (Paprotrain) induced senescence. Like in *vps36* morphants, up-regulation of senescence markers (SA-β-Gal, p16INK4A, p21, and SASP) was observed (fig. S19, J and K), indicating that although potentially perturbing distinct biological functions, the common effect of delaying cytokinesis is likely the cause of senescence.

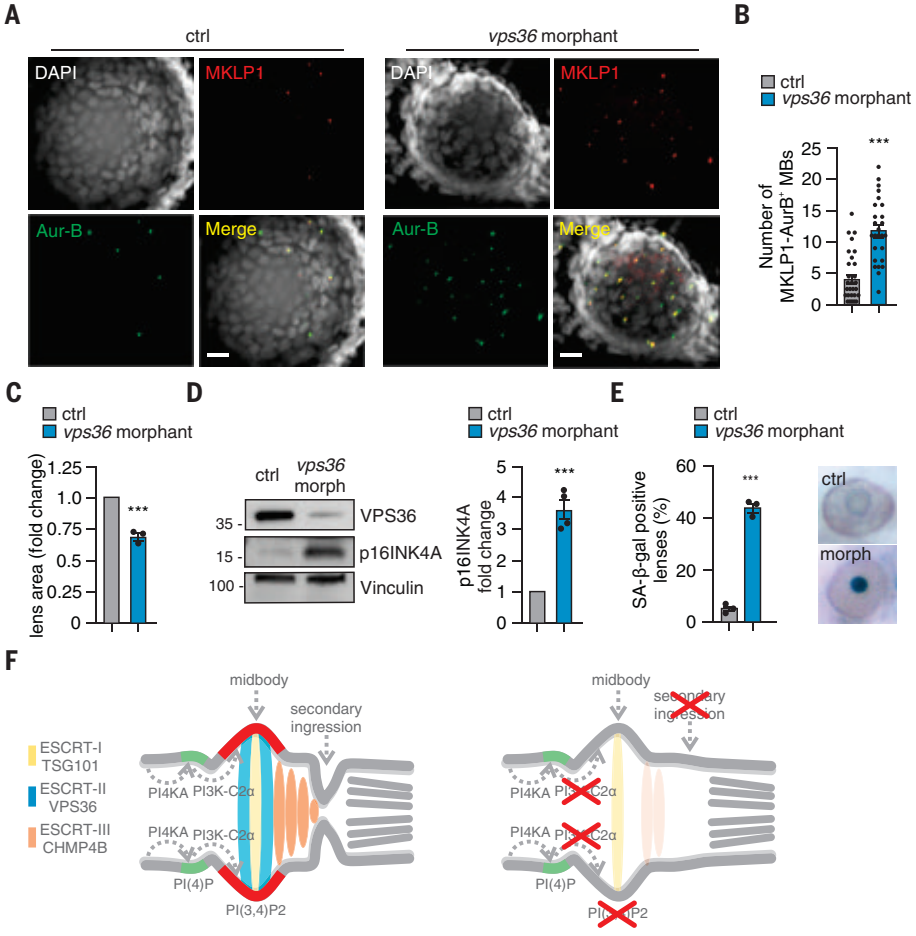


Fig. 7. Loss of VPS36 impairs abscission and leads to senescence in vitro and in vivo. (**A**) Confocal images of whole-mount immunofluorescence performed on 72-hpf embryos lens by using MKLP1 (red) and Aur-B (green) antibodies to stain midbody and TO-PRO-3 (gray) to stain nuclei. (**B**) Quantification of number of MKLP1 and Aur-B–positive midbody in 72-hpf embryos lens. (**C**) Measure of lens area in control and *vps36* morphants. (**D**) Immunoblot analysis and quantification of p16INK4A in control and *vps36* morphant 72-hpf embryos. (n = 4 pools of 15 embryos each). (**E**) Quantification and representative images of SA-β-gal intensity on the lens of control and morphant (*vps36*) 72-hpf embryos. (**F**) During cytokinesis, PI3K-C2α produced PI(3,4)P_2 (red) at the midbody by converting PI(4)P (light blue) synthetized by PI4KA. PI(3,4)P_2 triggers the recruitment and the stabilization of the ESCRT-II subunits VPS36 (green), which in turn contributed to the accumulation of ESCRT-III CHMP4B at the midbody. When PI3K-C2α is lost, the secondary ingression where abscission occurs does not form properly, resulting in impaired abscission. This leads to defective cytokinesis and early onset of senescence, which is particularly evident in the lens epithelium. If not previously specified, all results are shown as mean or representative picture of at least three independent experiments ± SEM. ***P < 0.001.

Discussion

Signs of premature aging—including high prevalence of stroke, bone fragility, and cataract—are a major phenotypical feature of *PIK3C2A*-null patients (*20*). Previous reports identify *PIK3C2A* within the evolutionarily conserved "CellAge" gene network in 24 different mammalian species that, when suppressed, leads to senescence (*41*). In line with these findings, in this work in fish, mice, and humans the loss of PI3K-C2α was directly linked to senescence, with a particularly high penetrance in the eye lens. The lack of the lipid product generated by PI3K-C2α appeared as the driver of a cytokinetic defect, leading to premature up-regulation of senescence markers in the lens. Loss of other enzymes that, like PI3K-C2α, control phosphoinositide turnover can lead to cataract development. For example, inactivating mutations in the two phosphatases OCRL and INPP5K result in early-onset cataracts (*42*, *43*). Both enzymes remove the phosphate in position 5 of PI(4,5)P_2 to generate PI(4)P, the main substrate of PI3K-C2α, indicating that a specific phosphoinositide shift critically protects lens cells from degenerative senescence.

Our observations in different vertebrate species converge on the notion that at the midbody, PI3K-C2α orchestrates the assembly of the ESCRT machinery required for abscission through its lipid product PI(3,4)P_2. The synthesis of specific phosphoinositide species is one of the mechanisms adopted by cells to selectively recruit proteins in defined cellular regions (*44*). Our observations suggest that binding to specific membrane components contributes to the precise positioning of the ESCRT III complex required to complete abscission. The

recruitment of ESCRT-III at the midbody depends on ALIX (*4*), but the presence of other mechanisms that trigger ESCRT III–mediated abscission have been observed (*9*, *11*). For example, ESCRT-III recruitment to the midbody can alternatively be mediated by the ESCRT-I and -II axis (*8*–*10*) and by the localization of the VPS22 subunit of ESCRT-II at the midbody, through the interaction with the ESCRT-I subunit TSG101 (*8*, *9*). ALIX itself is also localized at the midbody by both TSG101 association and the binding to a complex that contains syntenin, bridging ESCRT III to a transmembrane syndecan-4 molecule (*11*). However, how the core ESCRT-II subunit VPS36 could be recruited through this alternative pathway has remained elusive. In light of the inability of VPS22 to bind a select lipid (*14*) and thus localize to a specific subcellular domain, other ESCRT-II components must be involved in the targeting to the midbody-encasing plasma membrane. Our data demonstrate that VPS36 localizes to the midbody through the detection of a confined PI(3,4)P_2 pool.

Although the function of VPS36 is generally conserved throughout evolution, it substantially differs from yeast to mammals regarding the selective binding to phosphoinositides. A study based on comparison of the intact and GLUE domain–deleted ESCRT-II complexes further demonstrates that the human GLUE domain binds with high affinity not only to PI(3)P but also to a variety of other phosphoinositide isomers (*14*). In mammalian cells, PI(3)P is never found on the plasma membrane (*34*), where ESCRT polymerizes (*38*), thus indicating that another phosphoinositide must be involved in the localization at the midbody-enclosing membrane. This likely explains why the removal of PI(3,4)P_2 through the down-modulation of PI3K-C2α can lead to a substantial reduction in VPS36 amounts at the midbody. This modification might have evolved to support specific functions acting in parallel or in place of the ALIX pathway.

The finding that PI3K-C2α and ALIX are not always coexpressed indicates three modalities of ESCRT recruitment in abscission that can rely on either one alone or both of the two proteins together. In most cells, simultaneous presence of the two pathways allows partial compensation, and in cells such as fibroblasts and HeLa, the effect of the loss of PI3K-C2α is mitigated by ALIX activity, as suggested by limited CHMP4B localization at the midbody in the absence of PI3K-C2α. Nonetheless, the compensatory effect of ALIX is incomplete because delayed cytokinesis and senescence are still observed in *PIK3C2A*-null fibroblasts and HeLa cells, thus indicating a potential synergy between the two pathways. In the lens, where ALIX is substantially less expressed than in other tissues, cells mainly depend on PI3K-C2α. Consistent with this idea, ALIX-deficient mice do not show either eye abnormalities or premature aging (*7*), but loss-of-function mutations in ESCRT-III components, such as CHMP4B (*17*, *45*, *46*) and VPS4 (*18*, *19*), lead to altered cell division and premature aging with early-onset cataract. Other alterations of cytokinesis unrelated to defective ESCRT-III can equally result in the development of cataracts. For example, cataract can follow a dysfunction in lens cytoskeletal components acting in cytokinesis, such as vimentin mutation (*47*). The finding that loss of either PI3K-C2α or VPS36 leads to the same cytokinetic defect and the ensuing cellular senescence indicate that of the various functions that these two proteins play in the cell, their common and epistatic role in cytokinesis links them to senescence prevention. Our data that point to disturbance of cell division as a main driver of cataract development consolidate the causative link between aberrant cytokinesis and senescence. Defective cytokinesis has been generally associated with senescence through the activation of the senescence program after tetraploidization (*48*). However, our findings that p16INK4A starts to be expressed by cells blocked before abscission indicate that the senescence program could also be triggered by altered ESCRT function before tetraploidization, likely to prevent cell cycle reentry after refusion.

We defined a nonredundant molecular link between highly localized phosphoinositide signals and ESCRT assembly at the abscission site, together with an evolutionarily conserved mechanism that connects cytokinesis failure to senescence and eventually determines cataract, one of the most common conditions of the elderly population worldwide (Fig. 7F).

Materials and methods

Cell lines and primary cultures

HeLa and HEK293T, were purchased from ATCC (without further authentication) and cultured in DMEM GlutaMAXTM medium supplied with 10% fetal bovine serum (FBS) and 1% Penicillin-Streptomycin (10,000 U/mL). Cell lines used in this paper are not listed in the database of commonly misidentified cell lines maintained by ICLAC. Mouse embryonic fibroblasts (MEFs) were obtained as previously described (*35*). Fibroblast from patients were obtained and cultured as previously described (*20*). HLE-B3 cells were purchased from ATCC and cultured in MEM-α supplied with 20% fetal bovine serum (FBS), 1% Penicillin-Streptomycin (10,000 U/mL) and GlutaMAX. Cell lines were routinely tested for mycoplasma contamination.

Protein analysis

Cells and tissues were homogenized in lysis buffer (120 mM NaCl, 50 mM Tris-HCl pH=8.1% Triton X-100) supplemented with 25x protease inhibitor cocktail (Roche), 50 mM sodium fluoride and 1 mM sodium orthovanadate. Lysates were cleared by centrifugation at 13,000 rpm for 15 min at 4°C. Protein concentration was determined by Bradford method and supernatants were analyzed for immunoblotting or for immunoprecipitation (IP) with the indicated antibodies. Membranes probed with the indicated antibodies were then incubated with HRP conjugated secondary antibodies (anti-mouse used 1:10000, anti-rabbit 1:5000, Sigma) and developed with enhanced chemiluminescence (ECL, BD). For IP assays, 1 mg of precleared extracts were incubated with 1 μg of the indicated antibody at 4°C on a rotating rack. After 1.5 hours, 15 μl of protein G-Sepharose (Amersham Biosciences, Buckinghamshire, UK) were added for 30 min. Samples were collected by centrifugation (13000 rpm 1 min) and washed six-times with lysis buffer. Bound protein complexes were then eluted by adding 30 μl Laemmli sample buffer. For pull-down experiment, HEK293T cells homogenized in lysis buffer (120 mM NaCl, 50 mM Tris-HCl pH = 8.1% Triton X-100) supplemented with 25x protease inhibitor cocktail (Roche), 50 mM sodium fluoride and 1 mM sodium orthovanadate. Lysates were cleared by centrifugation at 13,000 rpm for 15 min at 4°C. 1 μg GST-GBD or 140 ng GST was incubated with 15 μl of protein G-Sepharose for 1 hour at 4°C in 1mg of cell lysate. Beads were washed four times with 1 ml of reaction buffer and analyzed by immunoblotting after the addition of 30 μl of Laemmli buffer. Bound protein complexes were then eluted by adding 30 μl Laemmli sample buffer

Antibodies

Anti-PI3K-C2α (#611046, BD Transduction Laboratories; #22028-1-AP, Proteintech; a rabbit polyclonal develop by VH (*33*)), anti-GFP (gift from Emilia Turco, University of Turin, Italy), anti α-tubulin (#2125, Cell Signaling; #3873, Cell Signaling), anti Myc-tag (#2276, Cell Signaling), anti TSG101 (GTX70255), anti PI(3)P (Z-P003, Echelon), anti PI(3,4)P_2 (Z-P034b, Echelon), anti PI(4)P (Z-P004, Echelon), anti PI(3,4,5)P3 (Z-P345B, Echelon) anti CHMP4B (ab105767, Abcam; sc82556 (C12), Santa Cruz Biotechnology), anti VPS36 (ab76331, Abcam; ab247016, Abcam; #043947, Sigma), anti GFP (ab291, Abcam), anti γ-tubulin (#T5326, GTU-88 Sigma), anti-MKLP-1 (sc-869, Santa Cruz), anti E-Cadherin (4A2, Cell Signaling), anti-ALIX (#92880, Cell Signaling; #634502, BioLegend), anti αA-Crystallin (#PA1-009, Thermo Fisher), anti-β Crystallin (Santa Cruz Biotechnology sc-376006), anti p16INK4A (#SAB4500072, Sigma; 10883-1-AP, Proteintech; #80772, Cell Signaling), anti p21 (Cell Signaling, #2947), anti-Aurora B (GTX132702, GeneTex; BD bioscience, #611082), anti Phospho-Histone H3 (Ser10) (#9701. Cell Signaling), anti GAPDH (Cell Signaling, #5174), anti Vdac (Cell Signaling, #4866), anti-Vinculin

(gift from Emilia Turco, University of Turin, Italy), anti-Citron Kinase (Transduction Laboratories, BD Biosciences, Franklin Lakes, NJ), anti-ALIX and CHMP4B (gift from Harald Stenmark, Institute for Cancer Research, The Norwegian Radium Hospital, Oslo, Norway).

Gene silencing and inhibitors

Plasmids containing shRNA sequences (arrest GIPZ lentiviral shRNA) against *PIK3C2A* were purchased from Thermo Scientific and used to generate lentiviral particles to stably infect cell lines as previously described (*35*). The Sh1 (V2LMM_73461) target sequence was 5′-GGCAAGATATGTTAGCTTT-3′, and the Sh2 (V2LMM_66190) target sequence was 5′-CAAAGTTTCTTTAACTCT-3′. Cells at early passages after infection (second to third) were used for experiments. siRNA treatment against *PIK3C2A* was previously described (*33*). siRNA against *VPS36* was purchased by Thermo Fischer (cd: 134802 20nM). siRNA against *OCRL* was purchased by Thermo Fischer (cd: 104448 20nM). siRNA against *CHMP6* was purchased by Sigma (esiRNA human CHMP6 EHU144401). The sequence of siRNA against *TSG101* was 5′-CCUCCAGUCUUCUCUCGUCUU-3′. The sequence of siRNA against *CHMP4B* was 5′-AUCGAUAAAGUUGAUGAGUUAUUUU-3′. The sequence of siRNA against *ALIX* was 5′CCUGGAUAAUGAUGAAGGA3′. The sequence of siRNA against *VPS36* was 5′- AAGUGAAUGCCAAUAUGAA-3′.

VPS34-IN1 was purchased by Sigma (#532628). Paprotrain was purchased from Abcam (ab144322). Blebbistatin was purchased from Sigma (#B0560). ZM 447439 was purchased from TOCRIS (#2458). The PI4K A1 inhibitor was used as described (*49*). For experiments in zebrafish, Tuebingen zygotes were treated with 40 μM PITCOIN1 in standard fish water. Medium was changed every 24 hours, embryos were sacrificed at 72 hpf for further analysis. For fibroblasts, 60 μM PITCOIN1 was added to culture medium every 24 hours for a maximum of four days. For inhibitor treatment, zebrafish embryos were treated from 48 to 72 hpf with blebbistatin (B05660, Sigma-Aldrich), paprotrain (512533, Sigma-Aldrich), and ZM447439 (189410, Sigma-Aldrich) at a concentration of 5, 100, and 40 μM, respectively.

Plasmids

Wild-type PI3K-C2α-GFP, kinase inactive (KD) PI3K-C2α-GFP, clathrin-binding deleted PI3K-C2α-GFP (Δ1–380) were previously described (*31*, *33*). PI3K-C2α-Myc construct was generated by cutting and pasting equivalent fragments from the PI3K-C2α-GFP mutants into pCDNA3.1 plasmid. Human VPS36 and CHMP4B were obtained by RLW and inserted in pEGFP empty vector. For mCherry-VPS36 plasmid generation, EGFP was replaced with mCherry by directly cloning mCherry sequence using restriction enzymes. GST-GBD and GFP-GBD were obtained by PCR amplification from WT PI3K-C2α of the predicted sequence and by cloning in either pGEX 5X1 or pEGFP-C2. GST-GBDQ1022A-T1025A-S1081A and GFP-GBDQ1022A-T1025A-S1081A were obtained by site-directed mutagenesis using QuikChange II Site-Directed Mutagenesis Kit (Agilent). GFP-VPS36 H0m and GFP-VPS36 yMut were generated by QuikChange II Site-Directed Mutagenesis Kit (Agilent). Recombinant VPS36 (WT and yMut), VPS25 and VPS22 were generated as described (*14*). mCherry-hALIX was a gift from James Hurley (Addgene plasmid # 21504; http://n2t.net/addgene:21504; RRID: Addgene_21504). CHMP6_GFP was a gift from Daniel Gerlich (Addgene plasmid # 31806; http://n2t.net/addgene:31806 ; RRID:Addgene_31806). All constructs were verified by restriction digest and automated DNA sequencing. HEK293T cells and HeLa cells were transfected by lipofection using Lipofectamine® 2000 (Life Technologies) or X-tremeGEN HP DNA Transfection Reagent (Roche, XTGHP-RO), according to the manufacturer's instructions. For silencing experiments, HeLa and HLE-B3 cells were transfected with 25 nM siRNAs for 48-72 hours using X-tremeGENE 360 Transfection Reagent (Roche) or Lipofectamine RNAiMAX (Invitrogen), according to the manufacturer's instructions.

Lipid sedimentation assay and pull-down using PI(3,4)P_2-coated beads

Lipid sedimentation assay was performed in accordance with what previously described in Im *et al.* 2008 (*14*). Pull-down using PI(3,4)P_2-coated beads was performed using cell lysate of HeLa cells transfected with the indicated plasmid, in accordance with manufacturer instructions (Echelon, P-B034a).

Immunofluorescence

Immunofluorescence was performed by ice cold methanol or PFA fixation of MEFs, human fibroblasts, HLE-B3 and HeLa cells followed by standard procedures (*31*, *35*). Cells were stained with DAPI, TO-PRO-3 iodide (Thermo Fisher) or Sytox Green (Thermo Fisher) and examined with either Zeiss Observer-Z1 microscope, equipped with the Apotome, Leica TCS-II SP5 or Leica TSC-II SP8 confocal microscope. Zebrafish embryos were fixed overnight in PFA 4%, nuclei were stained with TO-PRO-3 iodide (Thermo Fisher). Whole mounted embryos were examined with Leica TSC-II SP5 or Leica TSC-II SP8 confocal microscope. Raw images were digitally processed only to normalize the background and enhance the contrast. Z-stacks were acquired and processed with the Maximum Projection tool. 3D morphometric measurement and reconstruction was performed with Imaris (BitPlane, Zurich, Switzerland). Fluorescence intensity was calculated with ImageJ tools, modifying existing protocols (*50*), (https://theolb.readthedocs.io/en/latest/imaging/measuring-cell-fluorescence-using-imagej.html) and as outlined in fig. S20. Whisker plots show either 90/10, 95/5 or 97.5/2.5 percentile, and the median in the center line.

Time-lapse microscopy

For measurement of the time required to progress from anaphase onset to abscission, cells were taken in a humidified chamber, at 37°C and 5% CO2 and imaged every 4 min in accordance with what previously described (*31*). For time-lapse of GFP-VPS36, mCherry-VPS36 and GFP-CHMP4B, either Zeiss Observer-Z1 microscope or Leica TSC-II SP8 confocal microscope were used and HeLa cells were imaged every 30 s for at least 10 min. For time-lapse of GFP-PI3K-C2α, Leica TSC-II SP8 confocal microscope was used, and HeLa cells were imaged every 2 min for at least 2 hours.

Cell synchronization and midbody purification

Cell synchronization and midbody purification were adapted from (*51*, *52*). Briefly, to synchronize cells in cytokinesis, we used a thymidine-nocodazole block and release procedure. To purify midbodies, 5×10^7 HeLa or HEK-293T cells were used. Cells were synchronized using the thymidine-nocodazole block. After nocodazole washout cells were incubated for 2 h in fresh medium containing 10 μM MG132 (Sigma-Aldrich) to further increase the effectiveness of the synchronization, and then incubated at 37 °C for 60 min after release from MG132. Before collection, 5 μg/ml taxol (Sigma-Aldrich) was added to the medium for 4 min to stabilize microtubules in vivo. Cells were then transferred into a Corning 50 ml centrifuge tube and collected by centrifugation at 200 × g for 5 min. Cells were then washed once in H2O and gently resuspended in 25 ml of swelling solution (1 mM PIPES pH 7.0, 1 mM MgCl2, 5 μg/ml taxol and Roche Complete Protease Inhibitors) and centrifuged at 200 × g for 5 min. The cell pellet was then resuspended in 50 ml of lysis buffer (1 mM PIPES pH 7, 1 mM EGTA, 1% NP-40, 5 μg/ml taxol, 3 U/ml DNAse I, 10 μg/ml RNAse A, 1 U/ml micrococcal nuclease, and Roche Complete Protease Inhibitors) and vortexed vigorously. After the addition of 0.3 volumes of cold 50 mM 2-(N-mopholino)ethanesulfonic acid (MES) pH 6.3, the sample was incubated on ice for 20 min and then centrifuged at 200 × g for 10 min at 4 °C. The supernatant was transferred to a new tube and centrifuged at 650 × g for 20 min at 4 °C to pellet midbodies. The midbody pellet was then resuspended in 4 ml of 50 mM MES pH 6.3 and centrifuged through a 25 ml glycerol cushion (40% [w/v] glycerol diluted in 50 mM MES pH 6.3) at 2800 × g for 45 min at 4 °C. After removal of the glycerol cushion, the midbody pellet was washed with 2 ml of 50 mM MES pH 6.3,

transferred to a 15 ml conical tube and centrifuged at 2800 × g for 20 min at 4 °C. After removing as much liquid as possible, the midbody pellet was resuspended in Laemmli buffer for immunoblot analysis.

Zebrafish strains and treatments

All procedures using zebrafish (Danio rerio) were authorized by the Ethical Committee of the University of Torino and the Italian Ministry of Health. The wild-type fish strain Tuebingen was used. Adult fish were routinely maintained under a 14h light and 10h dark photoperiod at approximately 28°C, bred and genotyped according to standard procedures. Allelic transmission followed the expected mendelian ratios. Eggs were generated by natural mating, and following fertilization were collected, treated and maintained under a 12h light and 12h dark photoperiod, incubated at 28°C. Embryos and adult fish were sacrificed with a tricaine overdose. Embryos with the alleles $^{sa10124/+}$ and $^{sa12328/+}$ were created as part of the Zebrafish Mutation Project (*51*) and were obtained by in vitro fertilization from frozen sperm samples by the Zebrafish International Resource Center. The alleles sa10124 and sa12328 encode nonsense mutations in PI3K-C2α at amino acids 585 and 1236, respectively, that were confirmed by Sanger sequencing. Both nonsense mutations occurred prior to the end of the catalytic domain and were thus predicted to encode null alleles. We generated homozygous *pik3c2a*$^{sa12328/sa12328}$ and *pik3c2a*$^{sa10124/sa10124}$ zebrafish, as well as compound heterozygous *pik3c2a*$^{sa12328/sa10124}$ mutants by intercrossing heterozygous adults. As these alleles were generated by random ENU mutagenesis (*51*), analysis of the compound heterozygous *pik3c2a*$^{sa12328/sa10124}$ mutants minimized the likelihood of homozygosity for any unlinked ENU-induced mutations.

Cataract formation in zebrafish

Adult zebrafish were anesthetized with tricaine and examined. Coaxial illumination using a Leica M841 surgical microscope was used to visualize lenticular defects. Digital video recordings were made using a Panasonic GP-US932A HD camera system and later reviewed by an ophthalmologist without knowledge of the genotype of each animal. Optical sectioning by changing the z-axis focus was used to aid in the identification of cataracts. All zebrafish examined for cataracts were offspring of a cross between *pik3c2a*$^{sa10124/+}$ and *pik3c2a*$^{sa12328/+}$ fish.

Morpholino microinjection

Antisense Morpholino Oligonucleotides against *pik3c2a*, *vps36* and scrambled control were purchased from GeneTools (LLC, Philomat, OR, USA). The *pik3c2a*-MO sequence was 5′-TATGTGGGCCATGGTGTCAGCTCT -3′. The *vps36*-MO sequence was 5′- CATTTGTCCA-CATAAATCGGTCCAT -3′. The scrambled control-MO sequence was 5′-CCTCTTACCT-CAGTTACAATTTATA -3′ Tuebingen zygotes were collected at 1-cell stage and injected under stereological examination with 400 μM of *pik3c2a*-MO, *vps36*-MO or scrambled control-MO in the presence of Phenol Red for subsequent selection of the injected embryos.

Rescue experiments in zebrafish

For the rescue experiment, plasmids containing a GFP-tagged zebrafish transcript variant of *pik3c2a* (XM_021467821.1), modified with silent mutations in the region complementary to morpholino oligonucleotide, was constructed and packaged by VectorBuilder. For the mutagenesis, morpholino-resistant plasmids were obtained by site-directed mutagenesis using QuikChange II Site-Directed Mutagenesis Kit (Agilent) as previously described (*33*, *35*). Capped mRNA was transcribed with mMESSAGE mMACHINE T7 Transcription Kit (Invitrogen). Zebrafish embryos were then injected at 1-cell-stage with morpholino oligonucleotides and mRNA (500 ng/μL). Positive embryos were selected based on GFP-fluorescence.

Lens size measurement

Zebrafish embryos were treated with 0.003% 1-phenyl-2-thiourea (#P7629, Sigma) at 12 hpf to prevent formation of melanin pigment. Picture of fixed 72-hpf larvae were taken under stereological examination using fixed magnification. Images were analyzed using ImageJ tools. Lens size was quantified as ratio between lens and eye diameters.

SA-β-galactosidase assay

Cells or embryos were stained using Senescence Cells Histochemical Staining Kit (#CS00, Sigma), according to the manufacturer's instructions. Cell were fixed for 1 hour at room temperature, washed 3 times and incubated 3 hours in β-Gal staining solution at 37°C in absence of CO2. Zebrafish embryos were treated with 0.003% 1-phenyl-2-thiourea (#P7629, Sigma) at 12hpf to prevent formation of melanin pigment, then 72-hpf larvae were fixed at 4°C overnight, washed 3 times with PBS and incubated in β-Gal staining solution at 37°C for 12 hours in absence of CO2.

Flow cytometry analysis

Heads from 72-hpf zebrafish embryos were disaggregated by mechanical homogenization at 30°C in 0.25% trypsin containing 5 mg/mL collagenase. Enzymatic reaction was blocked by addition of DMED medium containing 10% FBS. Disaggregation products were washed in PBS and fixed by dropwise addition of absolute ethanol while vortexing reaching a final concentration of 70% ethanol, cells were incubated 30 min in ice. After further PBS washes, cells were resuspended in PI-staining solution containing 25 μg/mL propidium iodide (P4864, Sigma), 0.05% Triton-X 100 and 0.1 mg/mL RNAse in PBS and incubated at 37°C for 40 min. Samples were washed in PBS and analyzed using BD FACSVerse© flow cytometer. Cell cycle was analyzed quantifying incorporated propidium iodide using FCSalyzer ver.0.9.22-alpha (http://sourceforge.net/projects/FCSalyzer).

Quantitative RT-PCR

Total RNA was extracted using TRIzol reagent (Invitrogen, Carlsbad, CA). cDNA was synthesized from 1000 ng of total RNA using cDNA reverse transcription kits (Applied Biosystems, Foster City, CA). Relative mRNA level was analyzed by real time PCR (ABI 7900HT FAST Real-Time PCR system, Applied Biosystems, Foster City, CA) with SYBR Green master mix (Applied Biosystems). *GAPDH* gene was used as housekeeping control. The primers are listed in Table S3.

EdU assay

The EdU staining on cells was done using Click-iT® EdU Alexa Fluor® 647 HCS Assay (C10419, Invitrogen). Cells were incubated for 60 min in complete medium supplemented with 10 μM EdU. Cells labeling, fixation and detection was done following manufacturer's instructions.

Immunohistochemistry

Immunohistochemistry was performed on formalin-fixed, paraffin-embedded tissue sections after hydration. After blocking in TBS with 0.1% Triton-X 100, 5% BSA and 1% goat serum, primary antibody incubation was carried overnight in in humified chamber at 4°C, dilutions followed manufacturer's instructions. The Novolink Polymer detection system (Leica) was used for visualization, as described in manufacturer's instructions. Slides were then stained with hematoxylin and eosin then dehydrated, cleared and cover slipped.

Mouse strains and treatments

All mice were from C57/Bl6 background. They were born healthy and according to mendelian ratio. Animals were kept in SPF rooms in temperature and light controlled environment and fed standard chow diet ad libitum and had free access to water. All the animal use followed institutional animal welfare guidelines and legislation, as approved by the local Animal Ethics Committee (Comitato di Bioetica e Valutazione, Torino, Italy). *Pik3c2a*$^{hypo/hypo}$ were generated at Lexicon Pharmaceuticals by gene trapping as described (*23*). To generate *Pik3c2a*$^{fl/fl}$ mice, *Pik3c2a*$^{tm1a(EUCOMM)Hmgu}$ mouse embryonic stem cells (clone HEPD0636_

4_E04) were received from the International Mouse Phenotyping Consortium and used to generate chimeric mice. *Pik3ca*$^{fl/fl}$ mice were finally obtained from germ-line transmitters intercrossed with transgenic mice expressing the Flp recombinase under the β-actin promoter [B6.Cg-Tg(ACTFLPe)9205Dym/J, Jackson Laboratory] and eventually eliminating lacZ/neomycin cassettes from the *Pik3c2a* locus. Subsequently, *Pik3c2a*$^{fl/fl}$ mice were crossed with *Cre* transgenic mice [B6.Cg-Tg(CAG-cre/Esr1*)5Amc/J, Jackson Laboratory] to generate *Pik3c2a*$^{fl/fl}$;*Cre*$^{+}$ and *Pik3c2a*$^{fl/fl}$;*Cre*$^{-}$ mice. Cre recombinase was activated using 50 mg/kg of tamoxifen (Sigma-Aldrich) in corn oil, via a daily eye drop administration for 14 days. *Pik3c2a*$^{fl/fl}$;*Cre*$^{-}$ mice received the same tamoxifen treatment and were used as experimental controls.

Computational modelling

Structure predictions were performed using the I-Tasser server without additional constraints or templates (*52*). The following amino acid sequence of the helical domain of PI3K-C2α was used (Uniprot id: O00443, aa 861-1397), helical domain (aa. 861-1037) and PI3K/PI4K domain (aa. 1133-1397) are underlined while predicted sequence (P sequence - aa. 980-1203) is in bold: QHNLETLENDIKGKLLDILHKDSSLGLSKEDKAFLWEKRYYCFKHPNCLPKILASAPNWKWVNLAKTYSLLHQWPALYPLIALELLDSKFADQEVRSLAVTWIEAISDDELTDLLPQ**FVQALKYEIYLNSSLVQFLLSRALGNIQIAHNLYWLLKDALHDVQFSTRYEHVLGALLSVGGKRLREELLKQTKLVQLLGGVAEKVRQASGSARQVVLQRSMERVQSFFQKNKCRLPLKPSLVAKELNIKSCSFFSSNAVPLKVTMVNADPMGEEINVMFKVGEDLRQDMLALQMIKIMDKIWLKEGLDLRMVIFKCLSTGRDRGMVELVPASDTLRKIQVEYGVT**GSFKDKPLAEWLRKYNPSEEEYEKASENFIYSCAGCCVATYVLGICDRHNDNIMLRSTGHMFHIDFGKFLGHAQMFGSFKRDRAPFVLTSDMAYVINGGEKPTIRFQLFVDLCCQAYNLIRKQTNLFLNLLSLMIPSGLPELTSIQDLKYVRDALQPQTTDAEATIFFTRLIESSLGSIATKFNFFIHNLAQLR. Based on amino acid sequence, three dimensional atomic models were generated from multiple threading alignment and iterative structural assembly simulations. The output from the server run contained full-length secondary and tertiary structure predictions. An estimate of accuracy of the predictions is provided based on the confidence score of the modeling (*52*). The TM-scores and root mean square deviations (RMSDs) of the generated structures with respect to the template structures were calculated.

Crystal structure of human γ-tubulin was obtained from PDB deposited under ID: 3CB2 (resolved at 2.3 Å) (*53*). PI3K-C2α/γ-tubulin interaction was studied using High Ambiguity Driven biomolecular DOCK-ing (HADDOCK) webserver (https://wenmr.science.uu.nl/haddock2.4/). HADDOCK webserver offers protein–protein ab initio dockings based on random patch definition and center of mass restraints. Possible residues of interfaces in each structure were predicted using the Consensus Prediction Of Interface Residues in Transient complexes facility and integrated into HADDOCK. Then, two hundred complexes were generated by the HADDOCK program and clustered. Selection of the best complex was based on cluster size, HADDOCK score and electrostatic energy. Among the ten best clusters, we selected a cluster that was the second most populated cluster but was characterized by highest HADDOCK score and lowest electrostatic energy. PISA was used to analyze the protein-protein docking and binding interfaces. The illustrations and visualizations of the 3D models were produced in PyMOL (version 1.7.1.7).

Proteomic analysis of PI3K-C2α

HEK293 cells (0.8×10^6) were transfected with eGFP (mock), eGFP-PI3K-C2α (full length), or eGFP-PI3K-C2α ΔNT (amino acids 377-1686, i.e., a mutant lacking the clathrin binding N-terminal domain) using calcium-phosphate. 16 hours post-transfection cells were harvested and washed twice with ice-cold PBS. The cells were resuspended with 1ml lysis buffer (20 mM HEPES-NaOH pH 7.4, 130 mM NaCl, 10 mM NaF, 1 tablet/50 ml EDTA-free protease inhibitor cocktail, 0.05% Saponin). Cell debris was removed by centrifugation at 16,000 × g for 10 min. Cell lysates containing 1.5 mg of total protein were used for affinity-purification using 10 ml of GFP-Trap_MA beads (ChromoTek). Samples were incubated while rotating for 1h at 4°C. The beads were washed 3x with lysis buffer. Bound proteins were eluted with 2x SDS-PAGE loading buffer and separated by 4-12% SDS-PAGE. Each lane was cut into 12 slices, before in-gel digestion with trypsin was performed. Tryptic peptides were analyzed by reversed-phase capillary liquid chromatography system connected to an Orbitrap Elite mass spectrometer (ThermoScientific). The identification and label-free quantification of proteins was performed by MaxQuant software (Version 1.5.2.8). The LFQ intensities of the identified proteins were used to identify PI3K-C2α-associated proteins.

Development of PI3K-C2α inhibitor

We capitalized on the availability of active recombinant PI3K-C2α (Wang, Lo *et al.*, 2018) to conduct high-throughput screening of > 37,000 small molecules from an in-house library followed by iterative rounds of medicinal chemistry optimization and selectivity profiling. These experiments identified a potent PI3K-C2α-selective small molecule inhibitor that we named PITCOIN1 (for phosphatidylinositol three-kinase class two inhibitor 1). PITCOIN1 inhibits PI3K-C2α with an IC_{50} of 95 nM but is inactive against class I PI3Ks or Vps34 and does not display off-target activity toward a panel of > 100 kinases including related lipid kinases. Details are reported in a separate study in preparation.

General experimental approaches

All the statistical details of experiments can be found in the figure legends, in the Results and the Methods sections, including the statistical tests used, exact value of n, what n represents (cells, experiments) and precision measures (mean, median, SD, SEM, confidence intervals). No statistical methods were used to predetermine sample size. No samples or data points were excluded from the reported analyses. Samples were not randomized to experimental groups. Sample size was determined on the basis of our previous studies (Franco *et al.*, 2014, Franco *et al.*, 2015). The investigators were not blinded to allocation during experiments and outcome assessment.

Statistical analysis

Prism software (GraphPad) was used for statistical analysis. Significance was calculated with Student *t* test and one- or two-way analysis of variance tests (ANOVA) followed by Bonferroni's post hoc analysis, or Mantel Cox log-rank test where appropriate. Values are reported as the mean ± either standard error of the mean (SEM) or standard deviation (SD). $P < 0.05$ was considered statistically significant (∗), $P < 0.01$ very significant (∗∗), and extremely significant $P < 0.001$ (∗∗∗).

REFERENCES AND NOTES

1. F. Gulluni, M. Martini, E. Hirsch, Cytokinetic abscission: Phosphoinositides and ESCRTs direct the final cut. *J. Cell. Biochem.* **118**, 3561–3568 (2017). doi: 10.1002/jcb.26066; pmid: 28419521
2. O. Schmidt, D. Teis, The ESCRT machinery. *Curr. Biol.* **22**, R116–R120 (2012). doi: 10.1016/j.cub.2012.01.028; pmid: 22361144
3. M. A. Y. Adell, D. Teis, Assembly and disassembly of the ESCRT-III membrane scission complex. *FEBS Lett.* **585**, 3191–3196 (2011). doi: 10.1016/j.febslet.2011.09.001; pmid: 21924267
4. J. G. Carlton, M. Agromayor, J. Martin-Serrano, Differential requirements for Alix and ESCRT-III in cytokinesis and HIV-1 release. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10541–10546 (2008). doi: 10.1073/pnas.0802008105; pmid: 18641129
5. M. S. Bhutta, C. J. McInerny, G. W. Gould, ESCRT function in cytokinesis: Location, dynamics and regulation by mitotic kinases. *Int. J. Mol. Sci.* **15**, 21723–21739 (2014). doi: 10.3390/ijms151221723; pmid: 25429432
6. J. H. Hurley, ESCRTs are everywhere. *EMBO J.* **34**, 2398–2407 (2015). doi: 10.15252/embj.201592484; pmid: 26311197
7. M. H. Laporte *et al.*, Alix is required during development for normal growth of the mouse brain. *Sci. Rep.* **7**, 44767 (2017). doi: 10.1038/srep44767; pmid: 28322231
8. L. Christ, C. Raiborg, E. M. Wenzel, C. Campsteijn, H. Stenmark , Cellular functions and molecular mechanisms of the ESCRT membrane-scission machinery. *Trends Biochem. Sci.* **42**, 42–56 (2017). doi: 10.1016/j.tibs.2016.08.016; pmid: 27669649
9. L. Christ *et al.*, ALIX and ESCRT-I/II function as parallel ESCRT-III recruiters in cytokinetic abscission. *J. Cell Biol.* **212**, 499–513 (2016). doi: 10.1083/jcb.201507009; pmid: 26929449

10. I. Goliand, D. Nachmias, O. Gershony, N. Elia, Inhibition of ESCRT-II-CHMP6 interactions impedes cytokinetic abscission and leads to cell death. *Mol. Biol. Cell* **25**, 3740–3748 (2014). doi: 10.1091/mbc.e14-08-1317; pmid: 25232011
11. C. Addi *et al*., The Flemmingsome reveals an ESCRT-to-membrane coupling via ALIX/syntenin/syndecan-4 required for completion of cytokinesis. *Nat. Commun.* **11**, 1941 (2020). doi: 10.1038/s41467-020-15205-z; pmid: 32321914
12. J. H. Hurley, The ESCRT complexes. *Crit. Rev. Biochem. Mol. Biol.* **45**, 463–487 (2010). doi: 10.3109/10409238.2010.502516; pmid: 20653365
13. H. Teo *et al*., ESCRT-I core and ESCRT-II GLUE domain structures reveal role for GLUE in linking to ESCRT-I and membranes. *Cell* **125**, 99–111 (2006). doi: 10.1016/j.cell.2006.01.047; pmid: 16615893
14. Y. J. Im, J. H. Hurley, Integrated structural model and membrane targeting mechanism of the human ESCRT-II complex. *Dev. Cell* **14**, 902–913 (2008). doi: 10.1016/j.devcel.2008.04.004; pmid: 18539118
15. P. G. Sreekumar, D. R. Hinton, R. Kannan, The emerging role of senescence in ocular disease. *Oxid. Med. Cell. Longev.* **2020**, 2583601 (2020). doi: 10.1155/2020/2583601; pmid: 32215170
16. M. Matsuyama *et al*., Defect of mitotic vimentin phosphorylation causes microophthalmia and cataract via aneuploidy and senescence in lens epithelial cells. *J. Biol. Chem.* **288**, 35626–35635 (2013). doi: 10.1074/jbc.M113.514737; pmid: 24142690
17. Y. Zhou, T. M. Bennett, A. Shiels, A charged multivesicular body protein (CHMP4B) is required for lens growth and differentiation. *Differentiation* **109**, 16–27 (2019). doi: 10.1016/j.diff.2019.07.003; pmid: 31404815
18. C. Rodger *et al*., De novo VPS4A mutations cause multisystem disease with abnormal neurodevelopment. *Am. J. Hum. Genet.* **107**, 1129–1148 (2020). doi: 10.1016/j.ajhg.2020.10.012; pmid: 33186545
19. K. G. Seu *et al*., VPS4A mutations in humans cause syndromic congenital dyserythropoietic anemia due to cytokinesis and trafficking defects. *Am. J. Hum. Genet.* **107**, 1149–1156 (2020). doi: 10.1016/j.ajhg.2020.10.013; pmid: 33186543
20. D. Tiosano *et al*., Mutations in PIK3C2A cause syndromic short stature, skeletal abnormalities, and cataracts associated with ciliary dysfunction. *PLOS Genet.* **15**, e1008088 (2019). doi: 10.1371/journal.pgen.1008088; pmid: 31034465
21. V. Gorgoulis *et al*., Cellular senescence: Defining a path forward. *Cell* **179**, 813–827 (2019). doi: 10.1016/j.cell.2019.10.005; pmid: 31675495
22. F. Gulluni, M. C. De Santis, J. P. Margaria, M. Martini, E. Hirsch, Class II PI3K functions in cell biology and disease. *Trends Cell Biol.* **29**, 339–359 (2019). doi: 10.1016/j.tcb.2019.01.001; pmid: 30691999
23. D. P. Harris *et al*., Requirement for class II phosphoinositide 3-kinase C2α in maintenance of glomerular structure and function. *Mol. Cell. Biol.* **31**, 63–80 (2011). doi: 10.1128/MCB.00468-10; pmid: 20974805
24. J. D. Rhodes *et al*., Activation of the innate immune response and interferon signalling in myotonic dystrophy type 1 and type 2 cataracts. *Hum. Mol. Genet.* **21**, 852–862 (2012). doi: 10.1093/hmg/ddr515; pmid: 22062891
25. E. Morita *et al*., Human ESCRT and ALIX proteins interact with proteins of the midbody and function in cytokinesis. *EMBO J.* **26**, 4215–4227 (2007). doi: 10.1038/sj.emboj.7601850; pmid: 17853893
26. S. Tcherniuk *et al*., Relocation of Aurora B and survivin from centromeres to the central spindle impaired by a kinesin-specific MKLP-2 inhibitor. *Angew. Chem. Int. Ed.* **49**, 8228–8231 (2010). doi: 10.1002/anie.201003254; pmid: 20857469
27. J. A. Brill, R. Wong, A. Wilde, Phosphoinositide function in cytokinesis. *Curr. Biol.* **21**, R930–R934 (2011). doi: 10.1016/j.cub.2011.10.001; pmid: 22115464
28. H. Wang *et al*., Autoregulation of class II α PI3K activity by its lipid-binding PX-C2 domain module. *Mol. Cell* **71**, 343–351.e4 (2018). doi: 10.1016/j.molcel.2018.06.042; pmid: 30029007
29. G. D. Gupta *et al*., A dynamic protein interaction landscape of the human centrosome-cilium interface. *Cell* **163**, 1484–1499 (2015). doi: 10.1016/j.cell.2015.10.065; pmid: 26638075
30. M. L. Fanarraga, J. Bellido, C. Jaén, J. C. Villegas, J. C. Zabala, TBCD links centriologenesis, spindle microtubule dynamics, and midbody abscission in human cells. *PLOS ONE* **5**, e8846 (2010). doi: 10.1371/journal.pone.0008846; pmid: 20107510
31. F. Gulluni *et al*., Mitotic spindle assembly and genomic stability in breast cancer require PI3K-C2α scaffolding function. *Cancer Cell* **32**, 444–459.e7 (2017). doi: 10.1016/j.ccell.2017.09.002; pmid: 29017056
32. A. P. Sagona *et al*., PtdIns(3)P controls cytokinesis through KIF13A-mediated recruitment of FYVE-CENT to the midbody. *Nat. Cell Biol.* **12**, 362–371 (2010). doi: 10.1038/ncb2036; pmid: 20208530
33. Y. Posor *et al*., Spatiotemporal control of endocytosis by phosphatidylinositol-3,4-bisphosphate. *Nature* **499**, 233–237 (2013). doi: 10.1038/nature12360; pmid: 23823722
34. T. Balla, Phosphoinositides: Tiny lipids with giant impact on cell regulation. *Physiol. Rev.* **93**, 1019–1137 (2013). doi: 10.1152/physrev.00028.2012; pmid: 23899561
35. I. Franco *et al*., PI3K class II α controls spatially restricted endosomal PtdIns3P and Rab11 activation to promote primary cilium function. *Dev. Cell* **28**, 647–658 (2014). doi: 10.1016/j.devcel.2014.01.022; pmid: 24697898
36. S. L. Alam *et al*., Structural basis for ubiquitin recognition by the human ESCRT-II EAP45 GLUE domain. *Nat. Struct. Mol. Biol.* **13**, 1029–1030 (2006). doi: 10.1038/nsmb1160; pmid: 17057716
37. S. Hirano *et al*., Structural basis of ubiquitin recognition by mammalian Eap45 GLUE domain. *Nat. Struct. Mol. Biol.* **13**, 1031–1032 (2006). doi: 10.1038/nsmb1163; pmid: 17057714
38. N. Elia, G. Fabrikant, M. M. Kozlov, J. Lippincott-Schwartz, Computational model of cytokinetic abscission driven by ESCRT-III polymerization and remodeling. *Biophys. J.* **102**, 2309–2320 (2012). doi: 10.1016/j.bpj.2012.04.007; pmid: 22677384
39. B. Mierzwa, D. W. Gerlich, Cytokinetic abscission: Molecular mechanisms and temporal control. *Dev. Cell* **31**, 525–538 (2014). doi: 10.1016/j.devcel.2014.11.006; pmid: 25490264
40. J. A. Schiel *et al*., FIP3-endosome-dependent formation of the secondary ingression mediates ESCRT-III recruitment during cytokinesis. *Nat. Cell Biol.* **14**, 1068–1078 (2012). doi: 10.1038/ncb2577; pmid: 23000966
41. R. A. Avelar *et al*., A multidimensional systems biology analysis of cellular senescence in aging and disease. *Genome Biol.* **21**, 91 (2020). doi: 10.1186/s13059-020-01990-9; pmid: 32264951
42. M. A. De Matteis, L. Staiano, F. Emma, O. Devuyst, The 5-phosphatase OCRL in Lowe syndrome and Dent disease 2. *Nat. Rev. Nephrol.* **13**, 455–470 (2017). doi: 10.1038/nrneph.2017.83; pmid: 28669993
43. M. Wiessner *et al*., Mutations in INPP5K, encoding a phosphoinositide 5-phosphatase, cause congenital muscular dystrophy with cataracts and mild cognitive impairment. *Am. J. Hum. Genet.* **100**, 523–536 (2017). doi: 10.1016/j.ajhg.2017.01.024; pmid: 28190456
44. C. Cauvin, A. Echard, Phosphoinositides: Lipids with informative heads and mastermind functions in cell division. *Biochim. Biophys. Acta* **1851**, 832–843 (2015). doi: 10.1016/j.bbalip.2014.10.013; pmid: 25449648
45. A. P. Sagona, I. P. Nezis, H. Stenmark, Association of CHMP4B and autophagy with micronuclei: Implications for cataract formation. *BioMed Res. Int.* **2014**, 974393 (2014). doi: 10.1155/2014/974393; pmid: 24741567
46. V. Alfred, T. Vaccari, When membranes need an ESCRT: Endosomal sorting and membrane remodelling in health and disease. *Swiss Med. Wkly.* **146**, w14347 (2016). doi: 10.4414/smw.2016.14347; pmid: 27631343
47. H. Goto, M. Inagaki, New insights into roles of intermediate filament phosphorylation and progeria pathogenesis. *IUBMB Life* **66**, 195–200 (2014). doi: 10.1002/iub.1260; pmid: 24659572
48. A. Panopoulos *et al*., Failure of cell cleavage induces senescence in tetraploid primary cells. *Mol. Biol. Cell* **25**, 3105–3118 (2014). doi: 10.1091/mbc.e14-03-0844; pmid: 25143403
49. N. Bojjireddy *et al*., Pharmacological and genetic targeting of the PI4KA enzyme reveals its important role in maintaining plasma membrane phosphatidylinositol 4-phosphate and phosphatidylinositol 4,5-bisphosphate levels. *J. Biol. Chem.* **289**, 6120–6132 (2014). doi: 10.1074/jbc.M113.531426; pmid: 24415756
50. S. B. Thoresen *et al*., ANCHR mediates aurora-B-dependent abscission checkpoint control through retention of VPS4. *Nat. Cell Biol.* **16**, 550–560 (2014). doi: 10.1038/ncb2959; pmid: 24814515
51. L. Capalbo *et al*., The midbody interactome reveals unexpected roles for PP1 phosphatases in cytokinesis. *Nat. Commun.* **10**, 4513 (2019). doi: 10.1038/s41467-019-12507-9; pmid: 31586073
52. A. R. Skop, H. Liu, J. Yates 3rd, B. J. Meyer, R. Heald, Dissection of the mammalian midbody proteome reveals conserved cytokinesis mechanisms. *Science* **305**, 61–66 (2004). doi: 10.1126/science.1097931; pmid: 15166316
53. R. N. Kettleborough *et al*., A systematic genome-wide analysis of zebrafish protein-coding gene function. *Nature* **496**, 494–497 (2013). doi: 10.1038/nature11992; pmid: 23594742
54. A. Roy, A. Kucukural, Y. Zhang, I-TASSER: A unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738 (2010). doi: 10.1038/nprot.2010.5; pmid: 20360767
55. L. M. Rice, E. A. Montabana, D. A. Agard, The lattice as allosteric effector: Structural studies of alphabeta- and γ-tubulin clarify the role of GTP in microtubule assembly. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 5378–5383 (2008). doi: 10.1073/pnas.0801155105; pmid: 18388201

ACKNOWLEDGMENTS

We acknowledge M. Gai and the Open Lab of Advance Microscopy (OLMA@MBC) for technical assistance. **Funding:** This work was supported by Associazione Italiana Ricerca Cancro (AIRC; 21875 to E.H.), Leducq Foundation (19CVD02 to E.H.), PRIN (20177E9EPY to E.H.), Fondazione Ricerca Molinette (to E.H.), Fondazione Pezcoller-SIC "Patrizia Coser" (to F.G.), Fondazione Italiana per la Ricerca sul Cancro (FIRC 22558 to J.P.M. and FIRC 22248 to M.C.D.S.), Deutsche Forschungsgemeinschaft (DFG) (TRR186/A08 to V.H. and SCHU 3314/1-1 to M.S.), NeuroCure Cluster of Excellence (Charité Universitätsmedizin Berlin, 10117 Berlin, Germany to V.H.), National Institute of Diabetes and Digestive and Kidney Diseases (DK119305 to D.A.B.), and Sigma Xi research foundation (to A.C.). **Author contributions:** F.G. designed and performed research, analyzed data, and wrote the manuscript; L.P., H.L., I.C., S.J.C., A.M., P.K., A.C., J.F., A.Y., W.-T.L., N.T.S., M.C.D.S, and J.P.M. performed research and analyzed data; O.V., D.R.P., H.N.B., D.T., R.L.W., M.T., T.B., B.M.D., M.S., M.S.W., A.G., M.M., B.D.P., V.H, E.B, G.R.M., and D.A.B. provided critical reagents and supervised the work; E.H. designed the research, supervised the work and wrote the manuscript. **Competing interests:** E.H. and A.G. are cofounder and board member of Kither Biotech, a pharmaceutical product company developing PI3K inhibitors for the treatment of respiratory diseases not in conflict with statements made in this article. The other authors declare no conflict of interest. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. PITCOIN1 is subject to a European patent application (application no. PCT/EP2019/065009), and its distribution is dependent on completion of a materials transfer agreement with the FMP.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk0410
Figs. S1 to S20
Tables S1 to S3
MDAR Reproducibility Checklist
Movies S1 to S7

17 June 2021; accepted 26 October 2021
10.1126/science.abk0410

RESEARCH ARTICLE SUMMARY

STRUCTURE PREDICTION

Computed structures of core eukaryotic protein complexes

Ian R. Humphreys†, Jimin Pei†, Minkyung Baek†, Aditya Krishnakumar†, Ivan Anishchenko, Sergey Ovchinnikov, Jing Zhang, Travis J. Ness, Sudeep Banjade, Saket R. Bagde, Viktoriya G. Stancheva, Xiao-Han Li, Kaixian Liu, Zhi Zheng, Daniel J. Barrero, Upasana Roy, Jochen Kuper, Israel S. Fernández, Barnabas Szakal, Dana Branzei, Josep Rizo, Caroline Kisker, Eric C. Greene, Sue Biggins, Scott Keeney, Elizabeth A. Miller, J. Christopher Fromme, Tamara L. Hendrickson, Qian Cong*‡, David Baker*‡

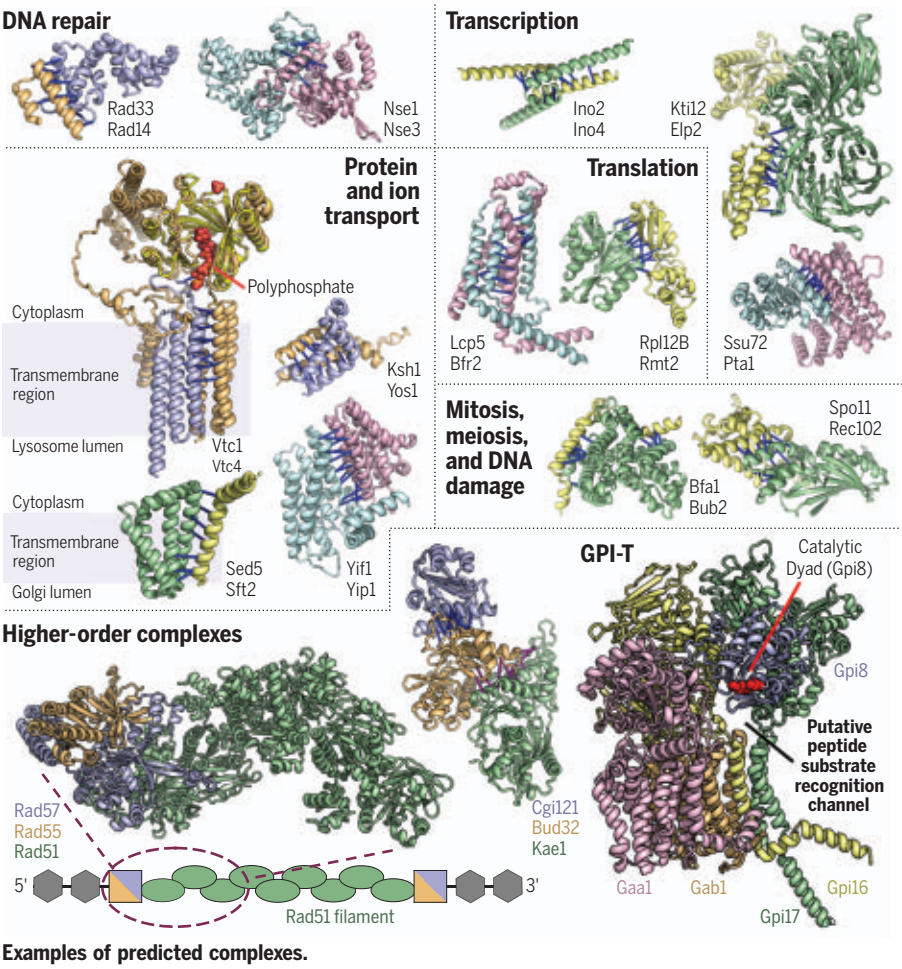
INTRODUCTION: Protein-protein interactions play critical roles in biology, but the structures of many eukaryotic protein complexes are unknown, and there are likely many interactions not yet identified. High-throughput experimental methods such as yeast two-hybrid and affinity-purification mass spectrometry have been used to identify interactions in multiple organisms, but there are inconsistencies between different datasets, and the methods do not provide high-resolution structural information. Here, we use deep learning methods to systematically identify and build structures for the protein complexes that mediate key processes in eukaryotes.

RATIONALE: Interacting proteins often coevolve, and in prokaryotes, evolutionary information can be used to identify interactions on the proteome scale at an accuracy higher than that of experimental screens. Extending this method to eukaryotes is complicated because there are fewer genome sequences available, resulting in weaker coevolutionary signals. The deep learning methods RoseTTAFold and AlphaFold, have a rich understanding of protein sequence-structure relationships, and so could help overcome this limitation.

RESULTS: We developed a coevolution-guided protein interaction identification pipeline that incorporates a rapidly computable version of RoseTTAFold with the slower but more accurate AlphaFold to systematically evaluate interactions between 8.3 million pairs of yeast proteins. RoseTTAFold alone has comparable performance in identifying protein-protein interactions to that of large-scale experimental methods; combination with AlphaFold increases identification accuracy. In total, we constructed models for 106 previously unidentified assemblies and 806 that were structurally uncharacterized.

These complexes provide rich insights into a range of biological processes from transcription, translation, and DNA repair to protein transport and modification. For example, Rad51 plays a pivotal role in DNA repair through homologous recombination, and mutations are associated with Fanconi anemia and cancer in humans. Rad55 and Rad57 are positive regulators of Rad51 assembly on single-stranded DNA. Our Rad55–Rad57–Rad51 complex model suggests that Rad55–Rad57 can bind at the 5' end of the Rad51 single-stranded DNA filament and may stabilize the filament conformation of Rad51. Glycosylphosphatidylinositol transamidase (GPI-T) is a pentameric enzyme complex that catalyzes the attachment of GPI anchors to the C terminus of proteins. GPI-T is structurally uncharacterized, and mutations in subunits of the complex have been implicated in neurodevelopmental disorders and cancer in humans. Our model of the five-protein assembly shows that the previously identified catalytic dyad is positioned adjacent to a channel formed by three other subunits that could function in C-terminal GPI-T signal peptide recognition.

CONCLUSION: Our approach extends the range of large-scale deep learning–based structure modeling from monomeric proteins to protein assemblies. Following up on the many new interactions and complex structures should advance the understanding of a wide range of eukaryotic cellular processes and provide new targets for therapeutic intervention. Our results herald a new era of structural biology in which computation plays a fundamental role in both interaction discovery and structure determination. ■



Examples of predicted complexes.

The list of authors and their affiliations is available in the full article online.
*Corresponding author. Email: qian.cong@utsouthwestern.edu (Q.C.); dabaker@uw.edu (D.B.)
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
Cite this article as I. R. Humphreys *et al.*, *Science* **374**, eabm4805 (2021). DOI: 10.1126/science.abm4805

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm4805

RESEARCH ARTICLE

STRUCTURE PREDICTION

Computed structures of core eukaryotic protein complexes

Ian R. Humphreys[1,2]†, Jimin Pei[3,4]†, Minkyung Baek[1,2]†, Aditya Krishnakumar[1,2]†, Ivan Anishchenko[1,2], Sergey Ovchinnikov[5,6], Jing Zhang[3,4], Travis J. Ness[7]‡, Sudeep Banjade[8], Saket R. Bagde[8], Viktoriya G. Stancheva[9], Xiao-Han Li[9], Kaixian Liu[10], Zhi Zheng[10,11], Daniel J. Barrero[12], Upasana Roy[13], Jochen Kuper[14], Israel S. Fernández[15], Barnabas Szakal[16], Dana Branzei[16,17], Josep Rizo[4,18,19], Caroline Kisker[14], Eric C. Greene[13], Sue Biggins[12], Scott Keeney[10,11,20], Elizabeth A. Miller[9], J. Christopher Fromme[8], Tamara L. Hendrickson[7], Qian Cong[3,4]*§, David Baker[1,2,21]*§

Protein-protein interactions play critical roles in biology, but the structures of many eukaryotic protein complexes are unknown, and there are likely many interactions not yet identified. We take advantage of advances in proteome-wide amino acid coevolution analysis and deep-learning–based structure modeling to systematically identify and build accurate models of core eukaryotic protein complexes within the *Saccharomyces cerevisiae* proteome. We use a combination of RoseTTAFold and AlphaFold to screen through paired multiple sequence alignments for 8.3 million pairs of yeast proteins, identify 1505 likely to interact, and build structure models for 106 previously unidentified assemblies and 806 that have not been structurally characterized. These complexes, which have as many as five subunits, play roles in almost all key processes in eukaryotic cells and provide broad insights into biological function.

Yeast two-hybrid (Y2H), affinity-purification mass spectrometry (APMS), and other high-throughput experimental approaches have identified many pairs of interacting proteins in yeast and other organisms (*1*–*5*), but there are discrepancies between sets generated using the different methods and considerable false-positive and false-negative rates (*6*–*8*). Because residues at protein-protein interfaces are expected to coevolve, the likelihood that any two proteins interact can be assessed by identifying and aligning the ortholog sequences of the two proteins in many different species, joining them to create paired multiple sequence alignments (pMSAs), and then determining the extent to which changes in the sequences of orthologs for the first protein covary with ortholog sequence changes for the second (*9*, *10*). Such amino acid coevolution has been used to guide modeling of complexes for cases in which the structures of the partners are known (*11*, *12*) and to systematically identify pairs of interacting proteins in prokaryotes with an accuracy higher than that of experimental screens (*9*). Recent deep-learning–based advances in protein structure prediction (*13*, *14*) have the potential to increase the power of such approaches as they now enable accurate modeling not only of protein monomer structures but also protein complexes (*13*).

We set out to combine proteome wide coevolution-guided protein interaction identification with deep-learning–based protein structure modeling to systematically identify and determine the structures of eukaryotic protein assemblies (Fig. 1A). We faced several challenges in directly applying to eukaryotes the statistical methods we had found effective in identifying coevolving pairs in prokaryotes (*8*). First, far fewer genome sequences are available for eukaryotes than prokaryotes: The average number of orthologous sequences (excluding nearly identical copies with >95% sequence identity) is on the order of 10,000 for bacterial proteins but 1000 for eukaryotic proteins. Thus, multiple sequence alignments for pairs of eukaryotic proteins contain fewer diverse sequences, making it more difficult for statistical methods to distinguish true coevolutionary signal from the noise. Second, eukaryotes in general have a larger number of genes, making comprehensive pairwise analysis more computationally intensive and increasing the background noise. Third, mRNA splicing in eukaryotes further increases the number of protein species, resulting in errors in gene predictions and complicating sequence alignments. Fourth, eukaryotes underwent several rounds of genome duplications in multiple lineages (*15*), and it can be difficult to distinguish orthologs from paralogs, which is important for detecting coevolutionary signal because the protein interactions of interest are likely to be conserved in orthologs in other species but less so in paralogs.

To mitigate the first three challenges, we chose to predict protein complexes for the yeast *Saccharomyces cerevisiae* as the starting point because there are a large number of fungal genomes (*16*), the genome is relatively small (6000 genes in total), and there is relatively little mRNA splicing (*17*). Furthermore, because the interactome of yeast has been extensively studied, there is a "gold standard" set (see materials and methods) of known interactions to evaluate the accuracy of predicted interactions and structures.

To distinguish orthologs from paralogs, we started from OrthoDB (*18*), a hierarchical catalog of orthologs across 1271 eukaryote genomes, and supplemented each orthologous group with sequences from 4325 eukaryote proteomes that we assembled from the National Center for Biotechnology Information (https://www.ncbi.nlm.nih.gov/genome) and the Joint Genome Institute (*19*). Among these, 2026 are fungal proteomes spanning 14 phyla (47 classes). We compared the sequences for each protein in each of the additional 4325 proteomes against those of the most closely related species in the OrthoDB database and used the reciprocal best hit criterion (*20*) to identify orthologs (fig. S1); these were then added to the corresponding orthologous group. A complication is that each species frequently contains multiple proteins belonging to the same orthologous group, leading to ambiguity in determining which protein should be included. These multiple copies may represent alternatively spliced forms of the same gene, parts of the same gene that were split

[1]Department of Biochemistry, University of Washington, Seattle, WA, USA. [2]Institute for Protein Design, University of Washington, Seattle, WA, USA. [3]Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX, USA. [4]Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX, USA. [5]Faculty of Arts and Sciences, Division of Science, Harvard University, Cambridge, MA, USA. [6]John Harvard Distinguished Science Fellowship Program, Harvard University, Cambridge, MA, USA. [7]Department of Chemistry, Wayne State University, Detroit, MI, USA. [8]Department of Molecular Biology and Genetics, Weill Institute for Cell and Molecular Biology, Cornell University, Ithaca, NY, USA. [9]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. [10]Molecular Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [11]Gerstner Sloan Kettering Graduate School of Biomedical Sciences, New York, NY, USA. [12]Howard Hughes Medical Institute, Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [13]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA. [14]Rudolf Virchow Center for Integrative and Translational Bioimaging, University of Würzburg, Würzburg, Germany. [15]Department of Structural Biology, St Jude Children's Research Hospital, Memphis, TN, USA. [16]IFOM, the FIRC Institute of Molecular Oncology, Via Adamello 16, 20139, Milan, Italy. [17]Istituto di Genetica Molecolare, Consiglio Nazionale delle Ricerche (IGM-CNR), Via Abbiategrasso 207, 27100, Pavia, Italy. [18]Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX, USA. [19]Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, TX, USA. [20]Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [21]Howard Hughes Medical Institute, University of Washington, Seattle, WA, USA.

***Corresponding author. Email: qian.cong@utsouthwestern.edu (Q.C.); dabaker@uw.edu (D.B.)**

†These authors contributed equally to this work. ‡Present address: Sanofi, Cambridge, MA, USA. §These authors contributed equally to this work.

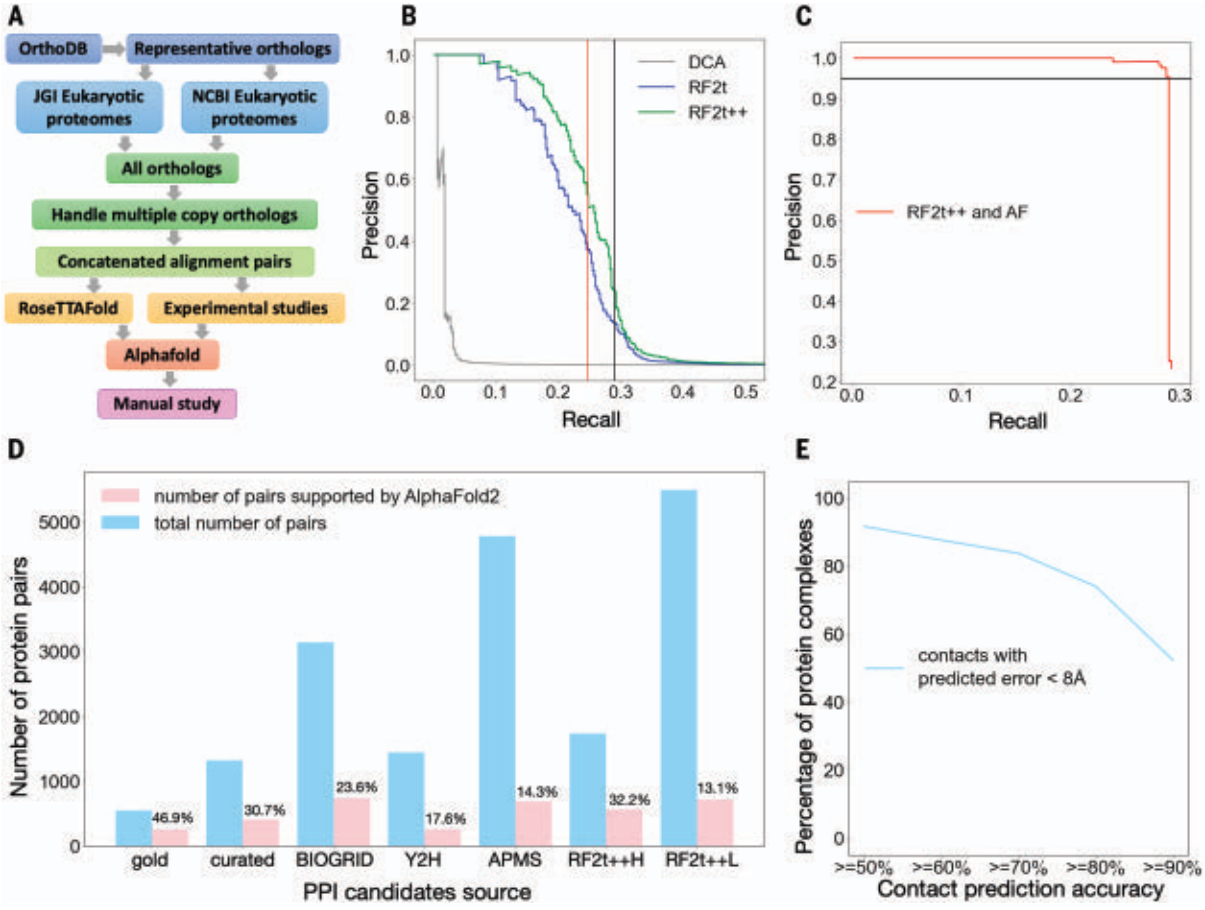


Fig. 1. Evaluation of protein interaction and structure prediction accuracy. (**A**) The PPI screen pipeline. (**B**) Performance (precision at different levels of recall) of different methods in picking out gold standard PPIs from the set of 4.3 million pMSAs [precision: number of true positives above a cutoff divided by the total number of pairs above this cutoff; recall: number of true positives above cutoff divided by the total number of true positives (gold standard PPIs)]. Pairs were ranked by the top coevolution score or contact probability between residue pairs. DCA: direct coupling analysis. RF2t: top contact probability between residues of two proteins by RF two-track model. RF2t++, optimized RF2t (see materials and methods). RF2t++ predictions better than the cutoff shown in vertical black line (RF2t++L in Fig. 1C) were processed with AF; recall of gold standard PPIs at this cutoff is 29%, and precision is 23%. RF2t++ results with a more stringent cutoff (red vertical line) are also shown in Fig. 1C (RF2t++H). (**C**) AF contact probability ranking of complexes selected by RF2t++ in (B); complexes with scores above the horizontal black line were selected for further analysis. (**D**) Number of high-scoring (top contact probability >0.67) AF predictions in PPI sets from different sources. (**E**) Distribution of percent of AF predicted interprotein contacts with predicted error <8 Å found in contact (<8 Å) in closely related experimental structures.

into multiple pieces because of errors in gene prediction, or recent gene expansions specific to certain lineages. We dealt with these possibilities by keeping only the longest isoform of each gene, merging pieces of the same gene, and selecting the copy with the highest sequence identity to single-copy orthologs in other species. For 4090 out of ~6000 yeast proteins, we were able to assign a single-copy yeast protein to orthologs in other species, and we generated pMSAs for all 4090 × 4089/2 = 8,362,005 pairwise combinations of these proteins (fig. S2). We focused on 4,286,433 pairs with alignments containing over 200 sequences to increase prediction accuracy and less than 1300 amino acids to accelerate computation (fig. S3).

In a first set of calculations, we found that even with the advantages of *S. cerevisiae* and improved ortholog identification, the statistical method (direct coupling analysis, DCA) we had used in our previous coevolution-guided protein-protein interaction (PPI) screen in prokaryotes (*9*) [the more accurate GREMLIN (*11*) method is too slow for this] could not effectively distinguish a gold standard set of 768 yeast protein pairs known to interact (*5*) (http://interactome.dfci.harvard.edu/S_cerevisiae/) from the much larger set (768,000 pairs) of primarily noninteracting pairs (Fig. 1B, gray curve, area under the curve: 0.016). Progress required a more accurate and sensitive, but still rapidly computable, method to evaluate protein interactions based on pMSAs.

We explored the application of the deep-learning–based structure prediction methods, RoseTTAFold (RF) and AlphaFold (AF), to this problem. Even though RF was originally trained on monomeric protein sequences and structures, it can accurately predict the structures of protein complexes given pMSAs with a sufficient number of sequences (*13*). We found that a lighter-weight (10.7 million parameters) RF two-track model (figs. S4 and S5) provided a good trade-off between compute time and accuracy: The model requires 11 s (about 100 times faster than AF) to process a pMSA of 1000 amino acids on a NVIDIA TITAN RTX graphic processing unit, and it can effectively distinguish gold standard PPIs among much larger sets of randomly paired proteins. The very short time required to analyze an individual pMSA made it possible to process all 4.3 million pMSAs. This method considerably outperformed DCA in distinguishing gold standard interactions from random pairs (Fig. 1B, blue curve, area under the curve:

0.219), using the highest predicted contact probability over all pairs of residues in the two proteins as a measure of the propensity for two proteins to interact (fig. S6). Performance was further improved (Fig. 1B, green curve, area under the curve: 0.248) by correcting overestimations of predicted contact probabilities between the C-terminal residues of the first protein and the N-terminal residues of the second protein, and of predicted interactions for a subset of proteins showing hub-like interactions with many other proteins (see materials and methods and figs. S7 and S8). The much better performance of RF than DCA likely stems from the extensive information on protein sequence-structure relationships embedded in the RF deep neural network; DCA, by contrast, operates solely on protein sequences with no underlying protein structure model.

We next explored whether AF residue-residue contact predictions could further distinguish interacting from noninteracting protein pairs. Like RF, AF was trained on monomeric protein structures, but given the good results with two-track RF on protein complexes and the higher accuracy of AF [also a two-track network followed by a three-dimensional (3D) structure module] on monomers, we reasoned that it might similarly have higher accuracy than RF on complexes; to enable modeling of protein complexes using AF, we modified the positional encoding in the AF script (see materials and methods). AF was too slow to be applied to the entire set of 4.3 million pMSAs [this would require 0.1 to 1 million graphics processing unit (GPU) hours]; instead we applied AF to the 5495 protein pairs with the highest RF support (indicated by the black vertical line in Fig. 1B). Using the highest AF contact probability over all residue pairs as a measure of interaction strength, we found that the combination of RF followed by AF provided excellent performance (Fig. 1C and figs. S9 and S11). Almost all the gold standard pairs were ranked higher than the negative controls, allowing selection of a set of 715 candidate PPIs with an expected precision of 95% at an AF contact probability cutoff of 0.67 (black horizontal line in Fig. 1C); we refer to this RF plus AF procedure as the de novo PPI screen, and the resulting set of predicted interactions, the de novo PPI set, below.

Owing to the trade-off between compute time and accuracy, and the necessity of setting a stringent threshold to avoid large numbers of false positives given the very large number of total pairs, we were concerned that some interacting proteins might not coevolve sufficiently to be identified robustly in our all-versus-all RF screen. Given the excellent performance of AF in distinguishing gold standard interactions among the RF filtered pairs, we also applied AF to pMSAs for PPIs reported in the literature, including those identified in high-throughput experimental screens. Similarly to our de novo PPI screen procedure, we considered protein pairs with AF contact probability larger than 0.67 to be confident interacting partners. We found that 47% of the gold standard PPIs were confidently predicted, with lower ratios (31 and 24%) for candidate PPIs from the literature (http://interactome.dfci.harvard.edu/S_cerevisiae/download/LC_multiple.txt) (*3*) or supported by low-throughput experiments according to BIOGRID (*21*) (Fig. 1D). The ratio of confidently predicted PPIs is even lower for protein pairs identified by Y2H (18%) or APMS (14%) screens (table S1), consistent with the known larger fraction of false positives in large-scale experimental screens (*8*, *22*). The fast RF two-track model used in the de novo screen has an accuracy comparable to or better than that of the large-scale experimental screens when assessed in this way: With a high-stringency RF cutoff (indicated by the red vertical line in Fig. 1B), the fraction of confidently predicted pairs among PPIs identified by RF is 32%, similar to the accuracy of low-throughput experiments; with a lower stringency cutoff (indicated by the black vertical line in Fig. 1B), this fraction becomes closer to that of the large-scale experimental screens, but somewhat fewer true PPIs are missed than with the higher cutoff (Fig. 1D).

In total, we identified 715 likely interacting pairs from the "de novo RF → AF" screen, and 1251 from the "pooled experimental sets → AF" screen, of which 461 overlap, resulting in a total of 1505 PPIs (see figs. S11 to S13 for interface size and secondary-structure distributions for the predicted complex structures). Out of these, 699 have been structurally characterized, 700 have some supporting experimental data from literature and databases, and 106 have not, to our knowledge, been previously described. To evaluate the accuracy of the predicted 3D structure of protein complexes, we used as a benchmark the 699 pairs with experimental structures in the Protein Data Bank (PDB). For 92% of these pairs, at least 50% of confident (predicted aligned error <8 Å) AF-predicted contacts are present in the experimental structures (Fig. 1E and fig. S14). The models do miss many contacts observed in the experimental structures, however, likely owing to lower residue-residue coevolution (fig. S15).

With these benchmark results providing confidence in the accuracy of the new complex interaction predictions and 3D models of the predicted complexes, we analyzed the structure models for the 806 complexes for which high-resolution structural information was not available. We classified these models into groups on the basis of their biological functions and provide examples of complexes in each functional class in Figs. 2 to 4. A first set of complexes are involved in maintenance and processing of genetic information: DNA repair, mitosis and meiosis checkpoints, transcription, and translation (Fig. 2). A second set of complexes play roles in protein translocation, transport through the secretory pathway, the cytoskeleton, and cell organelles (Fig. 3). A third set of complexes are involved in metabolism (Fig. 4). Examples of protein complexes in which proteins of unknown function are predicted to interact with well-characterized ones are shown in Fig. 4: These interactions provide hints about the function of the uncharacterized proteins and could help identify new components of previously characterized assemblies. In cases where three or more proteins were predicted to mutually interact, we generated models of the full assemblies by using as input a sequence alignment for the entire complex (see materials and methods). Examples of these larger assemblies are shown in Fig. 5; in most cases, the pairwise interactions are quite similar to those for the independently built binary complexes, but simultaneous modeling of the full complex has the advantage of allowing conformational changes that could accompany full assembly.

It is not possible to analyze the functional implications of all of the new complexes in a single paper. Instead, as an illustration of the insights that can be gained from these, we focus on a few selected examples in the following sections. To enable broader study of the functional implications of the full set of models, we have made them available at https://modelarchive.org/doi/10.5452/ma-bak-cepc and additional information is provided in the supplementary Excel file.

Complexes involved in DNA homologous recombination and repair

The homologous recombination required for accurate chromosome segregation during meiosis is initiated by DNA double-strand breaks made by Spo11 (*23*). Spo11 is essential for sexual reproduction in most eukaryotes (*24*, *25*), but mechanistic insight has been limited by a deficit of high-resolution structural information. We predict the structures of complexes of Spo11 with its essential partners Ski8 and Rec102 (Fig. 2 and figs. S16 and S17). The predicted Spo11–Ski8 structure is supported by cross-linking and mutagenesis data (*26*, *27*). Our model resembles a previous model based on the Ski3–Ski8 complex, with Ski8 contacting a sequence in Ski3 that is similar to the sequence $QREIF_{380}$ in Spo11 (*27*, *28*) (fig. S17A), but suggests a more extensive interaction surface than previously appreciated (*29*, *30*) (fig. S17, B and C). Rec102 was proposed to be a remote homolog of the transducer domain of the Top6B subunit of

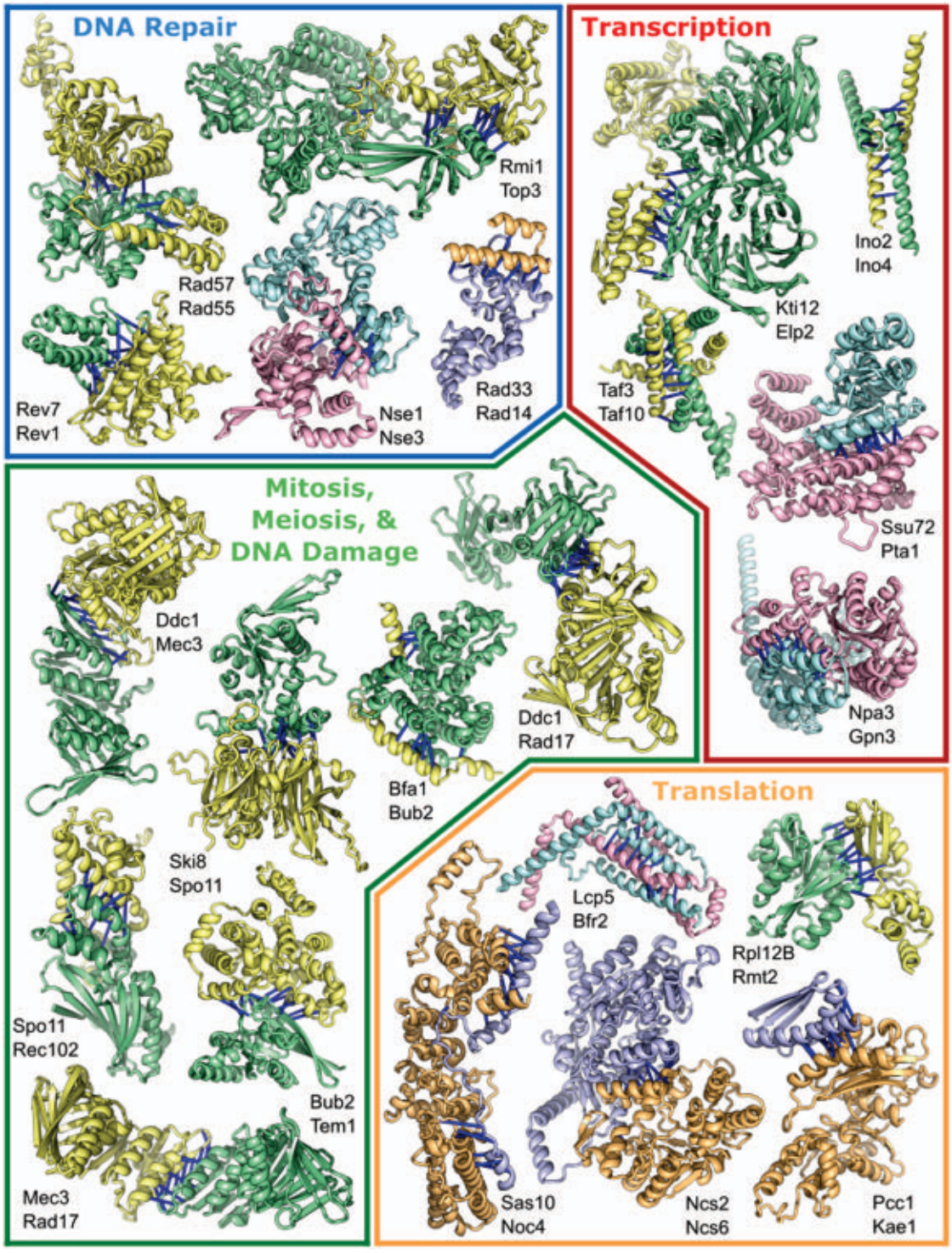


Fig. 2. Protein complexes involved in transcription, translation, and DNA repair. Top predicted residue-residue contacts are indicated with bars. Pair color indicates the method of identification: pairs from the "pooled experimental sets → AF" screen are in yellow and green, pairs from the "de novo RF → AF" screen are in blue and light orange; and pairs present in both datasets are teal and pink. Full names of these proteins are in table S2.

archaeal topoisomerase VI (*31*), which couples adenosine 5′-triphosphate–dependent dimerization of Top6B subunits to DNA cleavage by Top6A subunits (*32*). Our predicted Rec102–Spo11 complex resembles the Top6A–Top6B interface: a four-helix bundle consisting of two C-terminal helices from Rec102 and two helices from Spo11 (the first helix of the winged helix domain (WHD) plus a more N-terminally located helix) (fig. S17D). Alanine substitutions in this portion of Rec102 disrupt interaction with Spo11 and block meiotic recombination in vivo (*27*). The model clarifies the Spo11 portion of this interface, which was not well structured in previous

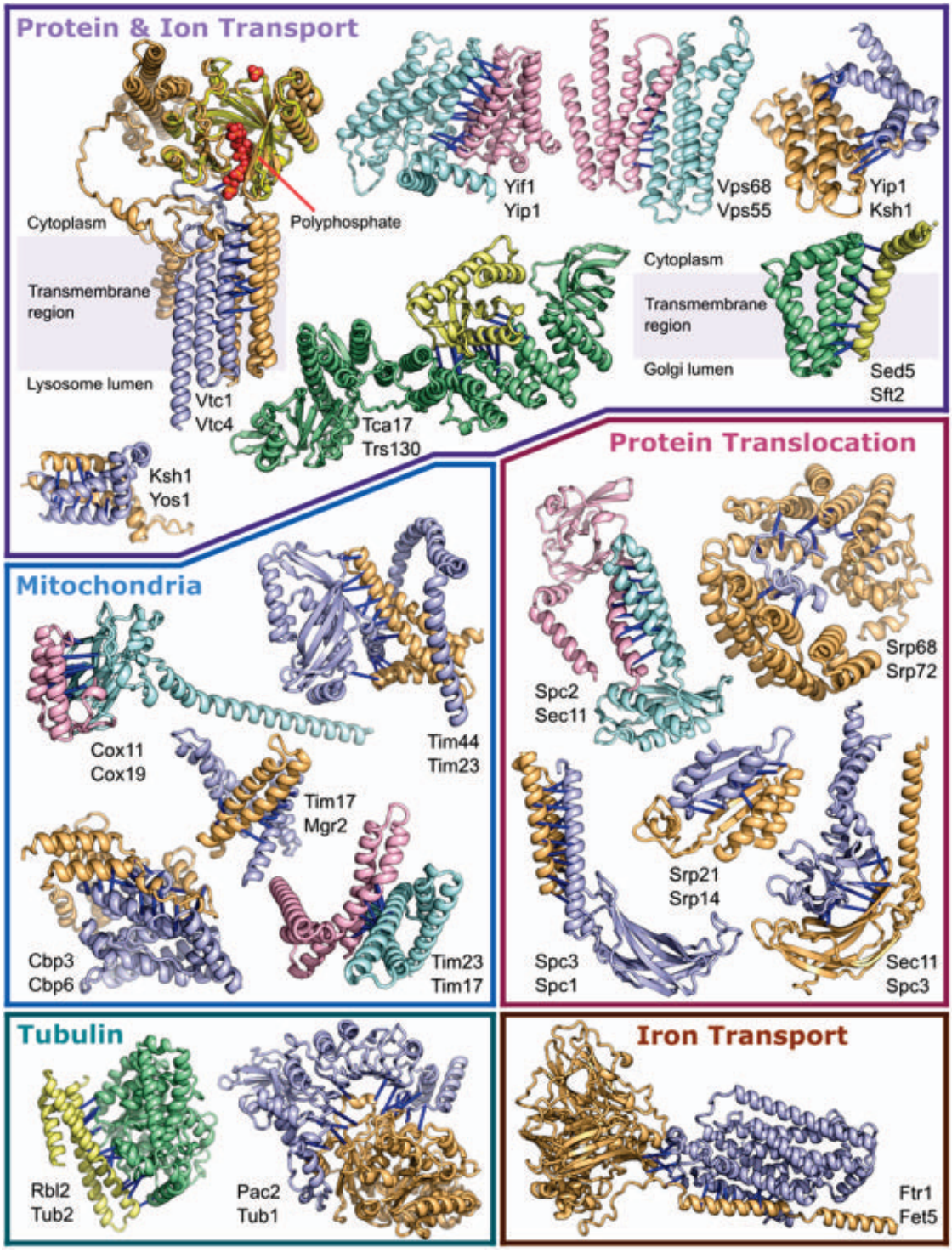


Fig. 3. Protein complexes involved in molecule transport, membrane translocation, and mitochondria. Bars and coloring as in Fig. 2. Full names for proteins are in table S3. Membrane-spanning regions are annotated on Vtc1-Vtc4 and Sed5-Sft2. Top left: model of Vtc1-Vtc4 complex, with superimposed crystal structure (PDB: 3G3Q, chain A) of the VTC4 (bright yellow) with phosphate bound (red balls).

homology models (*27*, *31*). Both Rec102 and Top6B have long, helical arms that feed into the Spo11 interface; our model predicts a different angle for this arm and contains a kink that corresponds to a conserved sequence motif $EYPMVF_{192}$ in *Saccharomyces* that is missing in both archaeal TopoVI and mammals (fig. S17, D and E). Mutations in this region can suppress *rec104* conditional alleles (*33*), suggesting that this part of Rec102 is important for integrating Rec104 function into the Spo11 core complex.

The highly conserved Rad51 protein, which is central to DNA repair, carries out key reactions during homologous recombination, and mutations in human paralogs are associated with Fanconi anemia and multiple types of cancer (*34*). Rad51 paralogs can be positive regulators

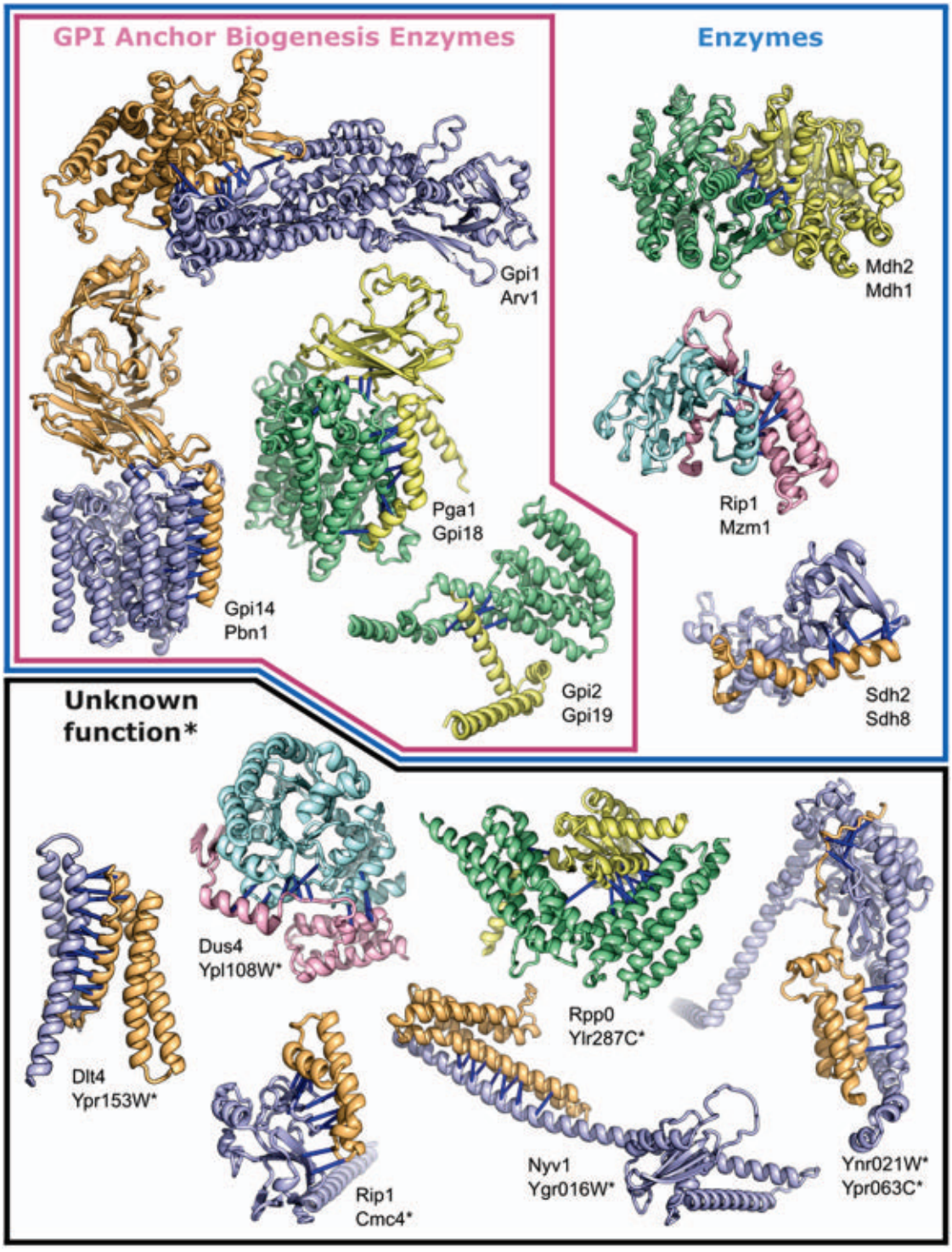


Fig. 4. Protein complexes involved in metabolism, GPI anchor biosynthesis, or including a protein of unknown function. Coloring is as in Figs. 2 and 3. Proteins annotated in the Uniprot database as uncharacterized proteins are denoted with an asterisk. Full names for these proteins are in table S4.

of Rad51 activity (*35*); in yeast, the Rad51 paralogs Rad55 and Rad57 form a stable homodimer that accelerates assembly of Rad51 filaments on single-stranded DNA (ssDNA) during homologous recombination through a transient interaction with Rad51 (*36*). The lack of structural data for the Rad55–Rad57 complex and its interface with Rad51 has limited mechanistic understanding of this process. We generated a model of the trimeric Rad55–Rad57–Rad51 complex, which in combination with the known Rad51 filament structure (*37*), suggests that Rad55–Rad57 binds at the 5′ end of the Rad51 filament where it could promote growth of the Rad51 filament in a directional manner (Fig. 5B and fig. S18).

Nucleotide excision repair (NER) requires a search for lesions in DNA that is mediated

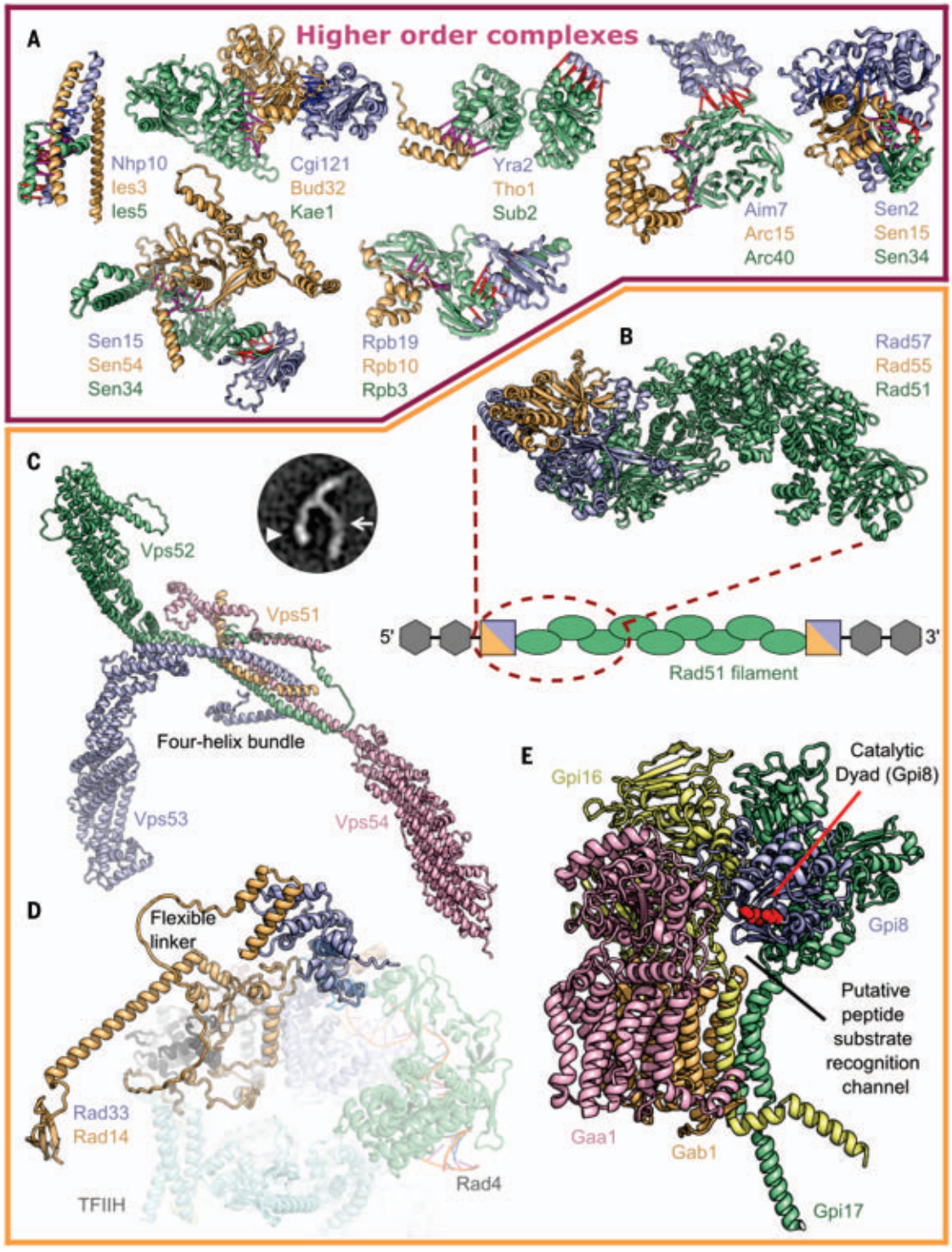


Fig. 5. Higher-order protein complexes. (**A**) Top predicted residue-residue contacts for trimers are indicated with bars. Bar color corresponds to the interacting protein pair; protein 1:2 are blue, 1:3 are red, 2:3 are purple. Full names of each protein within the complex are in table S5. (**B**) Model of Rad55–Rad57–Rad51 and cartoon depiction of placement of this complex in the larger Rad51 filament. Additional information is in fig. S18. (**C**) GARP complex model constructed by predicting structure of central hetero-oligomeric helical bundle, and superimposing models of individual components onto this. 2D class average of GARP complex with minor adaptation (*77*); reprinted by permission from Springer Nature Customer Service Center GmbH. Alternative GARP models are in fig. S24. (**D**) Rad33–Rad14 complex model superimposed onto previously determined TFIIH/Rad4–Rad23–Rad33 complex structure (7k04). See fig. S19 for additional details. (**E**) GPI-T pentamer model highlighting a possible peptide substrate recognition channel adjacent to the catalytic dyad. See fig. S27 for additional details.

by a conserved complex containing Rad4 (XPC—Xeriderma pigmentosum group C protein in humans), Rad23 (HR23B), and Rad33 (Centrin2) in yeast. The Rad4–Rad23–Rad33 complex is essential for global genome NER and is the major player in initial damage recognition (*38*). Rad14 (XPA) is recruited at a later stage and activates the helicase Rad3 (XPD) subunit of the general transcription and DNA repair factor IIH complex (TFIIH, consisting of Rad3, Ssl2, Ssl1, Tfb1, Tfb2, Tfb4, and Tfb5) through the release of the TFIIK (CAK) complex following interactions with the TFIIH subunits Tfb5 (p8) and Ssl2 (XPB), and double-stranded DNA (*39*). The structures of Rad14 that are currently available only comprise the extended DNA binding domain and lack the N and C terminus, where the latter interacts with Tfb5. We generated a model of the complex between full-length Rad14 and Rad33 that resolves much of the current structural ambiguity in this system (Fig. 2 and fig S19B), shedding light on how Rad14 may be recruited to the Rad4–Rad23–Rad33 complex. Placing this model into a cryo–electron microscopy (EM) map comprising XPA (Rad14) and TFIIH bound to DNA

(39) suggests how the Rad14 C terminus, which fits into previously unmodeled density, interacts with TFIIH. The long central helix observed in the Centrin2 (Rad33) structure (40) is kinked about 90° in our Rad33–Rad14 complex model (fig. S19B); both conformations are feasible and are compatible for the interaction with Rad14. In a recent cryo-EM structure of the TFIIH/Rad4–Rad23–Rad33 initial recognition complex (41), only the C- terminal part of Rad33 was determined. Superposition of Rad33 in the Rad33–Rad14 complex model onto this structure (Fig. 5D) shows how Rad14 can interact with the Rad4–Rad23–Rad33 recognition complex (38, 42) while maintaining the TFIIH interaction, bridging the steps of initial damage recognition and damage verification. Our model suggests that Rad14 and Rad4 can be present at the same time in the repair cascade; cross-talk between these important proteins could modulate downstream events.

Complexes involved in translation and ribosome regulation

Throughout evolution, the eukaryotic machinery for protein production has expanded in size and complexity (43), which facilitated the development of sophisticated mechanisms for the regulation of gene expression at the post-transcriptional level (44) and increased integration with the cellular environment (45). The expanded complexity of the eukaryotic translational machinery came at the cost of a highly complex process for ribosome maturation (46). We generate models of complexes that had not been structurally characterized previously that involve components of the translation apparatus (Fig. 2 and fig. S20). Two complexes, Rpl12B–Rmt2 and Rpl7A–Fpr4, involving enzymes that introduce protein modifications such as arginine methylations or proline isomerization (47), provide insight into mechanisms that expand the chemical diversity of ribosomal proteins at functional sites (48) and possibly regulate translation (49). A complex between components of the U3 ribosome-maturation factor and a protein involved in the regulation of glycerol, Lcp5–Sgd1 (50), could play a role in coupling translation with metabolism. A complex between eukaryotic initiation factor 2B (eIF2B), an auxiliary factor for eIF2 recycling after guanosine 5′-triphosphate hydrolysis, and transcriptional factor regulator Dig2 could help couple translation and transcription: The delivery of the first aminoacyl-tRNA (Met-tRNA$_i^{Met}$) is a key event in eukaryotic translation regulation by the GTPase eIF2 (51), and targeting eIF2 through its nucleotide exchanger eIF2B is a basal mechanism of translation regulation. This possible cross-talk between ribosome-maturation pathways and metabolic sensors, and translation initiation regulators such as eIF2, with transcription factors suggests exciting new avenues to further map the highly integrated nature of translation within eukaryotic cells.

Complexes involving ubiquitin and small ubiquitin-like modifier (SUMO) ligases

Reversible covalent modifications of proteins with ubiquitin and SUMO modulate protein-protein interactions, cellular localization, and stability (52). SUMO E3 ligases facilitate SUMO transfer, and Siz1, Siz2, Mms21, and Zip3 are the known SUMO ligases in budding yeast (52). Our model of the Siz2 and Mms21 SUMO ligase complex (fig. S21A) suggests that both E3s could act jointly to modify DNA-associated substrates, perhaps through the DNA binding SAP domain of Siz2 (53) or involving the Mms21 (Nse2)–containing Smc5–6 complex, which modulates DNA recombination, replication, and repair (54, 55). The Smc5–6 complex contains another RING-finger E3 ligase–like subunit, Nse1 (56), that interacts with Nse3 and Nse4. Our model of the yeast Nse1–Nse3–Nse4 complex (fig. S21B) is similar to a structure determined for the *Xenopus laevis* complex, despite the sequences of the yeast and *Xenopus* proteins being too distant for similarity to be detectable by BLAST.

SUMO-targeted ubiquitin ligases (STUbLs) are ubiquitin ligases that recognize SUMO-modified proteins. A STUbL consisting of the Slx8 ubiquitin ligase and the associated protein Slx5 functions in proteasome-mediated turnover of several proteins associated with DNA replication, repair, and chromosome structure (57–59). Our model of the Slx5-Slx8 complex (fig. S21C) provides insight into how these two proteins may collectively recognize their substrates. In addition, we generated a lower-confidence but intriguing model of a previously undescribed complex between Slx8 and Cue3 [coupling of ubiquitin conjugation to endoplasmic reticulum (ER) degradation protein 3] (fig. S21D), possibly linking ubiquitination of substrates to protein degradation in ER.

Complexes involved in chromosome segregation

The heterodecameric complex DASH/Dam1 (Dam1c) is composed of 10 proteins—Ask1, Dad1, Dad2, Dad3, Dad4, Dam1, Duo1, Hsk3, Spc19, and Spc34—which come together to form a "T" shape and can further oligomerize into rings (60, 61). During mitosis, these heterodecamers strengthen the attachment between kinetochores and microtubules (62) by oligomerizing to form either partial or complete rings around microtubules and further contacting kinetochore components (63–65). Microtubules are required for in vivo ring formation, but a structure of the Dam1c ring complex from *Chaetomium thermophilum* was determined in the absence of microtubules using monovalent salts (66). We generated structure models of nine binary complexes (Dad2–Ask1, Dad2–Hsk3, Dad2–Spc1, Dad4–Hsk3, Dam1–Duo1, Duo1–Dad1, Spc19–Dad1, Spc34–Duo1, and Spc34–Spc19) that encompass several members of Dam1c (fig. S22). These complexes are largely consistent with the Dam1c structure, suggesting that the findings from the thermophile structure can likely be extended to *S. cerevisiae*. We went beyond previous structural data by predicting the structure of a potential interdecamer interaction between a loop on Spc19 and the N terminus of Dad1, which could be important for ring formation in vivo (66).

Complexes involved in molecule transport and membrane trafficking

The small-membrane protein Ksh1 is essential for growth and conserved across eukaryotes, and plays an unknown role in protein secretion (67). We predicted structures of complexes between Ksh1 and two membrane proteins reported to form a complex: Yos1 and Yip1. This complex also includes Yif1 and interacts with Rab GTPases (68) (Fig. 3). These structures suggest that Ksh1 is a fourth member of this enigmatic complex that is essential to the secretory pathway and explain how Ksh1 can play a role in secretion despite its small size of 72 amino acids.

The vacuolar transporter chaperone (VTC) is a five-subunit complex that synthesizes polyphosphate to regulate cellular phosphate concentrations (69). Structures are only known for some soluble portions of this complex, including the catalytic domain of the Vtc4 subunit (70). Our model of the previously not structurally characterized Vtc1–Vtc4 subcomplex suggests that the cytosolic active site is positioned by the complex to feed the polyphosphate product through a membrane pore into the lumen of the lysosome (Fig. 3).

The ESCRT-III complex is involved in a number of cellular membrane remodeling pathways, including receptor down-regulation, membrane repair, and cell division (71, 72). Our predicted interface between the Vps2 and Vps24 subunits of the ESCRT-III complex resembles the polymerization interface of a different ESCRT-III subunit, Snf7 (73), providing insight into the roles of these previously uncharacterized ESCRT-III subunits and highlighting the generality of this mode of interaction in ESCRT-III complexes. Notably, previously unpublished mutations (fig. S23) in Vps24 that prevent ESCRT function in multivesicular body sorting are located on the predicted interface between Vps2 and Vps24, supporting our model and the functional importance of the Vps2–Vps24 interaction. Vps55 and Vps68 are conserved membrane proteins that are important for endosomal cargo sorting; our

predicted structure (Fig. 2) of their interaction provides clues about the mechanism of their function (*74*).

The GARP complex is a multisubunit tethering complex (MTC) that mediates docking and fusion of vesicles with the Golgi apparatus (*75*). Our approach generated models for binary complexes involving the four GARP subunits, and we further modeled the entire complex (fig. S24A). In this model, the four subunits assemble through a four-helix bundle. In each of the three larger subunits, Vps52, Vps53, and Vps54, C-terminal domains comprising "CATCHR" folds emanate from the bundle. This architecture resembles portions of the cryo-EM structure of the Exocyst complex, a distinct MTC that mediates fusion of vesicles at the plasma membrane (*76*), which possesses two separate four-helix bundles organizing its eight subunits. In our prediction, the "CATCHR" domains appear to be somewhat flexibly linked to the central four-helix bundle, and hence we overlaid the structure predictions for Vps52, Vps53, and Vps54, respectively, onto the central four-helix bundle (Fig. 5C and fig. S24B). The resulting model has a marked resemblance to previously published 2D classes (fig. S24C) from a negative-stain EM analysis of the GARP complex (*77*). These predictions will facilitate structure-guided experiments to elucidate the mechanism of MTC function.

Golgi-resident protein, Grh1, forms a tethering complex with Uso1 and Bug1 that interacts with the COPII coat protein complex, Sec23–Sec24. The tether is thought to participate in COPII vesicle capture (*78*, *79*), but the mechanism remains unclear. The C terminus of Grh1 contains a predicted intrinsically disordered region (IDR) with a net positively charged cluster and a triple-proline motif (fig. S25, A and B). Our model of the Sec23–Grh1 complex contains an interface between the Sec23 gelsolin domain and the PPP motif of Grh1 (*80*), and an interface between the Grh1 IDR and Sec23 involving a disorder-to-helical transition (fig. S25C). A similar multivalent interface also drives interaction between Sec23 and the COPII coat scaffolding protein, Sec31 (*81*). Our model suggests that the combinatorial multivalent interaction between Grh1 and Sec23 may compete with the interaction between Sec31 and Sec23 to promote vesicle uncoating; consistent with this model, Grh1 is recruited to glutathione *S*-transferase (GST)–Sec23, dependent on the IDR, and competes for Sec31 binding (fig. S25D).

SNARE [soluble *N*-ethylmaleimide-sensitive factor (NSF) attachment protein (SNAP) receptor] proteins drive intracellular membrane fusion between transport vesicles and organelles (*82*). Our predicted complex structure between the SNARE Sed5 and the uncharacterized transmembrane protein Sft2 unexpectedly predicted an interaction between transmembrane domains of the two proteins (Fig. 3). SNARE localization is thought to occur through interactions of cytoplasmic domains with cytoplasmic sorting factors, but this prediction, together with genetic evidence (*83*), suggests that SNARE localization or function may be subject to additional mechanisms through interactions with transmembrane protein regulators. Membrane fusion requires the formation of a four-helix bundle (called the SNARE complex) between the vesicle SNARE and the target membrane SNAREs (*84*, *85*). The bundle is formed by the SNARE motifs, which are 60 to 70 amino acids with heptad repeats and the ability to form coiled-coil structures. Models of binary complexes of SNARE-motif–containing proteins frequently differ from their classic conformation in the SNARE four-helical bundle (fig. S26A), probably because all four chains are required to form the stable complex (*86*). Indeed, modeling the four SNARE proteins (Ufe1, Use1, Sec20, and Sec22) that are known to mediate the fusion between Golgi-derived retrograde transport vesicles with ER (*87*) together resulted in a complex that resembles a typical SNARE complex (*84*) (fig. S26, B and C). This example highlights the potential pitfalls of modeling only binary complexes when the functional assembly involves more than two chains.

GPI transamidase complex

Glycosylphosphatidylinositol transamidase (GPI-T) is a pentameric enzyme complex of unknown structure (*88*–*90*) that catalyzes the attachment of GPI anchors to the C terminus of specific substrate proteins, based on recognition of a C-terminal signal peptide (*91*). GPI-T catalyzes the removal of this signal sequence, replacing it with a new amide bond to an ethanolamine phosphate in the GPI anchor. The five subunits of *S. cerevisiae* GPI-T are Gpi8 (which contains the catalytic active site), Gpi16, Gaa1, Gpi17, and Gab1 (*88*, *92*, *93*). Our large-scale modeling approach generated models for the following binary complexes: Gpi8–Gpi17, Gab1–Gaa1, Gab1–Gpi17, and Gaa1–Gpi16. We subsequently modeled the full-length, pentameric GPI-T in one shot, starting from the sequences of all components (Fig. 5E). Several features of this model are consistent with previous characterization of this enzyme. *S. cerevisiae* GPI-T can be purified as a core heterotrimer, containing only Gpi8, Gpi16, and Gaa1 (*92*); our GPI-T model confirms extensive interactions between the soluble domains of these three subunits. This model also recapitulates the disulfide bond between Gpi8 (Cys^{85}) and Gpi16 (Cys^{202}), previously characterized for human GPI-T (*94*) [the existence of this disulfide bond in yeast GPI-T has been called into question (*90*)]. Gaa1 is essential for binding of the GPI anchor to GPI-T (*95*), and the hydrophobic Gab1 is also predicted to participate in anchor recognition (*88*). Our model positions the transmembrane regions of Gaa1 and Gab1 against each other. The catalytic dyad in Gpi8 (Cys^{199} and His^{157}) faces these transmembrane domains, and abuts a highly conserved face of Gaa1, proposed to recognize the GPI anchor glycans (*96*, *97*). In our model, the positions of these subunits are consistent with binding of the GPI anchor to position its modifying amine in the Gpi8 active site for catalysis. Gpi16 is immediately adjacent to these interactions and is likely involved in anchor recognition. The functional role of Gpi17 has been elusive, but our model now suggests that Gpi17, together with Gpi8 and Gpi16, forms a recognition channel for the C-terminal GPI-T signal peptide (fig. S27) adjacent to the catalytic dyad (Fig 5E). In vivo, GPI-T is expected to be a dimer of pentamers, with dimerization occurring on one face of the caspase-like Gpi8 subunit (*92*, *97*, *98*). This decameric complex was too large for us to model computationally; however, the pentameric complex that we present here leaves open the dimerization face of Gpi8 consistent with probable dimerization. It also suggests that Gaa1 and Gpi17 would participate in dimerization of this enzyme. In humans, mutations in GPI-T subunits are associated with neurodevelopmental disorders (*99*). Each subunit contributes to different cancer mechanisms, in some cases by perturbing GPI anchoring of specific receptors and in others by separating from GPI-T to alter disparate signal transduction pathways (*89*). Now, with a structural model in hand, these mechanisms can be examined at a molecular level.

Limitations of the current method

As with any new method, it is important when interpreting the results (our large set of predicted complex structures) to keep in mind the limitations of the approach. First, our study is not comprehensive, so conclusions should not be drawn about absences; in particular, we eliminated proteins that arose from recent duplication owing to difficulty in identifying orthologs in other organisms, and thus only surveyed two-thirds of the entire yeast proteome. Second, the approach likely misses interactions restricted to a small set of organisms, or that vary rapidly during evolution, owing to weaker coevolutionary signals. Third, the approach likely works less well for transient interactions that generally involve smaller and weaker interfaces, which may be under lower selective pressure, in particular those involving intrinsically disordered regions, which are poorly represented in the PDB. The majority of known interactions identified by our approach are likely obligate assemblies and involve ordered structural elements. Fourth, interactions between single hydrophobic or amphipathic

helices, such as single transmembrane helices or coiled coils, may be overpredicted (in initial studies of human complexes, interactions solely between single-pass transmembrane regions appear to be over-represented). Fifth, and perhaps most important, for proteins that form high-order obligate protein complexes, binary complex models may be quite inaccurate, as illustrated by the SNARE example.

Conclusion

Our approach extends the range of large-scale deep-learning–based structure modeling from monomeric proteins to protein assemblies. As highlighted by the above examples, following up on the many new complexes presented here should advance understanding of a wide range of eukaryotic cellular processes and provide new targets for therapeutic intervention. The methods can be extended directly to large-scale mapping of interactions in the human proteome, but considerably more compute time will be required given the much larger total number of protein pairs, and models may be somewhat less accurate owing to weaker coevolutionary signal for the subset of human proteins specific to higher eukaryotes and for the many closely related paralogs arising from gene duplication. Investigating interactions of individual proteins or subsets of proteins— for example, deorphanization of orphan receptors—should be immediately accessible using our approach provided there are sufficient sequence homologs. Training RF and AF on protein complexes should further improve performance of both methods (*100*), particularly for protein pairs with fewer homologs and/or weaker and more transient interactions, and reduce the dependence on ortholog identification. Together with the advances in monomeric structure prediction, our results herald a new era of structural biology in which computation plays a fundamental role in both interaction discovery and structure determination.

Methods

As described in detail in the supplementary materials and methods, we developed a multistep bioinformatics and deep learning pipeline for identifying pairs of proteins likely to interact and modeling the 3D structures of the corresponding protein complexes. The steps of this pipeline are illustrated schematically in Fig. 1A. First, comprehensive orthologous groups of genes were generated and yeast genes were mapped to these groups; second, multiple sequence alignments of orthologous sequences were generated for each pair of yeast proteins; third, contact probability was computed for each protein pair using RoseTTAFold; and fourth, interaction probability was reevaluated, and complex structures were modeled using AlphaFold. The experimental data-guided PPI screening pipeline is very similar except that in the third stage, instead of using RoseTTAFold, we used experimental data primarily derived from large-scale screens to identify PPI candidates.

REFERENCES AND NOTES

1. T. Ito *et al*., A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569–4574 (2001). doi: 10.1073/pnas.061034498; pmid: 11283351
2. S. R. Collins *et al*., Toward a comprehensive atlas of the physical interactome of Saccharomyces cerevisiae. *Mol. Cell. Proteomics* **6**, 439–450 (2007). doi: 10.1074/mcp.M600381-MCP200; pmid: 17200106
3. T. Reguly *et al*., Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. *J. Biol.* **5**, 11 (2006). doi: 10.1186/jbiol36; pmid: 16762047
4. P. Uetz *et al*., A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* **403**, 623–627 (2000). doi: 10.1038/35001009; pmid: 10688190
5. H. Yu *et al*., High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008). doi: 10.1126/science.1158684; pmid: 18719252
6. O. Kuchaiev, M. Rasajski, D. J. Higham, N. Przulj, Geometric de-noising of protein-protein interaction networks. *PLOS Comput. Biol.* **5**, e1000454 (2009). doi: 10.1371/journal.pcbi.1000454; pmid: 19662157
7. A. M. Edwards *et al*., Bridging structural biology and genomics: Assessing protein interaction data with known complexes. *Trends Genet.* **18**, 529–536 (2002). doi: 10.1016/S0168-9525(02)02763-4; pmid: 12350343
8. J. P. Mackay, M. Sunde, J. A. Lowry, M. Crossley, J. M. Matthews, Protein interactions: Is seeing believing? *Trends Biochem. Sci.* **32**, 530–531 (2007). doi: 10.1016/j.tibs.2007.09.006; pmid: 17980603
9. Q. Cong, I. Anishchenko, S. Ovchinnikov, D. Baker, Protein interaction networks revealed by proteome coevolution. *Science* **365**, 185–189 (2019). pmid: 31296772
10. A. G. Green *et al*., Large-scale discovery of protein interactions at residue resolution using co-evolution calculated from genomic sequences. *Nat. Commun.* **12**, 1396 (2021). doi: 10.1038/s41467-021-21636-z; pmid: 33654096
11. S. Ovchinnikov, H. Kamisetty, D. Baker, Robust and accurate prediction of residue-residue interactions across protein interfaces using evolutionary information. *eLife* **3**, e02030 (2014). doi: 10.7554/eLife.02030; pmid: 24842992
12. T. A. Hopf *et al*., Sequence co-evolution gives 3D contacts and structures of protein complexes. *eLife* **3**, e03430 (2014). doi: 10.7554/eLife.03430; pmid: 25255213
13. M. Baek *et al*., Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021). doi: 10.1126/science.abj8754; pmid: 34282049
14. J. Jumper *et al*., Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021). doi: 10.1038/s41586-021-03819-2; pmid: 34265844
15. A. Meyer, M. Schartl, Gene and genome duplications in vertebrates: The one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.* **11**, 699–704 (1999). doi: 10.1016/S0955-0674(99)00039-3; pmid: 10600714
16. I. V. Grigoriev *et al*., MycoCosm portal: Gearing up for 1000 fungal genomes. *Nucleic Acids Res.* **42** (D1), D699–D704 (2014). doi: 10.1093/nar/gkt1183; pmid: 24297253
17. M. Spingola, L. Grate, D. Haussler, M. Ares Jr., Genome-wide bioinformatic and molecular analysis of introns in Saccharomyces cerevisiae. *RNA* **5**, 221–234 (1999). doi: 10.1017/S1355838299981682; pmid: 10024174
18. E. M. Zdobnov *et al*., OrthoDB in 2020: Evolutionary and functional annotations of orthologs. *Nucleic Acids Res.* **49** (D1), D389–D393 (2021). doi: 10.1093/nar/gkaa1009; pmid: 33196836
19. A. Clum *et al*., DOE JGI Metagenome Workflow. *mSystems* **6**, e00804-20 (2021). doi: 10.1128/mSystems.00804-20; pmid: 34006627
20. D. P. Wall, H. B. Fraser, A. E. Hirsh, Detecting putative orthologs. *Bioinformatics* **19**, 1710–1711 (2003). doi: 10.1093/bioinformatics/btg213; pmid: 15593400
21. R. Oughtred *et al*., The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* **30**, 187–200 (2021). doi: 10.1002/pro.3978; pmid: 33070389
22. H. Huang, B. M. Jedynak, J. S. Bader, Where have all the interactions gone? Estimating the coverage of two-hybrid protein interaction maps. *PLOS Comput. Biol.* **3**, e214 (2007). doi: 10.1371/journal.pcbi.0030214; pmid: 18039026
23. S. Keeney, C. N. Giroux, N. Kleckner, Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* **88**, 375–384 (1997). doi: 10.1016/S0092-8674(00)81876-0; pmid: 9039264
24. B. de Massy, Initiation of meiotic recombination: How and where? Conservation and specificities among eukaryotes. *Annu. Rev. Genet.* **47**, 563–599 (2013). doi: 10.1146/annurev-genet-110711-155423; pmid: 24050176
25. H. Murakami, S. Keeney, Regulating the formation of DNA double-strand breaks in meiosis. *Genes Dev.* **22**, 286–292 (2008). doi: 10.1101/gad.1642308; pmid: 18245442
26. C. Arora, K. Kee, S. Maleki, S. Keeney, Antiviral protein Ski8 is a direct partner of Spo11 in meiotic DNA break formation, independent of its cytoplasmic role in RNA metabolism. *Mol. Cell* **13**, 549–559 (2004). doi: 10.1016/S1097-2765(04)00063-2; pmid: 14992724
27. C. Claeys Bouuaert *et al*., Structural and functional characterization of the Spo11 core complex. *Nat. Struct. Mol. Biol.* **28**, 92–102 (2021). doi: 10.1038/s41594-020-00534-w; pmid: 33398171
28. F. Halbach, P. Reichelt, M. Rode, E. Conti, The yeast ski complex: Crystal structure and RNA channeling to the exosome complex. *Cell* **154**, 814–826 (2013). doi: 10.1016/j.cell.2013.07.017; pmid: 23953113
29. S. Steiner, J. Kohli, K. Ludin, Functional interactions among members of the meiotic initiation complex in fission yeast. *Curr. Genet.* **56**, 237–249 (2010). doi: 10.1007/s00294-010-0296-0; pmid: 20364342
30. S. Tessé, A. Storlazzi, N. Kleckner, S. Gargano, D. Zickler, Localization and roles of Ski8p protein in Sordaria meiosis and delineation of three mechanistically distinct steps of meiotic homolog juxtaposition. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 12865–12870 (2003). doi: 10.1073/pnas.2034282100; pmid: 14563920
31. T. Robert *et al*., The TopoVIB-Like protein family is required for meiotic DNA double-strand break formation. *Science* **351**, 943–949 (2016). doi: 10.1126/science.aad5309; pmid: 26917764
32. K. D. Corbett, P. Benedetti, J. M. Berger, Holoenzyme assembly and ATP-mediated conformational dynamics of topoisomerase VI. *Nat. Struct. Mol. Biol.* **14**, 611–619 (2007). doi: 10.1038/nsmb1264; pmid: 17603498
33. L. Salem, N. Walter, R. Malone, Suppressor analysis of the Saccharomyces cerevisiae gene REC104 reveals a genetic interaction with REC102. *Genetics* **151**, 1261–1272 (1999). doi: 10.1093/genetics/151.4.1261; pmid: 10101155
34. M. R. Sullivan, K. A. Bernstein, RAD-ical New Insights into RAD51 Regulation. *Genes* **9**, 629 (2018). doi: 10.3390/genes9120629; pmid: 30551670
35. J. San Filippo, P. Sung, H. Klein, Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.* **77**, 229–257 (2008). doi: 10.1146/annurev.biochem.77.061306.125255; pmid: 18275380
36. U. Roy *et al*., The Rad51 paralog complex Rad55-Rad57 acts as a molecular chaperone during homologous recombination. *Mol. Cell* **81**, 1043–1057.e8 (2021). doi: 10.1016/j.molcel.2020.12.019; pmid: 33421364
37. A. B. Conway *et al*., Crystal structure of a Rad51 filament. *Nat. Struct. Mol. Biol.* **11**, 791–796 (2004). doi: 10.1038/nsmb795; pmid: 15235592
38. K. Sugasawa, J. Akagi, R. Nishi, S. Iwai, F. Hanaoka, Two-step recognition of DNA damage for mammalian nucleotide excision repair: Directional binding of the XPC complex and DNA strand scanning. *Mol. Cell* **36**, 642–653 (2009). doi: 10.1016/j.molcel.2009.09.035; pmid: 19941824
39. G. Kokic *et al*., Structural basis of TFIIH activation for nucleotide excision repair. *Nat. Commun.* **10**, 2885 (2019). doi: 10.1038/s41467-019-10745-5; pmid: 31253769
40. J. R. Thompson, Z. C. Ryan, J. L. Salisbury, R. Kumar, The structure of the human centrin 2-xeroderma pigmentosum group C protein complex. *J. Biol. Chem.* **281**, 18746–18752 (2006). doi: 10.1074/jbc.M513667200; pmid: 16627479
41. T. van Eeuwen *et al*., Cryo-EM structure of TFIIH/Rad4-Rad23-Rad33 in damaged DNA opening in nucleotide excision

repair. *Nat. Commun.* **12**, 3338 (2021). doi: 10.1038/s41467-021-23684-x; pmid: 34099686
42. T. Riedl, F. Hanaoka, J.-M. Egly, The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J.* **22**, 5293–5303 (2003). doi: 10.1093/emboj/cdg489; pmid: 14517266
43. S. Klinge, F. Voigts-Hoffmann, M. Leibundgut, N. Ban, Atomic structures of the eukaryotic ribosome. *Trends Biochem. Sci.* **37**, 189–198 (2012). doi: 10.1016/j.tibs.2012.02.007; pmid: 22436288
44. A. G. Hinnebusch, I. P. Ivanov, N. Sonenberg, Translational control by 5′-untranslated regions of eukaryotic mRNAs. *Science* **352**, 1413–1416 (2016). doi: 10.1126/science.aad9868; pmid: 27313038
45. J. A. Saba, K. Liakath-Ali, R. Green, F. M. Watt, Translational control of stem cell function. *Nat. Rev. Mol. Cell Biol.* **22**, 671–690 (2021). doi: 10.1038/s41580-021-00386-2; pmid: 34272502
46. S. Klinge, J. L. Woolford Jr.., Ribosome assembly coming into focus. *Nat. Rev. Mol. Cell Biol.* **20**, 116–131 (2019). doi: 10.1038/s41580-018-0078-y; pmid: 30467428
47. K. M. Mulvaney *et al*., Molecular basis for substrate recruitment to the PRMT5 methylosome. *Mol. Cell* **81**, 3481–3495.e7 (2021). doi: 10.1016/j.molcel.2021.07.019; pmid: 34358446
48. Z. L. Watson *et al*., Structure of the bacterial ribosome at 2 Å resolution. *eLife* **9**, e60482 (2020). doi: 10.7554/eLife.60482; pmid: 32924932
49. J. M. Matecki *et al*., Human METTL18 is a histidine-specific methyltransferase that targets RPL3 and affects ribosome biogenesis and function. *Nucleic Acids Res.* **49**, 3185–3203 (2021). doi: 10.1093/nar/gkab088; pmid: 33693809
50. F. Dragon *et al*., A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* **417**, 967–970 (2002). doi: 10.1038/nature00769; pmid: 12068309
51. L. R. Kenner *et al*., eIF2B-catalyzed nucleotide exchange and phosphoregulation by the integrated stress response. *Science* **364**, 491–495 (2019). doi: 10.1126/science.aaw2922; pmid: 31048491
52. S. Jentsch, I. Psakhye, Control of nuclear activities by substrate-selective and protein-group SUMOylation. *Annu. Rev. Genet.* **47**, 167–186 (2013). doi: 10.1146/annurev-genet-111212-133453; pmid: 24016193
53. I. Psakhye, S. Jentsch, Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. *Cell* **151**, 807–820 (2012). doi: 10.1016/j.cell.2012.10.021; pmid: 23122649
54. D. Menolfi, A. Delamarre, A. Lengronne, P. Pasero, D. Branzei, Essential Roles of the Smc5/6 Complex in Replication through Natural Pausing Sites and Endogenous DNA Damage Tolerance. *Mol. Cell* **60**, 835–846 (2015). doi: 10.1016/j.molcel.2015.10.023; pmid: 26698660
55. S. Agashe *et al*., Smc5/6 functions with Sgs1-Top3-Rmi1 to complete chromosome replication at natural pause sites. *Nat. Commun.* **12**, 2111 (2021). doi: 10.1038/s41467-021-22217-w; pmid: 33833229
56. G. De Piccoli, J. Torres-Rosell, L. Aragón, The unnamed complex: What do we know about Smc5-Smc6? *Chromosome Res.* **17**, 251–263 (2009). doi: 10.1007/s10577-008-9016-8; pmid: 19308705
57. I. Psakhye, F. Castellucci, D. Branzei, SUMO-Chain-Regulated Proteasomal Degradation Timing Exemplified in DNA Replication Initiation. *Mol. Cell* **76**, 632–645.e6 (2019). doi: 10.1016/j.molcel.2019.08.003; pmid: 31519521
58. A. Waizenegger *et al*., Mus81-Mms4 endonuclease is an Esc2-STUbL-Cullin8 mitotic substrate impacting on genome integrity. *Nat. Commun.* **11**, 5746 (2020). doi: 10.1038/s41467-020-19503-4; pmid: 33184279
59. I. Psakhye, D. Branzei, SMC complexes are guarded by the SUMO protease Ulp2 against SUMO-chain-mediated turnover. *Cell Rep.* **36**, 109485 (2021). doi: 10.1016/j.celrep.2021.109485; pmid: 34348159
60. J. J. L. Miranda, P. De Wulf, P. K. Sorger, S. C. Harrison, The yeast DASH complex forms closed rings on microtubules. *Nat. Struct. Mol. Biol.* **12**, 138–143 (2005). doi: 10.1038/nsmb896; pmid: 15640796
61. S. Westermann *et al*., Formation of a dynamic kinetochore-microtubule interface through assembly of the Dam1 ring complex. *Mol. Cell* **17**, 277–290 (2005). doi: 10.1016/j.molcel.2004.12.019; pmid: 15664196
62. C. L. Asbury, D. R. Gestaut, A. F. Powers, A. D. Franck, T. N. Davis, The Dam1 kinetochore complex harnesses microtubule dynamics to produce force and movement. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9873–9878 (2006). doi: 10.1073/pnas.0602249103; pmid: 16777964
63. V. H. Ramey *et al*., Subunit organization in the Dam1 kinetochore complex and its ring around microtubules. *Mol. Biol. Cell* **22**, 4335–4342 (2011). doi: 10.1091/mbc.e11-07-0659; pmid: 21965284
64. J. O. Kim *et al*., The Ndc80 complex bridges two Dam1 complex rings. *eLife* **6**, e21069 (2017). doi: 10.7554/eLife.21069; pmid: 28191870
65. C. T. Ng *et al*., Electron cryotomography analysis of Dam1C/DASH at the kinetochore-spindle interface in situ. *J. Cell Biol.* **218**, 455–473 (2019). doi: 10.1083/jcb.201809088; pmid: 30504246
66. S. Jenni, S. C. Harrison, Structure of the DASH/Dam1 complex shows its role at the yeast kinetochore-microtubule interface. *Science* **360**, 552–558 (2018). doi: 10.1126/science.aar6436; pmid: 29724956
67. F. Wendler *et al*., A genome-wide RNA interference screen identifies two novel components of the metazoan secretory pathway. *EMBO J.* **29**, 304–314 (2010). doi: 10.1038/emboj.2009.350; pmid: 19942856
68. M. Heidtman, C. Z. Chen, R. N. Collins, C. Barlowe, Yos1p is a novel subunit of the Yip1p-Yif1p complex and is required for transport between the endoplasmic reticulum and the Golgi complex. *Mol. Biol. Cell* **16**, 1673–1683 (2005). doi: 10.1091/mbc.e04-10-0873; pmid: 15659647
69. Y. Desfougères, R. U. Gerasimaitė, H. J. Jessen, A. Mayer, Vtc5, a Novel Subunit of the Vacuolar Transporter Chaperone Complex, Regulates Polyphosphate Synthesis and Phosphate Homeostasis in Yeast. *J. Biol. Chem.* **291**, 22262–22275 (2016). doi: 10.1074/jbc.M116.746784; pmid: 27587415
70. M. Hothorn *et al*., Catalytic core of a membrane-associated eukaryotic polyphosphate polymerase. *Science* **324**, 513–516 (2009). doi: 10.1126/science.1168120; pmid: 19390046
71. M. Vietri, M. Radulovic, H. Stenmark, The many functions of ESCRTs. *Nat. Rev. Mol. Cell Biol.* **21**, 25–42 (2020). doi: 10.1038/s41580-019-0177-4; pmid: 31705132
72. J. H. Hurley, ESCRTs are everywhere. *EMBO J.* **34**, 2398–2407 (2015). doi: 10.15252/embj.201592484; pmid: 26311197
73. S. Tang *et al*., Structural basis for activation, assembly and membrane binding of ESCRT-III Snf7 filaments. *eLife* **4**, e12548 (2015). doi: 10.7554/eLife.12548; pmid: 26670543
74. C. Schluter *et al*., Global analysis of yeast endosomal transport identifies the vps55/68 sorting complex. *Mol. Biol. Cell* **19**, 1282–1294 (2008). doi: 10.1091/mbc.e07-07-0659; pmid: 18216282
75. S. Siniossoglou, H. R. Pelham, An effector of Ypt6p binds the SNARE Tlg1p and mediates selective fusion of vesicles with late Golgi membranes. *EMBO J.* **20**, 5991–5998 (2001). doi: 10.1093/emboj/20.21.5991; pmid: 11689439
76. K. Mei *et al*., Cryo-EM structure of the exocyst complex. *Nat. Struct. Mol. Biol.* **25**, 139–146 (2018). doi: 10.1038/s41594-017-0016-2; pmid: 29335562
77. H.-T. Chou, D. Dukovski, M. G. Chambers, K. M. Reinisch, T. Walz, CATCHR, HOPS and CORVET tethering complexes share a similar architecture. *Nat. Struct. Mol. Biol.* **23**, 761–763 (2016). doi: 10.1038/nsmb.3264; pmid: 27428774
78. R. Behnia, F. A. Barr, J. J. Flanagan, C. Barlowe, S. Munro, The yeast orthologue of GRASP65 forms a complex with a coiled-coil protein that contributes to ER to Golgi traffic. *J. Cell Biol.* **176**, 255–261 (2007). doi: 10.1083/jcb.200607151; pmid: 17261844
79. M. Schuldiner *et al*., Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005). doi: 10.1016/j.cell.2005.08.031; pmid: 16269340
80. W. Ma, J. Goldberg, TANGO1/cTAGE5 receptor as a polyvalent template for assembly of large COPII coats. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10061–10066 (2016). doi: 10.1073/pnas.1605916113; pmid: 27551091
81. V. G. Stancheva *et al*., Combinatorial multivalent interactions drive cooperative assembly of the COPII coat. *J. Cell Biol.* **219**, e202007135 (2020). doi: 10.1083/jcb.202007135; pmid: 32997735
82. T. C. Südhof, J. E. Rothman, Membrane fusion: Grappling with SNARE and SM proteins. *Science* **323**, 474–477 (2009). doi: 10.1126/science.1161748; pmid: 19164740
83. S. Conchon, X. Cao, C. Barlowe, H. R. Pelham, Got1p and Sft2p: Membrane proteins involved in traffic to the Golgi complex. *EMBO J.* **18**, 3934–3946 (1999). doi: 10.1093/emboj/18.14.3934; pmid: 10406798
84. R. B. Sutton, D. Fasshauer, R. Jahn, A. T. Brunger, Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution. *Nature* **395**, 347–353 (1998). doi: 10.1038/26412; pmid: 9759724
85. R. Jahn, R. H. Scheller, SNAREs—Engines for membrane fusion. *Nat. Rev. Mol. Cell Biol.* **7**, 631–643 (2006). doi: 10.1038/nrm2002; pmid: 16912714
86. J. Rizo, Mechanism of neurotransmitter release coming into focus. *Protein Sci.* **27**, 1364–1391 (2018). doi: 10.1002/pro.3445; pmid: 29893445
87. L. Burri *et al*., A SNARE required for retrograde transport to the endoplasmic reticulum. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9873–9877 (2003). doi: 10.1073/pnas.1734000100; pmid: 12893879
88. Y. Hong *et al*., Human PIG-U and yeast Cdc91p are the fifth subunit of GPI transamidase that attaches GPI-anchors to proteins. *Mol. Biol. Cell* **14**, 1780–1789 (2003). doi: 10.1091/mbc.e02-12-0794; pmid: 12802054
89. D. G. Gamage, T. L. Hendrickson, GPI transamidase and GPI anchored proteins: Oncogenes and biomarkers for cancer. *Crit. Rev. Biochem. Mol. Biol.* **48**, 446–464 (2013). doi: 10.3109/10409238.2013.831024; pmid: 23978072
90. L. Yi *et al*., Disulfide Bond Formation and N-Glycosylation Modulate Protein-Protein Interactions in GPI-Transamidase (GPIT). *Sci. Rep.* **7**, 45912 (2017). doi: 10.1038/srep45912; pmid: 28374821
91. P. Moran, I. W. Caras, A nonfunctional sequence converted to a signal for glycophosphatidylinositol membrane anchor attachment. *J. Cell Biol.* **115**, 329–336 (1991). doi: 10.1083/jcb.115.2.329; pmid: 1717483
92. P. Fraering *et al*., The GPI transamidase complex of Saccharomyces cerevisiae contains Gaa1p, Gpi8p, and Gpi16p. *Mol. Biol. Cell* **12**, 3295–3306 (2001). doi: 10.1091/mbc.12.10.3295; pmid: 11598210
93. K. Ohishi, N. Inoue, T. Kinoshita, PIG-S and PIG-T, essential for GPI anchor attachment to proteins, form a complex with GAA1 and GPI8. *EMBO J.* **20**, 4088–4098 (2001). doi: 10.1093/emboj/20.15.4088; pmid: 11483512
94. K. Ohishi, K. Nagamune, Y. Maeda, T. Kinoshita, Two subunits of glycosylphosphatidylinositol transamidase, GPI8 and PIG-T, form a functionally important intermolecular disulfide bridge. *J. Biol. Chem.* **278**, 13959–13967 (2003). doi: 10.1074/jbc.M300586200; pmid: 12582175
95. S. Vainauskas, A. K. Menon, A conserved proline in the last transmembrane segment of Gaa1 is required for glycosylphosphatidylinositol (GPI) recognition by GPI transamidase. *J. Biol. Chem.* **279**, 6540–6545 (2004). doi: 10.1074/jbc.M312191200; pmid: 14660601
96. U. Meyer, M. Benghezal, I. Imhof, A. Conzelmann, Active site determination of Gpi8p, a caspase-related enzyme required for glycosylphosphatidylinositol anchor addition to proteins. *Biochemistry* **39**, 3461–3471 (2000). doi: 10.1021/bi992186o; pmid: 10727241
97. D. G. Gamage *et al*., The soluble domains of Gpi8 and Gaa1, two subunits of glycosylphosphatidylinositol transamidase (GPI-T), assemble into a complex. *Arch. Biochem. Biophys.* **633**, 58–67 (2017). doi: 10.1016/j.abb.2017.09.006; pmid: 28893510
98. J. L. Meitzler, J. J. Gray, T. L. Hendrickson, Truncation of the caspase-related subunit (Gpi8p) of Saccharomyces cerevisiae GPI transamidase: Dimerization revealed. *Arch. Biochem. Biophys.* **462**, 83–93 (2007). doi: 10.1016/j.abb.2007.03.035; pmid: 17475206
99. T. T. M. Nguyen *et al*., Bi-allelic Variants in the GPI Transamidase Subunit PIGK Cause a Neurodevelopmental Syndrome with Hypotonia, Cerebellar Atrophy, and Epilepsy. *Am. J. Hum. Genet.* **106**, 484–495 (2020). doi: 10.1016/j.ajhg.2020.03.001; pmid: 32220290
100. R. Evans *et al*., Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034 [Preprint]. 4 October 2021. doi: 10.1101/2021.10.04.463034
101. M. Baek, L. Heo, R. Ndem, neilfleckSCRI, RosettaCommons/RoseTTAFold: RoseTTAFold update: Including the simpler version for PPI screening, Zenodo (2021). doi: 10.5281/zenodo.5639837

ACKNOWLEDGMENTS

We thank E. Horvitz, N. V. Grishin, H. Park, and J. H. Thomas for helpful discussions; L. Goldschmidt and A. Guillory for computing resource management; and L. Stewart for logistical support. Additionally, we are grateful to M. Bard, T. N. Davis, D. G. Drubin,

M. J. Dunham, S. D. Emr, F. Hughson, J. Hurley, K. Murakami, N. Nakamura, E. Nogales, R. Schekman, S. Shan, S. Showman, K. Sugasawa, and S. Suzuki for their correspondence and biological expertise. We thank S. Burley, B. Vallat, and J. Westbrook at the RCSB Protein Data Bank and T. Schwede, G. Tauriello, A. Waterhouse, and S. Bienert at SWISS-MODEL for hosting our model structures at the ModelArchive. **Funding:** This work was supported by Microsoft (M.B., D.B., and Azure compute time and expertise), Amgen (D.B. and I.R.H.), Southwestern Medical Foundation (J.P. and Q.C.), the Washington Research Foundation (M.B. and Q.C.), the Howard Hughes Medical Institute (D.B., S.B., S.K., and generous compute time on Janelia), National Science Foundation (NSF) Cyberinfrastructure for Biological Research (CIBR, Award DBI 1937533 to D.B. and I.A.), CPRIT training grant (RP210041 to J.Z.), UK Medical Research Council (MRC_UP_1201/10 to E.A.M.), HHMI Gilliam Fellowship (D.J.B.), the Deutsche Forschungsgemeinschaft (KI-562/11-1 and KI-562/7-1 to C.K.), NIH/NIGMS (R21AI156595 to S.O., R35GM136258 to J.C.F., R35NS097333 to J.R., R35GM118026 and R01CA221858 to E.C.G.), HHMI fellowship of the Damon Runyon Cancer Research Foundation (DRG2273-16 to S.B. and DRG2389-20 to K.L.), AIRC investigator and the European Research Council Consolidator (IG23710 and 682190 to D.Br.), and the Defense Threat Reduction Agency (HDTRA1-21-1-0007 to D.B.). We also thank The National Energy Research Scientific Computing Center (NERSC) for providing computing time (project m3962 at NERSC). **Author contributions:** Q.C. and D.B. conceived the research; J.P. and Q.C. prepared the sequence alignments used in the screen; M.B. implemented the RoseTTAFold pipeline; M.B. and S.O. repurposed AlphaFold for complex modeling; J.P., J.Z., and Q.C. designed the PPI screening procedure; I.R.H., M.B., I.A., and Q.C. carried out the screen; I.R.H., A.K., and Q.C. analyzed and presented the results; I.R.H., A.K., Q.C., and D.B. coordinated the collaborative efforts; T.J.N., S.B., S.R.B., V.G.S., X.H.L., K.L., Z.Z., D.J.B., U.R., J.K., I.S.F., B.S., D.B., J.R., C.K., E.C.G., S.B., S.K., E.A.M., J.C.F., and T.L.H. provided biological insights on specific examples; Q.C. and D.B. drafted the manuscript and all other authors contributed to the description of specific examples; all authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Structures of highly confident pairs with accompanying pMSAs and metadata are available at the ModelArchive: https://modelarchive.org/doi/10.5452/ma-bak-cepc. RoseTTAFold two-track version is available at https://github.com/RosettaCommons/RoseTTAFold or Zenodo (*101*). AlphaFold was fetched from https://github.com/deepmind/alphafold on 16 July 2021 (v2.0.0). Code for a GPU implementation of DCA and the modifications to the AlphaFold predictions script are provided in supplementary materials and methods.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.amb4805
Materials and Methods
Supplementary Text
Figs. S1 to S27
Tables S1 to S5
References (*102*–*116*)
Descriptions of all predicted protein-protein interactions

20 September 2021; accepted 2 November 2021
Published online 11 November 2021
10.1126/science.abm4805

RESEARCH ARTICLE SUMMARY

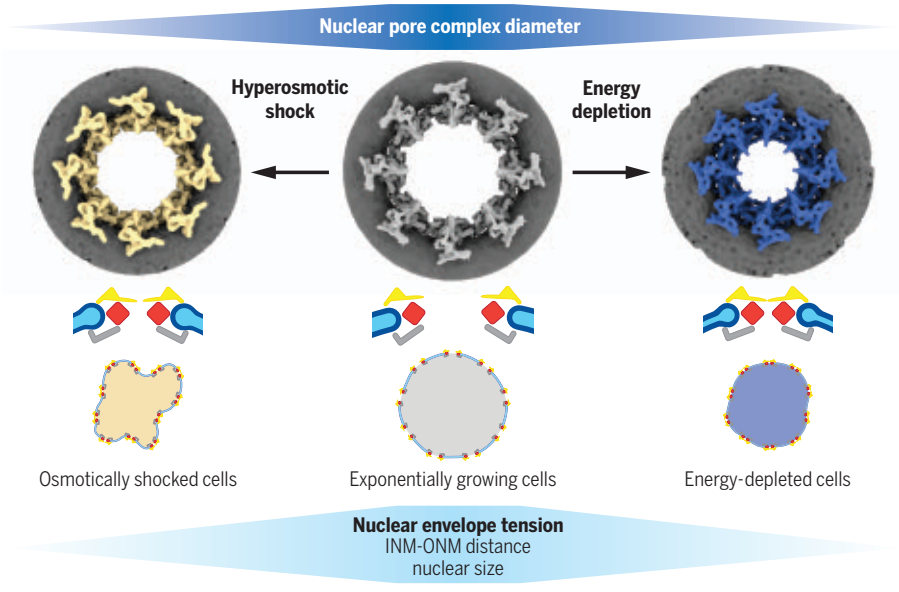
CELL BIOLOGY

Nuclear pores dilate and constrict in cellulo

Christian E. Zimmerli†, Matteo Allegretti†, Vasileios Rantos, Sara K. Goetz, Agnieszka Obarska-Kosinska, Ievgeniia Zagoriy, Aliaksandr Halavatyi, Gerhard Hummer, Julia Mahamid, Jan Kosinski*, Martin Beck*

INTRODUCTION: The nucleus harbors and protects the genetic information in eukaryotes. It is surrounded by the nuclear envelope, which separates the nucleoplasm from the cytosol. The double-membrane system of the nuclear envelope is physically connected to the cytoskeleton and chromatin on either side. It senses and signals mechanical stimuli. Nuclear pore complexes (NPCs) mediate nucleocytoplasmic exchange. They fuse the inner and outer nuclear membranes of the nuclear envelope to form an aqueous central channel. Their intricate cylindrical architecture consists of ~30 nucleoporins. The nucleoporin scaffold embraces the membrane fusion topology in a highly dynamic fashion. Variations in NPC diameter have been reported, but the physiological circumstances and the molecular details remain unknown.

RATIONALE: To investigate the dynamics of NPC conformation in vivo, we structurally analyzed NPCs in *Schizosaccharomyces pombe* cells. We used cryo–electron tomography with subsequent subtomogram averaging to quantify the diameter of NPCs in cells exposed to different stress stimuli. We found a dilated NPC scaffold in exponentially growing cells that constricted under conditions of energy depletion and hyperosmotic shock. Structural analysis in combination with integrative structural modeling revealed that the NPC scaffold underwent large-scale movements and bending during constriction. Thereby, the volume of the central channel was lessened to about one-half. Reduced nucleocytoplasmic diffusion was apparent. Scaffold nucleoporins do not have any known motor activity that would allow them to exert forces. Thus, it appears likely that the observed dynamics of the diameter of the NPC were the result of forces that are laterally applied within the nuclear envelope. Under conditions of both energy depletion and hyperosmotic shock, cellular and nuclear volumes were reduced such that nuclear shrinkage led to an excess of nuclear membranes. We therefore hypothesized that the NPC scaffold is mechanosensitive. We surmised that a reduction of nuclear envelope membrane tension resulted in a consequent NPC constriction into a conformational ground state.



NPCs constrict under conditions of cellular energy depletion and hyperosmotic shock. Cryo–electron microscopy maps of different NPC conformations are shown isosurface rendered as seen from the cytosol in exponentially growing (light gray), hyperosmotically shocked (yellow), and energy-depleted cells (blue) (top). Nuclear membranes are shown in dark gray. Cartoons indicate the respective conformational changes. Upon exposure to hyperosmotic shock and energy depletion, NPCs constrict to smaller central channel diameters. The nuclear volume is decreased, and the distance between the inner and outer nuclear membranes is reduced (bottom), which is indicative of a loss of nuclear envelope membrane tension. INM, inner nuclear membrane; ONM, outer nuclear membrane.

RESULTS: We outline several predictions of a conceptual model in which nuclear envelope membrane tension regulates NPC diameter. In such a scenario, nuclear volume and NPC diameter should be dependent on each other. The rigidity of the NPC scaffold should counteract its dilation. On the basis of membrane elastic theory, we predicted that osmotic pressure in the nuclear envelope lumen and membrane tension in the nuclear envelope act equivalently. As a consequence, the distance between the inner and outer nuclear membranes should grow linearly with both. We set out to test those predictions experimentally. We found that nuclear shrinkage correlated with NPC constriction. Recovery experiments established that this behavior was reversible. Constricted NPCs dilated when shifted back to normal medium. Nuclear volumes recovered, and cells remained viable. In exponentially growing cells, genetic perturbation of the NPC scaffold resulted in further dilation. We segmented the inner and outer nuclear membranes in cryo–electron tomograms and quantified their distance. As expected, the diameter of the NPC directly correlated with the distance between the inner and outer nuclear membranes.

CONCLUSION: Our data strongly suggest that the NPC scaffold is mechanosensitive and that membrane tension regulates its diameter. The data link the conformation of the NPC to the mechanical status of the nuclear envelope. This finding has implications for various cellular processes during which the nucleus may be exposed to mechanical forces. This includes cell differentiation and migration, nuclear envelope maintenance in mechanically active tissues, the import of very large cargos such as viral capsids, metastasis, and changing osmotic conditions. Under such conditions, nuclear envelope membrane tension will change globally or locally. The resulting changes in the diameter of the NPC may have functional consequences that remain to be investigated further in the future. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: jan.kosinski@embl.de (J.K.); martin.beck@biophys.mpg.de (M.B.)
†These authors contributed equally to this work.
Cite this article as C. E. Zimmerli *et al.*, *Science* **374**, eabd9776 (2021). DOI: 10.1126/science.abd9776

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abd9776

RESEARCH ARTICLE

CELL BIOLOGY

Nuclear pores dilate and constrict in cellulo

Christian E. Zimmerli[1,2,3]†, Matteo Allegretti[1,3]†, Vasileios Rantos[4,5], Sara K. Goetz[1,2], Agnieszka Obarska-Kosinska[3,5], Ievgeniia Zagoriy[1], Aliaksandr Halavatyi[6], Gerhard Hummer[7,8], Julia Mahamid[1], Jan Kosinski[1,4,5]*, Martin Beck[1,3]*

In eukaryotic cells, nuclear pore complexes (NPCs) fuse the inner and outer nuclear membranes and mediate nucleocytoplasmic exchange. They are made of 30 different nucleoporins and form a cylindrical architecture around an aqueous central channel. This architecture is highly dynamic in space and time. Variations in NPC diameter have been reported, but the physiological circumstances and the molecular details remain unknown. Here, we combined cryo–electron tomography with integrative structural modeling to capture a molecular movie of the respective large-scale conformational changes in cellulo. Although NPCs of exponentially growing cells adopted a dilated conformation, they reversibly constricted upon cellular energy depletion or conditions of hypertonic osmotic stress. Our data point to a model where the nuclear envelope membrane tension is linked to the conformation of the NPC.

Nuclear pore complexes (NPCs) bridge the nuclear envelope (NE) and facilitate nucleocytoplasmic transport. Across the eukaryotic kingdom, ~30 different genes encode for NPC components, termed nucleoporins (Nups). Although specialized Nups have been identified in many species, extensive biochemical and structural studies have led to the consensus that the core scaffold inventory is conserved. It consists of several Nup subcomplexes that come together in multiple copies to form an assembly of eight asymmetric units, called spokes, that are arranged in a rotationally symmetric fashion (*1*). The Y-complex (also called the Nup107 complex) is the major component of the outer rings [the nuclear ring (NR) and the cytoplasmic ring (CR)], which are placed distally into the nuclear and cytoplasmic compartments. The inner ring complex scaffolds the inner ring [(IR) also called the spoke ring] that resides at the fusion plane of the nuclear membranes. It consists of the scaffold Nups 35, 155, 188, and 192 as well as the Nsp1 complex. The IR forms a central channel lined with phenylalanine-glycine (FG) repeats containing Nups that interact with cargo complexes. The Nup159 complex (also called the P-complex) asymmetrically associates with the Y-complex of the CR and mediates mRNA export. Despite these common features of quaternary structure, in situ structural biology studies have revealed that the higher-order assembly is variable across the eukaryotic kingdom (*1*, *2*).

NPC architecture is conformationally highly dynamic, and variations in NPC diameter have been observed in various species and using different methods (*3*–*7*). Dilated states have been observed in intact human cells (*3*, *8*, *9*), contrasting with the constricted state in semi-purified NPCs (*10*–*12*). It has been shown that dilation is part of the NPC assembly process (*13*, *14*). However, it remains controversial whether NPC dilation and constriction play a role during active nuclear transport (*15*) and whether the dilation is required to open up peripheral channels for the import of inner nuclear membrane (INM) proteins (*16*–*18*). It has been argued that the constricted state may be a result of purification (*4*, *8*). It is difficult to conceive that such large-scale conformational changes can occur on similar time scales as individual transport events (*19*, *20*), which would be the essence of a physical gate. Nevertheless, several cues that potentially could affect NPC diameter have been suggested, such as exposure to mechanical stress, mutated forms of importin β, varying Ca^{2+} concentrations, or exposure to hexanediol (*7*, *21*–*26*). The biological relevance of these cues remains elusive because the analysis of NPC architecture under physiological conditions is experimentally very challenging. These previous studies did not explore NPC dilation and constriction and its functional cause and consequences within intact cellular environments, nor did they structurally analyze the conformational changes of nuclear pores in high molecular detail.

Here, we demonstrate that *Schizosaccharomyces pombe* NPCs (SpNPCs) constrict in living cells under conditions of energy depletion (ED) or hyperosmotic shock (OS), which is concomitant with a reduction of NE membrane tension. Using in cellulo cryo–electron microscopy (cryo-EM) and integrative structural modeling, we captured a molecular movie of NPC constriction. Our dynamic structural model suggests large-scale conformational changes that occur by movements of the spokes with respect to each other but largely preserve the arrangement of individual subcomplexes. Previous structural models obtained from isolated NEs (*10*–*12*, *27*–*29*) thereby represent the most constricted NPC state.

In cellulo cryo-EM map of the SpNPC

To study NPC architecture and function in cellulo at the best possible resolution and structural preservation, we explored various genetically tractable model organisms for their compatibility with cryo–focused ion beam (cryo-FIB) specimen thinning, cryo–electron tomography (cryo-ET), and subtomogram averaging (STA). *Saccharomyces cerevisiae* cells were compatible with high-throughput generation of cryo-lamellae and acquisition of tomograms. STA of their NPCs resulted in moderately resolved structures (*4*). By contrast, a larger set of cryo-tomograms from *Chaetomium thermophilum* cells did not yield any meaningful averages, possibly because their NPCs displayed a very large structural variability. We therefore chose to work with *S. pombe* cells that are small enough for thorough vitrification; offer a superior geometry for FIB milling compared with *C. thermophilum*, with the advantage of covering multiple cells; and, compared with *S. cerevisiae*, have a higher number of NEs and NPCs per individual cryo-lamella and tomogram, leading to increased data throughput (fig. S1).

To obtain a high-quality cryo-EM map of SpNPCs, we prepared cryo-FIB–milled lamellae of exponentially growing *S. pombe* cells and acquired 178 tomograms from which we extracted 726 NPCs. Subsequent STA resulted in an in cellulo NPC average of very high quality in both visible features and resolution (Fig. 1 and figs. S2 and S3). Systematic fitting of the *S. pombe* IR asymmetric unit model, built based on the *S. cerevisiae* NPC, resulted in precisely one highly significant fit, as expected (figs. S4A and S5A and Materials and methods). The subsequent refinement with integrative modeling led to a structural model that explains most of the observed EM density in the IR (Fig. 1B, fig. S5B, and movie S1). The IR architecture appears reminiscent to NPC structures of other eukaryotes (fig. S6), further corroborating its evolutionary conservation (*2*, *30*) (see table S1 for nomenclature of Nups across different species). Systematic fitting revealed that the NR of the SpNPC is composed of two concentric Y-complex rings (Fig. 1A,

[1]Structural and Computational Biology Unit, European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, Germany. [2]Collaboration for joint PhD degree between EMBL and Heidelberg University, Faculty of Biosciences, 69120 Heidelberg, Germany. [3]Department of Molecular Sociology, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany. [4]Centre for Structural Systems Biology (CSSB), 22607 Hamburg, Germany. [5]EMBL Hamburg, 22607 Hamburg, Germany. [6]Advanced Light Microscopy Facility, EMBL, 69117 Heidelberg, Germany. [7]Department of Theoretical Biophysics, Max Planck Institute of Biophysics, 60438 Frankfurt am Main, Germany. [8]Institute of Biophysics, Goethe University Frankfurt, 60438 Frankfurt am Main, Germany.
*Corresponding author. Email: jan.kosinski@embl.de (J.K.); martin.beck@biophys.mpg.de (M.B.)
†These authors contributed equally to this work.

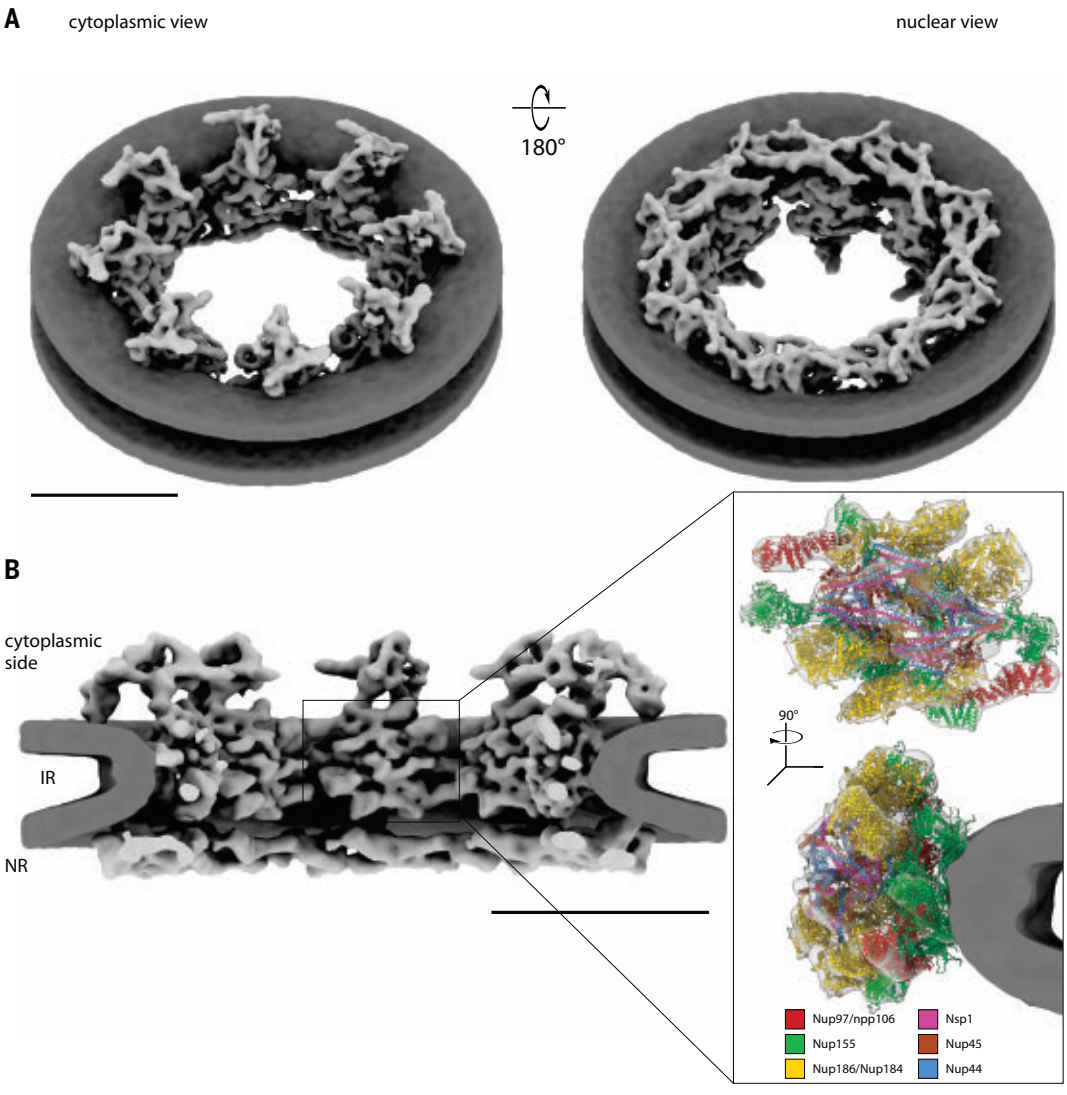


Fig. 1. In cellulo cryo-EM map of the SpNPC. (**A**) Isosurface rendered views of the SpNPC eightfold assembly as seen from the cytoplasm (left) and the nucleoplasm (right) (with membranes in dark gray and protein in light gray). Note, the eightfold symmetric assembly represents an artificial NPC conformation; within living cells, NPCs adopt more-flexible conformations deviating from a perfect circle. Although the cytoplasmic view (left) reveals eight disconnected protein entities instead of a CR, the nuclear view (right) shows two concentric nuclear Y-complex rings. Scale bar, 50 nm. (**B**) Same as (A) but shown as a cutaway view. Although the asymmetric curvature of the nuclear membranes and the arrangement of the cytoplasmic side is unprecedented in other species, the IR architecture is highly conserved, as highlighted in the inset (see also fig. S6). Scale bar, 50 nm. The inset shows one individual asymmetric IR unit map as obtained during STA (see Materials and methods); the density threshold is chosen at a level comparable to the entire eightfold assembly. In the side view (inset, below), the membrane is shown for reference. Fitting of the IR Nup homology models explains most of the observed electron optical density.

fig. S4B, and movie S1) as in vertebrates and algae but as opposed to the single ring observed in *S. cerevisiae* (fig. S3A) (*4*, *11*, *31*). Integrative modeling of the entire Y-complex in the NR revealed a Y-complex architecture with the typical head-to-tail oligomerization (Fig. 2A). It recapitulates the domain orientations and interactions known from other species. However, *S. pombe* Y-complex Nups do localize to the NR, in contrast to previous proposals (*32*).

Closer inspection of the cytoplasmic side of the cryo-EM map revealed an unanticipated architectural outline—it did not form an apparent ring. Instead, we observed eight spatially separated entities (Fig. 1A), which serve as anchor points for the mRNA export platform (*4*, *33*), the Nup159 complex (Fig. 2B). Although the dynein-arm characteristic for the *S. cerevisiae* NPC (*34*) is lacking, the Nup159 complex resembles its *S. cerevisiae* counterpart in shape (fig. S7A). Systematic fitting and subsequent refinement with integrative modeling revealed that only a single Y-complex vertex fits into the density observed at the cytoplasmic side (Fig. 2B, fig. S4C, and movie S1). The density potentially accounting for Nup107 and SpNup131-SpNup132 was missing (figs. S4D and S7B) and could not be recovered by local refinement (fig. S7C). To independently confirm the identity of the observed vertex-like density, we analyzed *nup37*Δ and *nup37*Δ-*ely5*Δ strains. The binding of both Nup37 and Ely5 to Nup120 has been shown previously in vitro (*35*, *36*), and, as expected, density was missing in the respective positions of all Y-complexes (Fig. 2, C and D, and figs. S8 and S9). A density that could accommodate the Ely5 homology model was missing also in the cytoplasmic Y-complex, suggesting that Ely5 is present in *S. pombe* at both sides, unlike its human counterpart Elys (*11*, *37*) (Fig. 2D and movie S1). Otherwise, the NPC architecture remained mostly unchanged, despite some increased flexibility in the Nup120 arm of the outer nuclear Y-complex pointing to a decreased NR stability under knockout conditions. These results independently identify the cytoplasmic structure as bona fide Y-complex vertex but lacking the EM density beyond the edge of SpNup189C. This is consistent with previous work suggesting a split of the SpNup189C-Nup107 interface (*32*, *38*). Alternatively, this observation may be explained by flexibility, although hinges within the Y-complex were identified at different locations (*10*, *39*).

ED is concomitant with a constriction of the central channel

Previous cryo-EM structures of NPCs obtained from isolated NEs (*10*–*12*, *27*–*29*) or by detergent extraction (*40*) had a smaller diameter compared with those obtained from intact cells (*3*, *4*, *31*, *41*). We therefore hypothesized that NPC diameter may depend on the biochemical energy level, which is diminished

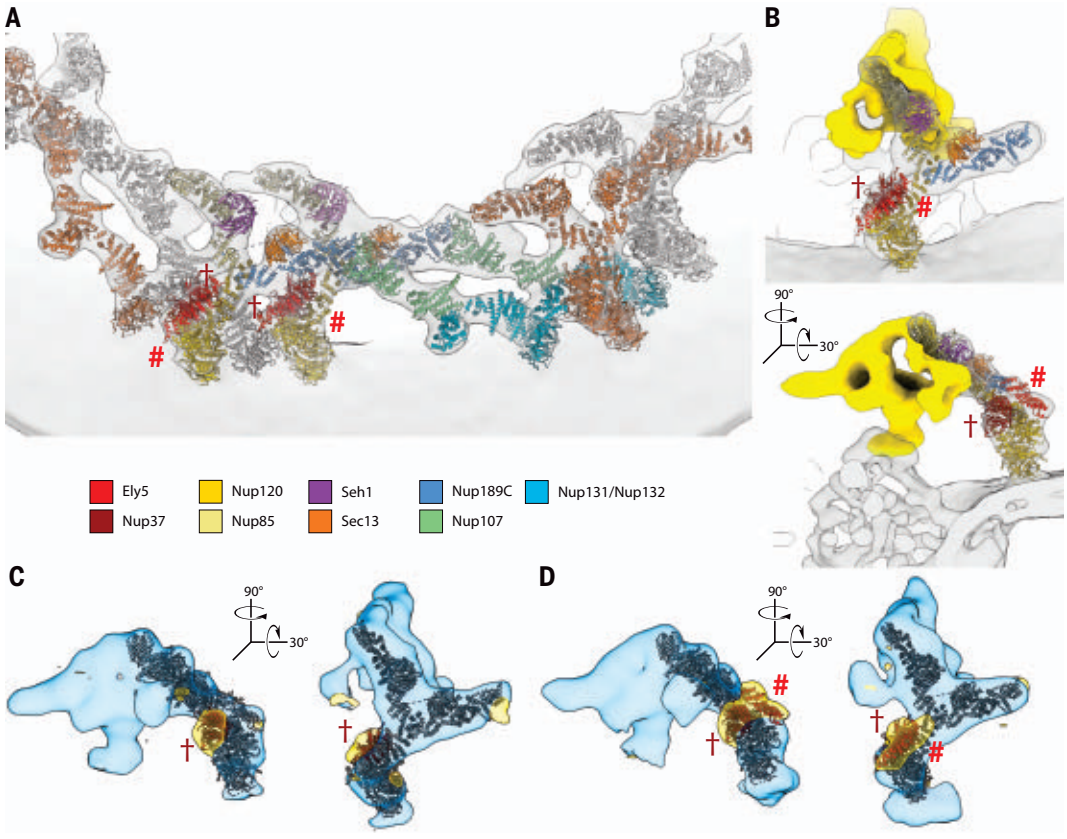


Fig. 2. Architecture of SpNPC outer rings. (**A**) Systematic fitting and integrative modeling of all *S. pombe* Y-complex Nups reveal a head-to-tail arrangement with two concentric Y-complex rings on the nuclear side of the SpNPC, as in the human NPC (*11*). The cryo-EM map of a NR segment is shown rendered as an isosurface in transparent light gray. The adjacent inner Y-complexes are shown in gray, and the outer Y-complexes are shown in orange. The homology models of SpNup131 and SpNup132 fit to the Y-complex tail region equally well, rendering these two proteins indistinguishable by our approach. (**B**) Integrative model of the cytoplasmic protein entities. The fit of the Y-complex vertex explains most of the observed density. The mRNA export platform as identified in (*4*, *33*) is segmented in yellow. (**C**) The *nup37*Δ cryo-EM map is shown in blue (threshold: 0.1) and overlaid with the difference map (yellow; threshold: 0.175) of the WT and *nup37*Δ maps, both filtered to 27 Å. The missing density in the long arm of the Y-vertex coincides with the position of Nup37 (indicated by the dark red dagger symbol) in the Y-complex vertex (dark gray) (see also fig. S8). (**D**) The *nup37*Δ-*ely5*Δ double knockout map (blue; threshold: 0.08) overlaid with the corresponding difference map (yellow; threshold: 0.188) comparing the WT and double knockout map at 34 Å. Differences are apparent at the location of both Nup37 (dark red dagger symbol) and Ely5 (light red number symbol) with respect to the fitted Y-complex model (dark gray) (see also fig. S9).

in preparations of isolated NEs or NPCs but may also be reduced within intact cells—e.g., during stress conditions. We first set out to analyze NPCs in *S. pombe* cells under conditions of ED, namely after 1 h of exposure to nonhydrolyzable 2-deoxy-glucose in combination with the respiratory chain inhibitor antimycin A, as previously established (*42*) (see Materials and methods). Recovery experiments demonstrated the viability of cells exposed to ED, and the subcellular architecture of the respective cells remained intact, as apparent by cryo-ET (fig. S10). Cells of various organisms including *S. pombe* show a rapid shut down of active nuclear transport and mRNA export when depleted of adenosine 5′-triphosphate (ATP) because of a concomitant reduction of guanosine 5′-triphosphate (GTP) levels, which leads to the loss of the nucleocytoplasmic RanGTP-RanGDP (GDP, guanosine diphosphate) gradient (*43*–*46*). To confirm a loss of active nucleocytoplasmic transport under ED conditions, we used live-cell imaging of *S. pombe* cells expressing a green fluorescent protein (GFP) variant tagged with a nuclear localization signal (NLS) or nuclear export signal (NES) on its N and C terminus (NLS-GFP-NLS or NES-GFP-NES), which show a nearly exclusive nuclear or cytoplasmic localization under control conditions (fig. S11, A to D). After 30 min of ED, most of the NLS-GFP-NLS localized to the cytoplasm, whereas NES-GFP-NES equilibrated into the nucleus (fig. S11, A to D). This was reverted in the vast majority of cells when recovered in glucose control medium (fig. S11, E to G), thus underlining their viability.

We analyzed 292 NPCs structurally by STA under these conditions and found a considerable constriction of the central channel diameter. However, the averages appeared blurred, and manual inspection of the data indicated a large variation of diameters. To generate a conformationally more-homogeneous ensemble, we manually assigned the NPCs from the ED dataset into two classes with central channel diameters of <50 nm and >50 nm and refined them separately (corresponding to 533 and 1012 subunits, respectively) (Fig. 3 and fig. S12, A to C). Both conformations of the ED state showed a smaller NPC diameter compared with the conformation observed in NPCs of exponentially growing cells, from here on referred to as control conditions. The intermediate conformation was ~65-nm wide at the IR compared with ~70 nm under control conditions, whereas the most-constricted conformation showed a diameter of ~50 nm, comparable to the diameter observed in isolated NEs (*10*–*12*, *27*–*29*).

To better understand how NPCs accommodate such massive conformational changes on the molecular level, we systematically fitted individual subcomplexes (fig. S13) and built structural models of the three different diameter states on the basis of the cryo-EM maps (fig. S14) using integrative modeling (fig. S5B). Both systematic fitting and integrative modeling showed that the relevant subunits are present in the control state and the ED states. At the cytoplasmic side and the NR, changes in the curvature of the Y-complexes and inward bending of the mRNA export platform toward the center of the pore were apparent (movies S4 to S6). By contrast, the central channel constriction of the IR is more elaborate and mediated by a lateral displacement of the eight spokes that move as independent entities to constrict or dilate the IR (fig. S14 and movies S4 to S6). In the control state, ~3- to 4-nm-wide gaps are formed in-between the neighboring

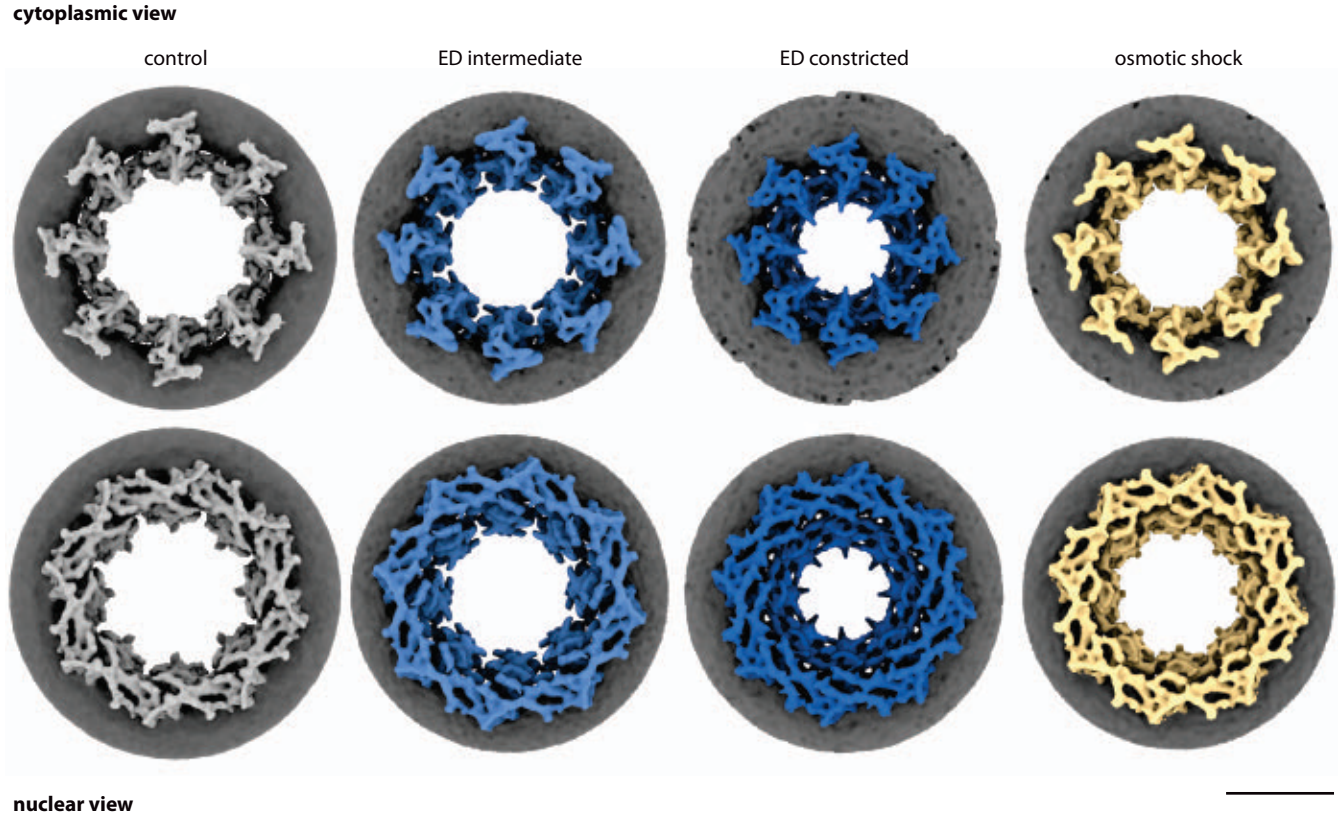


Fig. 3. NPCs constrict during ED and OS. Cytoplasmic and nuclear views of NPC conformations during control, intermediate, and fully constricted ED and OS conditions, illustrating the overall conformational change leading to a central channel diameter constriction from ~70 to <50 nm. The cytoplasmic and IR spokes move as individual entities and contribute the most to the central channel diameter change, whereas the NR constricts to a lesser extent. Scale bar, 50 nm.

spokes. By contrast, in the constricted state, the spokes form extensive contacts (Fig. 4, A and B) equivalent to those in the previously published structures of the human NPC in isolated NEs (*12*). It is plausible that such opening and closing of peripheral gaps may regulate the translocation of INM proteins (*16*–*18*). Notably, under conditions of ED, additional density is arching out into the lumen of the NE (Fig. 4, A, B, and D), contrasting control conditions. It has been previously proposed that such luminal structures are formed by Pom152 (ScPom152 or HsGP210) (*47*, *48*). In terms of their shape, the observed arches are reminiscent of those observed in isolated *Xenopus laevis* NEs (*28*). Our data imply that the luminal ring conformation becomes more prominent upon constriction. Whether this has any mechanical benefits to keep NPCs separated (*28*), or limit the maximal dilation or constriction, remains unclear.

Hyperosmotic shock results in constriction of the central channel

We noticed that ED was concomitant with a substantial reduction of nuclear size and speculated that it could be functionally related to NPC constriction. To test this hypothesis, we triggered nuclear shrinkage independently from ED. We caused a OS by exposing cells to 1.2 M sorbitol. This treatment leads to a pronounced loss of cellular and nuclear volume (*49*). A cryo-EM map from 200 NPCs (Fig. 3; Fig. 4, C and D; and fig. S15A) revealed a strongly constricted conformation. Individual structural features, such as tight lateral packing of IR segments and arching out of the luminal domain, were very similar to those observed with ED, with one notable exception: The N-terminal domain of Nup107 was now clearly resolved at the cytoplasmic side, with the expected angle of engagement of the Nup189C-Nup107 interface (fig. S15B) and consistent with the previously described hinge of the Y-complex (*39*). This observation may be interpreted as either compositional rearrangement or reduced flexibility. Notably, our findings emphasize that nuclear shrinkage correlates with NPC constriction independent of ED.

Passive nucleocytoplasmic transport is reduced under ED and OS conditions

To assess whether the NPC constriction under OS and ED conditions correlates with reduction of passive transport across the NE, we performed fluorescence recovery after photobleaching (FRAP) experiments of nuclei at different time points after ED in cells expressing freely diffusing GFP, as compared with control conditions (Fig. 5, A and B, and fig. S16, A and B) (see Materials and methods). GFP diffusion rates into the nucleus were significantly decreased upon ED (Fig. 5C), contrasting a minor, negligible effect observed within the cytoplasm (fig. S16C). Recovery experiments restored the initially observed diffusion rates (fig. S16D). Under OS conditions, nucleocytoplasmic diffusion decreased to a much lesser extent (Fig. 5, D to F). In contrast to ED, cytoplasmic diffusion also decreased (fig. S16, E to I) in line with previous reports, which demonstrate that OS strongly increases cellular crowding (*42*, *50*). Because ED also leads to a considerable cellular reorganization, including solidification of the cytoplasm, pH-shift, and water loss (*42*, *51*–*53*), it is difficult to disentangle whether the observed reduction in passive exchange rates is a direct consequence of NPC constriction. How exactly NPC constriction affects the passive and active transport of different types of cargos remains to be determined.

NE tension regulates NPC diameter

Scaffold Nups do not have any known motor activity that would allow them to exert forces.

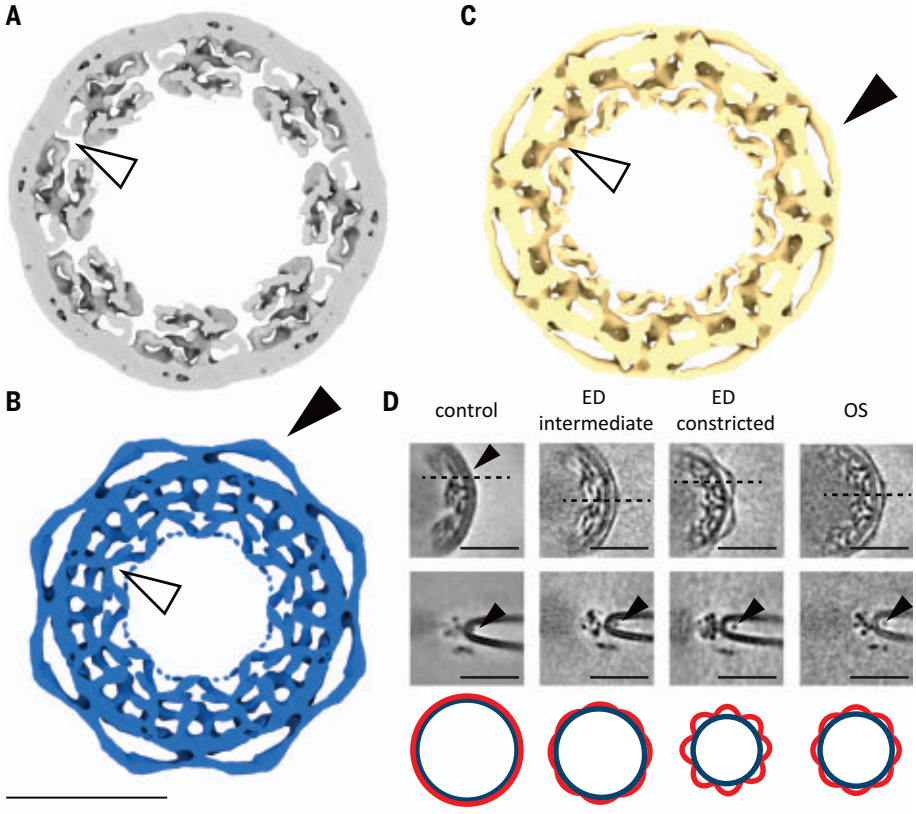


Fig. 4. Conformational changes of the luminal ring and IR during NPC constriction. (**A** to **C**) Central slice through an isosurface representation of NPCs under control (A), most-constricted ED (B), and OS (C) conditions. Under control conditions, IR spokes are clearly separated by 3- to 6-nm gaps and form extensive contacts during NPC constriction (white arrowheads). Luminal densities (black arrowheads) are arching into the NE lumen during NPC constriction. Scale bar, 50 nm. (**D**) Horizontal slices through the IR of cryo-EM maps of control, intermediate, and most-constricted ED and OS conditions (left to right). NPCs reveal the appearance of an arch-like density (black arrowheads) during constriction, likely corresponding to the C-terminal domains of Pom152, whereas the N terminus is attached to the NPC (*48*), which is consistent with previous work on isolated NEs (*28*). Dashed lines indicate the position of the orthogonal slices (below) with the luminal density marked by arrowheads. The cartoons (bottom) indicate how a constriction of the nuclear pore membrane (blue) leads to an arching out of Pom152 C-terminal luminal domains (red), whereas the N termini remain closely attached to the NE. Scale bar, 50 nm; slice thickness, 1.35 to 1.38 nm. In control conditions, the luminal densities are directly proximal to the NPC and thus not easily discernible from the membrane.

Thus, it appears more likely that NPC diameter changes result from mechanical forces that are applied externally. Membrane tension, as the energetic cost to expand the membrane area, results in lateral forces within the NE. We hypothesized that the rigidity of the NPC scaffold counteracts lateral forces imposed by NE tension. In such a scenario, an excess of membrane induced by nuclear shrinkage during OS and ED conditions would reduce tension and result in NPC constriction. This would be consistent with the fact that the most constricted conformation of NPCs is observed in isolated NEs where no mechanical forces act on the scaffold (*10*–*12*, *27*–*29*, *40*).

A model in which membrane tension pulls apart the NPC scaffold to dilate the central channel makes several predictions: (i) Perturbation of the structural integrity of the NPC scaffold should cause NPC dilation; (ii) the more rigid parts of the *S. pombe* scaffold, such as the NR, should be less responsive to lateral forces and constrict less compared with the more fragile part, namely the IR and cytoplasmic side, which in *S. pombe* does not form a rigid ring; (iii) nuclear shrinkage and deformation that are indicative of reduced NE tension (*54*, *55*) should coincide with NPC constriction; (iv) a reduced distance between the INM and outer nuclear membrane (ONM) should directly correlate with NPC constriction; and (v) recovery of nuclear volume, shape, and INM-ONM distance should restore the initial, dilated NPC diameters. Regarding the latter two predictions, it had been proposed that a reduced distance between the INM and ONM is indicative of reduced pressure within the NE lumen (*56*). Using membrane elastic theory, we show that osmotic pressure in the NE lumen and membrane tension in the NE act equivalently (supplementary text) and that the INM-ONM distance grows linearly with both.

To test predictions (i) and (ii), we systematically measured the NPC diameter across all investigated conditions based on centroids of opposite asymmetric subunits as obtained by STA (Fig. 6A and fig. S17, A to F) (see Materials and methods). Knockouts of the scaffold Nups Ely5 and Nup37 caused a dilation of the scaffold (Fig. 6A). We quantified the diameters of each ring under NPC-constricting conditions and found that although all three rings constrict significantly, the conformational changes are considerably smaller at the NR that has a more elaborate scaffold as opposed to the cytoplasmic side that does not form a closed ring (Fig. 6B and movies S4 to S6). These findings are consistent with our model.

To test a relationship of nuclear shrinkage and deformation with NPC diameter [prediction (iii)], we quantified nuclear volume and shape on the basis of three-dimensional (3D) reconstructions of nuclear membranes from confocal fluorescence light microscopy z-stacks and segmentation of nuclear membranes in cryo–electron tomograms for ED- and OS-treated cells. The nuclear volume was reduced in both cases and was reverted when cells were recovered from ED (Fig. 7, A to D, and fig. S18A). Nuclear shrinkage should lead to an excess of nuclear membranes if the total nuclear surface area remains constant, which we verified by quantification of the number of NPCs per NE surface in electron tomograms (Fig. 7E). Under conditions of OS, nuclei were also considerably deformed (Fig. 7, B, F, and G, and fig. S18, B and C). Thus, nuclear shrinkage is concomitant with NPC constriction in both OS and ED conditions (Fig. 7H). Nuclear deformation coincides with NPC constriction during OS but is less apparent during ED (Fig. 7, A and I, and fig. S18, C and D), which we attribute to the missing energy-dependent nuclear deformation forces derived from microtubules (*57*, *58*). Alternatively, the more consistent response of cells to OS compared with ED conditions may explain this effect.

Next, we investigated a relationship between the distance between the INM and ONM and the NPC diameter [prediction (iv); supplementary text]. Under both ED and OS conditions, the average INM-ONM distance was reduced from 22.6 nm in control cells to 15.8 and 19.5 nm, respectively, and reverted to control levels upon recovery from OS, further underlining a loss of NE tension (*56*, *59*) (Fig. 7J) and in agreement with our theoretical model (supplementary text). On the level

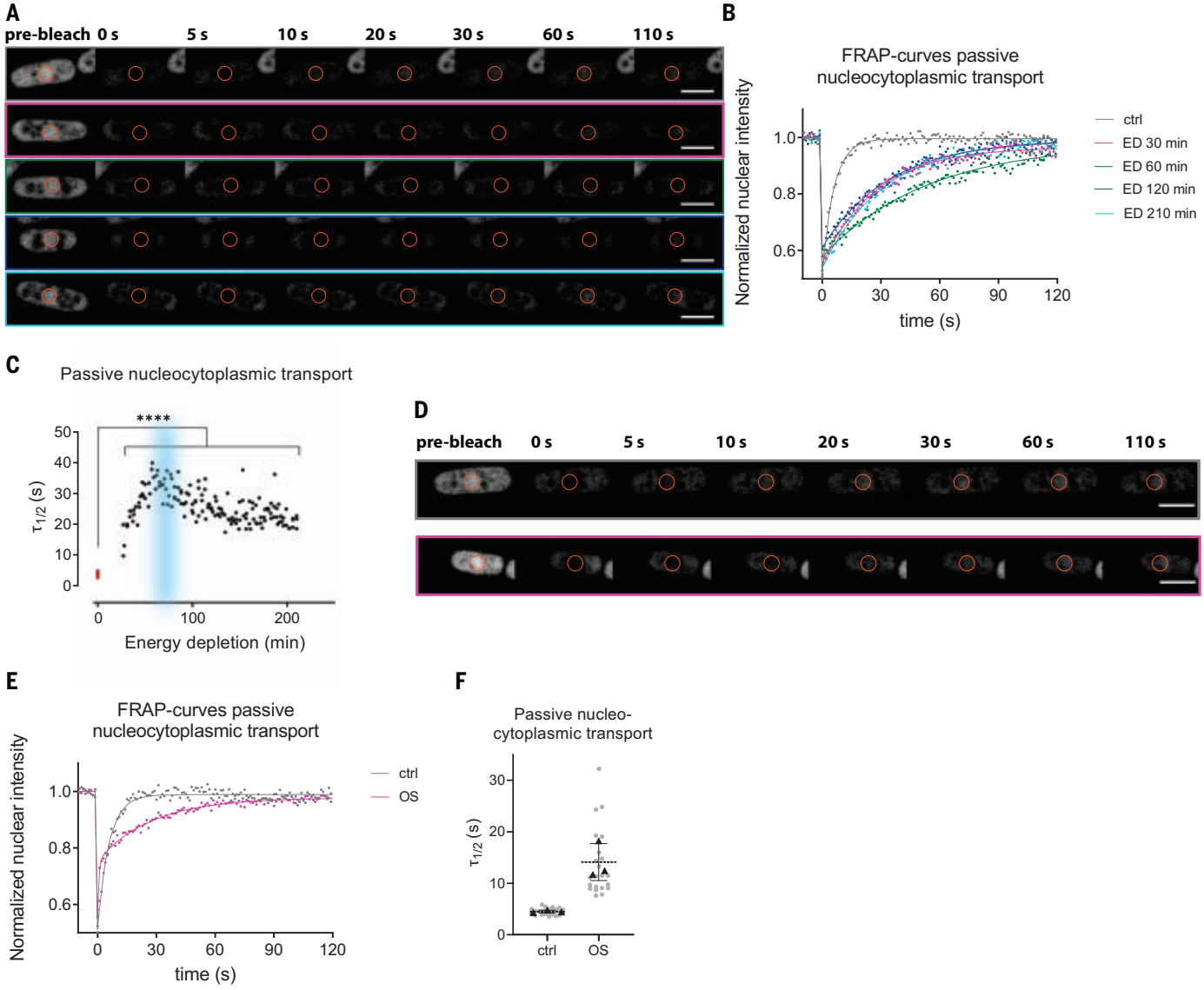


Fig. 5. FRAP experiments and quantification of passive nucleocytoplasmic transport during ED and OS. (**A**) Representative FRAP images from time series of passive nucleocytoplasmic diffusion acquired during control conditions (gray) and at 30 min (purple), 60 min (green), 120 min (dark blue), and 210 min (cyan) after ED (from top). Scale bars, 5 μm. (**B**) FRAP curves corresponding to time series in (A). ctrl, control. (**C**) FRAP recovery half-life times of nuclear signal from freely diffusing GFP at various time points during ED are significantly longer compared with control conditions (red dots). Passive transport of free GFP reaches a minimum after ~1 hour of ED and subsequently recovers slightly. The blue area shows the time point at which cryo-EM grids were prepared for structural analysis of ED NPCs. Data are from three experiments, and triangles indicate means of individual replicates (n = 24 cells in control and n = 154 cells in ED time series). Unpaired t test; ****P < 0.0001. (**D** and **E**) Representative snapshots of passive nucleocytoplasmic transport in control (top) and OS conditions (bottom) (D) and corresponding FRAP curves (E). The orange circles in (D) indicate the bleached area during FRAP experiments, and the time after bleaching is indicated on top. Scale bars, 5 μm. (**F**) Nuclear GFP signal shows a strong increase in recovery half-life time upon OS. Data are from three experiments, and triangles indicate means of individual replicates (n = cells 24 in control and n = 22 cells in OS).

of individual tomograms, the average NPC diameter correlated strongly with the average INM-ONM distance (Spearman correlation coefficient = 0.4803) (Fig. 7K).

Finally, we established that NPC constriction is reversible [prediction (v)]. Because OS conditions lead to a strong and homogeneous constriction of NPC diameters, we performed an OS recovery experiment. We analyzed 278 NPCs structurally by STA from cells that were previously exposed to a OS and shifted back to glucose control medium. Measurements of individual NPC diameters confirmed NPC diameter reversibility (Fig. 6A). As expected, the INM-ONM distance and nuclear volume recovered concomitantly, whereas the nuclear membrane surface–to-NPC ratio remained constant (Fig. 7, E, J, and K, and fig. S18D). These findings further underscore a model in which increasing NE tension causes NPC dilation, whereas a reduction of NE tension allows the NPC scaffold to relax and constrict (Fig. 8).

Conclusions

Our study reveals that NPCs within living cells populate a much larger conformational space than previously anticipated. Massive conformational changes mediate NPC constriction and dilation in response to physiological cues,

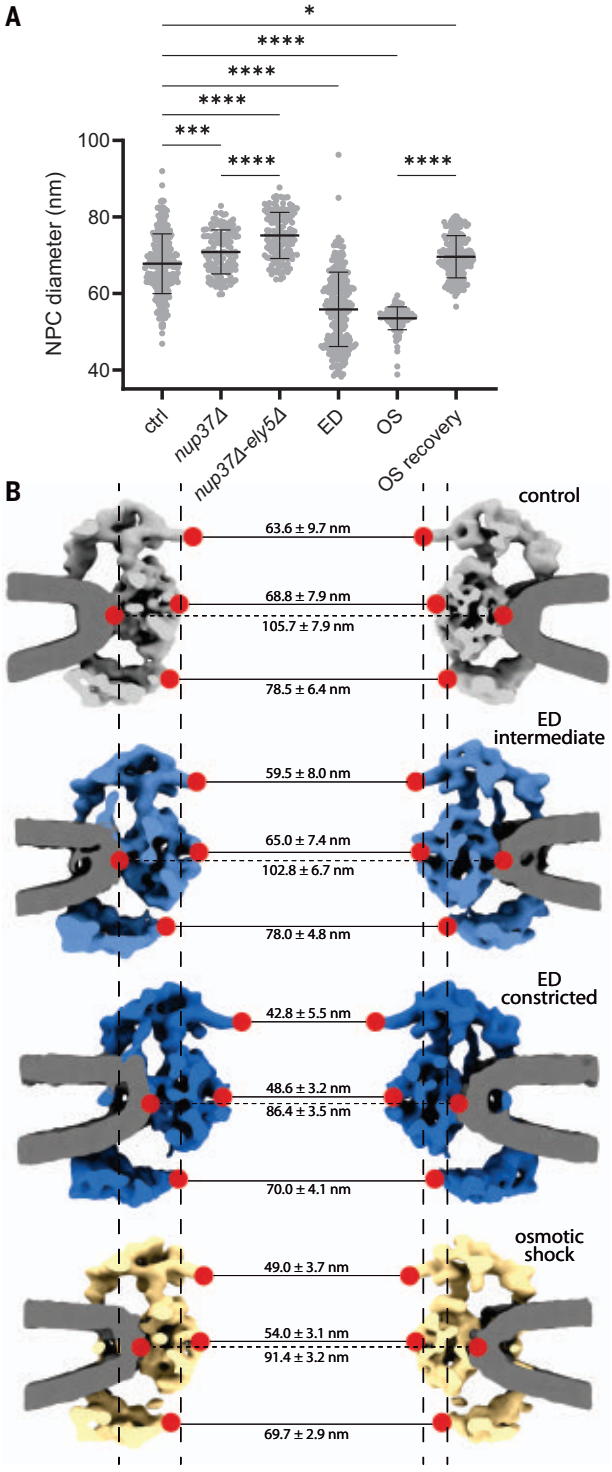


Fig. 6. NPC diameter measurements across different conditions. (**A**) Analysis of NPC central channel diameters measured at the equatorial center of the IR subcomplex based on subunit positions obtained by STA under control, *nup37*Δ, *nup37*Δ-*ely5*Δ, ED, OS, and OS recovery conditions. Data are from one or more experiments [means ± SDs; n = 270 NPCs (control), n = 129 NPCs (*nup37*Δ), n = 141 NPCs (*nup37*Δ-*ely5*Δ), n = 271 NPCs (ED), n = 141 NPCs (OS), and n = 197 NPCs (OS recovery)]. One-way analysis of variance (ANOVA) and Šidák's multiple comparison test; ****P < 0.0001; ***P < 0.001; *P < 0.05 (see also fig. S17). (**B**) Same as Fig. 3 but shown as cutaway side view and overlaid with diameter measurements of individually aligned rings. Measurements are taken at the INM-ONM fusion points (dashed horizontal lines) and the most centrally exposed scaffold points (solid horizontal lines). Positions outlining the central channel were chosen corresponding to the tip of the mRNA export platform in the cytoplasmic side, the equatorial center of the IR subcomplex, and the tip of the inner nuclear Nup85 arm. The vertical lines represent the positions of the respective measurement points in the control conformation. Data are from one or more experiments (means ± SDs; n = 270 NPCs in control, n = 136 NPCs in intermediate ED, n = 68 NPCs in most constricted ED, and n = 141 NPCs in OS).

such as the energy status of the cell or exposure to OS. Notably, ED resulted in a large heterogeneity of NPC constriction levels and two subpopulations with intermediate and strongly constricted NPC diameters. The average diameter observed under OS conditions was more defined at ~55 nm, with a relatively small standard deviation (Fig. 6A). This finding is consistent with the idea that OS affects membrane tension more directly than ED.

The observed conformational changes involve stretching the Y-complexes but are particularly pronounced at the IR, where the spokes move alongside the fused INM and ONM. Notably, the spokes do not move entirely as rigid bodies, but some conformational changes occur within the Nup155 and Nsp1 complex regions (movies S7 and S8). Those are, however, distinct from the previously proposed conformational sliding (*15*) and consistent with an overall preserved intrasubcomplex arrangement (*19*, *20*).

It has been suggested that mechanical forces directly applied to the NE could regulate transport of mechanosensitive cargo and alter NPC diameters (*25*, *60*, *61*). Our data point to a model where NE tension regulates NPC diameter (Fig. 8). During OS and ED, NPC constriction is dependent on nuclear shrinkage and a reduction in INM-ONM distance, which both indicate a reduced NE tension (Fig. 7). Furthermore, the conformational changes of the NPC scaffold appear to be dictated by the movement of the nuclear membranes. During dilation and constriction, the IR spokes move strictly together with the fused INM and ONM. Finally, the knockout of the nonessential Y-complex members Nup37 and Ely5 leads to an increased flexibility and NPC dilation compared with control cells (Fig. 6A), suggesting that the NPC counteracts lateral forces imposed by NE tension, which keep it in an open conformation. Thus, the loss of NE tension leads to a relaxation and constriction in both ED and OS conditions.

The unexpected split Y-complex arrangement of *S. pombe* breaks the long-standing dogma of a three-ringed architecture. This differential cytoplasmic and nuclear architecture, however, helps in assessing how the double–Y-complex arrangement contributes to the mechanical robustness of NPC architecture. The double–Y-complex ring arrangement of the NR constricts much less than the IR and the cytoplasmic structures. Knockouts of Y-complex components within the head-to-tail contact region destabilize the NR and lead to NPC dilation. These findings suggest that the Y-complex ring arrangement provides rigidity to the overall cylindrical architecture. It thus appears plausible that mechanical stress load on the NE—e.g., during cell migration or in osmotically variable environments that may result in unfavorable NE rupture events—imposed an evolutionary selection pressure for higher numbers of Y-complexes and ring formation in mammals or algae. By contrast, organisms less exposed to mechanical NE stresses, such as *S. cerevisiae* or *S. pombe*, reduced their Y-complex copy numbers and head-to-tail contacts during evolution. The NPC scaffold may have a wider range of functions beyond providing grafting sites for FG-Nups. It conformationally responds to mechanical cues and thus may have stabilizing but also mechanically sensory functions (*25*, *60*, *61*).

Materials and methods

S. pombe *culture and cryo-EM grid preparation*

Frozen stocks of *S. pombe* cells were freshly thawed and maintained on YES-agar plates

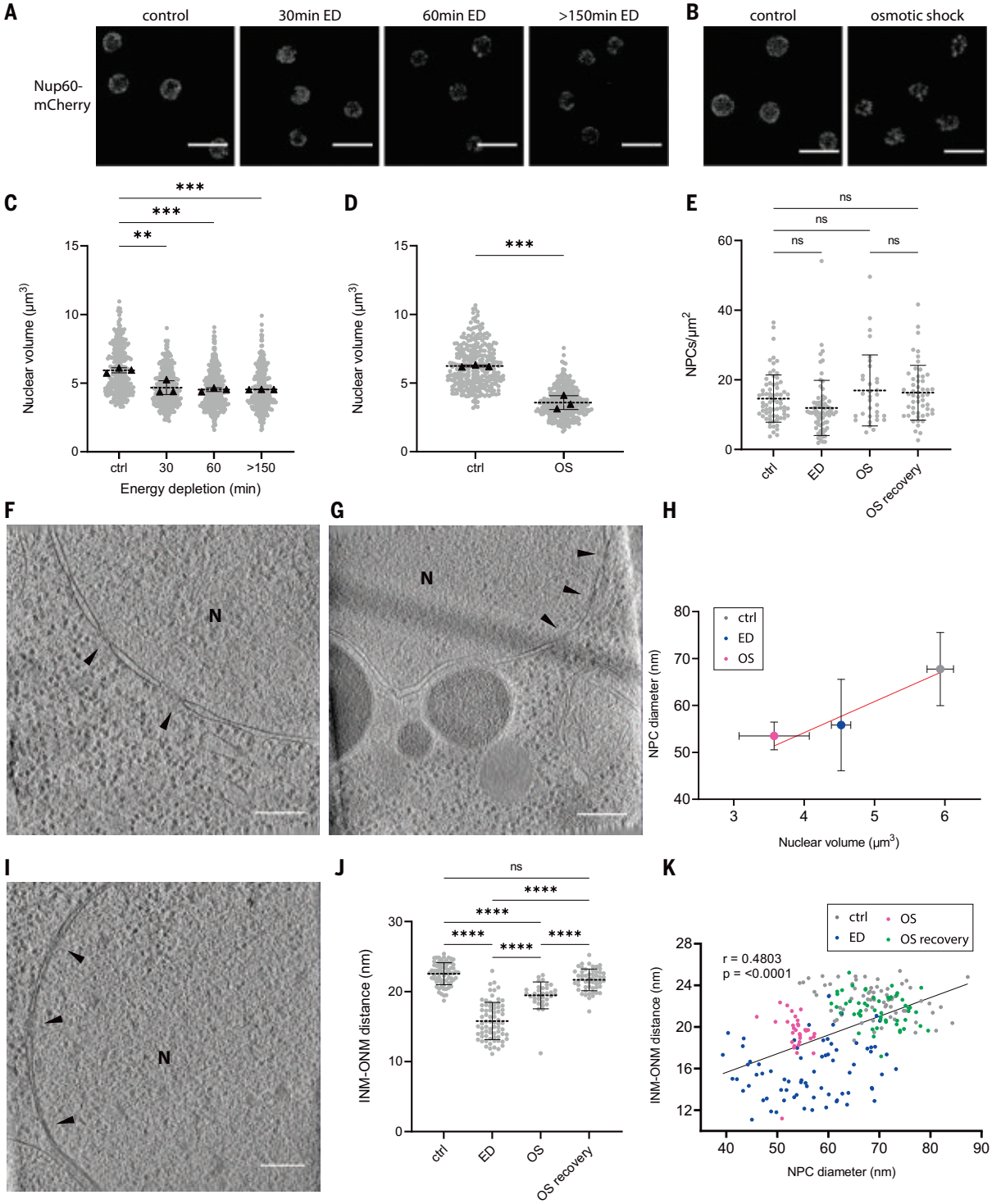


Fig. 7. NE tension is reduced during ED and OS and leads to NPC constriction. (**A**) Maximum projections of Nup60-mCherry channel from live-cell imaging confocal stacks of NLS-GFP-NLS and Nup60-mCherry–expressing cells used for nuclear shape and volume analysis. Scale bar, 5 μm. (**B**) Same as (A) but for cells expressing ubiquitous sfGFP and Nup60-mCherry under OS conditions. (**C** and **D**) Quantification of nuclear volume based on 3D reconstructions of Nup60-mCherry signal corresponding to (A) and (B). Data are from three experiments, and triangles indicate means of individual replicates (means ± SDs; n = 300 nuclei per condition under ED and n = 300 nuclei per condition under OS conditions). One-way ANOVA and Dunnett's multiple comparison test; ***P < 0.001; **P < 0.01. (**E**) Number of NPCs per estimated NE surface area under control, ED, OS, and OS recovery conditions (see Materials and methods). Data are from one or more experiments (means ± SDs; n = 71 tomograms under control, n = 72 tomograms under ED, n = 33 tomograms under OS, and n = 53 tomograms under OS recovery conditions). One-way ANOVA and Šidák's multiple comparison test; ns indicates not significant, P > 0.05. (**F** and **G**) Deconvolved slice through representative tomograms under control (F) and OS conditions (G). N marks the nucleus, and arrowheads mark NPCs. Scale bars, 200 nm. (**H**) Mean NPC diameter (see also Fig. 6A) plotted against mean nuclear volume [same as (C) and (D)] under and control and 60 min after ED and OS conditions. Linear regression

(red line) of mean nuclear volume values against scattered distribution of NPC diameters with means ± SDs and *n* values from (Fig. 6A) indicates a correlation between nuclear volume and NPC diameter. Data are from one or more experiments [coefficient of determination (R^2) = 0.3598]. (**I**) Deconvolved slice through representative tomograms under ED conditions. N marks the nucleus, and arrowheads mark NPCs. Scale bar, 200 nm. (**J**) Measurement of median INM-ONM distance per tomogram under control, ED, and OS conditions shows a clear decrease of INM-ONM distance during ED and a slight decrease during OS. Data are from one or more experiments (means ± SDs; *n* = 71 tomograms under control, *n* = 72 tomograms under ED, *n* = 33 tomograms under OS, and *n* = 53 tomograms under OS recovery conditions). One-way ANOVA with Tukey's multiple comparison test; ****$P < 0.0001$; ns, $P > 0.05$. (**K**) The mean NPC diameter per tomogram plotted against the median INM-ONM distance per tomogram shows a strong positive correlation. Data are from one or more experiments (*n* = 71 tomograms under control, *n* = 72 tomograms under ED, *n* = 33 tomograms under OS, and *n* = 53 tomograms under OS recovery conditions). Spearman correlation coefficient (*r*) and *P* values are plotted, and the black line represents a simple linear regression line.

(YES-broth from Formedium +20 g agarose/L) at 30°C for maximum 3 days and re-streaked on fresh YES-plates before liquid culture inoculation. Five to 10 ml YES-broth were inoculated with individual colonies and incubated over night at 30°C shaking at 200 rpm, where OD_{600} (the optical density of a sample measured at a wavelength of 600 nm) was kept below 1.0. Cryo-EM grid preparation was performed with a Leica EM GP plunger with a set chamber humidity of >90% and temperature of 30°C. A volume of 3.5 to 4 μl cell suspension with previously adjusted OD_{600} to 0.3 were applied to Cu200 or Au200 mesh R2/1 SiO_2 grids (Quantifoil) glow discharged on each side for 45 s before sample application. The grids were blotted with Whatman 597 paper for 1 to 2 s and plunge-frozen in liquid ethane or a mix of ethane-propane (60 and 40%) at −189°C or −194°C, respectively. A list of all yeast strains and their source used in this study can be found in table S2.

C. thermophilum *culture and cryo-EM grid preparation*

C. thermophilum (La Touche) *var. thermophilum* [from DSMZ, Braunschweig, Germany (no. 1495)] mycelium was maintained on CCM-plates, as described in (*62*). For cryo-EM grid preparation, Au200 mesh R2/1 SiO_2 grids (Quantifoil) were glow discharged on both sides for 45 s and placed on the CCM-plates at the edge of the mycelium growth rim with the SiO_2 foil facing up. Plates were incubated at 55°C until the mycelium covered at least half of the grid, typically 2 to ~3.5 hours. Grids were then mounted in the chamber of a Leica EM GP plunger with a set chamber humidity and temperature at >90% and 55°C, respectively. Before plunge freezing in liquid ethane at −186°C, 3.5 μl plunging phosphate buffer (5% trehalose, 0.013 M Na_2HPO_4, 0.045 M KH_2PO_4, pH 6.5) were applied to the sample and blotted for 2 s with Whatman 597 paper.

Generation of knockout and fluorescently tagged S. pombe *strains*

A *nup37*Δ knockout cassette containing a clonNat resistance marker corresponding to pFA6a-natMX6 (*63*), flanked by 70-basepair-long homologous sequences to the nup37 gene was synthesized by Geneart and transformed in a K972 h- wild-type (WT) *S. pombe* strain after a modified protocol from (*64*).

The *nup37*Δ-*ely5*Δ strain was generated by crossing the *nup37*Δ strain with an *ely5*Δ strain (*36*), provided by the Schwartz laboratory at MIT in Cambridge, Massachusetts, USA. Sporulation was triggered on sporulation plates prepared from SpoVB salts containing 8.2 g NaAc, 1.9 g KCl, 2.9 ml $MgSO_4$ 1 M, and 4.1 ml 5 M NaCl and 20 g agarose in 1 L H_2O. Tetrads of sporulated colonies were dissected in 40 μl sorbitol 1 M with 10 μl zymolyase-T100. Selected colonies were replica plated on selection plates containing 100 μg/ml clonNat (Jena Bioscience) or 100 μg/ml G418 (Sigma-Aldrich) and double gene knockout was confirmed by polymerase chain reaction (PCR).

To generate the strains CZ001 [expressing ubiquitous free super-folder GFP (sfGFP) with a Nup60-mCherry marker for FRAP analysis] and CZ007 (expressing NES-sfGFP-NES construct with a Nup60-mCherry marker), a Nup60-mCherry tagging cassette was amplified by PCR from isolated genomic DNA of the GD250 strain (*65*), provided by the Baum laboratory at MRC LMCB London, UK. The product was transformed into the sfGFP or the NES-sfGFP-NES expressing strain (AV0890 and AV1201, respectively) (*66*) obtained from the Yeast Genetic Resource Center (YGRC), as described above.

ED of S. pombe *cells*

Overnight cultures of *S. pombe* cells were grown at 30°C in 5 ml dextrose-free Edinburgh minimal medium (EMM) supplemented with 20 mM glucose (glucose-control medium) while cell density was kept below OD_{600} 0.8. Cells were collected and adjusted to obtain a cell amount of ~1 ml of cells with OD_{600} ~0.3 by centrifugation in Eppendorf tubes at 4000 rpm for 3 min at room temperature. Cells were then washed three times with 1 ml of ED buffer [20 mM 2-deoxy-glucose (Sigma-Aldrich) in dextrose free EMM and 10 μM antimycin A (Sigma-Aldrich)] and resuspended in 1 ml of ED buffer. Control samples were washed three times in glucose control medium. For EM-grid preparation cells were incubated for 1 hour at 30°C on a table-top shaker at 900 rpm. The cell concentration was adjusted to an OD_{600} of 0.2 before EM-grid preparation, as described above.

ED recovery experiments and colony counting viability assays

For recovery experiments, cells were energy depleted as described above and shifted back to glucose control medium ~1 hour after ED. Light microscopy experiments were performed ~20 min after shift back to energy rich medium.

For colony counting, three biological replicates of WT cells were cultivated and energy depleted for 1 hour, as described above. The corresponding control samples were washed in glucose control medium. All samples were then washed and taken up in glucose control medium and 1000 cells were plated on fresh YES plates in duplicates. Colonies were counted after incubation at 30°C for 3 days, and values were reported as mean survivability corrected for plating efficiency in control conditions. For spot assays replicates were prepared as described above and the concentration of each sample was adjusted to OD_{600} = 0.1. A serial dilution with 10x dilution steps was then spotted on a fresh YES plate.

OS of S. pombe *cells*

OS experiments were performed with *S. pombe* cells expressing ubiquitous sfGFP and mCherry-tagged Nup60 (strain CZ001). Cells were grown at 30°C in 5 ml dextrose-free Edinburgh minimal medium (EMM) supplemented with 20 mM glucose (glucose-control medium) while cell density was kept below OD_{600} 0.8. Cells for plunge freezing were grown to OD_{600} of 0.6 and concentrated 2x in 1.5 ml Eppendorf tubes by centrifugation at 4000 rpm for 3 min before OS, whereas cells for light microscopy analysis were grown to OD_{600} 0.4 before OS exposure. The OS was induced by diluting the cell suspension 1:1 with glucose control medium containing 2.4 M sorbitol, resulting in a OS exposure to 1.2 M sorbitol. Before plunge freezing cells were checked by fluorescent light microscopy to confirm the OS reaction and cryo-EM grids were subsequently prepared as described above 1 to 12 min after OS exposure. Light microscopy experiments were performed as described below. In OS recovery experiments cells were treated with 1.2 M sorbitol for ~10 min as described above and subsequently shifted back to glucose control medium where they were allowed to equilibrate for 30 to

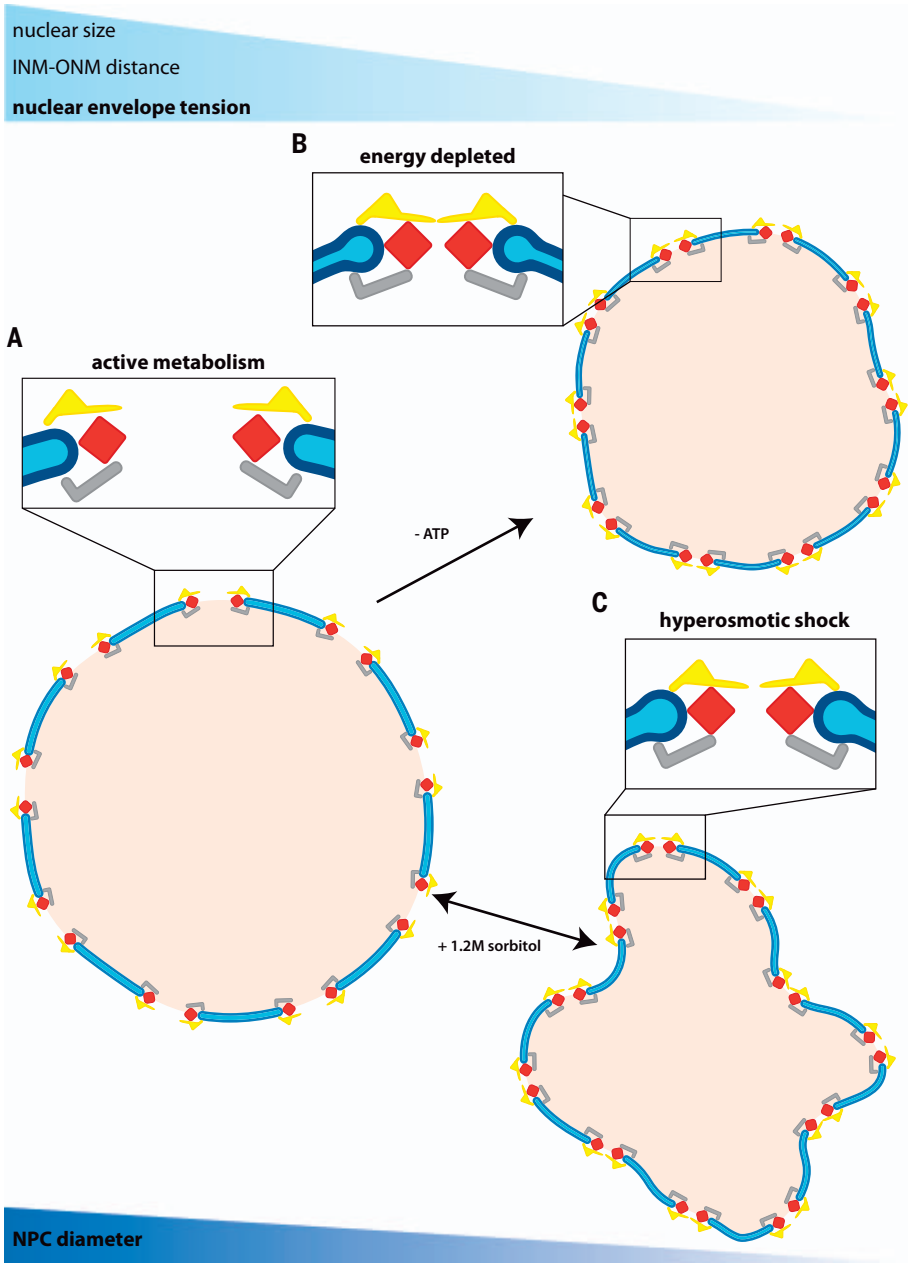


Fig. 8. Conceptual model of how of NE tension regulates NPC diameter. (**A**) NPCs adopt a dilated conformation in metabolically active cells. (**B** and **C**) Both ED (B) and OS (C) lead to a shrinkage of nuclear volume, a reduced INM-ONM distance, and an increased nuclear deformation, all of which point to a reduction in NE tension under both conditions. Such reduced NE tension leads to a constriction of the NPC scaffold architecture, where the central channel volume is reduced to approximately half compared with that in metabolically active cells.

40 min before cryo-EM grid preparation, as described above.

Live-cell fluorescence microscopy and FRAP imaging

Ibidi μ-Slide eight-well Ibidi-treat dishes were coated with 1 mg/ml Concanavalin A (Sigma-Aldrich) in phosphate-buffered saline (PBS) for 10 to 40 min at room temperature and washed three times with ddH_2O. Then, 300 μl of energy-depleted cells were seeded into the Ibidi μ-Slide dish and 150 μl of control cells were seeded and topped up with 150 μl glucose control medium to lower the amount of seeded cells and allow cells to keep dividing. Time of ED was measured from the first wash in ED buffer, typically ~12 min before finalizing the final wash and seeding eight-well dishes. Live-cell imaging and FRAP experiments were all performed on a laser scanning confocal microscope Zeiss LSM-780 equipped with an incubation chamber maintained at 30°C. All images were acquired with a Plan-Apochromat 63x/1.40 Oil differential interference contrast (DIC) objective.

For live-cell imaging of GD250 and CZ007 1532x1532 pixel z-stacks were acquired with a pixel size of 0.0887 μm, an optical section of 0.8 μm, and z-step of 0.38 μm using sequential line scanning and two line averaging rounds with 0.426 μs pixel-dwell time. GFP was excited with a 488 nm line of argon laser and detected with GaAsP spectral detector in the range 500 to 550 nm. mCherry was excited with a DPSS laser at 561 nm and detected with GaAsP spectral detector in the range 568 to 620 nm. Per image stack, 15 to 17 z-slices were acquired to cover the entire *S. pombe* cells. In total, three independent biological replicates were acquired in three independent experiments with each containing about five individual z-stacks. For energy-depleted cells, z-stacks were acquired at time intervals of 30 to 50 min, 55 to 70 min, and 150 to 180 min after ED (after first wash with ED buffer), and control conditions were acquired more than 2 hours after sample preparation. For energy-depleted cells, data acquisition of each time point was started with the first z-stack the given time point after ED. Z-stacks of osmotically shocked CZ001 cells were acquired 1 to 16 min after hypertonic medium exposure.

FRAP measurements of bleached nuclear and cytoplasmic GFP signal in the CZ001 strains were performed to measure passive nuclear transport and cytoplasmic diffusion, respectively.

Passive nucleocytoplasmic transport was measured by individual FRAP experiments performed in the GFP channel consisting of a time series of 140 152x90 pixel (13.6 μm × 8.1 μm) frames acquired over 138.45 s with an interval of 1 s between different frames, a frame time of 0.17 s, and a pixel size of 0.09 μm. The 488 nm line of the argon laser was used both for bleaching and for excitation during time lapse imaging. Bleaching was performed after acquiring 10 images. One bleaching iteration of 1.2 s was performed in a circular area with a diameter of 33 pixel (2.97 μm), previously assigned manually based on the Nup60-mCherry NE outline to cover the nuclear area. Per individual biological replicate independent time series were acquired on individual cells over a period 30 to 210 min after ED, i.e., after the first wash with ED buffer. In total, 48 to 55 FRAP experiments per biological replicate were performed consecutively with an interval of ~3 to 5 min between the individual measurements. Control measurements were performed after completion of

data acquisition on energy-depleted cells i.e., >210 min after sample preparation.

To measure cytoplasmic diffusion, FRAP datasets were acquired in the GFP channel with image dimensions of 80x30 pixel (13.6 μm × 5.1 μm), a pixel size of 0.17 μm, and 140 frames over a period of 2.11 s with a time interval between frames equal to 0.03 s. A rectangular area covering around half of the *S. pombe* cytoplasm and excluding the nucleus was bleached after acquiring 20 initial images. One bleaching iteration was performed in 20 ms with 100% power of 488 laser line and additional 40% of 405 nm laser line. In three independent biological replicates each, eight individual FRAP experiments at time points 30 min (27 to 44 min), 60 min (55 to 68 min), and >150 min (155 to 168 min) after ED were recorded on individual cells. Control measurements were performed after 180 min of sample preparation. FRAP experiments of CZ001 cells exposed to a OS were performed as described above for ED with the following modifications: Each measurement was performed in three biological and eight subsequent technical replicates ranging over a timeframe of 3 to 25 min after OS exposure. To accommodate for the significantly slower cytoplasmic diffusion upon OS, the imaging parameters described above for nuclear FRAP were used where a 33x33 pixel square was bleached within the cytoplasm and recovery was measured over 120 s with a 1-s interval.

Quantification of fluorescence microscopy images and FRAP analysis

To quantify the localization of NLS-GFP in GD250, z-stacks were maximum projected and the ratio of the average pixel intensity of a circular area within the nucleus compared with a similar area in the cytoplasm was measured manually in Fiji (*67*). The nuclear position was determined in the Nup60-mCherry channel. Cells that were not well attached to the eight-well dish and moving during z-stack acquisition or not entirely contained within the z-stack were removed before analysis, and cytoplasmic areas of clearly visible vacuoles were avoided. NES-GFP localization signal was quantified similarly as described above for NLS-GFP. To avoid signal bias from above and below the nucleus, only a single image slice approximately at the central plane of each cell was quantified.

Nuclear volume and sphericity of GD250 and CZ001 z-stacks were quantified based on 3D reconstructions of individual n of the Nup60-mCherry (with excluded areas of cells moving or not entirely contained within the imaged volume) using the 3DMembraneReconstruction (*68*) workflow in Mathwork's MATLAB.

FRAP curves were analyzed with FRAPAnalzer 2.1.0 (*69*). In brief, a double normalization was carried out normalizing against the background and the entire cellular surface as reference. Recovery half-life times were extracted by fitting a one exponential recovery equation to the normalized data. In the case of nuclear FRAP experiments under OS conditions, a two-exponential fitting was performed to account for the first (fast) recovery most likely corresponding to recovery of molecules bleached outside of the nucleus and a second (slower) recovery corresponding to the nuclear fluorescence recovery, the half-life time of the latter was reported as nuclear fluorescence half-life time. Moving cells and cells where the nuclear signal did not recover to a plateau were excluded from analysis.

Cryo-FIB milling

Plunge-frozen sample grids were FIB milled in an Aquilos FIB-SEM (Thermo Fisher) as previously described (*4*). In brief, samples were sputter coated with inorganic platinum at 10 kV voltage and 10 mPa argon gas pressure for ~10 s (Pt-sputtering). Subsequently a protective layer of organometallic platinum was deposited using the gas injection system (GIS-coating) for ~12 s. If needed, a second round of Pt-sputtering was performed to avoid charging during FIB milling. Between 5 and 10 lamellae per grid were milled in a step-wise fashion and polished to a final thickness of <280 nm using decreasing FIB current steps of 1 nA, 0.5 nA, 0.3 nA, and 50 pA. Before unloading the sample, an additional layer of Pt was deposited for 1 s to avoid charging effects during subsequent imaging in the transmission electron microscope (TEM).

Automated tomogram acquisition

The majority of WT *S. pombe* tomograms (136/178) as well as all tomograms from knockout strains were acquired on a Titan Krios G3 (Thermo Fisher), operating at 300 keV and equipped with a Gatan K2 Summit direct electron detector and energy-filter. All tilt-series (TS) were acquired in dose-fractionation mode at 4k × 4k resolution at a nominal pixel size of 3.45 Å. Automated TS acquisition was performed as described previously (*4*) using a dose-symmetric acquisition scheme (*70*) with an effective tilt range of −50° to 50°, considering a tilt-offset to compensate for the lamella angle (typically positive or negative 8° to 13°) (*4*), a 3° tilt increment, a total dose of 120 to 150 e/Å^2, and target defocus of −1.5 to −4.5 μm.

The remaining 42 WT tomograms and 15 of the total 76 tomograms from energy-depleted *S. pombe* cells were acquired on a Titan Krios (Thermo Fisher) equipped with a Gatan K2 Summit direct electron detector and energy-filter and a Volta potential phase plate (VPP) (*71*). Automated dose-symmetric tilt series acquisition was performed in 4k × 4k dose-fractionation mode with a nominal pixel size of 3.37 Å using a total dose of 100 to 125 e/Å^2 distributed over a tilt range of −60° to +60° or −50° to 50° with and a tilt step interval of 3° or 2°, respectively, a defocus range of −2 to −4 μm, and conditioning the VPP by exposure to the electron beam up to 5 min. An additional 11 tomograms of ED *S. pombe* cells were acquired using the same parameters but without usage of the VPP. All 57 tomograms of cells exposed to a OS were acquired with similar parameters including an effective tilt range of −50° to 50°, considering a tilt-offset to compensate for the lamella angle, a 3° tilt increment, a total dose of 120 to 150 e/Å^2, and target defocus of −2 to −5 μm.

The remaining 50 tomograms of ED *S. pombe* cells were acquired on a Titan Krios G3 (Thermo Fisher) operating at 300 keV equipped with a Gatan K3 direct electron detector and energy-filter operated in dose-fractionation mode and 5.7k × 4k resolution with a nominal pixel size of 3.425 Å. Automated dose-symmetric tilt series acquisition was performed with a nominal defocus range of −2 to −3.5 μm, an effective tilt range of −50° to 50° including a tilt offset to compensate for the lamella pretilt, a tilt step increment of 3°, and a total dose of 133 e/Å^2.

All 78 tomograms of cells recovered from a OS were acquired on a Titan Krios G2 (Thermo Fisher) equipped with a Gatan K3 direct electron detector and energy-filter with similar parameters as above including an effective tilt range of −50° to 50°, considering a tilt-offset to compensate for the lamella angle, a 3° tilt increment, a total dose of 142 e/Å^2, a target defocus of −2 to −5 μm, and a nominal pixel size of 3.372 Å. *C. thermophilum* tomograms were acquired as described above for WT *S. pombe* tomograms with a total dose of 130 to 140 e/Å^2 and a target defocus range of −2 to −4 μm.

Image preprocessing and tomogram reconstruction

Images were preprocessed, and contrast transfer function (CTF) was estimated as described previously (*4*). Tilt series were aligned automatically in a tailored workflow using the IMOD package including the patch-tracking functionalities (*72*). In detail, four-times binned (pixel size: 1.348 nm) and dose-filtered TS were low-pass filtered with seven empirically predefined low-pass filter parameters and seven initial patch-tracking attempts using 7x7 patches per tilt were performed in IMOD. The filter-set and patch tracking output yielding the lowest alignment residuals during the initial round of patch-tracking was selected for further iterative refinement. At each iteration, the contour with the largest residual error was removed before the next round of patch-tracking. This process was repeated until the overall residual dropped below 0.5 pixel or until a ratio of known/unknown in the IMOD tilt

align program dropped below a threshold of 10. The final aligned TS were reconstructed using the SIRT-like filtering option in IMOD (*73*) and manually inspected. Only tomograms yielding a high-quality alignment, both, visually and in terms of overall residual error (typically better than 1 pixel) were used for 3D CTF correction using the phase-flipping option in novaCTF (*74*) and subsequent STA workflow.

NE segmentation and tension analysis

Four-times binned 3D CTF-corrected tomograms were deconvolved in MATLAB using tom_deconv function (https://github.com/dtegunov/tom_deconv) and subsequently segmented using the convolutional neural network (CNN)–based tomogram segmentation workflow in EMAN2.2 (*75*) using a custom trained CNN to recognize NEs. The resulting segmentation was improved by applying the tensor voting–based membrane segmentation workflow TomoSegmemTV (*76*). The final binary segmented volume was manually inspected and curated, whereby false positive density such as NE connected ER was removed and tomograms resulting in overall inferior segmentation quality were excluded. Binary segmented NE volumes were subsequently loaded into MATLAB where a sphere was fitted to the coordinates of each individual segmentation. On the fitted sphere surface a sampling grid was defined with a kernel size (distance between sampling points) of 20 pixel (resulting in a subdivision of the spherical surface in ~27 nm × 27 nm patches). At each sampling point several measurements were performed.

First, at each sampling point the mean radius of all segmented voxels within a 20-pixel thick ray extending form the center of the fitted sphere through the sampling point were measured and the difference between the resulting mean NE radius and the fitted sphere radius was reported as deviation of the NE form a perfect sphere at each given sampling point. For each individual tomogram the variance or standard deviation of the deviation form a perfect sphere was then used as a measure of NE wobbliness (how much the NE deviates form a perfect sphere).

Second, the INM-ONM distance at each sampling point was measured by analyzing the distribution of radii of all voxels belonging to the NE segmentation within the 20-pixel-thick ray at each sampling point. Only radius distributions which showed a bimodal distribution were further used to perform a K-means clustering and obtain the mean NE radius of two classes. Sampling points which did not show a bimodal radius distribution were omitted. The shorter radius was then attributed to the INM radius, whereas larger value described the ONM radius at the given sampling point. The difference between the INM and ONM radius was reported as INM-ONM distance at any given sampling point, and the median over the entire tomogram was used as INM-ONM distance measurement for each individual tomogram.

Third, the total segmented NE surface was estimated by projecting the segmented NE voxels on the fitted sphere surface. Subsequently all sampling points covering at least one NE voxel within the 20-pixel radius were counted as 729 nm^2 (27 nm × 27 nm) membrane patch each and summed up to yield the total estimated NE surface. The number of initially picked NPCs (see below) per tomogram was divided by the corresponding estimated NE surface area to estimate the number of NPCs per square micrometer of NE.

Particle identification and STA

NPC coordinates and initial orientations were determined manually in four-times binned SIRT-like filtered (*73*) tomograms as described earlier (*4*). Additionally, a model describing the lamella slab geometry was generated manually by picking the coordinates of all eight corners of the lamella slab for each tomogram. Particle extraction, subtomogram alignment, and averaging (STA) was performed with novaSTA (*77*) from 3D CTF-corrected tomograms, as described previously (*4*). Initial alignment of NPCs was carried out using eight- and four-times binned particles. After obtaining an initial four-times binned average, coordinates of the eight individual NPC spokes were identified for symmetry-independent alignment, as described before (*6*). In brief, the coordinates of an individual subunit within the eightfold average was defined manually, and the remaining asymmetric subunit positions were calculated based on the eightfold symmetry (fig. S2A). The coordinates were then used to extract each individual asymmetric subunit form the original tomograms, where only subunits with coordinates retained in the lamella slab geometry model were retained to avoid the inclusion of subunits lying outside the FIB-milled lamellae. The individually extracted asymmetric subunits were then used for further alignment using a mask covering the entire asymmetric unit (cytoplasmic side, IR, and NR) (fig. S2B). After an initial subunit alignment, each subtomogram and its assigned orientation was inspected manually and misaligned or remaining false positive particles, i.e., subunits extracted outside of the FIB-milled lamellae, were removed. Subtomogram alignment was further continued using four- and two-times binned subtomograms, where the alignment was focused on the CR, IR, or NR using localized masks (fig. S2C). VPP data were excluded from further STA of the WT averages at the transition from bin4 to bin2 processing. If needed, the particle boxes were recentered around the individual rings to achieve better subtomogram alignment. In case of the WT subtomogram average, particles with low final cross correlation scores were removed for the final average. All subtomogram averages of WT, OS, and knockout datasets were b-factor sharpened empirically (*11*) and filtered to the given resolution determined on gold-standard FSC calculations at the 0.143 criterion. For ED averages, instead of b-factor sharpening, the amplitude of the final averages (containing VPP and defocus data) were matched to the amplitudes of a corresponding average generated by averaging only the defocus type subset of data (excluding all VPP data) with the EMAN2 (*78*) e2proc3d.py -matchto function, to overcome a previously described low-pass filter-like artifact occurring during STA of VPP data (*79*).

To generate the eightfold symmetric assemblies, the final individual ring maps were first fit to the initially aligned whole asymmetric unit map to find their correct relative position. The root mean square density value of all individual maps were scaled to similar values to allow appropriate representation of all three rings at a common threshold level. The resulting asymmetric unit was then assembled to an eightfold symmetric assembly based on the coordinates defined during the initial subunit extraction (see above).

Details about tomogram and particle numbers of each average can be found in table S3.

Difference map calculation

To calculate the difference map of the *nup37*Δ and *nup37*Δ-*ely5*Δ knockout to the WT, the maps of the individual knockout rings were filtered according to their reported resolution (*nup37*Δ cytoplasmic side: 27 Å, IR: 27 Å, and NR: 31 Å; cytoplasmic side: 35 Å, IR: 35 Å, and NR: 34 Å of *nup37*Δ-*ely5*Δ, respectively) (table S3 and figs. S8 and S9), and WT maps were filtered to the corresponding knockout resolution. To account for different b-factors and amplitudes, the power spectra of the WT maps were adjusted to the corresponding knockout map using RELION (*80*) relion_image_handler -adjust_power command. To obtain comparable difference maps across the density of the individual knockout rings, the WT cytoplasmic and nuclear map were scaled to match the root mean square of the corresponding WT IR map. The WT and corresponding knockout maps were then brought to the same reference frame by fitting the maps against each other and difference maps of the individual rings were calculated in UCSF Chimera (*81*). The final maps are shown at the same threshold and overlaid with the difference map (WT minus knockout) and the inverse (negative threshold) of the difference map, respectively (Fig. 2 and figs. S8 and S9)

NPC diameter and volume measurements

NPC diameters were measured with an in-house MATLAB script based on the final coordinates and orientations obtained for each individual subunit during STA. First a feature of interest in the subunit average was identified in UCSF Chimera (*81*) by placing a 1-voxel sphere mask on the point of interest. The offset between the center of the average and the mask was then used to calculate the coordinates of the area of interest within each averaged subunit in respect to the original tomograms. To calculate the diameter of each individual NPC, only NPCs with a subunit-occupancy of five or more were considered. For each individual NPC, vectors were derived connecting the opposing subunits. Based on those vectors, the center of each NPC was defined as the center of the intersection area between all vectors corresponding to a given NPC. The average distance between the newly identified center and each respective SU was chosen as a representative NPC radius for the given pore. By this means, an accurate average NPC radius measurement for a specific feature of interest within each individual NPC is obtained. Additionally, the z-translation of each subunit from the NPC center plane and the pairwise distance opposing subunits based on coordinates obtained from whole asymmetric subunit alignment were reported (fig. S17, A to G).

Homology modeling

Modeling templates were detected and selected using HHpred server (*82*). Sequence alignments for modeling were refined in Swiss PDB Viewer (*83*). The models were built based on the templates and alignments using Modeller (*84*). The templates used for modeling are listed in table S4.

Systematic fitting of SpNPC rigid bodies to cryo-ET maps

To assign the number of Y-complexes and IR subcomplexes, and to prepare the input for integrative modeling (fig. S5A), an unbiased global fitting approach was applied using several SpNPC structural models. Two of them were experimental x-ray structures previously published (*35*, *36*), whereas the remaining were homology modeled on the basis of *S. cerevisiae* and *C. thermophilum* NPC components (*20*, *85*–*88*) (table S4). All the aforementioned high-resolution structures were filtered to 40 Å before fitting. The resulting simulated model maps were subsequently fitted into individual segments of the SpNPC cryo-EM maps from this study by global fitting as implemented in UCSF Chimera (*81*). More precisely, structures of the Y-complex were fitted into cytoplasmic and NR segments of the cryo-EM SpNPC maps while the IR model was fitted in the IR. The segmented maps that were used for fitting did not include NE density to eliminate the possibility of fits substantially overlapping with the membrane.

All fitting runs were performed using 100,000 random initial placements with the requirement of at least 30% of the simulated model map to be covered by the SpNPC density envelope defined at a low threshold. For each fitted model, this procedure resulted in ~250 to 17,200 fits with nonredundant conformations upon clustering. The cross-correlation about the mean (cam score, equivalent to Pearson correlation) score from UCSF Chimera (*81*) was used as a fitting metric for each atomic structure, similarly to our previously published works (*4*, *11*, *12*, *31*). The statistical significance of every fitted model was evaluated as a *P* value derived from the cam scores. The calculation of *P* values was performed by first transforming the cross-correlation scores to z-scores (Fisher's z-transform) and centering, from which subsequently two-sided *P* values were computed using standard deviation derived from an empirical null distribution [based on all obtained nonredundant fits and fitted using fdrtool (*89*) R-package]. Finally, the *P* values were corrected for multiple testing with Benjamini-Hochberg (*90*). Figures were produced by UCSF Chimera (*81*) and UCSF ChimeraX (*91*).

Integrative modeling

Structural models of the SpNPCs were built using the integrative procedure (fig. S5) implemented in our Assembline software (*92*). Assembline implements a custom modeling pipeline based on Integrative Modeling Platform (IMP) (*93*) version 2.14 and Python Modeling Interface (PMI) (*94*). The pipeline is similar to the one we used previously for human and *S. cerevisiae* NPCs (*4*, *12*). The modeling procedure consisted of three steps (fig. S5B).

Step 1

First, ensembles of alternative models of the individual rings in the control NPC were constructed. To model the cytoplasmic Y-complexes and NR, the input homology models were divided into smaller rigid bodies (with cut points corresponding to boundaries of published crystal structures) and simultaneously fitted to the cryo-ET map using the global optimization step in Assembline. In this step, first, libraries of alternative nonredundant fits for each rigid body were generated using the systematic fitting procedure described above. Then, the rigid bodies were fitted simultaneously to the maps of individual rings by sampling alternative fits from these libraries using the simulated annealing Monte Carlo method. To cover a large landscape of possible models, the optimization was repeated 4000 and 40,000 times for the cytoplasmic side and NR, respectively, with each repetition leading to a candidate model. To achieve convergence and sampling exhaustiveness [assessed using the procedure by Viswanath *et al.* (*95*)], each repeat run was composed of 130,000 (cytoplasmic side) and 271,000 (NR) Monte Carlo steps at decreasing temperatures. The higher number of steps and models for NR was used because of the higher number of rigid bodies in this ring. The scoring function for the modeling optimization was a linear combination of the EM fit restraint represented as the *P* values of the precalculated rigid-body fits (from systematic fitting as described above), clash score (SoftSpherePairScore of IMP), connectivity distance between domains neighboring in sequence, a term preventing overlap of the protein mass with the NE, a restraint promoting the membrane-binding loops of Nup131, Nup132, and Nup120 to interact with the envelope, implemented using MapDistanceTransform of IMP [predicted by similarity to known or predicted ALPS motifs in human and *S. cerevisiae* homologs (*1*, *96*)], the restraint for Ely5 localization in the region identified based on the difference between the *nup37*Δ and *nup37*Δ-*ely5*Δ knockout NPC maps to the control NPC, and restraint for biochemically characterized interaction restraint between Ely5 and Nup120 (*36*). For the NR, a restraint promoting Nup37 localization from the difference between the *nup37*Δ NPC and the control NPC was also used. During the optimization, the structures were simultaneously represented at two resolutions: in Cα-only representation and in a coarse-grained representation, in which each 10-residue stretch was converted to a bead. The 10-residue bead representation was used for all restraints to increase computational efficiency except for the domain connectivity restraints, for which the Cα-only representation was used. Because the *P* values used for the EM restraint were derived from the original EM fit libraries generated with UCSF Chimera, the EM restraint can be regarded as an EM restraint derived from the full atom representation.

Owing to high structural conservation of the IR as determined by the systematic fitting, this ring was modeled by fitting a homology model of the entire IR built based on *S. cerevisiae* NPC, followed by the refinement with 50,000 Monte Carlo steps at a single temperature. The refinement involved the Monte Carlo simulated annealing optimization of the rigid bodies' rotations and translations, but this time without using the libraries of alternative nonredundant fits, allowing the rigid bodies to move in the EM map with small rotation and translation increments. The scoring function consisted of cross-correlation to the EM map (FitRestraint of IMP), domain connectivity restraint, the clash score, the envelope exclusion restraints as above, and a restraint for the membrane-binding loop of Nup155 (*12*).

Step 2

The above steps led to ensembles of separate models for the cytoplasmic side, NR, and IR. To build the model of the full NPC, the alternative fits of each subunit in the cytoplasmic side and NR models were extracted into a new set of fit libraries. Then, together with the top-scoring model of the IR, the global optimization step was repeated using the new fit libraries and now fitting all subunits from all three rings of the NPC simultaneously. The resulting models were refined as above for the IR, leading to an ensemble of models of the full NPC. An elastic network restraint was used to keep the Ely5-Nup120 orientation similar in all rings.

Step 3

To generate the models of the ED NPCs, the 1000 top-scoring models of the NPC under control conditions were fitted into the ED EM maps and optimized using the refinement procedure. The scoring function comprised the same restraints as for the refinement of the individual rings except for the restraint for Nup37 localization in the difference density. Each of the 1000 starting models was refined with 150,000 steps at five temperature levels leading to an ensemble of 1000 alternative models.

Model analysis

To assess convergence and exhaustiveness of sampling, and to assess the precision of sampling and of the models, we used the procedure by Viswanath *et al.* (*95*). The quantities were assessed at each modeling step of the procedure (fig. S5B). The sampling precision for the control, ED medium, and constricted diameter were 8.4, 12.6, and 12.8 Å, respectively. At this sampling precision, we obtained two to four clusters of models with the model precision between 7.2 and 9.7 Å (fig. S5B). The clusters differed mostly in the orientations of Ely5, Nup120, and the N-terminal propeller of Nup133. The model with the orientation of Ely5 consistent with the model of the human NPC (*11*) was used as a representative model for the figures. For clarity, a single model was shown in the figures and the top 10 models for each ensemble were shown in fig. S5.

REFERENCES AND NOTES

1. D. H. Lin, A. Hoelz, The Structure of the Nuclear Pore Complex (An Update). *Annu. Rev. Biochem.* **88**, 725–783 (2019). doi: 10.1146/annurev-biochem-062917-011901; pmid: 30883195
2. M. Beck, S. Mosalaganti, J. Kosinski, From the resolution revolution to evolution: Structural insights into the evolutionary relationships between vesicle coats and the nuclear pore. *Curr. Opin. Struct. Biol.* **52**, 32–40 (2018). doi: 10.1016/j.sbi.2018.07.012; pmid: 30103204
3. J. Mahamid *et al.*, Visualizing the molecular sociology at the HeLa cell nuclear periphery. *Science* **351**, 969–972 (2016). doi: 10.1126/science.aad8857; pmid: 26917770
4. M. Allegretti *et al.*, In-cell architecture of the nuclear pore and snapshots of its turnover. *Nature* **586**, 796–800 (2020). doi: 10.1038/s41586-020-2670-5; pmid: 32879490
5. G. J. Stanley, A. Fassati, B. W. Hoogenboom, Atomic force microscopy reveals structural variability amongst nuclear pore complexes. *Life Sci. Alliance* **1**, e201800142 (2018). doi: 10.26508/lsa.201800142; pmid: 30456374
6. M. Beck, V. Lučić, F. Förster, W. Baumeister, O. Medalia, Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. *Nature* **449**, 611–615 (2007). doi: 10.1038/nature06170; pmid: 17851530
7. J. Sellés *et al.*, Nuclear pore complex plasticity during developmental process as revealed by super-resolution microscopy. *Sci. Rep.* **7**, 14732 (2017). doi: 10.1038/s41598-017-15433-2; pmid: 29116248
8. A. P. Schuller *et al.*, The cellular environment shapes the nuclear pore complex architecture. *Nature* **598**, 667–671 (2021). doi: 10.1038/s41586-021-03985-3; pmid: 34646014
9. V. Zila *et al.*, Cone-shaped HIV-1 capsids are transported through intact nuclear pores. *Cell* **184**, 1032–1046.e18 (2021). doi: 10.1016/j.cell.2021.01.025; pmid: 33571428
10. K. H. Bui *et al.*, Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233–1243 (2013). doi: 10.1016/j.cell.2013.10.055; pmid: 24315095
11. A. von Appen *et al.*, *In situ* structural analysis of the human nuclear pore complex. *Nature* **526**, 140–143 (2015). doi: 10.1038/nature15381; pmid: 26416747
12. J. Kosinski *et al.*, Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016). doi: 10.1126/science.aaf0643; pmid: 27081072
13. S. Otsuka *et al.*, Postmitotic nuclear pore assembly proceeds by radial dilation of small membrane openings. *Nat. Struct. Mol. Biol.* **25**, 21–28 (2018). doi: 10.1038/s41594-017-0001-9; pmid: 29323269
14. S. Otsuka *et al.*, Nuclear pore assembly proceeds by an inside-out extrusion of the nuclear envelope. *eLife* **5**, e19071 (2016). doi: 10.7554/eLife.19071; pmid: 27630123
15. J. Koh, G. Blobel, Allosteric regulation in gating the central channel of the nuclear pore complex. *Cell* **161**, 1361–1373 (2015). doi: 10.1016/j.cell.2015.05.013; pmid: 26046439
16. A. C. Meinema *et al.*, Long unfolded linkers facilitate membrane protein import through the nuclear pore complex. *Science* **333**, 90–93 (2011). doi: 10.1126/science.1205741; pmid: 21659568
17. R. Ungricht, M. Klann, P. Horvath, U. Kutay, Diffusion and retention are major determinants of protein targeting to the inner nuclear membrane. *J. Cell Biol.* **209**, 687–703 (2015). doi: 10.1083/jcb.201409127; pmid: 26056139
18. A. Boni *et al.*, Live imaging and modeling of inner nuclear membrane targeting reveals its molecular requirements in mammalian cells. *J. Cell Biol.* **209**, 705–720 (2015). doi: 10.1083/jcb.201409133; pmid: 26056140
19. H. Chug, S. Trakhanov, B. B. Hülsmann, T. Pleiner, D. Görlich, Crystal structure of the metazoan Nup62•Nup58•Nup54 nucleoporin complex. *Science* **350**, 106–110 (2015). doi: 10.1126/science.aac7420; pmid: 26292704
20. T. Stuwe *et al.*, Architecture of the fungal nuclear pore inner ring complex. *Science* **350**, 56–64 (2015). doi: 10.1126/science.aac9176; pmid: 26316600
21. V. Shahin, T. Danker, K. Enss, R. Ossig, H. Oberleithner, Evidence for Ca^{2+}- and ATP-sensitive peripheral channels in nuclear pore complexes. *FASEB J.* **15**, 1895–1901 (2001). doi: 10.1096/fj.00-0838com; pmid: 11532969
22. D. Stoffler, K. N. Goldie, B. Feja, U. Aebi, Calcium-mediated structural changes of native nuclear pore complexes monitored by time-lapse atomic force microscopy. *J. Mol. Biol.* **287**, 741–752 (1999). doi: 10.1006/jmbi.1999.2637; pmid: 10191142
23. A. Rakowska, T. Danker, S. W. Schneider, H. Oberleithner, ATP-Induced shape change of nuclear pores visualized with the atomic force microscope. *J. Membr. Biol.* **163**, 129–136 (1998). doi: 10.1007/s002329900377; pmid: 9592077
24. L. Kastrup, H. Oberleithner, Y. Ludwig, C. Schafer, V. Shahin, Nuclear envelope barrier leak induced by dexamethasone. *J. Cell. Physiol.* **206**, 428–434 (2006). doi: 10.1002/jcp.20479; pmid: 16110478
25. I. Liashkovich, A. Meyring, A. Kramer, V. Shahin, Exceptional structural and mechanical flexibility of the nuclear pore complex. *J. Cell. Physiol.* **226**, 675–682 (2011). doi: 10.1002/jcp.22382; pmid: 20717933
26. R. D. Jäggi *et al.*, Modulation of nuclear pore topology by transport modifiers. *Biophys. J.* **84**, 665–670 (2003). doi: 10.1016/S0006-3495(03)74886-3; pmid: 12524319
27. M. Eibauer *et al.*, Structure and gating of the nuclear pore complex. *Nat. Commun.* **6**, 7532 (2015). doi: 10.1038/ncomms8532; pmid: 26112706
28. Y. Zhang *et al.*, Molecular architecture of the luminal ring of the *Xenopus laevis* nuclear pore complex. *Cell Res.* **30**, 532–540 (2020). doi: 10.1038/s41422-020-0320-y; pmid: 32367042
29. G. Huang *et al.*, Structure of the cytoplasmic ring of the *Xenopus laevis* nuclear pore complex by cryo-electron microscopy single particle analysis. *Cell Res.* **30**, 520–531 (2020). doi: 10.1038/s41422-020-0319-4; pmid: 32376910
30. M. C. Field, M. P. Rout, Pore timing: The evolutionary origins of the nucleus and nuclear pore complex. *F1000Research* **8**, 369 (2019). doi: 10.12688/f1000research.16402.1; pmid: 31001417
31. S. Mosalaganti *et al.*, In situ architecture of the algal nuclear pore complex. *Nat. Commun.* **9**, 2361 (2018). doi: 10.1038/s41467-018-04739-y; pmid: 29915221
32. H. Asakawa *et al.*, Asymmetrical localization of Nup107-160 subcomplex components within the nuclear pore complex in fission yeast. *PLOS Genet.* **15**, e1008061 (2019). doi: 10.1371/journal.pgen.1008061; pmid: 31170156
33. J. Fernandez-Martinez *et al.*, Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. *Cell* **167**, 1215–1228.e25 (2016). doi: 10.1016/j.cell.2016.10.028; pmid: 27839866
34. P. Stelter *et al.*, Molecular basis for the functional interaction of dynein light chain with the nuclear-pore complex. *Nat. Cell Biol.* **9**, 788–796 (2007). doi: 10.1038/ncb1604; pmid: 17546040
35. X. Liu, J. M. Mitchell, R. W. Wozniak, G. Blobel, J. Fan, Structural evolution of the membrane-coating module of the nuclear pore complex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16498–16503 (2012). doi: 10.1073/pnas.1214557109; pmid: 23019579
36. S. Bilokapic, T. U. Schwartz, Molecular basis for Nup37 and ELY5/ELYS recruitment to the nuclear pore complex. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 15241–15246 (2012). doi: 10.1073/pnas.1205151109; pmid: 22955883
37. B. A. Rasala, A. V. Orjalo, Z. Shen, S. Briggs, D. J. Forbes, ELYS is a dual nucleoporin/kinetochore protein required for nuclear pore assembly and proper cell division. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17801–17806 (2006). doi: 10.1073/pnas.0608484103; pmid: 17098863
38. S. W. Baï *et al.*, The fission yeast Nup107-120 complex functionally interacts with the small GTPase Ran/Spi1 and is required for mRNA export, nuclear pore distribution, and proper cell division. *Mol. Cell. Biol.* **24**, 6379–6392 (2004). doi: 10.1128/MCB.24.14.6379-6392.2004; pmid: 15226438
39. K. Kelley, K. E. Knockenhauer, G. Kabachinski, T. U. Schwartz, Atomic structure of the Y complex of the nuclear pore. *Nat. Struct. Mol. Biol.* **22**, 425–431 (2015). doi: 10.1038/nsmb.2998; pmid: 25822992
40. S. J. Kim *et al.*, Integrative structure and functional anatomy of a nuclear pore complex. *Nature* **555**, 475–482 (2018). doi: 10.1038/nature26003; pmid: 29539637
41. T. Maimon, N. Elad, I. Dahan, O. Medalia, The human nuclear pore complex as revealed by cryo-electron tomography. *Structure* **20**, 998–1006 (2012). doi: 10.1016/j.str.2012.03.025; pmid: 22632834
42. M. C. Munder *et al.*, A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy. *eLife* **5**, e09347 (2016). doi: 10.7554/eLife.09347; pmid: 27003292
43. W. D. Richardson, A. D. Mills, S. M. Dilworth, R. A. Laskey, C. Dingwall, Nuclear protein migration involves two steps: Rapid binding at the nuclear envelope followed by slower translocation through nuclear pores. *Cell* **52**, 655–664 (1988). doi: 10.1016/0092-8674(88)90403-5; pmid: 3125984
44. N. Shulga *et al.*, In vivo nuclear transport kinetics in *Saccharomyces cerevisiae*: A role for heat shock protein 70 during targeting and translocation. *J. Cell Biol.* **135**, 329–339 (1996). doi: 10.1083/jcb.135.2.329; pmid: 8896592
45. W. A. Whalen, J. H. Yoon, R. Shen, R. Dhar, Regulation of mRNA export by nutritional status in fission yeast. *Genetics* **152**, 827–838 (1999). doi: 10.1093/genetics/152.3.827; pmid: 10388805
46. E. D. Schwoebel, T. H. Ho, M. S. Moore, The mechanism of inhibition of Ran-dependent nuclear transport by cellular ATP depletion. *J. Cell Biol.* **157**, 963–974 (2002). doi: 10.1083/jcb.200111077; pmid: 12058015
47. R. W. Wozniak, G. Blobel, M. P. Rout, POM152 is an integral protein of the pore membrane domain of the yeast nuclear envelope. *J. Cell Biol.* **125**, 31–42 (1994). doi: 10.1083/jcb.125.1.31; pmid: 8138573
48. P. Upla *et al.*, Molecular Architecture of the Major Membrane Ring Component of the Nuclear Pore Complex. *Structure* **25**, 434–445 (2017). doi: 10.1016/j.str.2017.01.006; pmid: 28162953

49. S. Abuhattum *et al*., Intracellular Mass Density Increase Is Accompanying but Not Sufficient for Stiffening and Growth Arrest of Yeast Cells. *Front. Phys.* **6**, 131 (2018). doi: 10.3389/fphy.2018.00131
50. A. Miermont *et al*., Severe osmotic compression triggers a slowdown of intracellular signaling, which can be explained by molecular crowding. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5725–5730 (2013). doi: 10.1073/pnas.1215367110; pmid: 23493557
51. R. P. Joyner *et al*., A glucose-starvation response regulates the diffusion of macromolecules. *eLife* **5**, e09376 (2016). doi: 10.7554/eLife.09376; pmid: 27003290
52. G. Marini, E. Nüske, W. Leng, S. Alberti, G. Pigino, Reorganization of budding yeast cytoplasm upon energy depletion. *Mol. Biol. Cell* **31**, 1232–1245 (2020). doi: 10.1091/mbc.E20-02-0125; pmid: 32293990
53. M. B. Heimlicher *et al*., Reversible solidification of fission yeast cytoplasm after prolonged nutrient starvation. *J. Cell Sci.* **132**, jcs.231688 (2019). doi: 10.1242/jcs.231688; pmid: 31558680
54. A. J. Lomakin *et al*., The nucleus acts as a ruler tailoring cell responses to spatial constraints. *Science* **370**, eaba2894 (2020). doi: 10.1126/science.aba2894; pmid: 33060332
55. V. Venturini *et al*., The nucleus measures shape changes for cellular proprioception to control dynamic cell behavior. *Science* **370**, eaba2644 (2020). doi: 10.1126/science.aba2644; pmid: 33060331
56. A. Agrawal, T. P. Lele, Mechanics of nuclear membranes. *J. Cell Sci.* **132**, jcs229245 (2019). doi: 10.1242/jcs.229245; pmid: 31308244
57. P. T. Tran, L. Marsh, V. Doye, S. Inoué, F. Chang, A mechanism for nuclear positioning in fission yeast based on microtubule pushing. *J. Cell Biol.* **153**, 397–412 (2001). doi: 10.1083/jcb.153.2.397; pmid: 11309419
58. S. M. Schreiner, P. K. Koo, Y. Zhao, S. G. J. Mochrie, M. C. King, The tethering of chromatin to the nuclear envelope supports nuclear mechanics. *Nat. Commun.* **6**, 7159 (2015). doi: 10.1038/ncomms8159; pmid: 26074052
59. A. Agrawal, T. P. Lele, Geometry of the nuclear envelope determines its flexural stiffness. *Mol. Biol. Cell* **31**, 1815–1821 (2020). doi: 10.1091/mbc.E20-02-0163; pmid: 32583742
60. A. Elosegui-Artola *et al*., Force Triggers YAP Nuclear Entry by Regulating Transport across Nuclear Pores. *Cell* **171**, 1397–1410.e14 (2017). doi: 10.1016/j.cell.2017.10.008; pmid: 29107331
61. B. Enyedi, P. Niethammer, Nuclear membrane stretch and its role in mechanotransduction. *Nucleus* **8**, 156–161 (2017). doi: 10.1080/19491034.2016.1263411; pmid: 28112995
62. N. Kellner *et al*., Developing genetic tools to exploit *Chaetomium thermophilum* for biochemical analyses of eukaryotic macromolecular assemblies. *Sci. Rep.* **6**, 20937 (2016). doi: 10.1038/srep20937; pmid: 26864114
63. H. Amelina *et al*., Sequential and counter-selectable cassettes for fission yeast. *BMC Biotechnol.* **16**, 76 (2016). doi: 10.1186/s12896-016-0307-4
64. J. Bähler *et al*., Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* **14**, 943–951 (1998). doi: 10.1002/(SICI)1097-0061(199807)14:10<943::AID-YEA292>3.0.CO;2-Y; pmid: 9717240
65. G. Dey *et al*., Closed mitosis requires local disassembly of the nuclear envelope. *Nature* **585**, 119–123 (2020). doi: 10.1038/s41586-020-2648-3
66. A. Vještica *et al*., A toolbox of stable integration vectors in the fission yeast *Schizosaccharomyces pombe*. *J. Cell Sci.* **133**, jcs240754 (2020). doi: 10.1242/jcs.240754; pmid: 31801797
67. J. Schindelin *et al*., Fiji: An open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012). doi: 10.1038/nmeth.2019; pmid: 22743772
68. J. F. Williams, S. G. J. Mochrie, M. C. King, A versatile image analysis platform for three-dimensional nuclear reconstruction. *Methods* **157**, 15–27 (2019). doi: 10.1016/j.ymeth.2018.10.009; pmid: 30359725
69. A. Halavatyi, S. Terjung, in *Standard and Super-Resolution Bioimaging Data Analysis: A Primer*, A. Wheeler, R. Henriques, Eds. (Wiley, 2017), pp. 99–141.
70. W. J. H. Hagen, W. Wan, J. A. G. Briggs, Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017). doi: 10.1016/j.jsb.2016.06.007; pmid: 27313000
71. R. Danev, B. Buijsse, M. Khoshouei, J. M. Plitzko, W. Baumeister, Volta potential phase plate for in-focus phase contrast transmission electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15635–15640 (2014). doi: 10.1073/pnas.1418377111; pmid: 25331897
72. D. N. Mastronarde, S. R. Held, Automated tilt series alignment and tomographic reconstruction in IMOD. *J. Struct. Biol.* **197**, 102–113 (2017). doi: 10.1016/j.jsb.2016.07.011; pmid: 27444392
73. D. N. Mastronarde, Dual-axis tomography: An approach with alignment methods that preserve resolution. *J. Struct. Biol.* **120**, 343–352 (1997). doi: 10.1006/jsbi.1997.3919; pmid: 9441937
74. B. Turoňová, F. K. M. Schur, W. Wan, J. A. G. Briggs, Efficient 3D-CTF correction for cryo-electron tomography using NovaCTF improves subtomogram averaging resolution to 3.4 Å. *J. Struct. Biol.* **199**, 187–195 (2017). doi: 10.1016/j.jsb.2017.07.007; pmid: 28743638
75. M. Chen *et al*., Convolutional neural networks for automated annotation of cellular cryo-electron tomograms. *Nat. Methods* **14**, 983–985 (2017). doi: 10.1038/nmeth.4405; pmid: 28846087
76. A. Martinez-Sanchez, I. Garcia, S. Asano, V. Lucic, J. J. Fernandez, Robust membrane detection based on tensor voting for electron tomography. *J. Struct. Biol.* **186**, 49–61 (2014). doi: 10.1016/j.jsb.2014.02.015; pmid: 24625523
77. B. Turoňová *et al*., In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **370**, 203–208 (2020). doi: 10.1126/science.abd5223
78. G. Tang *et al*., EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007). doi: 10.1016/j.jsb.2006.05.009; pmid: 16859925
79. B. Turoňová *et al*., Benchmarking tomographic acquisition schemes for high-resolution structural biology. *Nat. Commun.* **11**, 876 (2020). doi: 10.1038/s41467-020-14535-2; pmid: 32054835
80. S. H. W. Scheres, RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012). doi: 10.1016/j.jsb.2012.09.006; pmid: 23000701
81. E. F. Pettersen *et al*., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
82. L. Zimmermann *et al*., A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *J. Mol. Biol.* **430**, 2237–2243 (2018). doi: 10.1016/j.jmb.2017.12.007; pmid: 29258817
83. N. Guex, M. C. Peitsch, SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997). doi: 10.1002/elps.1150181505; pmid: 9504803
84. A. Šali, T. L. Blundell, Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **234**, 779–815 (1993). doi: 10.1006/jmbi.1993.1626; pmid: 8254673
85. T. Stuwe *et al*., Nuclear pores. Architecture of the nuclear pore complex coat. *Science* **347**, 1148–1152 (2015). doi: 10.1126/science.aaa4136; pmid: 25745173
86. D. H. Lin *et al*., Architecture of the symmetric core of the nuclear pore. *Science* **352**, aaf1015 (2016). doi: 10.1126/science.aaf1015; pmid: 27081075
87. C. S. Weirich, J. P. Erzberger, J. M. Berger, K. Weis, The N-terminal domain of Nup159 forms a β-propeller that functions in mRNA export by tethering the helicase Dbp5 to the nuclear pore. *Mol. Cell* **16**, 749–760 (2004). doi: 10.1016/j.molcel.2004.10.032; pmid: 15574330
88. K. Yoshida, H. S. Seo, E. W. Debler, G. Blobel, A. Hoelz, Structural and functional analysis of an essential nucleoporin heterotrimer on the cytoplasmic face of the nuclear pore complex. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16571–16576 (2011). doi: 10.1073/pnas.1112846108; pmid: 21930948
89. K. Strimmer, fdrtool: A versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* **24**, 1461–1462 (2008). doi: 10.1093/bioinformatics/btn209; pmid: 18441000
90. Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995). doi: 10.1111/j.2517-6161.1995.tb02031.x
91. T. D. Goddard *et al*., UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018). doi: 10.1002/pro.3235; pmid: 28710774
92. V. Rantos, K. Karius, J. Kosinski, Integrative structural modelling of macromolecular complexes using Assembline. *bioRxiv* 2021.04.06.438590 [Preprint] (2021). doi: 10.1101/2021.04.06.438590
93. B. Webb *et al*., Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci.* **27**, 245–258 (2018). doi: 10.1002/pro.3311; pmid: 28960548
94. D. Saltzberg, C. H. Greenberg, S. Viswanath, I. Chemmama, B. Webb, R. Pellarin, I. Echeverria, A. Sali, in *Biomolecular Simulations*, M. Bonomi, C. Camilloni, Eds., *Methods in Molecular Biology, Volume 2022* (Humana, 2019), pp. 353–377.
95. S. Viswanath, I. E. Chemmama, P. Cimermancic, A. Sali, Assessing Exhaustiveness of Stochastic Sampling for Integrative Modeling of Macromolecular Structures. *Biophys. J.* **113**, 2344–2353 (2017). doi: 10.1016/j.bpj.2017.10.005; pmid: 29211988
96. G. Drin *et al*., A general amphipathic α-helical motif for sensing membrane curvature. *Nat. Struct. Mol. Biol.* **14**, 138–146 (2007). doi: 10.1038/nsmb1194; pmid: 17220896
97. J. Kosinski *et al*., Data for integrative modeling of the Nuclear Pore Complex from *Schizosaccharomyces pombe*, *Zenodo* (2021). doi: 10.5281/zenodo.5585949

ACKNOWLEDGMENTS

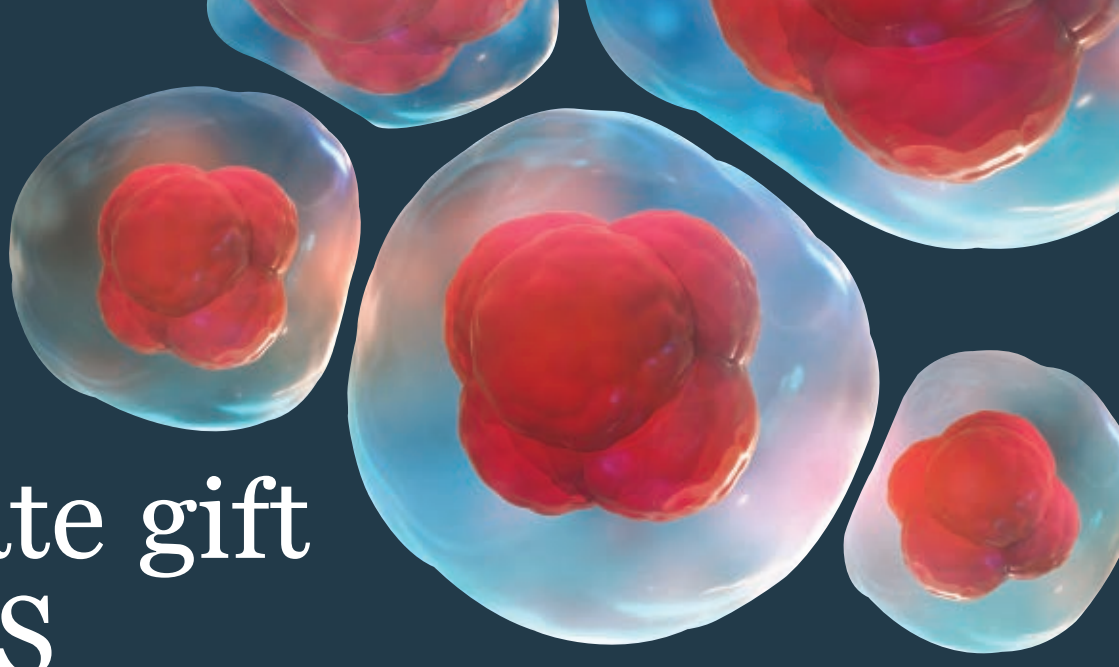
We thank B. Turoňová, N. Kellner, W. Hagen, F. Weiss, C. Tischer, J. Baumbach, S. Welsch, M. Linder, E. Hurt, E. Lemke, and all the members of the Mahamid, Kosinski, and Beck laboratories for advice and support. We thank E. Hurt, B. Baum, and T. Schwartz for providing yeast strains. We acknowledge support from the Electron Microscopy Core Facility, the Advanced Light Microscopy Facility, and IT services of EMBL Heidelberg and the Microscopy Core Facility of the Max Planck Institute of Biophysics, Frankfurt am Main. **Funding:** M.A. was funded by an EMBO long-term fellowship (ALTF-1389–2016). J.M. received funding from the European Research Council (ERC 3DCellPhase 760067). J.K. was supported by funding from the Federal Ministry of Education and Research of Germany (FKZ 031L0100). M.B. acknowledges funding by EMBL, the Max Planck Society, and the European Research Council (ComplexAssembly 724349). **Author contributions:** C.E.Z. conceived the project, designed and performed experiments, acquired all types of data, designed and established data analysis procedures, analyzed all types of data, and wrote the manuscript; M.A. conceived the project, designed and performed experiments, acquired data, analyzed data, and wrote the manuscript; V.R. analyzed data and wrote the manuscript; S.K.G. designed and performed experiments and acquired data; A.O.-K. analyzed data; I.Z. designed and performed experiments; A.H. designed and performed experiments; J.M. designed experiments and supervised the project; G.H. designed experiments, analyzed data, and derived the theoretical model based on membrane elastic theory; J.K. conceived the project, designed and established data analysis procedures, analyzed data, supervised the project, and wrote the manuscript; and M.B. conceived the project, designed experiments, supervised the project, and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Associated with the manuscript are accession numbers EMD-11373, EMD-11374, EMD-11375, and EMD-13081 (EM Data Bank, www.ebi.ac.uk/emdb/), as well as PDBDEV_00000094, PDBDEV_00000095, and PDBDEV_00000096 (PDB-Dev database, https://pdb-dev.wwpdb.org/). The code for the structural modeling and the input files are available on Zenodo (97).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abd9776
Supplementary Text
Figs. S1 to S18
Tables S1 to S4
References (*98*–*106*)
MDAR Reproducibility Checklist
Movies S1 to S7

24 July 2020; resubmitted 14 October 2021
Accepted 27 October 2021
Published online 11 November 2021
10.1126/science.abd9776

1848 SOCIETY
AAAS



An estate gift to AAAS

Going all the way back to 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

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

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

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

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

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

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

1848 SOCIETY
AAAS

PLEASE CONTACT ME AT:

Name: _____

Address: _____

City: _____ State: _____ Zip code: _____ Country: _____

Email: _____ Phone: _____

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLE SUMMARY

NEURODEVELOPMENT

Mouse and human share conserved transcriptional programs for interneuron development

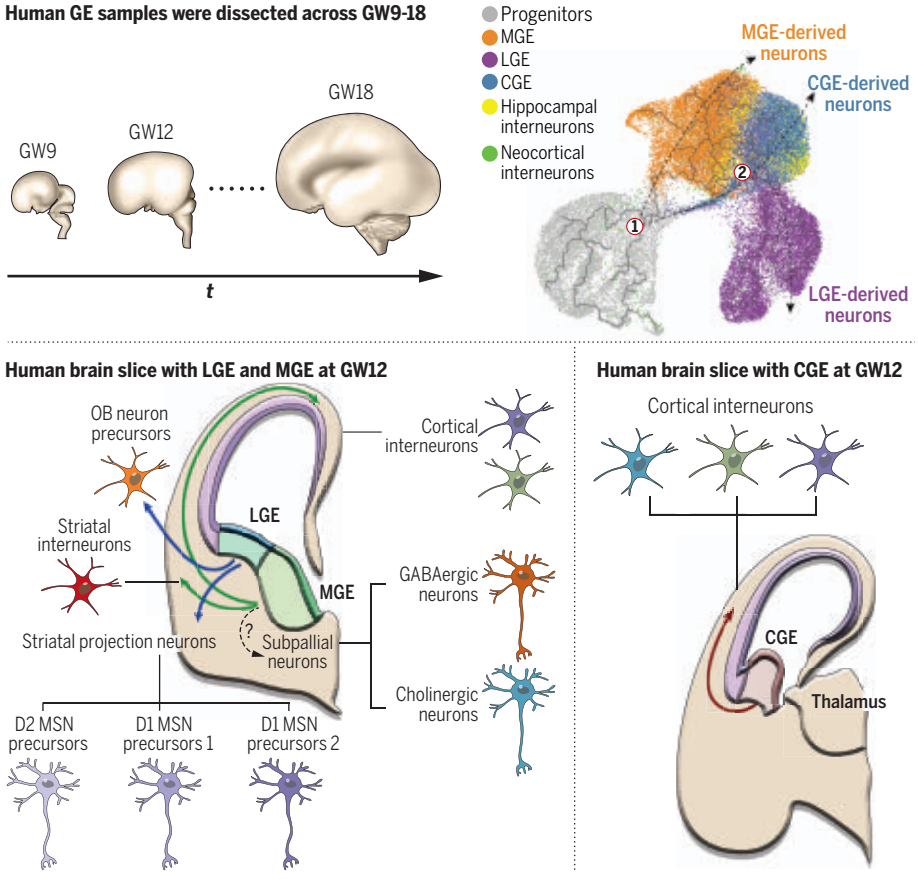
Yingchao Shi†, Mengdi Wang†, Da Mi†, Tian Lu, Bosong Wang, Hao Dong , Suijuan Zhong, Youqiao Chen, Le Sun, Xin Zhou, Qiang Ma, Zeyuan Liu, Wei Wang, Junjing Zhang, Qian Wu*, Oscar Marín*, Xiaoqun Wang*

INTRODUCTION: The cerebral cortex contains two main classes of neurons that derive from distinct structures in the developing telencephalon. Excitatory neurons originate from progenitor cells in the developing pallium, whereas γ-aminobutyric acid–expressing (GABAergic) interneurons derive from the ganglionic eminences, the transitory structures of the fetal brain that also give rise to the basal ganglia. The general organization and cellular architecture of the telencephalon are conserved among mammals, but its size and complexity vary enormously between rodents and humans. The extent to which differences in early regulatory mechanisms governing the development of the telencephalon shape fundamental differences in this brain structure in rodents and humans remains elusive.

RATIONALE: Despite the substantial progress in characterizing the development of excitatory neurons in the human cerebral cortex, our understanding of the generation of interneurons in the human ganglionic eminences is very limited. In this study, we used single-cell RNA sequencing to obtain the transcriptional profiles of 56,412 single cells in the developing human ganglionic eminences from the late first to the early second trimester of human development (gestational weeks 9 to 18) and applied trajectory inference methods to build the developmental trajectories of the main types of neurons generated in the human ganglionic eminences. We also revealed gene regulatory logic that is likely involved in cell fate specification.

RESULTS: We identified molecular features that characterize neural progenitor cells in the human ganglionic eminences. We found that the massive growth of the subventricular zone in the human ganglionic eminences during the second trimester is primarily supported by a large expansion of intermediate progenitor cells. We also revealed the molecular mechanisms underlying the regional specification of progenitor cells as well as the genetic programs driving divergent developmental trajectories in the medial, lateral, and caudal ganglionic eminences (MGE, LGE, and CGE, respectively). In particular, we delineated the developmental trajectories of olfactory bulb neurons, striatal and pallidal GABAergic projection neurons, striatal and cortical GABAergic interneurons, and cholinergic neurons. Despite the protracted development of human cortical interneurons, we found that their diversity is specified within the ganglionic eminences, long before these cells reach the developing cortex. Finally, we identified two populations of human interneurons with features that do not seem to be shared with rodents: a prospective subtype of MGE-derived fast-spiking interneuron and a large population of CGE-derived GABAergic interneurons.

CONCLUSION: Our findings reveal the molecular hierarchies governing the development of neurons generated in the human ganglionic eminences. Our results indicate that gene regulatory logic controlling their specification, migration, and differentiation is evolutionarily conserved in mouse and human. We anticipate that these data will advance our understanding of the regulatory mechanisms underlying human brain development. Considering the involvement of striatal and cortical GABAergic neurons in neurodevelopmental disorders such as autism and schizophrenia, our data should enable linking genetic variation to specific cell types to unravel the origin of neurodevelopmental disorders. ■



Developmental trajectories of neurons derived from the human ganglionic eminences. The main classes of neurons generated from the human medial, lateral, and caudal ganglionic eminences have now been inferred through single-cell transcriptomics. MSN, medium spiny neuron; OB, olfactory bulb; GW, gestational week; *t*, time.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: xiaoqunwang@ibp.ac.cn (X.W.); oscar.marin@kcl.ac.uk (O.M.); qianwu@bnu.edu.cn (Q.W.)
†These authors contributed equally to this work.
Cite this article as Y. Shi *et al.*, *Science* **374**, eabj6641 (2021). DOI: 10.1126/science.abj6641

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abj6641

RESEARCH ARTICLE

NEURODEVELOPMENT

Mouse and human share conserved transcriptional programs for interneuron development

Yingchao Shi[1]†, Mengdi Wang[1,2]†, Da Mi[3,4,5]†, Tian Lu[1,2], Bosong Wang[6], Hao Dong[1,2], Suijuan Zhong[6,7], Youqiao Chen[6], Le Sun[8], Xin Zhou[1], Qiang Ma[1,2], Zeyuan Liu[1,2], Wei Wang[1,2], Junjing Zhang[6], Qian Wu[6,7]*, Oscar Marín[3,4]*, Xiaoqun Wang[1,2,7,8,9]*

Genetic variation confers susceptibility to neurodevelopmental disorders by affecting the development of specific cell types. Changes in cortical and striatal γ-aminobutyric acid–expressing (GABAergic) neurons are common in autism and schizophrenia. In this study, we used single-cell RNA sequencing to characterize the emergence of cell diversity in the human ganglionic eminences, the transitory structures of the human fetal brain where striatal and cortical GABAergic neurons are generated. We identified regional and temporal diversity among progenitor cells underlying the generation of a variety of projection neurons and interneurons. We found that these cells are specified within the human ganglionic eminences by transcriptional programs similar to those previously identified in rodents. Our findings reveal an evolutionarily conserved regulatory logic controlling the specification, migration, and differentiation of GABAergic neurons in the human telencephalon.

The general organization and cellular architecture of the telencephalon are conserved among mammals, but its size and complexity vary enormously between rodents and primates. Between mouse and human, the cerebral cortex differs 1000-fold in size (*1*, *2*) and varies in the types, proportions, and distributions of cells (*3*–*5*). Global transcriptomic analyses have revealed differences in gene expression patterns between mouse and human (*6*–*10*), whereas single-cell transcriptomic analyses have found conservation in the cellular composition of the cerebral cortex. Most cell types found in rodents, monkeys, and humans are homologous (*11*, *12*). This raises the question of whether the distinctive features of the human telencephalon arise through fundamental changes in the gene regulatory networks controlling the development of this brain structure or through alternative mechanisms.

The cerebral cortex contains two main classes of neurons that derive from distinct structures in the developing telencephalon. Excitatory cortical neurons originate from progenitor cells in the developing pallium, whereas γ-aminobutyric acid–expressing (GABAergic) neurons are generated in the ganglionic eminences, the transitory structures of the fetal brain that also give rise to the basal ganglia (*13*, *14*). Although we have made substantial progress in elucidating the development of excitatory neurons in the human cortex (*14*, *15*), our understanding of the generation of GABAergic neurons in the medial, lateral, and caudal ganglionic eminences (MGE, LGE, and CGE, respectively) is very limited (*16*, *17*). For instance, single-cell transcriptomic studies have identified the molecular signatures of glutamatergic lineages in the developing cortex (*18*–*23*), but similar insights into the development of the human ganglionic eminences remain fragmentary (*10*, *24*). In this study, we investigated the transcriptional trajectories of cells in the developing human ganglionic eminences and found conservation in the genetic programs controlling the development of GABAergic neurons in mice and humans. Our study offers insights into the molecular regulation of neurogenesis and the mechanisms underlying the diversification of GABAergic neurons.

Cell diversity in the human ganglionic eminences

We used a droplet-based platform to study the transcriptomic profile of individual cells in the developing human ganglionic eminences. To this end, we dissected the ganglionic eminences from gestational weeks (GW) 9 to 18 (table S1), which overlap with the peak of neurogenesis in this region, and performed single-cell RNA sequencing (scRNA-seq). After quality control (fig. S1A), the transcriptional profiles of 56,412 single cells across GW9 to GW18 were obtained and analyzed collectively using unsupervised clustering (Fig. 1A and fig. S1B). Unsupervised clustering of cellular transcriptional identities by uniform manifold approximation and projection (UMAP) dimensionality reduction revealed the existence of 10 cell clusters (Fig. 1B and table S2), which we annotated using well-known cell type–specific markers (Fig. 1C and fig. S1C). The four main clusters correspond to progenitor cells and postmitotic cells from the MGE, LGE, and CGE (Fig. 1B). In addition, we identified smaller cell clusters containing oligodendrocyte progenitor cells (OPCs), microglia, endothelial cells, thalamic neurons, and pallial cells (Fig. 1B). The last two groups derive from tissues adjacent to the ganglionic eminences that were included in the dissection in the smallest samples (Fig. 1A). We used independent samples from two different stages to confirm data repeatability (fig. S1E) and validated the accuracy of unsupervised clustering using a sample in which the MGE, LGE, and CGE were manually dissected and analyzed independently (Fig. 1B). We also performed differential gene expression analysis to detect the genes that best distinguish dividing progenitors from postmitotic cells in the ganglionic eminences, as well as postmitotic cells from the MGE, LGE, and CGE (Fig. 1D, fig. S1D, and tables S3 and S4). Apart from known markers, we identified *NTRK2* (neurotrophic receptor tyrosine kinase 2), which encodes the BDNF (brain-derived neurotrophic factor) and NT-4 (neutrophin-4) receptor (*25*), as a marker of progenitor cells in the human ganglionic eminences (Fig. 1D). Immunohistochemistry staining confirmed that NTRK2 expression is strong in the progenitor domains of the LGE and CGE (Fig. 1E). We also identified genes that distinguish among postmitotic cells with MGE, LGE, and CGE identity, including multiple previously reported regional markers (Fig. 1D, fig. S1D, and table S3), as well as a small population of putative GABAergic neurons among thalamic cells (fig. S1F). Thus, the analysis of cellular transcriptomes revealed regional identities in the developing human ganglionic eminences along with some molecular signatures for specific cell types.

Conserved genetic regulation of progenitor cells

Neural progenitor cells in the ganglionic eminences (GE progenitors in Fig. 1B) include radial glial cells (RGCs) with neural stem characteristics and intermediate progenitor cells (IPCs), which derive from RGCs and

[1]State Key Laboratory of Brain and Cognitive Science, CAS Center for Excellence in Brain Science and Intelligence Technology (Shanghai), Institute of Biophysics, Chinese Academy of Sciences (CAS), BNU IDG/McGovern Institute for Brain Research, Beijing 100101, China. [2]College of Life Science, University of the Chinese Academy of Sciences, Beijing 100049, China. [3]Centre for Developmental Neurobiology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London SE1 1UL, UK. [4]MRC Centre for Neurodevelopmental Disorders, King's College London, London SE1 1UL, UK. [5]Tsinghua-Peking Center for Life Sciences, IDG/McGovern Institute for Brain Research, School of Life Sciences, Tsinghua University, Beijing 100084, China. [6]State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing 100875, China. [7]Chinese Institute for Brain Research, Beijing 102206, China. [8]Beijing Institute of Brain Disorders, Laboratory of Brain Disorders, Ministry of Science and Technology, Collaborative Innovation Center for Brain Disorders, Capital Medical University, Beijing 100069, China. [9]Guangdong Institute of Intelligence Science and Technology, Guangdong 519031, China.
***Corresponding author. Email: xiaoqunwang@ibp.ac.cn (X.W.); oscar.marin@kcl.ac.uk (O.M.); qianwu@bnu.edu.cn (Q.W.)**
†These authors contributed equally to this work.

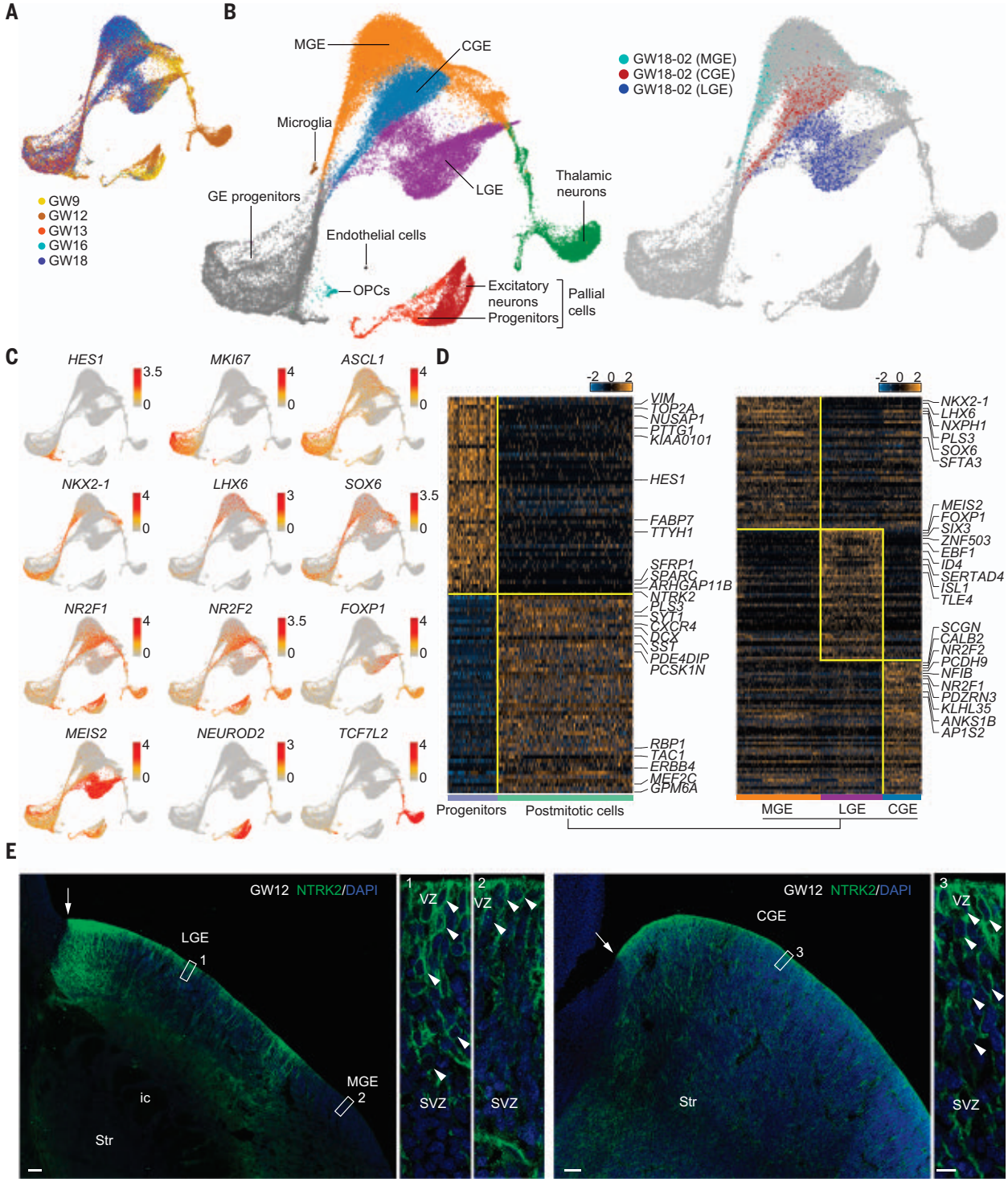


Fig. 1. Major sources of transcriptional heterogeneity among cells from the human ganglionic eminences. (**A**) Clustering of individual cells from different gestational weeks visualized by UMAP. (**B**) Annotation of cell clusters on the basis of gene expression. The schema on the right illustrates the distribution of cells obtained from individual dissections of the MGE, LGE, and CGE in relation to the unsupervised clustering of all cells. (**C**) Gene expression visualized by UMAP. Each dot represents an individual cell colored according to the expression level (red, high; gray, low). (**D**) Heatmaps illustrating DEGs enriched in progenitors and postmitotic cells in the ganglionic eminences (left) and among postmitotic cells in the MGE, LGE, and CGE (right). (**E**) Expression of NTRK2 in progenitor cells in the ganglionic eminences at GW12. The areas in white boxes are shown at high magnification. Scale bars, 200 μm (LGE and MGE), 100 μm (CGE), 10 μm (boxes 1 to 3). Str, striatum; ic, internal capsule; DAPI, 4′,6-diamidino-2-phenylindole.

are committed to the neuronal lineage (*16*). We used a list of established markers to delineate both types of progenitor cells in the human ganglionic eminences and visualized their developmental trajectories using pseudotime alignment (Fig. 2A). We then carried out differential gene expression analysis to unbiasedly identify genes whose expression best distinguishes RGCs and IPCs in the human ganglionic eminences (Fig. 2B, fig. S2A, and table S5). In addition to genes that are expressed in progenitor cells in the mouse ganglionic eminences, such as *NES* (nestin), *VIM* (vimentin), *HES1* [hes family basic helix-loop-helix (bHLH) transcription factor 1], and *ASCL1* (achaete-scute family bHLH transcription factor 1), we also found other genes characteristically enriched in progenitor cells. For example, RGCs express *NTRK2* and *FAM107A* (family with sequence similarity 107 member A), while IPCs express *TMSB10* (thymosin beta 10) and *KPNA2* (karyopherin subunit alpha 2) (Fig. 2, A and B).

Progenitor cells in the human ganglionic eminences are spatially organized into two adjacent niches, a relatively thin ventricular zone (VZ) and a large subventricular zone (SVZ) (fig. S2, B to D). In mice, RGCs and IPCs segregate between the VZ and SVZ, respectively. Analysis of gene expression trajectories across pseudo-lamina and pseudo-differentiation axes revealed a similar distribution for RGCs and IPCs in the human ganglionic eminences (Fig. 2C). This organization contrasts with the developing pallium, in which RGCs expressing *FAM107A* and *HOPX* (HOP homeobox) are the most abundant type of progenitor cells in the SVZ (*26*–*28*). We found that most progenitor cells expressing *FAM107A* are in the VZ of the human ganglionic eminences (Fig. 2C and fig. S2C). Although we also detected a small population of RGCs that coexpress *FAM107A* and *HOPX* (Fig. 2A), most progenitor cells in the human ganglionic eminences have the transcriptional signature of IPCs throughout the peak stages of neurogenesis (Fig. 2A).

We used unsupervised clustering to classify progenitor cells in the human ganglionic eminences into 10 transcriptionally distinctive clusters with different gene expression profiles (Fig. 2D and fig. S3A). We observed that these progenitor clusters could be readily segregated according to their regional identity using the expression of genes involved in the patterning of the ganglionic eminences in mice (*29*). For example, expression of *NKX2-1* (NK2 homeobox 1) and *SOX6* (SRY-box transcription factor 6) characterized progenitor cells in the MGE (Fig. 2, D and E; fig. S3B; and table S6). Expression of *PAX6* (paired box 6) was common to progenitor cells in the LGE and CGE, but CGE progenitors were further characterized by the expression of *PROX1* (prospero homeobox 1) and *NR2F2* (nuclear receptor subfamily 2 group F member 2) (Fig. 2, D and E, and figs. S2D and S3B). Further analysis identified *SIX3* (SIX homeobox 3) as a gene differentially expressed among LGE progenitors (Fig. 2E, fig. S3B, and table S6). We validated this later finding using immunohistochemistry and confirmed that SIX3 is highly enriched among dividing progenitor cells in the SVZ of the human LGE from GW10 to GW16 (Fig. 2, F and G). Because *Six3* plays a role in the generation of striatal medium spiny neuron which originate from LGE progenitors in mice (*30*), our results reinforce the notion that eminence-specific genetic programs similar to those described in rodents seem to be involved in the specification of progenitor cells in the human ganglionic eminences.

Developmental trajectories in the ganglionic eminences

In rodents, the ganglionic eminences give rise to different neuronal populations that populate multiple structures of the adult telencephalon (*30*). The MGE gives rise to GABAergic projection neurons for the globus pallidus, GABAergic interneurons destined for the striatum and cerebral cortex, and cholinergic neurons that remain in the basal telencephalon (*31*–*33*). The LGE primarily produces striatal medium spiny neurons (MSNs) and olfactory bulb (OB) interneurons (*31*, *32*). Finally, the CGE generates GABAergic projection neurons for the amygdala and other limbic system nuclei as well as a diversity of GABAergic interneurons that settle in the cerebral cortex (*34*).

We investigated whether cell diversification among neurons derived from the human ganglionic eminences follows developmental trajectories similar to those described in rodents. We first used a three-dimensional rendering of the distribution of human ganglionic eminence cells in UMAP to identify the relationships between progenitor cells and neurons (fig. S4A and movie S1). We noticed that while most MGE, LGE, and CGE neurons are spatially related to progenitor cells, one group of MGE neurons (which we named MGE-2) was not connected to any group of progenitor cells. This suggested that the progenitor cells of MGE-2 neurons might not have been captured in our dataset, perhaps because they predate the first stage we examined (GW9). We found that MGE-2 neurons express *NKX2-1*, *LHX8* (LIM homeobox 8), *GBX2* (gastrulation brain homeobox 2), and *ISL1* (ISL LIM homeobox 1) (fig. S4B), a combination of genes that are characteristic of subpallial projection neurons, which in the mouse are among the earliest neurons generated in the ganglionic eminences (*35*, *36*). Consistent with this notion, we observed that MGE-2 neurons largely derive from the GW9 and GW12 samples (fig. S4C) and that cells with these features are already present in the human subpallium at GW8 (fig. S4, D and E).

We next integrated our dataset with published scRNA-seq datasets of human neocortical and hippocampal interneurons from GW8 to GW27 (*19*, *20*, *37*), applied trajectory inference methods, and displayed the results via UMAP (Fig. 3, A and B). We excluded MGE-2 neurons from this analysis because their progenitor cells were not likely to be captured in our dataset. We found that most interneurons isolated from the developing neocortex and hippocampus cluster together with MGE and CGE neurons (Fig. 3A and fig. S5A), which is consistent with the view that most cortical interneurons derive from these structures (*17*). Referring to the inferred pseudotime trajectory, we identified the branch points that describe divergences in GE cells on the basis of discrepant gene expression (Fig. 3A and fig. S5B). The results of trajectory inference revealed that the postmitotic cells in the ganglionic eminences first diverge into distinct MGE and LGE or CGE lineages at branch point 1, and that the later lineage subsequently segregates into separate LGE and CGE trajectories at branch point 2 (Fig. 3A).

We then investigated the transcriptional programs driving the divergent developmental trajectories of these cells at each of the branching events. We found pleiotropic genes potentially driving the divergence of cells along each developmental trajectory (Fig. 3C and table S7). For example, *NKX2-1*, *LHX6* (LIM homeobox 6), and *NXPH1* (neurexophilin 1) are enriched in cells that follow the MGE trajectory at branch point 1, whereas *PAX6*, *MEIS2* (Meis homeobox 2), and *NR2F2* are prevalent in cells within the LGE and CGE branch (Fig. 3, D and E, and fig. S5, C and D). We also identified genes enriched in cells that follow LGE [*ZFHX3* (zinc finger homeobox 3), *FOXP1* (forkhead box P1), and *EBF1* (EBF transcription factor 1)] and CGE [*NFIX* (nuclear factor I X), *PROX1*, and *NR2F2*] trajectories at branch point 2, respectively (Fig. 3, D and E, and fig. S5, C and D).

Transcriptional control of cell specification in the LGE

We next sought to unveil the developmental trajectories of neurons generated in specific regions of the human ganglionic eminences. We first classified human postmitotic LGE cells using unsupervised clustering and performed differential gene expression analysis among the seven resulting clusters (Fig. 4A and fig. S6, A and B). We found that clusters L1 and L2 contain cells expressing *ISL1*, *EBF1*, and *TAC1* (tachykinin precursor 1), which are enriched in striatonigral (D1) MSNs, whereas L4 primarily consisted of cells expressing *PENK* (proenkephalin), a marker of striatopallidal (D2) MSNs. In addition, we observed that genes involved in the development of OB interneurons, such as *CHD7* (chromodomain helicase DNA binding protein 7), *ID2* (inhibitor of DNA

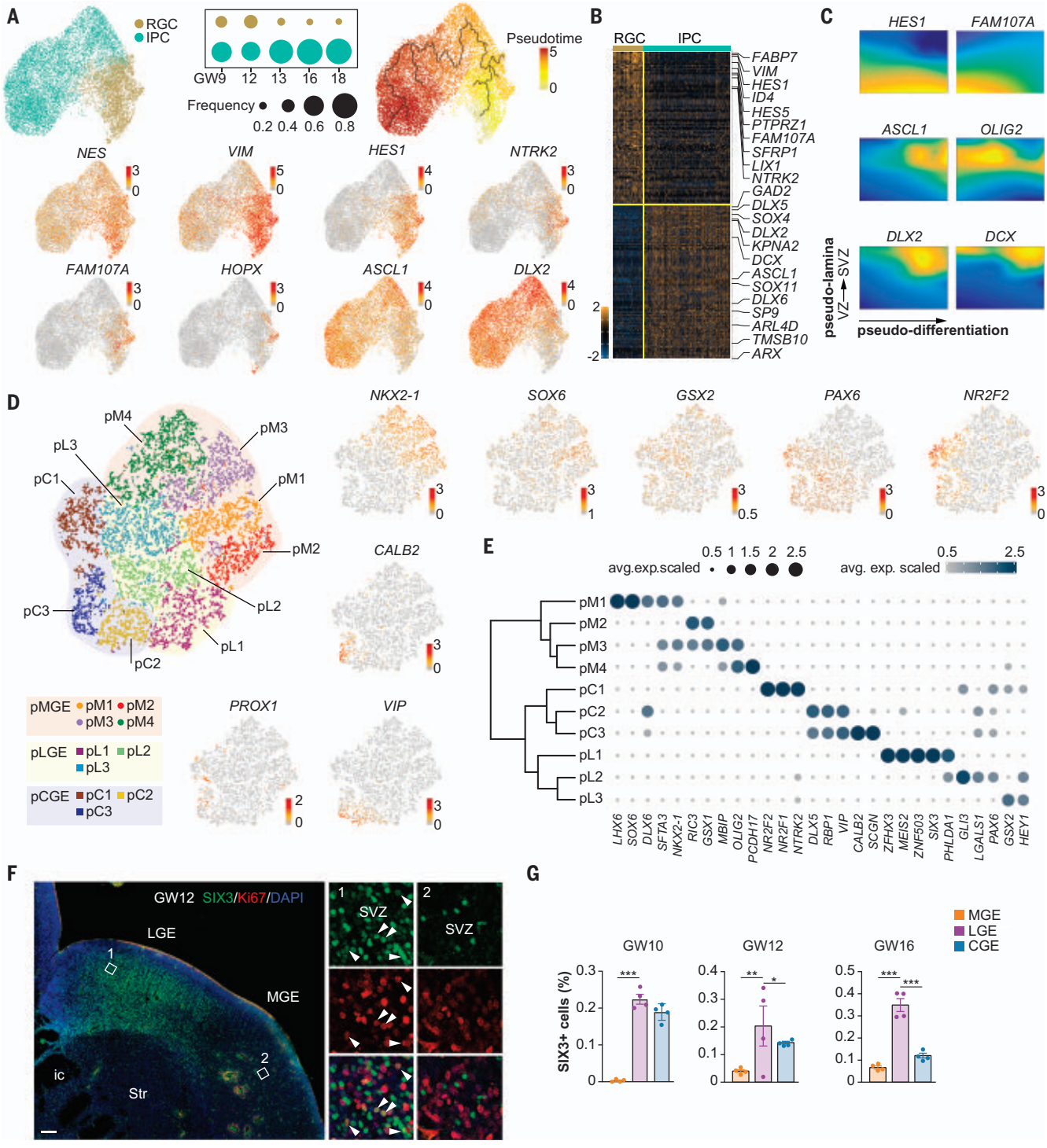


Fig. 2. Cell diversity and genetic regulation of progenitor cells in the human ganglionic eminences. (**A**) Clustering of individual progenitor cells in the human ganglionic eminences (top left) and gene expression visualized by UMAP. Each dot represents an individual cell colored according to the expression level (red, high; gray, low). The developmental trajectory of progenitor cells is analyzed via pseudotime alignment (top right). The inset illustrates the ratio of RGCs and IPCs among ganglionic eminence progenitor cells from GW9 to GW18. (**B**) Heatmap illustrating DEGs that distinguish between RGCs and IPCs in the human ganglionic eminences. (**C**) Gene expression patterns along pseudo-differentiation (*x* axis) and pseudo-lamina (*y* axis) coordinates. (**D**) Cell diversity among progenitor cells in the human ganglionic eminences visualized by t-distributed stochastic neighbor embedding (t-SNE). The regional identity of progenitor cells was established according to their characteristic patterns of gene expression. (**E**) DEGs among progenitor cells. The relationships among subclusters of GE progenitors is illustrated via dendrogram. The size of the dots and the color bar were scaled with the average expression of the corresponding genes. (**F**) Expression of SIX3 and Ki67 in the human LGE and MGE at GW12. The areas in white boxes are shown at high magnification. Scale bars, 200 μm (left), 20 μm (right). (**G**) Quantification of the cell ratio of SIX3+ cells in LGE, MGE, and CGE cells at GW10, GW12, and GW16, respectively. Data are presented as means ± SEM (n = 4 samples from three individual experiments; *P < 0.05, **P < 0.01, ***P < 0.001, one-way analysis of variance).

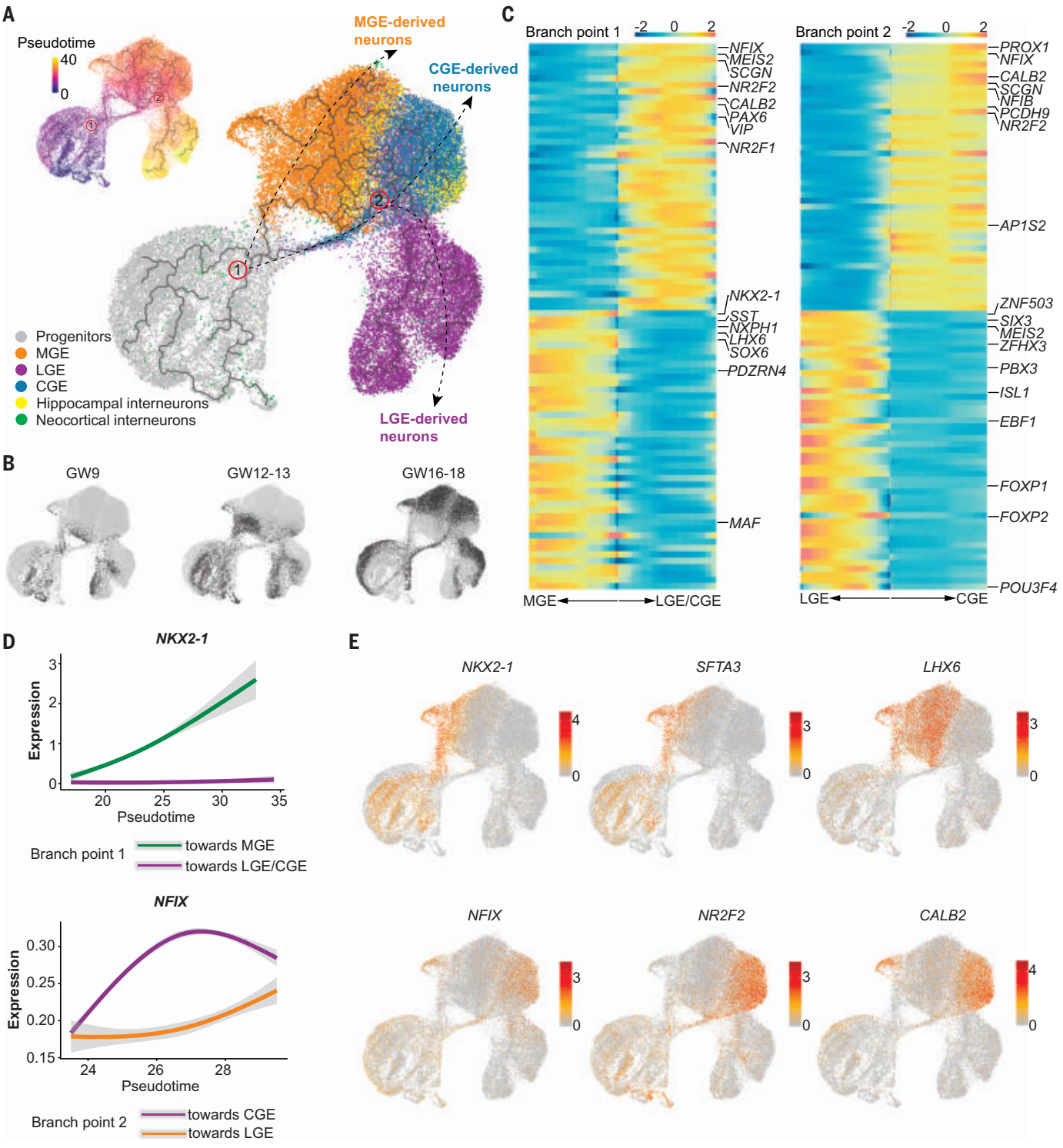


Fig. 3. Transcriptional programs underlying the developmental divergence of human ganglionic eminences. (**A**) The cells of human ganglionic eminences and developing cortical and hippocampal interneurons are integrated and visualized by UMAP with inferred trajectories. MGE-2 cells were excluded from this analysis. Potential differentiation trajectories are schematically depicted with arrows. The pseudotime of ganglionic eminence cells is visualized by UMAP (top left). (**B**) Cell distributions at different gestational stages are shown in dark gray. (**C**) Heatmaps illustrating genes linked to cell fate divergence at branch points 1 (left) and 2 (right). (**D**) The expression of *NKX2-1* and *NFIX* diverged at branch point 1 and 2, respectively. (**E**) The expression profile of genes related to cell fate divergence at branch points 1 and 2 is visualized by UMAP. Each dot represents an individual cell colored according to the expression level (red, high; gray, low).

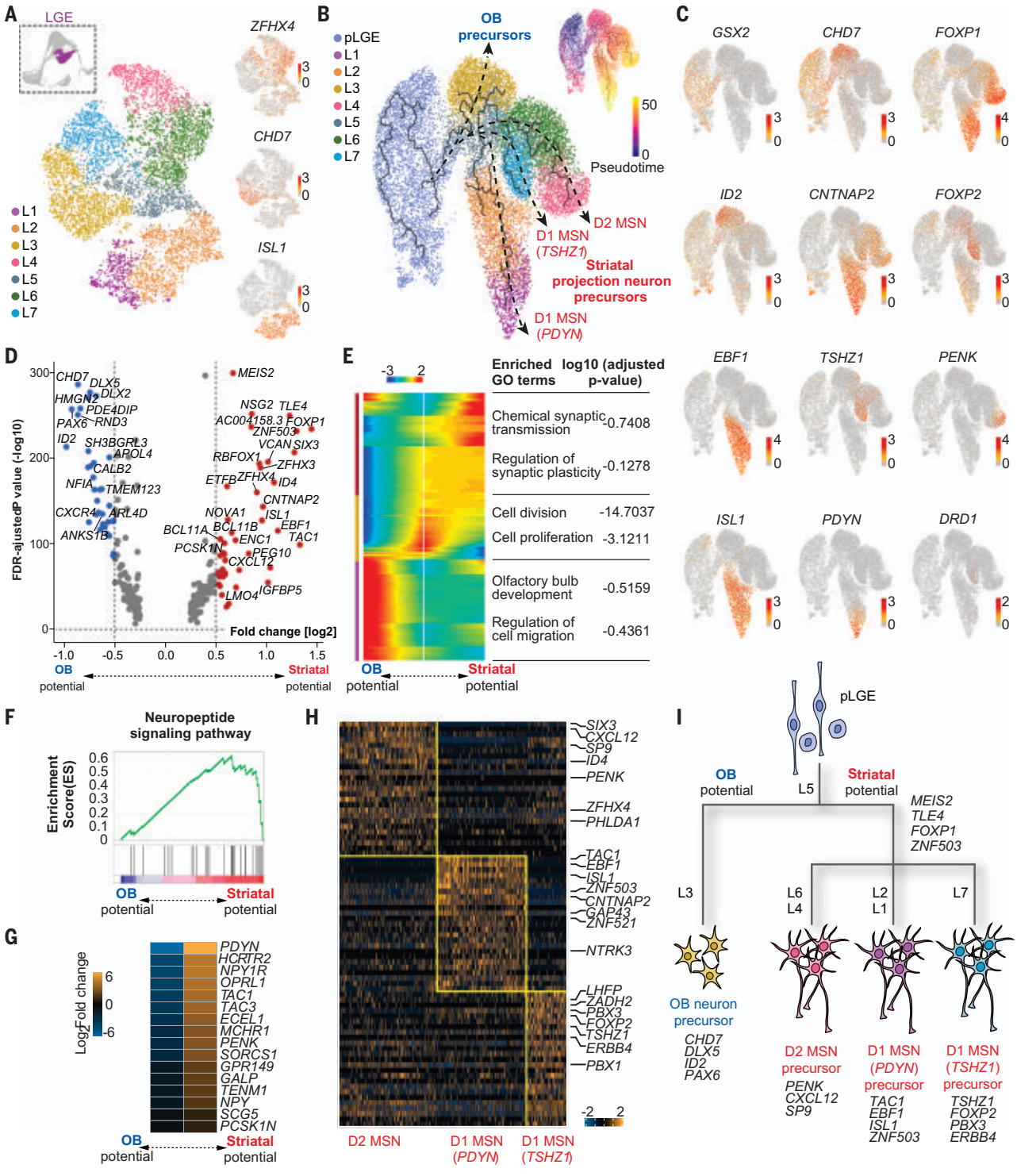


Fig. 4. Transcriptional regulation of LGE development. (**A**) Unsupervised clustering of human LGE postmitotic cells and gene expression patterns visualized by t-SNE. Each dot represents an individual cell colored according to the expression level (red, high; gray, low). (**B**) Developmental trajectory of LGE cells visualized by UMAP. Pseudotime for individual cells is also shown (top right). (**C**) Gene expression profile along the developmental trajectories of LGE cells visualized via UMAP. (**D**) Volcano plot of DEGs for LGE cells with OB and striatal potential. (**E**) Heatmap illustrating the bifurcation of gene expression along the developmental trajectory of LGE cells committed to OB and striatal fates. (**F** and **G**) The neuropeptide signaling pathway is enriched in LGE cells with striatal potential. (**H**) Heatmap illustrating DEGs enriched in *TSHZ1*+ D1 MSN, *PDYN*+ D1 MSNs, and D2 MSNs. (**I**) Schematic of hypothetical genetic programs underlying the early diversification of human LGE cells.

binding 2), and *PAX6*, were enriched among L3 cells (Fig. 4A and fig. S6, A and B).

We combined LGE progenitors and postmitotic cells to delineate their developmental trajectories and displayed the results using UMAP (Fig. 4B and fig. S6C). The results of this analysis revealed an early bifurcation of OB and striatal neuron fates, which are delineated by characteristic patterns of gene expression (Fig. 4, B and C). We performed differential gene expression analysis to identify genes potentially driving the divergence of OB and striatal lineages (Fig. 4D and table S8). Gene Ontology (GO) analysis of the differentially expressed genes (DEGs) revealed that GO terms associated with synaptic transmission and plasticity are enriched in LGE cells with striatal potential, whereas GO terms associated with OB development and cell migration are enriched in cells with OB potential (Fig. 4E).

We performed gene set enrichment analysis and found that the neuropeptide signaling pathway was distinctively enriched in the LGE cells with striatum potential (Fig. 4, F and G). We also observed that striatal-specific genes such as *EBF1* and *ISL1* were expressed at earlier stages of development than OB-specific genes such as *CHD7* (fig. S6D), suggesting that LGE cells with striatal potential mature earlier than cells with OB potential. To test this hypothesis, we defined a "maturation score" and mapped the trajectory of LGE postmitotic cells along this inferred trajectory (fig. S6E). This analysis confirmed higher maturation levels for cells with striatum potential than those with OB potential. Further analysis of the developmental trajectories of LGE cells with striatal potential revealed early divergence in three distinct fates, *TSHZ1*+ (teashirt zinc finger homebox 1)+ D1, *PDYN*+ (prodynorphin)+ D1, and D2 MSNs (Fig. 4B), each with a distinctive pattern of gene expression (Fig. 4H). Altogether, our analysis revealed the genetic programs underlying the early diversification of human LGE cells (Fig. 4I).

Transcriptional control of cell specification in the MGE

To explore neuronal diversity among newborn MGE cells, we classified postmitotic cells from this region using unsupervised clustering (Fig. 5A) and performed differential gene expression analysis among the seven resulting clusters (Fig. 5B and fig. S7, A and B). We found distinctive patterns of gene expression among these cell populations. For example, the largest cluster, M2, contains cells expressing *LHX6* and *SOX6*, two transcription factors that are critical for the development of MGE-derived cortical interneurons (*38*–*41*), and *CXCR4* (C-X-C motif chemokine receptor 4) and *ERBB4* (erb-b2 receptor tyrosine kinase 4), which encode guidance receptors regulating the tangential migration of interneurons from the ganglionic eminences toward the developing cortex (*42*–*46*). In contrast, M5 contains cells expressing *NKX2-1*, *LHX8*, *ISL1*, and *GBX2*, which are characteristic of cholinergic neurons in the subpallium (Fig. 5B) (*33*).

To determine developmental relationships among these clusters, we integrated MGE progenitors and postmitotic neurons with the datasets of developing human cortical and hippocampal interneurons, applied trajectory inference methods (with the exception of MGE-2 cells), and displayed the results using UMAP (Fig. 5C). This analysis revealed that M1 primarily contains undifferentiated precursors that are transitioning toward the acquisition of a distinct cell fate (Fig. 5C). The remaining clusters segregated along two separate trajectories. Branch 1 primarily contained cells from cluster M2 along with most of the developing neocortical and hippocampal interneurons (Fig. 5, C and D). Branch 2 consisted of the small M3 cluster, which contains cells expressing *ETV1* (ETS variant transcription factor 1), *CRABP1* (cellular retinoic acid binding protein 1), and *ANGPT2* (angiopoietin 2) (Fig. 5, C and D, and fig. S7B). This cluster also exhibited high levels of *CXCR4* and *ERBB4* expression (Fig. 5D), suggesting that it may include neurons tangentially migrating to the cortex and consistent with the fact that some more-mature neocortical and hippocampal interneurons also mapped to this branch (Fig. 5C). We also analyzed cell diversity within MGE-2 cells, which comprised a heterogeneous group of cells from the M4, M5, M6, and M7 clusters. We found very few neocortical and hippocampal interneurons among MGE-2 cells (Fig. 5C), which reinforced the idea that cells in these clusters are fated to develop into subpallial neurons. Consistently, we found that these cells segregated into prospective GABAergic (M4 and M7) and cholinergic (M5 and M6) fates with *GAD1* (glutamate decarboxylase 1) and *LHX8* expression, respectively (Fig. 5, D and E), which likely correspond to globus pallidus and cholinergic projection neurons of the basal telencephalon, along with a small number of striatal interneurons. Further analysis of differential gene expression revealed distinctive transcriptional programs that underlie the main developmental trajectories of diversification of human MGE neurons (Fig. 5E).

Early specification of cortical interneurons

As many 60 different transcriptional identities have been identified among GABAergic neurons in the cerebral cortex of mice and humans (*11*, *12*). To investigate the developmental mechanisms underlying the emergence of interneuron diversity in the adult human neocortex, we focused on the MGE, which gives rise to two main subclasses of cortical GABAergic neurons, parvalbumin-expressing (PV+) and somatostatin-expressing (SST+) cells (*47*). To this end, we integrated cells from the previously identified embryonic M2 and M3 clusters, which likely contain most of MGE-derived tangentially migrating interneuron precursors, with *LHX6*+ cells from two published single-nucleus RNA-seq datasets of the adult human cortex (*11*, *12*) (fig. S8A), and visualized the combined dataset using UMAP (Fig. 5F and fig. S8B). Gene expression profiles indicated that adult *LHX6*+ neurons were allocated to three main subclasses—LHX6+ LAMP5 (lysosomal associated membrane protein family member 5)+, PV+, and SST+ cells (fig. S8B)—which could be further subdivided into multiple types, including PV+ basket cells, PV+ chandelier cells, SST+ Martinotti cells, SST+ non-Martinotti cells, and SST+ long-range projection neurons (Fig. 5F). We then used this classification of adult neurons to annotate M2 and M3 MGE cells on the basis of their transcriptional similarity and found that most of these cells could be assigned to one of the six types identified in the adult cortex (Fig. 5F). Differential gene expression analysis among the embryonic neurons fated to become distinct subclasses of PV+ and SST+ neurons led to the identification of early developmental markers for these populations, many of which continued to be expressed in the corresponding adult interneurons (fig. S8C). This analysis suggested that interneuron diversification, at least at the level of the main interneuron types, is evident in the human MGE long before these cells reach the developing cortex.

We next searched for evidence of fate specification at the level of individual subtypes among human MGE cells. To this end, we focused on SST+ neurons, as they are perhaps the best characterized among the different transcriptional identities observed in the adult cortex (*11*, *12*). We used unsupervised clustering to classify adult SST+ neurons into 13 distinct transcriptional identities: one subtype of long-range projection neuron, six subtypes of Martinotti cells, and six subtypes of non-Martinotti cells (fig. S9, A to F). We then integrated the adult and embryonic datasets and assigned prospective identities to embryonic SST+ cells on the basis of their transcriptional similarity with adult SST+ interneurons (fig. S9G). Using this approach, we identified 11 distinct transcriptional identities (MGE_{SST} 1 to 11) among the embryonic SST+ interneurons that exhibit different patterns of gene expression (fig. S9, G and H). We also conducted an independent, pairwise comparison analysis to define homologies among the embryonic and adult SST+ identities. This analysis revealed that 7 of the 11 embryonic SST+ transcriptional identities identified in the MGE (MGE_{SST} 1, 2, 3, 5, 6, 7, and 10) matched one-to-one with adult SST+ transcriptional subtypes, including one subtype of long-range projection neuron, four Martinotti subtypes, and two non-Martinotti subtypes (Fig. 5G). Our analysis also revealed that four of

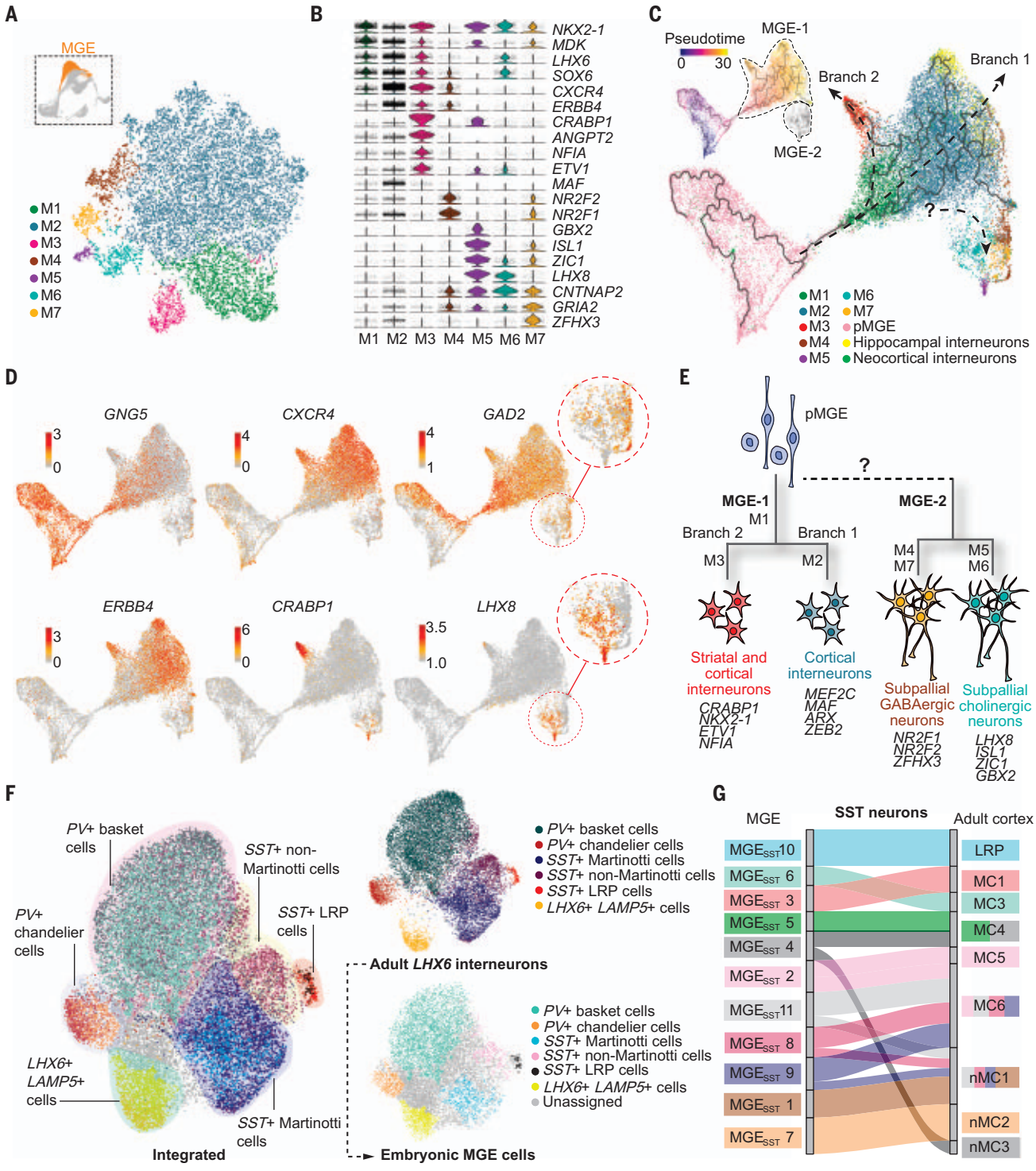


Fig. 5. Transcriptional control of early cell specification in the human MGE. (**A**) Neuronal diversity of postmitotic cells in human MGE visualized via t-SNE. (**B**) Violin plots of DEGs among MGE clusters. (**C**) Developmental trajectory of MGE cells are inferred via monocle analysis (with the exception of MGE-2 cells) and visualized by UMAP. Pseudotime of MGE cells in branches 1 and 2 is shown (top left). (**D**) Gene expression along the developmental trajectories of MGE cells is visualized via UMAP. MGE-2 cells comprise GABAergic and cholinergic subpallial neurons, according to the expression of *GAD2* and *LHX8*, respectively. Each dot represents an individual cell and colored according to the expression level (red, high; gray, low). (**E**) Schematic of hypothetical genetic programs underlying the early diversification of human MGE cells. The developmental trajectory linking pMGE (progentiors of the MGE) to MGE-2 cells is uncertain (dotted line) owing to the lack of related progenitor cells in our dataset. (**F**) Integration of postmitotic human MGE cells (M2 and M3) and *LHX6*+ adult human cortical interneurons is visualized by UMAP. MGE cells were annotated according to the classification of adult cortical interneurons on the basis of transcriptional similarities. (**G**) Riverplot illustrating the relationship between MGE cell clusters and adult SST+ interneurons. The size of the bars of MGE cells is normalized to cell numbers. LRP, long-range projection neurons; MC, Martinotti cells; nMC, non-Martinotti cells.

the embryonic SST+ transcriptional identities (MGE_{SST} 4, 8, 9, and 11) share similarities with both the Martinotti cell and non-Martinotti cell lineages (Fig. 5G), perhaps reflecting a relatively immature state. Furthermore, we could not identify embryonic correlates of four adult SST+ transcriptional identities, which are primarily defined by the expression of ion channel genes barely detectable in embryonic interneurons (fig. S9, E and F). Altogether, these results suggest that individual interneuron cell types—as defined by their transcriptional identity—are specified shortly after birth in the human embryonic MGE.

Distinctive features in developing human interneurons

Despite sharing many cellular features with other species, human cortical interneurons also exhibit some distinctive traits, including morphological and gene expression specializations (*11*, *12*). To investigate whether any of these distinctive characteristics are already embedded in the ganglionic eminences during embryonic development, we integrated our scRNA-seq dataset of human ganglionic eminence cells (GW9 to GW18) with two published datasets of mouse ganglionic eminence cells from embryonic days 12.5 to 14.5 (*48*, *49*) (fig. S10, A and B) and identified 19 cell groups using unsupervised clustering (Fig. 6A, fig. S10C, and table S9). We then calculated the normalized cell ratio of human and mouse cells for each cluster and detected two clusters, 1 and 19, that primarily consisted of human CGE and MGE cells, respectively (Fig. 6B and fig. S10D). We analyzed gene expression in these two predominantly human cell groups and found that cells in each of these clusters exhibit specific markers, such as *SCGN* (secretagogin, EF-hand calcium binding protein), *CALB2* (calbindin 2) and *NR2F2* in cluster 1, and *CRABP1*, *ETV1*, and *NKX2-1* in cluster 19 (Fig. 6C).

We classified human postmitotic CGE cells into three main groups (C1 to C3) using unsupervised clustering (fig. S10E). Cell constitution analysis revealed that most CGE cells at these stages belong to C1 and C2, which exhibit the highest levels of *SCGN* and *CALB2* expression and correspond to cluster 1 (fig. S10F). Immunohistochemical analyses confirmed that cells expressing secretagogin and calretinin, the proteins encoded by *SCGN* and *CALB2*, respectively, concentrate in the SVZ of the human CGE, where they often colocalize with NR2F2 (Fig. 6D and fig. S11A). In contrast, we detected very few SCGN+ cells in the mouse CGE (fig. S11B). In the adult human cortex, secretagogin-expressing neurons are GABAergic and often contain calretinin (Fig. 6E and fig. S11C).

We performed cell constitution analysis for cluster 19 and found that it mainly consisted of M3 cells, characterized by the expression of *CRABP1* and *NKX2-1* (fig. S10G). At GW16, cells coexpressing CRABP1 and NKX2-1, with the typical morphology of migrating interneurons (*46*, *50*), were found in the striatum and en route to the pallium (Fig. 6F). In addition, CRABP1/NKX2-1+ cells were also found in the developing human neocortex (Fig. 6F and fig. S12A). Crabp1/Nkx2-1+ cells could be detected in the mouse subpallium but not in the developing or adult cortex (fig. S12, B and C). In contrast, CRABP1+ cells coexpressing PV were detected in the adult human cortex (Fig. 6G), which suggests that they may represent a population of fast-spiking interneurons.

Discussion

In this study, we investigated the transcriptional identity of cells in the developing human ganglionic eminences as well as the gene regulatory networks controlling their fate specification. Using single-cell transcriptomics and trajectory inference methods, we built spatiotemporal maps of gene expression through the early second trimester of human development (GW9 to GW18) that recapitulate the developmental trajectories of the main classes of neurons generated in the subpallium, including OB neurons, striatal and pallidal GABAergic projection neurons, cholinergic projection neurons and interneurons, and striatal and cortical GABAergic interneurons. Our results revealed conserved mechanisms underlying the early diversification of subpallial neurons in mouse and human. Although the specific expression of some key factors in the developing human brain, such as *NKX2-1*, *SIX3*, *SCGN*, *CALB2*, *CRABP1*, *ISL1*, and *LHX8*, has been validated via immunostaining, further experiments should investigate their precise involvement in this process.

Newborn neurons are generated from RGCs and IPCs in both ganglionic eminences and developing cortex. The larger primate neocortex is due to a greater number of RGCs, especially basal RGCs in the SVZ, which in rodents is mostly populated by IPCs (*14*, *51*, *52*). Although it has been suggested that a shift toward more basal RGCs may also drive the greater neurogenesis in human than in mouse ganglionic eminences (*16*), we found that the massive growth of the human ganglionic eminence SVZ during the second trimester is primarily due to greater numbers of IPCs. Thus, for the human, the main progenitor cell populations driving neurogenesis in the ganglionic eminences are different from that in the developing cortex.

Previous work has shown that the expression patterns of several transcription factors in the primate ganglionic eminences are very similar to those found in rodents (*17*, *29*). Our transcriptomic analysis reveals that these conserved features extend beyond a handful of key factors and include the complex gene regulatory networks controlling neuronal specification in the human ganglionic eminences. Cortical interneurons in the human ganglionic eminences express the same receptors that have been shown to regulate the tangential migration of interneurons in rodents (*53*). This reveals that the conserved features of human and mouse interneuron development extend beyond the early specification of progenitor cells and involve essential aspects of their subsequent migration and differentiation. Although the roles of some key transcription factors in orchestrating interneuron development have been studied in the mouse, their function in human interneuron development requires experimental validation.

Cortical interneurons are a heterogeneous group of neurons with diverse morphologies, connectivity, biochemistry, and physiological properties (*54*, *55*). In mice, newborn interneurons are transcriptionally heterogeneous within a few hours of becoming postmitotic, and these early transcriptional signatures correlate with those found in specific types of interneurons in the adult cortex (*49*). Our study reveals that interneuron diversification also begins in the human ganglionic eminences, long before these cells reach the developing cortex. For instance, many of the transcriptional identities found among SST+ interneurons in the adult human cortex are identifiable in the MGE. In some cases, however, the transcriptional signature of developing interneurons does not allow a clear correlation with specific adult identities. This is the case for newborn CGE interneurons, most of which are jointly characterized by the expression of *CALB2* and *SCGN*, much like developing CGE interneurons in rodents are characterized by the expression of *Htr3a* (5-hydroxytryptamine receptor 3A) (*56*). *SCGN* encodes a calcium-binding protein similar in sequence and structure to calbindin and calretinin that seems particularly abundant in the developing human brain (*57*, *58*). In the adult human cortex, *SCGN* expression seems to be confined to a small population of GABAergic interneurons whose function remains to be established. We also identified a cluster of human MGE-derived interneurons characterized by the expression of *CRABP1*, which does not seem to have a clear counterpart in rodents. In the mouse, *Crabp1* is expressed by some fast-spiking interneurons in the developing striatum (*59*). In the human, we found that migrating CRABP1-expressing interneurons end up in the adult human cortex and coexpress NKX2-1 and PV. This observation suggests that CRABP1+ cells may represent a population of MGE-derived fast-spiking interneurons in the human cortex.

Although the general cellular architecture of inhibitory cell types is largely conserved, extensive differences seem to exist in the relative proportions, laminar distributions, and gene expression patterns of specific types of

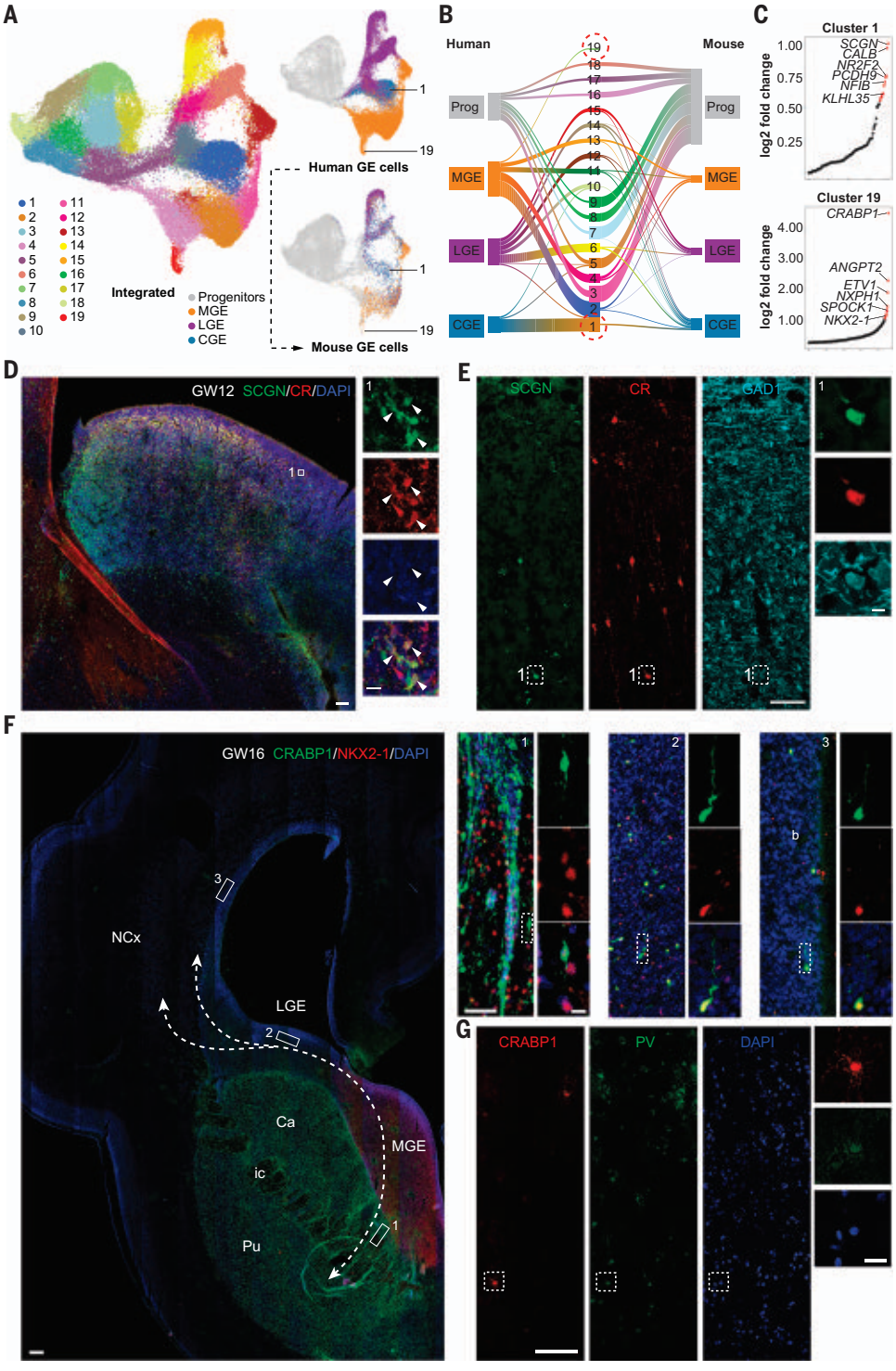


Fig. 6. Distinctive features of human ganglionic eminences. (**A**) The datasets of human and mouse ganglionic eminence cells are integrated on the basis of shared sources of variation and visualized by UMAP. (**B**) Riverplot illustrating relationships between human and mouse ganglionic eminence cell clusters. Two human-specific clusters (1 and 19) are highlighted by circles. (**C**) Specific gene expression in human-specific clusters 1 and 19. (**D**) Expression of SCGN and CR in the human CGE at GW12. The area in the white box is shown at high magnification. Scale bars, 100 μm (left) and 20 μm (right). (**E**) Expression of SCGN, CR, and GAD1 in the adult human cortex. The area in the white box is shown at high magnification. Scale bars, 100 μm (left) and 10 μm (right). (**F**) Expression of CRABP1 and NKX2-1 in the embryonic human brain at GW16. The regions in the white boxes are shown at high magnification. The dotted lines illustrate potential migratory routes for CRABP1+ and NKX2-1+ cells. Scale bars, 500 μm (left), 50 μm (boxes 1 to 3, right), 10 μm (boxes 1 to 3, left). NCx, neocortex; Pu, putamen; Ca, caudate. (**G**) Expression of CRABP1 and PVALB in the adult human cortex. The area in the white box is shown at high magnification.

interneurons in the adult cortex of mice and humans. For instance, *LHX6*+ *LAMP5*+ interneurons are much more abundant in the human than in the mouse adult cortex (*11*, *12*). Our findings reveal that human *LHX6*+ *LAMP5*+ interneurons originate in the MGE like their mouse counterparts and shared a common developmental trajectory, which suggests that the differences observed between mouse and human may arise through changes in the dynamics of neurogenesis (*16*) or in the interaction of developing interneurons with the local microenvironment, for instance by means of programmed cell death (*60*).

Neurodevelopmental disorders have overlapping phenotypes and genetics, suggestive of common deficits. Changes in striatal and cortical GABAergic neurons have been extensively documented in autism and schizophrenia (*61*), yet it is presently unclear whether genetic variation associated with these disorders converges on specific types of inhibitory cells because we lack a detailed understanding of their transcriptional trajectories in the developing human brain. Our study sheds light on the development of human GABAergic neurons and should enable the linking of gene

variation and cell types in neurodevelopmental disorders.

Materials and methods summary

Detailed materials and methods are provided in the supplementary materials. In brief, human fetal ganglionic eminence (GE) samples across gestational weeks (GW) 9 to 18 were collected from Beijing Anzhen Hospital with approval from the Reproductive Study Ethics Committee of Beijing Anzhen Hospital and the institutional review board (ethics committee) of the Institute of Biophysics. Single-cell RNA-sequencing was performed with an Illumina sequencing platform. The outcome reads were aligned to the human reference genome hg19 with Cell Ranger (10X genomics). Doublets were removed by performing the Scrublet pipeline on the raw gene-by-cell expression matrix from each sample (*62*). Batch correction was conducted on the reduced dimensions with the R function of fastMNN (*63*). Seurat was adopted to perform normalization, dimension reduction, unsupervised clustering, and differentially expressed genes identification (*64*, *65*). The Seurat integration method was used to determine the developmental divergence of human GE cells, the early specification of cortical interneurons in human GE, as well as distinctive features of developing human interneurons. The developmental trajectory of human GE cells was constructed using the R package of monocle 3 (*66*–*68*). K-nearest neighbors analysis (knn) was used to render the putative interneuron identities to MGE cells. Immunofluorescence staining was performed on human and mouse brain slices and images were acquired with an Olympus confocal microscope.

REFERENCES AND NOTES

1. S. Herculano-Houzel, The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. *Proc. Natl. Acad. Sci. U.S.A.* **109** (suppl 1.), 10661–10668 (2012). doi: 10.1073/pnas.1201895109; pmid: 22723358
2. J. DeFelipe, The evolution of the brain, the human nature of cortical circuits, and intellectual creativity. *Front. Neuroanat.* **5**, 29 (2011). doi: 10.3389/fnana.2011.00029; pmid: 21647212
3. J. DeFelipe, L. Alonso-Nanclares, J. I. Arellano, Microstructure of the neocortex: Comparative aspects. *J. Neurocytol.* **31**, 299–316 (2002). doi: 10.1023/A:1024130211265; pmid: 12815249
4. E. Boldog *et al*., Transcriptomic and morphophysiological evidence for a specialized human cortical GABAergic cell type. *Nat. Neurosci.* **21**, 1185–1195 (2018). doi: 10.1038/s41593-018-0205-2; pmid: 30150662
5. N. A. Oberheim *et al*., Uniquely hominid features of adult human astrocytes. *J. Neurosci.* **29**, 3276–3287 (2009). doi: 10.1523/JNEUROSCI.4707-08.2009; pmid: 19279265
6. H. Zeng *et al*., Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell* **149**, 483–496 (2012). doi: 10.1016/j.cell.2012.02.052; pmid: 22500809
7. T. E. Bakken *et al*., A comprehensive transcriptional map of primate brain development. *Nature* **535**, 367–375 (2016). doi: 10.1038/nature18637; pmid: 27409810
8. M. Hawrylycz *et al*., Canonical genetic signatures of the adult human brain. *Nat. Neurosci.* **18**, 1832–1844 (2015). doi: 10.1038/nn.4171; pmid: 26571460
9. Y. Zhu *et al*., Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, eaat8077 (2018). doi: 10.1126/science.aat8077; pmid: 30545855
10. A. M. M. Sousa *et al*., Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358**, 1027–1032 (2017). doi: 10.1126/science.aan3456; pmid: 29170230
11. R. D. Hodge *et al*., Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68 (2019). doi: 10.1038/s41586-019-1506-7; pmid: 31435019
12. F. M. Krienen *et al*., Innovations present in the primate interneuron repertoire. *Nature* **586**, 262–269 (2020). doi: 10.1038/s41586-020-2781-z; pmid: 32999462
13. O. Marín, J. L. Rubenstein, A long, remarkable journey: Tangential migration in the telencephalon. *Nat. Rev. Neurosci.* **2**, 780–790 (2001). doi: 10.1038/35097509; pmid: 11715055
14. J. H. Lui, D. V. Hansen, A. R. Kriegstein, Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011). doi: 10.1016/j.cell.2011.06.030; pmid: 21729779
15. T. Namba, J. Nardelli, P. Gressens, W. B. Huttner, Metabolic regulation of neocortical expansion in development and evolution. *Neuron* **109**, 408–419 (2021). doi: 10.1016/j.neuron.2020.11.014; pmid: 33306962
16. D. V. Hansen *et al*., Non-epithelial stem cells and cortical interneuron production in the human ganglionic eminences. *Nat. Neurosci.* **16**, 1576–1587 (2013). doi: 10.1038/nn.3541; pmid: 24097039
17. T. Ma *et al*., Subcortical origins of human and monkey neocortical interneurons. *Nat. Neurosci.* **16**, 1588–1597 (2013). doi: 10.1038/nn.3536; pmid: 24097041
18. T. J. Nowakowski *et al*., Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017). doi: 10.1126/science.aap8809; pmid: 29217575
19. X. Fan *et al*., Single-cell transcriptome analysis reveals cell lineage specification in temporal-spatial patterns in human cortical development. *Sci. Adv.* **6**, eaaz2978 (2020). doi: 10.1126/sciadv.aaz2978; pmid: 32923614
20. S. Zhong *et al*., Decoding the development of the human hippocampus. *Nature* **577**, 531–536 (2020). doi: 10.1038/s41586-019-1917-5; pmid: 31942070
21. S. Zhong *et al*., A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018). doi: 10.1038/nature25980; pmid: 29539641
22. S. Mayer *et al*., Multimodal Single-Cell Analysis Reveals Physiological Maturation in the Developing Human Neocortex. *Neuron* **102**, 143–158.e7 (2019). doi: 10.1016/j.neuron.2019.01.027; pmid: 30770253
23. Z. Li *et al*., Transcriptional priming as a conserved mechanism of lineage diversification in the developing mouse and human neocortex. *Sci. Adv.* **6**, eabd2068 (2020). doi: 10.1126/sciadv.abd2068; pmid: 33158872
24. V. D. Bocchi *et al*., The coding and long noncoding single-cell atlas of the developing human fetal striatum. *Science* **372**, eabf5759 (2021). doi: 10.1126/science.abf5759; pmid: 33958447
25. M. Bothwell, Functional interactions of neurotrophins and neurotrophin receptors. *Annu. Rev. Neurosci.* **18**, 223–253 (1995). doi: 10.1146/annurev.ne.18.030195.001255; pmid: 7605062
26. S. A. Fietz *et al*., OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nat. Neurosci.* **13**, 690–699 (2010). doi: 10.1038/nn.2553; pmid: 20436478
27. D. V. Hansen, J. H. Lui, P. R. Parker, A. R. Kriegstein, Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554–561 (2010). doi: 10.1038/nature08845; pmid: 20154730
28. A. A. Pollen *et al*., Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015). doi: 10.1016/j.cell.2015.09.004; pmid: 26406371
29. N. Flames *et al*., Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *J. Neurosci.* **27**, 9682–9695 (2007). doi: 10.1523/JNEUROSCI.2750-07.2007; pmid: 17804629
30. Z. Xu *et al*., SP8 and SP9 coordinately promote D2-type medium spiny neuron production by activating *Six3* expression. *Development* **145**, dev165456 (2018). doi: 10.1242/dev.165456; pmid: 29967281
31. L. Sussel, O. Marin, S. Kimura, J. L. Rubenstein, Loss of Nkx2.1 homeobox gene function results in a ventral to dorsal molecular respecification within the basal telencephalon: Evidence for a transformation of the pallidum into the striatum. *Development* **126**, 3359–3370 (1999). doi: 10.1242/dev.126.15.3359; pmid: 10393115
32. H. Wichterle, J. M. Garcia-Verdugo, D. G. Herrera, A. Alvarez-Buylla, Young neurons from medial ganglionic eminence disperse in adult and embryonic brain. *Nat. Neurosci.* **2**, 461–466 (1999). doi: 10.1038/8131; pmid: 10321251
33. Y. Zhao *et al*., The LIM-homeobox gene *Lhx8* is required for the development of many cholinergic neurons in the mouse forebrain. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9005–9010 (2003). doi: 10.1073/pnas.1537759100; pmid: 12855770
34. S. Nery, G. Fishell, J. G. Corbin, The caudal ganglionic eminence is a source of distinct cortical and subcortical cell populations. *Nat. Neurosci.* **5**, 1279–1287 (2002). doi: 10.1038/nn971; pmid: 12411960
35. O. Marín, S. A. Anderson, J. L. R. Rubenstein, Origin and molecular specification of striatal interneurons. *J. Neurosci.* **20**, 6063–6076 (2000). doi: 10.1523/JNEUROSCI.20-16-06063.2000; pmid: 10934256
36. S. Nóbrega-Pereira *et al*., Origin and molecular specification of globus pallidus neurons. *J. Neurosci.* **30**, 2824–2834 (2010). doi: 10.1523/JNEUROSCI.4023-09.2010; pmid: 20181580
37. X. Fan *et al*., Spatial transcriptomic survey of human embryonic cerebral cortex by single-cell RNA-seq analysis. *Cell Res.* **28**, 730–745 (2018). doi: 10.1038/s41422-018-0053-3; pmid: 29867213
38. G. Neves *et al*., The LIM homeodomain protein Lhx6 regulates maturation of interneurons and network excitability in the mammalian cortex. *Cereb. Cortex* **23**, 1811–1823 (2013). doi: 10.1093/cercor/bhs159; pmid: 22710612
39. Y. Zhao *et al*., Distinct molecular pathways for development of telencephalic interneuron subtypes revealed through analysis of Lhx6 mutants. *J. Comp. Neurol.* **510**, 79–99 (2008). doi: 10.1002/cne.21772; pmid: 18613121
40. E. Azim, D. Jabaudon, R. M. Fame, J. D. Macklis, SOX6 controls dorsal progenitor identity and interneuron diversity during neocortical development. *Nat. Neurosci.* **12**, 1238–1247 (2009). doi: 10.1038/nn.2387; pmid: 19657336
41. R. Batista-Brito *et al*., The cell-intrinsic requirement of *Sox6* for cortical interneuron development. *Neuron* **63**, 466–481 (2009). doi: 10.1016/j.neuron.2009.08.005; pmid: 19709629
42. N. Flames *et al*., Short- and long-range attraction of cortical GABAergic interneurons by neuregulin-1. *Neuron* **44**, 251–261 (2004). doi: 10.1016/j.neuron.2004.09.028; pmid: 15473965
43. M. C. Tiveron *et al*., Molecular interaction between projection neuron precursors and invading interneurons via stromal-derived factor 1 (CXCL12)/CXCR4 signaling in the cortical subventricular zone/intermediate zone. *J. Neurosci.* **26**, 13273–13278 (2006). doi: 10.1523/JNEUROSCI.4162-06.2006; pmid: 17182777
44. G. López-Bendito *et al*., Chemokine signaling controls intracortical migration and final distribution of GABAergic interneurons. *J. Neurosci.* **28**, 1613–1624 (2008). doi: 10.1523/JNEUROSCI.4651-07.2008; pmid: 18272682
45. G. Li *et al*., Regional distribution of cortical interneurons and development of inhibitory tone are regulated by Cxcl12/Cxcr4 signaling. *J. Neurosci.* **28**, 1085–1098 (2008). doi: 10.1523/JNEUROSCI.4602-07.2008; pmid: 18234887
46. J. A. Bagley, D. Reumann, S. Bian, J. Lévi-Strauss, J. A. Knoblich, Fused cerebral organoids model interactions between brain regions. *Nat. Methods* **14**, 743–751 (2017). doi: 10.1038/nmeth.4304; pmid: 28504681
47. L. Lim, D. Mi, A. Llorca, O. Marín, Development and functional diversification of cortical interneurons. *Neuron* **100**, 294–313 (2018). doi: 10.1016/j.neuron.2018.10.009; pmid: 30359598
48. C. Mayer *et al*., Developmental diversification of cortical inhibitory interneurons. *Nature* **555**, 457–462 (2018). doi: 10.1038/nature25999; pmid: 29513653
49. D. Mi *et al*., Early emergence of cortical interneuron diversity in the mouse embryo. *Science* **360**, 81–85 (2018). doi: 10.1126/science.aar6821; pmid: 29472441
50. F. J. Martini *et al*., Biased selection of leading process branches mediates chemotaxis during tangential neuronal migration. *Development* **136**, 41–50 (2009). doi: 10.1242/dev.025502; pmid: 19060332
51. M. Betizeau *et al*., Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. *Neuron* **80**, 442–457 (2013). doi: 10.1016/j.neuron.2013.09.032; pmid: 24139044
52. N. Kalebic, W. B. Huttner, Basal progenitor morphology and neocortex evolution. *Trends Neurosci.* **43**, 843–853 (2020). doi: 10.1016/j.tins.2020.07.009; pmid: 32828546
53. O. Marín, Cellular and molecular mechanisms controlling the migration of neocortical interneurons. *Eur. J. Neurosci.* **38**, 2019–2029 (2013). doi: 10.1111/ejn.12225; pmid: 23651101
54. A. Kepecs, G. Fishell, Interneuron cell types are fit to function. *Nature* **505**, 318–326 (2014). doi: 10.1038/nature12983; pmid: 24429630

55. R. Tremblay, S. Lee, B. Rudy, GABAergic interneurons in the neocortex: From cellular properties to circuits. *Neuron* **91**, 260–292 (2016). doi: 10.1016/j.neuron.2016.06.033; pmid: 27477017
56. B. Rudy, G. Fishell, S. Lee, J. Hjerling-Leffler, Three groups of interneurons account for nearly 100% of neocortical GABAergic neurons. *Dev. Neurobiol.* **71**, 45–61 (2011). doi: 10.1002/dneu.20853; pmid: 21154909
57. C. S. Raju *et al*., Secretagogin is expressed by developing neocortical GABAergic neurons in humans but not mice and increases neurite arbor size and complexity. *Cereb. Cortex* **28**, 1946–1958 (2018). doi: 10.1093/cercor/bhx101; pmid: 28449024
58. A. Alzu'bi, G. J. Clowry, Multiple origins of secretagogin expressing cortical GABAergic neuron precursors in the early human fetal telencephalon. *Front. Neuroanat.* **14**, 61 (2020). doi: 10.3389/fnana.2020.00061; pmid: 32982702
59. A. B. Muñoz-Manchado *et al*., Diversity of interneurons in the dorsal striatum revealed by single-cell RNA sequencing and PatchSeq. *Cell Rep.* **24**, 2179–2190.e7 (2018). doi: 10.1016/j.celrep.2018.07.053; pmid: 30134177
60. F. K. Wong *et al*., Pyramidal cell regulation of interneuron survival sculpts cortical networks. *Nature* **557**, 668–673 (2018). doi: 10.1038/s41586-018-0139-6; pmid: 29849154
61. O. Marín, Interneuron dysfunction in psychiatric disorders. *Nat. Rev. Neurosci.* **13**, 107–120 (2012). doi: 10.1038/nrn3155; pmid: 22251963
62. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst.* **8**, 281–291.e9 (2019). doi: 10.1016/j.cels.2018.11.005; pmid: 30954476
63. L. Haghverdi, A. T. L. Lun, M. D. Morgan, J. C. Marioni, Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.* **36**, 421–427 (2018). doi: 10.1038/nbt.4091; pmid: 29608177
64. T. Stuart *et al*., Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019). doi: 10.1016/j.cell.2019.05.031; pmid: 31178118
65. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018). doi: 10.1038/nbt.4096; pmid: 29608179
66. C. Trapnell *et al*., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014). doi: 10.1038/nbt.2859; pmid: 24658644
67. X. Qiu *et al*., Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017). doi: 10.1038/nmeth.4402; pmid: 28825705
68. J. Cao *et al*., The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019). doi: 10.1038/s41586-019-0969-x; pmid: 30787437

ACKNOWLEDGMENTS

Funding: This work was supported by the National Key Research and Development Program of China (2019YFA0110101 and 2017YFA0103303 to X.W.; 2017YFA0102601 to Q.W.), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA16020601 and XDB32010100 to X.W.), the National Natural Science Foundation of China (NSFC) (81891001 and 31771140 to X.W.), the BUAA-CCMU Big Data and Precision Medicine Advanced Innovation Center Project (BHME-2019001 to X.W.), the Wellcome Trust (215556/Z/19/Z to O.M.), and Collaborative Research Fund of Chinese Institute for Brain Research (2020-NKX-PT-03). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. **Author contributions:** X.W., O.M., and Q.W. conceived of the project and supervised research. T.L., B.W., and L.S. collected the samples. Y.S., Q.M., and Z.L. performed the scRNA-seq experiments. M.W., D.M., S.Z., and H.D. analyzed the scRNA-seq data. Y.S., Y.C., X.Z., and W.W. prepared the samples and carried out the immunofluorescent staining and image data analysis. J.Z. assisted with animal maintenance. Y.S., M.W., D.M., Q.W., O.M., and X.W. wrote the manuscript. All authors edited and proofed the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The scRNA-seq data used in this study have been deposited in the Gene Expression Omnibus (GEO) under the accession number GSE135827. The supplementary materials contain additional data. All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj6641
Materials and Methods
Figs. S1 to S12
Tables S1 to S9
References (69–74)
MDAR Reproducibility Checklist
Movie S1

25 May 2021; accepted 22 October 2021
10.1126/science.abj6641

RESEARCH ARTICLES

CORONAVIRUS

Protection against SARS-CoV-2 Beta variant in mRNA-1273 vaccine–boosted nonhuman primates

Kizzmekia S. Corbett[1]†‡, Matthew Gagne[1]†, Danielle A. Wagner[1], Sarah O' Connell[1], Sandeep R. Narpala[1], Dillon R. Flebbe[1], Shayne F. Andrew[1], Rachel L. Davis[1], Barbara Flynn[1], Timothy S. Johnston[1], Christopher D. Stringham[1]§, Lilin Lai[2], Daniel Valentin[3], Alex Van Ry[3], Zackery Flinchbaugh[3], Anne P. Werner[1], Juan I. Moliva[1], Manjari Sriparna[1]¶, Sijy O'Dell[1], Stephen D. Schmidt[1], Courtney Tucker[1]#, Angela Choi[4], Matthew Koch[4], Kevin W. Bock[5], Mahnaz Minai[5], Bianca M. Nagata[5], Gabriela S. Alvarado[1], Amy R. Henry[1], Farida Laboune[1], Chaim A. Schramm[1], Yi Zhang[1], Eun Sung Yang[1], Lingshu Wang[1], Misook Choe[1], Seyhan Boyoglu-Barnum[1], Shi Wei[1], Evan Lamb[1], Saule T. Nurmukhambetova[1], Samantha J. Provost[1], Mitzi M. Donaldson[1], Josue Marquez[1], John-Paul M. Todd[1], Anthony Cook[3], Alan Dodson[3], Andrew Pekosz[6], Eli Boritz[1], Aurélie Ploquin[1], Nicole Doria-Rose[1], Laurent Pessaint[3], Hanne Andersen[3], Kathryn E. Foulds[1], John Misasi[1], Kai Wu[4], Andrea Carfi[4], Martha C. Nason[7], John Mascola[1], Ian N. Moore[5], Darin K. Edwards[4], Mark G. Lewis[3], Mehul S. Suthar[2], Mario Roederer[1], Adrian McDermott[1], Daniel C. Douek[1], Nancy J. Sullivan[1], Barney S. Graham[1]*, Robert A. Seder[1]*

Neutralizing antibody responses gradually wane against several variants of concern (VOCs) after vaccination with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccine messenger RNA-1273 (mRNA-1273). We evaluated the immune responses in nonhuman primates that received a primary vaccination series of mRNA-1273 and were boosted about 6 months later with either homologous mRNA-1273 or heterologous mRNA-1273.β, which encompasses the spike sequence of the B.1.351 Beta variant. After boost, animals had increased neutralizing antibody responses across all VOCs, which was sustained for at least 8 weeks after boost. Nine weeks after boost, animals were challenged with the SARS-CoV-2 Beta variant. Viral replication was low to undetectable in bronchoalveolar lavage and significantly reduced in nasal swabs in all boosted animals, suggesting that booster vaccinations may be required to sustain immunity and protection.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) vaccines, such as mRNA-1273, BNT162b2, AD26.COV2.S, AZD1222, and NVX-CoV2373, which deliver the Wuhan-1 Spike (S) protein, are between 70 and 94% effective against symptomatic D614G or B.1.1.7 Alpha (α) SARS-CoV-2 variant infection (*1*–*5*). Several SARS-CoV-2 variants of concern (VOCs) show reduced sensitivity to neutralizing antibodies induced by vaccination or prior infection. The B.1.617.2 Delta (δ) variant is notable for its high transmissibility and an intermediate level of reduced neutralization by WA-1 immune sera (*6*, *7*). The B.1.351 Beta (β) variant and the P.1 Gamma (γ) variant show the greatest reduction in neutralizability by the vaccine recipient or convalescent patient sera (*8*–*13*).

Since mRNA-1273 was granted Emergency Use Authorization (EUA) in December 2020 and deployed globally, many months have elapsed since the first individuals were vaccinated, and the durability of vaccine protection remains a concern. Additionally, the question of whether an additional booster vaccination is necessary, particularly as new VOCs become more prevalent, is clinically relevant. mRNA-1273 vaccination in humans reportedly induces antibody responses that persist beyond 6 months (*14*), but serum-neutralizing activity was significantly reduced against β and δ variants (*7*, *15*, *16*). Recent data show that boosting fully vaccinated individuals with mRNA-1273 several months after initial vaccine priming with mRNA-1273, mRNA-1273.β (S sequence based on the β variant), or a 1:1 combination of both vaccines (mRNA-1273.211) showed significantly increased neutralizing titers against the prototypic D614G virus and all variants, including β (*17*). Although these in vitro functional data are promising, there are no available efficacy data to define the level of antibodies that is sufficient to mediate protection against infection because the correlates of protection against VOCs are undetermined. Moreover, in nonhuman primates (NHPs), a higher threshold of neutralizing antibodies was reportedly needed for upper compared with lower airway protection (*18*), suggesting that a boost could be important to reach this threshold and contribute to reducing virus transmission.

NHPs have proven useful for vaccine development by facilitating the analysis of immunogenicity and protection against SARS-CoV-2 (*19*–*24*). Here, NHPs immunized with the clinically relevant mRNA-1273 vaccine regimen (100 μg × 2) were boosted ~6 months later with either mRNA-1273 (homologous) or mRNA-1273.β (heterologous) vaccines. Antibody, B cell, and T cell responses were assessed temporally after vaccination and infection to determine how boosting with mRNA-1273 or mRNA-1273.β influences the magnitude and quality of antibody responses against WA-1-, β-, and δ-specific responses and protection against SARS-CoV-2 β challenge. The data show that an additional, delayed boost with mRNA vaccines matched to either the original vaccine strain or to the heterologous challenge strain can have a marked and sustained effect on increasing the breadth of neutralization against all VOCs tested and protection in both the lower and upper airways against the β variant.

Results

Boosting increases mRNA-1273–induced serum antibody responses

The primary focus of this study was to assess the effect of homologous or heterologous boosting with the original WA-1 mRNA-1273 or mRNA-1273.β vaccines, respectively, on immunogenicity and protection against infection with the β variant. NHPs that received the clinically relevant primary regimen of mRNA-1273 (100 μg at weeks 0 and 4) were boosted with either mRNA-1273 or mRNA-1273.β vaccines ~6 months later (fig. S1). WA-1 and β S-specific or receptor-binding domain (RBD)–specific serum antibody responses were assessed at weeks 6, 29, 31, and 37. These time points were selected because they correspond to peak (*19*), memory before boost, 2 weeks after boost, and time of challenge ("prechallenge") responses, respectively. WA-1 and β S-specific immunoglobulin G (IgG) geometric mean titers (GMTs) were 53,000 and 39,000 area under the curve (AUC), respectively, at the peak time point after a primary regimen of mRNA-1273. These responses dropped fourfold and sixfold,

[1]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [2]Center for Childhood Infections and Vaccines of Children's Healthcare of Atlanta, Department of Pediatrics, Department of Microbiology and Immunology, Emory Vaccine Center, Emory University, Atlanta, GA 30322, USA. [3]BIOQUAL Inc., Rockville, MD 20850, USA. [4]Moderna Inc., Cambridge, MA 02139, USA. [5]Infectious Disease Pathogenesis Section, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA. [6]Department of Microbiology and Immunology, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD 21205, USA. [7]Biostatistics Research Branch, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA.
***Corresponding author. Email: rseder@mail.nih.gov (R.A.S.); bgraham@nih.gov (B.S.G.)**
†These authors contributed equally to this work. ‡Present address: Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA. §Present address: Duke University School of Medicine, Medical Scientist Training Program, Durham, NC 27705, USA. ¶Present address: Medical College of Virginia, Richmond, VA 23223, USA. #Present address: Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA.

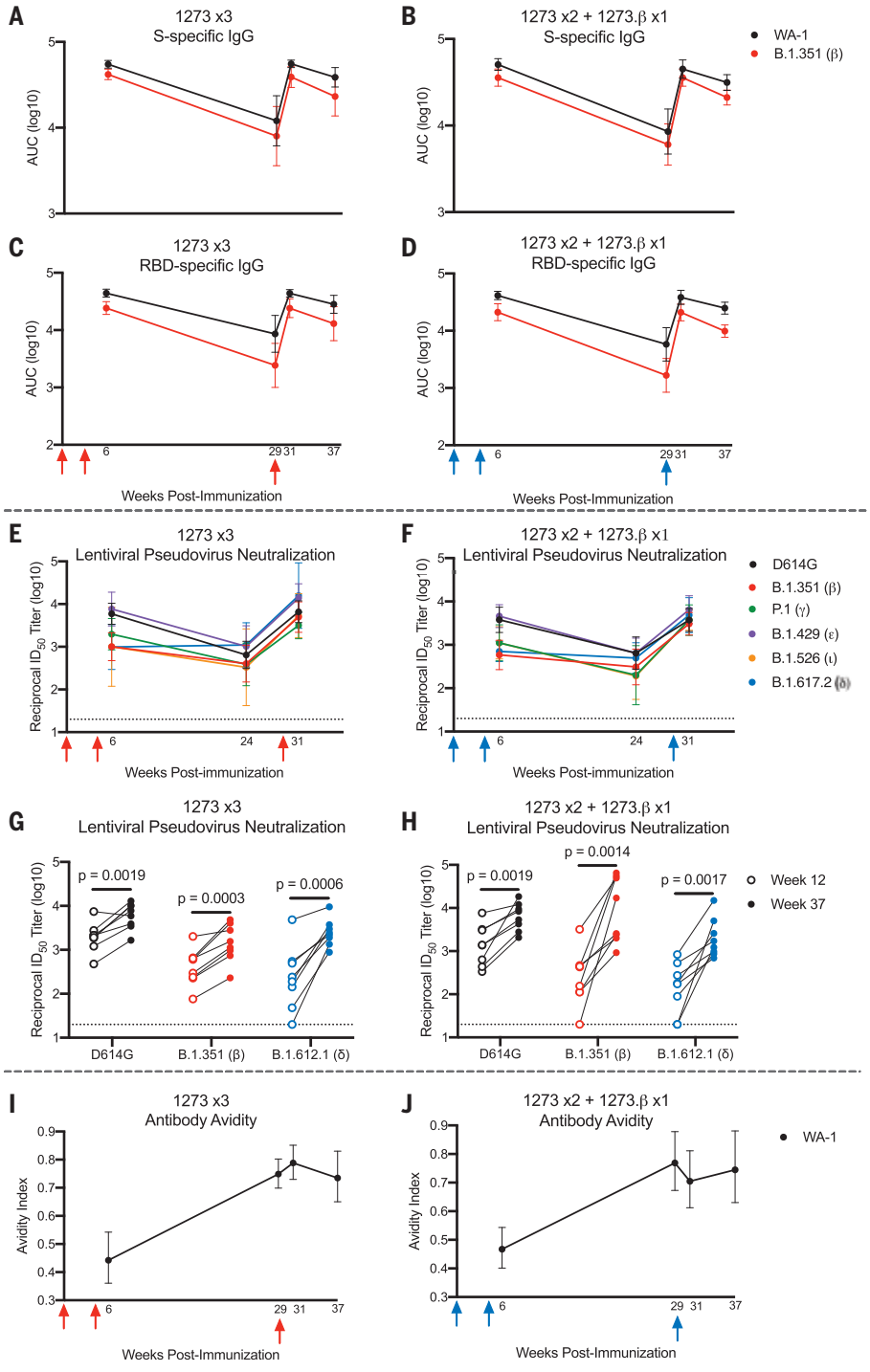


Fig. 1. Temporal serum antibody responses to SARS-CoV-2 variants. Rhesus macaques (n = 8 per group) were immunized as described in fig. S1. (**A** to **D**) Sera collected at weeks 6, 29, 31, and 37 were assessed for SARS-CoV-2 WA-1 (black) and β (red) S-specific [(A) and (B)] and RBD-specific [(C) and (D)] IgG by multiarray ELISA. (**E** and **F**) Sera collected at weeks 6, 24, and 31 were assessed for D614G (black), β, P.1-γ (green), B.1.429-ε (purple), B.1.526-ι (orange), and B.1.617.2-δ (blue) lentiviral-based pseudovirus neutralization. (**G** and **H**) Sera collected at weeks 12 and 37 were assessed for D614G, β, and δ lentiviral-based pseudovirus neutralization. (**I** and **J**) Sera collected at weeks 6, 24, and 31 were assessed for WA-1 S-specific antibody avidity. For (A) to (F) and (I) and (J), lines represent GMTs and geometric error. Arrows point to immunization time points. For (G) and (H), symbols represent individual NHPs. Dotted lines indicate neutralization assay limits of detection. Significance was measured by paired t test.

respectively, after ~6 months (week 29). Boosting with mRNA-1273 or mRNA-1273.β induced fourfold and about sixfold increases in WA-1 and β S-specific IgG titers such that, at 2 weeks after the memory boost, WA-1 and β S-specific IgG GMTs were restored to peak levels (Fig. 1, A and B). By week 37, S-specific IgG GMTs remained >30,000 and >20,000 AUC, respectively, for WA-1 and β (Fig. 1, A and B), which, for WA-1, translated to ~1400 international units/ml (table S1). RBD-specific responses displayed similar kinetic trends (Fig. 1, C and D). Together, these data show that there is no difference in the ability of mRNA-1273 or mRNA-1273.β vaccines to boost primary mRNA-1273–elicited S-binding IgG responses.

Next, we assessed the kinetics of neutralizing antibody responses using a lentiviral-based neutralizing antibody assay for D614G, the benchmark strain, as well as several globally circulating variants [β, δ, P.1-γ, B.1.429-Epsilon (ε), and B.1.526-Iota (ι)]. Consistent with previous data in NHPs and humans (*7*, *15*), mRNA-1273 elicited a variant-dependent hierarchy of neutralizing antibody responses as determined by a lentiviral-based pseudovirus assay. At peak, the median inhibitory dilution (ID_{50}) GMTs for D614G was 4700, followed by 5900, 1500, 1100, 830, and 770 for ε, γ, ι, δ, and β, respectively, representing onefold to sixfold decreases compared with D614G. At the memory time point (week 24), neutralizing antibody titers decreased to 640 ID_{50} GMTs for D614G and 810, 280, 250, 740, and 350 ID_{50} GMTs for ε, γ, ι, δ, and β, respectively. There was a greater fold reduction in neutralization titers from week 6 to week 24 for D614G compared with β and δ ($P < 0.0001$) (Fig. 1, E and F). Similar observations were made using D614G- and β VSV–based pseudovirus (fig. S2, A and B) or live virus neutralization assays (fig. S2, C and D). Last, WA-1 S-specific antibody avidity, one measure of affinity maturation, was significantly increased over the 6 months after the primary mRNA-1273 vaccination series in these NHPs ($P < 0.0001$) (Fig. 1, I and J). This indicates that continued affinity maturation occurs after primary vaccination, leading to an increase in antibody quality, and suggests that, for some variants, affinity maturation of mRNA-1273–elicited neutralizing antibody responses may occur over time, even in the absence of continued antigen exposure.

Six months after the primary vaccination series, either a homologous (mRNA-1273) or a heterologous (mRNA-1273.β) boost induced, on average, a 12-fold increase in pseudovirus-neutralizing titers for all the variants 2 weeks later, resulting in, e.g., 5000 and 3000 β-specific ID_{50} GMTs after mRNA-1273 or mRNA-1273.β boosts, respectively (Fig. 1, E and F). These responses to variants were significantly higher

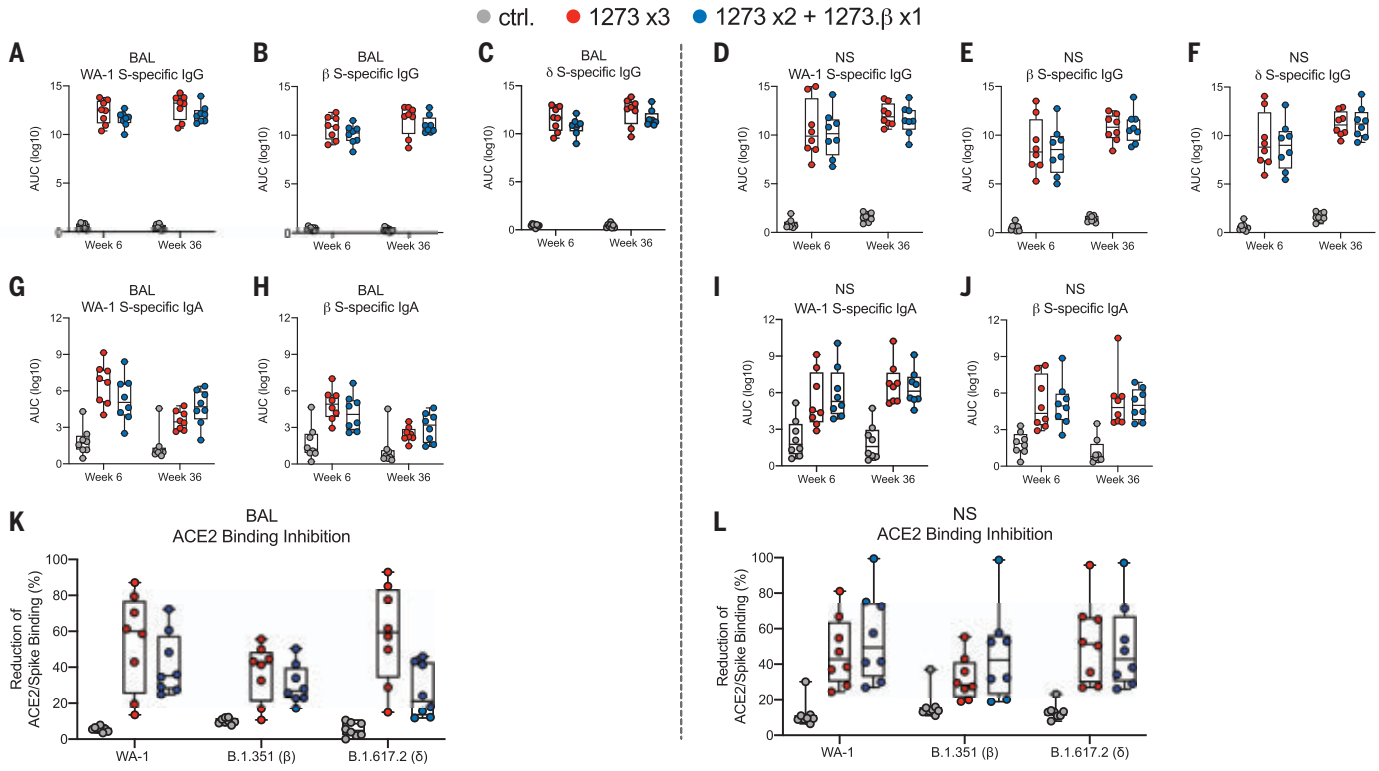


Fig. 2. Mucosal antibody responses to SARS-CoV-2 variants. Rhesus macaques (n = 8 per group) were immunized as described in fig. S1. BAL (left panels) and NS (right panels) collected at weeks 6 and 36 were assessed for SARS-CoV-2 WA-1 [(**A**), (**D**), (**G**), and (**I**)], β [(**B**), (**E**), (**H**), and (**J**)], and δ [(**C**) and (**F**)] S-specific IgG [(A) to (F)] and IgA [(G) to (J)] by multiarray ELISA, where applicable. (**K** and **L**) Week 36 BAL and NS were assessed for inhibition of ACE2 binding to WA-1, β S, and δ [(K) and (L)]. Boxes and horizontal bars denote interquartile ranges (IQRs) and medians, respectively; whisker end points are equal to the maximum and minimum values. Circles represent individual NHPs.

than at week 6 for both boost groups; e.g., week 31 δ-specific neutralizing antibody responses were fivefold greater than week 6 responses (P = 0.002 and 0.003 for mRNA-1273 and mRNA-1273.β boosts, respectively). This postboost increase in lentiviral pseudovirus-neutralizing titers was confirmed using D614G- and β VSV–based pseudovirus neutralization (fig. S2, A and B) and live virus neutralization assays (fig. S2, C and D). To assess the durability of neutralizing antibody responses after boost, we compared week 12, 8 weeks after the primary vaccination series, with week 37, 8 weeks after boost. D614G, β, and δ pseudovirus-neutralizing antibody responses were significantly higher at week 37 than at week 12 (Fig. 1, G and H). These data show that, despite significant waning of mRNA-1273–elicited neutralizing antibody responses against both β and δ over the 6 months after the primary vaccine regimen, boosting with either homologous or heterologous mRNA-1273 restored and increased the potency and breadth of neutralizing antibody responses.

Vaccine boosting increases antibody responses in the airway

Vaccination-induced antibodies localized in the upper and lower airways may play a role in the initial control of SARS-CoV-2 replication (*18*, *25*, *26*). Therefore, we extended analysis beyond circulating antibody responses to assess WA-1, β, and δ S-specific IgG responses in bronchoalveolar lavage (BAL) and nasal swabs (NS) at weeks 6 and 36. At week 6, two doses of mRNA-1273 elicited WA-1, β, and δ S-specific GMTs of 13, 11, and 11 AUC, respectively, in the BAL (Fig. 2, A to C). Similar GMTs (11, 10, and 9 AUC, respectively) of S-specific IgGs were detected in NS samples (Fig. 2, D to F). After a boost with either mRNA-1273 or mRNA-1273.β, the levels of BAL and NS S-specific IgG were also 10 to 12 AUC GMTs (Fig. 2, A to F). WA-1 and β S-specific IgA was also detected in BAL and NS (Fig. 2, G to J). To assess functional antibodies in these mucosal samples, we used an ACE2-binding inhibition assay. In the BAL of NHPs boosted with mRNA-1273, ACE2 binding to WA-1, β, and δ S proteins was reduced by a mean of 54, 37, and 58%, respectively (Fig. 2K) and by 47, 32, and 53%, respectively, in NS (Fig. 2L). Similarly, in BAL and NS of NHPs that were boosted with mRNA-1273.β, ACE2 binding to WA-1 S was reduced by 42 and 55% and for β S proteins by 30 and 46%, respectively (Fig. 2, K and L). These data confirm that either homologous or heterologous memory vaccination can induce upper and lower airway functional antibodies that are likely necessary to mitigate lower airway illness and transmission.

Primary mRNA-1273.β vaccination induces serum and mucosal antibody responses

A secondary focus of this study was to evaluate the immunogenicity of mRNA-1273.β as a primary vaccine regimen because this may have relevance for the design of future vaccines in naïve individuals; this group additionally served as a homologous vaccine control for the β challenge (fig. S1). S-specific IgG GMTs for WA-1 and β were 5100 and 8900, respectively, 2 weeks after the first vaccination of mRNA-1273.β and increased fourfold to fivefold after a second vaccination (fig. S3A). Two doses of mRNA-1273.β resulted in D614G and β live virus ID_{50} GMTs of 198 and 788, respectively (fig. S3B). Potent S-specific antibody responses against WA-1 or β were also detected in BAL (fig. S3C) and NS (fig. S3D). Mucosal antibody responses were further exhibited by ACE2-binding inhibition, where mRNA-1273.β induced BAL antibodies that reduced ACE2 binding to WA-1 and β S by a median of 21 and 48%, respectively (fig. S3E). In NS, the median reduction of ACE2 binding to WA-1 and β S binding was 30 and 56%, respectively

(fig. S3F). Together, these data show mRNA-1273.β given as a primary regimen elicits higher β-specific responses compared with WA-1.

Serum antibody repertoire elicited by homologous or heterologous prime boost

To evaluate the impact of homologous (WA-1) or heterologous (β) boosting antigen on serum antibody epitope specificity, the absolute value [in response units (RUs)] and the relative proportion (percent competition) of serum antibodies against 16 distinct antigenic sites (table S2) on WA-1 SARS-CoV-2 S was measured using surface plasmon resonance (SPR). We evaluated serum antibody specificity at week 35 (6 weeks after boost), to allow sufficient time for B cell expansion after exposure to naïve β-S antigen in heterologous-boosted animals. The epitope specificity spanned the S2, N-terminal domain (NTD), S1, and RBD subdomains of the S protein, and the breadth was similar whether the boost was homologous or heterologous (Fig. 3A). The specificity of serum antibodies was qualitatively similar across all animals within each group for the NTD subdomain (average SD 32.1 to 36.9) but not for RBD subdomain (average SD 50.2 to 56.1) (Fig. 3A). Between vaccine groups, there was more NTD site A specificity in homologous- than in heterologous-boosted animals (Fig. 3A). Both absolute and relative serum reactivity to NTD site A [monoclonal antibody (mAb) 4-8; table S2] were higher in homologous- than heterologous-boosted animals at all time points after boost (Fig. 3B), suggesting that mutations found in the β S protein (Δ242-244, R246I) may render the heterologous boost unable to recall a primary mRNA-1273 vaccination memory B cell response to this antigenic site. By contrast, the RBD site H (mAb A23-97.1; table S2) showed differences in relative but not absolute serum reactivity, with heterologous boost inducing higher reactivity than homologous boosting at all postmemory boost time points (Fig. 3C). Because A23-97.1 binds equivalently to both WA-1 and β S (fig. S4), the higher relative (but not absolute) reactivity against this site for β is likely caused by the decreased contribution of NTD-A reactivity, as shown by the ratio of mRNA-1273.β to mRNA-1273 relative reactivity (fig. S5). Overall, mRNA-1273 to β ratios decreased over time, indicating a contraction of the naïve β-directed response and a return to a memory, long-lived epitope profile (fig. S5) and suggesting that heterologous boost does not alter absolute epitope reactivity over time. By contrast, after primary vaccination with mRNA-1273.β, we observed significantly increased absolute serum reactivity to RBD sites B, C, D, and F (fig. S6) compared with homologous boost, with sites C and D (A19-46.1 and A19-61.1, respectively; table S2) being associated with strong neutralization potency against β (*27*). These data highlight the differences between primary vaccination and heterologous boost with mRNA-1273.β and indicate that primary vaccination plays a key role in shaping the overall serum antibody epitope repertoire.

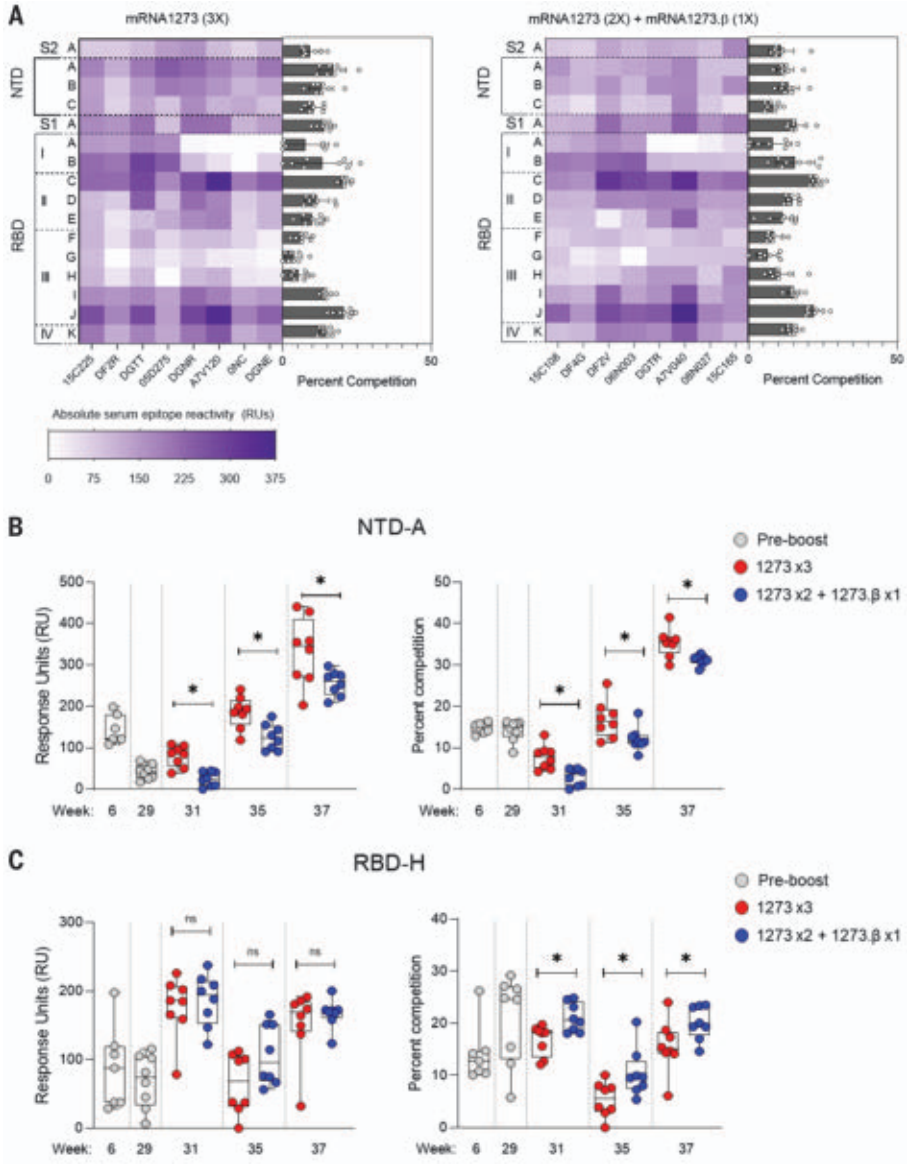


Fig. 3. Serum epitope mapping analysis. Rhesus macaques (n = 8 per group) were immunized according to fig. S1. (**A**) Serum epitope fingerprints at week 35 (6 weeks after boost). Absolute serum epitope reactivity (RUs) is indicated from light to dark purple. Columns represent individual NHPs; rows represent antigenic site. Bar graphs on the right depict relative serum reactivity (percent competition) for each antigenic site with each animal represented by an open circle (n = 8 per group). (**B** and **C**) Absolute (RUs) and relative (percent competition) serum reactivity for epitopes which show significant ($P \leq 0.05$) differences between vaccination groups at all time points. NTD-A site is represented by mAb 4-8, RBD site H is represented by mAb A23-97.1 (see table S2). Significant ($P \leq 0.05$) differences were determined by unpaired t test between vaccination groups at a single time point.

Vaccine boosting expands mRNA-1273–induced memory B cell responses

The durability of vaccine-induced antibody responses is an integral component of the pandemic response as efforts continue to mitigate the ongoing spread of SARS-CoV-2. Durable humoral immunity is driven by the ability to generate and sustain memory B cells and long-lived plasmablasts (*28*), and recent data in humans show the potential for such responses after vaccination with mRNA or primary infection (*29*, *30*). To define how B cell specificity changes over time in mRNA-1273–immunized NHPs and to assess how mRNA-1273 priming

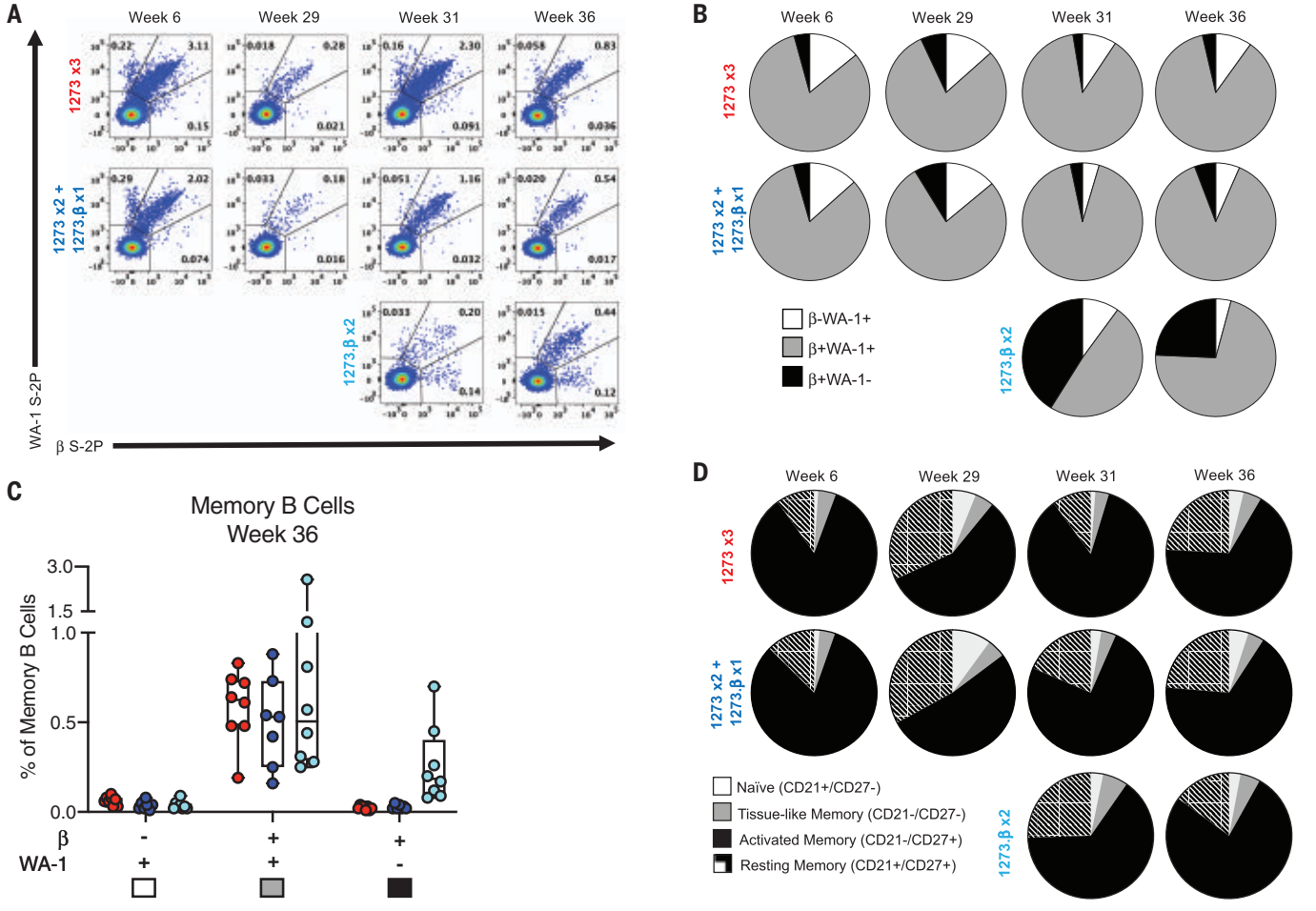


Fig. 4. Frequency of SARS-CoV-2 S-specific B cells in PBMCs. Rhesus macaques (n = 8 per group) were immunized as described in fig. S1. PBMCs were collected at weeks 6, 29, 31, and 36 and analyzed by flow cytometry. (**A**) Representative flow cytometry plots from one NHP per vaccine group depicting the frequency of memory B cells that bind the WA-1 S and/or β S, at various time points after vaccination. (**B**) Pie charts representing the fraction of S-specific B cells that are single positive for the WA-1 (white, β-WA-1$^+$), double positive for WA-1 and β (gray, β$^+$WA-1$^+$), and single positive for β (black, b$^+$WA-1$^-$) for all NHPs in each group. (**C**) Frequency of β-WA-1$^+$, β$^+$WA-1$^+$, and β$^+$WA-1$^-$ memory B cells at week 36 in each of the vaccine groups. Boxes and horizontal bars denote the IQRs and medians, respectively; whisker end points are equal to the maximum and minimum values. Circles represent individual NHPs. (**D**) Pie charts representing the fraction of S-specific B cells that are naïve (white, CD21$^+$/CD27$^-$), tissue-like memory (gray, CD21$^-$/CD27$^-$), activated memory (black, CD21$^-$/CD27$^+$), and resting memory (black and white, CD21$^+$/CD27$^+$) for all NHPs in each group.

imprints homologous or heterologous boost, we performed temporal analysis of WA-1 and β S-specific memory B cells (fig. S7). After the primary vaccination series with mRNA-1273, at week 6, the frequency of memory B cells expressing antibody receptors dual reactive for both WA-1 and β S was 2 to 3%, with a much lower proportion of single WA-1- or β-specific B cells (Fig. 4, A and B). Six months later, at week 29, there was an ~10-fold reduction in the frequency of double-positive WA-1 and β S-specific memory B cells (Fig. 4, A and B), but these were restored (~10-fold increase) after a boost with mRNA-1273 or mRNA-1273.β (Fig. 4). By contrast, boosting did not cause an increase in the frequency of single-positive WA-1 or β S-specific B cells (Fig. 4). There was also expansion of a resting memory B cell population over the 6 months after the primary mRNA-1273 vaccination series; a similar population of resting memory B cells was present in NHPs 7 weeks after boost (week 36) with homologous or heterologous mRNA-1273 (Fig. 4D). This finding demonstrates a rapid recall response of primary vaccination–induced B cells and coincides with an increase in neutralizing antibody responses (Fig. 1, E to J). The NHPs that underwent primary mRNA-1273.β vaccination had a memory B cell response that also consisted of WA-1 and β S-specific double-positive specificity but a higher proportion of single-positive β S-specific cells (Fig. 4), consistent with β-specific skewing of antibody responses after mRNA-1273.β primary vaccination (fig. S3). These data confirm the observation from serum antibody epitope mapping that both homologous and heterologous boosting can efficiently expand memory B cell responses that are maintained after primary vaccination.

Vaccine boosting restimulates mRNA-1273–induced T_H1 and T_{FH} T cell responses

mRNA-1273 has been shown to elicit CD4 responses composed of T helper 1 (T_H1) and T follicular helper (T_{FH}) cells and a lower frequency of CD8 T cells in both humans and NHPs (*18*, *19*, *31*). However, the definition of longitudinal T cell development and the potential for boosting contracted memory T cells have not yet been reported in mRNA-1273–vaccinated NHPs. Here, consistent with prior studies, S-specific T_H1 and T_{FH} responses, but

not T_H2 responses, were induced at week 6, 2 weeks after mRNA-1273 vaccination (Fig. 5). In these NHPs, T_H1 and T_{FH} responses were contracted by week 29 but were both boosted by either mRNA-1273 or mRNA-1273.β (Fig. 5, A, C, and D). Week 36 responses revealed that mRNA-1273.β vaccination also induced S-specific T_H1 and T_{FH} responses, but not T_H2 responses (Fig. 5). These data suggest that homologous and heterologous mRNA vaccine boosts are equally capable of restimulating T_H1 and T_{FH} cell responses.

NHPs are protected in the upper and lower airway against SARS-CoV-2 β challenge

NHPs were challenged at week 38, which was ~9 weeks after vaccine boost in animals receiving three immunizations, and ~5 weeks after the primary mRNA-1273.β vaccination series (i.e., two vaccinations), with a total dose of 2×10^5 plaque-forming units (PFUs) of SARS-CoV-2 β by intratracheal and intranasal routes (fig. S1 and Fig. 6A). Two days after challenge, control NHPs that received mock mRNA had a median of ~6 $\log_{10}$ SARS-CoV-2 envelope (E) subgenomic RNA (sgRNA _E) copies/ml in BAL. By contrast, all eight NHPs that received primary mRNA-1273.β vaccination had undetectable BAL sgRNA_E. Day 2 after challenge, four of eight NHPs that received primary mRNA-1273 vaccination with homologous mRNA-1273 boost had undetectable BAL sgRNA _E, and seven of eight NHPs boosted with heterologous mRNA-1273.β had undetectable BAL sgRNA_E. By day 4, all boosted NHPs, with the exception of one mRNA-1273–boosted animal, had undetectable BAL sgRNA_E (Fig. 6B). In NS, at days 2 and 4 after challenge, both groups of boosted NHPs showed significantly less (~3 to 4 $\log_{10}$ reduction) sgRNA_E compared with controls ($P = 0.0014$ and 0.0004 for mRNA-1273 and mRNA-1273.β, respectively). By day 7, the majority (13 of 16) of boosted NHPs had undetectable sgRNA_E compared with control NHPs, for which NS sgRNA_E persisted at a median of 4 $\log_{10}$ NS sgRNA_E (Fig. 6C).

Although sgRNA is a sensitive measurement of viral replication that can be used diagnostically, we also determined the ability of live virus to be propagated from postchallenge BAL and NS samples, an alternative measure of viral load that is relevant for indicating lung disease or transmission from the upper airway. On day 2 after challenge, all boosted animals had low to undetectable [i.e., <4 $\log_{10}$ median tissue culture infectious dose ($TCID_{50}$)/ ml] of culturable SARS-CoV-2 in BAL (Fig. 6D). In NS, SARS-CoV-2 was unculturable for three of eight or five of eight NHPs that were boosted with mRNA-1273 or mRNA-1273.β, respectively, which was similar to the mRNA-1273.β primary vaccination group (Fig. 6E). sgRNA and viral titers were highly correlated (Fig. 6, F and G); in fact, there was no culturable SARS-CoV-2 from BAL or NS that had sgRNA_E levels <1.0×10^4 RNA copies/ml (Fig. 6F) or <4.7×10^3 RNA copies/swab (Fig. 6G), respectively. Nucleocapsid (N)-specific sgRNA measurements followed the same trend, albeit with higher detection sensitivity (fig. S9).

We also evaluated lung samples for pathology and detection of viral antigen 7 to 9 days after SARS-CoV-2 β challenge. Inflammation was minimal to mild and was similar across lung samples from vaccinated NHPs, with rare cases showing a moderate to severe response (fig. S10 and table S3). The inflammatory changes in the lung were characterized by a mixture of macrophages and polymorphonuclear cells present within some alveolar spaces and mild to moderate expansion of alveolar capillaries with mild type II pneumocyte hyperplasia at day 7 after challenge to changes more consistent with lymphocytes, histiocytes, and fewer polymorphonuclear cells associated with more prominent and expanded alveolar capillaries, occasional areas of perivascular and peribronchiolar inflammation, and type II pneumocyte hyperplasia at days 8 to 9 after challenge. SARS-CoV-2 antigen was detected in six of six control NHPs evaluated at days 7, 8, and 9 after infectious challenge, with three of six showing antigen present in multiple lobes. In vaccinated NHPs, antigen was detected in only one of 24 animals and in a single lobe (mRNA-1273x3; table S3). These results, together with sgRNA data, confirm that boosting mRNA-1273–vaccinated NHPs limits β viral replication in the lower and upper airway.

Concluding remarks

The SARS-CoV-2 mRNA-1273 vaccine shows between 90 and 100% protection against WA-1 (*2*), α, or β variants (*32*) when administered as two doses 4 weeks apart and assessed within a 2-month window. Kinetic analyses of antibody responses after vaccination with mRNA-1273 (*7*) or BNT162b2 (*30*) show reduction of neutralizing activity from the peak humoral response after the second immunization through day 209 (*7*, *17*). Consistently, the β variant is the most neutralization resistant of all VOCs to date. However, when we initiated these studies, it was not yet established how this resistance would influence the durability of protective efficacy in humans against β. Of additional concern are recent clinical reports from Israel showing that cohorts immunized with the BioNTech/Pfizer COVID-19 vaccine, BNT162b2, >6 months previously may present

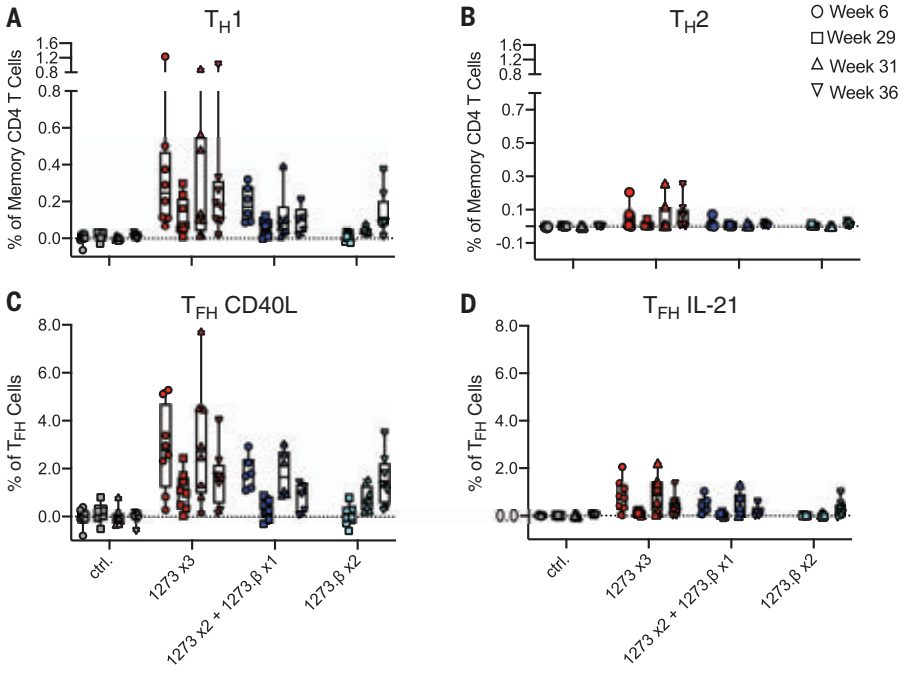


Fig. 5. T cell responses after mRNA-1273 vaccination. Rhesus macaques were immunized as described in fig. S1. Intracellular staining (ICS) was performed on PBMCs at weeks 6, 29, 31, and 36 to assess T cell responses to SARS-CoV-2 S protein peptide pools S1 and S2. Responses to S1 and S2 individual peptide pools were analyzed. (**A**) T_H1 responses (IFN-γ, IL-2, or TNF), (**B**) T_H2 responses (IL-4 or IL-13), and (**C**) T_{FH} CD40L up-regulation (T_{FH} cells) were gated on central memory $CXCR5^+PD\text{-}1^+ICOS^+$ CD4 T cells. (**D**) T_{FH} responses (IL-21). Boxes and horizontal bars denote IQRs and medians, respectively; whisker end points are equal to the maximum and minimum values. Circles represent individual NHPs. Dotted lines are set to 0%.

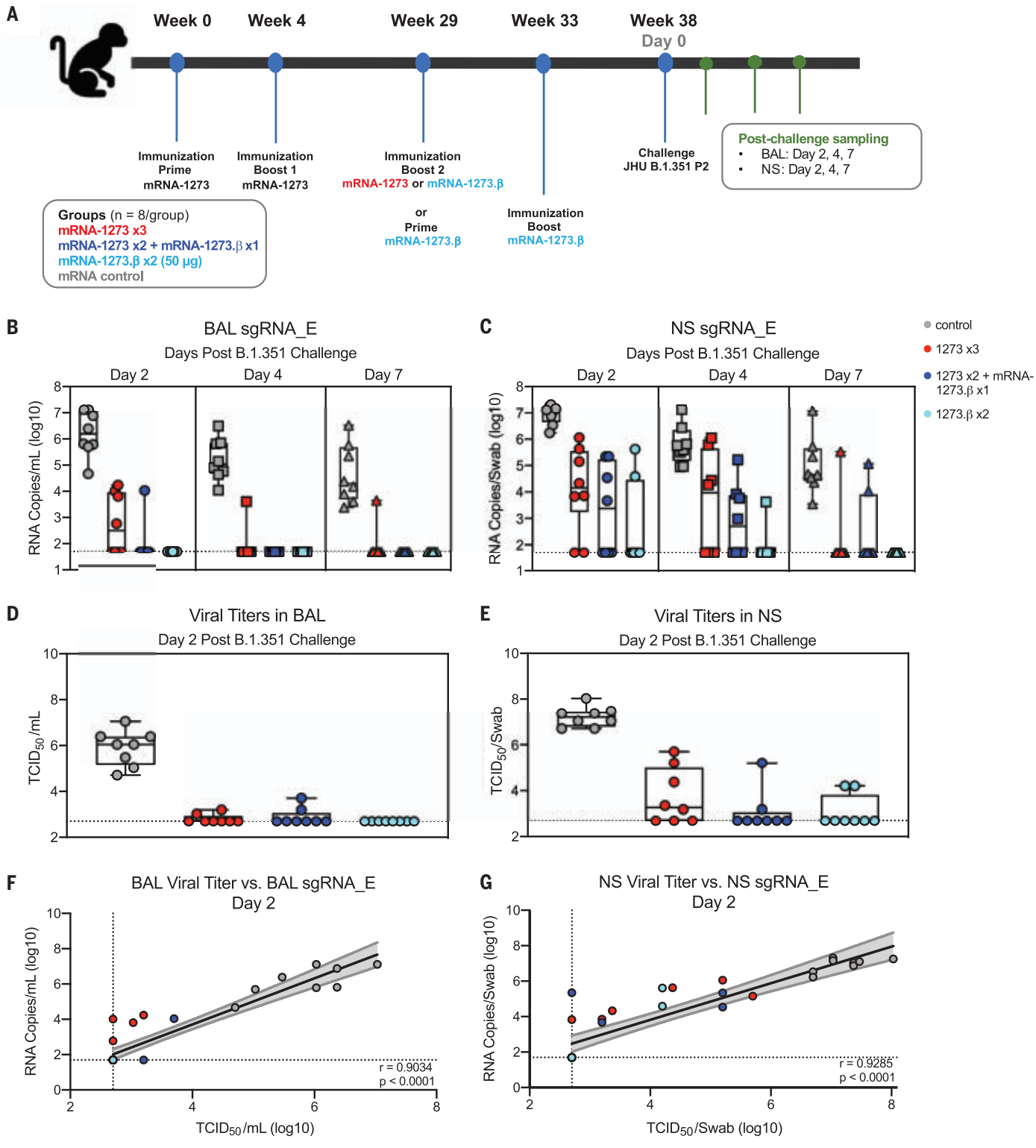


Fig. 6. Efficacy of mRNA-1273 against upper and lower respiratory B.1.351 viral replication. (**A**) Abbreviated schema: Rhesus macaques were immunized and challenged. (**B** to **E**) BAL [(B) and (D)] and NS [(C) and (E)] samples were collected on day 2 (circles), day 4 (squares), and day 7 (triangles) after challenge, where applicable. Viral replication was assessed by detection of SARS-CoV-2_E sgRNA [(B) and (C)], and viral titers were assessed by $TCID_{50}$ assay [(D) and (E)]. For (B) to (E), boxes and horizontal bars denote the IQRs and medians, respectively; whisker end points are equal to the maximum and minimum values. For (**F**) and (**G**), plots show correlations between viral titers and sgRNA_E in BAL (F) and NS (G) 2 days after challenge. Black and gray lines indicate linear regression and 95% confidence interval, respectively. *r* and *P* represent Spearman's correlation coefficients and corresponding *P* values, respectively. Symbols represent individual NHPs and may overlap for equal values. Dotted lines indicate assay limits of detection.

with symptomatic disease or increased frequency of severe disease when infected with the δ variant (*33*). Like β, δ also includes mutations in S that reduce neutralization sensitivity by vaccinee sera (*34*). Thus, the combination of lack of vaccination, waning immunity, reduced neutralization capacity against VOCs, and increased transmissibility of δ are key factors responsible for the current status of the ongoing COVID-19 pandemic. It may be important to boost antibody responses, especially against VOCs, to sustain and increase protection against severe disease, particularly in at-risk cohorts, and to reduce the potential for mild infection and transmission.

Current assessments of COVID-19 vaccine efficacy in both preclinical and clinical studies are concerned with whether an additional boost after the primary vaccine regimen should be matched to the most epidemiologically relevant VOCs or if homologous boosting with the original vaccine will be sufficient to generate broadly neutralizing antibody responses against VOCs. Here, we show that boosting NHPs ~6 months after their primary vaccination series with either mRNA-1273 or mRNA-1273.β significantly increased serum-neutralizing activity against all VOCs before the boost and mediated high-level protection in the upper and lower airways against β challenge. Neutralizing titers against all VOCs tested were significantly higher at both 2 and 8 weeks after the boost compared with the same time points after the primary vaccination series (*17*). These data suggest that boosting could lead to a higher antibody setpoint, potentially altering the need and timing for additional boosting. Last, we could not discern a significant difference in neutralizing antibody titers or reduction in viral replication between the homologous and variant mRNA boost. These data suggest that for the currently circulating strains, the ancestral WA-1 S encoded by mRNA vaccines will significantly enhance immunity and confer high-level protection. It remains possible that in the future, viral evolution of SARS-CoV2 may require a different boost regimen, and this will be an area of continued investigation.

Longitudinal assessment of neutralizing antibody responses showed a relatively slower decay in serum-neutralizing activity against VOCs compared with the benchmark D614G strain. These data suggest that although humoral antibody responses are an established correlate for short-term protection in NHPs and humans (*18*, *35*), deeper analysis may be needed to define correlates of protection with contracted memory responses. These data also propose the possibility that despite a quantitative decrease in the magnitude of antibody responses over time after vaccination, the loss of functional activity may be partially mitigated by affinity maturation and improved quality of the response. It was recently reported that there may be discordance in antibody evolution between convalescent individuals and those who receive an mRNA vaccine. Specifically, the report indicated less affinity maturation and a limited increase in breadth to the RBD subdomain in isolated mAbs between 2 and 5 months after mRNA vaccination (*30*). Our assessment of polyclonal serum antibody affinities (avidity) using the intact WA-1 S protein showed that vaccination with mRNA-1273 led to an increase in avidity over time, confirming a qualitative improvement in the antibody response in NHPs (*29*, *36*, *37*).

The increase in antibody responses after boost suggests that there is significant B cell memory induced by primary mRNA vaccination that can be rapidly recalled after the boost (*29*). Here, after primary vaccination with mRNA-1273, the peak response yielded a total frequency of WA-1 and β-specific memory B cells of ~3%, and the majority (~80%) of these responses recognized both S proteins, with a small proportion specific for WA-1 or β only. After 6 months, the magnitude of the responses contracted but were rapidly restored to a frequency of 3% after boosting. The relative frequency of B cells specific for WA-1, β, or both did not change after boost with either the homologous or heterologous mRNA, suggesting that priming with mRNA-1273 imprinted the B cell repertoire. Additionally, a substantial proportion of the S-specific B cells 7 weeks after boost were resting memory B cells, which suggests that boosting rapidly restores the resting memory B cell population that is seen after primary mRNA-1273 vaccination and is consistent with the sustained antibody responses after boost.

Naïve animals primed with only mRNA-1273.β showed between 25 and 40% of the total B cell response specific for β only. Likewise, serum antibody epitope profiling demonstrated a repertoire that was qualitatively similar after either homologous or heterologous memory boost after mRNA-1273 primary vaccination, whereas a primary vaccination series with mRNA-1273.β yielded a unique repertoire (both qualitative and quantitative) with marked increases in reactivity to epitopes associated with broad and potent neutralization of VOCs. These data have implications for how future mRNA vaccines can be designed to imprint B cell repertoires in naïve individuals for increased potency and breadth of neutralizing activity (*38*).

Although the data presented here focus on the role of boosting vaccine responses with mRNA, as a proof of principle, we also show that priming with mRNA-1273.β in naïve NHPs induced potent neutralizing antibody responses against β and provided high-level protection in the upper and lower airways after challenge. Parallel studies in preclinical rodent models have shown that priming with a variant mRNA or a bivalent vaccine containing the WA-1 and the β variant induces higher neutralizing responses to D614G and a number of variants tested relative to either mRNA-1273 or mRNA-1273.β alone (*38*). These data highlight how vaccines in unprimed individuals may be changed depending on the evolution of the virus in the future.

The primary aim of this study was to assess whether homologous or heterologous mRNA-1273 boost in naïve animals that underwent a primary vaccine regimen elicits comparable protective immune responses to inform clinical development, and the experimental design was powered to specifically assess this aim given our limited number of available animals. Recently, broad cross-clade neutralizing antibodies were isolated from SARS-CoV-1 survivors who then received the BNT162b2 vaccine (*39*), so it is notable that the close relatedness of the homologous and heterologous mRNA-1273 boosters used here may limit our ability to fully gauge postboost cross-reactivity. Another limitation of this study is that we could not directly establish whether boosting increased protection compared with vaccinated animals that did not receive a boost. As to whether the boost increased protection, in our recent NHP study (*40*), animals received a primary vaccination series of 100 μg of mRNA-1273 and were challenged at 8 weeks after vaccination with the same β isolate used here (JHU B.1351 P2). There was a 1-$\log_{10}$ reduction in sgRNA in the upper airway, and the β-specific neutralizing antibody ID_{50} GMT was 300 at the time of challenge. Here, we show that 8 weeks after boost, β-specific neutralizing GMTs were ~3000, and there was a 3- to 4-log reduction in sgRNA in the upper airway. These data strongly suggest that the increased neutralizing responses induced by the additional boost leads to better protection in the upper airway.

In conclusion, the data reported here show that a homologous mRNA-1273 boost can significantly improve both the breadth and potency of neutralizing antibody responses and confer protection against upper and lower airway infection to a heterologous challenge virus, which is relatively resistant to in vitro neutralization. The potential clinical utility of a boost would be to sustain high-level protection against severe disease and possibly limit the duration and extent of mild infection in the setting of waning immunity, especially in the elderly and others with preexisting health conditions or poor response to vaccination. We have previously shown that there is a higher threshold for antibody-mediated protection in the upper airway (*18*), which would be more important for limiting mild infection and transmission than for protecting the lower airways to limit severe disease. Thus, boosting will potentially provide more

durable and broader immunity and could reduce morbidity and mortality until greater population immunity is achieved.

Materials and Methods

Preclinical mRNA and lipid nanoparticle production

A sequence-optimized mRNA encoding prefusion-stabilized SARS-CoV-2 S protein containing 2 proline stabilization mutations (S-2P) (*41*, *42*) for WA-1 and β was synthesized in vitro and formulated as previously reported (*17*, *43*). Control mRNA "UNFIX-01 (Untranslated Factor 9)" was synthesized on the basis of the sequence in table S4 and similarly formulated into lipid nanoparticles.

Rhesus macaque model

Animal experiments were performed in compliance with all pertinent US National Institutes of Health regulations and approval from the Animal Care and Use Committees of the Vaccine Research Center and BIOQUAL Inc. (Rockville, MD). Studies were conducted at BIOQUAL Inc. The experimental details of VRC-20-857.3b (fig. S1) are similar to prior studies (*18*, *19*, *26*, *40*). Briefly, 3- to 15-year-old rhesus macaques of Indian origin were stratified into groups based on sex, age, and weight. Animals were immunized with mRNA-1273 at weeks 0 and 4 with a dose of 100 μg intramuscularly in 1 ml of formulated in phosphate-buffered saline (PBS) into the right hindleg. Placebo-control animals were administered control mRNA. At week 29 (~25 weeks after the second immunization), a group of animals was boosted with 50 μg of mRNA-1273 or 50 μg of mRNA-1273.β. An additional group of animals was immunized at weeks 29 and 33 with 50 μg of mRNA-1273.β. At week 38 (9 weeks after the homologous or heterologous mRNA boost or 5 weeks after the mRNA-1273.β prime and boost) all animals were challenged with a total dose of 2×10^5 PFUs of SARS-CoV-2 β (JHU P2) as previously described (*40*). The viral inoculum was administered as 1.5×10^5 PFUs in 3 ml intratracheally and 0.5×10^5 PFU in 1 ml intranasally in a volume of 0.5 ml into each nostril. Pre- and postchallenge sample collection is detailed in fig. S1.

Quantification of SARS-CoV-2 sgRNA

BAL and NS subgenomic SARS-CoV-2 E mRNA was quantified by reverse transcription polymerase chain reaction (RT-PCR) as previously described (*26*). Subgenomic SARS-CoV-2 N mRNA was quantified similarly, as described in (*40*). The lower limit of quantification was 50 copies.

$TCID_{50}$ quantification of SARS-CoV-2 from BAL

Viral load ($TCID_{50}$) from BAL samples was calculated using previously described methods (*40*).

Histopathology and immunohistochemistry

After challenge, on days 7 to 9, animals were euthanized, and lung tissue was processed and stained with hematoxylin and eosin for routine histopathology and analyzed for detection of SARS-CoV-2 virus antigen as previously described (*40*). Inflammation was often characterized by a mixture of inflammatory cells including polymorphonuclear cells, macrophages, and lymphocytes; cellular infiltrates were often observed in association with perivascular areas, alveolar interstitium, and surrounding small- and medium-sized airways. Antigen-positive foci were often associated with alveolar-lining cells (pneumocytes), immune cells, and infrequently with the airway epithelium of small- and medium-sized airways. All samples were blinded and evaluated by a board-certified veterinary pathologist.

Multiplex MSD ELISA for serum antibody responses

For 10-plex meso scale discovery (MSD) enzyme-linked immunosorbent assay (ELISA), 96-well plates were precoated with SARS-CoV-2 S-2P (*41*) and RBD proteins from multiple variants, SARS-CoV-2 N protein, and bovine serum albumin (BSA) and supplied by the manufacturer (Meso Scale Diagnostics). Determination of serum antibody binding was performed as previously described (*7*), and reagent details are provided in table S5. All calculations were performed within Excel and GraphPad Prism software version 7.0. Readouts are provided as AUC. For 4-plex ELISA, 96-well plates were precoated with SARS-CoV-2 S-2P, RBB, N, and a BSA control in each well. Determination of serum antibody binding was performed as previously described (*7*, *18*, *26*). Calculated electrochemiluminescence immunoassay (ECLIA) parameters to measure binding antibody activities included interpolated concentrations or assigned arbitrary units (per milliliter) read from the standard curve. International units (IU) were established for each antigen on the basis of parallelism between the MSD reference standard and the World Health Organization international standard. The S-specific IgG lower limit of detection was 0.3076 IU/ml, and the RBD-specific IgG lower limit of detection was 1.5936 IU/ml.

MSD ELISA for mucosal antibody responses

Total S-specific IgG in BAL and NS was determined by MSD ELISA as previously described (*26*), with the following minor change: MSD panel 13 was used.

Serum antibody avidity assay

Avidity was assessed using a sodium thiocyanate (NaSCN)–based avidity ELISA against SARS-CoV-2 S-2P as previously described (*18*, *26*). The avidity index was calculated using the ratio of IgG binding to S-2P in the absence or presence of NaSCN and is reported as the average of two independent experiments, each containing duplicate samples.

Serum antibody epitope definition

Serum epitope mapping competition assays were performed using a Biacore 8K+ (Cytiva) SPR spectrometer. Following the manufacturer's protocol, anti-histidine IgG1 antibody was immobilized on Series S Sensor Chip CM5 (Cytiva) through primary amine coupling using a His capture kit (Cytiva). His-tagged SARS-CoV-2 WA.1 S-2P was captured on active sensor surface.

Competitor mAbs or negative control antibodies at set concentrations were injected over both active and reference surfaces to saturation. The competitor human IgG mAbs used included S2-specific mAb S652-112; NTD-specific mAbs 4-8, S652-118, and N3C; S1-specific mAb A20-36.1; and RBD-specific mAbs B1-182, CB6, A20-29.1, A19-46.1, LY-COV555, A19-61.1, S309, A23-97.1, A19-30.1, A23-80.1, and CR3022. NHP sera were flowed over both active and reference sensor surfaces for 40 min. Then, 1× PBS-P+ (Cytiva) was used as the running buffer and diluent for all samples. Active and reference sensor surfaces were regenerated between each analysis cycle using 10 mM glycine, pH 1.5 (Cytiva).

Before analysis, sensorgrams were aligned to Y (RUs = 0), beginning at the serum association phase using Biacore 8K Insights Evaluation Software (Cytiva). Reference-subtracted relative "analyte binding late" report points (RUs) were collected and used to calculate absolute and relative competition. Relative percent competition (% C) was calculated using the following formula: % C = {1 – (100 * [(RU in presence of competitor mAb)/(RU in presence of negative control mAb)]}. Absolute competition (ΔRUs) was calculated with the following formula: ΔRUs = [(RUs in presence of negative control mAb) – (RUs in presence of competitor mAb)]. Results are reported as absolute serum epitope reactivity and percent competition, and statistical analysis was performed using unpaired, two-tailed *t* test (GraphPad Prism software version 8). All assays were performed in duplicate, with average data points for each animal represented on the corresponding graphs.

SPR-binding assay

Series S CM5 Sensor Chips (Cytiva) were activated by immobilizing anti-histidine IgG1 antibody on surface using a His capture kit (Cytiva) according to the manufacturer's protocol. His-tagged SARS-CoV-2 WA.1 or β S protein containing S-2P mutations was captured on an active sensor surface at a set concentration for 10 min. mAbs were injected over both active and reference surfaces and allowed to

bind to saturation. Graphs represent reference-subtracted relative "analyte binding late" report points (RUs), determined by aligning sensorgrams to Y (RUs) = 0, beginning at the mAb association stage using Biacore 8K Insights Evaluation Software (Cytiva). All assays were performed using a Biacore 8K+ (Cytiva) SPR spectrometer.

ACE2-binding inhibition assay

ACE2-binding inhibition was completed, as previously described (*7*), using 1:5 diluted BAL and NS samples.

Lentiviral pseudovirus neutralization assay

Pseudotyped lentiviral reporter viruses were produced by the cotransfection of plasmids encoding SARS-CoV-2 proteins from multiple variants, a luciferase reporter, lentivirus backbone, and human transmembrane protease serine 2 (TMPRSS2) genes as previously described (*17*, *44*). Reagent details are shown in table S5. Sera were tested, in duplicate, for neutralizing activity against the pseudoviruses, and percent neutralization was calculated in GraphPad Prism software version 9.0.2. The lower limit of quantification was 1:40 ID_{50}.

VSV pseudovirus neutralization assay

SARS-CoV-2 pseudotyped recombinant VSV-ΔG-firefly luciferase viruses were made by co-transfection of plasmid expressing full-length S, and subsequent infection with VSVΔG-firefly-luciferase and neutralization assays were completed on sera samples as previously described (*15*, *45*); reagent details are described in table S5. The lower limit of quantification was 1:40 ID_{50}.

Focus reduction neutralization test

Focus reduction neutralization tests were performed on sera samples, in duplicate, as previously described (*40*). Reagent details are provided in table S5. For samples that did not neutralize 50% of virus at the limit of detection, "5" was plotted and used for geometric mean calculations.

B cell probe-binding assay

Cryopreserved peripheral blood mononuclear cells (PBMCs) were thawed and washed in wash buffer (4% heat-inactivated newborn calf serum/0.02% NaN_3/phenol-free RPMI). Cells were incubated with streptavidin-BV605 (BD Biosciences)–labeled β S-2P and streptavidin-BUV661 (BD Biosciences)–labeled WA-1 S-2P for 30 min at 4°C, washed twice in 1× PBS, and incubated with Aqua live/dead fixable dead cell stain (Thermo Fisher Scientific) for 20 min at room temperature. After two washes in wash buffer, cells were incubated with primary antibodies for 20 min at room temperature. The following antibodies were used (monoclonal unless indicated): IgD fluorescein isothiocyanate (FITC, goat polyclonal, Southern Biotech), IgM PerCP-Cy5.5 (clone G20-127, BD Biosciences), IgA Dylight 405 (goat polyclonal, Jackson ImmunoResearch), CD20 BV570 (clone 2H7, BioLegend), CD27 BV650 (clone O323, BioLegend), CD14 BV785 (clone M5E2, BioLegend), CD8 BUV395 (clone RPA-T8, BD Biosciences), CD16 BUV496 (clone 3G8, BD Biosciences), CD4 BUV737 (clone SK3, BD Biosciences), CD19 APC (clone J3-119, Beckman Coulter), IgG Alexa Fluor 700 (clone G18-145, BD Biosciences), CD3 APC-Cy7 (clone SP34.2, BD Biosciences), CD38 PE (clone OKT10, Caprico Biotechnologies), CD21 PE-Cy5 (clone B-ly4, BD Biosciences), and CXCR5 PE-Cy7 (clone MU5UBEE, Thermo Fisher Scientific). Cells were washed twice in wash buffer, and residual red blood cells were lysed using BD FACS Lysing Solution (BD Biosciences) for 10 min at room temperature. After two additional washes, cells were fixed in 0.5% formaldehyde (Tousimis Research) All antibodies were previously titrated to determine the optimal concentration. Samples were acquired on an BD FACSymphony flow cytometer and analyzed using FlowJo software version 10.7.2 (BD Biosciences).

Intracellular cytokine staining

Cryopreserved PBMCs were thawed and rested overnight in a 37°C, 5% CO_2 incubator. The next morning, cells were stimulated with SARS-CoV-2 S protein peptide pools (S1 and S2) that were matched to the vaccine insert [composed of 158 and 157 individual peptides, respectively, as 15mers overlapping by 11 amino acids in 100% dimethyl sulfoxide (DMSO), JPT Peptides] at a final concentration of 2 μg/ml in the presence of 3 mM monensin for 6 hours. Negative controls received an equal concentration of DMSO instead of peptides (final concentration 0.5%). Intracellular cytokine staining was performed as described previously (*46*). The following mAbs were used: CD3 APC-Cy7 (clone SP34.2, BD Biosciences), CD4 PE-Cy5.5 (clone S3.5, Invitrogen), CD8 BV570 (clone RPA-T8, BioLegend), CD45RA PE-Cy5 (clone 5H9, BD Biosciences), CCR7 BV650 (clone G043H7, BioLegend), CXCR5 PE (clone MU5UBEE, Thermo Fisher), CXCR3 BV711 (clone 1C6/CXCR3, BD Biosciences), PD-1 BUV737 (clone EH12.1, BD Biosciences), ICOS Pe-Cy7 (clone C398.4A, BioLegend), CD69 ECD (cloneTP1.55.3, Beckman Coulter), interferon-γ (IFN-γ) Ax700 (clone B27, BioLegend), interleukin-2 (IL-2) BV750 (clone MQ1-17H12, BD Biosciences), IL-4 BB700 (clone MP4-25D2, BD Biosciences), tumor necrosis factor (TNF)–FITC (clone Mab11, BD Biosciences), IL-13 BV421 (clone JES10-5A2, BD Biosciences), IL-17 BV605 (clone BL168, BioLegend), IL-21 Ax647 (clone 3A3-N2.1, BD Biosciences), and CD154 BV785 (clone 24-31, BioLegend). The Aqua live/dead fixable dead cell stain kit (Thermo Fisher Scientific) was used to exclude dead cells. All antibodies were previously titrated to determine the optimal concentration. Samples were acquired on a BD FACSymphony flow cytometer and analyzed using FlowJo software version 9.9.6 (TreeStar). The gating schema is represented in fig. S8.

Statistical analysis

Graphs show data from individual animals, with dotted lines indicating assay limits of detection. Groups were compared for viral load and immune responses using Welch's *t* tests; data were analyzed on the log_{10} scale for viral loads and appropriate immune assays. Changes over time were summarized using the change on the log_{10} scale, and statistical significance was determined using paired *t* tests. When more than two variants were compared, as in Fig. 1, A to F, a Holm's adjustment for multiple comparisons was used on the set of *P* values to determine significance. There were no adjustments for multiple comparisons across different analyses in this descriptive study. Comparisons drawn between vaccination groups for serum antibody epitope analysis were performed using unpaired *t* tests, within each time point evaluated. Analyses were performed in R software version 4.0.2 and GraphPad Prism software versions 8.2 and 9.0.2.

REFERENCES AND NOTES

1. F. P. Polack *et al.*, *N. Engl. J. Med.* **383**, 2603–2615 (2020).
2. L. R. Baden *et al.*, *N. Engl. J. Med.* **384**, 403–416 (2021).
3. M. Voysey *et al.*, *Lancet* **397**, 99–111 (2021).
4. J. Sadoff *et al.*, *N. Engl. J. Med.* **384**, 2187–2201 (2021).
5. P. T. Heath *et al.*, *N. Engl. J. Med.* **385**, 1172–1183 (2021).
6. C. Liu *et al.*, *Cell* **184**, 4220–4236.e13 (2021).
7. A. Pegu *et al.*, *Science* **373**, 1372–1377 (2021).
8. Q. Li *et al.*, *Cell* **182**, 1284–1294.e9 (2020).
9. H. Tegally *et al.*, *Nature* **592**, 438–443 (2021).
10. C. K. Wibmer *et al.*, *Nat. Med.* **27**, 622–625 (2021).
11. E. Andreanoc *et al.*, bioRxiv 424451 [Preprint] (2020).
12. M. Hoffmann *et al.*, *Cell* **184**, 2384–2393.e12 (2021).
13. W. F. Garcia-Beltran *et al.*, *Cell* **184**, 2372–2383.e9 (2021).
14. N. Doria-Rose *et al.*, *N. Engl. J. Med.* **384**, 2259–2261 (2021).
15. K. Wu *et al.*, *N. Engl. J. Med.* **384**, 1468–1470 (2021).
16. T. Tada *et al.*, bioRxiv 452771 [Preprint] (2021).
17. K. Wu *et al.*, medRxiv 2021.2005.2005.21256716 [Preprint] (2021).
18. K. S. Corbett *et al.*, *Science* **373**, eabj0299 (2021).
19. K. S. Corbett *et al.*, *N. Engl. J. Med.* **383**, 1544–1555 (2020).
20. M. Guebre-Xabier *et al.*, *Vaccine* **38**, 7892–7896 (2020).
21. N. B. Mercado *et al.*, *Nature* **586**, 583–588 (2020).
22. J. Yu *et al.*, *Science* **369**, 806–811 (2020).
23. P. J. Klasse, D. F. Nixon, J. P. Moore, *Sci. Adv.* **7**, eabe8065 (2021).
24. N. van Doremalen *et al.*, *Nature* **586**, 578–582 (2020).
25. E. S. Winkler *et al.*, *Cell* **184**, 1804–1820.e16 (2021).
26. J. R. Francica *et al.*, *Sci. Transl. Med.* **13**, eabi4547 (2021).
27. L. Wang *et al.*, *Science* **373**, eabh1766 (2021).
28. G. D. Victora, M. C. Nussenzweig, *Annu. Rev. Immunol.* **30**, 429–457 (2012).
29. J. S. Turner *et al.*, *Nature* **595**, 421–425 (2021).

30. A. Cho *et al*., *bioRxiv*, 2021.2007.2029.454333 (2021).
31. N. Pardi *et al*., *J. Exp. Med.* **215**, 1571–1588 (2018).
32. H. Chemaitelly *et al*., *Nat. Med.* **27**, 1614–1621 (2021).
33. Y. M. Bar-On *et al*., *N. Engl. J. Med.* **385**, 1393–1400 (2021).
34. S. Gazit *et al*., medRxiv 2021.2006.2029.21259579 [Preprint] (2021).
35. D. S. Khoury *et al*., *Nat. Med.* **27**, 1205–1211 (2021).
36. C. Gaebler *et al*., *Nature* **591**, 639–644 (2021).
37. Z. Wang *et al*., *Nature* **595**, 426–431 (2021).
38. K. Wu *et al*., bioRxiv 439482 [Preprint] (2021).
39. C.-W. Tan *et al*., *N. Engl. J. Med.* **385**, 1401–1406 (2021).
40. K. S. Corbett *et al*., *Nat. Immunol.* **22**, 1306–1315 (2021).
41. D. Wrapp *et al*., *Science* **367**, 1260–1263 (2020).
42. J. Pallesen *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, E7348–E7357 (2017).
43. K. J. Hassett *et al*., *Mol. Ther. Nucleic Acids* **15**, 1–11 (2019).
44. L. A. Jackson *et al*., *N. Engl. J. Med.* **383**, 1920–1931 (2020).
45. M. A. Whitt, *J. Virol. Methods* **169**, 365–374 (2010).
46. M. M. Donaldson, S. F. Kao, K. E. Foulds, *Cytometry A* **95**, 261–263 (2019).

ACKNOWLEDGMENTS

We thank J. Stein and M. Young for technology transfer and administrative support, respectively; members of the National Institutes of Health (NIH) National Institute of Allergy and Infectious Diseases (NIAID) Vaccine Research Center (VRC) Translational Research Program, including H. Bao, E. McCarthy, J. Noor, A. Taylor, and R. Woodward, for technical and administrative assistance with animal experiments; H. Mu and M. Farzan for the ACE2-overexpressing 293 cells; B. Zhang for the production and purification of published mAbs; T. Ruckwardt, E. Phung, I.-T. Teng, A. Olia, O. Abiona, and A. DiPiazza for contributions to probe production and validation; and M. Brunner and M. Whitt for kind support on recombinant VSV-based SARS-CoV-2 pseudovirus production. **Funding:** This work was supported by the Intramural Research Program of the VRC, NIAID, and NIH; the Department of Health and Human Services, Office of the Assistant Secretary for Preparedness and Response, Biomedical Advanced Research and Development Authority (contract no. 75A50120C00034); the Undergraduate Scholarship Program, Office of Intramural Training and Education, Office of the Director, NIH (K.S.C.); an Emory Executive Vice President for Health Affairs Synergy Fund Award (M.S.S.); the Pediatric Research Alliance Center for Childhood Infections and Vaccines and Children's Healthcare of Atlanta (M.S.S.); and the Woodruff Health Sciences Center 2020 COVID-19 CURE Award (M.S.S.). **Author contributions:** K.S.C., M.G., D.W., S.O., S.R.N., D.R.F., S.F.A., R.L.D., B.F., T.S.J., C.D.S., L.L., D.V., A.V.R., Z.F., A.P.W., J.I.M., M.S., S.O., S.D.S, C.T., A.C., M.K., K.W.B., M.M., B.M.N., G.S.A., A.R.H., F.L., C.A.S., E.S.Y., L.W., E.L., S.T.N., S.J.P., M.M.D., J.M., J-P.M.T., A.C., A.D., N.D., A.P., L.P., H.A., K.E.F., J.M., K.W., D.K.E., M.C.N., J.M., I.N.M., A.C., M.G.L., M.S.S., M.R., A.M., D.C.D., N.J.S., B.S.G., and R.A.S. designed, completed, analyzed, and/or supervised experiments. Y.Z., L.W., S.B-B., M.C., W.S., A.P., and E.B. provided critical published reagents and analytic tools. K.S.C., M.C.N., N.J.S., M.R., B.S.G, and R.A.S. wrote the manuscript. K.S.C., D.W., M.S., K.E.F., S.F.A., and G.A. prepared the figures and tables. All authors contributed to discussions about and editing of the manuscript. **Competing interests:** K.S.C. and B.S.G. are inventors on US Patent 10,960,070 B2 and International Patent Application WO/2018/081318 entitled "Prefusion Coronavirus Spike Proteins and Their Use." K.S.C. and B.S.G. are inventors on US Patent Application 62/972,886 entitled "2019-nCoV Vaccine." J.M., L.W., C.A.S., J.R.M, D.D, N.J.S., A.R.H., W.S., Y.Z., and M.R. are inventors on US Patent Application 63/147,419 entitled "Antibodies Targeting the Spike Protein of Coronaviruses." K.S.C., B.S.G, L.W., W.S., and Y.Z. are inventors on multiple US Patent applications entitled "Anti-Coronavirus Antibodies and Methods of Use." A.C., M.K., A.C., and D.K.E. are employees of Moderna. M.S.S. serves on the scientific board of advisors for Moderna. D.V., A.V.R., Z.F., A.C., L.P., H.A., and M.G.L. are employees of BIOQUAL Inc. **Data and materials availability:** All data are available in the main text or the supplementary materials. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS
science.org/doi/10.1126/science.abl8912
Figs. S1 to S10
Tables S1 to S5
MDAR Reproducibility Checklist

11 August 2021; accepted 15 October 2021
Published online 21 October 2021
10.1126/science.abl8912

CORONAVIRUS

Membrane fusion and immune evasion by the spike protein of SARS-CoV-2 Delta variant

Jun Zhang[1,2]†, Tianshu Xiao[1,2]†, Yongfei Cai[1,2], Christy L. Lavine[3], Hanqin Peng[1], Haisun Zhu[4], Krishna Anand[4], Pei Tong[5,6], Avneesh Gautam[5,6], Megan L. Mayer[7,8], Richard M. Walsh Jr.[7,8], Sophia Rits-Volloch[1], Duane R. Wesemann[5,6], Wei Yang[4]‡, Michael S. Seaman[3], Jianming Lu[9,10], Bing Chen[1,2]*

The Delta variant of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has outcompeted previously prevalent variants and become a dominant strain worldwide. We report the structure, function, and antigenicity of its full-length spike (S) trimer as well as those of the Gamma and Kappa variants, and compare their characteristics with the G614, Alpha, and Beta variants. Delta S can fuse membranes more efficiently at low levels of cellular receptor angiotensin converting enzyme 2 (ACE2), and its pseudotyped viruses infect target cells substantially faster than the other five variants, possibly accounting for its heightened transmissibility. Each variant shows different rearrangement of the antigenic surface of the amino-terminal domain of the S protein but only makes produces changes in the receptor binding domain (RBD), making the RBD a better target for therapeutic antibodies.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of the COVID-19 pandemic (*1*). The strain responsible for the initial outbreak, Wuhan-Hu-1 (*1*), was the basis for first-generation vaccine development. We previously characterized two early variants of concern (VOC): Alpha and Beta (*2*). The Delta variant (*3*) (also known as lineage B.1.617.2) was first detected in India and was quickly characterized as a VOC and has since outcompeted other variants to become a globally dominant strain within several months. It is estimated to be about twice as transmissible as Wuhan-Hu-1 (*4*, *5*). Infection by the Delta variant appears to have a shorter incubation period with a viral load ~1000 times greater in the first positive Polymerase Chain Reaction (PCR) test than earlier variants (*6*). It remains uncertain whether it causes more severe disease (*7*, *8*), but it does have some resistance to immunity elicited by first-generation vaccines (*9*–*12*). Another VOC, Gamma (lineage B.1.1.28 or P.1), has spread in Brazil and some other countries (*13*, *14*). A third variant, Kappa (lineage B.1.617.1), also first reported in India, remains a variant of interest (VOI) but had only a limited surge (*15*, *16*). It is critical to understand the molecular mechanisms of the increased transmissibility and immune evasion of variants to guide intervention strategies.

SARS-CoV-2 is an enveloped, positive-stranded RNA virus that enters a host cell by fusing its lipid bilayer with the target cell membrane. The fusion reaction is facilitated by the virus-encoded trimeric spike (S) protein after it binds to the host angiotensin converting enzyme 2 (ACE2). The S protein is produced as a single-chain precursor, and processed by a host furin-like protease into the receptor-binding fragment S1 and the fusion fragment S2 (fig. S1) (*17*). After engaging with ACE2 on the host cell surface, the S protein is cleaved by a second cellular protease in S2 (S2' site; fig. S1) (*18*), initiating S1 dissociation and a cascade of S2

[1]Division of Molecular Medicine, Boston Children's Hospital, Harvard Medical School, 3 Blackfan Street, Boston, MA 02115, USA. [2]Department of Pediatrics, Harvard Medical School, 3 Blackfan Street, Boston, MA 02115, USA. [3]Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Boston, MA, 02215, USA. [4]Institute for Protein Innovation, Harvard Institutes of Medicine, 4 Blackfan Circle, Boston, MA 02115, USA. [5]Division of Allergy and Clinical Immunology, Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA. [6]Ragon Institute of MGH, MIT, and Harvard, Boston, MA 02115, USA. [7]The Harvard Cryo-EM Center for Structural Biology, Harvard Medical School, 250 Longwood Avenue, Boston, MA 02115, USA. [8]Department of Biological Chemistry and Molecular Pharmacology, Blavatnik Institute, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115, USA. [9]Codex BioSolutions, Inc., 401 Professional Drive, Gaithersburg, MD 20879, USA. [10]Department of Biochemistry and Molecular and Cellular Biology, Georgetown University School of Medicine, 3900 Reservoir Road, N.W., Washington, D.C. 20057, USA.
***Corresponding author. Email: bchen@crystal.harvard.edu**
†These authors contributed equally to this work. ‡Present address. GV20 Therapeutics LLC, 1 Broadway Floor 14, Cambridge, MA 02142, USA.

refolding events to drive membrane fusion (*19*, *20*). S1 contains four domains: NTD (N-terminal domain), RBD (receptor-binding domain), and two CTDs (C-terminal domains), protecting the central helical bundle structure of the prefusion S2. The RBD can adopt either a 'down' conformation for a receptor-inaccessible state, or an 'up' conformation for a receptor-accessible state (*21*); movement of the RBD allows the virus to protect the critical receptor-binding site from host immune responses (*21*, *22*).

Intensive studies on the S protein have advanced our knowledge of SARS-CoV-2 entry substantially (*23*–*26*). Here, we have characterized the full-length S proteins of the Delta, Kappa, and Gamma variants, and determined their structures by cryogenic electron microscopy (cryo-EM). Comparison of the structure, function, and antigenicity of Delta S with those of Gamma and Kappa—as well as the previously characterized Alpha and Beta variants (*2*)—provides molecular insights into the mechanisms of the heightened transmissibility and enhanced immune evasion of the most contagious form of SARS-CoV-2 since its initial outbreak.

Membrane fusion by Delta S is substantially faster than that of other variants

To characterize the full-length S proteins with the sequences derived from natural isolates of the Gamma (hCoV-19/Brazil/AM-992/2020), Kappa (hCoV-19/India/MH-NEERI-NGP-40449/2021), and Delta (hCoV-19/India/GJ-GBRC619/2021) variants (fig. S1), we transfected HEK293 cells with the respective expression constructs and compared their fusion activities with that of the full-length S construct of their parental strain (G614 or B.1 variant) (*27*). All S proteins were expressed at comparable levels (fig. S2A). Kappa S had <5% cleavage between S1 and S2 compared with ~40% cleavage for other variants at the time the cells were harvested, suggesting that the P681R mutation (found in Delta and Kappa) near the furin cleavage site does not increase furin processing. The extent of cleavage in Delta is not substantially altered from that in its parent strains (fig. S2A). The cells producing these S proteins fused efficiently with ACE2-expressing cells, as expected (fig. S2B). The fusion activity of Kappa S was ~50% that of other S proteins at a low transfection level, presumably as a result of the low furin cleavage, but the difference diminished at high transfection levels (fig. S2B).

To test whether more efficient fusion accounts for Delta S transmissibility, we performed a time-course experiment with a cell-cell fusion assay, with both S and ACE2 transfected at high levels (fig. S3A). We found no notable differences in fusion activity among G614, Alpha, Beta, Gamma, Delta, and Kappa. Notably, Delta S-expressing cells fused with the negative-control HEK293 cells more efficiently than other variants, particularly at longer time points (Fig. 1A and fig. S3B). HEK293 cells, expressing a minimal level of endogenous ACE2, are used as negative controls when not transfected by the ACE2 expression construct in our standard 2-hour fusion protocol (*28*). The same pattern was reproduced when small amounts of ACE2 were introduced in HEK293 cells, but the differences diminished when the ACE2 transfection level increased (Fig. 1B and fig. S3C). These data suggest that Delta S can enter a host cell expressing low levels of ACE2 more efficiently than other variants.

We performed a similar time-course experiment by using murine leukemia virus (MLV)–based pseudoviruses expressing the cytoplasmic tail-truncated S constructs to facilitate incorporation into particles (*29*, *30*). The infection was initiated by mixing the viruses and target cells, and the viruses were washed out at each time point. The Delta variant established

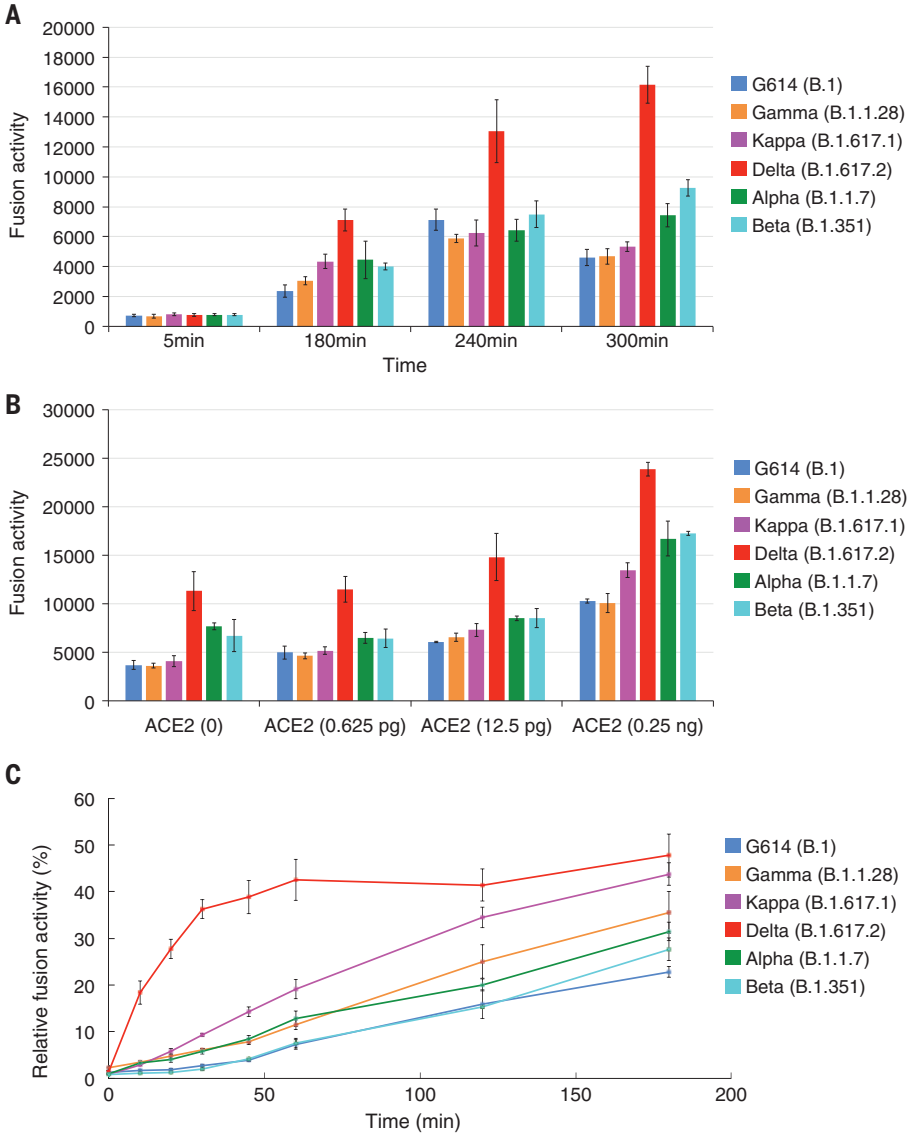


Fig. 1. More efficient membrane fusion by the Delta variant than other variants. (**A**) Time course of cell-cell fusion mediated by various full-length S proteins, as indicated by the labels, with HEK293 cells with no exogenous ACE2. (**B**) Cell-cell fusion mediated by various full-length S proteins with HEK293 cells transfected with low levels (0 to 0.25 ng) of ACE2 expression constructs. (**C**) Time course of infection HEK293-ACE2 cells by MLV-based, pseudotyped viruses by various SARS-CoV-2 variant S constructs containing a CT deletion in a single cycle. Infection was initiated by mixing viruses and target cells, and viruses were washed out at each time point as indicated. The full time course and concentration series are shown in fig. S3. The experiments were repeated at least three times, with independent samples each giving similar results.

infection much more rapidly in the first 60-min period than did any other variant, when infectivity was normalized to its maximum level (Fig. 1C). The other variants caught up over time, reaching their maximum levels at 8 hours (fig. S3D). Some viruses—including Delta—reproducibly showed lower measurements for the no wash-out controls than those measured at the 8-hour time point, consistent with some cytotoxicity reducing the reporter-gene expression. Taken together, these findings suggest that the Delta variant can infect a target cell more rapidly than the other variants tested, either by more effective attachment or by faster fusion kinetics.

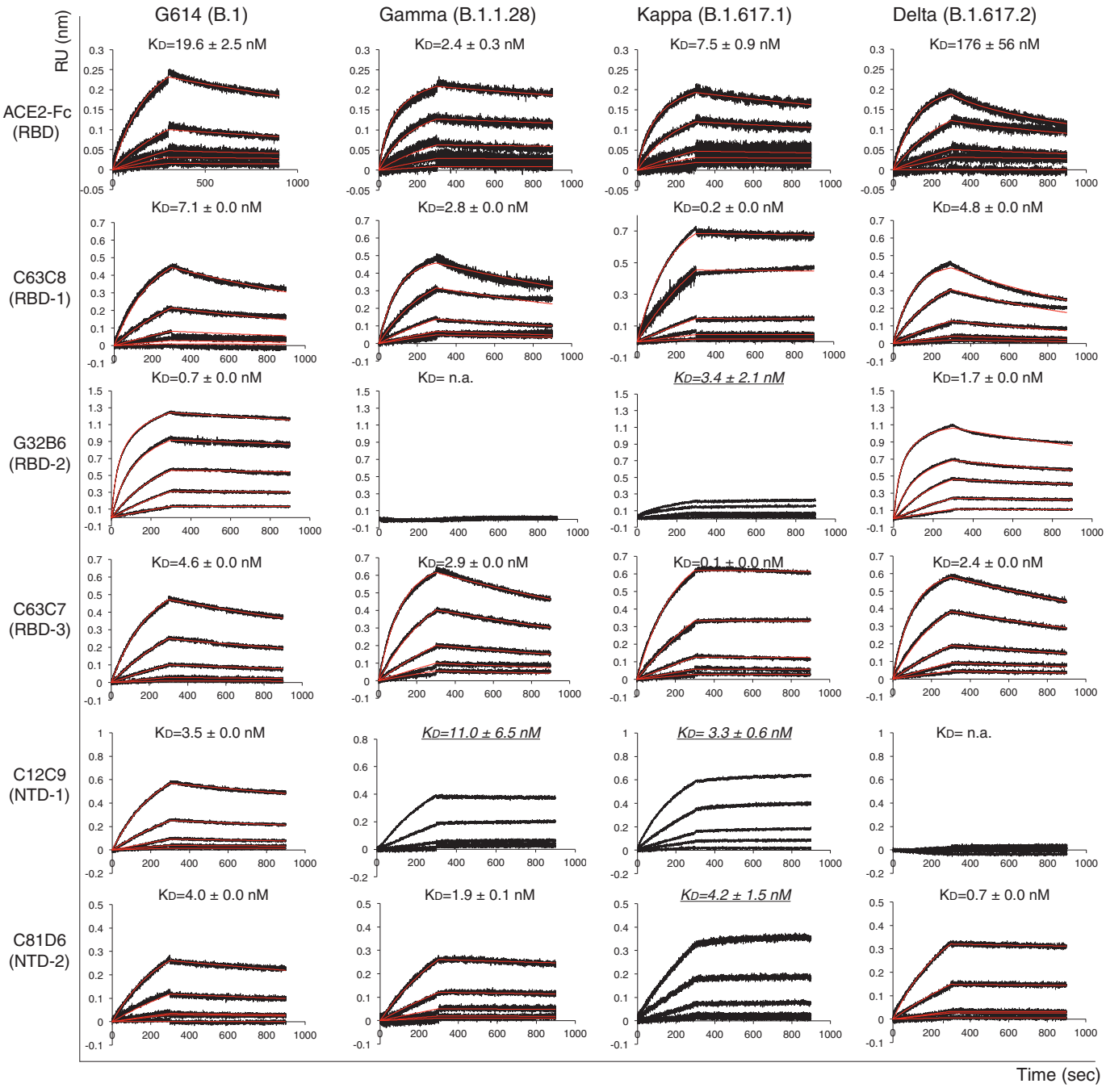


Fig. 2. Antigenic properties of purified full-length SARS-CoV-2 S proteins. Biolayer interferometry (BLI) analysis of the association of prefusion S trimers derived from the G614 "parent" strain (B.1) and the Gamma (B.1.1.28), Kappa (B.1.617.1), and Delta (B.1.617.2) variants with soluble ACE2 constructs and with a panel of antibodies representing five epitopic regions on the RBD and NTD (see fig. S4A) (*32*). For ACE2 binding, purified S proteins were immobilized to AR2G biosensors and dipped into wells containing ACE2 at various concentrations. For antibody binding, various antibodies were immobilized to AHC biosensors and dipped into wells containing each purified S protein at different concentrations. Binding kinetics were evaluated by a 1:1 Langmuir model except for dimeric ACE2 and antibody G32B6 targeting the RBD-2, which were analyzed by a bivalent binding model. All K_D values for multivalent interactions with antibody IgG or dimeric ACE2 and trimeric S protein are the apparent affinities with avidity effects. Sensorgrams are in black and fits are in red. Binding constants highlighted by underlines were estimated by steady state analysis as described in the materials and methods. RU, response unit. Binding constants are summarized both here and in table S1. All experiments were repeated at least twice with essentially identical results.

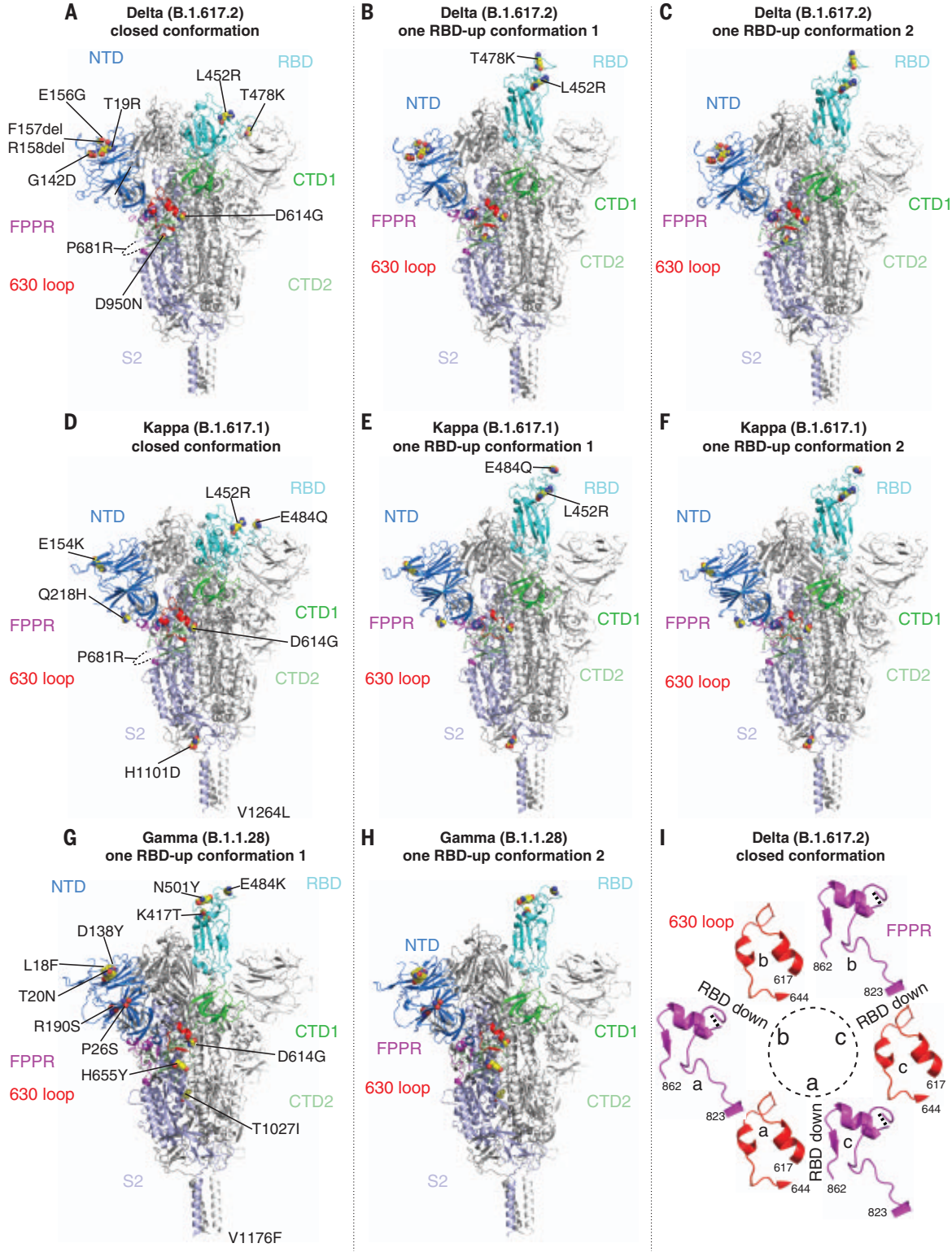


Fig. 3. Cryo-EM structures of full-length SARS-CoV-2 S proteins from the Delta, Kappa, and Gamma variants. (**A** to **C**) The structures of the closed prefusion conformation and two one-RBD-up conformations of the Delta S trimer are shown in the ribbon diagrams, with one protomer colored as follows: NTD in blue, RBD in cyan, CTD1 in green, CTD2 in light green, S2 in light blue, the 630 loop in red, and the FPPR in magenta. (**D** to **F**) The structures of the closed prefusion conformation and two one-RBD-up conformation of the Kappa S trimer are shown in the ribbon diagrams, with the same color scheme as in (A). (**G**) and (**H**) The structures of the two one-RBD-up conformations of the Gamma S trimer are shown in the ribbon diagrams with the same color scheme as in (A). All mutations in the three variants, as compared to the original virus (D614), are highlighted in the sphere model. (**I**) Structures in the Delta closed trimer of segments (residues 617 to 644) containing the 630 loop (red) and segments (residues 823 to 862) containing the FPPR (magenta) from each of the three protomers (A), (B), and (C). The position of each RBD is indicated. Dashed lines indicate gaps in the chain trace (disordered loops).

Biochemical and antigenic properties of intact S proteins from the variants

We added a C-terminal strep-tag to the full-length S proteins of the Gamma, Kappa, and Delta variants (fig. S4A), and expressed and purified them by the procedures established for the Wuhan-Hu-1 S trimer (*28*). The Gamma protein eluted in three distinct peaks, corresponding to the prefusion S trimer, post-fusion S2 trimer, and dissociated S1 monomer, respectively (*28*), as shown by Coomassie-stained SDS-PAGE analysis (fig. S4B). The prefusion trimer accounted for <40% of the total protein, similar to the profile of the Wuhan-Hu-1 protein, indicating that this trimer is not very stable. Although the Kappa protein eluted in one major peak corresponding to the prefusion trimer, there was a considerable amount of aggregate on the leading side and a large shoulder on the trailing side, suggesting that the protein is also conformationally heterogeneous. Moreover, a large fraction of the protein remained uncleaved (fig. S4B), confirming that the furin cleavage is inefficient despite the P681R mutation. By contrast, the Delta protein eluted in a single symmetrical peak of the prefusion trimer showing little aggregation or dissociation, and it appears to be the most stable trimer preparation among all the full-length S proteins that we have examined (fig. S3B) (*31*). Negative stain EM confirmed these results (fig. S5). SDS-PAGE analysis showed that the Delta trimer peak primarily contained the cleaved S1-S2 complex with a cleavage level very similar to that of the G614 and Beta S proteins (*2*, *31*), indicating that P681R has little effect on the furin cleavage.

To analyze the antigenicity of these S trimers, we measured their binding to soluble ACE2 proteins and S-directed monoclonal antibodies isolated from COVID-19 convalescent individuals by biolayer interferometry (BLI). The selected antibodies recognize distinct epitopic regions on the S trimer, as defined by antibody competition for binding, designated RBD-1, RBD-2, RBD-3, NTD-1, NTD-2, and S2 (fig. S6A) (*32*). The binding of the Gamma variant to the receptor was substantially stronger than that of its G614 parent, regardless of the ACE2 oligomeric state (Fig. 2, fig. S6B, and table S1), likely because of its mutations (K417T, E484K, and N501Y) in the RBD. ACE2 affinities for Kappa and Delta were intermediate between those of the G614 and Gamma trimers, with Kappa and Delta closer to Gamma and G614, respectively, except for binding of Delta with dimeric ACE2, which had an unexpectedly higher off rate than the other variants (Fig. 2). These data were largely confirmed by monomeric RBD preparations instead of the S trimers, except that the Kappa RBD showed slightly higher ACE2 affinity than the Gamma RBD (fig. S6B and table S1). ACE2 did not dissociate more rapidly from the Delta RBD than it did from the Gamma and Kappa RBDs; one possible explanation for the apparently weaker affinity of the Delta trimer for the ACE2 dimer than other variants is that ACE2 binding induces S1 dissociation. These results suggest that the RBD mutations of the Gamma variant enhance receptor recognition, whereas those in Kappa (L452R and E484Q) and Delta (L452R and T478K) have a smaller effect on ACE2 affinity. The dimeric ACE2 appears to be more effective in inducing S1 dissociation from the Delta S trimer than from any other variant.

All selected antibodies had reasonable affinities for the G614 trimer (Fig. 2, fig. S6B, and table S1). The Gamma variant lost binding to the two RBD-2 antibodies (G32B6 and C12A2) and to one NTD-1 antibody (C83B6), but retained binding to the NTD-1 antibody C12C9 with somewhat reduced affinity, suggesting that these two NTD-1 antibodies target overlapping but distinct epitopes (*32*). Its affinities for the remaining antibodies were the same as those of the G614 trimer. Binding of the Kappa trimer showed unrealistically slow off-rates for several antibodies (Fig. 2 and fig. S6B), presumably because of aggregation and conformational heterogeneity. Qualitatively, it had substantially weakened binding to the RBD-2 antibodies and the NTD-1 antibody C83B6, but with wildtype or even enhanced affinity for another NTD-1 antibody (C12C9). Thus, the changes in the Kappa antigenicity show trends similar to those of Gamma S. Delta S

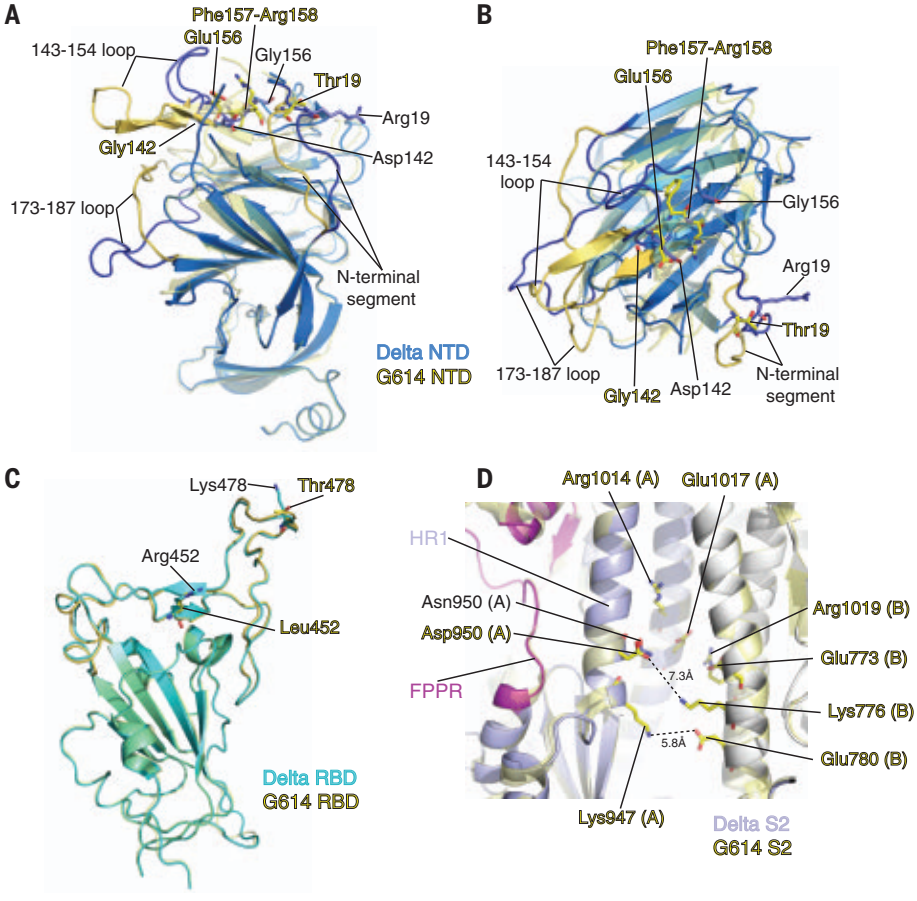


Fig. 4. Structural impact of mutations in Delta S. (**A**) Superposition of the NTD structure of the Delta S trimer (blue) with the NTD of the G614 S trimer (PDB ID: 7KRQ) (yellow). Locations of mutations T19R, G142D, E156G, and deletion of F157 and R158 are indicated; these residues are shown in the stick model. The N-terminal segment, as well as loops 143 to 154 and 173 to 187, are rearranged between the two structures and highlighted in darker colors. (**B**) Top view of panel (A). (**C**) Superposition of the RBD structure of the Delta S trimer (cyan) with the RBD of the G614 S trimer (yellow). Locations of mutations L452R and T478K are indicated; these residues are shown in the stick model. (**D**) A close-up view of superposition of the Delta S2 (light blue) with the S2 of the G614 S trimer (yellow) near residue 950. Locations of the D950N mutation and charged residues in the vicinity including Lys^{947}, Arg^{1014}, and Glu^{1017} from protomer A and Glu^{773}, Lys^{776}, Glu^{780}, and Arg^{1019} from the protomer B are indicated. All aforementioned residues are shown in the stick model.

only lost binding to the two NTD-1 antibodies with little change in affinities for the other antibodies, including those targeting the RBD (Fig. 2, fig. S6B, and table S1). The BLI data were also largely consistent with the binding results for the membrane-bound S trimers measured by flow cytometry (fig. S7).

We next analyzed the neutralization potency of these antibodies and trimeric soluble ACE2 (*33*) by measuring the extent to which they blocked infection by S variants in an HIV-based pseudovirus assay. For most antibodies, the neutralization potency correlated with their binding affinity for the membrane-bound or purified S proteins (table S2). C81D6 and C163E6 recognized two nonneutralizing epitopes in the NTD-2 and S2, respectively, and did not neutralize any of the pseudoviruses. Thus, the mutations in the Gamma and Kappa variants have a greater effect on antibody sensitivity than those in the Delta variant.

Overall structures of the intact S trimers of the Delta, Kappa and Gamma variants

We determined the cryo-EM structures of the full-length S trimers with the unmodified sequences of the Delta, Kappa, and Gamma variants, according to our established procedures (*2*, *28*, *31*). 3D classification gave three distinct classes each for both the Delta and Kappa trimers, representing one closed prefusion conformation and two one-RBD-up conformations, respectively. There were two classes for the Gamma trimer, representing two one-RBD-up conformations. These structures were refined to 3.1 to 4.4 Å resolution (figs. S8 to S14 and table S3). There are no major changes in the overall architectures of the full-length variant S proteins when compared with that of the parental G614 S trimer in the corresponding conformation (Fig. 3 and fig. S15) (*31*). The furin cleavage site (residues 682 to 685) at the S1-S2 boundary, including the P681R substitution, was not visible in these maps.

We have proposed that the FPPR (fusion peptide proximal region; residues 828 to 853) and 630 loop (residues 620 to 640) are control elements, and that shifts in their positions modulate the S stability and structural rearrangements (*28*, *31*). For the Delta and Kappa variants, the FPPR and 630 loop configurations are largely consistent with the distribution in the G614 trimer; all are structured in the RBD-down conformation, whereas only one FPPR and 630-loop pair is ordered in the one-RBD-up conformations. The density for FPPR residues 841 to 847 of the Delta S in the closed state is weak, probably because of slight (1 to 2Å) downward shifts of the CTD1 and RBD, which may weaken the FPPR packing (fig. S15). No class representing the closed conformation was identified for Gamma S from three independent datasets (fig. S12), suggesting this conformation is not well occupied by that variant; however, one FPPR and 630-loop pair is structured in the one-RBD-up conformations of Gamma, probably stabilizing the cleaved S trimers before receptor engagement. In all three variants, the distinct one-RBD-up structures differ only by the degree to which the up-RBD and the adjacent NTD of its neighboring protomer shift away from the central threefold axis (fig. S15). An N-linked glycan at Asn^{343} has been implicated in a gating role for facilitating RBD opening (*34*). Its density is stronger in the maps of all the new variants, particularly Delta and Kappa, than that in the G614 map (fig. S16). The distal end of this glycan contacts the neighboring RBD, forming a ring-like density and apparently stabilizing the three-RBD-down conformation. Nonetheless, it remains unclear why the Gamma prefusion trimer dissociates, the Kappa trimer tends to aggregate, and the Delta trimer is the most stable of the three.

Structural consequences of mutations in the Delta variant

We superposed the structures of the Delta S trimer onto the G614 trimer in the closed conformation, aligning them by the S2 region

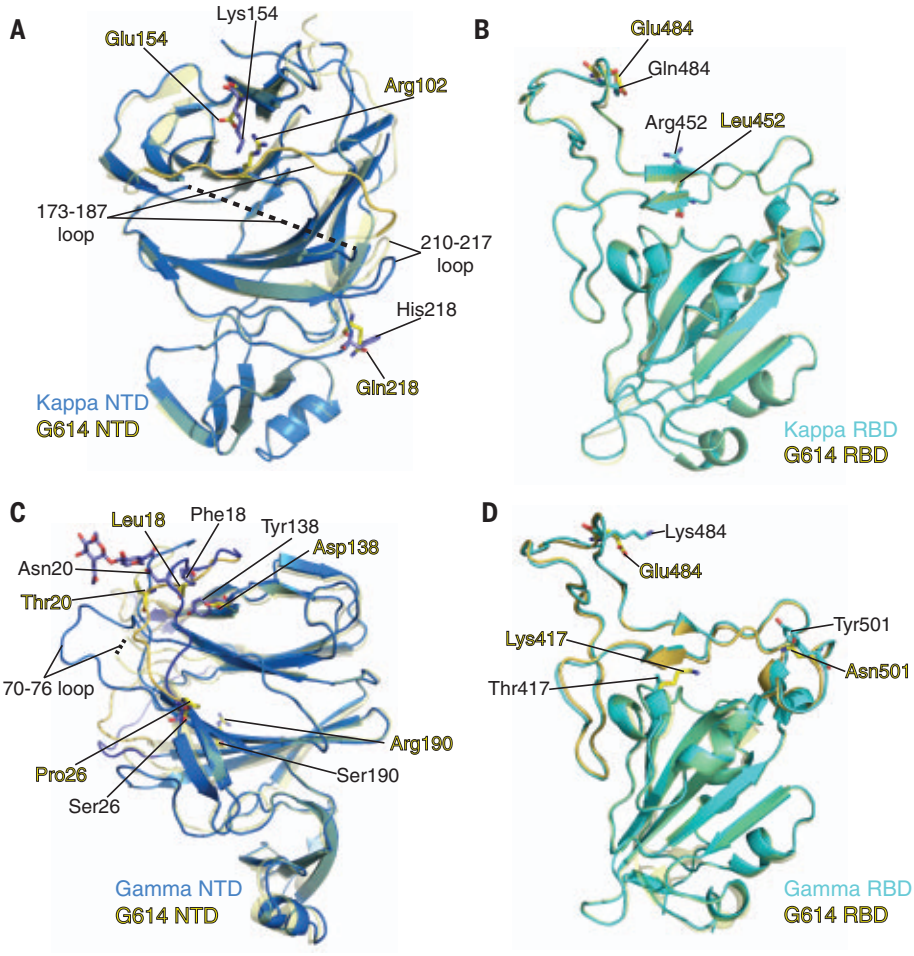


Fig. 5. Structural impact of mutations in the Kappa and Gamma S proteins. (**A**) Superposition of the NTD structure of the Kappa S trimer (blue) with the NTD of the G614 S trimer (yellow). Locations of mutations E154K and Q218H, as well as Arg^{102} which forms a salt bridge with Glu^{154} in the G614 structure, are indicated; these residues are shown in the stick model. The 173 to 187 loop in the G614 trimer is highlighted in a darker color; it becomes disordered in the Kappa trimer. (**B**) Superposition of the RBD structure of the Kappa S trimer (cyan) with the RBD of the G614 S trimer (yellow). Locations of mutations L452R and E484Q are indicated; these residues are shown in the stick model. (**C**) A view of superposition of the NTD structures of the Gamma (blue) and G614 (yellow; PDB ID: 7KRR) S trimers in the one-RBD-up conformation. Locations of mutations L18F, T20N, P26S, D138Y, and R190S are indicated, as well as the N-linked glycan attached to Asn^{20} in the Gamma structure; these residues are shown in the stick model. (**D**) Superposition of the RBD structure of the Gamma S trimer (cyan) with the RBD of the G614 S trimer (yellow). Locations of mutations K417T, E484K, and N501Y are indicated, and these residues are shown in the stick model.

(fig. S15) and revealing the most prominent differences in the NTD, which contains three point mutations (T19R, G142D, and E156G) and a two-residue deletion (F157del and R158del). When the two NTDs are aligned (Fig. 4A and 4B), the mutations reshape the 143-154 loop, which contains an N-linked glycan (N149) and forms part of the NTD-1 epitopes (*35*–*38*), projecting it away from the viral membrane. They also reconfigure the N-terminal segment and the 173-187 loop, substantially altering the antigenic surface near the NTD-1 epitopes, and consistent with loss of binding and neutralization by NTD-1 antibodies (Fig. 2, fig. S6, and table S2). There are no major structural rearrangements in the Delta RBD with two mutations, L452R and T478K (Fig. 4C). These residues are not in the ACE2 contacting surface, and have little influence on the receptor binding (fig. S17) (*39*). Neither binding nor neutralization of the Delta variant by most anti-RBD antibodies tested here have changed, suggesting that the two residues are also not in any major neutralizing epitopes. No obvious structural alterations were observed from the D950N substitution in HR1 (heptad repeat 1) of S2 (Fig. 4D), with multiple pairs of charged residues in the vicinity that could stabilize the packing between S2 protomers in the prefusion conformation.

Structural impact of the mutations in the Kappa and Gamma variants

There are only two mutations (E154K and Q218H) in the Kappa NTD (Fig. 5A). Glu^{154} forms a salt bridge with Arg^{102} in the G614 trimer (*31*); E154K substitution results in an unfavorable interaction with Arg^{102}, possibly leading to a disordered 173 to 187 loop nearby in the Kappa trimer. Residue 218 is surface-exposed and on the opposite side from the neutralizing epitopes. Q218H may contribute to rearrangement of the 210 to 217 and 173 to 187 loops (Fig. 5A). There are two RBD mutations (L452R and E484Q) in Kappa (Fig. 5B), which do not alter the overall structure of the domain. Glu^{484} forms a salt bridge with ACE2 Lys^{31} in the RBD-ACE2 complex (fig. S17) (*40*, *41*). The E484Q substitution loses the salt bridge, but hydrogen bonds between Gln^{484} and ACE2 Lys^{31} might compensate and thus account for a small increase in ACE2 binding affinity. L452R is unlikely to substantially affect ACE2 binding (fig. S17). The mutation H1101D in S2 caused little local change (fig. S18A), and V1264L is not visible in our structures.

Structural changes in the Gamma NTD caused by the mutations (L18F, T20N, P26S, D138Y, and R190S) were evident in the EM maps (fig. S14). All mutations except for R190S are located near the N-terminal segment and contribute to reconfiguration of its extended structure (Fig. 5C). The new conformation of the N-terminal segment appears to stabilize the 70 to 76 loop, disordered in most known S trimer structures (*21*, *28*, *42*). T20N has created a new glycosylation site and Asn^{20} is indeed glycosylated in Gamma (Fig. 5C). These changes apparently also shift the 143 to 154 and 173 to 187 loops (fig. S18B), leading to relatively large-scale rearrangement of the antigenic surface of the Gamma NTD. The three RBD mutations (K417T, E484K, and N501Y) in Gamma also produce no major structural rearrangements (Fig. 5D). N501Y increases receptor-binding affinity, which may be counteracted by K417T and E484K because of loss of ionic interactions with ACE2 (fig. S17) (*2*, *43*–*46*). K417T and E484K are probably responsible for loss of binding and neutralization of Gamma by antibodies that target the RBD-2 epitopes (*45*, *47*, *48*). H655Y in the CTD2 did not change the local structure (fig. S18C), but its location near the N terminus of the cleaved S2 suggests a role in destabilizing the Gamma S trimer. Finally, T1027I did not lead to any major changes in S2 (fig. S18D), and V1176F is in a disordered region.

The Delta variant of SARS-CoV-2 has rapidly replaced the previously dominant variants—including Alpha—which is itself ~60% more transmissible than the Wuhan-Hu-1 strain (*49*–*51*). Delta thus appears to have acquired an enhanced capacity for propagating in human cells. Several hypotheses have been proposed to explain its heightened transmissibility, including mutations in the RBD enhancing receptor engagement (*52*), P681R substitution near the S1-S2 boundary leading to more efficient furin cleavage (*53*, *54*), and changes in its RNA polymerase increasing viral replication. We cannot rule out the possibility that mutations in the viral replication machinery characteristic of Delta (e.g., G671S in nsp12) may increase the production of genomic RNA, but viral assembly into mature virions would require many other factors to achieve the >1,000-fold greater viral load in infected patients. We have not detected any notable increase in ACE2 binding by either the full-length Delta S trimer or its RBD fragment, nor have we observed more efficient cleavage in the Delta S than any other variants. Indeed, the furin cleavage is already very efficient in G614, Alpha, Beta, and Delta (*2*, *31*, *54*), and may no longer be a rate-limiting step for all these variants.

We have identified two properties, so far only found in the Delta variant, that might account for its transmissibility. First, when the Delta S protein is expressed on the cell surface at a saturating level, those cells fuse more efficiently with target cells that produce lower levels of ACE2 than do cells of any other variant (*2*). When the ACE2 expression level increases, the differences among the variants diminish. Second, the pseudoviruses containing the Delta S construct enter the ACE2-expressing cells more rapidly than other variants. These data suggest that the Delta S protein has evolved to optimize the fusion step for entering cells expressing low levels of the receptor. This optimization may explain why the Delta variant can transmit upon relatively brief exposure and infect many more host cells rapidly, leading to a short incubation period and greater viral load during infection. One caveat is that all our experiments were performed in vitro; additional studies with authentic viruses will be needed to confirm our findings in more clinically relevant settings.

The RBD and NTD are the two major sites on the S trimer targeted by neutralizing antibodies (*32*, *36*, *55*, *56*). The three strains studied here show once again how different variants can use different strategies to remodel their NTD and evade host immunity. One notable implication is that the NTD function does not require specific structural elements or sequences because the surface loops, β strands in the core structure, and even some N-linked glycans can be rearranged in different ways without compromising viral infectivity. By contrast, the overall structure of the RBD has been strictly preserved among all variants, and reoccurring surface mutations appear to be limited to a number of sites, consistent with its critical role in receptor binding. We therefore suggest that therapeutic antibodies or universal vaccines should not target the NTD, as escape from anti-NTD antibodies appears to be of little cost to the virus.

REFERENCES AND NOTES

1. P. Zhou *et al.*, *Nature* **579**, 270–273 (2020).
2. Y. Cai *et al.*, *Science* **373**, 642–648 (2021).
3. P. Mlcochova *et al.*, *Nature* **599**, 114–119 [Preprint] (2021).
4. R. Earnest *et al.*, *medRxiv*, 2021.10.06.21264641 [Preprint] (2021).
5. J. Dagpunar, *medRxiv*, 2021.2006.2003.21258293 [Preprint] (2021).
6. B. Li *et al.*, *medRxiv*, 2021.07.07.21260122v2 [Preprint] (2021).
7. A. Sheikh, J. McMenamin, B. Taylor, C. Robertson, *Lancet* **397**, 2461–2462 (2021).
8. D. N. Fisman, A. R. Tuite, CMAJ **193** (42) E1619-E1625 [Preprint] (2021).
9. D. Planas *et al.*, *Nature* **596**, 276–280 (2021).
10. V. V. Edara *et al.*, *N. Engl. J. Med.* **385**, 664–666 (2021).
11. J. Lopez Bernal *et al.*, *N. Engl. J. Med.* **385**, 585–594 (2021).
12. T. Tada *et al.*, *bioRxiv*, 2021.2007.2019.452771 [Preprint] (2021).
13. N. R. Faria *et al.*, *Science* **372**, 815–821 (2021).
14. M. Hoffmann *et al.*, *Cell* **184**, 2384–2393.e12 (2021).
15. P. D. Yadav *et al.*, *bioRxiv*, 2021.2005.2005.442760 [Preprint] (2021).
16. V.-V. Edara *et al.*, *N. Engl. J. Med.* **385**, 664–666 [Preprint] (2021).
17. B. J. Bosch, R. van der Zee, C. A. de Haan, P. J. Rottier, *J. Virol.* **77**, 8801–8811 (2003).
18. M. Hoffmann *et al.*, *Cell* **181**, 271–280.e8 (2020).
19. J. K. Millet, G. R. Whittaker, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15214–15219 (2014).
20. M. A. Tortorici, D. Veesler, *Adv. Virus Res.* **105**, 93–116 (2019).
21. D. Wrapp *et al.*, *Science* **367**, 1260–1263 (2020).
22. J. Shang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 11727–11734 (2020).

23. Y. Huang, C. Yang, X. F. Xu, W. Xu, S. W. Liu, *Acta Pharmacol. Sin.* **41**, 1141–1149 (2020).
24. C. Zhu *et al.*, *J. Med. Virol.* **93**, 5729–5741 (2021).
25. N. Murgolo *et al.*, *PLOS Pathog.* **17**, e1009225 (2021).
26. R. Peng, L. A. Wu, Q. Wang, J. Qi, G. F. Gao, *Trends Biochem. Sci.* **46**, 848–860 (2021).
27. B. Korber *et al.*, *Cell* **182**, 812–827.e19 (2020).
28. Y. Cai *et al.*, *Science* **369**, 1586–1592 (2020).
29. J. Yu *et al.*, *J. Virol.* **95**, e00044–e21 (2021).
30. T. Giroglou *et al.*, *J. Virol.* **78**, 9007–9015 (2004).
31. J. Zhang *et al.*, *Science* **372**, 525–530 (2021).
32. P. Tong *et al.*, *Cell* **184**, 4969–4980.e15 (2021).
33. T. Xiao *et al.*, *Nat. Struct. Mol. Biol.* **28**, 202–209 (2021).
34. T. Sztain *et al.*, *Nat. Chem.* **13**, 963–968 (2021).
35. G. Cerutti *et al.*, *Cell Host Microbe* **29**, 819–833.e7 (2021).
36. X. Chi *et al.*, *Science* **369**, 650–655 (2020).
37. M. McCallum *et al.*, *Cell* **184**, 2332–2347.e16 (2021).
38. D. Li *et al.*, *Cell* **184**, 4203–4219.e32 (2021).
39. M. McCallum *et al.*, *Science*, eabl8506 (2021).
40. J. Lan *et al.*, *Nature* **581**, 215–220 (2020).
41. J. Shang *et al.*, *Nature* **581**, 221–224 (2020).
42. A. C. Walls *et al.*, *Cell* **183**, 1735 (2020).
43. S. M. Gobeil *et al.*, *Science* **373**, eabi6226 (2021).
44. T. N. Starr *et al.*, *Cell* **182**, 1295–1310.e20 (2020).
45. M. Yuan *et al.*, *Science* **373**, 818–823 (2021).
46. T. J. Yang *et al.*, *Nat. Struct. Mol. Biol.* **28**, 731–739 (2021).
47. M. Ferrari *et al.*, *J. Virol.* **95**, e0068521 (2021).
48. P. Wang *et al.*, *Nature* **593**, 130–135 (2021).
49. E. Volz *et al.*, *Nature* **593**, 266–269 (2021).
50. M. U. G. Kraemer *et al.*, *Science* **373**, 889–895 (2021).
51. N. G. Davies *et al.*, *Science* **372**, eabg3055 (2021).
52. S. Kim *et al.*, *bioRxiv*, 2021.2007.2023.453598 [Preprint] (2021).
53. A. Saito *et al.*, *bioRxiv*, 2021.2006.2017.448820 [Preprint] (2021).
54. Y. Liu *et al.*, *bioRxiv*, 2021.2008.2012.456173 [Preprint] (2021).
55. L. Liu *et al.*, *Nature* **584**, 450–456 (2020).
56. C. O. Barnes *et al.*, *Cell* **182**, 828–842.e16 (2020).

ACKNOWLEDGMENTS

We thank the SBGrid team for technical assistance, K. Arnett for support and advice on the BLI experiments, and S. Harrison and A. Carfi for critical reading of the manuscript. EM data were collected at the Harvard Cryo-EM Center for Structural Biology of Harvard Medical School. We acknowledge support for COVID-19–related structural biology research at Harvard from the Nancy Lurie Marks Family Foundation and the Massachusetts Consortium on Pathogen Readiness (MassCPR). **Funding:** This work was supported by Fast grants by Emergent Ventures (to B.C. and D.R.W.), COVID-19 Award by MassCPR (to B.C. and D.R.W.), and NIH grants AI147884 and AI141002 (to B.C.), AI127193 (to B.C. and James Chou), and AI39538 and AI165072 (to D.R.W). **Author contributions:** B.C., J.Z., T.X., and Y.C. conceived the project. Y.C. expressed and purified the full-length S proteins with help from H.P. and carried out negative stain EM. T.X. performed BLI and cell-cell fusion experiments. J.Z. prepared cryo grids and performed EM data collection with contributions from M.L.M. and R.M.W., processed the cryo-EM data, and built and refined the atomic models. J.L. created all the expression constructs and performed the neutralization assays with the MLV-based pseudoviruses. C.L.L. and M.S.S. performed the neutralization assays with the HIV-based pseudoviruses. H.Z., K.A., and W.Y. performed the flow cytometry experiments. P.T., A.G., and D.R.W. produced the anti-S monoclonal antibodies. S.R.V. contributed to cell culture and protein production. All authors analyzed the data. B.C., J.Z., T.X., and Y.C. wrote the manuscript with input from all other authors. **Competing interests:** W.Y. serves on the scientific advisory boards of Hummingbird Bioscience and GO Therapeutics and is currently an employee of GV20 Therapeutics LLC. All other authors declare no competing interests. **Data and materials availability:** The atomic structure coordinates are deposited in the RCSB Protein Data Bank (PDB) under the accession numbers 7SBK, 7SBL, 7SBO, 7SBP, 7SBQ, 7SBR, 7SBS, and 7SBT; the electron microscopy maps have been deposited in the Electron Microscopy Data Bank (EMDB) under the accession numbers EMD-24981, EMD-24982, EMD-24983, EMD-24984, EMD-24985, EMD-24986, EMD-24987, and EMD-24988. All materials generated during the current study are available from the corresponding author under an MTA with Boston Children's Hospital. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl9463
Materials and Methods
Figs. S1 to S18
Tables S1 to S3
References (*57*–*69*)
MDAR Reproducibility Checklist

15 August 2021; accepted 22 October 2021
Published online 26 October 2021
10.1126/science.abl9463

CATALYSIS

Sabatier principle of metal-support interaction for design of ultrastable metal nanocatalysts

Sulei Hu and Wei-Xue Li*

The stability of supported nanocatalysts is crucial to meeting environmental and energy challenges and necessitates fundamental theory to relieve trial-and-error experimentation and accelerate lab-to-fab translation. Here, we report a Sabatier principle of metal-support interaction for stabilizing metal nanocatalysts against sintering based on the kinetic simulations of 323 metal-support pairs using scaling relations from 1252 energetics data. Too strong of an interaction is shown to trigger Ostwald ripening, whereas too weak of an interaction stimulates particle migration and coalescence. High-throughput screening of supports enables the sintering resistance of nanocatalysts to reach the Tammann temperature on homogeneous supports and far beyond it on heteroenergetic supports. This theory, which is substantiated by first-principles neural network molecular dynamics simulations and experiments, paves the way for the design of ultrastable nanocatalysts.

The stability of nanomaterials is a matter of life(time) and death in various nanotechnologies, especially for heterogeneous metal nanocatalysts, to meet energy and environmental requirements (*1*). The instability, among other issues, caused by the thermal- and/or chemical-induced sintering (*2*) of smaller metal nanoparticles (NPs) into larger NPs with shrinking atomic utilization severely decreases productivity, delays the lab-to-fab translation of highly active nanocatalysts (*3*), and requires plant shutdown for catalyst replacement or regeneration with large capital costs (*4*). The rational design of nanocatalysts with sufficient thermal resistance and operando lifetime, in addition to high activity and selectivity, has important economic and scientific value, even for the general nanoscience field (*5*).

Sintering of metal NPs proceeds through the formation and diffusion of metal atoms and/or metal-reactant complexes [Ostwald ripening (OR)] (*6*, *7*) and/or through particle migration and coalescence (PMC) on the support (Fig. 1A) (*8*). To mitigate sintering, physical approaches such as spatial confinement and geometric shielding (*9*) have been used, which involve the incorporation of NPs into one-dimensional (1D) tubular or 3D micro- or mesoporous materials (*10*) and the full or partial encapsulation of NPs with less-reactive thin films with microchannels (*11*). The cost of these approaches is the blockage of active sites and the impeded transport of reactants. Chemically, alteration of metal-support interaction (MSI) is used to stabilize NPs and control the sintering kinetics (*12*), which is, however, a complex function of the support composition and surface orientation and defects (*13*), interface structure, lattice mismatch, mixing and alloy phase formation, charge rearrangement (*14*), and reaction conditions (*15*). This fact results in a large gap between the sintering kinetics and the underlying MSI and challenges the screening and/or optimization of supports, thereby necessitating expensive trial-and-error experimentation (*3*). To bridge this gap, it is vital to identify the corresponding MSI descriptor and determine the governing rule for the sintering kinetics and develop a design theory to engineer the metal-support interface for ultrastable nanocatalysts.

Here, we report a Sabatier principle of MSI for the stability of supported metal NPs against sintering based on the linear scaling relationship between the particle adhesion energy and atom binding energy with supports, two MSI descriptors for the corresponding OR and PMC processes. Sintering kinetic simulations

Hefei National Laboratory for Physical Sciences at the Microscale, School of Chemistry and Materials Science, Key Laboratory of Surface and Interface Chemistry and Energy Catalysis of Anhui Higher Education Institutes, CAS Center for Excellence in Nanoscience, iChEM, University of Science and Technology of China, Hefei, China.
***Corresponding author. Email: wxli70@ustc.edu.cn**

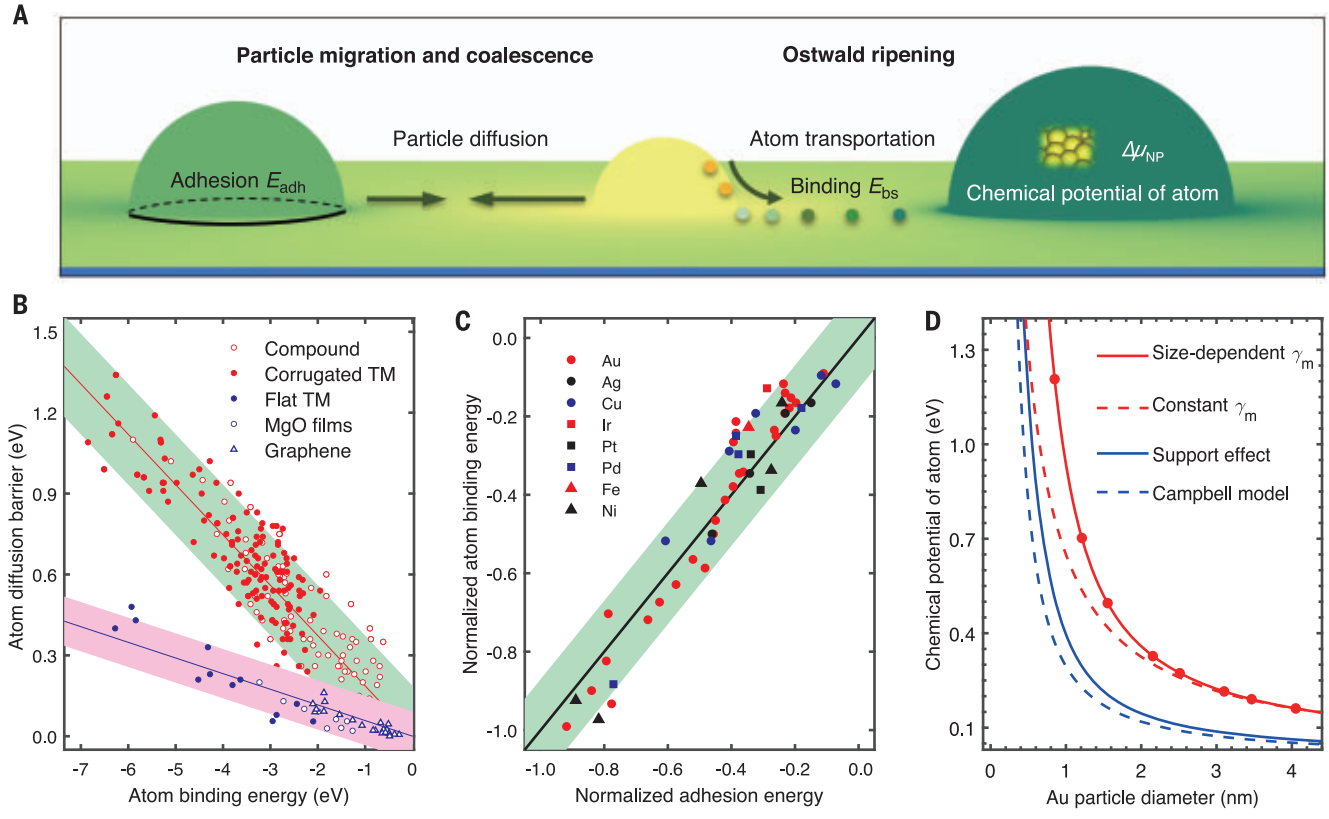


Fig. 1. Sintering mechanism and scaling relations for the energetics. (**A**) Schematic of PMC and OR of supported metal NPs. (**B**) The diffusion barrier E_{ds} and binding energy E_{bs} of metal atoms on corrugated surfaces, including oxide compounds (red open circles) and TMs (red solid circles) in the green region, and on flat surfaces including TM(111) (blue solid circles), MgO films (blue open circles), and graphene on metal (blue open triangles) in the purple region (SM and data S1). The colored regions indicate the distribution range of the data. (**C**) Metal particle adhesion energy E_{adh} (normalized by two times the surface energy γ_m of the bulk metal) versus metal atom binding energy E_{bs} (normalized by the metal bulk cohesive energy E_c) (data S2). The colored region indicates the distribution range of the data. (**D**) Chemical potential $\Delta\mu_{NP}$ of the atoms in unsupported spherical NPs under constant (red dashed line) and size-dependent (red solid line) γ_m values and $\Delta\mu_{NP}$ value for CeO_{2-x}-supported Au NPs with a contact angle of 43° (blue solid line) and that based on Campbell's model (*23*) (blue dashed line) as a function of the diameter of the unsupported spherical NPs.

of 10 transition metals (TMs) and 91 supports (323 metal-support pairs in total) using 1252 energetics data were conducted, and a universal volcanic dependence of the sintering kinetics on MSI was revealed. It was found that the optimal MSI corresponding to the highest stability on homogeneous supports should be neither too strong nor too weak. Too strong of a MSI triggered rapid OR, whereas too weak of a MSI stimulated facile PMC, both of which severely worsened the stability. For supports with the optimal MSI to NPs, the sintering onset temperature is about half of the melting temperature T_m of the bulk metal for typical NPs (~3 nm), substantiated by the long-reported empirical Tammann temperature. The revealed Sabatier principle enables the high-throughput screening of heteroenergetic supports to break the scaling relationship and boost the sintering resistance of the supported NPs far beyond the Tammann temperature. This theory, which is substantiated by molecular dynamics (MD) simulations based on the first-principles neural network potential (*16*) and available experiments (*17*), paves the way for the design of ultrastable nanocatalysts.

Scaling relationships

The size growth rate of NPs through OR is determined by the total activation energy E_{act} (OR) = $E_{bs} - E_c + E_{ds}$ describing the formation ($E_{bs} - E_c$) of the ripening monomer (herein metal atom) with respect to the cohesive energy E_c (negative) of bulk metal and its diffusion on the support (*18*). Both the binding energy E_{bs} (negative) and diffusion barrier E_{ds} (positive) of the metal atom on the support depend on the MSI. To identify the appropriate MSI descriptor, we retrieved 32 TM atoms and 61 supports (292 metal-support pairs in total) from the literature [see supplementary materials (SM) and data S1 to S6], including various pristine oxide compounds, TMs, and thin films, whose E_{bs} and E_{ds} were calculated using density functional theory (DFT). For most of the considered surfaces, especially corrugated oxides and TM surfaces, E_{ds} falls statistically in a region exhibiting a linear scaling relation with E_{bs}, that is, $E_{ds} = k_{ds}E_{bs}$ [slope (k_{ds}) = –0.18, coefficient of determination (R^2) = 0.86] (green region in Fig. 1B). Note that the flat surfaces that fall statistically in the purple region have a smaller slope. The results could be rationalized by the fact that the diffusion barrier E_{ds} of the metal atom is proportional to its binding strength E_{bs}, with a slope determined by, among other features, the surface corrugation (see SM for more details). This relationship simplifies the activation energy as

$$E_{act}(\mathrm{OR}) = kE_{bs} - E_c \quad (1)$$

where k = 0.82 and E_{bs} is used as the MSI descriptor for the OR kinetics below.

The size growth rate of NPs through PMC is determined by the particle diffusion barrier $E_{act}(\mathrm{PMC}) = E_{act}^{m} - S^{m}E_{adh}$ consisting of

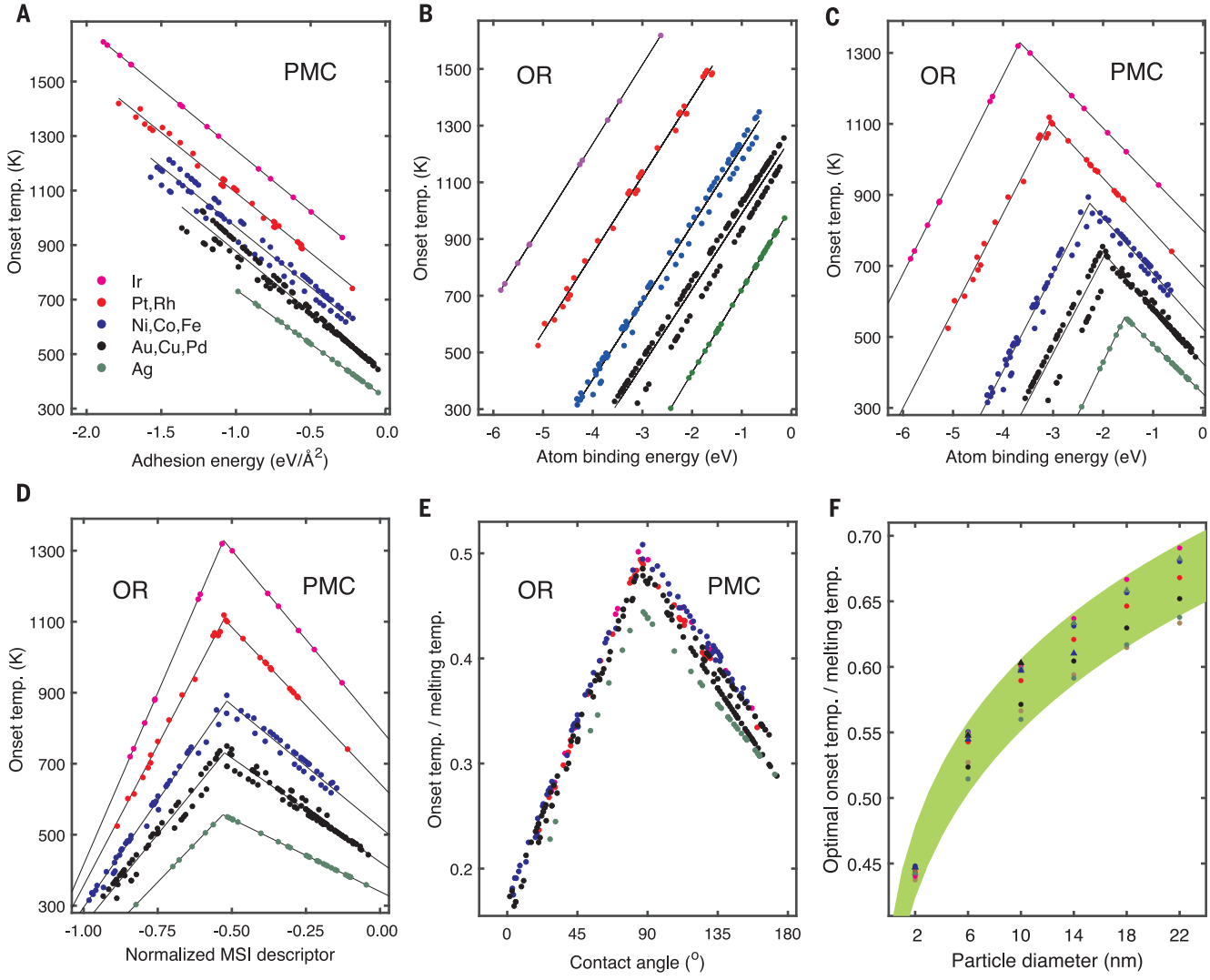


Fig. 2. Volcanic curves of the stability of NPs as a function of MSI. (**A**) Calculated onset temperature T_{on} through PMC versus the metal particle adhesion energy $S^m E_{adh}$ during temperature-programmed kinetics at a heating rate of 1 K/s. On each support, the NPs had the same volume as that of the unsupported spherical NPs with an average diameter of 3.0 ± 0.30 nm. (**B**) T_{on} through OR versus the metal atom binding energy E_{bs}. (**C**) Sintering T_{on}, corresponding to the lower T_{on} between PMC and OR, versus E_{bs}. (**D**) Sintering T_{on} versus the normalized MSI descriptors of $E_{adh}/2\gamma_m$ or $-E_{bs}/E_c$. (**E**) Sintering T_{on} normalized by the melting temperature T_m of the bulk metal versus the contact angle. (**F**) Peak T_{on} normalized by T_m versus the initial average diameter. The shaded region indicates the distribution range of the data. All the data are listed in the SM. Sensitivity analysis of the scaling relationships used and the impact of their uncertainties on the volcanic curves and trend behavior can be found in the SM.

the self-activation energy E_{act}^m of the metal atom on the metal particle surface and the cost of the particle adhesion energy E_{adh} (negative) to the support (where S^m is the area per atom) (*19*, *20*). E_{act}^m was found to be proportional to the corresponding E_c (fig. S1), $E_{act}^m = -0.33E_c$ (R^2 = 0.95), based on the DFT values from more than 200 TM pristine surfaces retrieved from the literature (SM and data S1). Thus, the particle diffusion barrier becomes

$$E_{act}(\text{PMC}) = -S^m E_{adh} - 0.33 E_c \quad (2)$$

where E_{adh} as the MSI descriptor for PMC can be calculated by DFT or experimentally extracted. We retrieved the measured contact angle α of 188 metal-support pairs, including 10 late TMs from the literature (data S2), and calculated the corresponding E_{adh} using Young-Dupre's equation and the experimental surface energy of bulk metal. Depending on the metal-support pairs, E_{adh} varied at a wide range of values from −23 to −327 meV/Å^2, though the calorimetric experiments showed that E_{adh} is sensitive to the cleanliness of the surface (*21*).

For a given metal-support pair, the corresponding E_{bs} and E_{adh} are highly correlated. To figure out the correlation, we constructed a series of Ag-supported MgO(100) films (*22*) with thicknesses from 1 to 8 monolayers (fig. S2) and calculated their E_{bs} and E_{adh} toward the Au atom and film using DFT (data S3), respectively. Both values decreased with the film thickness and exhibited a linear scaling relationship, $S^m E_{adh} \cong E_{bs}/3$ (R^2 = 0.94) (fig. S3A). Moreover, we considered another 42 metal-support pairs retrieved from the literature, and the corresponding E_{adh} from experiments and E_{bs} from DFT were both available (data S3), from which nearly the same relation as above was found (fig. S3B). To rationalize the results, we note that a given support adhering strongly to a metal particle would be strongly bound to the metal atom, and the energy gain

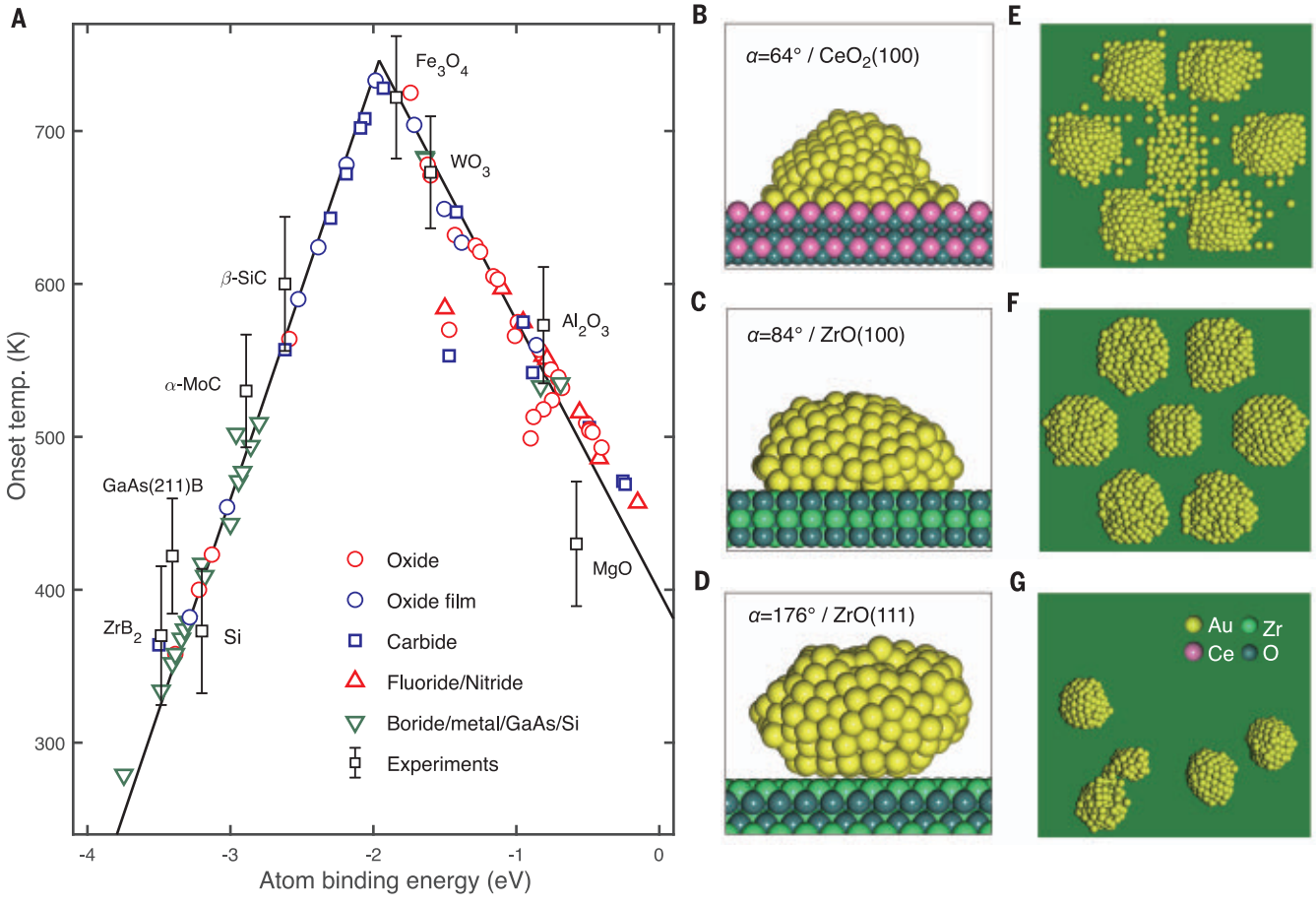


Fig. 3. Volcanic curves of supported Au NPs and MD simulations. (**A**) Calculated onset temperature of Au NPs (3 nm) versus Au atom binding energy onto various supports. The black squares indicate the experimental values with error bars of ±46 K based on the uncertainty in the size distribution and loading (data S4). The error bars indicate the uncertainty of the experimental temperature. (**B** to **G**) MD snapshots using the first-principle neural network potential at 800 K at 10^5 time steps of 1 fs for one Au_{299} cluster [(B) to (D)] and seven Au clusters (one Au_{69} and six Au_{157}) [(E) to (G)] on Ce-terminated CeO_2(100) [(B) and (E)], Zr-terminated ZrO(100) [(C) and (F)], and Zr-terminated ZrO(111) [(D) and (G)].

for the adhesion (per atom) of the particle with a larger coordination is smaller than that of the metal atom to the support. Note that for the bulk metals themselves, there are two relevant quantities with E_{bs} and E_{adh}, that is, the cohesive energy E_c and surface energy γ_m. These two quantities were found to follow a similar scaling relationship, $2S^m\gamma_m \cong -1/3E_c$ (fig. S4). Normalizing the previous relationship by the latter one, a dimensionless scaling relationship

$$E_{adh}/2\gamma_m = -E_{bs}/E_c \quad (3)$$

was introduced (R^2 = 0.89; see SM for more details), and a diagonal line crossing the origin is depicted in Fig. 1C. For the metal-support pair of interest, when either E_{adh} or E_{bs} was available, Eq. 3 was applied to estimate the unavailable one (fig. S5). With these paired E_{adh} and E_{bs} values, E_{act}(OR) and E_{act}(PMC) were calculated for the kinetic simulation below (figs. S6 and S7).

The chemical potential $\Delta\mu_{NP}$ of metal atoms in NPs with respect to the bulk metal is crucial for the rate of both OR and PMC. According to the Gibbs-Thomson equation, $\Delta\mu_{NP}$ is proportional to the γ_m of the metal NPs, whose value differs from the bulk value at small size owing to the increased ratio of the lower-coordinated atoms exposed. The change in the surface coordination was estimated by counting various constructed Wulff-shaped NPs of different sizes (fig. S8), and, based on this, a size-dependent γ_m was established (fig. S9). For a NP with a given volume on different supports, different E_{adh} leads to different α and radius R of curvature, and the corresponding $\Delta\mu_{NP}$ becomes support-dependent throughout this work (see SM for more details). Using CeO_{2-x}-supported Au NPs as an example, the resulting $\Delta\mu_{NP}$ values (blue solid line) in Fig. 1D were notably lower than those of the unsupported NPs (red solid line), consistent with Campbell's model (blue dashed line) (*23*).

Sintering volcano on homogeneous supports

We now turn to the temperature-programmed OR and PMC kinetic simulations of the above 323 metal-support pairs with the same typical initial average diameter of 3.0 ± 0.3 nm (spherical NPs before contact with the supports of interest). During heating, the onset temperature T_{on} corresponding to a 10% decrease in particle number is defined to estimate the thermal resistance against PMC and OR (methodology details in the SM). The simulated T_{on} for PMC was found to linearly increase with the absolute value of S^mE_{adh} (Fig. 2A and fig. S10). Depending on the TMs, the enhanced MSI strength improves the thermal resistance to PMC by 400 to 700 K owing to the increase of E_{act}(PMC) (fig. S7). Consistently, the diffusion coefficient of the supported NPs decreases by five to six orders of

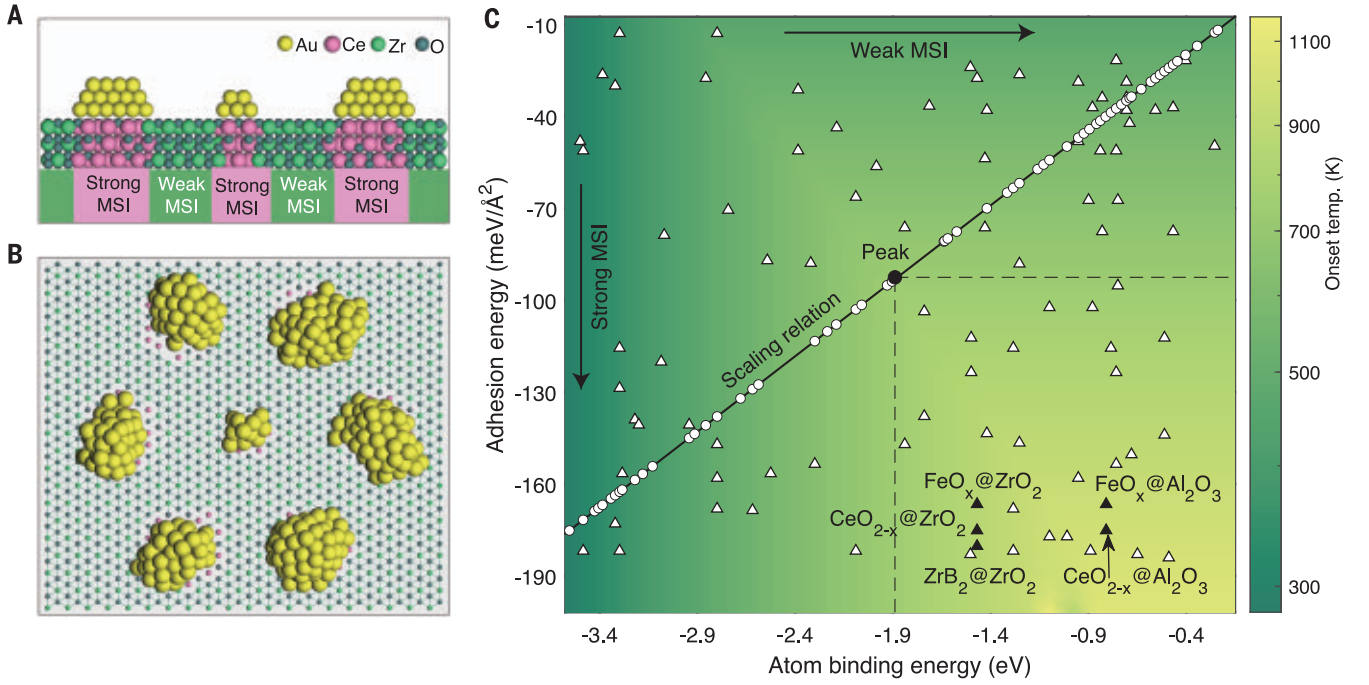


Fig. 4. Screening of heteroenergetic supports beyond the volcanic curves. (**A**) Heteroenergetic support S@W [CeO_{2-x}/ZrO_2(111)] containing a small area of domains S (CeO_{2-x}) with large absolute E_{adh} to Au clusters surrounded by a large area of domains W (ZrO_2) with small absolute E_{bs} to Au atoms. The colored regions in Fig. 4A are just used to distinguish the space distribution of the domain S and domain W. (**B**) Snapshot of the MD simulation after 80 ps (800 K) for one Au_{19} and six Au_{55} clusters on the CeO_{2-x} domains of CeO_{2-x}@ZrO_2(111). (**C**) Onset temperature of sintering of the Au NPs (3 nm) supported on the S domains of the S@W support versus the E_{adh} value (*y* axis) of the Au NPs onto S domains and the E_{bs} value (*x* axis) of the Au atoms onto W domains.

magnitude (fig. S11). By contrast, the simulated T_{on} for OR linearly decreased with the absolute value of E_{bs}, as shown in Fig. 2B and fig. S12. This trend was observed because E_{act} (OR) decreased with the MSI strength (fig. S6), and the thermal resistance to OR weakened accordingly.

Because both the OR and PMC contribute to the sintering, the effective T_{on} for the overall sintering is determined by the faster one with the lower T_{on}. Considering the counter-dependence of the OR and PMC rate on MSI and the constraint of the scaling relationship between E_{adh} and E_{bs} (Eq. 3), a volcanic dependence of the effective T_{on} was immediately seen and plotted in Fig. 2C and fig. S13 with respect to E_{bs} or S^mE_{adh}, respectively. At the left side of the volcanic peak where the corresponding MSI was strong, OR had a lower T_{on} and faster rate than PMC, irrespective of the TMs considered. On the right side with weak MSI, PMC had a lower T_{on} and dominated the sintering process. At the volcanic peak with the optimal MSI (neither too strong nor too weak), the rates of OR and PMC became balanced, but neither was facile. The optimum MSI and corresponding peak temperature showed a great dependence on the metal composition: The stronger the optimal MSI was (−3.73 eV for Ir versus −1.67 eV for Ag) , the higher the peak T_{on} was (1330 versus 540 K). To unravel the underlying physics, we used the normalized descriptors of $E_{adh}/2\gamma_m$ or $-E_{bs}/E_c$ to replot Fig. 2C. The optimal MSI after normalization became insensitive to the composition and approximated to a constant value of −0.53 (Fig. 2D). The constant optimal MSI could be derived approximately by setting E_{act}(OR) = E_{act}(PMC) (fig. S14), highlighting the key role of the elementary activation in these two processes to the overall sintering.

The T_{on} value for each TM (Fig. 2D) was further normalized by the melting temperature T_m of the bulk metal counterparts and then replotted versus the corresponding α value using Young-Dupre's equation in Fig. 2E. All the volcanic curves were found to be nearly overlapping, with a peak T_{on} at $0.47T_m \pm 0.02T_m$. This overlap was not unexpected because both T_m and the peak T_{on} depended similarly on the E_c of the bulk metals (fig. S15). To see the size effect on the volcanic curves, we increased the initial average diameter of the unsupported NPs from 2 to 22 nm. The optimal MSI was not found to be affected by particle size (fig. S16). However, the peak T_{on} for these 10 TMs increased from $0.44T_m \pm 0.01T_m$ to $0.67T_m \pm 0.03T_m$ (Fig. 2F). These peak T_{on} values were substantiated by the Tammann temperature ($\sim 0.5T_m$) of the bulk metal counterparts as the long-reported feature that empirically characterized the sintering temperature of materials. Under reaction conditions, the reactant might stabilize the metal atoms to form metal-reactant complexes on the support (*6*), leading to a metal-reactant interaction–dependent volcanic curve, as discussed in the SM.

Sintering volcano for Au NPs on homogeneous supports

As proof of principle, the sintering of Au NPs that is sensitive to MSI (*24*) was investigated. The calculated T_{on} values over 82 pristine surfaces were plotted versus the corresponding E_{bs} values in Fig. 3A. Ionic compounds, including oxides, fluorides, and nitrides, attained a relatively weak MSI with Au and were mainly distributed on the right side of the volcanic curve. Carbides with noble metal features that bind modestly to Au were distributed on both sides. However, supports that were dominated by metallic or covalent bonds, such as metal, boride, semiconductor, or oxide films, exhibited a strong MSI with Au and were distributed mainly on the left side. Figure 3A also presents the available experimental data for comparison. Specifically, for MgO, Al_2O_3 and WO_3 located on the right side (fig. S17, A and B), sintering was experimentally observed to proceed through PMC, as predicted by the above kinetic theory. By contrast, for GaAs(211)B and ZrB_2 on the left side, sintering was observed

to proceed through OR, also in agreement with the theory (fig. S17, C and D). Fe_3O_4(111) had the optimal MSI to Au and was predicted with a peak temperature of 743 K. Indeed, when annealing at a temperature of 773 K, scanning tunneling microscopy (STM) revealed simultaneous Au particle agglomeration and atom migration, suggesting that PMC and OR occurred at a comparable rate and that neither of them was facile (*25*). Consistent trends between theory and experiment were also found for industry-related Pt and Rh nanocatalysts (fig. S18). Moreover, data from STM experiments also showed a preferential OR process for Pd_{19} clusters on Rh(111) with a strong metal-metal bond at the interface (*26*). The agreement with experiments substantiated the trend behavior and underlying mechanism revealed by the kinetic simulations.

The Sabatier principle was further supported by MD simulations using the first-principles neural network potential (*16*). Twelve oxide surfaces with different compositions and surface structures were constructed to cover a wide range of MSI to Au. (figs. S19 to S23) One small Au_{69} and six large Au_{157} clusters were supported on the constructed surfaces (10 nm by 10 nm), and MD simulations were performed at 800 K for 100 ps. For Ce-terminated CeO_2(100) interacting strongly with Au at the left side (α = 64°; Fig. 3B), MD simulations revealed detachment of Au atoms from the small Au_{69} cluster and diffusion toward large Au_{157} clusters and no cluster diffusion was observed during the simulation time (movie S1), indicating a typical OR feature. By contrast, for Zr-terminated ZrO(111) interacting weakly with Au (α = 176°; Fig. 3D), the small Au_{69} cluster diffused toward the large Au_{157} cluster (Fig. 3G) and there was no atom detachment found, a typical feature involved in PMC (movie S3). For ZrO(100) near the optimal MSI (α = 84°; Fig. 3C), neither atom detachment nor cluster migration was observed within the simulation time (Fig. 3F and movie S2), consistent with the volcanic peak exhibiting a higher stability against sintering.

Beyond Tammann temperature on bifunctional heteroenergetic supports

A further increase in thermal stability beyond the volcanic peak, which is crucial for subnanometer metal clusters and high-temperature catalytic reactions, necessitates breaking of the scaling relationship between E_{adh} and E_{bs} (Eq. 3). This feature, among others, can be achieved by constructing a bifunctional heteroenergetic support S@W, where support S at the nanoscale with strong adhesion (large absolute E_{adh}) pins metal NPs to suppress PMC and support W with weak binding (small absolute E_{bs}) prevents the formation of metal atoms to inhibit OR. As an example, we constructed a heteroenergetic support of CeO_{2-x}@ZrO_2(111) (Fig. 4A), where CeO_{2-x}(111) domains containing 15% oxygen vacancies (*24*) adhering strongly with Au (fig. S24, A to D) were embedded into the matrix of ZrO_2(111) binding weakly with Au (fig. S24, E to H). Small Au_{19} and six Au_{55} clusters were deposited on separated CeO_{2-x}(111) domains with a size comparable to that of the Au clusters. On homogeneous ZrO(100) with optimal MSI as confirmed above (Fig. 3, C and F), MD simulation at 800 K indicated that these small clusters became unstable even at the beginning (fig. S25). However, on CeO_{2-x}@ZrO_2(111), neither cluster migration nor atom detachment from Au_{19} were observed (Fig. 4B and movie S4). Note that a dual-oxide support was proposed to increase the diffusion barrier of metal NPs or atoms across the support surface (*2*, *27*).

The revealed Sabatier principle and scaling relationships can enable the high-throughput screening of the heteroenergetic support S@W. For the Au NPs of ~3 nm in Fig. 3A, any two of the 82 supports can be regarded as S and W domains in turn, generating 6724 combinations. For each combination, the lower T_{on} between that attributed to PMC on the S domains and that attributed to OR on the W domains determines the effective T_{on} of sintering. These T_{on} values yield a 2D sintering map with respect to E_{adh} for S (y axis) and E_{bs} for W (x axis) (Fig. 4C). The diagonal line from the bottom left to the top right corresponds to the volcanic curve in Fig. 3A with a peak T_{on} of 743 K. Notably, at the bottom right quarter of the 2D map that includes 1681 heteroenergetic supports with large $|E_{adh}|$ for S and small $|E_{bs}|$ for W, all the effective T_{on} exceeded 743 K and had a maximum even up to 1140 K (~0.85T_m). Among others, CeO_{2-x}@ZrO_2 and CeO_{2-x}@Al_2O_3 were predicted to appear in this area, and the effective T_{on} were 1050 and 1100 K, respectively. Because the support W that bound weakly with metal atoms is also reluctant to form the metal-reactant complexes, influence of reaction conditions on chemical stability of supported metal NPs through the reactant-promoted OR on the corresponding heteroenergetic S@W could be relieved (see SM for more details). Indeed, CeO_{2-x}@ZrO_2 as a crucial additive has been applied in commercial automotive exhaust catalysts to improve the overall stability under practical application (*28*). Highly durable three-way catalysts as small as a dozen atoms in size were realized on CeO_{2-x}@Al_2O_3 under severe hydrothermal aging conditions at 1173 K, and the corresponding structure can maintain its stability for 24 hours (*17*). Considering that the new active sites might form at the boundaries between nanocatalyst and support, the heteroenergetic supports provide a platform to simultaneously improve the stability, activity, and selectivity of supported nanocatalysts.

REFERENCES AND NOTES

1. H. Chen *et al*., *ACS Cent. Sci.* **6**, 1617–1627 (2020).
2. Y. Dai, P. Lu, Z. Cao, C. T. Campbell, Y. Xia, *Chem. Soc. Rev.* **47**, 4314–4331 (2018).
3. A. K. Datye, M. Votsmeier, *Nat. Mater.* **20**, 1049–1059 (2021).
4. M. Rahmati, M. S. Safdari, T. H. Fletcher, M. D. Argyle, C. H. Bartholomew, *Chem. Rev.* **120**, 4455–4533 (2020).
5. S. L. Scott, *ACS Catal.* **8**, 8597–8599 (2018).
6. R. Ouyang, J. X. Liu, W. X. Li, *J. Am. Chem. Soc.* **135**, 1760–1771 (2013).
7. Q. Wan *et al*., *Nanoscale* **10**, 17893–17901 (2018).
8. E. M. Dietze, F. Abild-Pedersen, P. N. Plessow, *J. Phys. Chem. C* **122**, 26563–26569 (2018).
9. X. Yang *et al*., *Nat. Commun.* **10**, 1611 (2019).
10. J. Zhang *et al*., *Nat. Catal.* **1**, 540–546 (2018).
11. H. Tang *et al*., *Sci. Adv.* **3**, e1700231 (2017).
12. T. W. van Deelen, C. Hernández Mejía, K. P. de Jong, *Nat. Catal.* **2**, 955–970 (2019).
13. Y. Y. Zhou *et al*., *Nat. Catal.* **3**, 454–462 (2020).
14. Q. Fu *et al*., *Science* **328**, 1141–1144 (2010).
15. J. Hulva *et al*., *Science* **371**, 375–379 (2021).
16. S. D. Huang, C. Shang, P. L. Kang, X. J. Zhang, Z. P. Liu, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **9**, e1415 (2019).
17. H. Jeong *et al*., *Nat. Catal.* **3**, 368–375 (2020).
18. S. L. Hu, W. X. Li, *ChemNanoMat* **4**, 510–517 (2018).
19. S. L. Hu, W. X. Li, *Sci. China Technol. Sci.* **62**, 762–772 (2019).
20. P. Yin *et al*., *Nat. Commun.* **12**, 4865 (2021).
21. S. L. Hemmingson, C. T. Campbell, *ACS Nano* **11**, 1196–1203 (2017).
22. M. Sterrer, T. Risse, M. Heyde, H. P. Rust, H. J. Freund, *Phys. Rev. Lett.* **98**, 206103 (2007).
23. C. T. Campbell, Z. T. Mao, *ACS Catal.* **7**, 8460–8466 (2017).
24. J. A. Farmer, C. T. Campbell, *Science* **329**, 933–936 (2010).
25. G. S. Parkinson, *Surf. Sci. Rep.* **71**, 272–365 (2016).
26. Y. Fukamori *et al*., *ChemCatChem* **5**, 3330–3341 (2013).
27. P. Lu, C. T. Campbell, Y. Xia, *Nano Lett.* **13**, 4957–4962 (2013).
28. C. K. Lambert, *Nat. Catal.* **2**, 554–557 (2019).
29. S. L. Hu, W.-X. Li, Sabatier principle of metal-support interaction for design of ultrastable metal nanocatalysts, Zenodo (2021); https://doi.org/10.5281/zenodo.5524646.

ACKNOWLEDGMENTS

We thank C. T. Campbell for comments. **Funding:** This work was supported by the National Key R&D Program of China (2018YFA0208603), the National Natural Science Foundation of China (91945302 and 21903077), the Chinese Academy of Sciences (QYZDJ-SSW-SLH054), K. C. Wong Education (GJTD-2020-15), and the China Postdoctoral Science Foundation. We thank the Supercomputing Center of the University of Science and Technology of China and the National Supercomputing Center in Zhengzhou. **Author contributions:** Conceptualization: W.-X.L.; Supervision: W.-X.L.; Methodology: S.H.; Investigation: S.H., W.-X.L.; Visualization: S.H.; Writing: W.-X.L., S.H. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions are present in the paper and the supplementary materials and are deposited at Zenodo (*29*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi9828
Materials and Methods
Supplementary Text
Figs. S1 to S33
Tables S1 to S4
References (*30*–*53*)
Movies S1 to S4
Data S1 to S6

13 April 2021; resubmitted 23 August 2021
Accepted 6 October 2021
Published online 4 November 2021
10.1126/science.abi9828

HYDROPHOBICITY

Charge transfer across C–H···O hydrogen bonds stabilizes oil droplets in water

Saranya Pullanchery[1], Sergey Kulik[1], Benjamin Rehl[1], Ali Hassanali[2]*, Sylvie Roke[1,3,4]*

The hydrophobic–water interface plays a key role in biological interactions. However, both the hydrophobic–water interfacial molecular structure and the origin of the negative zeta potential of hydrophobic droplets in water are heavily contested. We report polarimetric vibrational sum-frequency scattering of the O–D and C-H stretch modes of 200-nanometer hexadecane oil droplets dispersed in water. An unusually broad spectral distribution (2550 to 2750 per centimeter) of interfacial water molecules that were not hydrogen bonded to other water molecules was observed, as well as a blue shift in the vibrational frequency of the interfacial hexadecane C-H stretch modes. Oil and water frequency shifts correlated with the negative electrostatic charge. Molecular dynamics simulations demonstrated that the unexpected strong charge-transfer interactions arose from interfacial C–H···O hydrogen bonds.

Hydrophobicity represents a fundamental physical property that determines a variety of processes in aqueous media, such as protein folding, self-assembly, and aggregation (*1*, *2*). Submicrometer-sized hydrophobic oil droplets or particles in water are an important model system for understanding how hydrophobicity works. However, both the charge and the structure of these interfaces are highly debated. Since the late 19th and early 20th century (*3*–*6*), hydrophobic nanodroplets and air/gas bubbles have been prepared and investigated. Early experiments (*4*, *5*) reported the surprising observation that oil droplets or air bubbles in water exhibit a negative zeta (ζ) potential. The ζ potential is interpreted as the electrostatic potential at the slip plane of the droplet (*7*), which is the plane that separates moving molecules from stationary ones. There are two main explanations for the negative charge on oil droplets. The earliest explanation invokes the adsorption of hydroxide (OH^-) ions, because these are the only ionic sources of negative charge in neat water. However, the adsorption of OH^- has not been spectroscopically verified (*5*, *8*, *9*), and most theoretical studies have not found any thermodynamic stabilization for OH^- at hydrophobic interfaces (*10*, *11*). OH^- is a small, nonpolarizable ion that prefers to be hydrated instead of being adsorbed at an interface. More recently, the negative charge was explained by a charge-transfer mechanism involving electron density that is present at the interface because of an asymmetric hydrogen (H)–bond distribution in the aqueous interfacial region (*6*). This picture has recently been refined by molecular dynamics (MD) simulations (*12*) predicting explicit charge transfer between water and oil molecules. Charge transfer implies the transfer of charge between two molecules, and electrostatic polarization is a measure of the displacement of bound charge. Although these two concepts are related, polarization is often assumed to be confined within molecules.

Although the model of charge transfer suggests that water is key to the existence of the charge on oil droplets, the structure of this water next to oil is equally controversial. Water in contact with extended planar interfaces of nonpolar liquids or gases has been investigated by vibrational sum-frequency spectroscopy (*13*, *14*), an inherently surface-specific spectroscopic technique. However, the spectra recorded from the hydrophobic–water interface vary substantially throughout the literature (see the supplementary materials, section S1 and fig. S1, for a summary of data and experiments). As a consequence, the structure of water next to a hydrophobic liquid material remains unknown. In the case of water in contact with oil droplets, only the structure of the oil has been investigated using vibrational sum-frequency scattering (*6*, *15*). The vibrational spectrum of water in contact with oil droplets has been recorded only partially (*9*, *16*), with inconclusive results concerning the H-bond network of water.

Here, the structure of water in contact with hydrophobic droplets dispersed in water was unraveled and the source of the negative charge was investigated. The interfacial H-bond network of hexadecane droplets in water was measured using polarimetric vibrational sum-frequency scattering (SFS). Spectral peak ratios showed that the H-bonding network was stronger at the oil–water interface than at the air–water interface. Furthermore, polarization analysis revealed a broad spectral region from ~2550 to ~2750 cm^{-1}, which originated from interfacial water molecules that interacted with the oil phase. These spectral features and interactions between water and oil suggest charge transfer from water O atoms to the C–H groups in the oil. This charge transfer was spectroscopically confirmed by a blue shift in the C–H modes of the interfacial oil molecules. MD simulations confirmed this observation and revealed the presence of improper C–H···O H bonds between water and oil. We showed two distinct ways to inhibit the charge transfer from water to oil molecules, leading to a near-zero or positive ζ potential. In both cases, the C–H frequency shifts of the oil phase were inverted. Thus, the charge transfer from interfacial water to oil molecules is the origin of the negative charge of bare oil droplets in water and is responsible for their stabilization.

Results and Discussion

Investigating the water structure at the interface of pure oil droplets

Nanodroplets of n-hexadecane (C16) with an average diameter of 200 nm were prepared in neat heavy water (D_2O) using ultrasonication (see the supplementary materials, section S2 and fig. S2). The vibrational SFS (Fig. 1A) spectra of droplets were measured in the O–D and C–H stretch regions using both SSP [i.e., S-polarized sum frequency (SF), S-polarized visible, and P-polarized infrared (IR)] and PPP polarization combinations. Because the IR beam passed through the liquid phase, it was absorbed by water when it had a frequency content that matched the O–D stretch modes. Therefore, when the IR pulse traveled through the sample cell (Fig. 1B), each droplet was probed by an IR pulse that had a different spectral shape and intensity, resulting in strongly modified SFS spectra (*17*). For this reason, it was long deemed impossible to measure the surface vibrational spectrum of water for particles dispersed in water. However, recently, we determined the appropriate light–matter interactions that allowed us to devise a method to retrieve the true surface response ($|\Gamma^{(2)}|^2$) from the measured SF spectral intensity (I_{SF}) (*17*). The procedure is described in the supplementary materials, section S2. $\Gamma^{(2)}$ is the effective second-order particle susceptibility that describes the spectral interfacial response of droplets dispersed in solution and is proportional to the square root of the measured intensity (*18*–*20*) (see the supplementary materials, section S2).

The resulting $|\Gamma^{(2)}|^2$ spectrum of 2 vol% hexadecane droplets in D_2O (Fig. 1C, blue trace) showed the vibrational response of the O–D stretch modes of water and thus revealed

[1]Laboratory for Fundamental BioPhotonics, Institute of Bioengineering (IBI), School of Engineering (STI), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. [2]International Centre for Theoretical Physics, 34100 Trieste, Italy. [3]Institute of Materials Science and Engineering (IMX), School of Engineering (STI), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland. [4]Lausanne Centre for Ultrafast Science, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland.
*Corresponding author. Email: sylvie.roke@epfl.ch (S.R.); ahassana@ictp.it (A.H.)

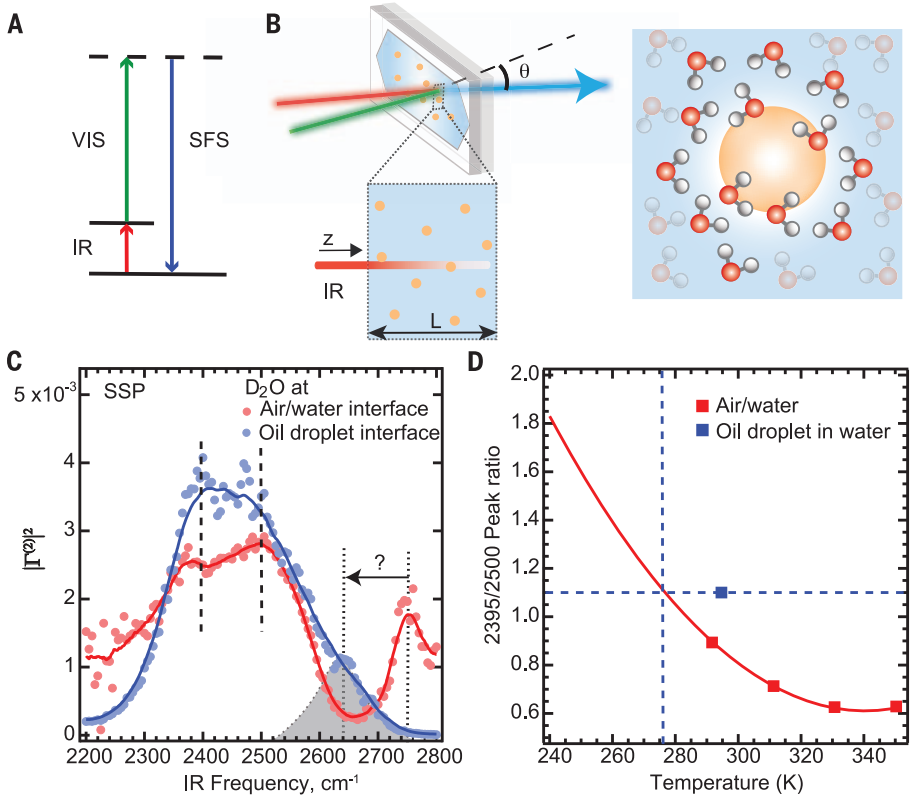


Fig. 1. Water structure at bare oil droplet surface. (**A**) Energy-level diagram of the SFS experiment involving simultaneous excitation with an infrared (IR) and visible (VIS) beams. (**B**) Sketch of the vibrational SFS experiment on oil droplets in water. θ is the scattering angle, L is the optical path length, and the magnified sample cross section shows the attenuation of IR intensity as the beam travels through the sample. (**C**) SFS spectra of hexadecane droplets in D_2O (blue) and reflection sum-frequency generation spectra of the planar D_2O–air interface [red; (*14*, *21*)]. The solid lines are guides to the eye. The dashed lines indicate the frequencies of low- and high-frequency H-bonded O–D peaks. The dotted lines refer to water molecules that are not H bonded to water. The spectra were recorded using the SSP polarization combination. The gray shaded region summarizes the result of this study that there is a broad distribution of non–water–H-bonded water molecules. (**D**) Ratio between the low-(2395 cm^{-1}) and high-(2500 cm^{-1}) frequency bands of the SF spectrum of the planar D_2O–air interface as a function of temperature (red markers) with a quadratic polynomial fit to a lower temperature range (red line) (*21*). The ratio at the oil droplet surface is shown in blue.

the structure of the H-bond network around the hydrophobic oil droplet. The spectral SF intensity between 2200 and 2800 cm^{-1} reported on water molecules that had an anisotropic structure induced by the hexadecane droplet interface. For comparison, Fig. 1C also shows the interfacial SF water spectrum of the extended planar air–water interface (red trace) (*21*). The SF spectra of Fig. 1C had two broad features around 2395 and 2500 cm^{-1} in common (dashed lines), and a narrow peak at 2745 cm^{-1} was only visible in the air–water SF spectrum. The O–D stretch frequency decreased with increasing H-bond strength. Therefore, the feature at 2500 cm^{-1} reported on water molecules that were more weakly H bonded to other water molecules than the ones that were attributed to the feature at 2395 cm^{-1}. The peak at 2745 cm^{-1} in the air–water spectrum originated from interfacial O–D groups that were not H bonded (*22*, *23*). Both SF spectra contained similar 2395 and 2500 cm^{-1} features (Fig. 1C, dashed lines), but they differed at higher wave numbers.

To gain insight into the relative H-bonding strength between the two interfaces, we examined the peak ratio of the 2395 and 2500 cm^{-1} features. This ratio is temperature dependent when measured at the air–water interface (*14*, *21*) and provides insight into the relative amount of stronger versus weaker H-bonded water (Fig. 1D). Lowering the temperature at the air–water interface increased the peak ratio of the 2395 and 2500 cm^{-1} modes (*14*, *21*), indicating that the population of stronger H bonds increased over that of weaker H bonds (Fig. 1D, red data). The blue marker in Fig. 1D indicates the ratio found for water at the oil droplet interface, which was higher than that of the air–water interface of the same temperature. Extrapolating this temperature dependence, we determined that the H-bonding network at the oil droplet surface at room temperature (293 K) was equivalent to that of an air–water interface near the freezing point (277 K, the temperature at which bulk water has its highest density). This implied a more enhanced H-bonded network near the oil–water interface.

In addition to stronger H bonding, there was also a clear difference in the high-frequency side of the spectrum. The non–H-bonded O–D mode at 2745 cm^{-1} was not detected at the oil droplet–water interface. Inspecting the water spectrum adjacent to the oil droplets, a broadening of the high-frequency side was seen (Fig. 1C, dotted line), which could represent a red shift and broadening of the sharp feature at 2745 cm^{-1} in the air–water spectrum. Water molecules that were not H bonded to other water molecules (having so-called "free O–D" modes) should exhibit specific polarization dependences depending on the spatial symmetry of the mode. Their presence could therefore be detected even if they were not spectrally separated from other water molecules.

Polarimetry to assess molecular structure and interactions

Vibrational modes of molecular groups with different symmetry properties have different polarization dependences on the interacting optical fields. A free water molecule exhibits C_{2v} symmetry, yet if the two O–D groups of a water molecule are asymmetric because of H bonding or other interactions, then each O–D group has $C_{\infty v}$ symmetry. The relative signal strength measured by different polarization combinations depends on molecular orientation and symmetry. As it turns out, the symmetry species of the water molecules can be approximated as symmetric C_{2v} O–D stretch modes ($C_{2v\text{-}ss}$), $C_{\infty v}$ O–D stretch modes, and asymmetric C_{2v} stretch O–D stretch modes ($C_{2v\text{-}as}$), despite the water structure being more complex than that (for a description of this analysis, see the supplementary materials, section S3 and fig. S5). Therefore, the polarization intensity ratios for interfacial vibrational modes with these different symmetries were calculated taking all possible molecular tilt angles into account.

The black line in Fig. 2A shows the frequency-dependent SSP/PPP intensity polarization ratio. The red, green, and blue bands in Fig. 2A represent the computed intensity ratios for the $C_{2v\text{-}ss}$ modes, the $C_{\infty v}$ O–D modes, and

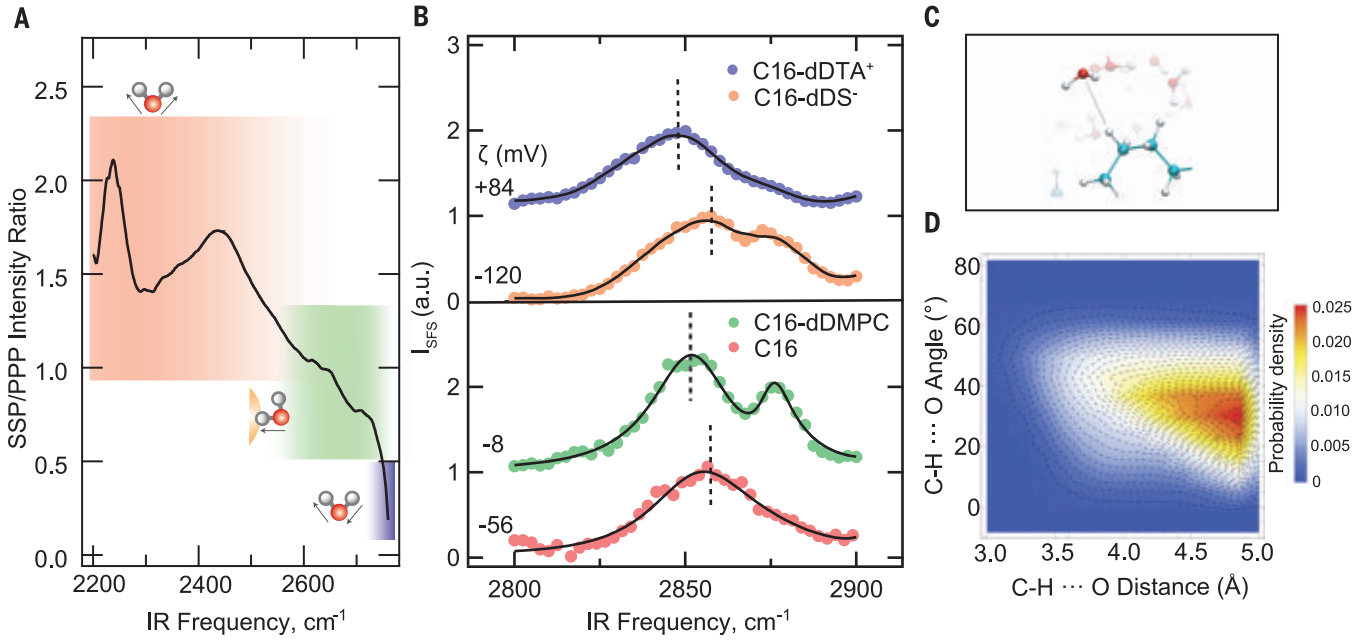


Fig. 2. Interfacial C–H···O H bonds. (**A**) SFS polarimetry. Shown is the SSP/PPP intensity ratio of the O–D spectrum as a function of IR frequency (black line). The shaded regions correspond to the calculated SSP–PPP intensity ratio of the $C_{2v\text{-}ss}$ (red), $C_{2v\text{-}as}$ (blue), and $C_{\infty v}$ (green) modes. The frequency regions spanned by the calculated ratios were determined by comparison with the experimentally measured ratio. The vertical spread of the ratios was computed assuming any molecular orientation and takes into account a 10% uncertainty in the hyperpolarizability values, a scattering angle spread of 20°, and a droplet size distribution as described in the supplementary materials, section S3. (**B**) C–H peak shifts caused by charge transfer. The bottom panel contains the C–H spectra of bare hydrogenated oil droplets (C16, red) and hydrogenated oil droplets covered with d-DMPC (C16-dDMPC, green). The top panel contains the C–H spectra of hydrogenated oil droplets covered with deuterated SDS (C16-dDS$^-$, orange) and d-DTAB (C16-dDTA$^+$, blue). The corresponding ζ-potential values are given next to each spectrum. All spectra were recorded using the SSP polarization combination. The solid black lines are smoothed data as guides to the eye. The dashed vertical lines represent the center frequency of the CH_2-ss mode, implying the presence or absence of water-to-hexadecane charge transfer. (**C**) MD simulations showing C–H···O H bonds. A typical snapshot illustrates a C–H···O H-bonding–like configuration. (**D**) Two-dimensional probability density distribution associating the distance between the C atoms of oil (dodecane) and the O atoms of the water molecules nearby (*x* axis) with the corresponding angle between the C–H bond and the C–O bond vectors (*y* axis).

the $C_{2v\text{-}as}$ modes, respectively. Although the equations for computing the intensity ratios are different for scattering and reflection (*24*), the general selection rule that the SSP polarization combination generates a higher intensity than the PPP polarization combination for the $C_{2v\text{-}ss}$ modes was retained (*25*). The computed intensity ratios were compared with the measured ones to determine which frequency ranges reported primarily on which type of molecular symmetry. From Fig. 2A, we observed that the red band that reports on the $C_{2v\text{-}ss}$ mode spanned the largest frequency range of the measured data, from 2200 to 2550 cm^{-1}. Additionally, the measured and computed intensity ratios from 2550 to 2745 cm^{-1} agreed with the presence of $C_{\infty v}$ O–D stretch modes. In the region where the red and the green bands overlapped, both modes were present. At frequencies higher than 2745 cm^{-1}, the intensity ratios pointed to a small number of $C_{2v\text{-}as}$ modes. The high frequency of these water molecules suggested that there were no H-bonded water molecules to which H bonds could be donated. Because the SSP/PPP intensity ratio for this mode depended on the molecular orientation, we were able to determine that these water molecules were oriented with their O atom facing the oil (see the supplementary materials, section S4 and fig. S5). Furthermore, for the oil droplet interface, the frequency range of water molecules with $C_{\infty v}$ symmetry (i.e., lacking water–H bonds) was much broader (up to ~2550 to 2750 cm^{-1}). The center frequency of this range was at ~2650 cm^{-1}, which was ~100 cm^{-1} lower compared with the air–water interface. This red shift and broadening indicated an unexpected strong interaction between oil and water.

C–H···O H bonds

Generally, non–H-bonded O–D stretches appear as a sharp peak at ~2745 cm^{-1} at the air–water interface (Fig. 1C). Raman hydration shell spectroscopy has reported significant broadening and red shifts of non–H-bonded water up to ~100 cm^{-1} (*26*, *27*). The sizable frequency shift of ~100 cm^{-1} shown in Fig. 1C, highlighted by the gray peak shape, suggested that charge must be transferred from the water O–D groups to the hexadecane C–H groups in the oil phase, because these were the only candidates for accepting electron density. If there were such a strong interaction, then we would expect that the vibrational modes of the oil molecules would experience a (blue) shift in the opposite direction as that of the high-frequency O–D modes (*28*, *29*). These shifts should be accompanied by structural changes, such as the presence of C–H···O H bonds. To investigate this, the SFS spectra of interfacial C–H modes were recorded and MD simulations were performed.

Figure 2B (bottom panel) presents the SFS spectra of the interfacial C–H modes, showing both the symmetric CH_2 (CH_2-ss) and CH_3 (CH_3-ss) stretch modes around ~2850 and ~2878 cm^{-1}, respectively. SFS spectra were recorded from the oil phase of bare oil droplets of hexadecane in water (Fig. 2B, red spectrum) and compared with oil droplets that were covered with a monolayer of deuterated 1,2-dimyristoyl-sn-glycero-3-phosphocholine (d-DMPC), a zwitterionic lipid (Fig. 2B, green

spectrum). The projected area of a single d-DMPC lipid within this monolayer was in the range of 0.65 to 0.75 nm^2 (see the supplementary materials, section S5). The CH_2-ss mode occurred at 2851 cm^{-1} for the d-DMPC–covered hexadecane droplets, whereas it occurred at 2856 cm^{-1} for the bare oil droplets in water. A discussion of other spectral changes is given in the supplementary materials, section S5. The vibrational modes of the lipids were not at resonance in this spectral window. The spectral shift in the CH_2-ss mode is highlighted by the vertical dashed lines in Fig. 2B. The CH_2 modes of the oil directly in contact with D_2O were clearly blue shifted, consistent with a transfer of negative charge from the water to the oil. This shift was not observed when d-DMPC was present because the full monolayer prevented the water from interacting with the oil.

To investigate the possible interactions that may emerge at the oil–water interface, MD simulations were performed for dodecane-water interfaces (see the supplementary materials, materials and methods). Figure 2C shows a typical snapshot of a dodecane–water interface in which the water and oil molecules interacted with one another, with the conformation reminiscent of an H bond between water and oil. In acting as an acceptor of a weak H bond from the C–H bond, the O lone pairs on water molecules donated some electron density, leaving the dodecane molecules with a negative charge. Figure 2D shows a two-dimensional probability density distribution correlating the distances (x axis) and angular distributions (y axis) associated with possible H bonds that form between the water and the C–H groups of dodecane. The graph shows that a broad spectrum of interactions developed. Specifically, there was a significant population of water molecules in which the C–H bond vector pointed to the O lone-pair electrons in slightly distorted geometries (angle <20°) with larger distances (>~3.5 Å) than for water–water H bonds. The water molecules that were within 5 Å from the oil phase oriented with their O atoms toward the CH_2 groups of the oil (see the supplementary materials, section S6 and fig. S6), in agreement with the polarimetric tilt angle analysis of the high-frequency water molecules (see the supplementary materials, section S4 and fig. S5). We thus concluded that C–H···O H bonds were responsible for the transfer of charge between water and oil and led to spectral shifts in both the interfacial water and interfacial oil molecules. C–H···O H bonds were discovered several decades ago (*28*) and were identified as being important structural factors in nucleic acid and protein structures, as well as in enzymatic reactions involving C–H groups and water molecules (*30*). To our knowledge, however, C–H···O H bonds have never been associated with hydrophobicity or the interaction between water and oil, even though C–H modes of small molecules dissolved in water are known to display blue shifts in their C–H stretch frequencies (*31*). The blue shift in C–H modes arises from C–H bond contraction (*28*, *29*). In H bonds in water, charge transfer is the primary contributor to all interactions, leading to bond lengthening and a red shift. C–H···O bonds have relatively weak charge transfer, and thus the Pauli repulsion between the filled C–H and O orbitals dominates the interactions (*29*), resulting in a blue shift (for details, see the supplementary materials, section S7). The charge transferred from water to oil was in the range of 0.025 to 0.05 electrons/dodecane molecule (*12*). Although the energetic stabilization of single C–H···O H bonds is rather weak (a fraction of the thermal energy), the collective contribution of such small interactions can lead to a sufficient buildup of charge on the oil droplet, as predicted by Poli *et al.* (*12*).

Oil–water charge-transfer interactions explain droplet stability

The oil droplets in this study had negative ζ potential values of −56 ± 10 mV. Adding an insulating layer of d-DMPC lipids removed the frequency shift and the negative charge, reducing the ζ potential to −8 ± 6 mV (*32*) (Fig. 2B, green). To further test whether the frequency shifts in the C–H modes were indeed correlated with the interfacial charge, hexadecane droplets with positively and negatively charged deuterated surfactants were prepared using d-dodecyltrimethylammonium bromide ($d\text{-}DTA^+$) and d-sodium dodecyl sulfate ($d\text{-}DS^-$), respectively. $d\text{-}DS^-$ and $d\text{-}DTA^+$ molecules formed dilute monolayers with a projected area per molecule of >4.25 nm^2 for $d\text{-}DS^-$ (*33*) and >5.00 nm^2 for $d\text{-}DTA^+$ (*34*) at their respective critical micelle concentrations. Adding $d\text{-}DS^-$ up to the critical micelle concentration increased the magnitude of the ζ potential further to −120 ± 10 mV. The CH_2-ss mode remained at 2856 cm^{-1} (Fig. 2B, orange), because $d\text{-}SD^-$ adsorbed to the oil droplet surface in a dilute concentration without perturbing the oil molecules (*15*). Therefore, the absence of a blue shift in the CH_2-ss frequency of $d\text{-}DS^-$–covered droplets compared with bare oil droplets indicated that adsorbed $d\text{-}DS^-$ molecules left enough oil–water contact points free to retain charge transfer. Adding positively charged $d\text{-}DTA^+$ up to the critical micelle concentration increased and inverted the sign of the ζ potential to +84 ± 10 mV. $d\text{-}DTA^+$ adsorbed to the oil droplet surface in an equally diluted fashion as $d\text{-}DS^-$ by inserting its alkyl tails and part of the head group into the oil phase (*35*), reverting the charge of the oil droplets to positive and thereby negating the charge transfer. Concomitantly, the CH_2-ss mode frequency shifted to 2847 cm^{-1} (Fig. 2B, blue). Thus, reversal of charge also removed the blue shift observed for the bare oil droplet C–H modes. A possible molecular mechanism that relates charge transfer to negative ζ potential is described in the supplementary materials, section S8 and fig. S7. The negation or retention of water-to-oil charge transfer in these droplet systems was correlated with their stability. Oil droplets covered with positively charged or neutral surfactants were significantly less stable than negatively charged droplets that retained the charge transfer on bare oil droplets (*34*).

Conclusions

In summary, interfacial water structure in contact with hydrophobic oil droplets was characterized by charge-transfer interactions between water and oil molecules that stem from improper C–H···O H bonds. Water at the oil droplet surface had a stronger H-bonding network compared with the planar air–water interface. The spectral shifts of interfacial O–D and C–H modes provided evidence for charge transfer from water to oil that explained the negative charge and stability of bare oil droplets in water.

REFERENCES AND NOTES

1. D. Chandler, *Nature* **437**, 640–647 (2005).
2. P. Ball, *Chem. Rev.* **108**, 74–108 (2008).
3. J. K. Beattie, A. M. Djerdjev, *Angew. Chem. Int. Ed.* **43**, 3568–3571 (2004).
4. J. C. Carruthers, *Trans. Faraday Soc.* **34**, 300–307 (1938).
5. N. Agmon *et al.*, *Chem. Rev.* **116**, 7642–7672 (2016).
6. R. Vácha *et al.*, *J. Am. Chem. Soc.* **133**, 10204–10210 (2011).
7. R. J. Hunter, *Zeta Potential in Colloid Science: Principles and Applications* (Academic Press, 1981).
8. P. B. Petersen, R. J. Saykally, *J. Phys. Chem. B* **109**, 7976–7980 (2005).
9. J.-S. Samson, R. Scheu, N. Smolentsev, S. W. Rick, S. Roke, *Chem. Phys. Lett.* **615**, 124–131 (2014).
10. R. Vácha, V. Buch, A. Milet, J. P. Devlin, P. Jungwirth, *Phys. Chem. Chem. Phys.* **9**, 4736–4747 (2007).
11. Y.-L. S. Tse, C. Chen, G. E. Lindberg, R. Kumar, G. A. Voth, *J. Am. Chem. Soc.* **137**, 12610–12616 (2015).
12. E. Poli, K. H. Jong, A. Hassanali, *Nat. Commun.* **11**, 901 (2020).
13. L. F. Scatena, M. G. Brown, G. L. Richmond, *Science* **292**, 908–912 (2001).
14. S. Strazdaite, J. Versluis, E. H. G. Backus, H. J. Bakker, *J. Chem. Phys.* **140**, 054711 (2014).
15. H. B. de Aguiar, M. L. Strader, A. G. F. de Beer, S. Roke, *J. Phys. Chem. B* **115**, 2970–2978 (2011).
16. A. P. Carpenter, E. Tran, R. M. Altman, G. L. Richmond, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9214–9219 (2019).
17. S. Kulik, S. Pullanchery, S. Roke, *J. Phys. Chem. C* **124**, 23078–23085 (2020).
18. S. Roke, M. Bonn, A. V. Petukhov, *Phys. Rev. B Condens. Matter Mater. Phys.* **70**, 115106 (2004).
19. A. G. F. de Beer, S. Roke, *Phys. Rev. B Condens. Matter Mater. Phys.* **75**, 245438 (2007).
20. J. I. Dadap, H. B. de Aguiar, S. Roke, *J. Chem. Phys.* **130**, 214710 (2009).
21. N. Smolentsev, W. J. Smit, H. J. Bakker, S. Roke, *Nat. Commun.* **8**, 15548 (2017).
22. Q. Du, R. Superfine, E. Freysz, Y. R. Shen, *Phys. Rev. Lett.* **70**, 2313–2316 (1993).
23. I. V. Stiopkin *et al.*, *Nature* **474**, 192–195 (2011).
24. A. G. F. de Beer, S. Roke, *J. Chem. Phys.* **132**, 234702 (2010).
25. H.-F. Wang, W. Gan, R. Lu, Y. Rao, B.-H. Wu, *Int. Rev. Phys. Chem.* **24**, 191–256 (2005).

26. J. G. Davis, K. P. Gierszal, P. Wang, D. Ben-Amotz, . *Nature* **491**, 582–585 (2012).
27. R. Scheu *et al*., *Angew. Chem. Int. Ed.* **53**, 9560–9563 (2014).
28. P. Hobza, Z. Havlas, *Chem. Rev.* **100**, 4253–4264 (2000).
29. Y. Mao, M. Head-Gordon, *J. Phys. Chem. Lett.* **10**, 3899–3905 (2019).
30. S. Horowitz, R. C. Trievel, *J. Biol. Chem.* **287**, 41576–41582 (2012).
31. D. S. Wilcox, B. M. Rankin, D. Ben-Amotz, *Faraday Discuss.* **167**, 177–190 (2013).
32. H. I. Okur, Y. Chen, N. Smolentsev, E. Zdrali, S. Roke, *J. Phys. Chem. B* **121**, 2808–2813 (2017).
33. H. B. de Aguiar, A. G. F. de Beer, M. L. Strader, S. Roke, *J. Am. Chem. Soc.* **132**, 2122–2123 (2010).
34. E. Zdrali, Y. Chen, H. I. Okur, D. M. Wilkins, S. Roke, *ACS Nano* **11**, 12111–12120 (2017).
35. R. Scheu *et al*., *J. Am. Chem. Soc.* **136**, 2040–2047 (2014).
36. S. Pullanchery, S. Kulik, B. Rehl, A. Hassanali, S. Roke, Data for: Charge transfer across C–H···O hydrogen bonds stabilizes oil droplets in water, Zenodo (2021); http://doi.org/10.5281/zenodo.5549457.

ACKNOWLEDGMENTS

We thank H. Okur and H. B. de Aguiar for helpful discussions. S.R. thanks the Julia Jacobi Foundation. **Funding:** This work was supported by the Swiss National Science Foundation (grant 200021-182606-1 to S.P., S.K., B.R., and S.R.). **Author contributions:** S.P., S.K., and B.R. performed the experiments. S.P., S.R., S.K., B.R., and A.H. interpreted the data. A.H. performed the simulations. S.R. conceived and supervised the study. S.R., S.P., A.H., B.R., and S.K. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data in the manuscript and supplementary materials are available through Zenodo (*36*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj3007
Materials and Methods
Supplementary Text
Figs. S1 to S7
Table S1
References (*37*–*60*)

4 May 2021; accepted 19 October 2021
10.1126/science.abj3007

FOREST ECOLOGY

Multidimensional tropical forest recovery

Lourens Poorter[1]*, Dylan Craven[2], Catarina C. Jakovac[1,3], Masha T. van der Sande[1], Lucy Amissah[4], Frans Bongers[1], Robin L. Chazdon[5,6], Caroline E. Farrior[7], Stephan Kambach[8], Jorge A. Meave[9], Rodrigo Muñoz[1,9], Natalia Norden[10], Nadja Rüger[8,11,12], Michiel van Breugel[13,14,15], Angélica María Almeyda Zambrano[16], Bienvenu Amani[17], José Luis Andrade[18], Pedro H. S. Brancalion[19], Eben N. Broadbent[20], Hubert de Foresta[21], Daisy H. Dent[12,22], Géraldine Derroire[23], Saara J. DeWalt[24], Juan M. Dupuy[18], Sandra M. Durán[25,26], Alfredo C. Fantini[27], Bryan Finegan[28], Alma Hernández-Jaramillo[29], José Luis Hernández-Stefanoni[18], Peter Hietz[30], André B. Junqueira[31], Justin Kassi N'dja[32], Susan G. Letcher[33], Madelon Lohbeck[1,34], René López-Camacho[35], Miguel Martínez-Ramos[36], Felipe P. L. Melo[37], Francisco Mora[36], Sandra C. Müller[38], Anny E. N'Guessan[32], Florian Oberleitner[39], Edgar Ortiz-Malavassi[40], Eduardo A. Pérez-García[9], Bruno X. Pinho[37], Daniel Piotto[41], Jennifer S. Powers[42,43], Susana Rodríguez-Buriticá[10], Danaë M. A. Rozendaal[44,45], Jorge Ruíz[46], Marcelo Tabarelli[37], Heitor Mancini Teixeira[44,47,48], Everardo Valadares de Sá Barretto Sampaio[49], Hans van der Wal[50], Pedro M. Villa[51,52], Geraldo W. Fernandes[53], Braulio A. Santos[54], José Aguilar-Cano[10], Jarcilene S. de Almeida-Cortez[55], Esteban Alvarez-Davila[56], Felipe Arreola-Villa[36], Patricia Balvanera[36], Justin M. Becknell[57], George A. L. Cabral[37], Carolina Castellanos-Castro[10], Ben H. J. de Jong[58], Jhon Edison Nieto[10], Mário M. Espírito-Santo[59], Maria C. Fandino[60], Hernando García[10], Daniel García-Villalobos[10], Jefferson S. Hall[13], Alvaro Idárraga[61], Jaider Jiménez-Montoya[62], Deborah Kennard[63], Erika Marín-Spiotta[64], Rita Mesquita[65], Yule R. F. Nunes[59], Susana Ochoa-Gaona[58], Marielos Peña-Claros[1], Nathalia Pérez-Cárdenas[36], Jorge Rodríguez-Velázquez[36], Lucía Sanaphre Villanueva[66,18], Naomi B. Schwartz[67], Marc K. Steininger[68], Maria D. M. Veloso[59], Henricus F. M. Vester[69], Ima C. G. Vieira[70], G. Bruce Williamson[65,71], Kátia Zanini[38], Bruno Hérault[72,73,74]

Tropical forests disappear rapidly because of deforestation, yet they have the potential to regrow naturally on abandoned lands. We analyze how 12 forest attributes recover during secondary succession and how their recovery is interrelated using 77 sites across the tropics. Tropical forests are highly resilient to low-intensity land use; after 20 years, forest attributes attain 78% (33 to 100%) of their old-growth values. Recovery to 90% of old-growth values is fastest for soil (<1 decade) and plant functioning (<2.5 decades), intermediate for structure and species diversity (2.5 to 6 decades), and slowest for biomass and species composition (>12 decades). Network analysis shows three independent clusters of attribute recovery, related to structure, species diversity, and species composition. Secondary forests should be embraced as a low-cost, natural solution for ecosystem restoration, climate change mitigation, and biodiversity conservation.

Tropical forests are converted at alarming rates to other land uses (*1*), yet they also have the potential to regrow naturally on abandoned agricultural fields and pastures. Widespread land abandonment because of fertility loss, migration, or alternative livelihood options has led to a rapid increase in the extent of regrowing forests. Currently, regrowth covers as much as 28% (2.4 million km^2) of the neotropics alone (*2*). Regrowing secondary forests (SFs) form a large and important component of human-modified tropical landscapes and have the potential to play a key role in biodiversity conservation (*3*), climate change mitigation (*2*), and landscape restoration (*4*). A holistic, quantitative understanding of the recovery of multiple SF functions is needed to inform and design effective policies that benefit nature and people from local to global scales.

In this study, we assess the resilience of 12 forest attributes to recover from agriculture and pasture use. Resilience is the ability of a system to absorb disturbances and return to its previous state (*5*). Resilience is driven by two underlying components: the ability to resist disturbance and the ability to recover after disturbance (*6*). We defined "resistance" as the difference between the value of the forest attribute at the start of succession and the average old-growth forest (OGF) values [compare (*7*)], which reflects the combined legacies of previous forest and previous land use, and "recovery" as the ability to return to OGF attribute values after succession. Succession is defined as a change in vegetation structure, species composition (SC), and ecosystem functioning over time after a disturbance (*8*). Secondary succession occurs on previously vegetated lands when a disturbance removes most of the aboveground vegetation and can proceed at fast rates due to legacy effects of previous forest or previous land use, such as a developed soil, seed bank, remnant trees, and resprouting stumps. Successional pathways are, to some extent, predictable but, because of local stochastic factors, are also, to some extent, uncertain (*9*).

Most successional theories have focused on specific ecosystem attributes, such as SC (*10*), species richness (SR) (*11*), forest structure (*12*), or soils (*13*), but they have rarely been conceptually integrated and assessed together. Recovery of these different ecosystem attributes (i.e., dimensions) is likely to depend on one another. For example, rapid recovery of biomass may lead to high litter production and decomposition and, hence, rapid recovery of soil organic carbon.

Chronosequence studies allow us to infer long-term trends in forest recovery by comparing forests with similar land-use history that differ in age since agricultural or pasture abandonment. Single-site studies have assessed the recovery of multiple attributes [summarized in (*14*)], and several synthetic analyses have assessed the recovery of single attributes. They have found that ecosystem functioning, such as nitrogen fixation, recovers fast [in about three decades (*15*)],

whereas aboveground biomass (AGB) and SR recover more slowly [three to seven decades (*16*, *17*)] and SC recovers slowest [i.e., centuries (*17*)]. To date, we lack a comprehensive understanding on how multiple attributes differ in recovery rates and how recovery of these attributes is interrelated.

In this study, we analyze how 12 forest attributes recover during secondary succession and how their recovery is interrelated. We focused on four complementary groups of attributes that capture successional changes in soils [bulk density (BD), carbon (C), and nitrogen (N)], ecosystem functioning [community nitrogen fixers, wood density (WD), and specific leaf area (SLA)], forest structure [AGB, maximum tree diameter, and structural heterogeneity (SH)], and diversity and composition (SR, species diversity, and similarity to OGF). These four groups are key components of ecosystem functioning (*18*), and knowledge of their recovery during succession is a prerequisite for the formulation of global policies on biodiversity conservation, climate change mitigation, and forest restoration. We ask (i) how multiple forest attributes recover during succession, (ii) how their relative recovery is interrelated, and (iii) whether one (or several) attribute(s) can be used as a simple proxy for multidimensional recovery. We advance previous analyses by (i) including a wider range of forest attributes for a larger number of sites (77) compared with those in previous studies, (ii) developing and applying an original conceptual framework to model forest recovery, (iii) examining how recovery among forest attributes is interrelated, and (iv) identifying simple indicators to monitor the progress of forest restoration.

We compiled original chronosequence data from three continents, 77 sites, 2275 plots, and 226,343 stems, spanning the major environmental and latitudinal gradients in the lowland neotropics and West Africa [(*18*); Fig. 1D and table S1]. Chronosequences do not monitor plots over time but rather substitute space for time to infer recovery. Plots were, on average, 0.1 ha, in which all woody plants were identified and measured for their stem diameter. Forest attributes were measured for 21 sites (for soils) up to 77 sites for the other variables. To quantify to what extent SF attributes recover toward OGF values, recovery was modeled for each chronosequence as a process in which SF values return exponentially to OGF values (Fig. 1A). When available, OGF plots were used to estimate OGF chronosequence reference values (supplementary text, section S1). For each study site, relative forest recovery was expressed as the similarity (ranging between 0 and 100%) between the predicted values for SF plots and OGF plots, thereby enabling direct comparisons of recovery across forests and attributes (Fig. 1, A and B). To assess how recovery of different forest attributes was connected during succession and which attributes can serve as proxies for multidimensional recovery, we carried out a network analysis (Fig. 1C).

Pace of recovery

Forest attributes differ in their starting values after land abandonment (i.e., resistance) and subsequent recovery (Fig. 2). Starting values varied from 1 to 90% (Fig. 3A), recovery after 20 years (R_{20y}) varied from 33 to 100% (Fig. 3B), and recovery time (RT) to 90% of OGF values varied from 0 to 120 years (Fig. 3D). The ranking in recovery of the four different groups is maintained when recovery is evaluated in terms of intrinsic recovery rate (λ) instead of percentage of recovery (fig. S2). In the coming sections, we first briefly introduce each group of attributes. See (*18*) for a detailed explanation of their importance and how they recover during succession.

Soil functioning was evaluated in terms of organic C, N, and BD of the topsoil. Soil C concentration scales positively with soil organic matter content and, hence, with nutrients in organic material and water holding capacity. Abandoned agricultural fields and pastures may have low soil C because of combustion during slash and burn (*19*). Soil N concentration is an indicator of soil fertility and may be low in abandoned fields because of uptake by crops and cattle, volatilization, erosion, and leaching (*19*). Soil BD is soil dry mass over soil volume and may be high because of

[1]Forest Ecology and Forest Management Group, Wageningen University, Wageningen, Netherlands. [2]Centro de Modelación y Monitoreo de Ecosistemas, Universidad Mayor, Santiago, Chile. [3]Departamento de Fitotecnia, Universidade Federal de Santa Catarina. Rod. Admar Gonzaga, Florianópolis, SC, Brazil. [4]CSIR-Forestry Research Institute of Ghana, KNUST, Kumasi, Ghana. [5]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA. [6]Tropical Forests and People Research Centre, University of the Sunshine Coast, Maroochydore DC, QLD, Australia. [7]University of Texas at Austin, Austin, TX, USA. [8]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. [9]Departamento de Ecología y Recursos Naturales, Facultad de Ciencias, Universidad Nacional Autónoma de México, Coyoacán, Mexico City, Mexico. [10]Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, Colombia. [11]Department of Economics, University of Leipzig, Leipzig, Germany. [12]Smithsonian Tropical Research Institute, Ancón, Balboa, Panama. [13]SI ForestGEO, Smithsonian Tropical Research Institute, Ancón, Balboa, Panama. [14]Yale-NUS College, Singapore, Singapore. [15]Department of Biological Sciences, National University of Singapore, Singapore, Singapore. [16]Center for Latin American Studies, University of Florida, Gainesville, FL, USA. [17]UFR Agroforesterie, Université Jean Lorougnon Guédé Daloa, Daloa, Côte d'Ivoire. [18]Centro de Investigación Científica de Yucatán A.C. Unidad de Recursos Naturales, Colonia Chuburná de Hidalgo, Mérida, Yucatán, Mexico. [19]Department of Forest Sciences, "Luiz de Queiroz" College of Agriculture, University of São Paulo, Piracicaba, São Paulo, Brazil. [20]Spatial Ecology and Conservation Lab, School of Forest Resources and Conservation, University of Florida, Gainesville, FL, USA. [21]UMR AMAP, Institut de Recherche pour le Développement (IRD), Montpellier, France. [22]Biological and Environmental Sciences, University of Stirling, Stirling, UK. [23]CIRAD, UMR EcoFoG (AgroParistech, CNRS, INRAE, Université des Antilles, Université de la Guyane), Campus Agronomique, Kourou, French Guiana. [24]Department of Biological Sciences, Clemson University, Clemson, SC, USA. [25]Earth and Atmospheric Sciences Department, University of Alberta, Edmonton, AB, Canada. [26]Department of Ecology and Evolutionary Biology, University of Minnesota, St. Paul, MN, USA. [27]Universidade Federal de Santa Catarina, Brazil. [28]CATIE-Centro Agronómico Tropical de Investigación y Enseñanza, Turrialba, Costa Rica. [29]Neotropical Primate Conservation Colombia, Bogotá, Colombia. [30]Institute of Botany, University of Natural Resources and Life Sciences, Vienna, Austria. [31]Institut de Ciència i Tecnologia Ambientals, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Barcelona, Spain. [32]Departement of Bioscience, University Felix Houphouet-Boigny, Abidjan, Côte d'Ivoire. [33]College of the Atlantic, Bar Harbor, ME, USA. [34]World Agroforestry Centre, ICRAF, United Nations Avenue, Gigiri, Nairobi, Kenya. [35]Universidad Distrital Francisco José de Caldas, Facultad de Medio Ambiente y Recursos Naturales, Bogotá, Colombia. [36]Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Morelia, Michoacán, Mexico. [37]Departamento de Botânica, Universidade Federal de Pernambuco, Recife, Brazil. [38]Departamento de Ecologia, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil. [39]Department of Ecology, University of Innsbruck, Innsbruck, Austria. [40]Instituto Tecnológico de Costa Rica, Escuela de Ingeniería Forestal, Cartago, Costa Rica. [41]Centro de Formação em Ciências Agroflorestais, Universidade Federal do Sul da Bahia, Itabuna, BA, Brazil. [42]Department of Ecology, Evolution and Behavior, University of Minnesota, St. Paul, MN, USA. [43]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, USA. [44]Plant Production Systems Group, Wageningen University and Research, Wageningen, Netherlands. [45]Centre for Crop Systems Analysis, Wageningen University and Research, Wageningen, Netherlands. [46]Programa de Estudios de Posgrado en Geografia, Convenio Universidad Pedagogica y Tecnológica de Colombia-Instituto Geografico Agustin Codazzi, Bogotá, Colombia. [47]Farming Systems Ecology, Wageningen University, Wageningen, Netherlands. [48]Copernicus Institute, Utrecht University, Utrecht, Netherlands. [49]Departamento de Energia Nuclear–CTG, Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil. [50]Departamento de Agricultura, Sociedad y Ambiente, El Colegio de la Frontera Sur – Unidad Villahermosa, Centro, Tabasco, México. [51]Program of Botany, Departamento de Biologia Vegetal, Laboratório de Ecologia e Evolução de Plantas, Universidade Federal de Viçosa, Viçosa, Minas Gerais, Brazil. [52]Fundación para la Conservación de la Biodiversidad (PROBIODIVERSA), Mérida, Mérida, Venezuela. [53]Ecologia Evolutiva e Biodiversidade/DBG, ICB, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. [54]Federal University of Paraíba, João Pessoa, Brazil. [55]Departamento de Botânica–CCB, Universidade Federal de Pernambuco, Pernambuco, Brazil. [56]Escuela ECAPMA – Universidad Nacional Abierta y a Distancia, Bogotá, Colombia. [57]Environmental Studies Program, Colby College, Waterville, ME, USA. [58]Department of Sustainability Science, El Colegio de la Frontera Sur, Lerma, Campeche, Mexico. [59]Departamento de Biologia Geral, Universidade Estadual de Montes Claros, Montes Claros, Minas Gerais, Brazil. [60]Fondo Patrimonio Natural para la Biodiversidad y Areas Protegidas, Bogota, Colombia. [61]Fundación Jardín Botánico de Medellín, Herbario JAUM, Medellín, Colombia. [62]Instituto de Biología, Universidad de Antioquia, Antioquia, Colombia. [63]Department of Physical and Environmental Sciences, Colorado Mesa University, Grand Junction, CO, USA. [64]Department of Geography, University of Wisconsin–Madison, Madison, WI, USA. [65]Biological Dynamics of Forest Fragments Project, Environmental Dynamics Research Coordination, Instituto Nacional de Pesquisas da Amazonia, Manaus, Amazonas, Brazil. [66]Consejo Nacional de Ciencia y Tecnologia, Centro del Cambio Global y la Sustentabilidad, Tabasco, Mexico. [67]Department of Geography, University of British Columbia, Vancouver, BC, Canada. [68]Department of Geographical Sciences, University of Maryland, College Park, MD, USA. [69]Institute for Biodiversity and Ecosystem Dynamics (IBED), University of Amsterdam, Amsterdam, Netherlands. [70]Museu Paraense Emilio Goeldi, Belém, Pará, Brazil. [71]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA, USA. [72]CIRAD, UPR Forêts et Sociétés, Yamoussoukro, Côte d'Ivoire. [73]Forêts et Sociétés, Université Montpellier, CIRAD, Montpellier, France. [74]Institut National Polytechnique Félix Houphouët-Boigny, INP-HB, Yamoussoukro, Côte d'Ivoire.
***Corresponding author. Email: lourens.poorter@wur.nl**

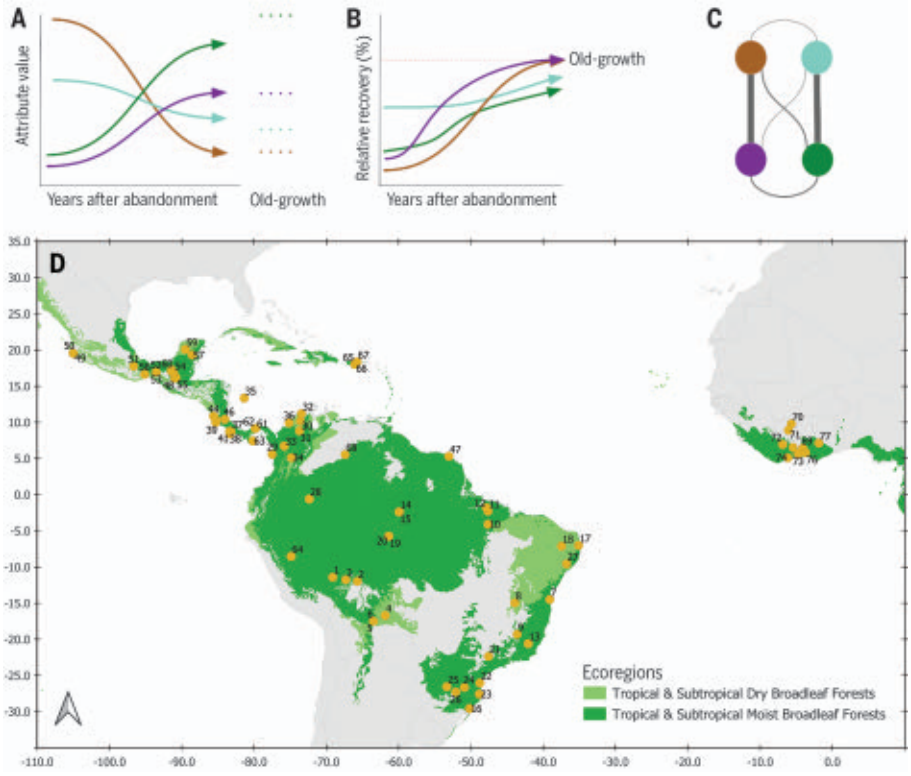


Fig. 1. Study approach to analyze recovery of different forest attributes. (**A** to **C**) Absolute recovery of SF attributes toward OGF values (A) can be standardized to relative recovery rates (B), which express how close each SF attribute is to OGF values, thereby allowing direct comparisons across attributes, such as in network analyses (C), which show how recovery is coordinated across forest attributes. The widths of paths among attributes indicate the strength of the coordination. The different colors indicate attribute category: soil (brown), plant functioning (purple), structure (green), and diversity (turquoise). (**D**) Map of the 77 study sites in the neotropics and West Africa (for site numbers, see table S1). Potential forest cover is shown in green.

soil compaction by agricultural practices and livestock.

We expected soil recovery to occur more slowly than vegetation recovery because soil recovery depends on leaf and root litter inputs. Yet recovery of soil attributes was surprisingly fast [compare (*7*)], with an R_{20y} of 98 to 100% and an RT of 1 to 9 years (Fig. 3, B and D). Just after land abandonment of agriculture or pasture (t_0), the starting values of C, N, and BD were relatively high (62 to 90%; Fig. 3A), which indicates that they are less affected by slashing or burning than aboveground vegetation, contain more legacies of previous land use, and have a high resistance to disturbance. Most of our data come from regrowth after light- to mid-intensity land uses during which soil degradation is not extreme. Soils may also recover quickly due to rapid recovery of the soil biotic community, because slash-and-burn management has transferred nutrients from the aboveground vegetation to the soil, or because productive grass roots and nitrogen-fixing herbs have increased soil C and N (*19*).

Soil C recovered in ~5 years to 90% of OGF values, probably because it is weakly affected by aboveground disturbances associated with land-use change, such as fire and clearing. A meta-analysis found that soil C of SF was similar to that of OGF and did not change during succession (*20*). Most soil nutrients may recover quickly because plants may acquire nutrients from deeper soil layers, because of high litter production early in succession due to ample light availability, and because of the high rates of leaf and root turnover of pioneer species (*21*). Litter quality may also be higher early in succession, because pioneers tend to have high concentrations of leaf nutrients (*22*) and nitrogen fixers are especially abundant (*15*) and active early in succession (*23*). Recovery of phosphorus may be slow because it can only be replenished through atmospheric deposition and mineral weathering (*7*). The observed fast soil recovery is important for the sustainability of shifting cultivation agriculture, which coincides with agronomic studies that indicate that a fallow period of more than 8 to 10 years allows agricultural productivity to be maintained (*21*).

Plant functioning was evaluated in terms of basal area–weighted community WD and SLA and the percentage basal area of nitrogen-fixing trees. WD is the stem-wood dry mass divided by stem volume, and it increases tissue longevity and carbon residence time in trees and forests. SLA is the leaf area divided by the leaf mass. It reflects leaf display cost and scales positively with photosynthetic capacity and forest productivity and negatively with leaf longevity. WD and SLA change during secondary succession because pioneer species are typically replaced by later-successional species with opposite trait values (*24*). Nitrogen fixation indicates the potential for biological nitrogen input to trees and forests. Nitrogen fixation is generally high early in succession when irradiance is high and trees can support their nitrogen-fixing symbionts with carbohydrates (*23*) and declines over time as forests regrow (*15*), light availability in the stand drops, and nitrogen fixation becomes too costly (*23*).

Recovery of ecosystem processes depends on the characteristics of species that make up the community. Although SC may recover slowly, we expected plant functioning to recover at an intermediate pace because many OGF species have similar (i.e., redundant) trait values. We found that plant functioning recovers surprisingly fast (R_{20y} of 82 to 100% and an average RT of 3 to 27 years; Fig. 3, B and D). During succession, short-lived pioneer species (with life spans of 10 to 30 years and extreme trait values) are rapidly replaced by later-successional species that are functionally similar to one another but different from pioneer species (*25*), which leads to a fast functional recovery. Additionally, resprouting is a common mode of regeneration on abandoned fields, which explains why the functional composition rapidly resembles that of the previous OGF [(*26*); Fig. 3A]. Finally, fast recovery also occurs because traits such as SLA and WD never have values of zero and, therefore, start closer to OGF values (85 and 76%, respectively) than, for example, the proportion of nitrogen-fixing trees (40%; Fig. 3A).

Forest structure was evaluated in terms of AGB, maximum tree size (Dmax), and SH. AGB is a strong driver of ecosystem processes (*27*) and important for carbon storage and climate change mitigation (*16*). Dmax reflects the presence of large trees that have a high conservation value, providing habitat and food for many organisms. SH refers to the tree size variation in a plot; it increases light capture and ecosystem productivity (*18*) and contributes to biodiversity conservation by providing a habitat for different species.

We expected forest structure to recover faster than soil and trait attributes because all trees and species contribute to forest structure, but we found that it occurs at an intermediate pace (R_{20y} of 33 to 83% and an average RT of 27 to 119 years; Fig. 3, B and D), probably because it often starts close to

zero. SH recovered at an intermediate pace, probably because it increases with Dmax and because it reflects a gradual transition from even-aged forest that establishes just after land abandonment toward uneven-aged forest with continuous regeneration and multiple cohorts. Recovery of Dmax took more time because it depends on the identity and growth of individual trees. AGB had the slowest recovery because large trees drive AGB (*28*) and because of low productivity in later successional stages (*16*). The RT of 12 decades for AGB is substantially longer than the seven decades we previously estimated, owing to differences in the number of study sites (77 versus 43) and modeling approach (*18*).

Diversity was evaluated in terms of SR, Simpson diversity (SD), and SC. SR is directly relevant for conservation, because it indicates the number of locally co-occurring species. SD indicates the diversity of common species and reflects successional shifts in community structure from young forests dominated by few pioneer species to diverse forests with many rare species. SC indicates to what extent the SC in an area (i.e., the identity of species and their relative abundance) resembles that of an OGF and thus indicates the quality of diversity and the value of SFs for the conservation of old-growth species. SR, species diversity, and SC usually start close to zero (Fig. 3A) with few or no woody plants, owing to biomass and species removal for previous land use, and increase over time as seeds germinate from the seed bank and new species arrive and get established.

Recovery of species diversity and SC occurred at an intermediate to slow pace. SR recovered fastest (R_{20y} = 78%, RT = 37 years) because early in succession, biodiversity can be high because both light-demanding early-successional and shade-tolerant later-successional species coexist (*11*, *29*). SD recovered more slowly (R_{20y} = 69%, RT = 59 years) because it takes time before competition leads to more equal species abundances. SC recovered slowest (R_{20y} = 33%, RT = 120 years) because it depends on overcoming dispersal and recruitment limitations, the accumulation of rare shade-tolerant species, and tree turnover (which takes decades to centuries). Recovery in SC varied substantially across sites (as indicated by wide credibility intervals; Fig. 3D), possibly because sites vary in the drivers of succession, such as land-use history, the number and identity of remnant trees, proximity to seed sources, resprouting ability, and the proportion of wind-dispersed tree species in the community (*17*).

We used a chronosequence approach to infer long-term recovery because few studies have monitored succession over time. Our approach assumes that all plots within a chronosequence had similar starting conditions and follow a similar recovery trajectory, which is not necessarily the case (*9*, *30*). SF chronosequence studies that also monitored dynamics over time showed that dynamic pathways in species diversity, SC, and species structure generally matched chronosequence trends but also showed deviations for some plots (*31*, *32*). Instantaneous trends could show a faster increase for diversity and similar trends for composition and basal area compared with chronosequence predictions (*31*, *32*). Hence, the patterns we observed in this study should be corroborated by long-term studies that monitor SF dynamics over time.

Network properties and proxies for multidimensional recovery

We hypothesized that recovery of different forest attributes would be positively correlated, because recovery of certain attributes (e.g., biomass) can facilitate that of others (e.g., soil C) or can only occur when recovery of other attributes occurs simultaneously. We performed network analyses of relative recovery of multiple attributes after 20 years. The first network analysis was based on pairwise correlations among all 12 attributes and showed that recovery of attributes occurred in parallel (Fig. 4A), with the highest expected influence (i.e., many links with other attributes) for SC, followed by the three structural attributes and soil C (Fig. 4C). This was also confirmed by the results of a principal components analysis, which showed similar associations between recovery of different forest attributes (fig. S3).

The second network analysis was based on partial correlations—i.e., accounting for the variation explained by other attributes—thus showing independent, causal links between attributes. We focused on seven forest attributes that were measured at most study sites (N = 74). We found two clusters of attributes whose recovery is likely to be causally linked (Fig. 4B). First, recovery of the three structural attributes was highly connected, because large trees (Dmax) lead to large SH and contribute disproportionally to forest biomass

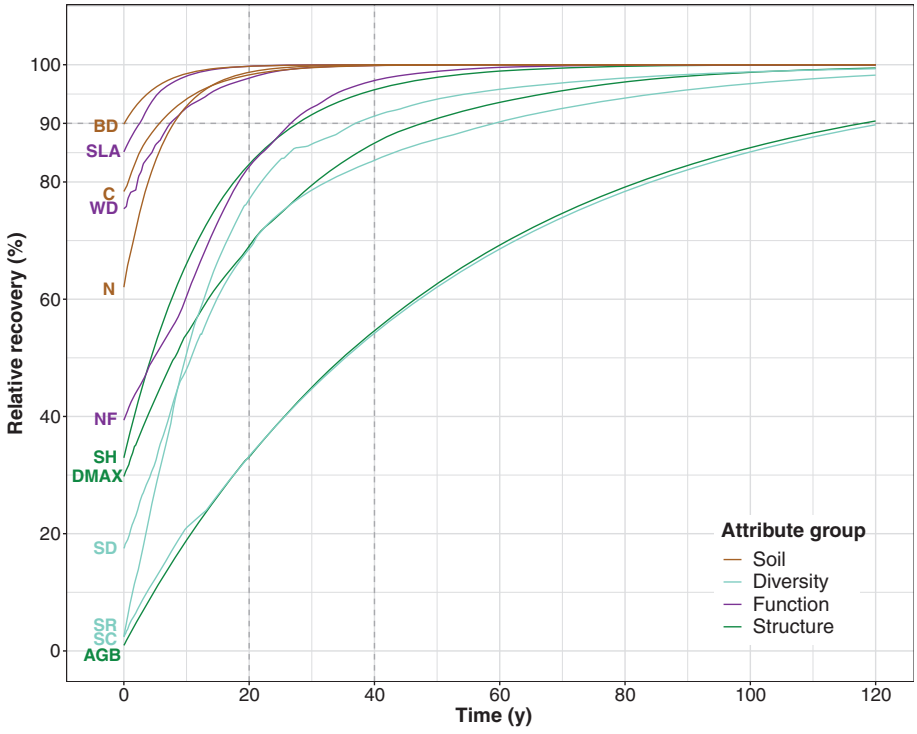


Fig. 2. Predicted relative recovery trajectories over time for 12 forest attributes. The attributes are related to soil (brown), plant functioning (purple), structure (green), and diversity (turquoise). Relative recovery is expressed for each attribute as the similarity (in percentage) between the predicted age-dependent SF value and the OGF value. For some attributes, absolute values increase over time (e.g., AGB), whereas for other attributes, the absolute values generally decrease over time (e.g., BD) (compare Fig. 1A). Here, we show similarity with OGF values, which, by definition, increases over time (compare Fig. 1B). Succession often starts with some remnant trees and soil legacies, and some attributes (SLA and WD) can never be zero, which explains why most attributes do not start at zero (see main text). Dashed lines indicate relative recovery at 20 years, recovery at 40 years, and RT until 90% recovery toward OGF values (see Fig. 3). Recovery trajectories are across-site median values were estimated by using Bayesian models for each attribute (see supplementary text S1). C, soil C; N, soil N; NF, proportional basal area of nitrogen-fixing species; SLA, community-weighted mean SLA; WD, community-weighted mean WD.

(AGB). Forests with more biomass also have a more complex structure. Second, recovery in SR and SD were positively linked, because both increase during succession when species arrive.

When the analysis was repeated for the 43 sites for which SC was also included, a third cluster emerged that showed that recovery in SC, WD, and nitrogen fixation were linked (fig. S4). This may be explained by the fact that succession in SC is underlain by concomitant changes in WD because in wet forests, pioneer species with low WD are replaced by OGF species with high WD, whereas in dry forests, pioneer species with high WD are replaced by OGF species with low WD (*24*).

The clustering of forest attributes into multiple groups suggests that recovery of different forest attributes is shaped by different drivers or processes. For example, recovery of biodiversity attributes may be driven by the landscape context (*17*), land-use history, and the availability of seed trees and dispersal vectors, whereas recovery of structural attributes may be driven by resource availability [i.e., water availability, soil fertility (*16*), and remnant trees].

We hypothesized that AGB would be the best predictor of multidimensional recovery because ecosystem processes and flux rates strongly depend upon the amount of vegetation. Instead, we found that recovery of Dmax had the highest influence (Fig. 4D), indicating that it is strongly linked with other forest structural attributes. The largest tree can be one that regenerated during succession or is a remnant from previous land use. Remnant trees may act as nuclei of forest regeneration (*33*) and kickstart succession (*34*) because they improve microclimate and soil conditions, attract frugivorous seed dispersers (*35*), and favor regeneration of old-growth species. This network structure of forest recovery may be affected by future climate change, but at this stage, we cannot predict how.

To assess and monitor ecosystem recovery, we aimed to identify indicators that change continuously during succession and that are closely correlated with recovery of other ecosystem properties and functions. To be operational, the indicators should be easy to measure, scalable, and cost-effective to implement and use (*36*). To serve simultaneously as resilience indicators, they should be slowly changing variables that underlie ecosystem capacity to recover (*5*). We identified three clusters of forest attributes, related to structure, species diversity, and SC (Fig. 4D and fig. S4B). Dmax and SH are robust indicators of structural recovery; they take, respectively, 5 and 2.5 decades to recover, have a central position in the multidimensional recovery network, and can easily be measured and monitored, either in the field or by remote sensing (*37*). Recovery of Dmax is strongly linked to recovery of AGB, which is more time consuming to measure, but Dmax is weakly linked to recovery of biodiversity attributes (Fig. 4B). Hence, SR should be used as an additional indicator, because it is more closely linked

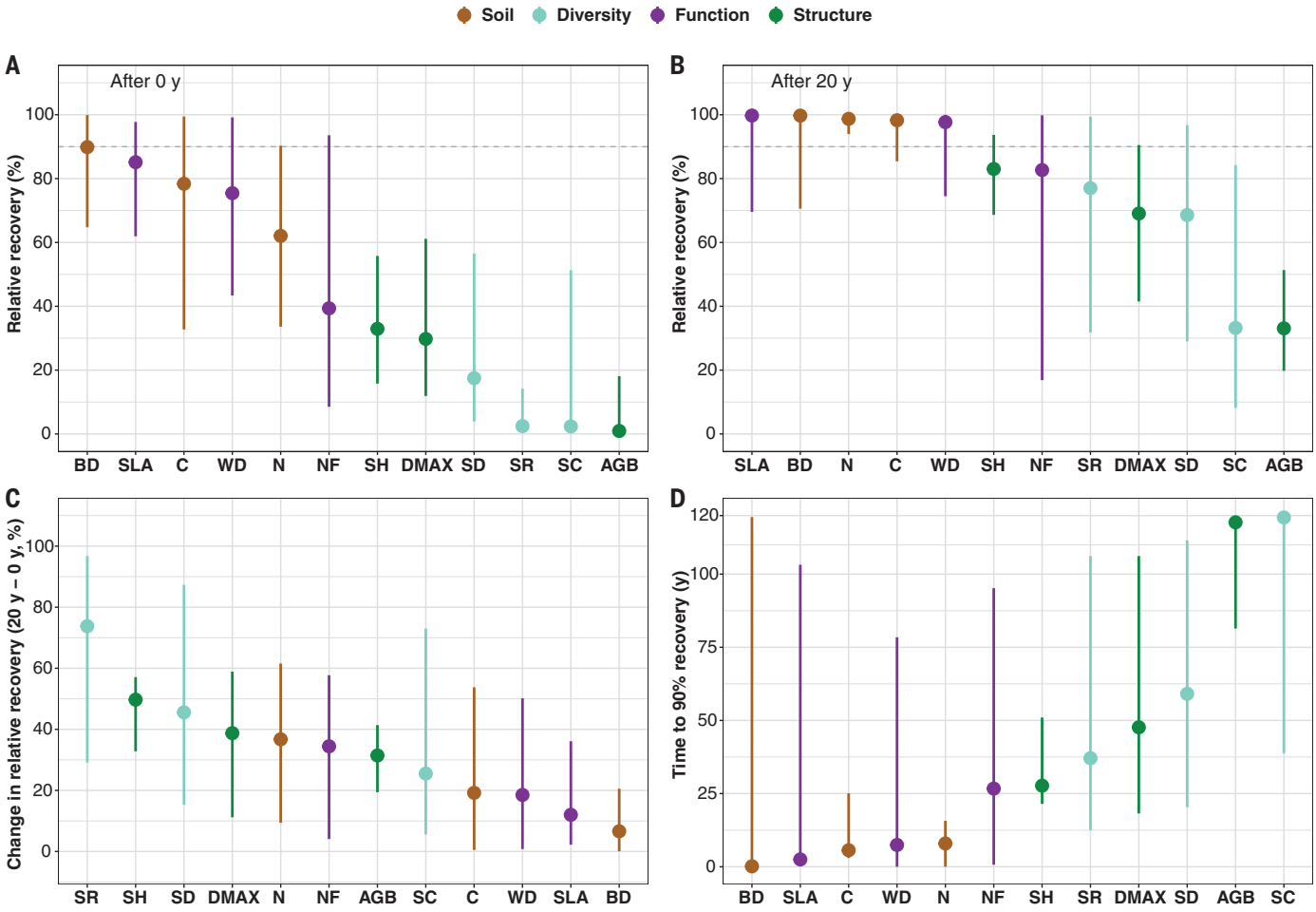


Fig. 3. Relative recovery of SF attributes. (**A**) Recovery at land abandonment (0 year). (**B**) Recovery at 20 years. (**C**) Difference in recovery between 0 and 20 years ($R_{20y} - R_{0y}$). (**D**) RT until 90% recovery toward OGF values. Relative recovery is expressed as the similarity (in percentage) between the SF value and the OGF value. Medians and 95% credible intervals are based on site predictions; *N* = 21 sites for soil attributes, 31 for SLA, 46 for SC, and 77 for all the other attributes. The credible intervals nearly correspond to the range of observations, and in (D), it has been truncated to 120 years to increase resolution. The attributes are related to soil, plant functioning, structure, and diversity and are ranked from fast (left) to slow (right).

to the recovery of other biodiversity attributes. A large number of species also ensures that there is a large diversity in species responses to environmental conditions, which increases the adaptive capacity of ecosystems to deal with environmental change (*38*). When these slower indicators have recovered, faster attributes, such as soil and plant functioning, will have recovered as well (Fig. 3D).

Resilience

We assessed the resilience of forest attributes on the basis of the relative starting value at agricultural abandonment (t_0; resistance) and subsequent recovery rates (λ) during secondary succession. Aboveground attributes such as structure and diversity had low starting values because of the nearly complete removal of woody vegetation for agricultural use, whereas soil attributes had high starting values because of belowground legacies (Fig. 3A). We found that resistance and recovery were positively correlated [correlation coefficient (r) = 0.78, P = 0.0026; fig. S5), which partially explains why some attributes recover quickly and others slowly. All 12 attributes recovered close to their predisturbance values within ~120 years (Figs. 2 and 3D), which is notably fast given that tropical forests are complex in terms of structure, SR, evenness, and plant interactions (*6*). Fast forest recovery during secondary succession can be explained by the many legacies and the relatively productive, warm, and wet conditions of most study sites. We show that tropical forests are resilient to agricultural use, provided that agricultural use has not been too long, intense (*39*), or extensive and that there is sufficient forest in the surrounding area to provide seeds (*40*). Average RTs of the 12 attributes varied from <1 to 12 decades. To assess ecological resilience, the attributes and time frames that are considered [compare (*5*)] are therefore crucial, because soil scientists would consider forests to be highly resilient on the basis of soil legacies, whereas conservationists would consider forests to have low resilience on the basis of the slow recovery of SC.

Applied implications

SFs cover large areas and provide multiple services to local and global stakeholders (*3*). Their fast multidimensional recovery has important implications for ecosystem restoration, climate change mitigation, and biodiversity conservation. Rapid recovery of plant functioning suggests that restoration of ecosystem functioning (such as productivity) should be similarly rapid, because it is underpinned by the traits used in this study (*41*). Rapid soil N and C recovery indicate that natural regrowth provides an inexpensive, nature-based solution to restore the fertility of agricultural lands. Rapid recovery of soil C is crucial for climate change mitigation. The soil C pool exceeds that of biomass and is more persistent because it is less affected by aboveground disturbances, such as fire and clearing. Rapid recovery in SR means that SFs form an important biodiversity

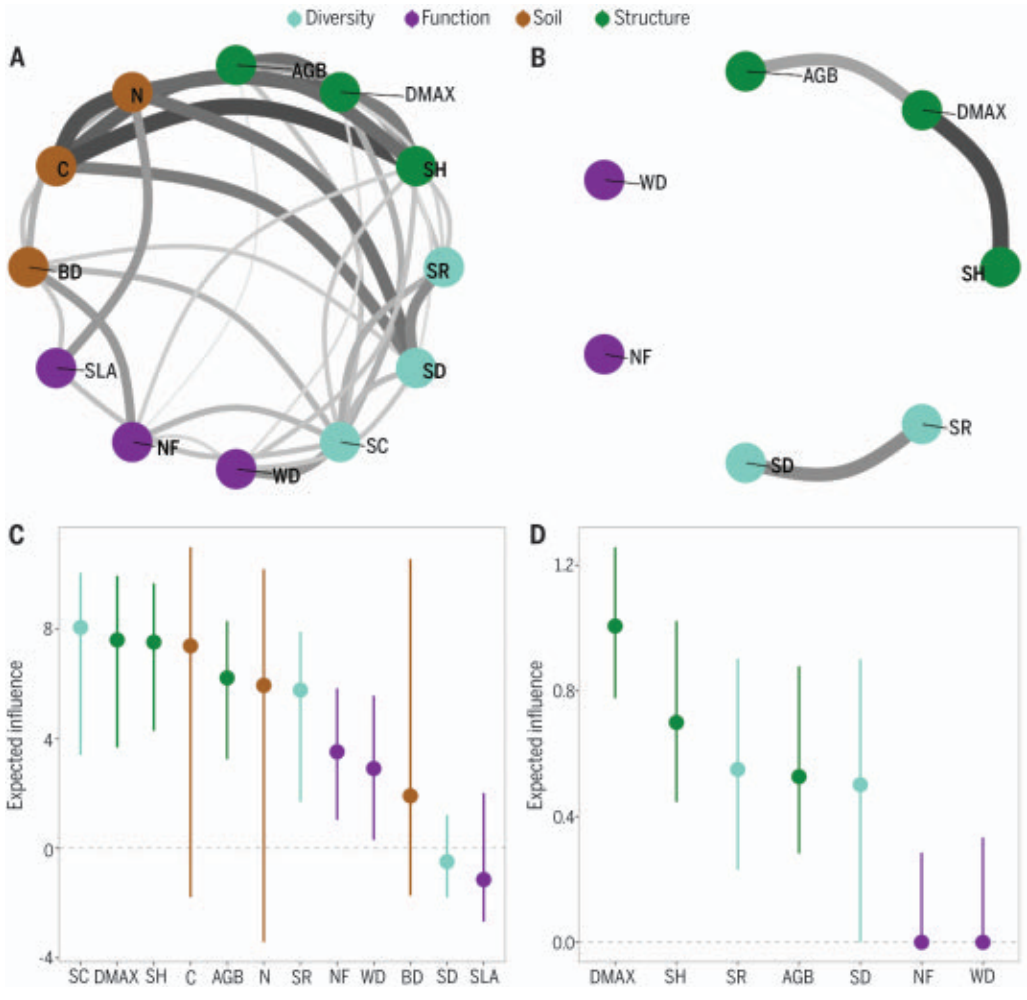


Fig. 4. Network analysis for relative recovery of attributes after 20 years. (**A** to **D**) The networks are based on Pearson's correlations for 12 attributes (left panels) and partial correlations for seven attributes (right panels). (A) and (B) show connectivity among SF attributes. (C) and (D) show expected influence of individual attributes on the network. The correlation network to the left indicates how attributes are associated with one another and uses, for each pairwise correlation, the maximum number of sites possible (N = 17 to 77; table S2). The partial correlation network indicates the direct links between two attributes, independent from others, and is based on a subset of attributes available for most sites (N = 74 sites). In (A) and (B), the line thickness indicates the strength of the (partial) correlation, and lines indicate significant pairwise correlations (i.e., with a 95% confidence interval that does not overlap with zero). The expected influence is the sum of the partial correlations between the target attribute and other attributes. Edge weights and expected influence were estimated from 10,000 bootstraps of the empirical network, and the bootstraps were used to calculate the 95% confidence intervals displayed in (C) and (D). Soil attributes were based on a smaller sample size (N = 21) and therefore had wider credibility intervals. The spinglass algorithm identified three clusters in (A) (AGB-Dmax-SH-C-N, BD-NF, and SC-SR-SD-SLA-WD) and two clusters in (B) (AGB-Dmax-SH and SR-SD). For the partial network, the correlation stability coefficient was 0.43 for edge weights and 0.51 for expected influence. Attributes are colored according to their category: soil (brown), plant functioning (purple), structure (green), and diversity (turquoise).

reservoir in human-modified landscapes and should be conserved (*17*). This will also lead to greater connectivity for plants and animals in fragmented, human-modified landscapes.

Although SFs show, on average, a rapid recovery, there is also substantial variation across the study region (see credibility intervals in Fig. 3), which indicates that some areas may show arrested succession because of a lack of seed sources or dominance of invasive grass, ferns, or woody species. Under such conditions, management practices for assisted natural regeneration—such as weeding, controlling invasive species, enrichment planting, and the establishment of ecological corridors—are needed to safeguard multidimensional recovery.

Given the local and global importance of SFs and their substantial R_{20y} (on average, 78%; range, 33 to 100%; Fig. 3B), we urge the embrace of SFs as a low-cost, nature-based solution to meet the United Nations' Sustainable Development goals and the United Nations' Decade on Ecosystem Restoration goals (*42*). Enabling policies combined with careful landscape planning should identify areas where SFs are best conserved and provide the most co-benefits while minimizing socioecological conflicts [e.g., (*3*, *4*)]. SFs should feature prominently in restoration portfolios, where older SFs and OGFs are conserved, severely degraded areas are actively restored, and young regrowth is protected from deforestation. For example, SFs can only deliver their full conservation potential if they are conserved for a sufficient amount of time, such that tree species can attain reproductive maturity and maintain viable populations (*43*). In addition, substantial gains are made when young, 20-year-old SFs are conserved for 20 years more, because AGB, SR, and similarity with OGFs increase by 15 to 22% (fig. S1C). All OGFs should be conserved because little remains; they harbor many distinctive OGF species and provide seed sources and dispersers to assure landscape resilience (*17*, *40*). To monitor passive and active restoration success, we recommend using simple indicators in different phases of succession, in which SH can be used for the first 25 years (Fig. 3D) and Dmax and SR in the following 25 years.

Conclusions

Our analysis shows that tropical forests and their soils are highly resilient because all attributes recover within 12 decades after low- to moderate-intensity land use. Recovery of soil attributes (<1 decade) and plant functional attributes (<2.5 decades) is very fast, followed by recovery of structure and diversity (2.5 to 6 decades), and recovery of AGB and SC is slowest (12 decades). Network analysis shows that recovery is multidimensional, with three clusters of attributes related to structure, SR, and SC. Monitoring of forest restoration could use Dmax, SH, and SR as complementary indicators of multidimensional recovery.

REFERENCES AND NOTES

1. Food and Agricultural Organization of the United Nations (FAO), "Global forest resources assessment 2020 – Key findings" (FAO, 2020).
2. R. L. Chazdon *et al*., *Sci. Adv.* **2**, e1501639 (2016).
3. P. H. S. Brancalion *et al*., *Sci. Adv.* **5**, eaav3223 (2019).
4. B. B. N. Strassburg *et al*., *Nat. Ecol. Evol.* **3**, 62–70 (2019).
5. S. Carpenter, B. Walker, J. M. Anderies, N. Abel, *Ecosystems* **4**, 765–781 (2001).
6. S. L. Pimm, *Nature* **307**, 321–326 (1984).
7. B. Ayala-Orozco *et al*., *Land Degrad. Dev.* **29**, 315–325 (2018).
8. J. P. Grime, *Plant Strategies, Vegetation Processes, and Ecosystem Properties* (Wiley, 2006).
9. N. Norden *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8013–8018 (2015).
10. D. Tilman, *Plant Strategies and the Dynamics and Structure of Plant Communities* (Monographs in Population Biology, Princeton Univ. Press, 1988).
11. J. H. Connell, *Science* **199**, 1302–1310 (1978).
12. C. D. Oliver, B. C. Larson, *Forest Stand Dynamics: Updated Edition* (Wiley, 1996).
13. L. R. Walker, D. A. Wardle, R. D. Bardgett, B. D. Clarkson, *J. Ecol.* **98**, 725–736 (2010).
14. R. L. Chazdon, *Second Growth: The Promise of Tropical Forest Regeneration in an Age of Deforestation* (Univ. Chicago Press, 2014).
15. M. Gei *et al*., *Nat. Ecol. Evol.* **2**, 1104–1111 (2018).
16. L. Poorter *et al*., *Nature* **530**, 211–214 (2016).
17. D. M. A. Rozendaal *et al*., *Sci. Adv.* **5**, eaau3114 (2019).
18. See supplementary materials.
19. E. Veldkamp, M. Schmidt, J. S. Powers, M. D. Corre, *Nat. Rev. Earth Environ.* **1**, 590–605 (2020).
20. P. A. Martin, A. C. Newton, J. M. Bullock, *Proc. Biol. Sci.* **280**, 20132236 (2013).
21. P. H. Nye, D. J. Greenland, *The Soil Under Shifting Cultivation* (Technical Communication No. 51, Commonwealth Agricultural Bureau, 1960).
22. L. Poorter, M. van de Plassche, S. Willems, R. G. A. Boot, *Plant Biol.* **6**, 746–754 (2004).
23. S. A. Batterman *et al*., *Nature* **502**, 224–227 (2013).
24. L. Poorter *et al*., *Nat. Ecol. Evol.* **3**, 928–934 (2019).
25. B. Finegan, *Trends Ecol. Evol.* **11**, 119–124 (1996).
26. M. F. Barros *et al*., *For. Ecol. Manage.* **482**, 118881 (2021).
27. M. Lohbeck, L. Poorter, M. Martínez-Ramos, F. Bongers, *Ecology* **96**, 1242–1252 (2015).
28. J. W. F. Slik *et al*., *Glob. Ecol. Biogeogr.* **22**, 1261–1271 (2013).
29. F. Bongers, L. Poorter, W. D. Hawthorne, D. Sheil, *Ecol. Lett.* **12**, 798–805 (2009).
30. E. A. Johnson, K. Miyanishi, *Ecol. Lett.* **11**, 419–431 (2008).
31. E. Lebrija-Trejos, J. A. Meave, L. Poorter, E. A. Pérez-García, F. Bongers, *Perspect. Plant Ecol. Evol. Syst.* **12**, 267–275 (2010).
32. R. L. Chazdon *et al*., *Philos. Trans. R. Soc. London Ser. B* **362**, 273–289 (2007).
33. G. Yarranton, R. Morrison, *J. Ecol.* **62**, 417–428 (1974).
34. A. E. N'Guessan *et al*., *For. Ecol. Manage.* **433**, 325–331 (2019).
35. A. D. Manning, J. Fischer, D. B. Lindenmayer, *Biol. Conserv.* **132**, 311–321 (2006).
36. T. H. Oliver *et al*., *Trends Ecol. Evol.* **30**, 673–684 (2015).
37. T. Jucker *et al*., *Glob. Change Biol.* **23**, 177–190 (2017).
38. B. Sakschewski *et al*., *Nat. Clim. Change* **6**, 1032–1036 (2016).
39. C. C. Jakovac, M. Peña-Claros, T. W. Kuyper, F. Bongers, *J. Ecol.* **103**, 67–77 (2015).
40. V. Arroyo-Rodríguez *et al*., *Biol. Rev. Camb. Philos. Soc.* **92**, 326–340 (2017).
41. L. Poorter *et al*., *Glob. Ecol. Biogeogr.* **26**, 1423–1434 (2017).
42. UN Environment Assembly (UNEA), "Resolution 73/284: United Nations Decade on Ecosystem Restoration (2021–2030)" (UNEA, 2019); https://undocs.org/A/RES/73/284.
43. M. van Breugel *et al*., *PLOS ONE* **8**, e82433 (2013).
44. D. Craven, dylancraven/MultiDimensional Recovery, Zenodo (2021); https://doi.org/10.5281/zenodo.4988848.

ACKNOWLEDGMENTS

This paper is a product of the 2ndFOR collaborative research network on SFs (www.2ndFOR.org) and the sDiv working group sUCCESS and is paper no. 7 of 2ndFOR. We thank the owners of the SF sites for access to their forests, all the people who have established and measured the plots, the institutions and funding agencies that supported them (see below), and M. Aide for data use. **Funding:** This research was supported by European Research Council-ERC (Advanced Grant PANTROP 834775 to L.P.); German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig (sDIV W7.20 sUCCESS to L.P., N.R., and M.v.B.) funded inter alia by the Deutsche Forschungsgemeinschaft (DFG; FZT-118); Netherlands Organisation for Scientific Research - NWO (ALW.OP241 to L.P., M.T.v.d.S., and C.C.J.; ALWOP.457 to F.B. and R.Mu.; ALW 863.15.017 to M.L.; and Veni.192.027 to M.T.v.d.S.); NWO-Fundação de Amparo á Pesquisa do Estado de São Paulo 17418 (NEWFOR) to P.H.B., F.B., and M.P.-C.; Agencia Nacional de Investigación y Desarrollo (FONDECYT Regular No. 1201347) to D.C.; Conselho Nacional de Desenvolvimento Científico e Tecnológico-CNPq (to A.C.F., I.C.G.V.; 308471/2017-2 to M.M.E.-S.; 308877/2019-5 to Y.R.F.N.; #312178/2019-0 to B.A.S.; CNPq 308778-2017-0 to I.C.G.V.; CNPq to G.W.F.; Universal 01/2016 to H.M.T.; 309659/2019-1 to S.C.M.; and SinBiose-REGENERA 442371/2019-5 to C.C.J. and R.Me.); Corredor Biológico La Gamba (COBIGA) to F.O.; Deutsche Forschungsgemeinschaft DFG (RU 1536/3-1 to N.R.); Fondo Mixto CONACYT-Gobierno del estado de Yucatán (FOMIX YUC-2008-C06-108863 to J.M.D. and J.L.H.-S.); Fundacão de Amparo à Pesquisa de Minas Gerais-FAPEMIG (to G.W.F.; PPM-00627-16 to Y.R.F.N.; PPM-00726-16 to M.M.E.-S.; PPM-00623-16 to M.D.M.V.; and APQ-03348-16 to H.M.T.); Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul-FAPERGS (2218–2551/ 12-2 to S.C.M.); Fundación Jardín Botánico de Medellín to A.I.; GEF/FONACIT to P.M.V.; German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, a research center of the German Research Foundation (DFG – FZT 118) (iDiv-Flexpool grant nos. 34600967 and 34600970 to N.R. and S.K.); Herbario JAUM to A.I.; Rainforest Luxemburg to F.O.; Secretaría de Educación Pública-Consejo Nacional de Ciencia y Tecnología, Ciencia Básica (SEP-CONACYT 2015-255544 to P.B. and F.M.); Stichting Het Kronendak to H.M.T. and H.F.M.V.; STRI, ForestGEO, Heising–Simons Foundation, HSBC Climate Partnership, Stanley Motta, Small World Institute Fund, the Hoch family, to J.S.H. and M.v.B.; Universidad de Antioquia to J.J.-M., Universidad Nacional Autónoma de México, Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (DPAGA–PAPIIT IN218416, DPAGA–PAPIIT IN217620 to J.A.M., E.A.P.-G. and R.Mu.; PAPIIT-UNAM IN211417 to P.B. and F.M.); Rufford Small Grants 19426-2 to F.M.; SENACYT Panama Grant (COL10-052) to D.H.D.; Tropenbos Foundation to H.F.M.V.; US National Science Foundation (no. 9208031 to D.H.D. and S.J.D.; EAR-1360391 to M.v.B.; and Graduate Fellowship to S.G.L.); Wageningen University and Research Interdisciplinary Research and Education Fund (FOREFRONT program) to F.B. and H.M.T.; and Yale-NUS College and MOE (through a startup grant and grant IG16-LR004) to M.v.B. A.H.-J. was supported by the LICCI project, funded by the European Research Council (FP7-771056-LICCI). This work contributes to the "María de Maeztu" Programme for Units of Excellence of the Spanish Ministry of Science and Innovation (CEX2019-000940-M). **Author contributions**: L.P., B.H., D.C., C.C.J., M.T.v.d.S., L.A., F.B., C.E.F., R.L.C., S.K., J.A.M., R.Mu., N.N., N.R., and M.v.B. conceived the idea; all authors but C.E.F., S.K., and N.R. contributed data; B.H., D.C., C.C.J., M.T.v.d.S., and L.P. analyzed the data; L.P. led the writing of the manuscript with the help of B.H., D.C., C.C.J., and M.T.v.d.S.; L.A., F.B., C.E.F., R.L.C., S.K., J.A.M., R.Mu., N.N., N.R., M.v.B., A.M.A.Z., B.A., J.L.A., P.B., P.H.S.B., E.N.B., H.d.F., D.H.D., G.D., S.J.D., J.M.D., S.M.D., A.C.F., C.E.F., B.F., A.H.-J., J.L.H.-S., P.H., A.H.-J., J.K., S.G.L., M.L., R.L.-C., M.M.-R., F.P.L.M., F.M., S.C.M., A.E.N., F.O., E.O.-M., E.A.P.-G., B.X.P., D.P., J.S.P., S.R.-B., D.M.A.R., J.R., M.T., H.M.T., E.V.d.S.B.S., H.F.M.V., H.v.d.W., P.M.V., and G.W.F. commented upon the results and the manuscript; and all authors approved submission of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data on relative recovery of 12 attributes and the code are available at Zenodo (*44*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abh3629
Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S3
References (*45*–*70*)

10 March 2021; accepted 6 October 2021
10.1126/science.abh3629

REPORTS

PLANT SCIENCE

Evolution of the grass leaf by primordium extension and petiole-lamina remodeling

A. E. Richardson[1,2,3]*, J. Cheng[1,4,5], R. Johnston[6,7], R. Kennaway[1], B. R. Conlon[6], A. B. Rebocho[1]†, H. Kong[4,5], M. J. Scanlon[6]*, S. Hake[2], E. Coen[1]*

The sheathing leaf found in grasses and other monocots is an evolutionary innovation, yet its origin has been a subject of long-standing debate. Here, we revisit the problem in the light of developmental genetics and computational modeling. We show that the sheathing leaf likely arose through *WOX*-gene-dependent extension of a primordial zone straddling concentric domains around the shoot apex. Patterned growth within this zone, oriented by two polarity fields, accounts for wild-type, mutant and mosaic grass leaf development, whereas zone contraction and growth remodeling accounts for eudicot leaf development. In contrast to the prevailing view, our results suggest that the sheath derives from petiole, whereas the blade derives from the lamina of the eudicot leaf, consistent with homologies proposed in the 19th century.

The grass leaf is a conundrum. Unlike a eudicot leaf, which typically has a broad lamina, narrow petiole, and basal stipules (Fig. 1, A to C), the grass leaf has a cylindrical sheath supporting a strap-like blade (Fig. 1, D to F). The encircling sheath, a derived feature of monocots (*1*, *2*), allows grasses to grow in height during the vegetative phase without extending stem internodes, keeping the apical meristem protected close to the ground.

Evolution of the sheathing leaf presents two problems. First, unlike eudicot leaf primordia, which occupy a fraction of the apical meristem circumference, sheathing leaf primordia extend to encircle the meristem (*1*, *3*, *4*) (Fig. 1, G to J). It is unclear how this extension arose. Second, the origins of sheath and blade are uncertain. The grass sheath was originally considered homologous to petiole and blade to lamina: the "petiole-sheath" hypothesis (*5*, *6*) (Fig. 1K). Later, the petiole-like parallel venation of grasses led to the idea that the grass leaf mainly derives from the petiole [phyllode theory (*7*–*9*); Fig. 1L] or from the petiole base: the current "petiole-leaf" hypothesis (*1*, *10*–*14*) (Fig. 1M). Here, we revisit these problems through developmental genetics and computational modeling.

The grass leaf primordium emerges from a primordium zone (PZ) (Fig. 1N), which lacks *KNOX* expression (*15*). The PZ straddles concentric domains that will give rise to the adaxial (upper) and abaxial (lower) regions of the leaf which meet at a midplane boundary (green) (*16*, *17*). The PZ is also subdivided mediolaterally into central, lateral, and marginal domains (*18*) (Fig. 1, O and P). Marginal identity depends on *NARROWSHEATH* genes (*NS1* and *NS2*), members of the *WUSCHEL-RELATED HOMEOBOX* (*WOX*) gene family (*19*, *20*). *ns1/2* double-mutant primordia do not fully encircle the apex and produce narrow leaves (*21*).

To understand how these domains control leaf morphogenesis and to clarify hypothesis predictions, we modeled their growth. In simulations, morphology is an emergent property that depends on how specified local growth rates interact with mechanical tissue constraints.

To simulate primordium emergence from a dome-shaped apex, we built on a recently proposed model based on growth oriented by two polarity fields (*22*): an orthoplanar field running orthogonal to the tissue surface and a planar polarity field running parallel to the tissue surface. We first tried to model growth with orthoplanar polarity alone. Orthoplanar polarity ran from the tissue surface toward the ad-abaxial midplane to orient primordium emergence and toward an axial domain to orient apex growth (Fig. 2, A and B). Growth rates were specified in two orientations: K_{OP}, parallel to orthoplanar polarity, and K_{PER}, perpendicular to orthoplanar polarity. Setting K_{PER} greater than K_{OP} in the PZ generated a ring-shaped primordium encircling the apex (Fig. 2, C and D, and figs. S1A and S2, A to D).

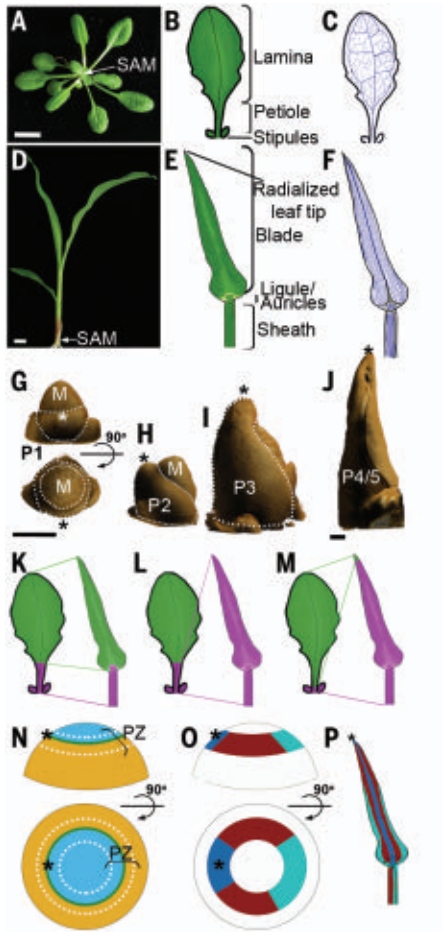


Fig. 1. Eudicot and grass leaf. (**A** to **F**) Eudicot *Arabidopsis* [(A) to (C)] and grass *Zea mays* [(D) to (F)]. (A) and (D) are seedlings. SAM, shoot apical meristem. Scale bar, 1 cm. (B) and (E) are mature leaf. (C) and (F) are venation patterns. (**G** to **J**) Optical projection tomography of maize leaf primordia. (G) is plastochron 1 (P1) viewed from the side or top down. (H) and (I) are P2 and P3 viewed from the side. (J) P4/P5 with wrapped margins (front view). M, meristem. Dotted line indicates the primordium. Scale bar, 100 μm. (**K** to **M**) Proposed homologies between eudicot and grass leaves. (**N** to **P**) Domains in the grass leaf primordium. (N) The primordial zone (PZ, dotted line) straddles the midplane (green line) between the abaxial (orange) and adaxial (blue) domains. (O) and (P) The central (blue), lateral (red), and marginal (cyan) domains in the PZ and the mature leaf [modified from (*18*)]. Asterisk indicates the presumptive midvein tip.

To generate a primordium that slopes down from the midvein tip (Fig. 1H), we modulated K_{PER} such that it decreased mediolaterally. The resulting primordium was sloped but lacked an upwardly growing tip (Fig. 2, E and F, and fig. S2, E and F), suggesting that planar polarity may be required for proper shaping.

[1]John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK. [2]Agricultural Research Service/US Department of Agriculture Plant Gene Expression Center, Albany, CA 94710, USA. [3]Institute of Molecular Plant Science, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, UK. [4]State Key Laboratory of Systematic and Evolutionary Botany, CAS Center for Excellence in Molecular Plant Sciences, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China. [5]College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China. [6]Plant Biology Section, School of Integrative Plant Science, Cornell University, Ithaca, NY 14853, USA. [7]The Elshire Group Limited, Palmerston North 4472, New Zealand.
***Corresponding author. Email: annis.richardson@ed.ac.uk (A.E.R.); mjs298@cornell.edu (M.J.S.); enrico.coen@jic.ac.uk (E.C.)**
†Present address: Enza Zaden, 1602 DB Enkhuizen, Netherlands.

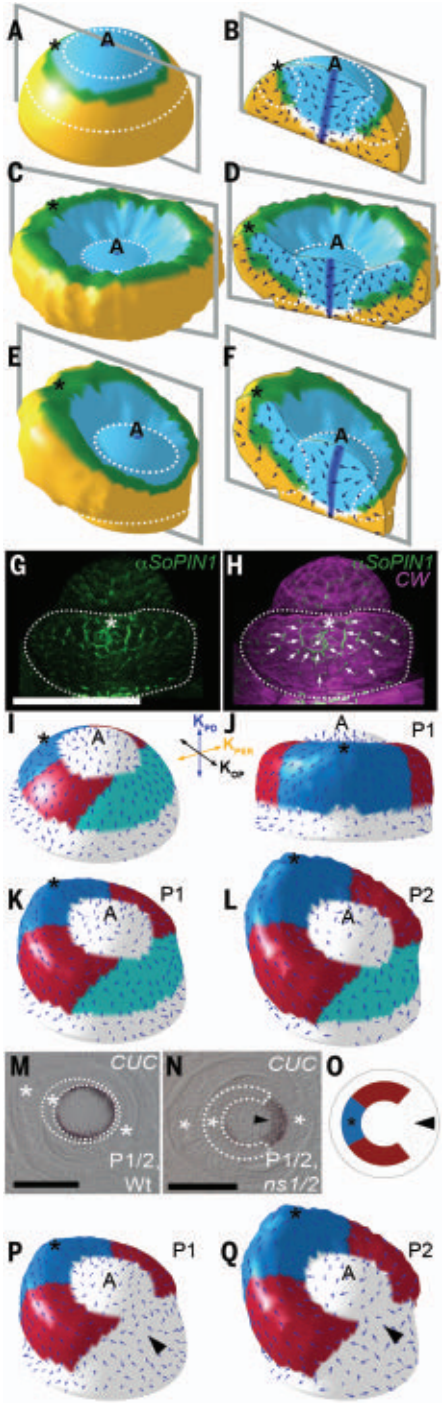


Fig. 2. Grass leaf primordium emergence models. (**A**) Meristem apex with abaxial (orange) and adaxial (blue) identities. The PZ (dotted line) straddles the abaxial-adaxial midplane (green). (**B**) Section through (A). Orthoplanar polarity (OP) (black arrows) runs from the surface toward midplane and axial (dark blue) domains. (**C** and **D**) Fate of (A) if K_{PER} in PZ is high. (**E** and **F**) Same as (C) and (D) but with K_{PER} increasing toward the midvein. (**G** and **H**) Whole-mount immunolocalization of SoPIN1 (green) in barley P1/P2 primordia without (G) or with (H) cell wall signal (CW, magenta). White arrows indicate the SoPIN1 polarity (n = 4). (**I**) Central (blue), lateral (red), and marginal (cyan) domains. PD polarity field (blue arrows) runs from the PZ boundary toward the presumptive midvein tip (*) and apex ("A"). Axes illustrate specified growth rate orientations. (**J** to **L**) Model output at P1 [rear (J) or oblique (K) views] and P2 (L). (**M** and **N**) *ZmCUC2* in situ hybridization in transverse sections of wild-type (M) and *narrowsheath1/2* (N) vegetative maize meristems (n = 4). Dotted line indicates the primordium. (**O**) *ns1/2* domains. (**P** and **Q**) PZ truncation by marginal domain removal (arrowhead). Scale bars, 100 μm. Asterisk indicates presumptive midvein tip.

To determine the orientation of planar polarity, we analyzed an early indicator of epidermal polarity in grasses: the auxin transporter SISTER-OF-PINFORMED1 (SoPIN1) (*23*). Whole-mount immunolocalization of SoPIN1 in barley (*24*) revealed epidermal polarity converging at the primordium midpoint (Fig. 2, G and H). We therefore introduced a proximodistal polarity field (blue arrows in Fig. 2I and fig. S3) pointing from the PZ boundary toward the midpoint. Local growth rates could then be specified in three orientations: (i) parallel to orthoplanar polarity (K_{OP}), (ii) parallel to proximodistal polarity (K_{PD}), and (iii) perpendicular to both (K_{PER}) (Fig. 2I). Low K_{OP} combined with modulation of K_{PD} and K_{PER} (figs. S1 and S2, G and H) generated a sloping ring primordium with a shape and polarity pattern resembling that observed experimentally (Fig. 2, I to L).

To test whether this model could account for *ns1/2*, we first determined PZ extent using the *CUP-SHAPED-COTYLEDON2* (*CUC2*) boundary gene (*25*). In the wild type, *CUC2* expression encircled the meristem, whereas in *ns1/2*, the PZ was truncated by a *CUC2* expression boundary (Fig. 2, M and N). To model the *ns1/2* mutant, we similarly truncated the PZ by removing the marginal domain (Fig. 2O). This removal generated a primordium morphology similar to that observed experimentally at this stage (*20*) (Fig. 2, P and Q).

We next studied the formation of sheath and blade. The sheath margin derives from an overlapping domain, evidenced by clonal sectors that mark both sheath margins, with unmarked regions in between (*3*, *21*) (Fig. 3A, yellow-green-yellow sector). To clarify how overlap arises, we localized *CUC2* expression after primordium emergence. Instead of a continuous ring (Fig. 2M), we observed a diagonal line of *CUC2* expression in the marginal domain, delimiting overlapping PZ ends (Fig. 3B). In *ns1/2*, the PZ had blunt ends delimited by *CUC2* (Fig. 3C). Thus, NS1/2 are needed to extend the PZ and establish overlapping ends.

We incorporated these findings into a model for later developmental stages by considering the primordium as a ring-shaped tissue with overlapping ends (Fig. 3D and fig. S4). Tissue was modeled as a sheet, with K_{PD} and K_{PER} corresponding to planar growth rates and K_{OP} to growth rate in sheet thickness. A clonal sector (Fig. 3D, yellow) was introduced to allow comparison with experimentally observed sectors.

Using growth patterns similar to those above generated a sloping primordium (Fig. 3E and fig. S5 and S6A), and subsequent modulation of K_{PD} and K_{PER} led to a wrapped primordium (Fig. 3F). SHEATH identity was then introduced (Fig. 3G), consistent with the timing of sheath margin emergence (*3*, *4*), and further modulated growth rates. The result was a leaf with typical grass morphology and a yellow-green-yellow sector (Fig. 3H).

As a further test of the model, we removed marginal identity. The result was a more open primordium shape (Fig. 3, I to K), a mature leaf with a narrow sheath and proximal blade (Fig. 3L), and a clonal sector marking a single sheath margin, all features observed experimentally in *ns1/2* mutants (*20*, *21*).

Taken together, our findings suggest two roles for NS1/2 in the marginal domain: (i) extension of the PZ and midplane to encircle the meristem and (ii) growth promotion perpendicular to the midplane to drive primordium emergence and planar growth, which is shaped through differential regulation of K_{PD} and K_{PER}.

To explore the relationship between grass and eudicot leaves, we modified the grass models to produce a eudicot leaf. In the eudicot *Arabidopsis thaliana*, *PRESSED FLOWER* (*PRS*) is the ortholog of maize *NS1/2*. *prs* mutants lack stipules, and *wox1* mutations

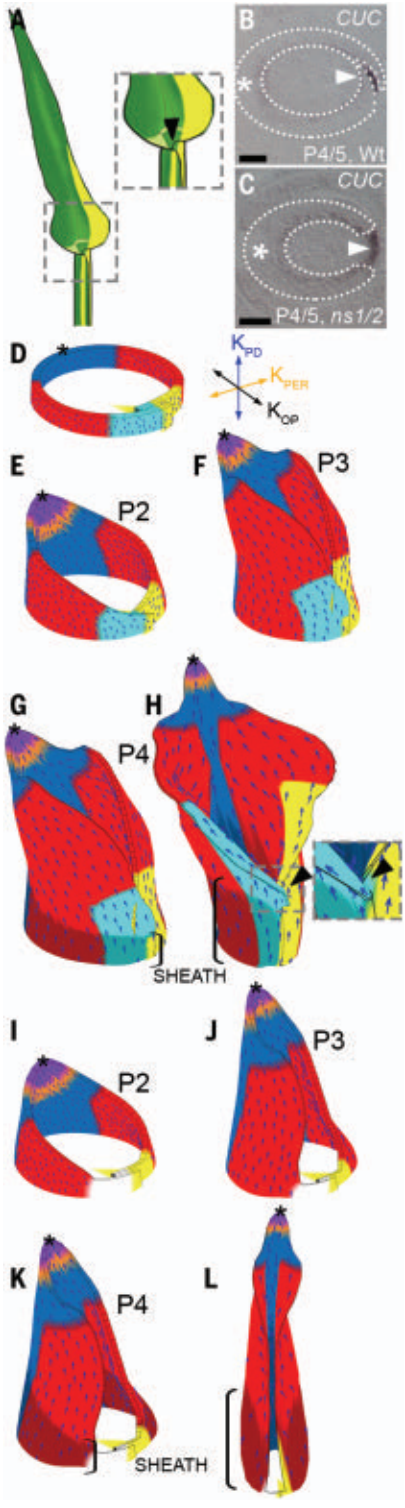


Fig. 3. Grass leaf tissue sheet model. (**A**) A clonal sector (yellow) can mark both margins of the leaf with an intervening unmarked region [green, arrowhead, adapted from (*21*)]. (**B** and **C**) *ZmCUC2* in situ hybridization in transverse sections of wild-type (B) and *narrowsheath1/2* (C) vegetative maize meristems (n = 4). Dotted line indicates P4/5. Arrowhead indicates the sheath margin. Scale bars, 100 μm. (**D** to **L**) Tissue sheet models. (D) the initial ring with overlapping margins, clonal sector (yellow), central (blue), lateral (red), and marginal (cyan) domains. PD polarity (blue arrows) runs from the PZ boundary towards the presumptive midvein tip (*). Axes illustrate specified growth rate orientations. (E) and (F) show model output at P2 and P3. Upper leaf domains are shown in orange and purple. (G) shows introduced SHEATH identity (dark overlay and bracket). (H) is the final output of the emerging leaf with sector marking both margins with intervening unmarked region (arrowhead). In (I) to (L), the marginal domain removal generates a nonwrapping primordium and a leaf with a narrow sheath and proximal blade, with the sector marking one sheath margin. Pesumptive midvein tip (*).

enhance this phenotype to produce narrow leaves (*26*, *27*). Early *prs wox1* primordia are narrower than those of the wild type and do not produce stipules (*28*) (fig. S7). We therefore modeled the *Arabidopsis* leaf primordium by contracting the PZ, assigning stipule identity to the marginal domain, and creating an outer lateral domain (Fig. 4, A to C). The result was a eudicot primordium (Fig. 4D and fig. S2, I and J). Removing the marginal domain gave the *prs* mutant (Fig. 4, E and F), and further removing the outer lateral domain gave *prs wox1* (Fig. 4, G and H).

To determine whether the model could account for mutants that lack ad-abaxial distinctions, we truncated the PZ to the central domain, replaced adaxial with abaxial identity, and replaced the midplane with an axial domain (Fig. 4, I to K, and fig. S2, K and L). This led to a radialized leaf, as observed in abaxialized mutants (*29*) (Fig. 4K). Thus, ad-abaxial genes may normally act to extend an axial domain to a midplane and promote planar growth (*22*, *27*).

To simulate later stages of eudicot leaf development, we first modeled the petiole-sheath hypothesis (Fig. 1K), with SHEATH corresponding to petiole, and BLADE to lamina (Fig. 4, L to O, and fig. S8). We next modeled the petiole-leaf hypothesis (Fig. 1M) by subdividing the primordium domain fated to form the grass leaf tip into two subdomains (Fig. 3, E to H, orange and purple), and inhibiting K_{PAR} proximal to this (Fig. 4, P to S, and fig. S9). In both models, growth was inhibited at the marginal-lateral boundary leading to stipule formation (Fig. 4, O and S, and figs. S6 C to F, S8, and S9).

These modeled hypotheses make different assumptions and predictions. The petiole-leaf hypothesis assumes additional proximal-distal domains and is therefore less parsimonious. The petiole-leaf hypothesis also predicts that petiole mainly derives from the middle of the early primordium (Fig. 4S, orange), whereas the petiole-sheath hypothesis predicts that petiole derives from the primordium base (Fig. 4O). Cell tracking shows that petiole derives from proximal primordium cells with high proximodistal growth rates, supporting the petiole-sheath prediction (*30*, *31*). The petiole-leaf hypothesis predicts that the *prs wox1* mutant has a narrow petiole base (Fig. 4T), whereas the petiole-sheath hypothesis predicts a narrow leaf (Fig. 4U), as was observed experimentally. The petiole-leaf hypothesis further predicts that homologs of petiole identity genes are expressed throughout the grass leaf, except the tip, whereas the petiole-sheath hypothesis predicts sheath-specific expression. Grass homologs of the *Arabidopsis* petiole identity gene *BLADE ON PETIOLE* (*BOP*) are expressed in sheath (Fig. 4V, maize TASSELS REPLACE UPPER EARS1, ZmTRU1) (*32*, *33*), and rice triple knock-out *bop* mutants lack sheath but not blade development (*32*, *33*). Taken together, these findings strongly support the petiole-sheath hypothesis.

We show how a common ground plan of identities may modulate specified growth rates to produce eudicot or grass leaf morphogenesis. In *Arabidopsis*, *WOX* genes act redundantly to extend the PZ and promote planar growth (*27*, *34*, *35*). Redundancy likely varies among eudicot species, because mutants in the *PRS* ortholog of tobacco, which normally lacks stipules, have very narrow leaves (*36*). A key step in grass evolution was the extension of primordium identity and *WOX* activity along the ad-abaxial boundary to encircle the apex. Further modulation of planar growth in the petiole and lamina domains led to grass sheath

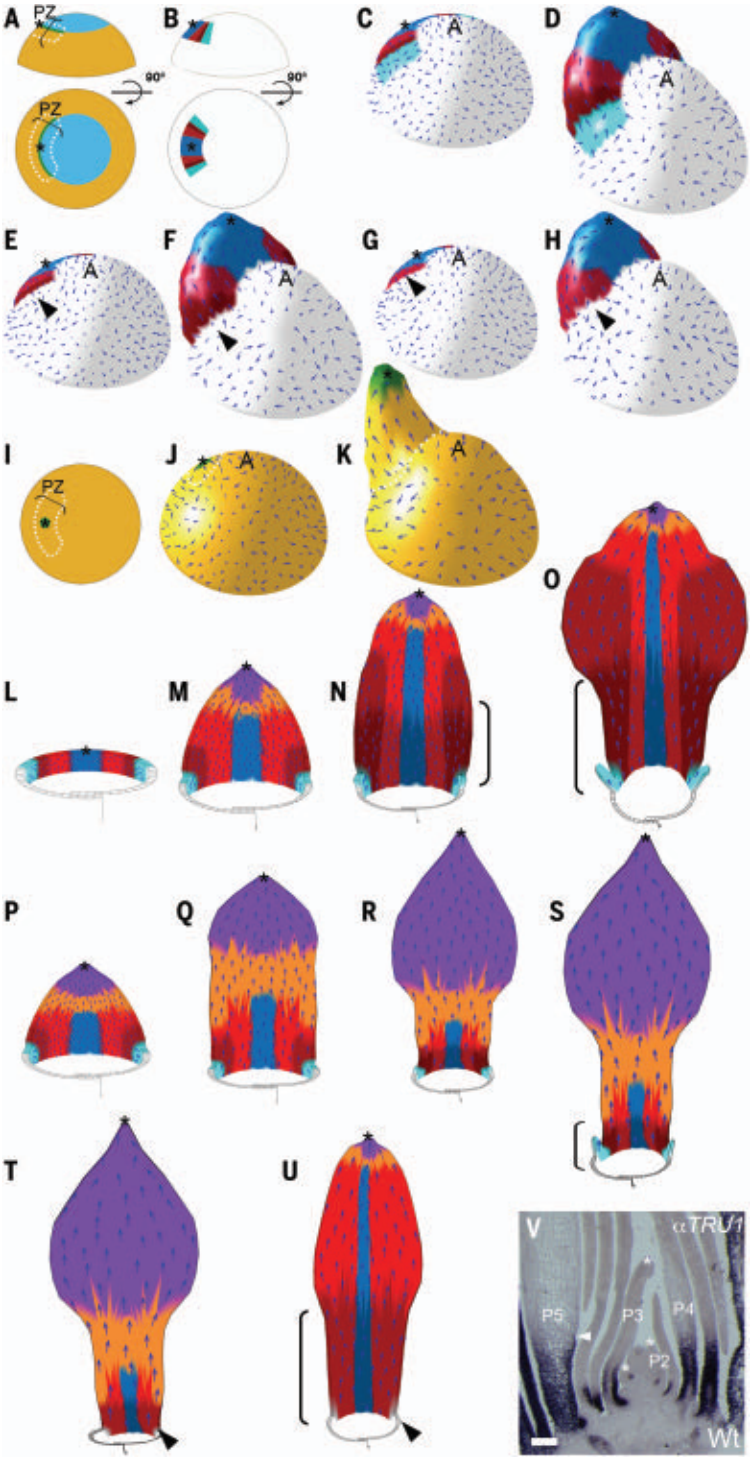


Fig. 4. Eudicot leaf models. (**A**) PZ (dotted line) straddles the midplane (green line) between the abaxial (orange) and adaxial (blue) domains. (**B**) PZ central (blue), lateral (light red), outer lateral (dark red), and marginal (cyan) domains. (**C** to **K**) Volumetric primordium emergence models. (C) and (D) are the wild type, (E) and (F) are *prs*, (G) and (H) are *prs wox1*, and (I) to (K) are the abaxialized mutant. Blue arrows indicate proximodistal polarity, asterisks the presumptive midvein tip, and "A" the apex. Arrowhead indicates the missing domains. (**L** to **U**) Tissue sheet models. Upper leaf domains are shown in orange and purple. (L) to (O) are the petiole-sheath hypothesis. (P) to (S) are the petiole-leaf hypothesis. (N) and (R) SHEATH identity (dark gray overlay, bracket) introduced at P4. In (T) and (U), the *prs wox1* mutant generates a narrow petiole base in the petiole-leaf model (T) and a narrow leaf in the petiole-sheath model (U). Arrowhead indicates the missing domains. (**V**) is ZmTRU1 immunolocalization on a maize vegetative shoot apex longitudinal section (n = 4). Arrowhead indicates the ligule. P1 to P6 are the primordia plastochrons. Scale bar, 100 μm. Asterisk indicates presumptive midvein tip.

and blade morphogenesis, consistent with the 19th-century view of homology (Fig. 1K). Other anatomical traits, such as venation patterns, may represent further elaborations rather than being primary indicators of homology. Our findings are comparable to those from animal evo-devo studies, in which a discarded hypothesis, the notion that the ventral side of insects corresponds to the dorsal sides of vertebrates, was reinstated in the light of fresh developmental genetic evidence (*37*). We further provide a mechanistic link between the developmental genes involved and their morphogenetic effects.

REFERENCES AND NOTES

1. D. R. Kaplan, *Q. Rev. Biol.* **48**, 437–457 (1973).
2. P. F. Stevens, Angiosperm phylogeny website, version 14 (2017); http://www.mobot.org/MOBOT/research/APweb/.
3. R. Johnston, S. Leiboff, M. J. Scanlon, *New Phytol.* **205**, 306–315 (2015).
4. B. C. Sharman, *Ann. Bot. (Lond.)* **6**, 245–282 (1942).
5. A. P. De Candolle, *Organographie Végétale* (Chez Deterville Libraire, 1827), vol. 1, book 2, pp. 285–288.
6. A. M. Grey, "The leaves," in *The Botanical Text-book: An Introduction to Scientific Botany* (Putnam, ed. 3, 1850), pp. 140–179.
7. A. Arber, *Ann. Bot. (Lond.)* **32**, 465–501 (1918).
8. A. R. Arber, *Monocotyledons: A Morphological Study* (Cambridge Univ. Press, 1925).
9. G. Henslow, *Ann. Bot.* **25**, 717–744 (1911).
10. W. Troll, *Chron. Bot.* **5**, 38–41 (1939).
11. F. Knoll, *Österreichische. Bot. Z.* **95**, 163–193 (1948).
12. W. Hagemann, *Bot. Jahrb.* **90**, 297–413 (1970).
13. D. R. Kaplan, *Am. J. Bot.* **57**, 331–361 (1970).
14. A. W. Eichler, *Zur Entwickelungsgeschichte des Blattes mit besonderer Berücksichtigung der Nebenblatt-Bildungen* (Elwert'sche Universitäts-Buchhandlung, 1861).
15. D. Jackson, B. Veit, S. Hake, *Development* **120**, 405–413 (1994).
16. M. T. Juarez, R. W. Twigg, M. C. P. Timmermans, *Development* **131**, 4533–4544 (2004).
17. F. T. S. Nogueira *et al.*, *PLOS Genet.* **5**, e1000320 (2009).
18. A. Hay, S. Hake, *Plant Physiol.* **135**, 300–308 (2004).
19. J. Nardmann, J. Ji, W. Werr, M. J. Scanlon, *Development* **131**, 2827–2839 (2004).
20. M. J. Scanlon, R. G. Schneeberger, M. Freeling, *Development* **122**, 1683–1691 (1996).
21. M. J. Scanlon, M. Freeling, *Dev. Biol.* **182**, 52–66 (1997).
22. C. D. Whitewoods *et al.*, *Science* **367**, 91–96 (2020).
23. D. L. O'Connor *et al.*, *PLOS Comput. Biol.* **10**, e1003447 (2014).
24. A. Richardson, A. B. Rebocho, E. Coen, *Plant Cell* **28**, 2079–2096 (2016).
25. M. Aida, T. Ishida, M. Tasaka, *Development* **126**, 1563–1570 (1999).
26. N. Matsumoto, K. Okada, *Genes Dev.* **15**, 3355–3364 (2001).
27. M. Nakata *et al.*, *Plant Cell* **24**, 519–535 (2012).
28. Z. Zhang *et al.*, *Curr. Biol.* **30**, 4857–4868.e6 (2020).
29. M. E. Byrne, *Curr. Opin. Plant Biol.* **8**, 59–66 (2005).
30. S. Fox *et al.*, *PLOS Biol.* **16**, e2005952 (2018).
31. D. Kierzkowski *et al.*, *Cell* **177**, 1405–1418.e17 (2019).
32. Z. Dong *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8656–E8664 (2017).
33. T. Toriba *et al.*, *Nat. Commun.* **10**, 619 (2019).
34. M. Vandenbussche *et al.*, *Plant Cell* **21**, 2269–2283 (2009).
35. H. Lin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 366–371 (2013).
36. J. I. Pueyo, J. P. Couso, *Curr. Opin. Genet. Dev.* **15**, 439–446 (2005).
37. A. Richardson *et al.*, Computational model codes for: Evolution of the grass leaf by primordium extension and petiole-lamina remodeling, Zenodo (2021); https://doi.org/10.5281/zenodo.5555063.

ACKNOWLEDGMENTS

We thank D. O'Connor, members of the Coen laboratory, C. Whitewoods, members of the Hake laboratory, S. Leiboff, C. L. Shaw, and A. Hudson for support and helpful discussions; T. Dao for help with the PI staining and tissue clearing method; and the microscopy and greenhouse staff at the JIC and the USDA and the Cornell Institute of Biotechnology's Imaging Facility. **Funding:** This work was supported by a bilateral NSF/BIO-BBSRC grant (BB/M023117/1 to S.H. and E.C.); the BBSRC (grants BBS/E/J/000PR9787, BBS/E/J/00000152, BB/L008920/1, and BBS/E/J/000PR9789 to E.C.); and the NSF (grant DEB-1457070 to M.S.). H.K. received funding from the Chinese Academy of Sciences (grants XDB27010304 and ZDBS-LY-SM022) for J.C. A.R. is funded by University of Edinburgh start-up funds. The Cornell

Institute of Biotechnology's Imaging Facility is supported by NYSTEM (C29155) and the NIH (S10OD018516) **Author contributions:** A.E.R., E.C., M.S., and S.H. wrote the manuscript. A.E.R. performed immunolocalizations and in situ hybridizations. A.E.R., J.C., and E.C. developed the models. R.J. performed in situ hybridizations. R.K. helped with the modeling. B.C. performed confocal imaging of the *Arabidopsis* leaf primordia. A.E.R., A.B.R., R.J., M.S., S.H., and E.C. developed the project. S.H., M.S., E.C., A.R., and H.K. provided funding. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Microscopy data are available through the Edinburgh University DataShare (https://datashare.ed.ac.uk/handle/10283/3987) in The Plant Shape Lab collection. All materials are available on request from the corresponding authors. The growing polarized tissue framework (GFtbox) software is freely available at: https://coensoft.jic.ac.uk. The computational model codes are available through GitHub (*37*) and https://coensoft.jic.ac.uk. All other data provided are in the main paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf9407
Materials and Methods
Supplementary Text: Model Descriptions
Figs. S1 to S10
Tables S1 to S4
References (*38*–*44*)
Movies S1 to S7
MDAR Reproducibility Checklist

1 December 2020; accepted 8 October 2021
10.1126/science.abf9407

SUPERCONDUCTIVITY

Discovery of segmented Fermi surface induced by Cooper pair momentum

Zhen Zhu[1]†, Michał Papaj[2]†, Xiao-Ang Nie[1], Hao-Ke Xu[1], Yi-Sheng Gu[1], Xu Yang[1], Dandan Guan[1], Shiyong Wang[1], Yaoyi Li[1], Canhua Liu[1], Jianlin Luo[3], Zhu-An Xu[4], Hao Zheng[1]*, Liang Fu[2]*, Jin-Feng Jia[1]*

A sufficiently large supercurrent can close the energy gap in a superconductor and create gapless quasiparticles through the Doppler shift of quasiparticle energy caused by finite Cooper pair momentum. In this gapless superconducting state, zero-energy quasiparticles reside on a segment of the normal-state Fermi surface, whereas the remaining Fermi surface is still gapped. We use quasiparticle interference to image the field-controlled Fermi surface of bismuth telluride (Bi_2Te_3) thin films under proximity effect from the superconductor niobium diselenide ($NbSe_2$). A small applied in-plane magnetic field induces a screening supercurrent, which leads to finite-momentum pairing on the topological surface states of Bi_2Te_3. We identify distinct interference patterns that indicate a gapless superconducting state with a segmented Fermi surface. Our results reveal the strong impact of finite Cooper pair momentum on the quasiparticle spectrum.

In the presence of supercurrent flow, Cooper pairs in a superconductor acquire finite momentum (*1*, *2*). This Cooper pair momentum $\mathbf{q}$ results in a Doppler shift of the energy of the Bogoliubov quasiparticle excitations (Fig. 1A)

$$E(\mathbf{k}) = \sqrt{(\epsilon_{\mathbf{k}} - \mu)^2 + \Delta^2} + \frac{1}{2} v_{\mathbf{k}} \cdot \mathbf{q} \qquad (1)$$

which is characterized by the Doppler shift energy scale $E_D \approx v_{\mathbf{k}} q/2$, where $v_{\mathbf{k}} = \partial \epsilon_{\mathbf{k}} / \partial \mathbf{k}$ is the velocity at momentum $\mathbf{k}$, $\epsilon_{\mathbf{k}}$ is the normal-state dispersion, μ is the chemical potential, and Δ is the superconducting gap in the absence of supercurrent. This Doppler shift changes the energy of the quasiparticles that move parallel or antiparallel to the supercurrent but does not affect the quasiparticles that move in the perpendicular direction, resulting in an anisotropic quasiparticle energy dispersion.

When the Doppler energy shift E_D exceeds Δ, the superconducting gap closes, and a segmented Fermi surface appears with zero-energy quasiparticles that are a superposition of electrons and holes (*2*–*4*). A schematic example of this behavior for a system with a circular Fermi surface in the normal state is shown in Fig. 1B. The supercurrent decreases the energy of quasiparticles moving parallel to it. With sufficiently large current, zero-energy quasiparticles first appear on a segment of the normal-state Fermi surface in the direction of the current. As the current increases, the momentum-space contour of these zero-energy quasiparticles grows, forming a banana-like shape and contributing to finite density of states (DOS) at $E = 0$. The solid and dashed lines in Fig. 1B indicate whether the excitation is more electron- or hole-like, respectively, in a given part of the contour. At the same time, the rest of the normal-state Fermi surface remains fully gapped because the Doppler shift energy has not overcome the superconducting gap there. Although the existence of supercurrent-induced gapless quasiparticles, known as the Volovik effect (*5*), has been established by tunneling (*6*, *7*) and specific heat measurements (*8*, *9*) on d- and s-wave superconductors, the Fermi surface of a current-carrying superconductor has not been observed directly. Detection of the segmented Fermi surface of Bogoliubov quasiparticles requires spectroscopic techniques with both energy and momentum resolution. Furthermore, the supercurrent that is necessary to close the superconducting gap is often larger than the depairing current at which the superconductor switches to a resistive state (*6*).

In this work, we overcome the aforementioned difficulty with an alternative approach and material platform. Instead of passing a transport current directly through a superconductor, we apply a small external magnetic field $\mathbf{B}_{ext}$ to produce a screening current near the surface by means of the Meissner effect. Then, the superconducting order parameter on the surface develops a spatially varying phase $\Delta(\mathbf{r}) = \Delta \exp(i\mathbf{q} \cdot \mathbf{r})$ at position $\mathbf{r}$, with $\mathbf{q}$ in the direction of the screening current that is perpendicular to the field $\mathbf{B}_{ext}$. Our platform of choice (Fig. 1C) consists of a thin film of Bi_2Te_3, the quintessential topological insulator (*10*, *11*), grown by molecular beam epitaxy on top of bulk crystal $NbSe_2$, an s-wave superconductor. The superconducting proximity effect from $NbSe_2$ induces a hard gap in Bi_2Te_3 at zero field, as demonstrated in previous scanning tunneling microscopy (STM) measurements (*12*). This material combination provides ideal synergy for creating and detecting the gapless superconducting state. $NbSe_2$ is a clean s-wave superconductor with a large superconducting gap and long London penetration depth λ_L that exceeds 100 nm (*13*). Notably, this means that a small in-plane magnetic field can create a large screening current and Cooper pair momentum on the surface, without introducing vortices in the measurement region.

Our platform is ideal to observe formation of the segmented Fermi surface, owing to a combination of two factors. A Bi_2Te_3 film under the proximity effect has a simple Fermi pocket surrounding the Γ point (*14*) that originates from the topological surface states. The proximity-induced gap in the top layer of Bi_2Te_3 is reduced compared with the bulk gap of $NbSe_2$, and the topological surface state has a substantially higher Fermi velocity (~2.6 eV·Å) (*14*) than the states of the parent

[1]School of Physics and Astronomy, Key Laboratory of Artificial Structures and Quantum Control (Ministry of Education), Shenyang National Laboratory for Materials Science, Tsung-Dao Lee Institute, Shanghai Jiao Tong University, Shanghai 200240, China. [2]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China. [4]Department of Physics, Zhejiang University, Hangzhou 310027, Zhejiang, China.
***Corresponding author. Email: haozheng1@sjtu.edu.cn (H.Z.); liangfu@mit.edu (L.F.); jfjia@sjtu.edu.cn (J.-F.J.)**
†These authors contributed equally to this work.

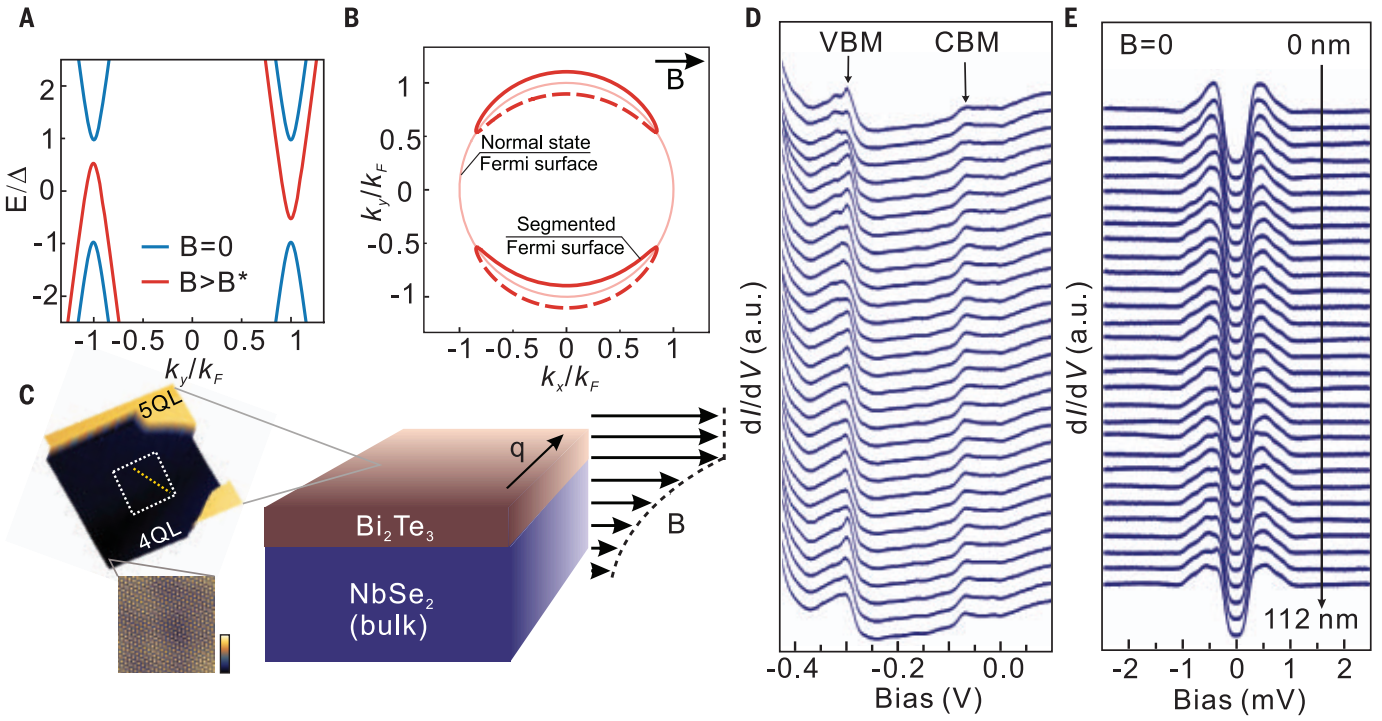


Fig. 1. Topography and characterization of proximitized Bi_2Te_3 thin films. (**A**) Quasiparticle dispersion around the superconducting gap without magnetic field (blue curves) and with magnetic field above B^* (red curves) that induces the gapless superconducting state (for parabolic dispersion $B^* \sim \Delta/ev\lambda_L$, where e is the electron charge and v is the Fermi velocity). (**B**) Schematic depiction of a segmented Fermi surface that arises as a result of the magnetic field B along the x axis. Solid red lines indicate electron-dominated states, whereas dashed red lines denote hole-dominated states. The pink line indicates the Fermi surface of the normal state. (**C**) Structure of the samples with four or five quintuple layers (QLs) of Bi_2Te_3 on top of $NbSe_2$. An external in-plane magnetic field $\mathbf{B}_{ext}$ induces a perpendicular screening supercurrent and Cooper pair momentum **q**. Insets show the topography of a large area (320 nm by 320 nm) of the thin film, with regions of varying thickness (top inset) and an atomic-resolution image (10 nm by 10 nm) of the Bi_2Te_3 lattice (bottom inset). The dotted white square indicates the area where the dI/dV maps in Fig. 3 are measured. (**D**) Differential conductance dI/dV spectra taken across the yellow line cut in (C) for a wide range of energies. The set point is at 0.1 V and 0.5 nA. CBM and VBM denote the conduction band minimum and valence band maximum, respectively, of the Bi_2Te_3 bulk bands. (**E**) The proximity-induced superconducting gap along the yellow line cut from (C). The set point is at 2.5 mV and 0.5 nA. From the position of the coherence peaks, we determine a superconducting gap $\Delta \approx 0.5$ meV. All STM measurements were conducted at a temperature of 40 mK. a.u., arbitrary units.

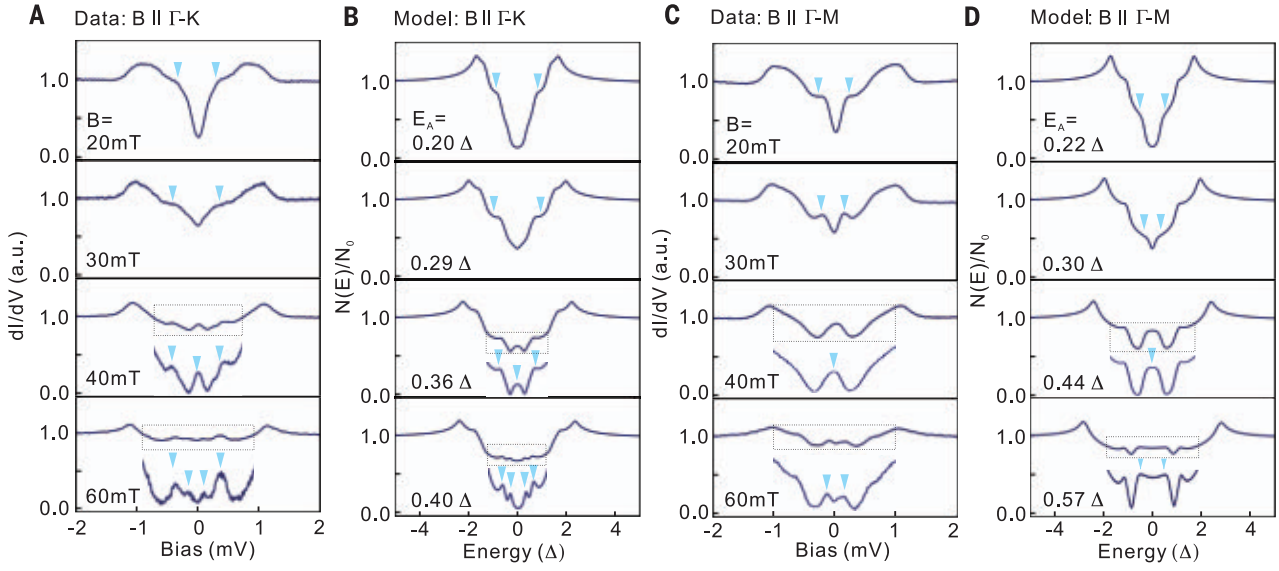


Fig. 2. DOS on the $Bi_2Te_3/NbSe_2$ surface under in-plane magnetic field. (**A**) Differential conductance dI/dV spectra and (**B**) theoretical DOS curves for increasing magnetic field along the ΓK direction. (**C**) Differential conductance dI/dV spectra and (**D**) theoretical DOS curves for increasing magnetic field along the ΓM direction. For both (A) and (C), the spectra are acquired at a set point of 2.5 mV and 0.5 nA and a temperature of 40 mK. The characteristic energy scale for theoretical calculations is $E_A = evA_{x,y}$, where $A_{x,y}$ is the magnetic vector potential (*22*). Arrowheads indicate in-gap features attributed to pockets of segmented Fermi surface, as depicted in movies S1 and S2. Dotted rectangles indicate regions magnified in the insets.

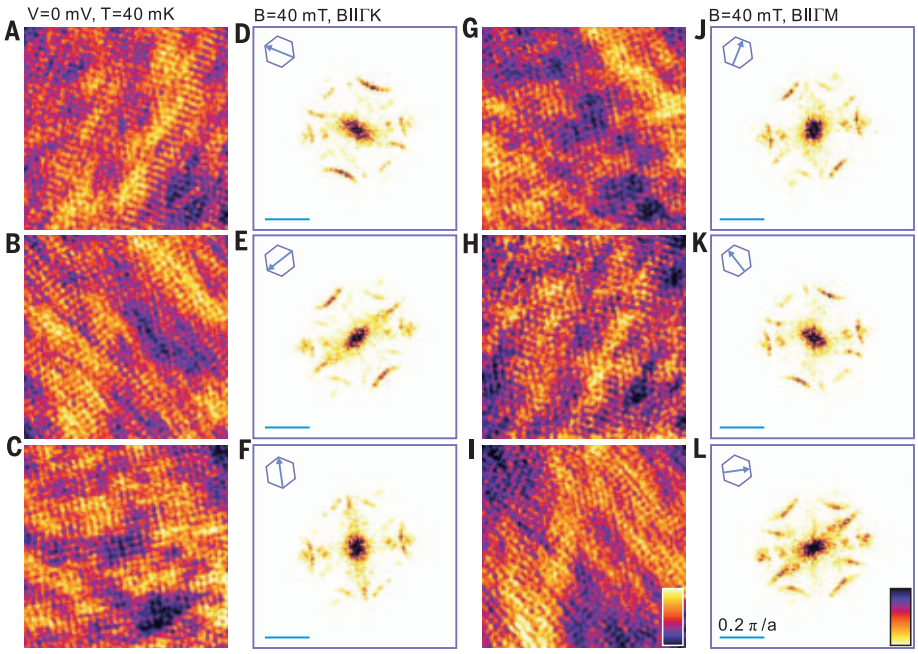


Fig. 3. Real- and momentum-space QPI patterns for six different orientations of the in-plane magnetic field. (**A** to **C**) *dI*/*dV* maps depicting the real-space QPI patterns with magnetic field oriented along the ΓK directions. (**D** to **F**) Momentum-space images, in which Fourier transforms of (A) to (C) each display two bright segments. Arrows represent the magnetic field along the ΓK directions. (**G** to **I**) As in (A) to (C), but with magnetic field oriented along the ΓM directions. (**J** to **L**) Fourier transforms of (G) to (I), which display four bright segments. All data were collected within the same 120-nm–by–120-nm area, at a bias voltage of 0 mV and temperature of 40 mK and under an in-plane magnetic field of 40 mT. a, crystal lattice constant.

superconductor (0.65 to 1.95 eV·Å, depending on the pocket and crystalline axis) (*15*). Therefore, although both $NbSe_2$ and the proximitized surface state of Bi_2Te_3 share the same Cooper pair momentum q, the substantial differences in Fermi velocities and superconducting gaps make it possible to choose a small magnetic field that closes the gap of the proximitized surface state but not that of the parent superconductor $NbSe_2$. The Bi_2Te_3/$NbSe_2$ heterostructure has previously been investigated for the presence of Majorana zero modes in vortex cores (*12*, *16*–*19*). A recent theory predicts that topological superconductivity hosting Majorana end states may be formed when parallel field-induced quasiparticles in proximitized topological insulators are confined into a quasi–one dimensional (1D) channel (*4*).

In the insets of Fig. 1C, we present the topography of our thin film with regions of varying thickness (top inset) and an atomic-resolution image of the Bi_2Te_3 lattice (bottom inset), demonstrating its high quality. The properties of Bi_2Te_3 films are appreciably affected by the number of quintuple layers (QLs), the 1-nm-thick basic building blocks of this crystal. Topological surface states are formed for thicknesses above three QLs (*20*); previous results show that a proximity-induced superconducting energy gap is present in the top surface for films up to 11 QLs thick (*12*). To optimize the superconductivity in the topological surface states, we perform all of our measurements on a four-QL area of the sample, denoted in Fig. 1C by the dotted white square of area 120 nm by 120 nm, away from the step edges. Figure 1, D and E, shows differential conductance dI/dV curves (I, current; V, voltage) along the line cut inside this region. These curves display a high degree of spatial uniformity over a wide range of energy scales from 0.1 to −0.43 eV (Fig. 1D). We identify the valence band maximum (VBM) and conduction band minimum (CBM) from dI/dV peaks at −0.3 and −0.07 eV, respectively. This allows us to infer the Fermi-level position at ~360 meV above the surface Dirac point (*14*). Near the Fermi level, we observe a hard, U-shaped superconducting energy gap $\Delta \approx 0.5$ meV at zero magnetic field, with no visible in-gap features across the line cut.

We now apply an in-plane magnetic field to the thin film and measure the differential conductance (dI/dV) to investigate the gapless superconducting state. Considering the strong hexagonal warping effect of Bi_2Te_3 surface states (*21*), we orient the magnetic field along two different high-symmetry directions ΓK and ΓM. The dI/dV spectra (Fig. 2) reveal a rich set of in-gap features, indicated by arrowheads. As the magnetic field is increased in small steps of 10 mT, multiple distinct peaks and shoulders appear and change rapidly. The in-gap spectrum also depends on the field direction. This behavior is in contrast to that of the surface of pristine $NbSe_2$ (fig. S7) (*22*), where the in-gap spectrum is featureless under the same magnetic field, displaying a hard gap with only minimal changes to the coherence peaks. Moreover, these observations contrast sharply with the tunneling spectra of conventional superconductors such as aluminum or lead, for which the magnetic field also causes the filling of the superconducting gap in a featureless manner (*7*), as do magnetic impurities (*23*). To understand the microscopic origin of the observed tunneling spectra, we perform theoretical calculations of the DOS at various field-induced Cooper pair momenta. The distinctive in-gap features and their evolution with the field in both directions are reproduced by our calculation that is based on the established model Hamiltonian for Bi_2Te_3 surface states including hexagonal warping (*22*). We can rule out the Zeeman effect as the primary origin of these effects because in order for the Zeeman energy to close the superconducting gap $g\mu_B B_{ext}/2 = \Delta \sim 0.5$ meV, where μ_B is the Bohr magneton) at 20 mT, the effective g-factor would have to be ~800—much higher than the values expected in various topological materials (*24*–*27*). On the other hand, the Doppler shift energy due to the screening current at 20 mT, estimated from the Fermi velocity and London penetration depth of $NbSe_2$, is close to the measured gap on the proximitized surface of Bi_2Te_3 at zero field (*22*). The observed in-gap features are therefore a consequence of supercurrent-induced quasiparticles. The evolution of spectral function at energies around the Fermi level with the field is calculated and depicted in movies S1 and S2.

To detect the segmented Fermi surface directly within momentum space, we scan the constant energy local DOS over the whole region of interest and perform a Fourier transform to obtain the quasiparticle interference (QPI) patterns (*28*, *29*). At energies far outside the superconducting gap, we observe strong and equal intensities at six segments, which are symmetrically placed along three equivalent ΓM directions (fig. S5E). This QPI pattern is independent of the direction or magnitude of magnetic field (fig. S6, E and J) and is similar to those observed in Bi_2Te_3 without superconductivity (*30*–*32*).

However, the QPI pattern becomes markedly different at the energies inside the superconducting gap. Pairs of real- and momentum-space images at zero energy are presented in Fig. 3 for six different orientations of magnetic field along high-symmetry directions at B_{ext} = 40 mT. In the real-space images, we observe 1D standing wave patterns, whose orientation

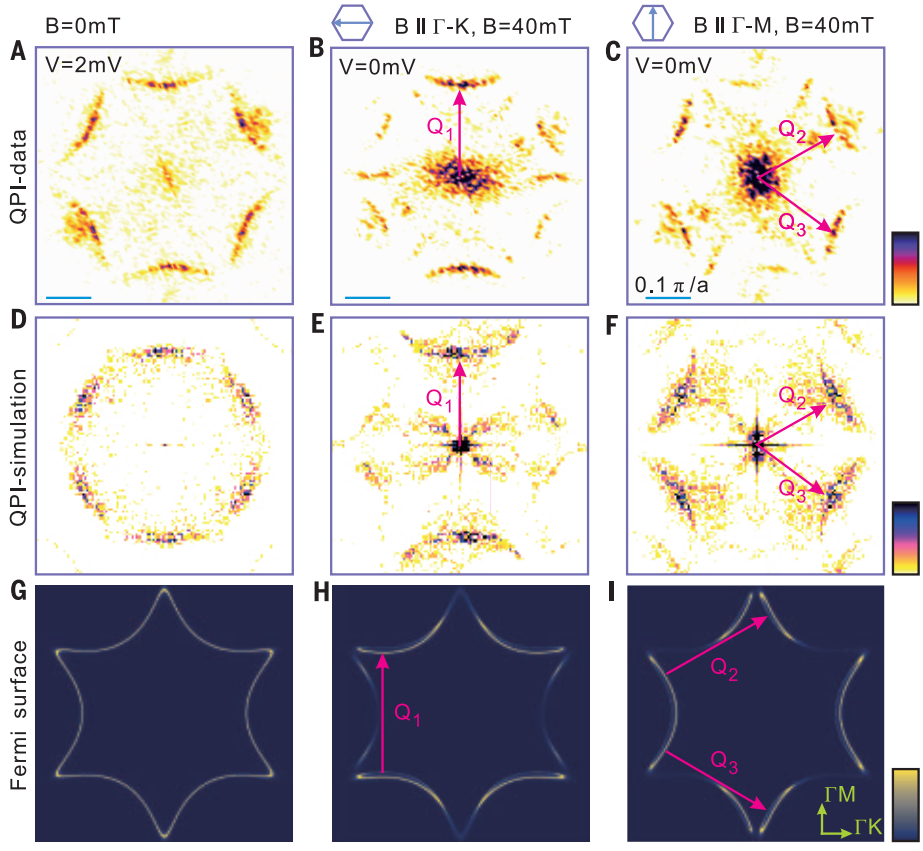


Fig. 4. QPI patterns of the segmented Fermi surface. (**A**) QPI at zero magnetic field outside of the superconducting gap. The six bright segments correspond to scattering between the tips of the normal-state Fermi surface. (**B**) QPI at B = 40 mT along ΓK at V = 0 mV. The two bright segments correspond to vertical scattering between the tips of the "star" indicated in (H). (**C**) QPI at B = 40 mT along ΓM at V = 0 mV. The four bright segments correspond to diagonal scattering between the tips of the "star" indicated in (I). (**D** to **F**) Numerical simulations (*22*) for QPI in a hexagonally warped disordered Dirac surface state corresponding to the magnetic fields as in (A) to (C). (**G** to **I**) Spectral function displaying Fermi surface contours corresponding to the magnetic fields as in (A) to (C). $\mathbf{Q}_1$, $\mathbf{Q}_2$, and $\mathbf{Q}_3$ indicate momentum transfer vectors.

changes with the magnetic field direction. These real-space patterns yield two distinct classes of Fourier images, respectively showing two and four of the six original high-intensity segments. When the field is applied along ΓK, two bright segments are found along the perpendicular ΓM direction, whereas the remaining four are strongly suppressed. When the field is directed along ΓM, the corresponding two segments are dark, whereas the other four segments are bright.

These QPI patterns can be understood directly from the picture of a segmented Fermi surface in the superconducting state under an in-plane magnetic field. Because of the direction-dependent pair-breaking effect, only a portion of the normal-state Fermi surface becomes gapless, and thus only the hotspots located in this gapless segment are activated for scattering at zero energy. To illustrate this, in Fig. 4, G to I, we present the spectral function of the normal state, the superconducting state with magnetic field along ΓK, and the superconducting state with magnetic field along ΓM, respectively. In the normal state, we observe a hexagonally warped contour, which gives rise to six symmetric segments in the QPI pattern (Fig. 4A) associated with scattering between the hotspots at the neighboring tips of the star. In the superconducting state at zero field, there is no Fermi surface, owing to the hard gap. Nevertheless, it reemerges at B_{ext} = 40 mT, as a result of the gap being filled with quasiparticle states. However, this new Fermi surface consists of only segments of the normal-state Fermi surface, whose size and location are controlled by the field strength and orientation (Fig. 4, H and I). Scattering between the available hotspots on the segmented Fermi surface gives rise to those bright segments in the observed QPI pattern (Fig. 4, B and C), as represented by the indicated momentum transfer vectors $\mathbf{Q}_i$. By contrast, the gapped hotspots cannot participate in quasiparticle scattering processes, leading to suppression of QPI intensity at corresponding wave vectors.

Notably, scattering between Bogoliubov quasiparticles on the segmented Fermi surface is strongly dependent on the superconducting coherence factors. Zero-energy quasiparticles that originate from the $E > 0$ and $E < 0$ branches at zero field are the symmetric and antisymmetric superpositions of electron and hole states, respectively. In the presence of nonmagnetic impurities, scattering between the two opposite branches is constructively enhanced, whereas scattering within each branch is destructively suppressed. For this reason, even though the wave vector $\mathbf{Q}_1$ connects the same pair of hotspots in both Fig. 4H and Fig. 4I, quasiparticle scattering at this wave vector is present in the former case but suppressed by coherence factors in the latter. This can be clearly seen in Fig. 4, B and C, indicating that the gapless excitations are Bogoliubov quasiparticles rather than normal electrons. To further substantiate the above theoretical analysis, we perform a full numerical simulation of the proximitized topological surface state under an in-plane magnetic field. By using recursive Green's functions (*22*), we calculate the local DOS in the presence of random disorder and construct its Fourier image. Our numerical QPI patterns (Fig. 4, D to F) show good agreement with the experimental data.

Our results reveal the strong impact of Cooper pair momentum caused by screening supercurrent on the quasiparticle energy dispersion. The observation of the segmented Fermi surface of Bogoliubov quasiparticles paves the way for further STM study of pair density wave and Fulde-Ferrell-Larkin-Ovchinnikov states in unconventional superconductors (*33*–*36*).

REFERENCES AND NOTES

1. P. Fulde, *Phys. Rev.* **137**, A783–A787 (1965).
2. P. Fulde, in *Tunneling Phenomena in Solids: Lectures Presented at the 1967/NATO Advanced Study Institute at Risö, Denmark*, E. Burstein, S. Lundqvist, Eds. (Springer, 1969), pp. 427–442; https://doi.org/10.1007/978-1-4684-1752-4_29.
3. N. F. Q. Yuan, L. Fu, *Phys. Rev. B* **97**, 115139 (2018).
4. M. Papaj, L. Fu, *Nat. Commun.* **12**, 577 (2021).
5. G. Volovik, *JETP Lett.* **58**, 469–473 (1993).
6. A. Anthore, H. Pothier, D. Esteve, *Phys. Rev. Lett.* **90**, 127001 (2003).
7. J. L. Levine, *Phys. Rev.* **155**, 373–378 (1967).
8. Y. Wang *et al.*, *Phys. Rev. B* **84**, 184524 (2011).
9. Y. Bang, *Phys. Rev. Lett.* **104**, 217001 (2010).
10. H. Zhang *et al.*, *Nat. Phys.* **5**, 438–442 (2009).
11. Y. Xia *et al.*, *Nat. Phys.* **5**, 398–402 (2009).
12. J.-P. Xu *et al.*, *Phys. Rev. Lett.* **112**, 217001 (2014).
13. J. D. Fletcher *et al.*, *Phys. Rev. Lett.* **98**, 057003 (2007).
14. Y. L. Chen *et al.*, *Science* **325**, 178–181 (2009).
15. T. Kiss *et al.*, *Nat. Phys.* **3**, 720–725 (2007).
16. A. D. K. Finck, C. Kurter, Y. S. Hor, D. J. Van Harlingen, *Phys. Rev. X* **4**, 041022 (2014).
17. A. Q. Chen *et al.*, *Nat. Commun.* **9**, 3478 (2018).
18. M. Bai *et al.*, *Phys. Rev. Mater.* **4**, 094801 (2020).
19. B. Jäck *et al.*, *Science* **364**, 1255–1259 (2019).
20. Y.-Y. Li *et al.*, *Adv. Mater.* **22**, 4002–4007 (2010).
21. L. Fu, *Phys. Rev. Lett.* **103**, 266801 (2009).
22. Additional experimental data and details of numerical simulations are available in the supplementary materials.

23. M. A. Woolf, F. Reif, *Phys. Rev.* **137**, A557–A564 (1965).
24. A. Wolos *et al.*, *Phys. Rev. B* **93**, 155114 (2016).
25. R. Y. Chen *et al.*, *Phys. Rev. Lett.* **115**, 176404 (2015).
26. S. Jeon *et al.*, *Nat. Mater.* **13**, 851–856 (2014).
27. Y.-S. Fu *et al.*, *Nat. Commun.* **7**, 10829 (2016).
28. Q.-H. Wang, D.-H. Lee, *Phys. Rev. B* **67**, 020511 (2003).
29. J. E. Hoffman *et al.*, *Science* **297**, 1148–1151 (2002).
30. W.-C. Lee, C. Wu, D. P. Arovas, S.-C. Zhang, *Phys. Rev. B* **80**, 245439 (2009).
31. H. Beidenkopf *et al.*, *Nat. Phys.* **7**, 939–943 (2011).
32. T. Zhang *et al.*, *Phys. Rev. Lett.* **103**, 266803 (2009).
33. P. Choubey *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 14805–14811 (2020).
34. Z. Du *et al.*, *Nature* **580**, 65–70 (2020).
35. P. M. R. Brydon, D. F. Agterberg, H. Menke, C. Timm, *Phys. Rev. B* **98**, 224509 (2018).
36. N. F. Q. Yuan, L. Fu, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2019063118 (2021).
37. Z. Zhu *et al.*, Replication data and theory code for: Discovery of segmented Fermi surface induced by Cooper pair momentum. *Zenodo* (2021); https://doi.org/10.5281/zenodo.5341932.

ACKNOWLEDGMENTS

We thank P. A. Lee and C.-K. Chiu for the helpful discussions and comments on the manuscript. **Funding:** The work at Shanghai Jiao Tong University was supported by the Ministry of Science and Technology of China (grant 2019YFA0308600) and NSFC (grants 11790313, 11521404, and 92065201). The work at Massachusetts Institute of Technology was supported by DOE Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, under awards DE-SC0018945 (theoretical analysis) and DE-SC0019275 (numerical simulation). L.F. is partly supported by the David and Lucile Packard Foundation. We also thank NSFC (grants 12104292, 11634009, 11874256, 11874258, 12074247, and 11861161003), the Ministry of Science and Technology of China (grant 2020YFA0309000), the Strategic Priority Research Program of Chinese Academy of Sciences (grant XDB28000000), the Science and Technology Commission of Shanghai Municipality (grants 2019SHZDZX01, 19JC1412701, and 20QA1405100), and the China Postdoctoral Science Foundation (grant BX2021184) for partial support. **Author contributions:** L.F. and J.-F.J. initiated the research. H.Z., J.-F.J., and L.F. supervised the research. Z.Z. conducted the experiment, with the help of X.-A.N., H.-K.X., Y.-S.G., X.Y., D.G., S.W., Y.L., and C.L. J.L. and Z.-A.X. prepared the $NbSe_2$ samples. M.P. and L.F. performed the theoretical analysis. M.P. and L.F. wrote the manuscript, with input from all authors. All authors discussed the results, interpretation, and conclusions. **Competing interests:** The authors declare no competing financial or nonfinancial interests. **Data and materials availability:** All experimental data and codes for theoretical analysis from this study are available at Zenodo (*37*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf1077
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*38*–*40*)
Movies S1 and S2

4 October 2020; accepted 15 October 2021
Published online 28 October 2021
10.1126/science.abf1077

QUANTUM CHEMISTRY

Pushing the frontiers of density functionals by solving the fractional electron problem

James Kirkpatrick[1]*†, Brendan McMorrow[1]†, David H. P. Turban[1]†, Alexander L. Gaunt[1]†, James S. Spencer[1], Alexander G. D. G. Matthews[1], Annette Obika[1], Louis Thiry[2], Meire Fortunato[1], David Pfau[1], Lara Román Castellanos[1], Stig Petersen[1], Alexander W. R. Nelson[1], Pushmeet Kohli[1], Paula Mori-Sánchez[3], Demis Hassabis[1], Aron J. Cohen[1,4]*

Density functional theory describes matter at the quantum level, but all popular approximations suffer from systematic errors that arise from the violation of mathematical properties of the exact functional. We overcame this fundamental limitation by training a neural network on molecular data and on fictitious systems with fractional charge and spin. The resulting functional, DM21 (DeepMind 21), correctly describes typical examples of artificial charge delocalization and strong correlation and performs better than traditional functionals on thorough benchmarks for main-group atoms and molecules. DM21 accurately models complex systems such as hydrogen chains, charged DNA base pairs, and diradical transition states. More crucially for the field, because our methodology relies on data and constraints, which are continually improving, it represents a viable pathway toward the exact universal functional.

Computing electronic energies underpins theoretical chemistry and materials science, and density functional theory (DFT) (*1*, *2*) promises an exact and efficient approach. However, there is a conundrum at the heart of DFT: The exact functional—mapping electron density to energy—is proven to exist, but little practical guidance is given on its explicit form. Approximations to the exact functional have enabled numerous investigations of matter at a microscopic level and rank as some of the most impactful works in the whole of science (*3*). Nevertheless, despite their design and success, pathological errors persist in these approximations, and it has been known for over a decade (*4*) that the root cause of many of these errors is the violation of exact conditions for systems with fractional electrons. In this work, we used deep learning to train a functional that respects these conditions and thus has excellent performance across main-group chemistry.

Since the early days of DFT, it has been clear that approximations improve when they satisfy more of the mathematical constraints of the exact functional and fit more systems. Seventeen known exact constraints (but not the fractional constraints) are satisfied by the strongly constrained and appropriately normed (SCAN) functional (*5*), which achieves unprecedented accuracy and predictiveness for bonded systems among the functionals that are not fitted to any bonded system. Recent work has also begun to address the fractional constraints, of particular interest being a localized correction on the orbitals (*6*, *7*). In parallel, machine learning has emerged as a powerful tool at the level of molecular modeling in chemistry (*8*, *9*) and has been recently applied to functional development (*10*, *11*). Proof-of-principle studies have shown that neural networks (*12*–*16*) can be trained on molecular data, but to date, they are not competitive in accuracy with traditional hand-designed functionals.

In this work, we present a functional, DM21 (DeepMind 21) that is state of the art on thorough benchmark evaluation and has qualitatively improved properties because it obeys two classes of constraints on systems with fractional electrons. The types of fractional constraints considered were fractional charge (FC) systems, with a noninteger total charge, and fractional spin (FS) systems, with noninteger spin magnetization. In both cases, the exact energy is a linear interpolation of the energy of the neighboring integer systems (*17*, *18*). FC and FS systems are fictitious, but real charge densities can include regions that have FC or FS character, and therefore, correctly modeling these idealized problems helps to ensure that functionals behave correctly in a wide variety of molecules and materials. The FC and FS linearity conditions have shown to be challenging to address with manual design of the functional, but they are easy to illustrate as examples. This situation is ideally suited to a deep learning framework, in which the constraints can be expressed as data and a functional can be trained to obey them and to reproduce the energy of molecular systems.

Our functional is illustrated in Fig. 1. Only the exchange-correlation term was learned and interfaced to a standard Kohn-Sham DFT code [PySCF (*19*)]. The functional was evaluated by integrating local energies computed by a multilayer perceptron (MLP), which took as input both local and nonlocal features of the occupied Kohn-Sham (KS) orbitals, and can be described as a local range-separated hybrid. To train the functional, the sum of two objective functions was used: a regression

[1]DeepMind, 6 Pancras Square, London N1C 4AG, UK. [2]Département d'informatique, ENS, CNRS, PSL University, Paris, France. [3]Departamento de Química and IFIMAC, UAM, 28049, Madrid, Spain. [4]Max Planck Institute for Solid State Research, 70569 Stuttgart, Germany.
***Corresponding author. Email: kirkpatrick@google.com (J.K.); aroncohen@google.com (A.J.C.)**
†These authors contributed equally to this work.

loss for learning the exchange-correlation energy itself and a gradient regularization term that ensured that the functional derivatives can be used in self-consistent field (SCF) calculations after training. For the regression loss, we used a dataset of fixed densities representing reactants and products for 2235 reactions, and the network was trained to map from these densities to high-accuracy reaction energies by means of a least-squares objective (Fig. 1B). Specifically, 1161 training reactions represented atomization, ionization, electron affinity, and intermolecular binding energies of small main-group, H-Kr molecules, and 1074 represented the crucial FC and FS densities only for the atoms H-Ar (supplementary materials, section 2.1). The fixed densities for the main-group molecules were obtained from a popular traditional functional [B3LYP (*20*)], and the energy labels were either obtained from literature (*21*–*25*) or based on in-house complete basis set CCSD(T) (coupled-cluster with single and double and perturbative triple excitations) calculations. More justification on the use of a fixed charge density is provided in the supplementary materials, section 4.3. For gradient training, perturbation theory gives the leading order change in energy, δE_{SCF}, after a single SCF iteration (supplementary materials, section 1.3.1). This energy change depends on the derivatives of the exchange-correlation functional (Fig. 1C), and adding $\delta E_{\mathrm{SCF}}^2$ to the training objective encourages the model to avoid making spuriously large orbital rotations away from reasonable orbitals during self-consistent iteration. This approach was considerably cheaper than supervising explicit self-consistent iterations during training (*26*) or Monte Carlo methods to supervise densities (*12*). Networks with gradients regularized in this way were able to run self-consistently on all reactions in large main-group benchmarks, and DM21 produced accurate molecular densities (supplementary materials, section 5).

After training, the behavior of the functional was analyzed, starting with the archetypal FC and FS systems shown in Fig. 2, A and B. We compared DM21 with SCAN and popular hybrid functionals B3LYP (*20*), M06-2X (*27*), and ωB97X (*28*), with all calculations carried out by using a modified version of PySCF (*19*). Generally, traditional functional approximations are convex with respect to the FC exact condition and concave with respect to the FS exact condition, with improved performance on FC coming at the cost of a larger error in FS, and vice versa. DM21 stands out in comparison as close to the correct behavior for both FC and FS. The functional was trained only on the exact conditions for bare atoms, but correct behavior was also seen on fragments of molecules for both FC and FS, albeit with a somewhat larger error. This result shows that DM21 has not simply memorized the training examples but has found features in the charge density of the atom data that usefully generalize to molecular systems.

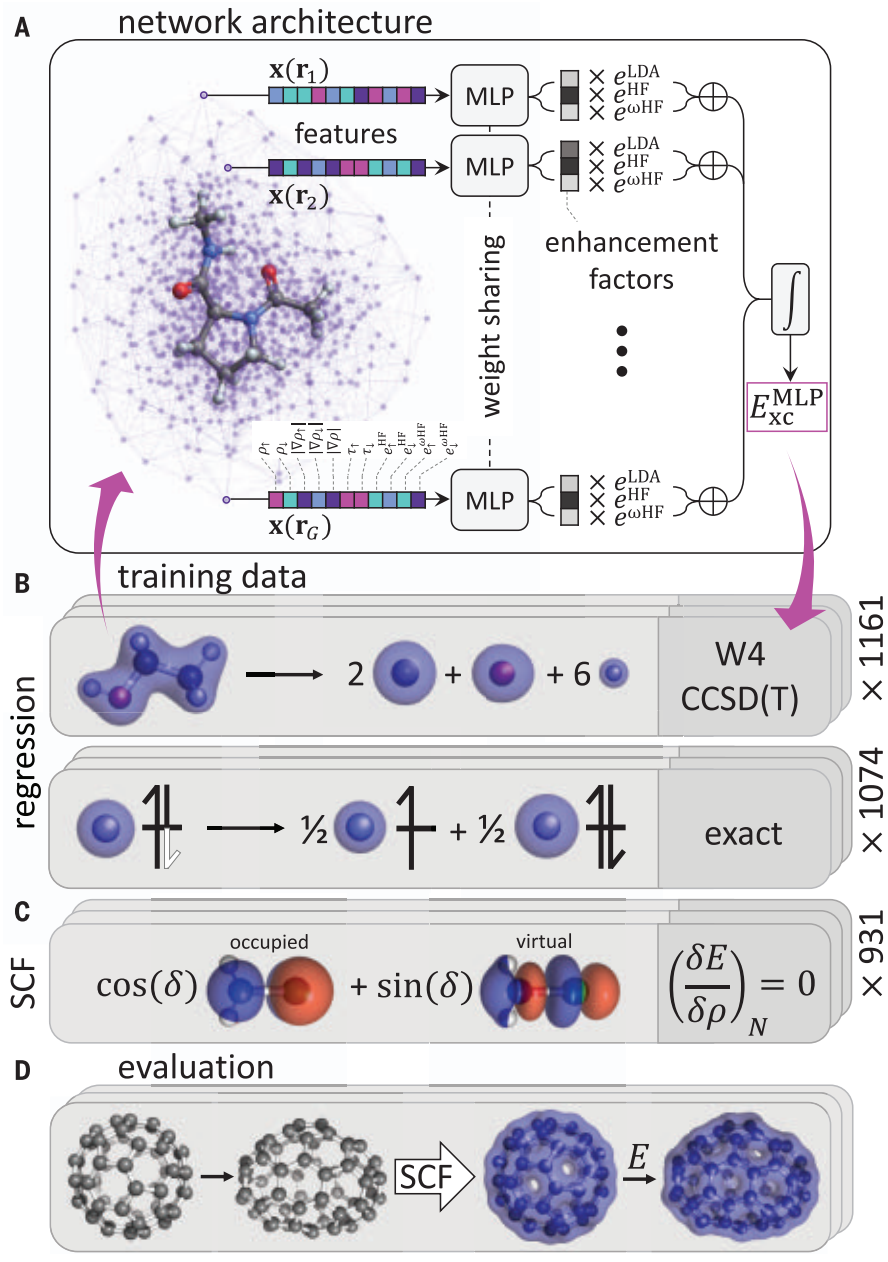


Fig. 1. Overview of the functional architecture and training. (**A**) Features of the electron density computed from KS orbitals are sampled on an atom-centered quadrature grid. Specifically, the input features are the spin-indexed charge density ρ, the norm of its gradient $|\nabla\rho|$, the kinetic energy density τ, and the (range-separated) local Hartree Fock exchange energy densities $e^{\omega\mathrm{HF}}$ and e^{HF}. These are fed through a shared MLP that predicts local enhancement factors for local density approximation and Hartree-Fock contributions to the local exchange-correlation energy density, which is then integrated over all space. A dispersion correction is then added to the functional. (**B**) The network is trained by using a dataset of KS input densities and high-accuracy energy labels for molecules and exact mathematical constraints. (**C**) The gradient of the learned functional at fixed electron number (*N*) is supervised by requesting that the supplied orbitals are a stationary point of the total energy with respect to unitary rotation of occupied and virtual orbitals (illustrated by angle δ). (**D**) Once trained, the functional can be deployed in self-consistent calculations. Numbers on the right indicate dataset sizes (excluding grid augmentations) for the DM21 functional.

Additional limitations of current functionals associated with FC and FS errors are incorrect description of bond breaking for charged and closed-shell neutral molecules,

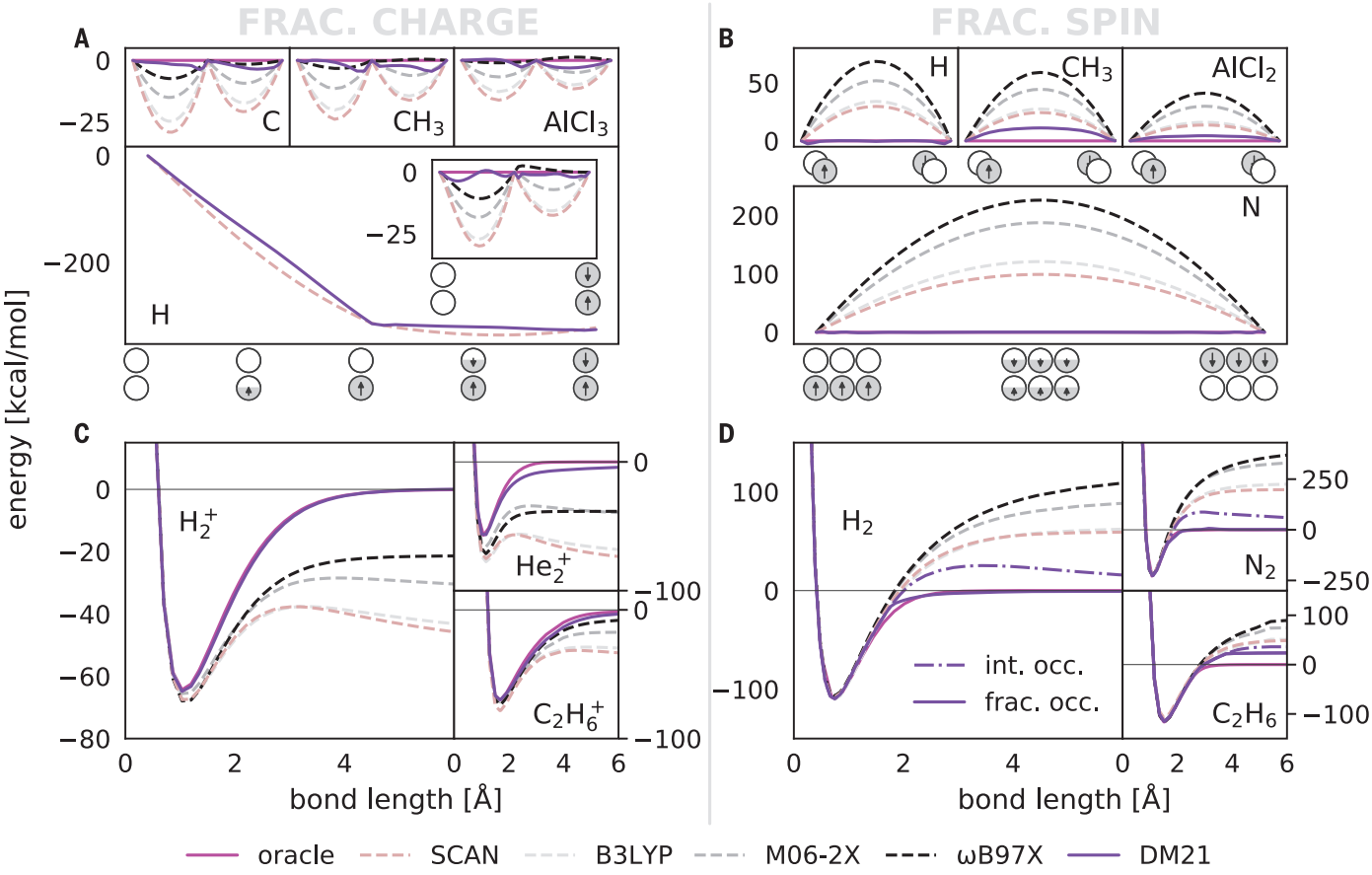
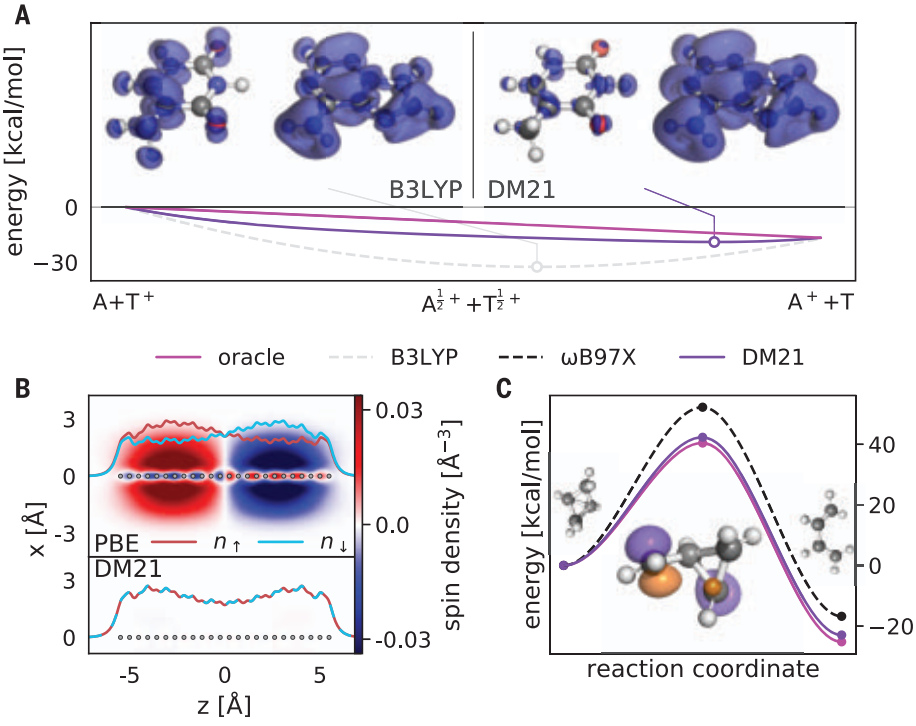


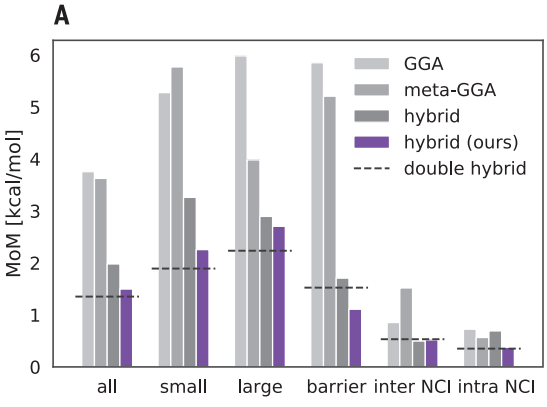
Fig. 2. Training on fractional electron constraints solves charge and spin localization and delocalization errors. (A) and (C) relate to the FC constraint, and (B) and (D) relate to the FS constraint. (**A**) DM21 correctly captures the piecewise linear energy of a H atom as the electron number is continuously varied. (Insets) Deviation from linear behavior for simple atoms (H and C), and small molecules. (**B**) DM21 correctly captures the constancy condition of energy upon interpolating between spin flipped solutions. Shown are the results for a quadruplet (N) and some doublets (H, CH_3, and $AlCl_2$). (**C**) Correct handling of the fractionally charged states generalizes to improved cation binding curves for DM21. The oracle is HF for H_2^+ and UCCSD(T) for He_2^+ and $C_2H_6^+$. (**D**) Improved performance on closed-shell bond breaking. DM21 gives the correct stretched limit but shows a bump at intermediate distances, which is corrected in a restricted optimization that allows fractional occupation of the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO). Oracles for these curves come from FermiNet QMC calculations, except for C_2H_6, which used UCCSD(T) at the basis set limit.

respectively. When dissociating a charged molecule, functionals with a convex error for FC artificially lower the energy by delocalizing charge; as such, they predict that—even at infinite separation—a charged molecule is bound. This limitation is the essence of the well-known charge delocalization error in DFT, and DM21 achieves the correct asymptote as in Fig. 2C. Related discussion on eigenvalues is available in the supplementary materials, section 6. Traditional functionals also grossly overestimate the energy of a stretched closed-shell molecule, whereas DM21 yields correct asymptotes (Fig. 2D). This overestimation is often described in terms of static correlation error under the interpretation that at large separation, there is near degeneracy of bonding and antibonding states that cannot be represented by a single reference method.

Following previous studies (*4*, *29*), we revisited this interpretation and instead suggest that the error is due to the overestimation of the energy for spin delocalized solutions: Closed-shell orbitals are not capable of artificially breaking spin symmetry and localizing spins, giving asymptotes that are too high for functionals with FS error. Additionally, we made a quantitative evaluation of the advantage of DM21 for bond breaking by using an accurate Quantum Monte Carlo bond breaking benchmark (BBB) (supplementary materials, section 8.1). For neutral molecules at intermediate distances, DM21 could exhibit a "hump" in the energy. This feature, seen before with other methods such as the random phase approximation (*30*), can be corrected with an extension to fractional occupation of the closed-shell orbitals (*31*). Of the functionals presented, optimization of the orbital occupations lowered the hump energy only for DM21 (supplementary materials, section 3.2).

Having established the improved FC and FS behavior of DM21 on textbook systems, how this behavior leads to improved description of subtle electronic structure in systems of scientific interest is illustrated in Fig. 3. Three systems from across the sciences were considered: charge delocalization in a DNA base pair, magnetic properties of a compressed hydrogen chain, and reaction barrier heights for a ring-opening intermediate with diradical character. Charge transport in DNA is a subject of considerable experimental and theoretical interest (*32*), and the distribution of the charge of an ionized base pair (adenine and thymine) is shown in Fig. 3A. A popular functional such as B3LYP predicts charge density delocalized over both base pairs, but this prediction is an artefact driven by the violations of FC conditions for the individual bases. Conversely, DM21 is much closer to the correct FC behavior, and charge is localized





B

	GMTKN55 MoM	BBB	QM9
SCAN(:D3$_{BJ}$)	3.63	52.90	3.57
ωB97X(-V)	2.45	70.28	2.13
M06-2X	2.08	64.50	2.20
PW6B95:D3$_0$	1.98	57.81	2.24
DM21	**1.50**	**13.20**	**1.66**

Fig. 3. Exact constraints improve performance on challenging chemistry. (**A**) Charge density for a singly ionized adenine-thymine base pair. B3LYP unphysically delocalizes charge on both base pairs despite adenine having a deeper ionization potential. Conversely, DM21 displays localization of charge on the adenine only. (**B**) Spin density for a compressed chain of 24 hydrogen atoms at 0.48 Å separation. The line density n for each spin channel is overlaid (supplementary materials, section 3.3). The PBE functional (*41*) breaks spin symmetry and leads to an apparent magnetization along the chain. This effect is also predicted by other functionals but is absent in DM21. (**C**) The conrotatory pathways of bicyclobutane isomerization. The HOMO of a single spin channel in an unrestricted calculation is shown for the transition states. Spin is delocalized across two atoms in the conrotatory path, requiring satisfaction of FS for accurate modeling. The oracle is diffusion Monte Carlo from (*42*).

Fig. 4. State-of-the-art performance by DM21 on benchmarks. All errors are in kcal/mol. (**A**) The MoM error metric in each class of reactions from GMTKN55. More details are available in the supplementary materials, section 8.2. DM21 is compared with functionals at rungs two to five, with strong GMTKN55 metrics from (*43*): revPBE:D3$_{BJ}$ (*44*), SCAN:D3$_{BJ}$, and PW6B95:D3$_0$ (*45*). The dashed black line indicates the performance of the double-hybrid functional DSD-PBEP86:D3$_{BJ}$ (*46*). (**B**) Performance of DM21 compared with the SCAN functional and the three best performing hybrid functionals on three benchmark sets: BBB (supplementary materials, section 8.1), GMTKN55, and QM9. The BBB benchmark measures mean absolute errors for selected first- and second-row diatomics relative to high-level UCCSD(T) (cation) and QMC (neutral) calculations. The neutral dimer DFT calculations use a restricted ansatz with integer occupation. QM9 errors are taken relative to high-level G4(MP2) theory. (:D3$_{BJ}$) indicates that the best of SCAN with or without D3 correction is reported on each metric and similarly for VV10 variants (*47*) of ωB97X.

on the adenine unit alone. The difficulties for traditional functionals to localize charges are well understood, but artificial spin localization errors associated with FS violation are less well studied.

In a highly compressed hydrogen chain, unrestricted DFT calculations with traditional functionals localizes spin on what appear to be antiferromagnetic domains (Fig. 3B). This observation has been used to bolster the evidence for a magnetic phase transition (*33*). Conversely, high-level wave function methods did not yield spin polarized solutions, suggesting that the symmetry breaking might be driven by errors in traditional functionals, and DM21 predicts a ground state with no spin symmetry breaking. In the example of ring opening in bicyclobutane (C_4H_6), the energy of the disrotatory transition state is highly overestimated in unrestricted calculations with functionals such as M06-2X and ωB97X but is correctly predicted by DM21. This is again linked to FS error; although the transition state is a singlet, DFT predicts partial localization of the spin with an intermediate $\langle S^2 \rangle$ value between 0 and 1. This means that functionals with incorrect FS behavior give large errors. This result is highly reminiscent of the way that energy is overestimated for closed-shell bond breaking. For C_4H_6, high-level reference calculations are available to verify that DM21 behaves correctly (*34*), but we also observed that hybrid functionals tend to overestimate

barrier heights for transition states with intermediate levels of spin polarization in other sets of reactions (supplementary materials, section 7.2), and this phenomeneon has also been observed in the literature for other cycloaddition reactions (*35*).

Last, beyond the treatment of FC and FS, analysis of DM21 is extended to consider broader classes of main-group chemistry contained in large benchmark sets. Shown in Fig. 4 is the summary performance of DM21 compared with existing functionals on the GMTKN55 benchmark (*36*), a set of subbenchmarks used to probe the behavior of functionals for several important chemical tasks that require extrapolation to systems very distinct from the training set. GMTKN55 includes systems that contain heavy atoms beyond Kr that were never seen during training and that therefore we would not normally recommend for DM21 [we evaluated these using pseudo-potentials following the method in (*36*)]. We calculated the mean absolute error for each subbenchmark and report the mean of these means (MoM), with additional reweighted scores presented in the supplementary materials, section 8.2. Overall, the behavior of DM21 is better than the best hybrid functional and approaches the performance of the much more expensive double-hybrid functionals. In particular, DM21 excels at the description of barrier heights, which is unexpected given that no transition states were present in the training data. Ablation of the training data revealed that the improved performance on this class stems from the FC and FS training data (supplementary materials, section 8.2). Additionally, DM21 considerably outperforms existing functionals on the mindless benchmark subset (MB16-43) of GMTKN55, a set of atomization energies for randomly generated geometries of atoms that was designed to test performance on out-of-distribution exotic geometries. Otherwise well-performing functionals have large errors for this dataset, such as ωB97-X (>30 kcal/mol error), as well as nonempirical functionals, such as SCAN (15 kcal/mol error), but DM21 had an error of <5 kcal/mol. This dataset is as far from the training and validation set as is possible, showing that our functional performs well even when applied to out-of-distribution generalization examples. To further stress this out-of-distribution generalization, benchmarking on the QM9 (*37*, *38*) dataset is shown in Fig. 4B. This is a collection of 133,857 enumerated isomers of organic molecules with up to nine heavy atoms, and again DM21 displays state-of-the-art performance when compared with that of existing functionals.

This work has successfully demonstrated that deep learning provides a framework for the development of improved functionals with new properties. Specifically, development of DM21 combined highly accurate chemical data and fractional electron constraints to address major shortcomings in prior functionals. This combination led to a better description of the quantum mechanical interaction of electrons, from simple atomization energies to complicated reaction barriers and exotic compressed hydrogen chains. This work has focused on main-group chemistry, but the methodology can easily be extended to incorporate new data and constraints that will allow even better functionals to be trained. To illustrate this flexibility, results from adding the uniform electron gas condition to the functional are provided in the supplementary materials, section 4.2. Many natural phenomena, from charge transfer excitations to the stripe phase in superconducting cuprates (*39*), rely on subtle effects dependent on the motion of charge and spin polarization, and we believe that the functional presented here, and the approach that we suggest, are central to improving our understanding of these and other properties of materials.

REFERENCES AND NOTES

1. P. Hohenberg, W. Kohn, *Phys. Rev.* **136** (3B), B864–B871 (1964).
2. W. Kohn, L. J. Sham, *Phys. Rev.* **140**, A1133–A1138 (1965).
3. R. Van Noorden, B. Maher, R. Nuzzo, *Nature* **514**, 550–553 (2014).
4. A. J. Cohen, P. Mori-Sánchez, W. Yang, *Science* **321**, 792–794 (2008).
5. J. Sun, A. Ruzsinszky, J. P. Perdew, *Phys. Rev. Lett.* **115**, 036402 (2015).
6. C. Li, X. Zheng, N. Q. Su, W. Yang, *Natl. Sci. Rev.* **5**, 203–215 (2018).
7. N. Q. Su, C. Li, W. Yang, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 9678–9683 (2018).
8. R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, *J. Chem. Theory Comput.* **11**, 2087–2096 (2015).
9. O. T. Unke *et al*., *Chem. Rev.* **121**, 10142–10186 (2021).
10. J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, K. Burke, *Phys. Rev. Lett.* **108**, 253002 (2012).
11. L. Li *et al*., *Int. J. Quantum Chem.* **116**, 819–833 (2016).
12. R. Nagai, R. Akashi, O. Sugino, *npj Comput. Mater.* **6**, 43 (2020).
13. A. V. Sinitskiy, V. S. Pande, Physical machine learning outperforms "human learning" in quantum chemistry. arXiv:1908.00971 [physics.chem-ph] (2019).
14. K. Ryczko, D. A. Strubbe, I. Tamblyn, *Phys. Rev. A* **100**, 022512 (2019).
15. Y. Chen, L. Zhang, H. Wang, E. Weinan, *J. Chem. Theory Comput.* **17**, 170–181 (2021).
16. S. Dick, M. Fernandez-Serra, *Nat. Commun.* **11**, 3509 (2020).
17. J. P. Perdew, R. G. Parr, M. Levy, J. L. Balduz Jr., *Phys. Rev. Lett.* **49**, 1691–1694 (1982).
18. A. J. Cohen, P. Mori-Sánchez, W. Yang, *J. Chem. Phys.* **129**, 121104 (2008).
19. Q. Sun *et al*., *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8**, e1340 (2018).
20. A. D. Becke, *J. Chem. Phys.* **98**, 5648–5652 (1993).
21. A. Karton, N. Sylvetsky, J. M. L. Martin, *J. Comput. Chem.* **38**, 2063–2075 (2017).
22. A. Karton, S. Daon, J. M. L. Martin, *Chem. Phys. Lett.* **510**, 165–178 (2011).
23. A. Karton, A. Tarnopolsky, J. M.L. Martin, *Mol. Phys.* **107**, 977–990 (2009).
24. B. Brauer, M. K. Kesharwani, S. Kozuch, J. M. L. Martin, *Phys. Chem. Chem. Phys.* **18**, 20905–20925 (2016).
25. M. K. Kesharwani, D. Manna, N. Sylvetsky, J. M. L. Martin, *J. Phys. Chem. A* **122**, 2184–2197 (2018).
26. L. Li, S. Hoyer, R. Pederson, R. Sun, E. D. Cubuk, P. Riley, K. Burke, Kohn-Sham equations as regularizer: building prior knowledge into machine-learned physics. arXiv:2009.08551 [physics.comp-ph] (2020).
27. Y. Zhao, D. G. Truhlar, *Theor. Chem. Acc.* **120**, 215–241 (2008).
28. J.-D. Chai, M. Head-Gordon, *J. Chem. Phys.* **128**, 084106 (2008).
29. A. J. Cohen, P. Mori-Sánchez, W. Yang, *Chem. Rev.* **112**, 289–320 (2012).
30. M. Fuchs, Y.-M. Niquet, X. Gonze, K. Burke, *J. Chem. Phys.* **122**, 094116 (2005).
31. A. D. Becke, *Density Functionals*, E. R. Johnson, Ed. (Springer, 2014), vol. 365, pp. 175–186.
32. J. C. Genereux, J. K. Barton, *Chem. Rev.* **110**, 1642–1662 (2010).
33. M. Motta *et al*., *Phys. Rev. X* **7**, 031059 (2017).
34. N. P. Bauman, J. Shen, P. Piecuch, *Mol. Phys.* **115**, 2860–2891 (2017).
35. H. V. Pham, K. N. Houk, *J. Org. Chem.* **79**, 8968–8976 (2014).
36. L. Goerigk, S. Grimme, *J. Chem. Theory Comput.* **6**, 107–126 (2010).
37. R. Ramakrishnan, P. O. Dral, M. Rupp, O. A. von Lilienfeld, *Sci. Data* **1**, 140022 (2014).
38. H. Kim, J. Y. Park, S. Choi, *Sci. Data* **6**, 109 (2019).
39. Y. Zhang *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 68–72 (2020).
40. J. Kirkpatrick *et al*., Zenodo (2021); doi:10.5281/zenodo.5482370.
41. J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
42. R. Berner, A. Lüchow, *J. Phys. Chem. A* **114**, 13222–13227 (2010).
43. L. Goerigk *et al*., *Phys. Chem. Chem. Phys.* **19**, 32184–32215 (2017).
44. Y. Zhang, W. Yang, *Phys. Rev. Lett.* **80**, 890 (1998).
45. Y. Zhao, D. G. Truhlar, *J. Phys. Chem. A* **109**, 5656–5667 (2005).
46. S. Kozuch, J. M. L. Martin, *Phys. Chem. Chem. Phys.* **13**, 20104–20107 (2011).
47. N. Mardirossian, M. Head-Gordon, *Phys. Chem. Chem. Phys.* **16**, 9904–9924 (2014).

ACKNOWLEDGMENTS

We thank C. Donner for technical assistance in running PySCF at scale, T. Back for help with project inception, K. Kavukcuoglu and H. Maclean for useful discussions, D. Rezende and O. Vinyals for reviewing the manuscript, and the rest of the DeepMind team for their support. **Author contributions:** J.K., B.M., D.H.P.T., A.L.G., and A.J.C. designed and built the DM21 network architecture and training datasets, with advice from P.M.-S.; J.K., B.M., D.H.P.T., A.L.G., and A.G.D.G.M. trained the functionals. J.K., B.M., D.H.P.T., A.L.G., J.S.S., A.G.D.G.M., P.M.-S., and A.J.C. evaluated the functionals on chemical systems and interpreted the results. J.S.S. and D.P. obtained the FermiNet oracles. J.K., D.H.P.T., L.T., M.F., A.W.R.N., A.J.C., P.K., and D.H. initiated the project. B.M., D.H.P.T., J.S.S., and S.P. contributed to software engineering. J.K., A.O., P.K., and D.H. managed the project. J.K., A.L.G., P.M.-S., and A.J.C. wrote the manuscript, with help from L.R.C. **Competing interests:** There is a pending patent application, US Provisional Application 63/135,223, on this work. The remaining authors declare no competing interests. **Data and materials availability:** The code and learned network weights for running our functionals on new systems using the open source PySCF package is available at https://github.com/deepmind/deepmind-research/tree/master/density_functional_approximation_dm21 and (*40*). The BBB benchmark, errors for DM21 on all reactions in GMTKN55, and all data underlying the figures are also available as digital materials (data files S1 to S3) in the supplementary materials. All other data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj6511
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S8
References (*48*–*73*)
Data Files S1 to S3

24 May 2021; accepted 16 September 2021
10.1126/science.abj6511

PHASE-CHANGE MEMORY

Elemental electrical switch enabling phase segregation–free operation

Jiabin Shen[1,2]†, Shujing Jia[1,2]†, Nannan Shi[3], Qingqin Ge[3], Tamihiro Gotoh[4], Shilong Lv[1], Qi Liu[5], Richard Dronskowski[6], Stephen R. Elliott[7,8], Zhitang Song[1]*, Min Zhu[1]*

Nonvolatile phase-change memory has been successfully commercialized, but further density scaling below 10 nanometers requires compositionally and structurally homogeneous materials for both the memory cell and the associated vertically stacked two-terminal access switch. The selector switches are mostly amorphous-chalcogenide Ovonic threshold switches (OTSs), operating with a nonlinear current response above a threshold voltage in the amorphous state. However, they currently suffer from the chemical complexity introduced by the quaternary or even more diverse chalcogenide compositions used. We present a single-element tellurium (Te) volatile switch with a large (≥11 megaamperes per square centimeter) drive current density, ~10^3 ON/OFF current ratio, and faster than 20 nanosecond switching speed. The low OFF current arises from the existence of a ~0.95–electron volt Schottky barrier at the Te–electrode interface, whereas a transient, voltage pulse–induced crystal-liquid melting transition of the pure Te leads to a high ON current. Our discovery of a single-element electrical switch may help realize denser memory chips.

The explosive growth of digital data urgently requires memory devices capable of ultrafast programming speeds and ultrahigh densities, yet these requirements cannot be fulfilled by mainstream dynamic random-access memory (DRAM) and flash devices (*1*, *2*). Nonvolatile phase-change memory (PCM), based on reversible (crystal-to-amorphous) structural changes of chalcogenides, exhibits a thousand times faster operation than flash memory. PCM also has a storage density (in terms of the parameter $4F^2$, where F denotes the feature size) that is comparable with flash, currently one of the main nonvolatile memory technologies (*3*, *4*). PCM-based products, dubbed Optane, were successfully launched into the market in 2017 by Intel and Micron (*5*). Gigabyte-scale Optane uses a 20-nm node technology and a three-dimensional (3D) stacked architecture (*5*). Further advances to fabricate even denser PCM devices can be achieved by continued dimensional downscaling to sub-10-nm feature sizes (*6*, *7*).

A full PCM device is composed of a memory cell and an associated access-switching device (also called a selector). Aggressive miniaturization of both components is needed (*8*). This requires not only state-of-the-art manufacturing but also that the materials used in both cells be (and remain) extremely compositionally and structurally homogeneous on the nanometer scale; this is necessary to avoid unwanted performance variations and also to guarantee long switching-operation lifetimes. Complex Ge-Sb-Te–based chalcogenides are likely used in the Optane PCM technology (*9*), but a different approach would be to use elemental-Sb phase-change nonvolatile memory material that is pure and intrinsically homogeneous, which has been previously explored (*10*).

Each memory-cell operation—including the most frequent, the read operation—needs to access the switch; therefore, excellent switch reliability and ultralong cycling lifetimes are required (*11*). As a consequence, the switching cell faces greater challenges for higher-density scaling. When using chalcogenides as the memory cell, albeit with different stoichiometries, two-terminal chalcogenide Ovonic threshold switches (OTSs) have successfully been used in 3D PCM products instead of charge-based Si diode and MOSFET (metal-oxide-semiconductor field-effect transistor) switches (*8*, *12*, *13*). As only amorphous chalcogenides exhibit the volatile OTS behavior, elements such as As, Si, and N are often introduced as dopants to improve the thermal stability to withstand a temperature of 400° to 450°C in the back-end-of-line (BEOL) fabrication process (*14*–*16*). OTS cells, being quaternary compounds (e.g., Ge-As-Se-Si) or having even more complex compositions, can suffer from compositional inhomogeneities (*17*–*19*). This challenge becomes more serious when decreasing the scale, as reduced dimensions result in a drastic degradation of thermal stability caused by interfacial hetero-crystallization (*20*, *21*). This may make the conventional doping strategy ineffective, unpredictable, and time consuming. However, the continuing exploration of Ge-, Si-, Al-, B-, and C-Te OTSs (especially Te-rich Ge) has provided some valuable clues for improving stability (*15*, *22*, *23*). Because pure elemental Ge, Si, Al, B, and C materials are widely known as semiconductors or metals, without any intrinsic switching tendency, this suggests that elemental tellurium plays a dominant role in the volatile-switching process. The crucial question then rests on the ability of pure Te to switch.

In this study, we prepared T-shaped switching devices with between 60- and 200-nm-sized electrodes made of pure Te films (with ≥99.99% purity). One cross-sectional transmission electron microscopy (TEM) image shows a Te device with a 60-nm TiN plug, above which we stacked a ~20-nm Te layer (Fig. 1A). Before deposition of the Te, we etched the device with Ar ions for 20 min to clean the surface oxide layer. Cleaning was confirmed by the fact that Te-free cells had a low resistance of between 150 ohms and 2 kilohms [see (*24*) for fabrication details and tables S1 and S2]. The Te layer is intrinsically homogeneous, without any oxide layer, which we confirmed by the overlapped corresponding energy-dispersive spectroscopy (EDS) mappings of W, Ti, Te, and Si (fig. S1). We measured the current-voltage (*I-V*) curves of Te devices under consecutive pulses (Fig. 1B). Initially, the Te device was in a high-resistance state (OFF state) through which a negligible current passed. As we increased the voltage to ~2.5 V (the firing or forming voltage, V_{fire}), the current suddenly increased to ~1 mA, and a negative differential resistance behavior appeared. Upon reducing the voltage to below ~0.9 V, the so-called holding voltage (V_{hold}), the cell returned to the high-resistance state. In the second sweep, we only needed ~1.2 V (threshold voltage, V_{th}) to trigger the transition from the OFF state to the ON state, and we obtained an ON current of ~0.4 mA. The switching behavior was completely reproducible, with a V_{th} variation of ~0.4 V (fig. S2). The typical, volatile *I-V* behavior directly proves that pure Te devices behave as switches (*15*, *22*). Intrinsically and completely different from the Te cells, elemental-Sb devices, with the same T-shaped device structure and 20 nm film thickness, exhibit a typically nonvolatile memory feature characterized by a nonvolatile *I-V* behavior (fig. S3), as previously reported (*10*). Interestingly, the values of programming voltages for the Te cells, including V_{fire}, V_{th}, and V_{hold}, hardly vary with the electrode size (Fig. 1C), with a maximum variation of 0.7 V (for V_{fire}) for more than 100 cells (fig. S4), thereby indicating predictable parameters for circuit design upon scaling.

[1]State Key Laboratory of Functional Materials for Informatics, Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, 200050 Shanghai, China. [2]University of the Chinese Academy of Sciences, Beijing 100029, China. [3]Thermo Fisher Scientific China, Shanghai 200050, China. [4]Department of Physics, Graduate School of Science and Technology, Gunma University, Maebashi 3718510, Japan. [5]Frontier Institute of Chip and System, Fudan University, Shanghai 200433, China. [6]Institute of Inorganic Chemistry, Chair of Solid-State and Quantum Chemistry, RWTH Aachen University, Aachen 52056, Germany. [7]Trinity College, University of Cambridge, Cambridge CB2 1TQ, UK. [8]Physical and Theoretical Chemistry Laboratory, University of Oxford, Oxford OX1 3QZ, UK.
*Corresponding author. Email: minzhu@mail.sim.ac.cn (M.Z.); ztsong@mail.sim.ac.cn (Z.S.)
†These authors contributed equally to this work.

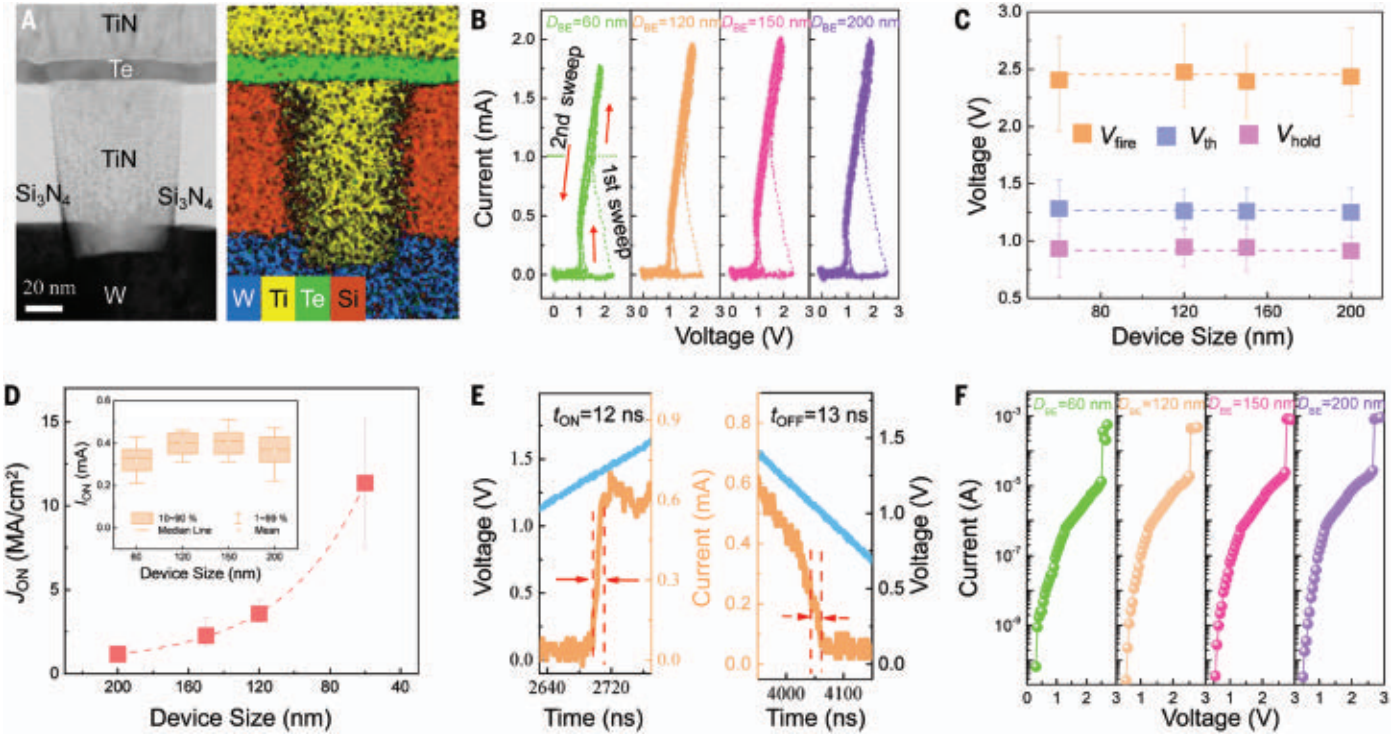


Fig. 1. Structure and performance of Te switching devices. (**A**) Cross-sectional TEM image of a Te cell with a ~60-nm TiN plug and corresponding EDS elemental mappings. The Te layer is ~20 nm thick. (**B**) Volatile *I-V* behavior of Te devices with various device sizes (60, 120, 150, and 200 nm) under consecutive voltage pulses. D_{BE} is the diameter of the bottom electrode (BE). (**C**) Statistical distribution of transition voltages (V_{fire}, V_{th}, and V_{hold}) in more than 100 Te devices. (**D**) The ON current density increases quadratically as the device size is reduced. A 60-nm cell already exhibits an ON current density of ~11 MA/cm^2. The inset shows the statistical distribution of ON currents measured at V_{th}, appearing to be size independent. (**E**) ON and OFF switching speeds, ~12 and ~13 ns, respectively, obtained from the measured currents (orange) by applying ~4-V triangular voltage pulses with rising and falling edges of 1 μs (blue). (**F**) DC *I-V* curves of Te devices in series with a 1-kilohm resistor. ON and OFF currents (measured at V_{th} and $V_{th}/2$) are ~0.5 mA and ~0.5 μA, respectively. The ratio of ON and OFF currents is thus ~10^3 and remains at this value while the electrode size changes.

The ON current, ~0.4 mA, was also size independent (Fig. 1D, inset). In other words, the ON current density, J_{ON}, quadratically increased with a decrease of the device size. We achieved a J_{ON} value of ~11 MA/cm^2 for a 60-nm device, which further increased to 14.9 MA/cm^2 for a 45-nm device (fig. S5). This is comparable to Te- and Se-based OTSs (fig. S6) and sufficient for programming PCM cells (requiring ≥10 MA/cm^2) (*15*, *22*, *25*). We found large J_{ON} values within ~12 ns (Fig. 1E and fig. S7), while the time for turning off the switch was ~13 ns. This switching time is almost comparable to DRAM (~10 ns), suggesting that Te switches may also serve in future DRAM-like PCM devices (*21*, *26*). To determine the leakage current, we performed direct-current *I-V* measurements (Fig. 1F) with a protective 1-kilohm resistor in series with the Te device. Initially, a current of <0.1 nA passed through the Te cell, and then the current exponentially increased upon increasing the voltage. We found an OFF current of ~0.45 μA at $V_{th}/2$ in the 60-nm device, resulting in a ~10^3 ON/OFF ratio.

By its very nature, the pure elemental-Te switching material circumvents any phase-segregation issues, which are often observed in failed switching cells. We achieved a 2×10^8–cycle endurance by using square voltage pulses (Fig. 2A and fig. S8). Our corresponding dynamical-response results clearly show that the Te switch was successfully turned on and off by each pulse during the 2×10^8–cycle endurance test (fig. S8). In addition, switching cells made from tellurium evidence a bidirectional threshold-switching characteristic (Fig. 2B), suggesting potential applications in other emerging memory types, such as resistive random-access memory (RRAM) (*27*). The Te films are still switchable down to 5-nm thickness without sacrificing the ON current (Fig. 2C) and OFF current (fig. S9). Surprisingly, the value of V_{th} decreases to ~1.1 V for both 5-nm and 10-nm Te devices, implying that the Te device does not follow an electric field–driven process, making it distinctly different from OTS cells.

Given a drive current density greater than 10 MA/cm^2, a switching speed on the nanosecond scale, and an ON/OFF current ratio of ~1000, as well as an intrinsically homogeneous composition, we consider the pure elemental-Te switch to be a very promising selector candidate for PCM applications. Moreover, the Te material is perfectly compatible with the materials in PCM memory cells, which are typically Ge-Sb tellurides, thereby enabling stacking-compatible integration (*4*, *28*). The final factor that needs to be considered is the thermal stability of Te switches; more specifically, whether they can withstand a thermal treatment of 400°C for 30 min, as required in the BEOL process (*14*–*16*). To study the thermal stability, we annealed 200-nm Te cells at 100° to 400°C for 30 min. We measured their *I-V* curves at room temperature after the thermal treatment (Fig. 2D). The Te switch still exhibits a clear threshold-switching behavior, even after the 400°C treatment. We observed a voltage increase of ~0.9 V in V_{fire}, together with a ~0.25 V increase for V_{th} and V_{hold}. Most importantly, the ON current decreased by only ~0.09 mA. The difference in switching voltage and ON current is mainly due to the increase of the TiN resistance under high-temperature annealing (table S3) (*29*). That said, the ON switching speed remains at ~15 ns (fig. S10). These results therefore directly show that our pure Te switch is BEOL compatible.

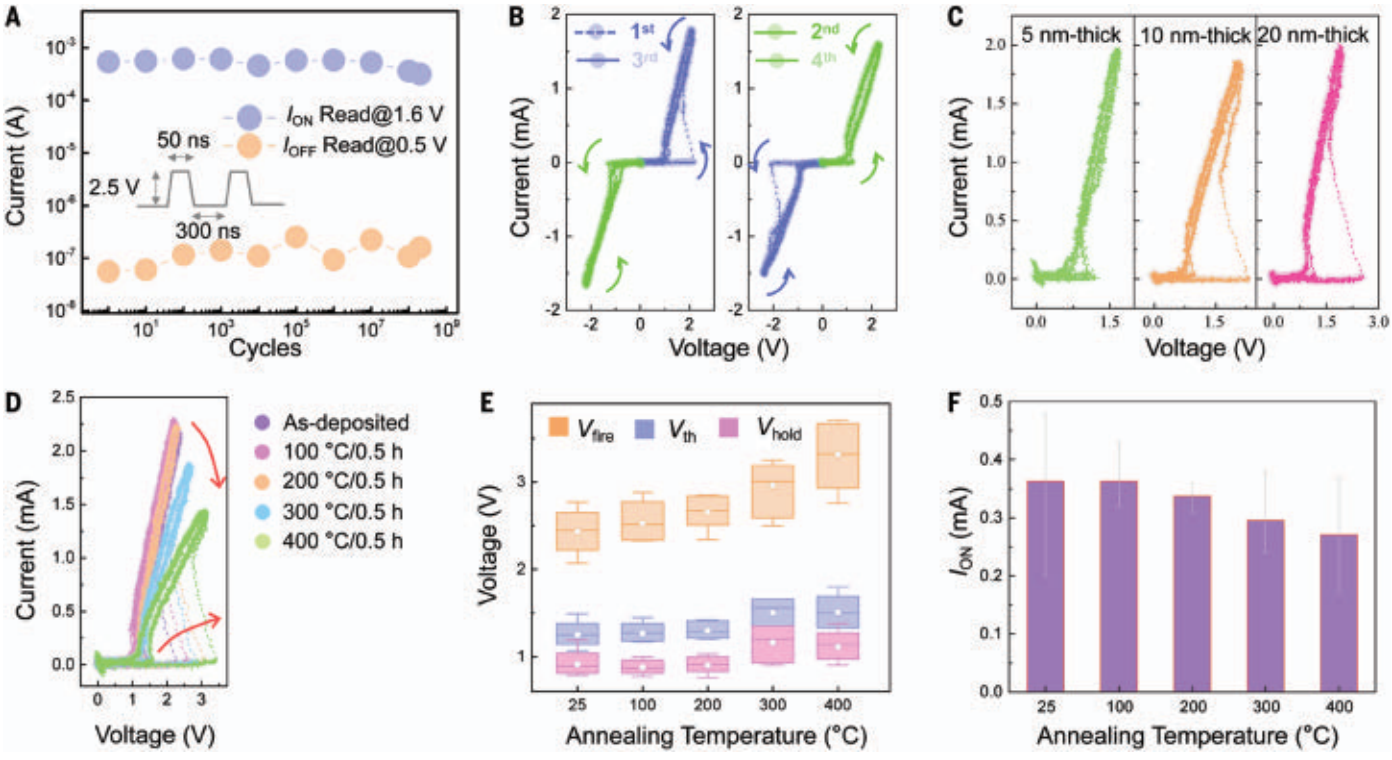


Fig. 2. Electrical properties of 200-nm Te switching devices, before and after thermal annealing. (**A**) Endurance of a Te device exceeding 2×10^8 cycles of applied voltage pulses. (**B**) Bidirectional threshold-switching characteristic of Te switches. The arrows point along the switching directions. (**C**) Volatile *I-V* behavior of Te devices with various Te thicknesses (5, 10, and 20 nm) under consecutive pulses. (**D**) *I-V* behavior of 200-nm Te devices annealed at different annealing temperatures (25°, 100°, 200°, 300°, and 400°C). (**E**) Statistical distribution of transition voltages before and after thermal annealing. (**F**) The ON current slightly decreases with increasing annealing temperature.

The microstructure of the Te cell after repeated operations, which we investigated with TEM (Fig. 3, A to D), explains the good thermal stability. The Te layer farthest away from the TiN plug is fully crystalline, adopting a trigonal structure (Fig. 3B). Further analysis of a Te device without programming (fig. S11) directly confirmed the crystalline state of the as-deposited Te layer, resulting from its ultralow crystallization temperature of –10°C (*30*, *31*). The Te layer right above the TiN plug is also crystalline but has a different orientation (Fig. 3C). Atomic EDS mapping of Te confirms the high purity of the Te film.

We measured *I-V* curves in the subthreshold region at 100°, 130°, 150°, 175°, and 200°C to answer the question why the crystalline Te-based cell possesses a high-resistance state (with a low leakage current) and volatile-switching behavior (Fig. 3E). We determined the activation energy, E_a, using $-\partial\log(I)/\partial(1/k_BT)$ (fig. S12, A to C), where k_B is the Boltzmann constant and T is the temperature (*32*). For voltages lower than ~0.8 V, these *I-V* curves are characterized by I being proportional to $\exp(V^{1/2})$ (fig. S12F) with a large activation energy E_{a1} of 0.85 eV at ~0 V, suggesting a Schottky-emission mechanism (*33*). Above 1.25 V, however, I is proportional to $\exp(V)$ after the disappearance of the Schottky barrier (Fig. 3E and fig. S12F), with an initial activation energy E_{a2} value of ~0.19 eV. By measuring the bandgap, E_g, the work function, and the Fermi level, E_F, of a 20-nm-thick Te film, as well as the work function of the TiN layer (Fig. 3F), the electronic energy-band diagrams for the Te/TiN interface are experimentally accessible (Fig. 3G). We found an E_g of 0.45 eV and a work function of 4.59 eV for the 20-nm Te layer, and we found a value of 3.82 eV for the work function of the TiN film. These values agree well with previously reported results: 0.33 to 0.49 eV (*34*), 4.65 eV (*35*), and 3.74 eV, respectively (*36*). For a p-type Te semiconductor (confirmed by the positive Seebeck coefficient, 8.8 mV/K) (*37*), the 0.77 eV difference in the work function between Te and TiN results in a large (0.95 eV) Schottky barrier [= 0.77 + (E_F – E_v)], agreeing well with the value of E_{a1} at ~0 V (i.e., only ~0.1 eV higher). Because of the intrinsic p-type doping in crystalline Te, a simple semiconductor-metal contact with a Schottky barrier can be naturally formed between Te and its adjacent electrode (*38*, *39*). A Schottky barrier was also found for Te-based PCM–metal interfaces (*40*). Hence, only an ultralow current, ~0.1 nA, can pass through the Te cell with a ~0.05 V bias, as we observed (Figs. 1F and 3E). In contrast, nonvolatile memory devices made of crystalline Sb have a low resistance of ~1.6 kilohms, owing to an absence of the Schottky barrier (fig. S3B). The Schottky barrier of the Te switch disappears as the voltage increases above 0.95 V, because then the semiconductor behavior of the crystalline Te film becomes dominant, as indicated by the perfect match of the 0.18 eV energy barrier (= E_F – E_V) with the Ea_2 value, ~0.19 eV. This observation also implies that the OFF current (and the ON/OFF ratio) can be further decreased (increased) by using electrodes with relatively different work functions, such as Pt (with a 5.65 eV work-function value) (*40*), or few-layer Te with >1 eV bandgaps (*41*).

We carried out in situ TEM investigations for monitoring the switching process of the Te cells (*42*). We prepared ~100- to 200-nm-diameter Te nanodevices (Fig. 4A) on a Cu TEM grid (fig. S13) (*24*), having the same sandwich structure (TiN/Te/TiN, seen in EDS mapping) as the T-shaped device (Fig. 1A). Similar to our T-shaped device, an as-deposited Te layer is in the trigonal phase (Fig. 4C) that we identified by the corresponding fast Fourier transform (FFT) image (Fig. 4D). We contacted the applied nanomanipulator, a tungsten tip, onto the Te nanodevice (Fig. 4B, inset), through

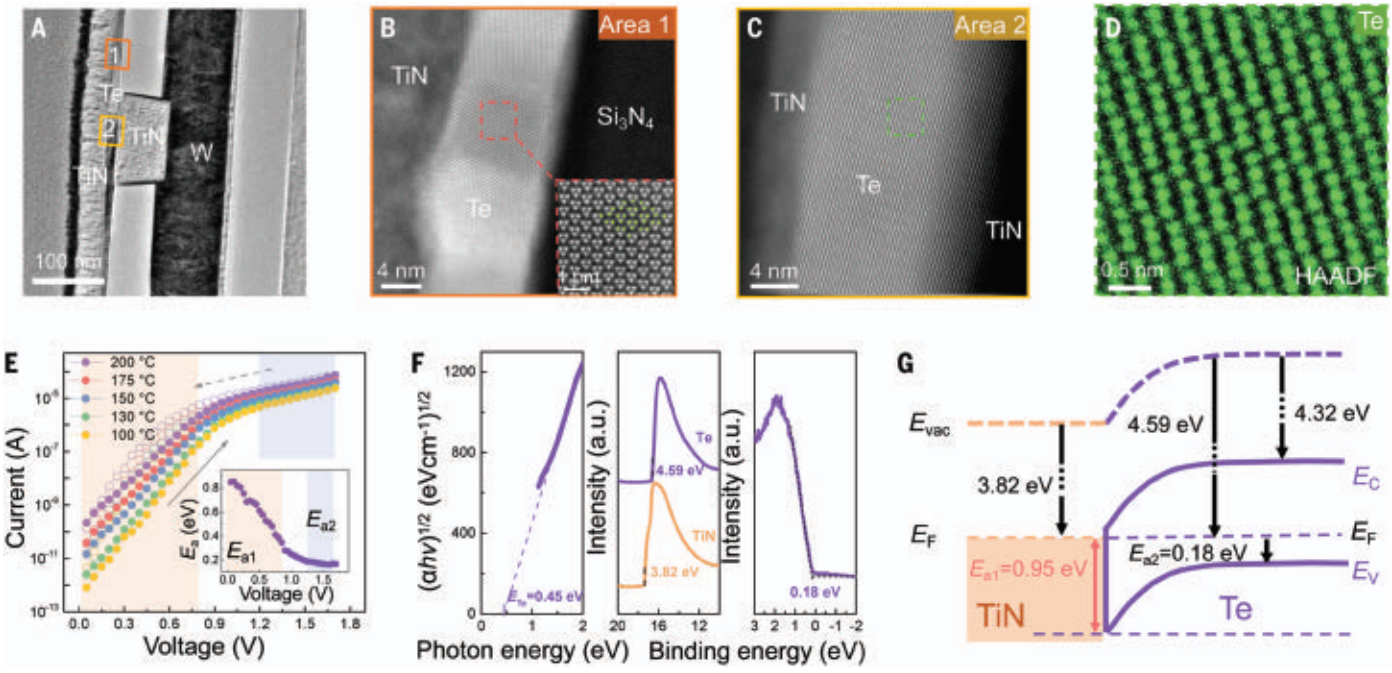


Fig. 3. Origin of the low OFF current in the Te switch. (**A**) Cross-sectional TEM image of the Te device after pulsed switching. (**B**) TEM image of the Te layer farther away from the TiN electrode. As-deposited Te is in the crystalline state owing to its very low crystallization temperature, and the inset shows Te to be trigonal ($a = b = 4.512$ Å and $c = 5.960$ Å). (**C**) TEM image of the Te layer above the TiN electrode. After switching, the Te layer remains trigonal but with a different crystal-growth orientation. Corresponding atomic mapping is shown in (**D**). (**E**) *I-V* characteristics of a Te switch in the subthreshold region, measured at 100°, 130°, 150°, 175°, and 200°C, from which the activation energy, E_a, can be obtained. Two *I-V* relationships, $I \propto \exp(V^{1/2})$ and $I \propto \exp(V)$, are found (fig. S12). (**F**) Bandgap, work functions, and Fermi level of a 20-nm-thick Te film. A Tauc plot of the optical-absorption coefficient, α, that is $(\alpha h\nu)^{1/2}$, versus photon energy, $h\nu$, of a crystalline Te film indicates a bandgap of 0.45 eV. The work function is 4.59 eV for a Te film, whereas it is 3.82 eV for the TiN electrode, as obtained from ultraviolet (UV) photoelectron spectra. UV-excited valence-band spectra show the Fermi-level energy (E_F) to be 0.18 eV for Te. a.u., arbitrary units. (**G**) Experimentally determined energy-band diagrams of the p-type Te/TiN interface, obtained from (F). E_{a1} arises from the Schottky barrier, whereas E_{a2} corresponds to the energy required to move hole carriers from E_F to the valence-band edge (E_v). E_{vac}, vacuum level; E_C, conduction-band edge.

which we recorded the *I-V* curves and corresponding times (Fig. 4B). The microstructural evolution of the Te layer during the switching process was synchronously recorded by a high-speed camera in the TEM (movies S1 and S2). The recording time occurs on the seconds time scale, much longer than the nanosecond-scale switching process. Still, the switching processes could be captured by the applied direct-current voltage (~100 ms every step).

At the beginning of the switching process (Fig. 4, G to J, and movie S1), the Te device was in the OFF state with negligible passing current. As time proceeded, we measured the corresponding currents (Fig. 4B, red arrows). The lattice of the Te film was slightly blurred owing to a slight trembling of the nanodevice during switching. Nonetheless, the Te film appeared to be melted as the voltage went beyond 2.5 V (Fig. 4, H to J), which can be directly observed in movies S1 and S2. The flow of the liquid caused a small void to appear in this unconfined nanodevice (Fig. 4E, red arrow). This unwanted void does not appear in the T-shape device after operation (Fig. 3A). Tellurium easily melts by electrical pulse–induced Joule heating owing to its low melting point of ~445°C (*37*). As liquid Te exhibits a metal-like low resistivity of ~3×10^{-4} ohms·cm (*42*), a submilliampere ON current can then pass through the device, as in the T-shaped Te device (Fig. 1). The 0.28 mA reduction in the ON current was caused by the larger protective resistor (10 kilohms) in series with the Te nanodevice. After removing the applied voltage, the liquid Te layer spontaneously recrystallized back into the trigonal structure (Fig. 4E) owing to its ultralow crystallization temperature (–10°C), with a nanosecond-scale crystallization speed (*29*, *43*), which in turn resulted in the recurrence of the OFF state. The recrystallized Te had a noticeably different crystalline orientation before (Fig. 4D) and after (Fig. 4F) switching, further confirming the crystalline-liquid-crystalline phase-transition mechanism in the pure elemental-Te switch. This process is also responsible for the ~31.5% thickness increase in the active Te layer above the TiN plug in the T-shaped device we observed and explained (*24*) after 3×10^8 cycles of operation when the device failed (fig. S14).

The underlying mechanism of the pure Te volatile switch intrinsically differs from that of chalcogenide-based nonvolatile PCMs, which depend on the amorphous-to-crystalline phase transition (fig. S3). The large Schottky barrier formed at the elemental-Te–electrode interface enables a low leakage current required for the switch cell, which is absent in the elemental-Sb nonvolatile memory device. Our Te devices are also unlike the OTS, in which the chalcogenide materials remain frozen in the amorphous state throughout the switching process. Not only may the elemental-Te switch we presented be considered a new switching type, we also expect more elements and/or compounds with low crystallization temperature and fast crystallization ability to possess such volatile switchability, but only if there is a sufficiently large difference in properties of the ON and OFF states. In Te, this difference is due to the metallic-like–liquid and semiconducting-crystal behavior. Integration of pure Te (as the switching selector cell) with pure Sb (as the nonvolatile memory cell) (*10*) would be an interesting combination, offering the potential for fabricating entirely single-element–based memory arrays. We have demonstrated that the device

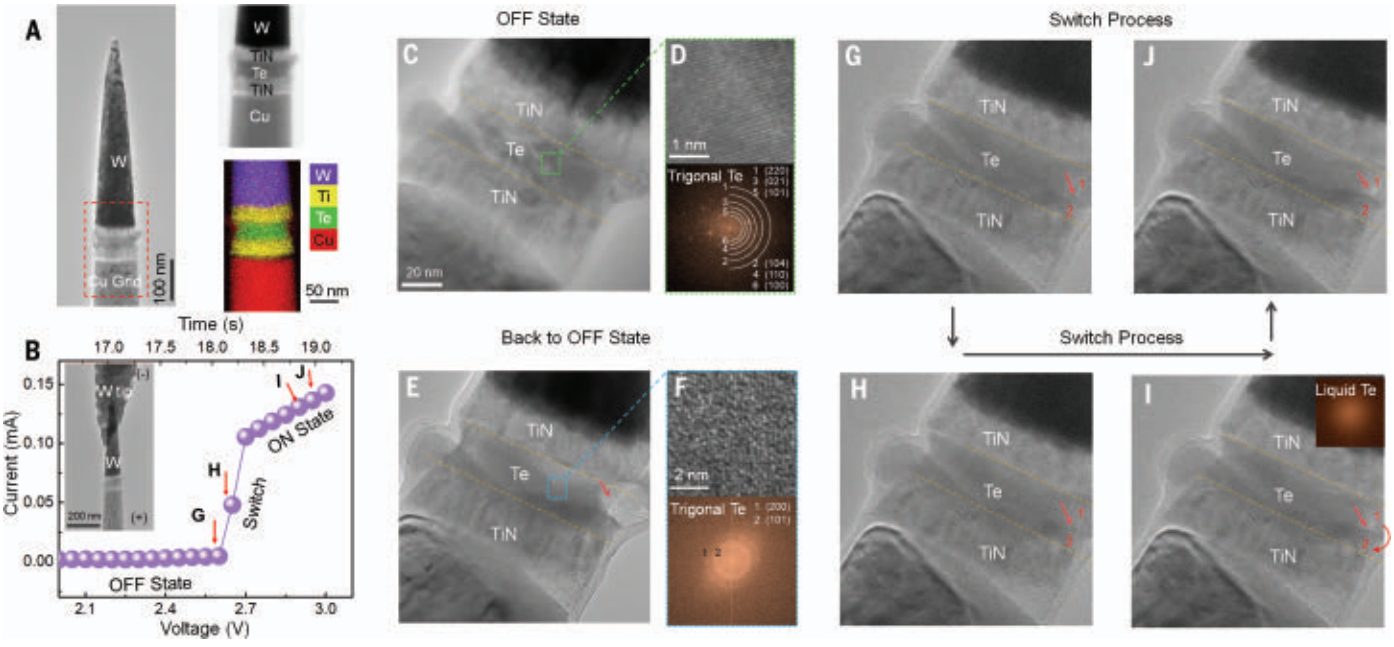


Fig. 4. Switching mechanism of the Te device obtained through in situ TEM observations. (**A**) TEM image of a TiN/Te/TiN nanodevice, resembling the T-shaped device. Enlarged view and corresponding mappings show that the nanodevice is a stack of Cu, TiN, Te, TiN, and W layers. (**B**) DC *I-V* curve of the Te nanodevice during switching. Critical time steps are also indicated by red arrows. A tungsten tip (cathode) is touching the nanodevice (anode), as shown in the inset, and applies the voltage. (**C** and **D**) Microstructure of the Te film in the nanodevice before switching. The Te film has a trigonal structure, identified by the FFT image (D). (**E** and **F**) Microstructure of the Te film after switching. Spontaneous recrystallization occurs after voltage removal. The recrystallized Te regrows into the same trigonal structure, yet with a different orientation (F). (**G** to **J**) Snapshots of TEM images (from movie S1) during switching. The initial Te device is in the OFF state with negligible passing current. As time proceeds, the corresponding currents are found [marked earlier (by red arrows) in (B)]. The Te film melts as the voltage exceeds 2.5 V [(H) to (J)], and the flow of the liquid causes a void in this unconfined nanodevice [red arrow in (E)].

performance of Te switches matches well with those of GST-based PCM cells, which do not deteriorate severely after series connection with PCM cells (fig. S15). The intrinsically homogeneous composition and BEOL compatibility of the Te switch not only cut the time-consuming material-optimization process but also enable aggressive size downscaling toward sub-10-nm dimensions for realizing 3D memory chips with the highest density.

REFERENCES AND NOTES

1. P. Zalden *et al.*, *Science* **364**, 1062–1067 (2019).
2. D. Ielmini, H.-S. P. Wong, *Nat. Electron.* **1**, 333–343 (2018).
3. F. Rao *et al.*, *Science* **358**, 1423–1427 (2017).
4. M. Wuttig, N. Yamada, *Nat. Mater.* **6**, 824–832 (2007).
5. B. Tallis, "Intel launches Optane Memory M.2 Cache SSDs for consumer market" (2017); www.anandtech.com/show/11227/intel-launches-optane-memory-m2-cache-ssds-for-client-market.
6. F. Xiong, A. D. Liao, D. Estrada, E. Pop, *Science* **332**, 568–570 (2011).
7. S.-H. Lee, Y. Jung, R. Agarwal, *Nat. Nanotechnol.* **2**, 626–630 (2007).
8. D. Kau *et al.*, in *2009 IEEE International Electron Devices Meeting (IEDM)* (2009), pp. 617–620.
9. W. Zhang, R. Mazzarello, M. Wuttig, E. Ma, *Nat. Rev. Mater.* **4**, 150–168 (2019).
10. M. Salinga *et al.*, *Nat. Mater.* **17**, 681–685 (2018).
11. S. Jia *et al.*, *Nat. Commun.* **11**, 4636 (2020).
12. M. Gill, T. Lowrey, J. Park, in *2002 IEEE International Solid-State Circuits Conference. Digest of Technical Papers (Cat. No.02CH37315)* (2002), pp. 202–203.
13. G. Servalli, in *2009 IEEE International Electron Devices Meeting (IEDM)* (2009), pp. 1–4.
14. A. Velea *et al.*, *Sci. Rep.* **7**, 8103 (2017).
15. J. Yoo *et al.*, *2018 IEEE Symp. VLSI Technol.* **20**, 207–208 (2018).
16. M. Zhu, K. Ren, Z. Song, *MRS Bull.* **44**, 715–720 (2019).
17. D. Garbin *et al.*, in *2019 IEEE International Electron Devices Meeting (IEDM)* (2019), pp. 823–826.
18. M. J. Lee *et al.*, *Nat. Commun.* **4**, 2629 (2013).
19. P. Noé *et al.*, *Sci. Adv.* **6**, eaay2830 (2020).
20. W. Wang *et al.*, *Sci. Rep.* **2**, 360 (2012).
21. D. Loke *et al.*, *Science* **336**, 1566–1569 (2012).
22. M. Anbarasu, M. Wimmer, G. Bruns, M. Salinga, M. Wuttig, *Appl. Phys. Lett.* **100**, 143505 (2012).
23. Y. Koo, S. Lee, S. Park, M. Yang, H. Hwang, *IEEE Electron Device Lett.* **38**, 568–571 (2017).
24. Materials and methods are available as supplementary materials.
25. H. Y. Cheng *et al.*, "Ultra-high endurance and low IOFF selector based on AsSeGe chalcogenides for wide memory window 3D stackable crosspoint memory" in *2019 IEEE International Electron Devices Meeting (IEDM)* (2019), pp. 37.3.1–37.3.4.
26. G. W. Burr *et al.*, *J. Vac. Sci. Technol. B Nanotechnol. Microelectron.* **28**, 223–262 (2010).
27. M. A. Zidan, J. P. Strachan, W. D. Lu, *Nat. Electron.* **1**, 22–29 (2018).
28. M. Salinga *et al.*, *Nat. Commun.* **4**, 2371 (2013).
29. B. Govoreanu *et al.*, *2017 IEEE Symp. VLSI Technol.* T92–T93 (2017).
30. H. Keller, J. Stuke, *Phys. Status Solidi B Basic Res.* **8**, 831–840 (1965).
31. B. Kramer, K. Maschke, L. D. Laude, *Phys. Rev. B* **8**, 5781–5793 (1973).
32. D. Ielmini, Y. Zhang, *J. Appl. Phys.* **102**, 054517 (2007).
33. P. R. Emtage, W. Tantraporn, *Phys. Rev. Lett.* **8**, 267–268 (1962).
34. X. Huang *et al.*, *Nano Lett.* **17**, 4619–4623 (2017).
35. M. Jałochowski, P. Mikołajczak, M. Subotowicz, *Phys. Status Solidi A Appl. Res.* **14**, K135–K137 (1972).
36. G. M. Matenoglou, L. E. Koutsokeras, P. Patsalas, *Appl. Phys. Lett.* **94**, 152108 (2009).
37. F. A. Blum Jr., B. C. Deaton, *Phys. Rev.* **137**, A1410–A1417 (1965).
38. Y. Wang *et al.*, *Nat. Electron.* **1**, 228–236 (2018).
39. Y. Du *et al.*, *Nano Lett.* **17**, 3965–3973 (2017).
40. Z. Zhang, Y. Guo, J. Robertson, *J. Appl. Phys.* **127**, 155301 (2020).
41. A. Kramer, M. L. Van de Put, C. L. Hinkle, W. G. Vandenberghe, *npj 2D Mater. Appl.* **4**, 10 (2020).
42. S. W. Nam *et al.*, *Science* **336**, 1561–1566 (2012).
43. A. S. Epstein, H. Fritzsche, K. Lark-Horovitz, *Phys. Rev.* **107**, 412–419 (1957).

ACKNOWLEDGMENTS

Funding: Financial support was provided by the National Key Research and Development Program of China (2017YFB0206101) and the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB44010000). M.Z. acknowledges support by the Hundred Talents Program (Chinese Academy of Sciences) and Shanghai Rising-Star Program (21QA1410800). **Author contributions:** Conceptualization: M.Z. Methodology: M.Z., J.S., and S.J. Investigation: J.S., S.J., N.S., Q.G., T.G., and S.L. Visualization: J.S. and M.Z. Funding acquisition: M.Z., Q.L., and Z.S. Project administration: M.Z., Q.L., and Z.S. Supervision: M.Z. Writing – original draft: M.Z. Writing – review & editing: M.Z., S.R.E., and R.D. **Competing interests:** The idea of a Te-based switch is protected by a Chinese patent (CN202010975902.2) and an international patent (PCT/CN2020/124585). **Data and materials availability:** All data needed to evaluate the conclusions in this paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi6332
Materials and Methods
Figs. S1 to S15
Tables S1 to S3
References (*44–54*)
Movies S1 and S2

20 April 2021; resubmitted 12 August 2021
Accepted 12 October 2021
10.1126/science.abi6332

OPTICAL IMAGING

Time-of-flight 3D imaging through multimode optical fibers

Daan Stellinga[1]†, David B. Phillips[2]†, Simon Peter Mekhail[1]†, Adam Selyem[3], Sergey Turtaev[4], Tomáš Čižmár[4,5], Miles J. Padgett[1]*

Time-of-flight three-dimensional (3D) imaging has applications that range from industrial inspection to motion tracking. Depth is recovered by measuring the round-trip flight time of laser pulses, typically using collection optics of several centimeters in diameter. We demonstrate near–video-rate 3D imaging through multimode fibers with a total aperture of several hundred micrometers. We implement aberration correction using wavefront shaping synchronized with a pulsed source and scan the scene at ~23,000 points per second. We image moving objects several meters beyond the end of an ~40-centimeters-long fiber of 50-micrometer core diameter at frame rates of ~5 hertz. Our work grants far-field depth-resolving capabilities to ultrathin microendoscopes, which we expect to have applications to clinical and remote inspection scenarios.

Multimode optical fibers (MMFs) represent an extremely efficient method of transporting light with a high spatial information density. They can support the propagation of thousands of spatial modes—i.e., optical field patterns that act as independent information channels—within a cross-sectional area similar to that of a human hair. These features have led to much interest in the deployment of MMFs as microendoscopes, enabling high-resolution imaging at the tip of a needle (*1*–*4*).

In this work, we extend the imaging capabilities of such microendoscopes to include depth information. Three-dimensional (3D) imaging through ultrathin MMFs promises an array of new applications, including the 3D inspection of the internal chambers of objects that are difficult to open, such as jet engines or nuclear reactors, and the 3D visualization of hollow viscous organs, which could help surgeons navigate inside the body during operations.

However, the compact form of MMFs comes at a cost: Optical signals are subject to modal dispersion, typically causing input coherent light patterns to be unrecognizably scrambled into speckle patterns at the output facet, formed entirely from nonballistic light that has scattered multiple times from the core-cladding interface (*5*–*7*). Fortunately, as long as a MMF remains in a fixed configuration, the scrambling process is deterministic and unchanging in nature, so a given input field will always produce the same output field. This means that the way a static MMF scrambles light can be represented by a linear matrix operator, known as a transmission matrix (TM), which maps any possible input field to the resulting output (*8*–*13*).

Measurement of the TM enables the calculation of how an input field should be preshaped to generate a desired output—for example a spot focused to a particular location. This method is known as wavefront shaping (*14*, *15*); thus, by illuminating the proximal facet with a sequence of carefully prepared input light fields, a focused spot can be raster scanned across the distal facet of a MMF. Scanning-based imaging can then be achieved by recording the total intensity of light returning through the fiber and correlating this with the position of the focus (*16*, *17*). Recently, wavefront shaping through MMFs has been employed for in vivo imaging of neurons deep inside the brains of mice (*2*–*4*)—an endeavor very challenging to achieve in any other such minimally invasive way. The imaging of objects some distance from the distal facet of a MMF is also possible (*9*, *18*). However, this is more demanding because the level of return signal falls off rapidly in proportion to the square of the object's distance.

Here, we augment MMF microendoscopy with time-of-flight (TOF) light detection and ranging (LIDAR) techniques to provide depth information alongside 2D reflectance images (*19*). TOF techniques recover depth by measuring the round-trip flight time of a laser pulse reflecting from the scene. To achieve this, we implement high-speed wavefront shaping synchronized with a subnanosecond pulsed laser source. Light fields entering the MMF are shaped using a digital micromirror device (DMD) operating at 22.7 KHz. We use knowledge of the TM linking the proximal (near) end of the MMF to the far field of the distal facet to raster scan the pulsed source across the field of view.

Extending wavefront shaping through MMFs from continuous-wave to pulsed illumination presents an additional complication—the potential for temporal pulse distortion as a result of chromatic and spatial mode dispersion. In our system, the latter form of dispersion dominates. Spatial mode dispersion can be understood by considering that an incident pulse will excite different spatial modes supported by the fiber, which travel at slightly different velocities (*20*, *21*).

We must ensure that the optical path difference accumulated by these modes as they propagate through the fiber remains below the coherence length of the pulse. This places a constraint on the temporal length of the pulse τ_p that can be used with a given geometry of step-index fiber, leading to (see supplementary materials, section 7 for derivation) the following

$$\tau_p \gg \frac{\mathrm{NA}^2 L}{c n_c} \tag{1}$$

where L is the fiber length, NA is the numerical aperture of the fiber, n_c is the refractive index of the core, and c is the speed of light in a vacuum. Here, we use a step-index fiber of $L \sim 0.4$ m, NA = 0.22, $n_c = 1.46$, and core radius $a = 25$ μm. Therefore, Eq. 1 yields $\tau_p \gg 44$ps. To ensure minimal pulse distortion, we choose a laser with a pulse duration of $\tau_p \sim 700$ ps, with central wavelength $\lambda = 532$ nm. We note that Eq. 1 provides a bound on the minimum pulse duration to achieve arbitrary wavefront shaping at the output without temporal distortion. However, each far-field focal spot only requires the excitation of spatial fiber modes with very similar phase velocities, so for this special case it may be possible to use a source with substantially shorter pulse duration than that implied by Eq. 1. We also note that the use of a graded-index fiber would place a lower limit on τ_p. Our fiber supports $N \sim (\pi a \mathrm{NA}/\lambda)^2 \sim 1000$ spatial modes per polarization, setting the lateral resolution to $\sim 4N = 4000$ independently resolvable features within each image, according to the Raleigh criterion (supplementary materials, section 8) (*18*, *22*).

Figure 1 shows a schematic of our prototype system and some example depth images. Each frame was recorded by raster scanning 4200 points in 200 ms. A single laser pulse is delivered to each spatial position in the field of view. A second fiber, placed alongside the illumination fiber, is used to collect the backscattered light. This collection fiber has a larger core diameter of 500 μm to increase the collection efficiency and thereby the working distance of the endoscope. The return signal from each image pixel is coupled directly to an avalanche photodiode (APD) and converted to a TOF histogram, referenced to the time that the outgoing pulse entered the illumination fiber.

[1]School of Physics and Astronomy, University of Glasgow, Glasgow G12 8QQ, UK. [2]School of Physics and Astronomy, University of Exeter, Exeter EX4 4QL, UK. [3]Fraunhofer Centre for Applied Photonics, Glasgow G1 1RD, UK. [4]Leibniz Institute of Photonic Technology, Albert-Einstein-Straße 9, 07745 Jena, Germany. [5]Institute of Scientific Instruments of the CAS, Královopolská 147, 612 64 Brno, Czech Republic.
*Corresponding author. Email: miles.padgett@glasgow.ac.uk
†These authors contributed equally to this work.

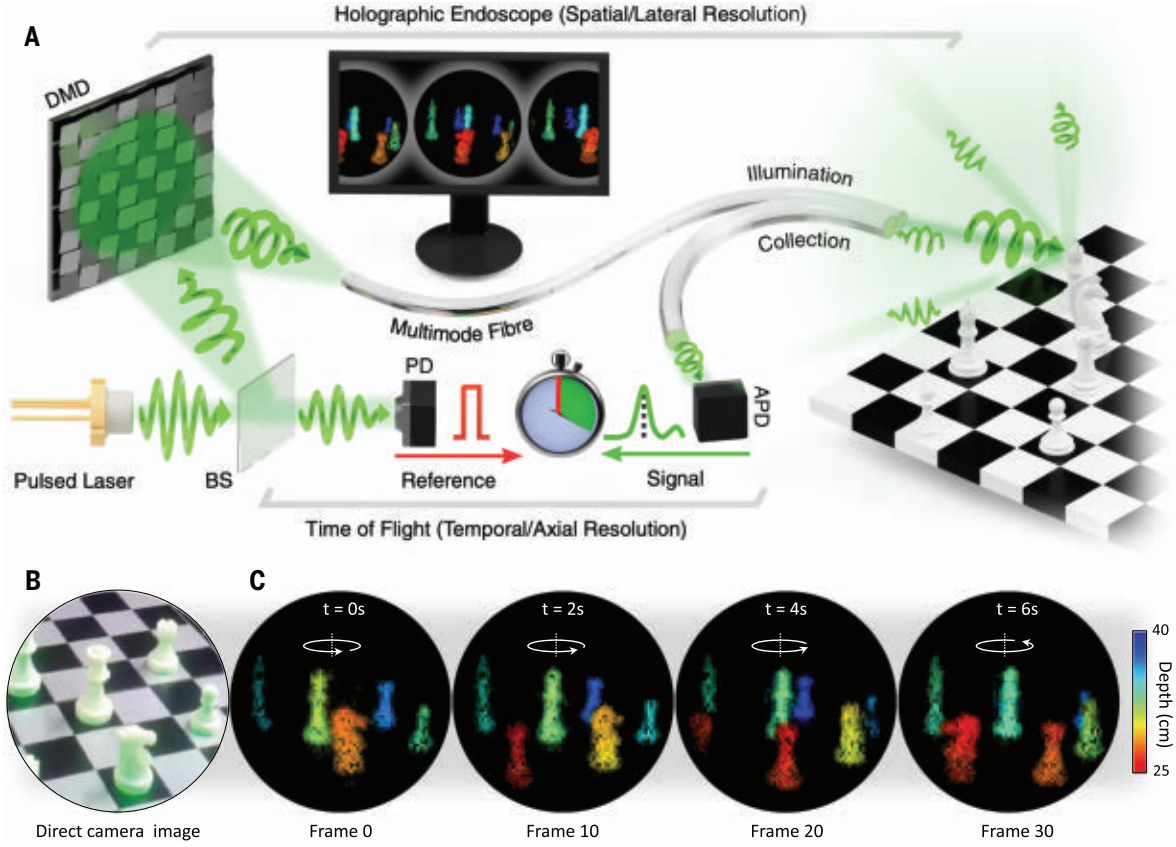


Fig. 1. Endoscopic LIDAR. (**A**) A schematic of the experimental setup. See supplementary materials, sections 1 to 3, for a detailed description of the TM acquisition, imaging procedure, and data processing. BS, beamsplitter; PD, photodiode. (**B**) A snapshot of the true scene being recorded. (**C**) Typical depth-resolved images obtained with our system. Each frame is captured in 200 ms. The frames show the pieces on a revolving chess board located at a depth of ~30 cm from the distal fiber facet, recorded at a frame rate of 5 Hz (supplementary materials, section 12, and movie S1). Scene depth is encoded in the color channels, and scene reflectivity is encoded in the transparency channel—thus, regions of the scene with low levels of return signal, and consequently a poorly estimated depth, are displayed with low brightness. The dark spots in the images are the result of singularities in the speckle reference used to measure the TM.

The histograms have 400-ps time bins, short enough to resolve the 700-ps pulses.

Depth information is recovered by locating the temporal position of the histogram peak, δt, which can be achieved with sub–time bin precision (here, we use Fourier up-sampling by a factor of 32; see supplementary materials, section 5 for details). The depth of each pixel, D, is a function of both δt and the angular coordinate of the pixel with respect to the optical axis of the fiber, θ

$$D(\delta t, \theta) = \frac{1}{2} c \delta t \cos\theta - n_c L \cos\theta \left(1 - \frac{\sin^2\theta}{n_c^2}\right)^{-\frac{1}{2}} \quad (2)$$

The first term on the right-hand side of Eq. 2 accounts for the different optical path lengths traveled by pulses in free space, whereas the latter term accounts for differences accumulated inside the fibers. See supplementary materials, section 4 for a derivation.

Figure 2 shows snapshots from movies of several dynamic scenes. Figure 2A shows consecutive frames of a swinging pendulum, and the depth map tracks the 3D motion of both the pendulum bob and its thread. Figure 2, B to E, shows scenes with more dynamic motion—some of the authors moving around at a progressively increasing distance from the distal facet, up to 2.5 m away. The full movies are available in the supplementary materials (supplementary materials, section 12, and movies S2 to S4). The depth precision of our system is largely determined by the level of return signal from each pixel, rendering it highly dependent on the reflectivity and depth of objects in the scene. We estimate a typical depth precision of ~1 cm in the results shown in Fig. 2.

To more-quantitatively assess the imaging performance of our prototype, we measure the depth precision (Fig. 3A) and the signal-to-noise ratio of the associated reflectivity maps (Fig. 3B) as a function of scene depth and radial position within the field of view. In these measurements, the optimal depth precision found was ~2.5 mm at a scene distance of 0.4 m, dropping off slightly with distance to ~6 mm at 2 m.

Because the number of resolvable features is fixed by the number of spatial modes supported by the MMF, both the lateral resolution and the diameter of the field of view grow linearly in proportion to the distance to the scene, l. The angular resolution θ_r is therefore constant with depth. We experimentally measure an angular resolution of ~16 mrad, close to the theoretical value $\theta_r \sim 0.61\lambda/a \sim 13$ mrad. The spot contrast ratio, defined as the ratio of power in the spot to the total projected power, is ~0.4. Both resolution and contrast are relatively constant across the field of view, with only a slight decrease toward the edges. Section 9 in the supplementary materials gives more details of these measurements.

There are several ways that the image quality can be enhanced in our current prototype system if necessary. Although τ_p does not directly limit the depth precision, shorter pulses, providing they can be properly sampled, will yield a higher depth precision. Images acquired

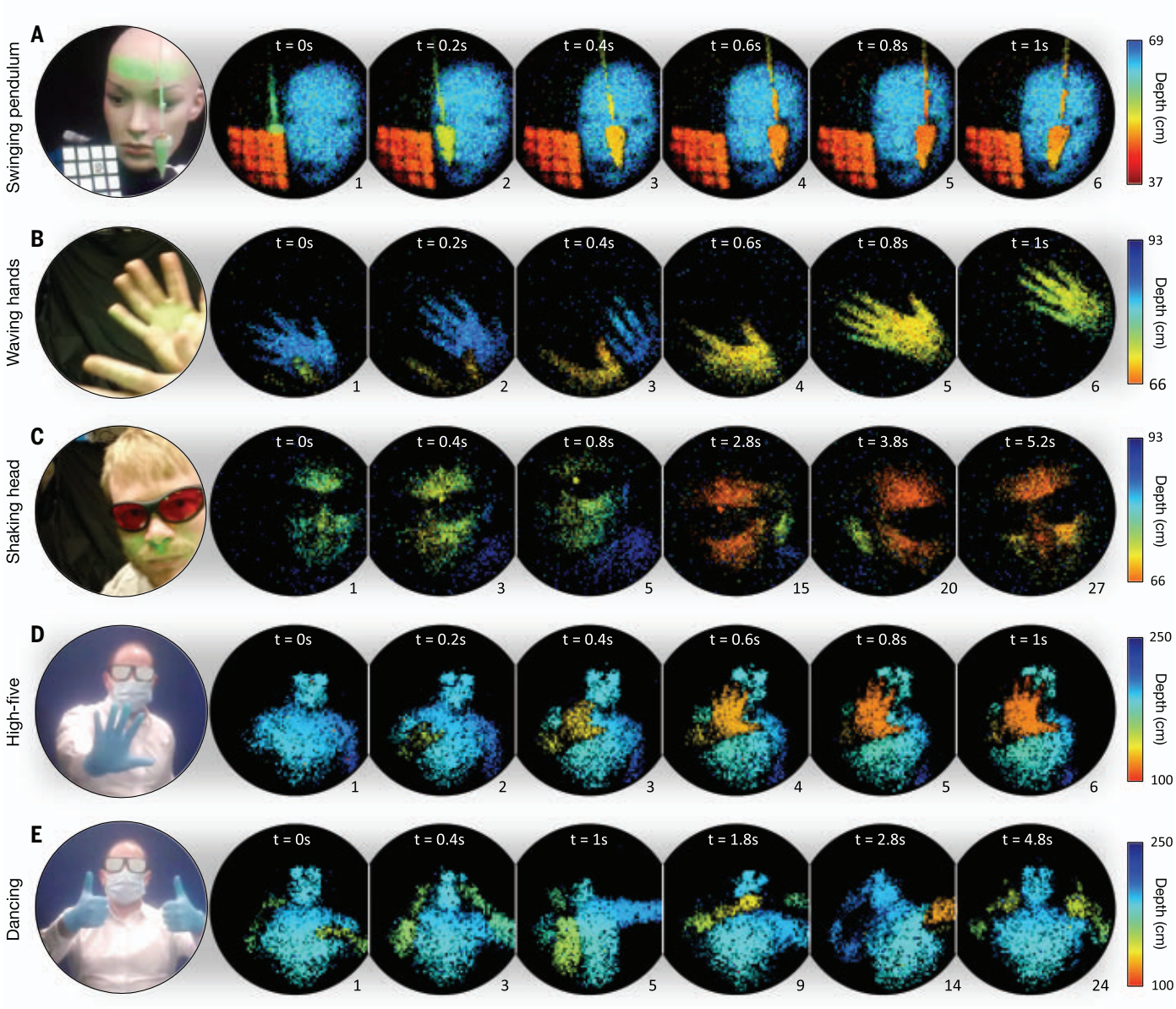


Fig. 2. Snapshots from depth-resolved movies at progressively increasing scene depths. Left-hand column shows a direct camera image of each scene. (**A**) Consecutive frames of a pendulum of ~1 m in length swinging adjacent to a mannequin head and a Rubik's cube at 37 to 69 cm (movie S2). A small amount of rolling shutter effect is visible, causing the bob to appear to point in the direction of motion as a result of the raster scan reaching the tip slightly later in the swing than the base. (**B** and **C**) Author Daan Stellinga waving his hands and shaking his head at 66 to 93 cm (movie S3). (**D** and **E**) Author Peter Mekhail giving a high-five and dancing at 1 to 2.5 m (movie S4). Here, human subjects wore eye protection, although we estimate the maximum average power illuminating an exposed pupil is well below eye-safety limits in our prototype (supplementary materials, section 11). The relative time is given at the top of each frame. The frame number is given at the bottom right of each frame.

at video rates, like those shown in Fig. 1 and Fig. 2, are reconstructed by assuming that the scene was illuminated with an ideally focused spot. However, if the actual projected patterns are measured during the calibration phase and do not change, a more sophisticated reconstruction algorithm can be used to incorporate this information. We refer to this hybrid method as TM-guided computational imaging: Signal-to-noise ratio is maximized by using the TM to concentrate power into a focused spot, and regularized matrix inversion is then used to account for residual background speckle. Figure 4 depicts a series of views of a static 3D image of a mannequin head reconstructed from data recorded over 2 s using TM-guided computational imaging. The depth error has been reduced to an extent that enables the contours of the face to be revealed. See supplementary materials, section 10, for more details about this reconstruction method.

Finally, we consider our work in the context of alternative 3D imaging approaches. There is an emerging class of clinical microendoscopes capable of recovering high-resolution depth images within tissue using optical coherence tomography (OCT) (*23*, *24*). These endoscopes range down to ~500 μm in diameter and typically deliver images with an

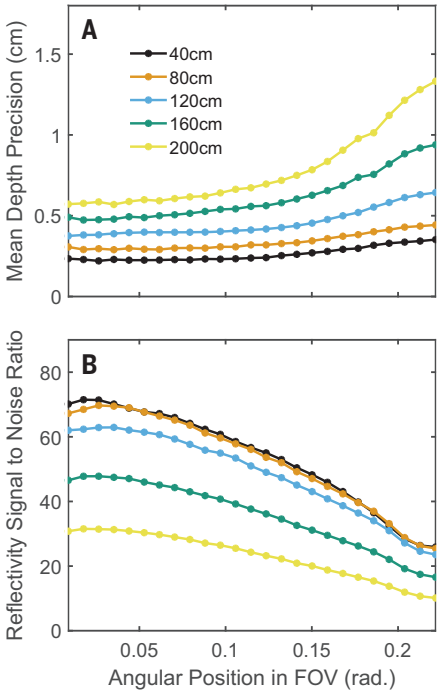


Fig. 3. System characterization. (**A**) Depth precision given by the standard deviation of repeated depth measurements and plotted as a function of scene depth and radial position in the field of view. Here, we used a movable flat white screen as a target object. (**B**) Signal-to-noise ratio of the associated reflectivities. FOV, field of view.

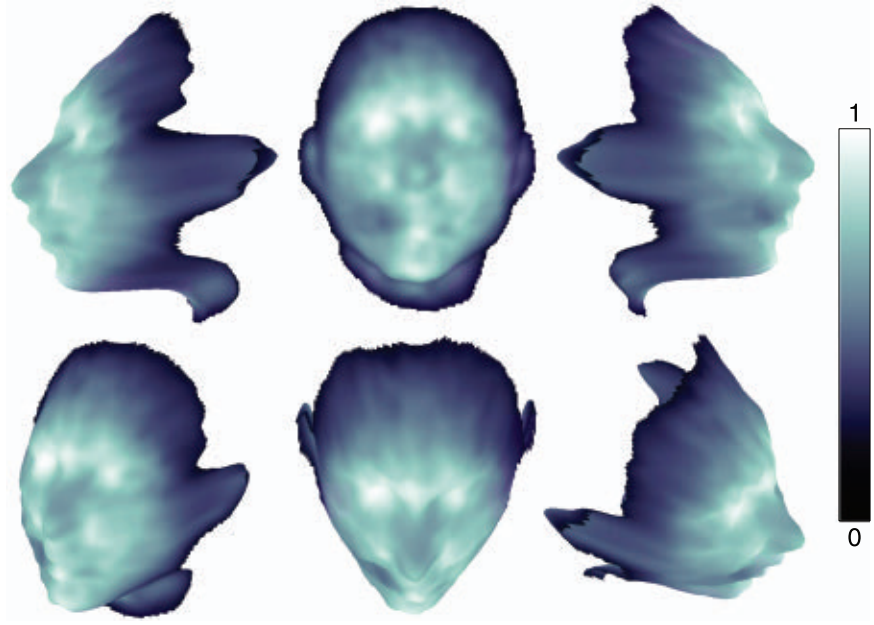


Fig. 4. 3D imaging with enhanced fidelity. A series of views of a 3D reconstruction of a mannequin head located at a range of 30 cm. Here, depth is represented as the surface profile, and reflectivity is rendered as the surface shading. The data for this reconstruction were acquired in 2 s.

axial resolution of ~10 μm. Light is delivered through a single mode fiber, and many designs feature bespoke micro-optics at the distal facet, giving a side view of the sample. Images are constructed pixel by pixel by mechanically retracting the endoscope in a spiral motion, which enables 3D tomography of the tissue surrounding the endoscope. Rather than imaging the cross section of vessels and similar structures, our system is designed to operate in a different regime, namely to image more-distant surfaces in the far field of the fiber output at near–video frame rates. As such, our system has a greater depth range, limited only by collection efficiency (*18*), and yet lower axial resolution, which is constrained by the shortest coherence length of the light it is possible to transmit through MMFs without spatiotemporal compensation (see Eq. 1). The use of a MMF that supports many spatial modes also enables high-speed DMD-based point scanning rather than mechanical scanning, allowing for high–frame rate operation while sacrificing flexibility of the endoscope itself. Emerging techniques to monitor the TM of flexible MMFs in real time with access to only the proximal end have potential to overcome this restriction (*25*, *26*).

There are also several methods to image in three dimensions through scattering media under development. Time-gated LIDAR is a well-established technique to see through obscurants by filtering out scattered light (by photon arrival time) to recover a signal dominated by ballistic photons that have traveled straight through the occluding object (*27*). More recently, non–line-of-sight 3D imaging around corners has been demonstrated, relying on streak or single-photon avalanche diode (SPAD) cameras to overcome signal mixing caused by several diffuse reflections from opaque walls (*28*).

Our approach is fundamentally different from these methods. Our measurements are formed purely from light that has forward scattered many times through a MMF, with no ballistic component remaining. We extract image data from this nonballistic light by precharacterization of a TM that enables the spatial control of the incident field to guide it through the MMF to its target. Our work also complements recent demonstrations of 3D imaging through thick randomly scattering media, in which diffuse scattering effects are inverted using ultrafast pulse detection (*29*) and optical sectioning through larger form factor multicore fibers using in situ distal holography (*30*). Our prototype TOF-based 3D imaging system provides an exciting new modality to MMF-based microendoscopy, with many potential applications to remote inspection and biomedical imaging in the life sciences.

REFERENCES AND NOTES

1. Y. Choi *et al.*, *Phys. Rev. Lett.* **109**, 203901 (2012).
2. S. Ohayon, A. Caravaca-Aguirre, R. Piestun, J. J. DiCarlo, *Biomed. Opt. Express* **9**, 1492–1509 (2018).
3. S. Turtaev *et al.*, *Light Sci. Appl.* **7**, 92 (2018).
4. S. A. Vasquez-Lopez *et al.*, *Light Sci. Appl.* **7**, 110 (2018).
5. J. W. Goodman, *J. Opt. Soc. Am.* **66**, 1145 (1976).
6. M. Plöschner, T. Tyc, T. Čižmár, *Nat. Photonics* **9**, 529–535 (2015).
7. M. C. Velsink, Z. Lyu, P. W. H. Pinkse, L. V. Amitonova, *Opt. Express* **29**, 6523–6531 (2021).
8. S. M. Popoff *et al.*, *Phys. Rev. Lett.* **104**, 100601 (2010).
9. T. Cižmár, K. Dholakia, *Nat. Commun.* **3**, 1027 (2012).
10. S. Popoff, G. Lerosey, M. Fink, A. C. Boccara, S. Gigan, *Nat. Commun.* **1**, 81 (2010).
11. I. N. Papadopoulos, S. Farahi, C. Moser, D. Psaltis, *Opt. Express* **20**, 10583–10590 (2012).
12. R. Di Leonardo, S. Bianchi, *Opt. Express* **19**, 247–254 (2011).
13. S. Li *et al.*, *Light Sci. Appl.* **10**, 88 (2021).
14. I. M. Vellekoop, A. P. Mosk, *Opt. Lett.* **32**, 2309–2311 (2007).
15. A. P. Mosk, A. Lagendijk, G. Lerosey, M. Fink, *Nat. Photonics* **6**, 283–292 (2012).
16. I. Gusachenko, M. Chen, K. Dholakia, *Opt. Express* **25**, 13782–13798 (2017).
17. J. Trägårdh *et al.*, *Opt. Express* **27**, 30055–30066 (2019).
18. I. T. Leite, S. Turtaev, D. E. Boonzajer Flaes, T. Čižmár, *APL Photonics* **6**, 036112 (2021).
19. B. Schwarz, *Nat. Photonics* **4**, 429–430 (2010).
20. J. Carpenter, B. J. Eggleton, J. Schröder, *Laser Photonics Rev.* **11**, 1600259 (2017).
21. W. Xiong, C. W. Hsu, H. Cao, *Nat. Commun.* **10**, 2973 (2019).
22. R. N. Mahalati, R. Y. Gu, J. M. Kahn, *Opt. Express* **21**, 1656–1668 (2013).
23. J. G. Fujimoto *et al.*, *Nat. Med.* **1**, 970–972 (1995).
24. J. Li *et al.*, *Light Sci. Appl.* **9**, 124 (2020).
25. S. Li, S. A. R. Horsley, T. Tyc, T. Čižmár, D. B. Phillips, *Nat. Commun.* **12**, 3751 (2021).
26. G. S. Gordon *et al.*, *Phys. Rev. X* **9**, 041050 (2019).
27. L. Wang, P. P. Ho, C. Liu, G. Zhang, R. R. Alfano, *Science* **253**, 769–771 (1991).
28. A. Velten *et al.*,*Nat. Commun.* **3**, 745 (2012).

29. D. B. Lindell, G. Wetzstein, . *Nat. Commun.* **11**, 4517 (2020).
30. N. Badt, O. Katz, Label-free video-rate micro-endoscopy through flexible fibers via Fiber Bundle Distal Holography (FiDHo). arXiv:2102.06482 [physics.optics] (2021).
31. D. Stellinga *et al*., Time of flight 3D imaging through multimode optical fibres, Enlighten: Research Data (Univ. Glasgow, 2021); https://doi.org/10.5525/gla.researchdata.1199.

ACKNOWLEDGMENTS

The authors thank I. Leite for useful discussions and S. Johnson and G. Gibson for practical advice in the laboratory. **Funding:** D.S., S.P.M., and M.J.P. acknowledge financial support from the Horizon 2020 project QSORT (766970) and Quantic (EP/M01326X/1). D.B.P. acknowledges financial support from the Royal Academy of Engineering and the ERC (804626). S.T. and T.Č. acknowledge financial support from MEYS, the ERDF (CZ.02.1.01/0.0/0.0/15_003/0000476), and the ERC (724530). M.J.P. thanks the Royal Society for financial support. **Author contributions:** D.S.: conceptualization, formal analysis, investigation, methodology, and writing – original draft; D.B.P.: conceptualization, formal analysis, visualization, and writing – original draft; S.P.M.: formal analysis, visualization, investigation, methodology, and writing – original draft; A.S.: software; S.T.: methodology, software, and visualization; T.Č.: methodology, conceptualization, formal analysis, writing – review and editing, funding acquisition, and supervision; M.J.P.: methodology, conceptualization, formal analysis, writing – review and editing, funding acquisition, and supervision. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data are available at (*31*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl3771
Materials and Methods
Figs. S1 to S4
References (*32*, *33*)
Movies S1 to S4

9 July 2021; accepted 19 October 2021
10.1126/science.abl3771

GRAPHENE

Detection of graphene's divergent orbital diamagnetism at the Dirac point

J. Vallejo Bustamante[1], N. J. Wu[1,6], C. Fermon[2], M. Pannetier-Lecoeur[2], T. Wakamura[1,7], K. Watanabe[3], T. Taniguchi[4], T. Pellegrin[1], A. Bernard[1], S. Daddinounou[1], V. Bouchiat[5], S. Guéron[1], M. Ferrier[1], G. Montambaux[1], H. Bouchiat[1]*

The electronic properties of graphene have been intensively investigated over the past decade. However, the singular orbital magnetism of undoped graphene, a fundamental signature of the characteristic Berry phase of graphene's electronic wave functions, has been challenging to measure in a single flake. Using a highly sensitive giant magnetoresistance (GMR) sensor, we have measured the gate voltage–dependent magnetization of a single graphene monolayer encapsulated between boron nitride crystals. The signal exhibits a diamagnetic peak at the Dirac point whose magnetic field and temperature dependences agree with long-standing theoretical predictions. Our measurements offer a means to monitor Berry phase singularities and explore correlated states generated by the combined effects of Coulomb interactions, strain, or moiré potentials.

Orbital magnetism results from the quantum motion of electrons in a magnetic field. At low energy, this motion leads to the Landau spectrum, which is, in most two-dimensional (2D) conductors, a harmonic oscillator–type spectrum with equally spaced levels separated by the cyclotron energy $\hbar\omega_c$ (*1*). As long as the material is nonsuperconducting, this spectrum causes a very small diamagnetic low-field susceptibility that is usually hidden by spin contributions. However, some materials, such as graphene, can display extraordinarily large diamagnetism. This was predicted in the theoretical work of McClure (*2*), who showed that graphene is diamagnetic at half filling (at the so-called Dirac point), with a divergent zero-field susceptibility (the derivative of the magnetization M with respect to the magnetic field B),

$$\chi_0(\mu) = \frac{\partial M}{\partial B} = -\frac{2e^2 v_F^2}{3\pi}\delta(\mu) \qquad (1)$$

where v_F is the Fermi velocity, e is the electronic charge, and the Fermi energy μ is zero at the Dirac point. This is all the more surprising because the density of states is zero at that point. The reason for this singular susceptibility stems from the electron-hole symmetric linear spectrum of Dirac relativistic electrons, which gives rise to a Landau spectrum quantized as $\pm\sqrt{nB}$ where n is a positive integer. The diamagnetic sign of the response is attributable to the existence of the zero-energy Landau level ($n = 0$), as recalled and sketched below [see also figure 5 of (*3*) and related comment]. This peculiar level is known to result from the Berry phase (*4*) of π acquired by the wave function pseudo-spin upon a revolution around a Dirac cone in reciprocal space (*5*). Therefore, the diamagnetic sign of the susceptibility at the Dirac point is a direct consequence of the π Berry phase. Indeed, it has been shown that slightly different models with a zero Berry phase lead to orbital paramagnetism at the Dirac point (*3*). To summarize, the divergence reflects the linear spectrum and the diamagnetic sign reflects the nontrivial geometry of the eigenstates via the Berry phase (*3*).

However, despite these striking predictions, the singular orbital magnetism of a single graphene flake remains undetected. The reason for this lies in at least three obvious experimental challenges. First, the magnetic signal of an atomic monolayer is extremely small. Second, the McClure singularity, originally computed for an ideal system without disorder at zero temperature and in the limit of zero magnetic field, is rounded when any of these conditions is relaxed (*6*–*9*). Finally, this orbital magnetism is expected to be hidden by the magnetism of spins originating from edges, vacancies, or impurities (*10*), which tends to become dominant at low temperatures. This may explain why magnetization measurements have to date only been performed on a macroscopic number of graphene flakes. In one case (*11*), the focus was mainly on the spin paramagnetism of induced vacancy– and resonant states–type defects, which were found to depend on the chemical doping of the samples. A second set of measurements (*12*) did focus on the diamagnetism, and found a diamagnetism larger than that of pure graphite by a factor of 3. The magnetization curves at high fields were found to be compatible with the $\sqrt{B}$ dependence predicted for the Dirac spectrum. However, in those experiments it was not possible to fix the doping, nor could the residual contribution of paramagnetic spins along the edges of the flakes be well controlled (*13*).

In the present experiment, by contrast, we measure the orbital moment of a single flake whose Fermi energy is precisely controlled. This is achieved by implementing several sensitivity-enhancing features detailed in (*14*). As shown in Fig. 1, our experiment consists of a graphene monolayer, encapsulated between two hexagonal boron nitride (hBN) 2D crystals, capacitively coupled to a top-gate electrode and positioned above a highly sensitive magnetic detector made of two giant magnetoresistance (GMR) strips (figs. S1 to S3) in a Wheatstone bridge configuration. One key asset is that whereas graphene's orbital magnetism responds to a field perpendicular to the graphene plane ("vertical" field), the resistance of the GMR detectors only depends on the in-plane field, and thus detects

[1]Université Paris-Saclay, CNRS, Laboratoire de Physique des Solides, 91405 Orsay, France. [2]SPEC, CEA, CNRS, Université Paris-Saclay, 91191 Gif-sur-Yvette, France. [3]Research Center for Functional Materials, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [4]International Center for Materials Nanoarchitectonics, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [5]Néel Institute, CNRS, 38000 Grenoble, France. [6]Université Paris-Saclay, CNRS, Institut des Sciences Moléculaires d'Orsay, Orsay, France. [7]NTT Basic Research Laboratories, NTT Corporation, Atsugi, Kanagawa, Japan.
***Corresponding author. Email: helene.bouchiat@u-psud.fr**

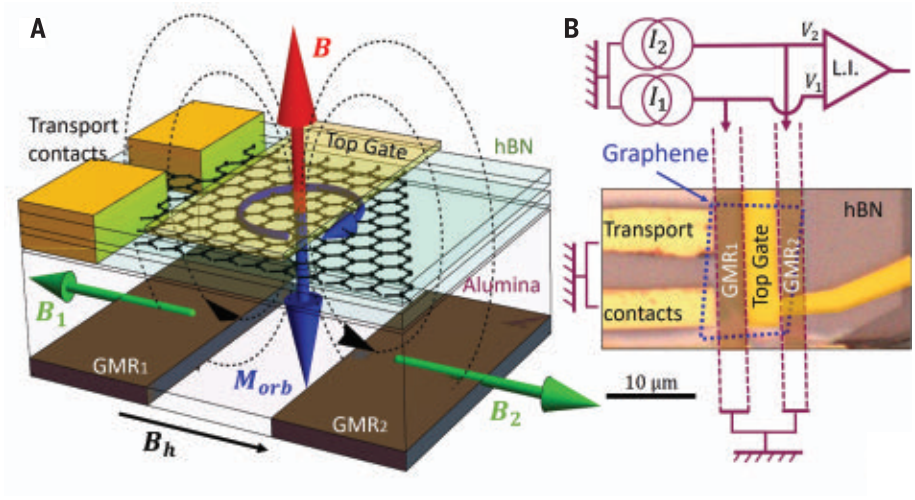


Fig. 1. Experimental setup. (**A**) Principle of the experiment. The orbital magnetization M_{orb} can be viewed as a current loop (blue circle) generated by a vertical magnetic field B and circulating around the graphene region covered by the gate electrode. It is detected by the two GMR detectors, which measure the horizontal components B_1 and B_2 (respectively on the detectors GMR_1 and GMR_2) of the magnetic field (black dashed lines) generated by this loop. The sensitivity is on the order of 0.1 nT (*14*). (**B**) Micrograph of the sample investigated; the gate voltage derivative of the orbital magnetization is measured via the difference between the DC current–biased GMR_1 and GMR_2 resistances with I_1 and I_2 adjusted so as to cancel the DC component of the voltage difference $V_1 - V_2$. The signal measured by a lock-in amplifier (L.I.) is the AC component of $V_1 - V_2$ at the modulation frequency of the gate voltage. There is no current applied to the graphene sample during the magnetization measurements.

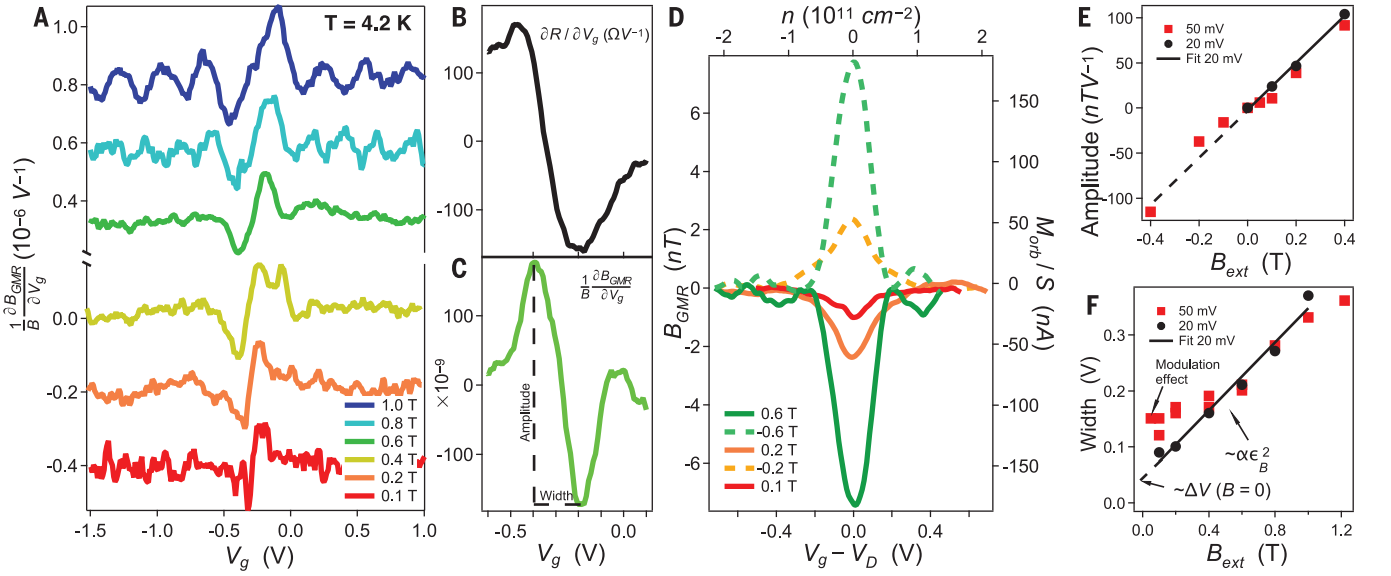


Fig. 2. Magnetization data. (**A**) Detected modulation of the GMR detector's resistance with an AC gate voltage modulation of 20 mV, as a function of the DC gate voltage V_g. The quantity plotted is $\partial B_{GMR}/\partial V_g$, where B_{GMR} is deduced from the signal on the calibrated GMR sensor divided by the applied vertical magnetic field B. Data are the average of 80 independent measurements. (**B**) Derivative with respect to gate voltage of the two-point resistance of graphene measured through the side electrodes, in the region of the Dirac point, with a gate voltage modulation of 50 mV. (**C**) For comparison, the GMR signal at −0.6 T using the same gate voltage modulation as in (B). The GMR peak is much narrower. (**D**) Numerical integration of the data plotted in (A) and fig. S4, yielding the magnetization per unit surface (in nA; right axis) and the magnetic field B_{GMR} detected by the GMR device (in nT; left axis) as a function of the gate voltage. (**E** and **F**) Field dependences of the GMR peak maxima and widths, as defined in (C), for gate voltage modulations of 20 mV (circles) and 50 mV (squares), and comparison with the linear variations expected theoretically (see Eqs. 5 and 6 and eqs. S20 to S27), using the scaling between the gate voltage and the square of the Landau energy ε_B^2 via the parameter α defined in Eq. 8. Deviations from linearity caused by excessive modulation amplitudes are visible for a 50-mV modulation.

the horizontal component of the field generated by the orbital current loops in the graphene (Fig. 1), all the while being insensitive to the applied vertical field. A second feature is the addition of a small AC modulation to the DC gate voltage, which in turn modulates the magnetization with respect to gate voltage and thus the resistance of the GMR detector. Beyond increasing the sensitivity, this modulation technique makes gate-independent magnetic signals invisible. Thanks to these experimental implementations, we were able to detect the derivative with respect to gate voltage of the diamagnetic McClure peak at low magnetic fields. We have also measured the crossover to the de Haas–van Alphen magnetic oscillations at higher fields.

Figure 2 shows the gate voltage derivative of the field induced by the graphene sample on the calibrated GMRs as a function of V_g for perpendicular magnetic fields between 0.1 and 1.2 T. We found an antisymmetric peak centered at V_g = −0.29 V, which we identified as the Dirac point by comparing to the position of the maximum in the resistance of the sample $R(V_g)$ (Fig. 2B and fig. S3). At low magnetic

fields, the antisymmetric peak detected in the GMR resistance is directly proportional to the derivative of the McClure peak with respect to the chemical potential (controlled by the gate voltage), as detailed in (*14*). The experimental detection of this peak and its evolution with magnetic field are the central result of our work. Both the peak width and amplitude increase linearly with field, as shown in Fig. 2, E and F. Above 0.6 T, $\partial M/\partial V_g(V_g)$ displays periodic oscillations in addition to the antisymmetric peak around the Dirac point. These oscillations are related to the expected de Haas–van Alphen oscillations of the magnetization, as discussed below.

The magnetization, shown in Fig. 2D, is obtained by the integration of the curves in Fig. 2A. The peak amplitude translates into a few nanoteslas induced in the GMR plane by graphene's orbital response to a 0.1-T perpendicular field. This illustrates the sensitivity of our experiment. The correspondence between this detected field, B_{GMR}, and magnetization is obtained by modeling the orbital magnetic moment as an effective current loop whose geometry is defined by the gated region of graphene (fig. S10). We find that positive magnetic fields produce a negative peak in magnetization, and vice versa, which is consistent with the expected diamagnetic response of graphene (*2*). The sign of the response was carefully determined via the sign of the response of the GMR sensor to a horizontal field of known orientation. We can assert that the signal cannot be attributed to gate voltage–dependent magnetism of paramagnetic impurities, given the absence of temperature dependence between 4.2 and 40 K (*15*) (see fig. S9). In addition, thanks to our gate modulation technique, we can exclude spurious contributions from impurities or defects in alumina or graphene, which would not depend on gate voltage. This contrasts with all previous measurements of graphene's magnetism, which were performed on large ensembles of flakes.

In the following, we compare our results to theoretical predictions, taking into account the variations of the chemical potential caused by charge inhomogeneity, and ignoring the smaller broadening due to temperature (*14*). Assuming a Gaussian distribution for the electrochemical potential μ' of standard deviation σ,

$$P_\sigma(\mu') = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\mu'^2}{2\sigma^2}\right) \qquad (2)$$

yields a smoothed susceptibility,

$$\chi_\sigma(\mu) = \int P_\sigma(\mu')\chi_0(\mu-\mu')d\mu' \qquad (3)$$

Then, the δ-peak of the susceptibility is broadened as

$$\chi_\sigma(\mu) = -\frac{2e^2v_F^2}{3\pi}P_\sigma(\mu) \qquad (4)$$

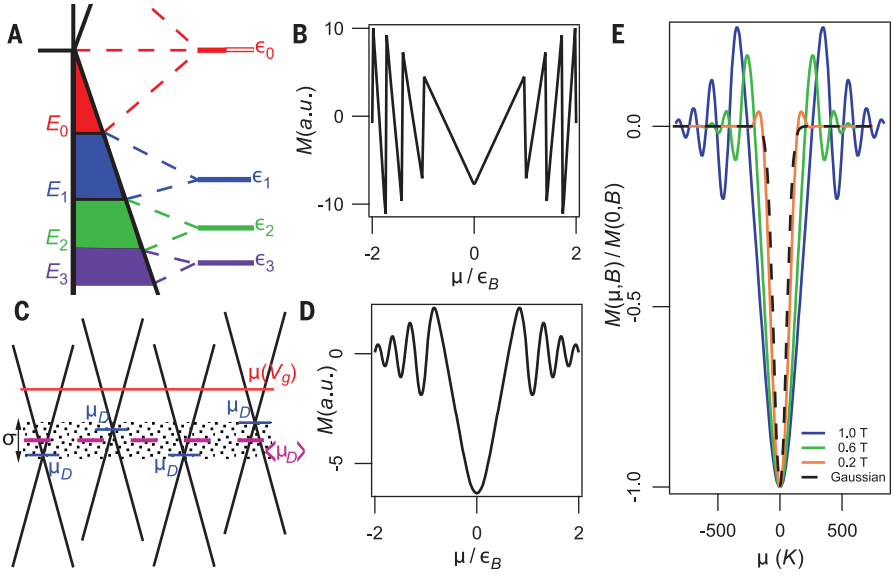


Fig. 3. Calculated chemical potential dependence of the orbital magnetization of graphene in a finite magnetic field. The calculations are based on Eq. 6; see (*14*) for more details. (**A**) Evolution of the graphene spectrum in a magnetic field [adapted from (*3*)]. The condensation of the continuous spectrum into Landau levels decreases the energy, except for the zero energy level whose contribution is predominant. Globally, the net result is an increase of the energy with the field—that is, a diamagnetic response [see also figure 5 of (*3*)]. (**B**) Without disorder, the magnetization, plotted as a function of the rescaled chemical potential μ/ε_B, exhibits discontinuities at the Landau level energies $\sqrt{n}\varepsilon_B$; a.u., arbitrary units. (**C**) Sketch illustrating the spatial distribution of electrochemical potentials $\mu' = \mu_D - \langle\mu_D\rangle$ where μ_D is the local Dirac point and $\langle\mu_D\rangle$ is its spatial average. (**D**) Rounding of $M(\mu/\varepsilon_B)$ by a Gaussian chemical potential distribution with a variance $\sigma = 0.1\varepsilon_B$. (**E**) Calculated $M(\mu)$ for different magnetic fields for $\sigma = 50$ K. At low fields, the oscillations disappear and the magnetization displays a Gaussian diamagnetic peak at $\mu = 0$. This peak is broadened by the magnetic field as soon as $\varepsilon_B \geq \sigma$.

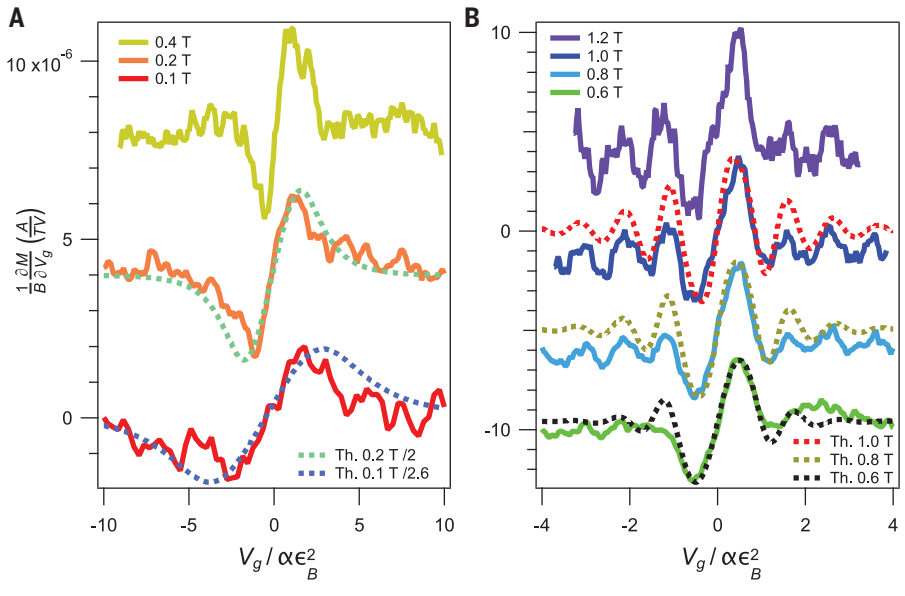


Fig. 4. Comparison of theory to experiment. Fit of detected AC magnetization response to a gate voltage modulation of 50 mV, as a function of the DC gate voltage divided by $\alpha\varepsilon_B^2 = 2\alpha e\hbar v_F^2 B$. Dashed lines show the theoretical gate dependence of $\partial M/\partial V_g$, with $\sigma_0 = 165$ K and $\sigma_\infty = 50$ K, including the extra rounding effect owing to the 50-mV AC gate modulation. In (**A**), the amplitude of the theoretical signal has been rescaled by a factor of 1/2.6 at 0.1 T and by a factor of 1/2 at 0.2 T to fit quantitatively the experimental data. In (**B**), the rescaling factors are closer to unity for higher fields for which the McClure peak is expected to be independent of σ_0.

The full field and chemical potential dependence of the magnetization, including the oscillations, is given by the derivative $M = -\partial\Omega/\partial B$ of the disorder-averaged grand potential $\Omega_\sigma(\mu, B)$ (*14*) (eqs. S20 to S23):

$$\Omega_\sigma(\mu, B) = \int P_\sigma(\mu')\Omega_0(\mu - \mu', B)d\mu' \quad (5)$$

with

$$\Omega_0(\mu, B) = \frac{\varepsilon_B^3}{4\pi^2\hbar^2c^2}\sum_{p>0}\frac{1}{p^{3/2}}\left[1 - 2S\left(2\sqrt{p}\frac{|\mu|}{\varepsilon_B}\right)\right] \quad (6)$$

The Landau levels at energies $\sqrt{n}\varepsilon_B$, with $\varepsilon_B = \sqrt{2e\hbar v_F^2 B}$, enter via the argument $\sqrt{p}|\mu|/\varepsilon_B$ where p is an integer, in the Fresnel function

$$S(x) = \int_0^x \sin\frac{\pi}{2}t^2 dt \quad (7)$$

The predicted disorder-averaged magnetization is displayed in Fig. 3E. With increasing field, it evolves from a sole diamagnetic McClure peak of width σ to a broader peak with additional oscillations, centered at $\mu_n/\varepsilon_B = \sqrt{n}$. Figure 3D demonstrates how charge disorder induces rounding and attenuates the oscillations.

To compare these predictions to experiment, we must also relate the gate voltage V_g to the chemical potential μ. Far from the Dirac point, this relation is quadratic, $V_g(\mu) - V_D = \alpha\mu^2\,\mathrm{sign}(\mu)$, with

$$\alpha = \frac{e/C_g}{\pi\hbar^2 v_F^2} \quad (8)$$

where V_D is the gate voltage at the Dirac point, and C_g is the geometrical capacitance per unit surface between graphene and gate, as determined from the V_g periodicity of the de Haas–van Alphen oscillations at high field (*14*). In contrast, close to the Dirac point, V_g varies linearly with μ, with a slope given by the standard deviation of μ disorder around the Dirac point, σ_0:

$$V_g(\mu) - V_D = \frac{4\sigma_0\mu}{\sqrt{2\pi}} \quad (9)$$

(eqs. S24 and S25). We find that the experimental data can be fit (see Fig. 4) using two constants, σ_0 = 165 K and σ_∞ = 50 K, which describe the μ distribution at low and high doping respectively (eqs. S26 and S27). The smaller value of σ_∞ is explained by the more efficient screening of charge impurities at high doping. We note that the two constants can practically be determined independently, given the high sensitivity of the decay of the de Haas–van Alphen oscillations to disorder, and the large broadening of the McClure peak induced by magnetic field (on the order of ε_B).

We find that $M(V_g)$ and $\partial M/\partial V_g$ depend on V_g, σ_0, and σ_∞, exclusively via the variables $V_g/\alpha\varepsilon_B^2$, σ_0/ε_B, and $\sigma_\infty/\varepsilon_B$. In particular, the variation as $\alpha\varepsilon_B^2$ of the $\partial M/\partial V_g$ peak's width, shown in Fig. 2F, is directly related to this scaling, which originates from the Dirac Landau spectrum of graphene.

Next, we compare the magnetization peaks measured at the Dirac point at 0.1 T and 0.2 T to theoretical expectations. We find that the predicted amplitude of the antisymmetric magnetization peak at the Dirac point $(1/B)(\partial M/\partial V_g)$ at low magnetic field, equal to $9.6 \times 10^{-6}\,A(TV)^{-1}$, is on the order of the experimental values, although larger by a factor 2 to 2.6. This is probably a consequence of the oversimplified model of the Gaussian distribution of electrochemical potentials we have used. This value corresponds to a diamagnetic magnetization two orders of magnitude larger than the Landau diamagnetism of a 2D free electron gas. Finally, deviations from the linearity between magnetization and magnetic field are expected when ε_B becomes much greater than σ_0, with a smooth crossover toward a $\sqrt{B}$ dependence (eqs. S20 to S23). Because the calibration of the GMR sensor becomes delicate in high perpendicular magnetic fields owing to the residual imperfect alignment of the magnetic field, these deviations from linearity cannot be precisely checked in the field range above 0.5 T where they are expected to occur.

We have detected the McClure singularity of low-field orbital magnetization of a single graphene monolayer at the Dirac point, which is the signature of the π Berry phase of electronic wave functions in graphene. This experiment should also enable the investigation of interband-induced Berry curvature anomalies (*16–18*) as well as Coulomb interaction effects in 2D materials such as graphene and its bilayer (*19*, *20*). Moreover, in contrast to the diamagnetic McClure peak observed here, a divergent paramagnetic orbital susceptibility (*21*) is expected at Van Hove singularities in the presence of moiré potentials of high periodicity. These moiré potentials also generate flat bands in the magic-angle twisted bilayer of graphene (*22*). An anomalous quantum Hall effect is then expected to appear as the result of Coulomb interactions leading to valley symmetry breaking (*23–25*) and orbital current loops in zero magnetic field. They are detectable via the orbital magnetic moments they would generate, as very recently shown in (*26*). The possibility of generating flat bands with a periodic array of strain has also been predicted (*27–29*). In fig. S12, we present data on a strained sample on which it was possible to detect a gate-dependent GMR signal at zero magnetic field. This preliminary result suggests that more controlled situations like that in (*30*) can be investigated. Such measurements could also be used to reveal the expected ballistic loop currents along the edges of 2D topological insulators (*31–34*).

REFERENCES AND NOTES

1. L. D. Landau, *Eur. Phys. J. A* **64**, 629–637 (1930).
2. J. W. McClure, *Phys. Rev.* **104**, 666–671 (1956).
3. A. Raoux, M. Morigi, J.-N. Fuchs, F. Piéchon, G. Montambaux, *Phys. Rev. Lett.* **112**, 026402 (2014).
4. M. V. Berry, *Proc. R. Soc. London Ser. A* **392**, 45–57 (1984).
5. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, A. K. Geim, *Rev. Mod. Phys.* **81**, 109 (2009).
6. H. Fukuyama, *Prog. Theor. Phys.* **45**, 704–729 (1971).
7. H. Fukuyama, *J. Phys. Soc. Jpn.* **76**, 043711 (2007).
8. M. Koshino, T. Ando, *Phys. Rev. B* **76**, 085425 (2007).
9. Y. Ominato, M. Koshino, *Phys. Rev. B* **87**, 115433 (2013).
10. O. V. Yazyev, *Rep. Prog. Phys.* **73**, 056501 (2010).
11. R. R. Nair *et al*., *Phys. Rev. Lett.* **105**, 207205 (2010).
12. Z. Li *et al*., *Phys. Rev. B* **91**, 094429 (2015).
13. L. Chen *et al*., *RSC Adv.* **3**, 13926 (2013).
14. See supplementary materials.
15. J. H. Chen, L. Li, W. G. Cullen, E. D. Williams, M. S. Fuhrer, *Nat. Phys.* **7**, 535–538 (2011).
16. D. Xiao, M. C. Chang, Q. Niu, *Rev. Mod. Phys.* **82**, 1959–2007 (2010).
17. A. Raoux, F. Piéchon, J.-N. Fuchs, G. Montambaux, *Phys. Rev. B* **91**, 085120 (2015).
18. A. Raoux, thesis, Université Paris Saclay (2017).
19. L. Ju *et al*., *Science* **358**, 907–910 (2017).
20. A. Principi, M. Polini, G. Vignale, M. I. Katsnelson, *Phys. Rev. Lett.* **104**, 225503 (2010).
21. G. Vignale, *Phys. Rev. Lett.* **67**, 358–361 (1991).
22. Y. Cao *et al*., *Nature* **556**, 80–84 (2018).
23. A. L. Sharpe *et al*., *Science* **365**, 605–608 (2019).
24. M. Serlin *et al*., *Science* **367**, 900–903 (2020).
25. J. Liu, Z. Ma, J. Gao, X. Dai, *Phys. Rev. X* **9**, 031021 (2019).
26. C. L. Tschirhart *et al*., *Science* **372**, 1323–1327 (2021).
27. J. W. F. Venderbos, L. Fu, *Phys. Rev. B* **93**, 195126 (2015).
28. C. C. Hsu, M. L. Teague, J. Q. Wang, N. C. Yeh, *Sci. Adv.* **6**, eaat9488 (2020).
29. Y. Zhang *et al*., *Phys. Rev. B* **102**, 081403 (2020).
30. A. Reserbat-Plantey *et al*., *Nano Lett.* **14**, 5044–5051 (2014).
31. C. L. Kane, E. J. Mele, *Phys. Rev. Lett.* **95**, 226801 (2005).
32. T. Yoda, T. Yokoyama, S. Murakami, *Sci. Rep.* **5**, 12024 (2015).
33. P. Potasz, J. Fernández-Rossier, *Nano Lett.* **15**, 5799–5803 (2015).
34. H. Kim *et al*., *NPG Asia Mater.* **8**, e271 (2016).
35. J. Vallejo, Zenodo (2021); https://zenodo.org/record/5541353#.YW0rIF5R1PY

ACKNOWLEDGMENTS

We thank E. Paul of SPEC-CEA for the GMR sensors patterning, and R. Deblock, A. Chepelianskii, F. Piéchon, J. N. Fuchs, F. Parmentier, R. Delagrange, A. Murani, and S. Sengupta for fruitful discussions. **Funding:** Supported by the BALLISTOP ERC 66566 advanced grant and the MAGMA ANR-16-CE29-0027-01 grant. Also supported by the Elemental Strategy Initiative conducted by the MEXT, Japan, grant JPMXP0112101001, JSPS KAKENHI grant JP20H00354, and CREST (JPMJCR15F3), JST (K.W. and T.T.). **Author contributions:** J.V.B. fabricated and positioned the graphene stack on the GMR device, optimized and ran the experiment, and worked on the interpretation and fits of the data. N.J.W. and T.W. helped on sample fabrication. K.W. and T.T. provided hBN single crystals. T.P., A.B., S.D., and V.B. contributed to the design, optimization, and calibration of the experiment. C.F. and M.P.-L. designed, fabricated, and optimized the GMR sensors. G.M. is responsible for the theoretical part of the work. M.F., S.G., and H.B. supervised the experimental work. J.V.B., T.W., C.F., M.P.-L., G.M., S.G., and H.B. contributed to the writing of the manuscript. **Competing interests:** There are no competing interests. **Data and materials availability:** Datasets and theory curves computed using mathematica are available on Zenodo (*35*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf9396
Materials and Methods
Supplementary Text
Figs. S1 to S17
References (*36–44*)

10 December 2020; resubmitted 14 August 2021
Accepted 19 October 2021
10.1126/science.abf9396

CALL FOR PAPERS

SPJ
SCIENCE PARTNER JOURNALS

Health Data Science

Health Data Science is an online-only Open Access Science Partner Journal published in affiliation with **Peking University (PKU)** and distributed by the **American Association for the Advancement of Science (AAAS)**. This journal is committed to promoting the appropriate application of cutting-edge technologies and analytic approaches for the health field, translating the value of health data into practice for the benefit of patients' health and populations' well-being and advancing health data science. *Health Data Science* publishes innovative, scientifically-rigorous original research, comprehensive reviews, insightful perspectives, viewpoints and editorials. The topics of interest include but are not limited to: data management, governance and provenance; comparative effectiveness research; ethics issues; public health and health policies; and more.

Submit your research to *Health Data Science* today!

Learn more at **spj.sciencemag.org/hds**

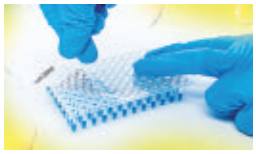
The Science Partner Journal (SPJ) program was established by the American Association for the Advancement of Science (AAAS), the nonprofit publisher of the *Science* family of journals. The SPJ program features high-quality, online-only, Open Access publications produced in collaboration with international research institutions, foundations, funders and societies. Through these collaborations, AAAS furthers its mission to communicate science broadly and for the benefit of all people by providing top-tier international research organizations with the technology, visibility, and publishing expertise that AAAS is uniquely positioned to offer as the world's largest general science membership society.

Visit us at **spj.sciencemag.org**

@SPJournals **@SPJournals**

OPEN ACCESS

ARTICLE PROCESSING CHARGES WAIVED THROUGH DECEMBER 2024



Microplate Adhesive Seals

Porvair Sciences announces a new range of fast-to-apply and highly effective microplate adhesive seals for labs that are short on time and may not have access to automated sealing instruments. Adhesive seals offer an inexpensive, easy way to seal microplates for transport or to prevent contamination and evaporation. Our adhesive films come in a variety of formats to suit most applications. All are manufactured from top-quality materials and tested to provide the strongest seal required with the necessary characteristics to perform in different situations, such as peeling after adhesion, piercing to recover samples, or being gas permeable but resistant to evaporation or solvents. In addition, clear seals are available for optical applications. For labs with lower plate-sealing requirements, the Porvair Hand Applicator is ideal for firmly applying pressure to self-adhesive film and foil sheets, resulting in a perfect seal every time. For higher-throughput microplate sealing requirements, we offer the Porvair Kbiosystems KAPS Adhesive Microplate Sealer and sealing devices.

Porvair Sciences
For info: +1-800-552-3696
www.microplates.com/adhesive-sealing-films

Extracellular Matrix for Cell Culture

Biochemically defined and animal-free, Mussel Adhesive Protein based matrix (MAPtrix) from AMSBIO provides a highly defined, controllable recombinant extracellular matrix (ECM) for cell culture, which can be tailored to the specific needs of particular cell types. Inexpensive and ready-to-use, MAPtrix produces a reliable, reproducible protein coating on a wide variety of surfaces, including glass, plastic, and biological. Used as a replacement for traditional basement membrane extracts in drug-delivery, surface-modified medical device, stem-cell, and tissue engineering scaffold applications, MAPtrix delivers superior cell-plating efficiency, improved cell morphology, and enhanced cell proliferation. MAPtrix incorporates small individual motifs from ECM proteins into recombinant proteins with a backbone of Mussel Adhesive Protein to attach to the substrate, thereby presenting the ECM motif for cultured cells to attach to. Incorporating mix-and-match genetically incorporated bioactive peptides, MAPtrix supplies an in-vivo-like surface structure that mimics the native extracellular environment, enabling the maintenance of cells under serum and feeder-free conditions.

AMS Biotechnology
For info: +1-617-945-5033
www.amsbio.com/maptrix-recombinant-ecm

Specialized Pipette Calibration Balances

Cubis MPS is a portfolio of specialized micro and semimicro balances designed to perform fast, ergonomic, and precise pipette calibrations. These highly specialized instruments are for customers that use their balance for pipette calibrations only and who need the fastest stabilization times and the most ergonomic handling. Cubis MPS features an automatic draft shield that includes a motion sensor. The sensor recognizes if a user wants to dispense and opens automatically; it closes again only if the user removes the pipette. With this system, misoperation is avoided completely, and the user can focus on the proper handling of the pipettes. Cubis MPS can be equipped with a climate tower and a temperature-controlled liquid reservoir. It includes workflows integrated to guide the user through the pipette calibration routine (according to ISO 8655), calculates weight values into volumes (incorporating environmental data such as temperature and barometric pressure), and evaluates according to ISO 8655 tolerances.

Sartorius
For info: +49-(0)-551-308-0
www.sartorius.com/en/products/weighing/pipette-calibration

Neuron Culture Medium

This product is a serum-free medium designed for primary culture of rat and murine neurons and is suited for culturing cells of the central nervous system. The product contains culture supernatant of rat glial cells. Mature neurons are produced in 14 days—approximately half the time required with conventional culture methods. Only one-fifth to one-tenth the number of cells are needed, as compared with conventional methods. Cells can be cultured with the culture medium alone, with no requirement for additional supplements.

FUJIFILM Wako Chemicals
For info: +81-(0)-6-6203-3741
labchem-wako.fujifilm.com/us/category/00821.html

GPC/SEC System for Macromolecule Characterization

The HK Series Multi-Detector GPC/SEC System from TESTA Analytical Solutions combines high-performance chromatographic detector technologies and powerful, easy-to-use software to unlock the secrets of even the most complex synthetic polymers and natural macromolecules. The system offers a wide range of detectors to choose from, including differential refractive index (DRI), multiangle light scattering (MALS), viscometry, ultraviolet, and flow measurement. This versatile system can be used to analyze a variety of polymers regardless of molecular weight range or solvent. The combination of these different sensitive detectors can deliver accurate, absolute molecular weight, size, and distribution information as well as useful data to help derive fundamental parameters such as radius of gyration, concentration, intrinsic viscosity, specific refractive index increment (*dn/dc*), and second virial coefficient. Each detector has independent temperature control to ensure that high levels of precision and reproducibility can be easily achieved every time. This powerful multi-detector gel permeation chromatography/size-exclusion chromatography (GPC/SEC) platform is supported by a powerful software suite that integrates highly intuitive operation with all the necessary tools to generate and present results from the collected raw data.

Testa Analytical Solutions
For info: +49-(0)-30-864-24076
www.testa-analytical.com/shk-series.html

siRNA Reagents

Accell siRNA (small interfering RNA) offers what no other RNA interference (RNAi) product can claim: delivery into difficult-to-transfect cells without transfection reagents, virus, or instrumentation. Achieve targeted gene silencing in neuronal, immune, or primary cells that are beyond the reach of conventional RNAi products due to toxicity caused by transfection reagents or undesirable viral responses. Predesigned Accell siRNA is available as individual reagents and in SMARTpool format. These reagents deliver proven performance in neuronal, immunological, primary, and other difficult-to-transfect cell types and are appropriate for use in vivo, due to stabilizing modifications. They also enable extended-duration knockdown with optimized continuous application.

Horizon Discovery Biosciences
For info: +1-800-235-9880
horizondiscovery.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.

SAINT LOUIS UNIVERSITY
SCHOOL OF MEDICINE

FACULTY POSITIONS
EDWARD A. DOISY DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY
SAINT LOUIS UNIVERSITY SCHOOL OF MEDICINE

Saint Louis University, a Catholic Jesuit institution dedicated to education, research, health care, and service, is seeking outstanding applicants for tenure-track faculty positions at the Assistant Professor level in the Department of Biochemistry and Molecular Biology (https://biochem.slu.edu/). The Department, founded by Nobel Laureate Edward A. Doisy, is growing the number of faculty to celebrate its forthcoming centennial in 2024. An outstanding track record of training and scholarly work in structural and/or molecular biology is strongly preferred. Please submit a cover letter, curriculum vitae, description of future research plans with a detailed 3-year budget and addresses of three references to https://slu.wd5.myworkdayjobs.com/Careers, Job #2021-02444.

Saint Louis University is an equal opportunity/affirmative action employer. All qualified candidates will receive consideration for the position applied for without regard to race, color, religion, sex, age, national origin, disability, marital status, sexual orientation, military/veteran status, gender identity, or other non-merit factors. We welcome and encourage applications from minorities, women, protected veterans, and individuals with disabilities (including disabled veterans). If accommodations are needed for completing the application and/or with the interviewing process, please contact Human Resources at 314-977-5847.

Who's the top employer for 2021?

Science Careers' annual survey reveals the top companies in biotech & pharma voted on by *Science* readers.

Read the article and employer profiles at sciencecareers.org/topemployers

Science 2021 TOP EMPLOYER

EPFL Empa

Faculty Position in Environmental Sensing Technologies

Joint Appointment between the Swiss Federal Laboratories for Materials Science and Technology (Empa) and the Ecole polytechnique fédérale de Lausanne (EPFL)

EPFL's School of Architecture, Civil and Environmental Engineering (ENAC) and the Swiss Federal Laboratories for Materials Science and Technology (Empa) invite applications for a tenured (Associate or Full) Professor in the Institute of Environmental Engineering. The appointee will participate in Empa's Functional Materials Department, and contribute to research and teaching activities within the EPFL Institute of Environmental Engineering.

The Institute of Environmental Engineering (IIE) in ENAC carries out basic and translational research spanning fundamental understanding of environmental systems and their resilience to design of adaption strategies. It covers a diverse portfolio in research, teaching and innovative technology development across transversal themes: understanding and adapting to climate change, distributed and dynamic environmental sensing, resource quality and its effects on ecosystem health, and human-environment interactions.

Empa's Functional Materials Department develops in highly interdisciplinary teams sustainable material and technology solutions for a climate-neutral society, as well as novel soft robotic systems for infrastructure maintenance and environmental monitoring.

The joint Empa/EPFL professor in Environmental Sensing Technologies will have acknowledged strengths in research and innovation related to sensing of relevant variables in water, soil or the atmosphere, as well as leveraging the resulting data for improved environmental sustainability and adaptation. We welcome a range of applications, from environmental specialists who leverage sensing technologies to mobile sensing specialists with interests in environmental questions. Areas of interest within these domains include, but are not limited to:

- Distributed and local sensing of (new) environmental variables via novel methodologies/technologies
- Multi-environment sensing of spatially and/or temporally varying phenomena using mobile platforms
- Integration of sensing with localization, estimation and optimization across multiple scales and terrains (air/soil/water)
- Unconventional sensing (e.g., combining living organisms and mechatronics)

The Empa/EPFL professor will lead an internationally recognized research program that leverages the opportunities offered by Empa and EPFL in cross-institutional, complementary research groups. The appointee will promote excellence in research and in undergraduate and graduate level teaching.

EPFL is a growing and well-funded institution with excellent experimental and computational infrastructure. Teaching and research at EPFL covers essentially the entire palette of engineering and science, and offers a fertile environment for research collaboration between different disciplines. Empa conducts cutting-edge research for the benefit of industry and the well-being of society. The EPFL and Empa environments are multilingual and multicultural, with English serving as a common interface.

The following documents are requested in PDF format: cover letter including a statement of motivation, curriculum vitae, publications list, concise statements of research and teaching interests, as well as the contact information of at least five references who are ready to supply their letter upon request.

To apply, please follow the application procedure at:
https://facultyrecruiting.epfl.ch/position/35848192

Formal evaluation of the applications will begin on **15 January 2022** and the search will continue until the position is filled.

Further enquiries should be made to:
Prof. Athanasios Nenes
Chair of the Search Committee
E-mail: SearchEnvSensingTech@epfl.ch

Additional information on Empa:
Dr. Tanja Zimmerman (tanja.zimmermann@empa.ch)
Head of Department Functional Materials

Additional information on Empa and EPFL: https://empa.ch, https://www.epfl.ch

EPFL and Empa are equal opportunity and family friendly employers. We are committed to increasing the diversity of their faculty, and strongly encourage women to apply.

By **Christopher J. Hernandez**

Sciencing while brown

I could not ignore the contrast. Working with a collaborator on a grant proposal, I was struck by our polar-opposite approaches: He used bold, confident language to describe exciting new possibilities our research would open up, while my sections were cautious and defensive, aimed at heading off potential criticism and convincing reviewers we were capable of performing the work. Although I had been a faculty member for 8 years, it had never occurred to me to write so confidently, and I later asked myself why. Upon reflection, I realized that my cautious, defensive writing style didn't just reflect my scientific training; it also had underpinnings in my background as a Mexican American.

I grew up in the agricultural heartland of California. My grandfather brought the family there from Mexico during World War II, responding to the need for farm labor. My parents were first generation college students who valued education and encouraged me to work hard in school.

As a young boy, I was keenly aware that if I wanted to succeed in school, I had to be sensitive to rules and authority. An infraction such as correcting an adult, exploiting a loophole, or pulling a prank might have been seen as cute if a white kid did it, but the same activity would get me labeled a troublemaker or lead to punishment. As I saw it, getting ahead meant I not only had to color within the lines, but had to make sure I was nowhere close to the lines.

"It has taken me time to unlearn my cautious approach."

By never pushing the limits, I became a model student and received admission to prestigious universities for my undergraduate and graduate studies. I ended up having access to career opportunities my family never even imagined. But my risk-averse approach had unintended consequences.

For starters, it factored heavily in my decision about what field to study at college. I opted for engineering, which felt safer than physics because there was a clear career path in industry if my dream of becoming an academic didn't work out. It also influenced my research, as I shied away from asking bold, risky questions. I viewed each opportunity in my career as too precious to potentially squander doing work that wouldn't lead to guaranteed results. Even when I became a faculty member, my lab's approach was "safe"—we addressed hypotheses that were likely true, but had yet to be proved.

My cautious approach kept my lab productive. But a downside became apparent when I met with the program officer overseeing a young investigator award I desperately wanted. He listened politely to my pitch, then told me the truth: "The topic area is just not right for this program; your research is too incremental." I realized that by choosing to do less-risky research, I had lost out on a chance at the flashy publications and prestigious awards that catapulted the careers of some of my peers.

At first, I thought I had focused on incremental work because that's a common approach in my subdiscipline. A few years later, though, my view changed after I sat on grant review panels and read several proposals by researchers in other fields whose approach was as cautious and defensive as mine. I noticed we had one thing in common: We are all people of color.

I realized then that marginalized groups face a dilemma in science: The cautious strategy they adopt to succeed can ultimately be a hindrance, keeping them from reaching the upper echelons of science. In my case it helped me become a solid, productive scientist doing useful work. But it also kept me from pushing the boundaries of scientific knowledge.

It has taken me time to unlearn my cautious approach. Getting tenure helped me feel freer. But I also made a conscious decision to take more risks even when that meant leaving my comfort zone. Over the past few years, my lab started several new lines of investigation, each entirely novel. And my grant proposals now focus on the implications of my work, rather than on defending my capabilities. If I could go back in time and send a message to my young investigator self, it would be: You've already proved you're capable; now's the time to take risks and change the world. ■

Christopher J. Hernandez is a professor at Cornell University.
Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

Ninth International Congress on Peer Review and Scientific Publication

Sept 8-10, 2022 | Swissôtel | Chicago, IL

Congress Advisory Board

Congress Directors
John P.A. Ioannidis, Meta-Research Innovation Center at Stanford (METRICS)

Michael Berkwits, *JAMA* and the JAMA Network

Congress Executive Director
Annette Flanagin, *JAMA* and the JAMA Network

European Director
Fiona Godlee, *BMJ*

European Coordinator
Theodora Bloom, *BMJ*

Associate Directors
Steve Goodman, Meta-Research Innovation Center at Stanford (METRICS)

Véronique Kiermer, *PLOS*

David Moher, Ottawa Hospital Research Institute

Director Emeritus
Drummond Rennie, University of California, San Francisco

Vivienne Bachelet, *Medwave,* Universidad de Santiago de Chile

Albert-László Barabási, Northeastern University

Howard Bauchner, Boston University School of Medicine

Lisa A. Bero, The University of Sydney

Patrick M. Bossuyt, University of Amsterdam

Lex Bouter, Vrije University

Isabelle Boutron, Université de Paris

An-Wen Chan, University of Toronto

Kay Dickersin, Johns Hopkins Bloomberg School of Public Health

Dan Evanko, American Association for Cancer Research

James Evans, The University of Chicago

R. Brian Haynes, McMaster University Faculty of Health Sciences

James Kigera, *Annals of African Surgery*

Sabine Kleinert, *The Lancet*

Christine Laine, *Annals of Internal Medicine*

José Florencio F. Lapeña Jr, University of the Philippines

Malcolm MacLeod, The University of Edinburgh

Emilie Marcus, University of California, Los Angeles

Ana Marušić, *Journal of Global Health*

Bahar Memani, Elsevier

Frank Miedema, University Medical Center Utrecht

Jigisha Patel, London, UK

Eric J. Rubin, *New England Journal of Medicine*

David Schriger, University of California, Los Angeles

Emily Sena, *BMJ Open Science*

Magdalena Skipper, *Nature* and Nature Research

Victoria Stodden, University of Illinois

Deborah Sweet, Cell Press

Sarah Tegen, American Chemical Society

Martin R. Tramèr, Geneva University Hospitals

Valda Vinson, *Science*

Erik Von Elm, Cochrane Switzerland

Eric-Jan Wagenmakers, University of Amsterdam

Jasmine Wallace, American Society for Microbiology

Rebecca Williams, National Library of Medicine

CALL FOR ABSTRACTS

Abstracts are due January 31, 2022

***JAMA, BMJ*, and METRICS invite abstracts for the Ninth International Congress on Peer Review and Scientific Publication.**

The Congress will provide a forum for the presentation and discussion of new research into the quality and credibility of peer review and scientific publication, to establish the evidence base on which scientists can improve the conduct, reporting, and dissemination of scientific research.

Abstracts reporting new research on any aspect of editorial and funding peer review, scientific publication, research and reporting practices, and information exchange will be considered.

Topics of Interest, include

- Bias
- Editorial and Peer Review Decision Making
- Research and Publication Ethics
- Improving the Quality of Reporting
- Models for Peer Review and Scientific Publication
- Dissemination of Scientific and Scholarly Information

More details, examples of Topics of Interest, Instructions for Preparing and Submitting Abstracts, and abstracts of previously presented research are available at **peerreviewcongress.org**

Abstracts may be submitted December 1, 2021 through January 31, 2022

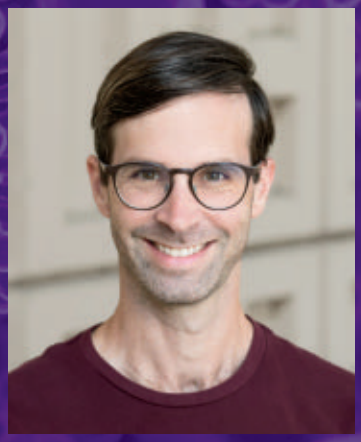
Questions? jama-peer@jamanetwork.org

JAMA Network™ thebmj METRICS META-RESEARCH INNOVATION CENTER AT STANFORD

peerreviewcongress.org

TRILLIONS OF MICROBES ONE ESSAY

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



Eran Blacher, Ph.D.
2021 Winner

NOSTER | Science
MICROBIOME
PRIZE

Apply by 1/24/22 at www.science.org/noster

Sponsored by Noster, Inc