**China's ambitious brain project achieves liftoff** p. 1368

**The future of virtual reality displays** p. 1376

**A liver protein protects the heart** pp. 1382 & 1399

# Science

$15
23 SEPTEMBER 2022
science.org

AAAS

## TERMITES AND TEMPERATURE

Wood decomposition across latitudes p. 1440

BII | Prize for Science | Innovation

# Bringing research to life–and science to market

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

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

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

The call for applications has just opened. Apply before November 1, 2022.
**www.bii.dk/scienceprize**

See film from the grand award ceremony 2022


Benedetto Marelli, Grand Prize Winner 2022

Presented by BII & *Science*

BII BioInnovation Institute Science

Apply before November 1, 2022
www.bii.dk/scienceprize

# CONTENTS

23 SEPTEMBER 2022 • VOLUME 377 • ISSUE 6613

1372

## NEWS

### IN BRIEF



### IN DEPTH



### FEATURES



## INSIGHTS

### PERSPECTIVES



### POLICY FORUM



### BOOKS *ET AL.*



1384 & 1431

PHOTOS: (LEFT TO RIGHT) MIKE BELLEME; ALAN MURPHY/MINDEN PICTURES

## LETTERS



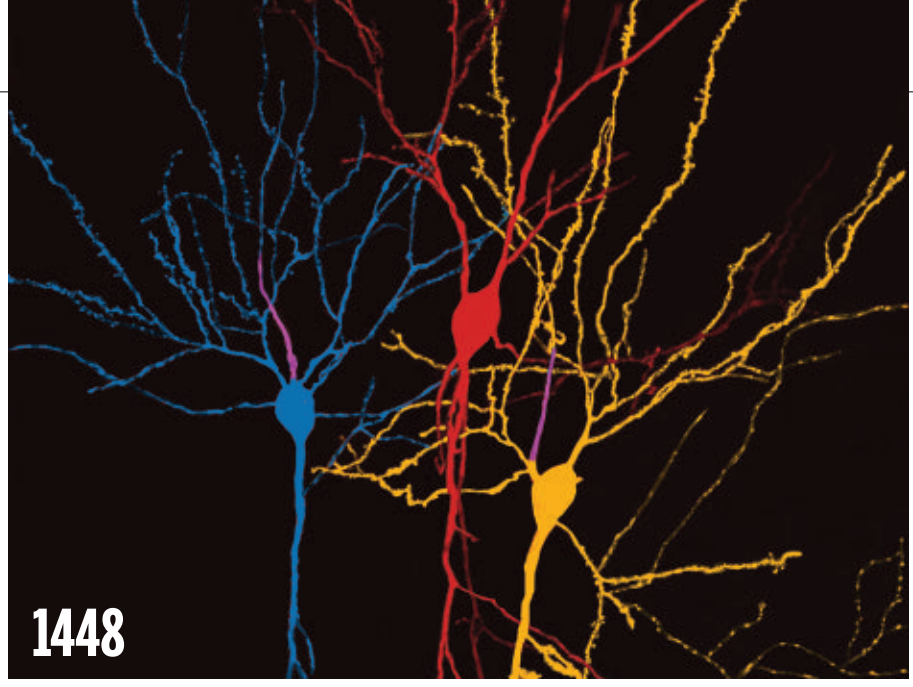
# RESEARCH

## IN BRIEF



## RESEARCH ARTICLES






Dendritic and soma-derived axons differentially contribute to hippocampal networks.



## REPORTS



## DEPARTMENTS



## ON THE COVER

Termites (*Microcerotermes* sp.) explore and consume wood, digesting it in their mound in Far North Queensland, Australia. A globally replicated experiment revealed that wood-feeding termites are far more sensitive to temperature and precipitation compared with wood-dwelling microbes. As temperatures warm, termites are likely to increase recycling rates of deadwood globally, especially in hot, dry places. See page 1440. *Image: Johan Larson*




**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.  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): $2212; 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.

CREDIT: A. ROZOV

EDITORIAL

# Public access is not equal access

**Sudip Parikh** is the chief executive officer of the American Association for the Advancement of Science (AAAS) and executive publisher of the *Science* journals. sparikh@aaas.org

**Shirley M. Malcom** is senior advisor and director of SEA Change at AAAS. smalcom@aaas.org

**Bill Moran** is the publisher of the *Science* journals at AAAS. bmoran@aaas.org

On 25 August, the White House Office of Science and Technology Policy provided guidance for scientific publishing aimed at making publications and their supporting data—the products of federally funded research—publicly available without an embargo by the end of 2025. The American Association for the Advancement of Science (AAAS, the publisher of *Science* and the *Science* family of journals) strongly supports this guidance. As written, several paths to public access remain possible. It will matter greatly to the scientific enterprise which become predominant.

As a scientific membership organization, AAAS looks at public access through the lens of scientists and engineers. We have experimented with various public access models over the past decade. The *Science* family has five subscription journals that libraries pay for access to content, and one journal for which authors pay an "article processing charge" to make the version of record of their paper freely available ["gold open access (OA)"]. All six journals publish excellent science and influential analyses, but their sustainability models differ. Each model supports the high quality that authors, readers, librarians, and funders expect us to provide through rigorous peer review shepherded by professional editors, careful editing, access to all relevant data, striking and informative visuals, and an engaging website. Importantly, we put substantial post-publication resources into preventing misinformation by informing accurate coverage of research through mainstream and social media.

From our experience, open and accessible data are essential to scientific integrity and reproducibility, and we require this accessibility immediately upon publication. Public access to trusted scientific information is also important, and situationally appropriate communication of accurate and understandable science with every audience is paramount. When any reader is unable to separate wheat from chaff, we must help by providing expertise to sift well-done from poorly done science.

Public access should foster a diverse universe of authors and readers regardless of their economic circumstances. This drives scientific excellence and public understanding. Some models for public access are bad for inclusivity. Gold OA journals, for which authors pay publication charges, work for senior scientists who are well-funded, tenured, and overwhelmingly male and white, but not so much for early-career scientists who may be poorly funded, not yet tenured, and much more diverse. Also disadvantaged are scientists at smaller schools, including historically Black colleges and universities, and in underfunded disciplines like math and the social sciences. Although it enables "open access" to readers, this model can be inequitable for many scientists and institutions.

Gold OA damages the scientific enterprise when it incentivizes a volume business model in which every paper is a quantum of revenue that must be published somewhere in a publisher's ecosystem. The perverse incentive for publishers is to accept more papers, which furthers academia's publish-or-perish mindset, makes predatory publishing more enticing, and dilutes the scientific literature. As a publisher of a gold OA journal, we've made the costly decision to maintain editorial quality and not accept papers just to meet financial targets—but we understand the temptation.

"Public access should foster a diverse universe of authors and readers..."

As a scientific membership society, AAAS seeks the best path forward for the enterprise it serves. We are actively seeking to balance the tensions between equitable access for readers and equitable access to publishing. As such, *Science* is made available through progressively priced licenses whereby larger, more research-intensive institutions pay more. We will soon provide immediate public access to all taxpayer-funded research through a policy called "green OA-zero day," which allows *Science* authors to post their "author accepted manuscript" (a fully peer-reviewed and revised version), without delay or incurring additional fees, in a public repository of their choice. This approach allows immediate public access without requiring authors to pay a publication charge, while maintaining the ability of *Science* to fulfill its mission of communicating groundbreaking research discoveries and illuminating the impact of research on society.

AAAS recognizes that its approach is not perfect and may not work for all journals, so we continue to explore other ideas. We are eager to work with the White House, funding agencies, and anyone else to implement policies that optimize equity for authors and readers. In the meantime, our approach ensures that world-changing science is published and placed into the public realm regardless of a scientist's geographic location, institutional affiliation, academic rank, or identity. We must not sew more structural inequity into the very fabric of the enterprise we seek to improve.

**–Sudip Parikh, Shirley M. Malcom, Bill Moran**

Published online 9 September 2022; 10.1126/science.ade8028

# Science AAAS





**MAIN HEADQUARTERS**
Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

**SCIENCE INTERNATIONAL**
Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

**SCIENCE CHINA**
Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

**SCIENCE JAPAN**
ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

**EDITORIAL**
science_editors@aaas.org

**NEWS**
science_news@aaas.org

**INFORMATION FOR AUTHORS**
science.org/authors/
science-information-authors

**REPRINTS AND PERMISSIONS**
science.org/help/
reprints-and-permissions

**MEDIA CONTACTS**
scipak@aaas.org

**MULTIMEDIA CONTACTS**
SciencePodcast@aaas.org
ScienceVideo@aaas.org

**INSTITUTIONAL SALES AND SITE LICENSES**
science.org/librarian

**PRODUCT ADVERTISING & CUSTOM PUBLISHING**
advertising.science.org/
products-services
science_advertising@aaas.org

**CLASSIFIED ADVERTISING**
advertising.science.org/
science-careers
advertise@sciencecareers.org

**JOB POSTING CUSTOMER SERVICE**
employers.sciencecareers.org
support@sciencecareers.org

**MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS**
science.org/subscriptions

**MEMBER BENEFITS**
aaas.org/membership/benefits




*Science* serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority or conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and book reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

EDITORIAL

# Communicating with clarity

It's easy to be discouraged by the apparently grim state of science communication. Politicians and charlatans are promoting conspiracy theories about everything from COVID-19 to climate change. Scientists are ignored by friends and relatives who would rather believe the pseudoscientific claims of a huckster with a podcast. Experts who enter the political and public milieu sometimes find themselves unable to communicate clearly, creating confusion and disdain. In the midst of this angst, Alan Alda has emerged as an unassailable voice of hope and optimism, firm in the belief that better science communication is attainable and can be transformative.

After rising to prominence in the 1970s portraying Captain Benjamin "Hawkeye" Pierce in the iconic television series, *M*A*S*H*, Alda became a continuous presence in television, theater, and movies. In the latter part of his life, he has devoted himself to improving communication of all kinds, especially science communication. For 14 years, he was the host of *Scientific American Frontiers*, a television program that produced interviews with scientists that are still used in classrooms today. In 2009, he founded the Alan Alda Center for Communicating Science at Stony Brook University, which conducts workshops and programs for scientists around the world who want to be better communicators. His podcast, *Clear+Vivid with Alan Alda*, is devoted to "connecting and communicating."

I spoke with Alda recently, and when I asked him if the main problem with scientists is that they are so caught up in the details of what they do that they forget to make connections with the people they're talking to, he encouraged me to think of it more positively: "I would say the details are so exciting, that it's hard not to remember that the details are not as meaningful to the person you're talking to if that person hasn't been at your side while you've been discovering." This optimistic outlook is what has driven Alda to devote his life to the idea that all scientists can and should be better communicators, not just those who are thrust into the public eye.

That's not to say that Alda is in denial about the challenges that surround this endeavor. "We can't do much about the politics that's hurt science," he acknowledged. "We can't do much about the whirlpool of internet communication that has not only hurt science, but every form of human intercourse.…But what we can do is communicate better." Doing so will require empathy, the establishment of personal connections, and remembering that those whom scientists are seeking to communicate with don't always have the same foundational knowledge that they do.

The magic happens when scientists dispense with the "guru on the hill" routines, Alda argues. What was once a lecture can then become a dialogue. "When I did *Scientific American Frontiers*, I felt that the reason the interviews were going so well was because they weren't conventional interviews," he told me. "I didn't come in with a list of questions that I read out. I went in just wanting to understand. And if I didn't understand, I said 'I don't get what you're telling me. Tell me again. Tell me in a different way.'"

> "Scientists can't expect to get their message across to nonscientists if they can't get it across to each other."

We talked a lot about the difficulty of conveying that science is a self-correcting work in progress. Alda agreed this is a critical hurdle. "[Journalists] have to get a story out and they have to make it readable," he argued. "There's a natural tendency to avoid ambiguity. Not just ambiguity, but nuance. Science has to make it clear that it's on a road to truth or it's on a road to understanding, and an express stop along the way isn't the whole thing."

I asked Alda to recount the biggest surprises he's encountered since he started this work. The first was how many scientists were eager for communications training. It took a while to get going, but "now they're knocking at our door," he marveled. Another unexpected outcome was that his work has led to better communication among scientists, not just between scientists and the public. We agreed that these things start at home: Scientists can't expect to get their message across to nonscientists if they can't get it across to each other.

**–H. Holden Thorp**

PHOTO: CAMERON DAVIDSON

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

10.1126/science.ade8683

# NEWS

> “You’re counting votes. Many of us are just counting bodies.”
>
> **Yale University epidemiologist Gregg Gonsalves,** tweeting about President Joe Biden’s remark that the pandemic “is over,” although about 400 Americans continue to die from the disease daily.

## IN BRIEF

*Edited by* **Jeffrey Brainard**

Culturing bacteria has long been a requirement for naming them, but a new approach may change that.

TAXONOMY

## Bacteria naming method relies on DNA

A controversial new system for naming bacteria and other prokaryotes relies only on their DNA, rather than laboratory cultures, to identify them. The approach, dubbed SeqCode and described this week in *Nature Microbiology*, promises to relieve a backlog created because so many microbial species are being revealed through DNA analyses. Under an existing protocol, the scientific community accepts a bacterium, or a prokaryote known as an archaeon, as real only if microbiologists grow the species in the lab and submit a pure “type” culture to at least two of the world’s facilities that keep microbes in perpetuity. Instead, SeqCode accepts a full or comprehensive set of a bacterium’s genome sequence data as its “type” material and outlines a protocol for assigning a Latin name. SeqCode software checks to make sure the DNA sequence is unique, and scientists evaluate whether the name was chosen according to guidelines. But it’s unclear whether the method will take hold. Some microbiologists refuse to accept a genome as sufficient evidence of a species’ existence.

## Disputed botany paper stands

SCIENTIFIC INTEGRITY | The journal *BMC Medicine* announced this month that it will not retract an influential but controversial 2013 paper by botanist Steven Newmaster of the University of Guelph (UG) that questioned the purity of herbal remedies. In 2021, eight scientists signed a complaint alleging that Newmaster was responsible for “missing, fraudulent, or plagiarized data” in three papers, including this one, which helped make him a sought-after expert and industry consultant. Independent specialists supported those concerns, as detailed in a *Science* investigation (3 February, p. 484). A UG investigation cleared Newmaster of misconduct in June, although it cited his failure “to apply the standards reasonably expected in research” for his work supporting the *BMC Medicine* article and others, including one that was retracted. Despite not retracting the herbal remedies paper, *BMC Medicine* retained a note posted in February alerting readers that doubts had arisen about the reliability of the paper’s data. Newmaster did not respond to a request for comment.

## Math professor given probation

RESEARCH SECURITY | An applied math professor at Southern Illinois University (SIU), Carbondale, will avoid prison and instead serve 1 year of probation in the most recent resolution of a case involving the U.S. government’s controversial China Initiative, which targeted U.S. academics, most of them of Chinese ancestry. In May, a jury found Mingqing Xiao not guilty of making a false statement to the government regarding his ties to Chinese institutions on a grant application. But Xiao was found guilty of filing incorrect tax returns and failing to report a foreign bank account, charges added to his original indictment. At Xiao’s sentencing this week, District Court Judge Staci Yandle said no purpose would be served by incarcerating him and that he posed no threat of reoffending. Xiao, who has been on paid administrative leave since his arrest in April 2021, told the judge he hopes to be reinstated by SIU and resume

PHOTO: TEK IMAGE/SCIENCE SOURCE

BIOLOGY

## King of the hill: 20 quadrillion ants

Ants were already estimated to be the most numerous insects. Now, a research team has developed the most comprehensive estimate to date of the number of individual ants, one that puts a new perspective on "teeming anthill." By combining data from 489 studies from around the world, the team pegged that figure at 20 quadrillion, or 20 followed by 15 zeroes. Although individual ants are light, that astronomical figure translates to a collective dry weight—the weight with all the fluids removed, which constitutes the total biomass of carbon—of 12 megatons, more than all wild birds and wild mammals combined, the team reports this week in the *Proceedings of the National Academy of Sciences*. The estimate is two to 20 times higher than previous ones, which varied because many were extrapolated from studies of ants in just one place or calculated based on an estimated percentage of ants relative to all insects. The new study relied on actual tallies of ants caught above ground but may be incomplete because it included no studies covering ants hidden in nests and lacked surveys from boreal forests, much of central Africa, and parts of Asia.

Army ants (*Eciton hamatum*) in Panama form a bridge during a raid.

PHOTO: CHRISTIAN ZIEGLER/MINDEN PICTURES

teaching and research. In February, the U.S. Department of Justice dropped the name China Initiative after concluding that the phrase has "fueled a narrative of intolerance and bias." The department has announced no new indictments of academic researchers since the name change.

## Europe protects more sea floor

FISHERIES | To prevent damage to sensitive marine habitats, the European Commission will next month close more than 16,000 square kilometers of shallow coastal waters in the northeastern Atlantic Ocean to bottom trawling. When fishing boats drag heavy nets along the sea floor to catch bottom dwellers, such as rose shrimp, they also kill other species and cloud the water with sediment. In 2016, the Commission banned bottom trawling below 800 meters in an area covering more than 4.9 million square kilometers to protect cold water coral reefs and the ecosystems where they live. The Commission's announcement last week extends the protections to EU waters between 400 and 800 meters off the coasts of four member states: France, Ireland, Portugal, and Spain. Scientists working on behalf of the International Council for the Exploration of the Sea drew on existing data to predict areas likely to contain vulnerable species, such as glass sponges and tube-dwelling anemones. Environmental groups welcomed the announcement, but fishing groups warned it would cost jobs.

THREE QS

## Bringing your kids to the field

Yale University paleoanthropologist Jessica Thompson has been towing her three children with her to field sites in Malawi for years. Her experiences—both challenging and rewarding—have led her to reflect on how fieldwork-intensive disciplines raise unique questions for researchers, especially mothers, with families. Should you bring kids with you? What if they trample fragile fossils or fall sick? She and colleagues have been surveying fellow scientists about professional imperatives to do fieldwork and their decisions about child care. She hopes the answers will support changes in practices to make fieldwork more manageable for researchers with families.

**Q: How have child care issues affected your ability to do fieldwork?**

**A:** When I was just starting, my oldest son was 1, and my parents would take him for me during the summer so I could focus on what I needed to do. Without that family support, there's just no way I could ever have done that. Because my partner works in the field with me, it means either my parents help out, or we bring [the children]. There's no other option. ... The most obvious obstacle is financial. The airfare costs of bringing multiple kids to Central Africa, where I work, rapidly add up. ... The local community loves the fact that we bring our kids. It opens doors in ways that would otherwise be completely closed to you, because it humanizes this group of scientists coming in.

**Q: How do kids on site affect the work?**

**A:** I worry about the morale of the other people at the field site. If there's this kid who's annoying them for some reason, are they going to feel like they don't want to be there? Or that they're not part of this family unit? ... If everyone is living in the field camp sharing housing and food, and your research grant is paying for all of that, how do you separate out people's personal expenses [for the children]?

**Q: How are these challenges affecting your area of science?**

**A:** We think being mothers is one of the reasons why women don't have a lot of representation among leadership in our field. I think it's pulling a lot of people out of the pipeline of field research and directing them instead to lab-based work.

PLANETARY SCIENCE

# Rover builds Mars rock cache for return plan

## Researchers plan to study organic-rich rocks in terrestrial labs, looking for past life

*By* **Eric Hand**

After collecting a dozen pinkie-size rock samples over its 18 months on Mars, the Perseverance rover has a message for planetary scientists: Your order is ready for pickup.

Next week, at a Mars community workshop, mission managers will reveal a plan to deposit 10 or 11 of the titanium sample tubes on the floor of Jezero crater, which held a lake billions of years ago. If NASA officials endorse the plan, the rover could begin to drop the samples as soon as November, assembling a cache that will play a key role in an ambitious plan to retrieve the first rocks from another planet. The Mars Sample Return (MSR) mission would use a small rocket to ferry rocks to an orbiting spacecraft that would deliver them to a special facility on Earth by 2033. There, laboratory researchers could follow up on the rover's tantalizing finding that many samples contain organic molecules—the building blocks of life—and learn whether they were made by living things.

The sample cache is actually MSR's backup plan. Plan A is for the rover to stow a larger set of 30 samples in its belly as it continues its scientific treasure hunt and deliver them to the return rocket around 2030. But if the rover gets stuck or fails along the way, researchers don't want to be left empty-handed. "Call it an insurance policy," says Susanne Schwenzer, a planetary mineralogist at the Open University and a member of the MSR campaign science group. "Once we have that cache on the ground we know we always have the option to pick it up."

For the rover team, establishing the backup cache is a milestone that shows how MSR—a dream of Mars scientists for a generation—is starting to come together. "The fact that we have reached this point is pretty amazing," says Ken Farley, the rover mission's project scientist and a geologist at the California Institute of Technology. "It's really getting real." The cache is also an inventory of rocks from the rover's 13 kilometers of exploration, extending from the crater floor where it landed to the edge of a fossilized river delta.

Some come from lava flows, a surprising and welcome discovery for rover scientists who were expecting to find mostly lakebed sediments on the crater floor. These igneous rocks contain radioactive elements such as uranium. Their decay provides a clock that Earth-based labs can use to date the moment when the rocks crystallized. Some of the volcanic rocks are thought to have been laid down before the delta, and some may have come after, so they could provide time bounds on the watery episode that created it.

Researchers also want to use lab tools to detect ancient magnetic fields frozen into certain volcanic minerals. Mars lacks a magnetic field today, but meteorites from the planet show traces of an ancient field. Its loss could have allowed water molecules to escape to space, explaining why Mars is so dry today. Dating when the magnetic field disappeared could bolster that theory, says Tanja Bosak, a geobiologist on the rover team at the Massachusetts Institute of Technology.

**In November 2021, the Perseverance rover cast its shadow among drill holes in a volcanic rock called Brac.**

The volcanic rocks might even hold signs of ancient life. Perseverance has already found that some contain carbonates and sulfates—a sign that hot water once percolated through the rocks, driving reactions favorable for early biochemistry. "There are water-rock interactions that would produce hydrogen and methane that could form a habitable environment," says Katherine French, an organic geochemist at the U.S. Geological Survey and member of the MSR campaign science group.

In the quest for past life, however, the fossilized river delta has always been the main attraction because of how sediments might preserve telltale signs. Those could be chemical: organic molecules adsorbed on clay minerals in the muds. They could even be physical: microbial fossils entombed as silt particles got cemented together over time. "The cell effectively gets sealed away from the processes that would degrade it," Bosak says.

PHOTO: NASA/JPL-CALTECH

In April, the rover arrived at the 40-meter-tall cliff at the delta's edge. Last week, the rover team revealed that one of the drilling targets there, a fine-grained mudstone, contained the highest concentration of organic molecules the rover has ever seen—a class of ring-shaped molecules called aromatics.

Further scrutiny on Earth could show whether living things made those molecules. Researchers will want to see whether they contain more of the light isotopes of carbon that life prefers, says Chris Herd, a planetary geologist on the rover team at the University of Alberta, Edmonton. "We're really looking for evidence of metabolism." Bosak wants to find even clearer signs of ancient life: the tough lipid molecules that can form cell walls. "You hope for an outline of a cell," she says. "You will never find peptides and proteins, but lipids can persist."

Rover managers want to add a few more samples to their collection before they drop the backup cache. Next week, they plan to drill at a site called Enchanted Lake, which has the potential to provide the finest grained delta rock of all. Soon after that, the rover will collect a sample of wind-deposited soil, which "integrates" information from across all of Mars, says Katie Stack Morgan, the mission's deputy project scientist at NASA's Jet Propulsion Laboratory. "We could be getting a truly global sample of the fine-grained dust that circulates on Mars." The team also wants the cache to include a tube containing nothing but air, an important resource for those who study the martian atmosphere.

Once the rover team has completed its cache and NASA has approved the plan, a small arm under the rover will begin to discharge the sample tubes. It's not going to drop them in a pile. Instead, the rover will spend about 2 months depositing them one by one, several meters apart, in a flat area of the crater. "It's like a billiards table," says Meenakshi Wadhwa, MSR principal scientist at Arizona State University, Tempe. "It's as good as it gets in terms of a place to land a sample retrieval mission."

Current plans call for a pair of autonomous helicopters, like the one Perseverance deployed last year, to collect individual samples and carry them to the 3-meter-tall rocket that will launch them into orbit. Farley says he's not worried about finding the tubes. "We will know to within a centimeter or so where they are."

If the rover remains healthy, of course, the backup cache may never make it to Earth. But psychologically, the cache will be a spur to proceed with the rest of the expensive, risky MSR scheme and an incentive to ensure it works flawlessly. "When we place that cache, that's sending a message," Bosak says, "that this is a returnable set of samples." ■

IMAGE: MOLEKUUL/SCIENCE SOURCE

BIOMEDICINE

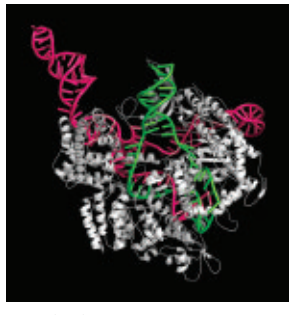
# CRISPR infusion eases symptoms in genetic disease

## Some patients no longer need drugs to prevent swelling

*By* **Jocelyn Kaiser**

In a medical first, an infusion of the CRISPR gene editor into the blood of three people with a rare genetic disease is easing their symptoms, a biotech company reports. The experimental treatment tamped down a liver protein that causes painful and potentially life-threatening bouts of swelling in the throat and limbs. Two people in the company's trial are doing so well after a single CRISPR injection that they no longer need drugs to control their condition.

The data were reported last week at a meeting in Berlin on the disease, called hereditary angioedema. The effort marks the second time the company, Intellia Therapeutics, has used CRISPR to inactivate a gene directly inside a person's body. But the Berlin presentation offers the first clear evidence of clinical benefits from infusing the tool, which can snip out or replace targeted bits of DNA.



CRISPR uses an enzyme (white) and guide RNA (pink) to cut DNA (green).

The clinical data are "impressive" and "an important achievement for the field," says Fyodor Urnov, a CRISPR researcher at the University of California, Berkeley.

CRISPR can treat blood disorders such as sickle-cell disease when a patient's cells are harvested, edited in a lab, and then returned to the body. Injecting it into the eye to treat certain blindness disorders is also showing tentative benefits. But infusing CRISPR into the blood to reach organs or cells that can't be treated outside the body is harder.

Last year, Intellia and partner Regeneron Pharmaceuticals reported that an in vivo CRISPR drug for a different rare genetic disease halted the buildup of liver proteins that can cause nerve pain, numbness, and heart problems. Although the knockdown of the protein appears to be long-lasting, the company hasn't yet revealed whether the patients' symptoms improved. For the trial in hereditary angioedema, however, the benefits emerged quickly, Intellia reports.

That disease results from mutations disabling a protein called C1-esterase inhibitor that's part of a signaling pathway controlling levels of bradykinin, a peptide hormone that causes blood vessels to leak fluid. In people with hereditary angioedema, stress, illness, or trauma can trigger high blood levels of bradykinin, producing severe swelling of the hands or feet, the abdomen, face, or even the throat, which can suffocate a person.

Drugs can help prevent these attacks by blocking a protein, kallikrein, that ramps up bradykinin levels. But CRISPR could permanently knock out the gene for kallikrein, allowing patients to avoid lifelong use of the drugs. So Intellia paired a RNA encoding CRISPR's DNA-cutting enzyme with a strand of RNA that guides the enzyme to the kallikrein gene and wrapped copies of both RNAs in a lipid nanoparticle. In the blood, the particles go to the liver, where kallikrein is made, and get sucked up its cells. They make the CRISPR enzyme, which snips the liver's gene for kallikrein.

Three hereditary angioedema patients who received a low dose of the CRISPR treatment saw their kallikrein blood levels drop by 65% on average at 8 weeks, clinical immunologist Hilary Longhurst of the University of Auckland reported at a meeting known as the Bradykinin Symposium. Two patients who were having one to three swelling attacks per month have had none since the treatment. A third who was having up to seven attacks per month stopped experiencing them after 10 weeks. The two patients who were taking drugs to prevent the swelling episodes no longer need them.

Kallikrein levels dropped even more, by 92%, in three patients treated more recently with a higher dose of CRISPR nanoparticles. The results suggest the company's early success is "not a fluke; it's reproducible," says John Leonard, Intellia president and CEO.

Longhurst's patients who are now attack-free tell her they no longer worry that a hard gym workout or an infection will trigger painful swelling. They call the CRISPR treatment "life-changing," she says. ■

A gene-edited macaque gets fed at the Chinese Academy of Sciences's Institute of Neuroscience.

PHOTO: JIN LIWANG/XINHUA VIA GETTY IMAGES

NEUROSCIENCE

# China's big brain project is finally gathering steam

## The $746 million program builds on the country's strengths in neuroscience—and its openness to primate research

*By* **Dennis Normile**

After 5 years of planning and debate, China has finally launched its ambitious contribution to neuroscience, the China Brain Project (CBP). Budgeted at 5 billion yuan ($746 million) under the latest 5-year plan, the CBP will likely get additional money under future plans, putting it in the same league as the U.S. Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) Initiative, which awarded $2.4 billion in grants through 2021, and the EU Human Brain Project, budgeted at $1.3 billion. The project "is really on the move," says one of its architects, neuroscientist Mu-ming Poo, head of the Chinese Academy of Sciences's (CAS's) Institute of Neuroscience (ION).

The details of the project remain murky. But China's researchers "seem to be building on their strengths, which is great," says neuroscientist Robert Desimone of the Massachusetts Institute of Technology, who collaborates with colleagues in China. The CBP focuses on three broad areas: the neural basis of cognitive functions, diagnosing and treating brain disorders, and brain-inspired computing. Monkey studies will play a key part in the research, and project leaders hope the virtual absence of animal rights activism in China will help lure talent from overseas. (Poo himself studied and worked in the United States for 40 years, including a decade at the University of California, Berkeley, and moved to China full-time in 2009.)

Neuroscience was first identified as a priority in China's 2016 Five-Year Plan, but soon became "a very contentious project," says Denis Simon, a China science policy expert at Duke University. "There was hefty debate and discussion about how to choose projects, set priorities, and allocate funds," Simon says. Deliberations dragged on until brain science was again designated as a priority field in the 2021 Five-Year Plan, adopted in March 2021. Funding for the CBP finally started to flow in December 2021, Poo says.

The acrimony continued. The money will be shared among 11 designated centers and about 50 research groups selected by an organizing committee that Poo heads. Neurobiologist Yi Rao, president of Capital Medical University, told *Science* all 11 selected institutes are represented on the committee, which creates conflicts of interest. "Everyone tends not to oppose the targeted projects proposed by others, so that the projects they support can also be adopted smoothly," he wrote in a 23 January social media post. Poo declined to comment on the criticism; several other neuroscientists in China did not answer emails seeking comment on the CBP.

Who will benefit most from the plan is hard to determine. *Science* could not find any official announcements of the awards to designated centers or grant recipients, and Poo declined to provide such numbers. Calls to China's Ministry of Science and Technology seeking information proved futile.

Still, Desimone says it's clear that the CBP, with its focus on treatments and basic work with primates, complements the EU and U.S. schemes. The BRAIN Initiative, announced in 2013, is more focused on tools and technologies. Europe's Human Brain Project started off, also in 2013, as a plan to build a computer model of the human brain, although its research objectives were broadened after that goal was criticized as unrealistic.

One Chinese strength the CBP plans to expand on is imaging. A group led by Qingming Luo, president of Hainan University, has refined and automated a technique named fluorescence micro-optical sectioning tomography (fMOST) to slice and image microns-thick ribbons of tissue from blocks of mouse brains. Computers reconstruct the data into 3D views of neurons and their connections. fMOST "provided a foundational data set for understanding and identifying the diverse cell types in the mouse brain," says Hongkui Zeng, director of the Allen Institute for Brain Science, which is collaborating with the U.S. BRAIN Initiative.

Now, Luo's team plans to do the same for the macaque brain, which is 200 times bigger, aiming to produce a "mesoscale connectome"—something like a wiring diagram. The effort will complement new brain-mapping programs under the BRAIN Initiative, says the initiative's director, John Ngai of the U.S. National Institutes of Health (see sidebar, p. 1369). He and Poo are discussing cooperation.

China is already a leader in another CBP focus area, the development of disease models in monkeys. Poo's team made headlines in 2019 by combining cloning with gene editing to produce five genetically identical macaques that lacked a key gene regulating the circadian clock. The cloning proved inefficient; the group used 325 gene-edited embryos and 65 surrogate females to create the five animals. But the gene deletion had dramatic effects: The animals exhibit sleep disorders, increased anxiety, and depression. Poo's group has also used gene editing to produce monkeys predisposed to Alzheimer's disease. Other ION researchers are developing

techniques for crippling genes in monkeys to induce symptoms of amyotrophic lateral sclerosis, says Stanford University neuroscientist Aaron Gitler, who studies ALS and has spent the past year on sabbatical at ION.

Poo intends to share his team's animal models. But because major airlines no longer carry nonhuman primates as cargo, researchers will have to visit the International Center for Primate Brain Research, which receives funding from the city of Shanghai and CAS and is not part of the CBP. The center is led by Poo and neuroscientist Nikos Logothetis, who over the past 2 years moved most of his team from the Max Planck Institute for Biological Cybernetics after his lab was targeted by animal rights activists.

Logothetis, who declined an interview request, is unlikely to have similar problems in China. "There is some but not a lot of concern about animals used in research, but there is no animal rights group focusing on this area," says Deborah Cao, an animal law and welfare expert at Griffith University, Nathan. Still, Chinese researchers are striving to "replace, reduce, and refine" animal experiments, says Ji Dai, a neuroscientist at CAS's Shenzhen Institutes of Advanced Technology. Even in China, policies on handling animals are getting stricter, he says.

For now, China is expanding the number of nonhuman primates in its research centers. The Kunming Institute of Zoology is completing a new facility that, with space for 5000 monkeys, will be China's largest, says Bing Su, a geneticist there. CAS institutes in Shanghai already have more than 1000 animals and may double or triple that number, Poo says. He adds a monkey breeding and research center in Hainan province may hold 20,000 animals a decade from now. For comparison, the United States's seven National Primate Research Centers hold 18,000 to 20,000 nonhuman primates.

The CBP may face other challenges. China's stringent zero–COVID-19 policy has led to draconian control measures; Shanghai, one of China's top research cities, shut down entirely in April and May. Such restrictions have "expats bailing out of China," Gitler says. "We'll see how that affects the ability of Mu-ming and others to recruit people."

Another question is whether political friction between China and the West will dent cooperation. Desimone sees little impact on life science collaborations at the moment. "But I don't have a crystal ball about international tensions in the future," he says. ■

With reporting by Bian Huihui.

## Next goal for U.S. initiative: A human brain atlas

*By* **Jocelyn Kaiser**

This week, the multibillion-dollar U.S. BRAIN Initiative took on its most ambitious challenge yet: compiling the world's most comprehensive map of cells in the human brain. Neuroscientists say the BRAIN Initiative Cell Atlas Network (BICAN), funded with $500 million over 5 years, will help them understand how the human brain works and how diseases affect it. The project "will transform the way we do neuroscience research for generations to come," says BRAIN Initiative Director John Ngai of the National Institutes of Health (NIH).

BRAIN, or Brain Research Through Advancing Innovative Neurotechnologies, was launched by then-President Barack Obama in 2013. It developed a program called the BRAIN Initiative Cell Census Network, resulting in a raft of papers in 2021. The studies combined data on gene expression, shapes, locations, and electrical activity of millions of cells to identify more than 100 cell types across the primary motor cortex—which coordinates movement—in mice, marmosets, and humans. Hundreds of researchers involved in the network are now completing a cell census for the rest of the mouse brain.

BICAN will characterize and map neural and nonneuronal cells across the entire human brain, which is 1000 times larger than a mouse brain. "It's using similar approaches but scaling up," says Hongkui Zeng, director of the Allen Institute for Brain Science, which won one-third of the BICAN funding. Other groups will add data from human brains across a range of ancestries and ages, including fetal development. "We will try to cover the breadth of human development and aging," says Joseph Ecker of the Salk Institute for Biological Studies, another major participant.

Another BRAIN effort announced this week will develop viral vectors and lipid nanoparticles that can alter or disable specific brain cells, which will help scientists study cell function and develop treatments. A third project, tracing how cells in mammalian brains are wired together, starts next year. Altogether, NIH has spent $2.5 billion so far on BRAIN, a figure it expects to reach $5.2 billion by the end of 2026.

BIOLOGY

# Genomes tell tales of spores versus seeds

## Genomes from ferns and a cycad reveal deep roots of plant reproduction

*By* **Elizabeth Pennisi**

The emergence of seed-producing plants more than 300 million years ago was an evolutionary watershed, opening new environments to plants and ultimately leading to the flowering plants that brighten our world and supply much of our food. But it was less of a leap than it seems, newly published DNA sequences suggest.

The genomes, from three fern species and a cycad, one of the oldest kinds of seed-bearing plants, show genes key to making seeds are the same as those in the spore-producing machinery of ferns, which emerged tens of millions of years earlier. They evidently existed in a common ancestor but were recruited into different reproductive functions as plants diverged.

The fern and cycad genomes, published in a series of papers over the past several months "fill the gap of the gene flow during plant evolution," says Shu-Nong Bai, a plant developmental biologist emeritus at Peking University who helped sequence a member of the maidenhair fern genus. "Evolutionary innovation [can] come from the alternative use of existing genes or networks, not new genes." The genomes also teach a second striking lesson: that plants acquired some of their genes not through mutation and selection, but straight from fungi or other microbes through a controversial process dubbed horizontal gene transfer.

Because of the daunting size of most fern genomes and the focus on crops such as rice, wheat, and maize, the majority of the more than 800 plant genomes sequenced so far have come from seed plants. Until now, just two were from ferns—ones with unusually small genomes. As a result, "We have only had a small snapshot of plant evolution," says Blaine Marchant, a plant evolutionary geneticist at Stanford University.

Thanks to advances in sequencing long stretches of DNA and reductions in costs, his team and three other groups have now

tackled ferns with more typical, large genomes as well as a species of cycad, a nonflowering plant with bare seeds, like those of pines and other conifers. "It is wonderful to finally see more diverse plant genomes being sequenced," says Jennifer Wisecaver, an evolutionary biologist at Purdue University.

The fern genomes, with some 30,000 genes each, reveal a panoply of genes previously tied to flowering plants, which evolved more than 200 million years later. For example, Marchant and his colleagues reported on 1 September in *Nature Plants* that a water fern, *Ceratopteris richardii*, has 10 members of a gene family known to control flowering time, seed germination, and flower shape in a small flowering plant, *Arabidopsis*. Their roles in the fern are unclear, but seven of these genes are active in leaves where spores are produced, suggesting they play a role in reproduction in ferns as well as in seed plants.

Jianbin Yan, a plant physiologist at the Chinese Academy of Agricultural Sciences's Agricultural Genomics Institute, and colleagues found similar parallels in a maidenhair fern, *Adiantum capillus-veneris*. Its DNA contains genes for transcription factors called EMS1 and TPD1, proteins that in maize and other seed plants regulate genes involved in pollen development, Yan's team reported in the same issue of *Nature Plants*. These pollen gene controllers are active in maidenhair's sporangia, the tissue where spores develop.

That fern's genome also contains a trio of genes that regulate seed development in flowering plants, adds Hongzhi Kong, a plant evolutionary developmental biologist at the Chinese Academy of Sciences's Institute of Botany. Ferns, Yan says, are "evolutionarily pivotal for a comprehensive understanding of the origin and diversification of the seed." The cycad genome contains similar networks, showing they were active in the earliest seed plants, notes Shouzhou Zhang, the botanist at the Fairy Lake Botanical Garden in Shenzan, who led its sequencing.

The new genomes shed light on one reason such insights were slow in coming: Ferns "are notoriously known to have gigantic genomes," says Fay-Wei Li, a plant evolutionary biologist at Cornell University. Researchers had assumed a process called whole genome duplication, in which an organism's complement of DNA is doubled during reproduction, explains their genome size. But, "We're not seeing the genome doubling footprint that we thought we would," says Paul Wolf, a plant geneticist at the University of Alabama, Huntsville. Instead, the ferns and the cycad gained the bulk of their DNA from the accumulation of mobile DNA—transposons and other genetic elements that infect genomes and multiply, or repetitive DNA, short sequences that got copied over and over again.

The four new genomes are also changing views about whether plants experience horizontal gene transfer. Microbes are known to swap genes all the time, helping them adapt to new conditions, but multicellular organisms seemed to borrow genes only rarely. However, the genomes of the ferns and cycad contain a surprising number of genes from bacteria and fungi. "It is remarkable that we see genes of bacterial and fungal origin in vascular plants," Kong says.

For example, the cycad sequenced has four copies of a fungal gene for a cytotoxin, a protein that can bore holes in foreign cells, and the *Ceratopteris* genome has 36 copies of another cytotoxin gene from a bacterium. These acquired genes could have bolstered their new hosts' defenses against pathogens or herbivores.

Verónica Di Stilio, a botanist at the University of Washington, Seattle, expects more surprises from the newly unveiled genomes. "Having reference genomes representative of each of the major plant lineages opens up so many possibilities," she says. "Genomes are tools, the tip of the iceberg." ■

*Ceratopteris richardii* (top) is one of three ferns to have its genome recently sequenced, revealing some of the genes behind the making of its spores (bottom).

PHOTOS: (TOP TO BOTTOM) CLEMENS ROESSNER; DAVID RANDALL

This replica of a late–sixth century helmet from the Sutton Hoo burial shows late Anglo-Saxon style.

PHOTO: GERNOT KELLER/THE BRITISH MUSEUM CC BY SA 2.5

ARCHAEOLOGY

# How the Anglo-Saxons settled England

## Genetic study of burials suggests whole families migrated to the island in the first millennium C.E.

*By* **Andrew Curry**

In the eighth century C.E., an English monk named Bede wrote the history of the island, saying Rome's decline in about 400 C.E. opened the way to an invasion from the east. Angle, Saxon, and Jute tribes from what is today northwestern Germany and southern Denmark "came over into the island, and they began to increase so much, that they became terrible to the natives."

But in the later 20th century, many archaeologists suspected Bede, writing centuries later, had exaggerated the invasion's scale. Instead, they envisioned a small migration of a warrior elite, who imposed their imported culture on the existing population. Now, a sweeping genomic study, published this week in *Nature*, suggests Bede may have been at least partly right. New DNA samples from 494 people who died in England between 400 and 900 C.E. show they derived more than three-quarters of their ancestry from Northern Europe.

The results address a long-standing debate about whether past cultural change signals new people moving in or a largely unchanged population adopting new technologies or beliefs. With the Anglo-Saxons, the data point strongly to migration, says University of Cambridge archaeologist Catherine Hills, who was not part of the research. The new data suggest "significant movement into the British Isles … taking us back to a fairly traditional picture of what's going on."

When 19th century archaeologists began to dig up Anglo-Saxon houses and burials, their finds seemed to confirm the outlines of Bede's tale. Around 450 C.E. in western England, Roman-style pottery, tools, and architecture dwindled; jewelry, swords, and houses began to resemble those found along the North Sea coast in what is today Germany and the Netherlands. Some styles evolved into spectacular forms in the new land, such as the helmets and weapons found at Sutton Hoo in eastern England.

"You can't deny there was a big shift in material culture—Roman Britain looks very different from the Anglo-Saxon period 200 years later," Hills says. In spite of that, "Most archaeologists have been critical of the idea of migration," rejecting it as an overly simplistic explanation for cultural change.

But the new DNA analysis revives it. Together with previously published DNA, samples from more than 20 cemeteries along England's eastern coast suggest a rapid, large-scale migration from Northern Europe, beginning by 450 C.E. at the latest. "Some Anglo-Saxon sites look almost 100% continental European," says co-author Joscha Gretzinger, a geneticist at the Max Planck Institute for Evolutionary Anthropology. "The only explanation is a large amount of people coming in from the North Sea zone."

The population shift brought huge cultural changes, some of which reverberate today. "There was a relatively dramatic period of language change," says University of Oxford archaeologist Helena Hamerow. Celtic languages and Latin soon gave way to Old English, a Germanic language that shares vocabulary with German and Dutch. "This suggests a significant number of Germanic speakers in lowland Britain," Hamerow says.

The Vikings who surged across the North Sea a few centuries later left fewer traces, accounting for about 6% of the genes of modern English people, compared with between 30% and 40% from the Anglo-Saxons.

That's not to say Bede got it completely right, either. The graves don't tell a clear story of armed conquest. Even people with little continental DNA were buried in Anglo-Saxon fashion, suggesting they willingly adopted the new culture. And the DNA shows both women and men immigrated, a finding supported by other researchers' results.

The team also found many individuals had a mixture of DNA from continental Europe and eastern Great Britain, suggesting intermarriage and integration lasted for centuries. One high status woman in her 20s with mixed ancestry was laid to rest near modern Cambridge under a prominent mound with silvered jewelry, amber beads, and a whole cow. Such evidence suggests more complexity than simple conquest, says co-author Duncan Sayer, a University of Central Lancashire archaeologist. "We're a million miles away from an invasion hypothesis—it's not a bunch of blokes getting in boats with weapons and conquering territory," he says.

Family relationships within cemeteries point to mass immigration as well. At one site, three generations of people with all Northern European DNA were buried close together. "I suspect there are families, or even small villages, getting up and moving," Sayer says, in line with evidence in northern Germany of settlements coming to sudden ends in the fifth or sixth centuries C.E. Researchers have proposed changing climate and pressure from other groups pushed people to migrate, and that the end of Roman control opened new opportunities in England.

Traces of western British and Irish ancestry in people buried on the continent suggest a reverse migration, too, with migrants' descendants moving back after generations in Great Britain. The results undercut the idea of Great Britain as an isolated island, upset only occasionally by invasions. "Actually, the North Sea was a highway, where people were coming and going," Hills says. "Maybe mobility is a more normal human state than we think." ■

# NATURAL INSPIRATION

## Martha Muñoz has shown that organisms can influence their own evolution—a lesson she's passing on to her students

*By* **Elizabeth Pennisi**; *Photography by* **Mike Belleme**

Martha Muñoz fell in love with nature as a child in New York City; now, she's studying salamanders at North Carolina's Grandfather Mountain.

On a hot day in July, evolutionary biologist Martha Muñoz is leading four undergraduate students on a scouting expedition in the Smoky Mountains of North Carolina. As they hike up a steep trail, Muñoz turns over rocks and pokes leaf litter to assess where they might find salamanders when they return that night. She quizzes the students about how the weather might affect their chances, then demonstrates how the crunch of leaves underfoot is an easy way to assess an area's dryness. Too much crunch means the salamanders won't be out that night.

When one student falls behind, Muñoz hangs back to lend a hand if needed. Aha Anderson has a balance disorder and apologizes for their slowness. "No apologies needed," Muñoz assures them. Later the crew will stop by Walmart to pick up a walking stick for Anderson. When another student, Jesús Buenrostro, proves squeamish about spiders, centipedes, and even grasshoppers, Muñoz reassures him with a few words in Spanish.

After dark, the group will return with headlamps, thermometers, and humidity sensors—and the goal of collecting 10 gray-cheeked salamanders to add to the growing salamander collection in Muñoz's lab at Yale University. They'll document the precise environment in which each one is found.

The southern Appalachians are a diversity hot spot for these creatures, but many of the roughly 30 species of lungless salamanders here look similar. Their environment also seems uniform, at least at first glance—creating a puzzle about how so many species could have evolved. Muñoz suspects subtle differences in behavior or habitat may have driven the salamanders to diversify, and she wants to figure out what they could be.

At 37, Muñoz has already won recognition for her discoveries about underappreciated influences on evolution, some of which buck classical thinking in the field. Her extensive studies with Caribbean lizards called anoles, for example, have provided some of the best empirical evidence that organisms can shape their evolutionary trajectory through their behavior, either speeding up or slowing down the evolution of physiological and morphological traits. She brings perspectives from multiple disciplines to evolutionary questions, says Robert Pringle, an evolutionary ecologist at Princeton University. "Her research is at the nexus of ecology, evolution, and physiology, and she has been in the vanguard of testing whether behavior acts as a drag on evolution or instead accelerates it," Pringle says.

Muñoz sees a parallel in her own career path. The daughter of Cuban refugees, she knows firsthand the challenges people from underrepresented groups face as they try to get a toehold in academic science. "There is power in knowing that we can take control of our own circumstances, that we can guide our futures," she says. "And there is even more power in knowing that this is a process that has unfolded for millions of years. It's not the exception; it's the norm."

With that in mind, not only does Muñoz work hard to influence evolutionary thinking, she also strives to make sure others have a chance to make their own impact, no matter their background. "In my home you could often hear, '*El éxito de uno de nosotros es el éxito de todos*'—the success of any of us is a success for all of us," she explains. "This is how I run my lab."

Muñoz fell in love with nature at an early age. She and a friend scaled the chain link fence at a local park, pretending they were climbing trees in the wilderness. "I dragged every adult I could find" to the American Museum of Natural History, the Bronx Zoo, and the New York Botanical Garden, where she could connect to the natural world.

In freshman biology at Boston University, she learned about the rapid diversification of animal species during the Cambrian explosion more than 500 million years ago. It "moved me to tears," she recalls, and inspired her to study evolutionary biology. She was accepted into a Ph.D. program at Harvard University, which had rejected her undergraduate and midcollege transfer applications. "I was so proud to be able to tell my parents I got into Harvard because then they relaxed—they knew they had done their part," she says.

The Smoky Mountains are home to about 30 species of lungless salamanders, including *Plethodon montanus*.

PHOTO: MIKE BELLEME

**MUÑOZ CREDITS** her grandmothers and parents for her work ethic and success. After fleeing Cuba in the 1970s, her maternal grandmother scrubbed toilets to keep Muñoz's mother and aunt housed and fed and later took care of Muñoz so her parents could work. The family eventually moved to a semidetached house in Queens near LaGuardia Airport, where despite regular insults from a racist neighbor, Muñoz found the diverse neighborhood exciting and inspiring. "We were all immigrants, all trying to get ahead," Muñoz recalls. To help out, Muñoz took a job as a cashier at the local Rite Aid, where she endured threats from angry patients being refused expired prescriptions, met customers who had to choose between food and medicine, and put up with condescending doctors. "There isn't anything about being a PI [principal investigator] that you can't learn by being in retail," she says.

At Harvard, she worked with evolutionary biologist Jonathan Losos, whose research on Caribbean anoles has become a classic example of how evolution can follow a predictable path. For decades Losos and his students have studied lizards introduced to new islands, finding that when faced with similar challenges, these newcomers often adapt by evolving similar characteristics (*Science*, 31 July 2020, p. 496).

Muñoz added a twist to this story with field research on anoles in the Dominican Republic, which boasts some of the region's highest peaks. Tropical lizards there can thrive at 3000 meters' elevation, where it can be bitter cold. Most researchers had assumed that when a tropical lizard expands to the top of a mountain, its body would change over generations to tolerate

Aha Anderson (center) labels a specimen bag for a snagged salamander as Jessica Coutee (left) records measurements on her phone and Jesús Buenrostro looks on.

PHOTO: MIKE BELLEME

the cold. But after comparing different species, Muñoz found little evidence of physiological differences that would confer cold tolerance. Instead, whereas sea-level anoles seek shelter from Sun in moist vegetation, the high-altitude lizards stayed warm by spending their days perched on boulders. They were "behaviorally nimble, exploiting Sun and shade to their advantage to stay optimally warm," Muñoz explains.

The mountain lizards' shift in behavior sped up morphological change, Muñoz found. Compared with their peers at low elevations, they had quickly evolved shorter hindlegs and flatter skulls that enabled them to hide from predators in narrow crevices in the rocks where they bask, she and her colleagues reported in 2017. The work showed a single behavior could slow one aspect of evolution, such as physiological changes in heat tolerance, and speed up another, such as the changes in anatomy she'd observed. "Far from being passive vessels at the mercy of their circumstances, organisms can influence evolution directly," she says.

That idea wasn't new, but prior to Muñoz few researchers had gone looking for empirical evidence. The influence of behavior on evolution "is an underemphasized problem that has not received nearly enough attention," says Harry Greene, an emeritus evolutionary biologist at the University of Texas, Austin. With her data, "Muñoz is causing us older folks to think harder."

After finishing her Ph.D., Muñoz did a postdoc at Duke University, where she explored another underappreciated influence on evolution: biomechanics. Duke integrative biologist Sheila Patek had been figuring out how predatory mantis shrimp evolved such fast, powerful forelimbs for crushing the shells of the snails they eat and snagging prey swimming by, and what influenced their evolutionary trajectories. These invertebrates use what's called a four-bar linkage, in which components of the forelimb act (mechanically speaking) as four "bars" connected end to end via movable joints to form a closed loop that can resemble a parallelogram. This arrangement abounds in nature and in human-engineered devices, such as locking pliers. Many researchers had assumed each bar had a similar influence on the forces produced and would be equally likely to evolve.

But that's not what Patek and Muñoz found. By comparing bar lengths in 36 species with known relationships on the mantis shrimp family tree, they showed the shortest bar often changed as a new species evolved. That bias most likely arose because the shortest bar has the most dramatic effect on mechanical output, amplifying force more than any of the other three when it got shorter.

Patek and Muñoz made a similar discovery in certain fish with four-bar linkages in their jaws. This arrangement enables wrasses, cichlids, and sunfish to snap open their mouths extra wide and suck in prey, and the proportions of the bars in these fish vary depending on whether their prey is fast moving or stationary. Fish that chase faster prey have shorter short bars that generate more force and enable them to snap prey faster, the researchers reported in 2018. Much like behavior, biomechanical principles can sculpt the rate, pattern, and direction of evolution, Muñoz says.

In 2020, Muñoz won the Society of Integrative and Comparative Biology's award for achievements in biomechanics. The following year she won the society's comparative physiology award, becoming the first researcher to win both. "She is able to integrate diverse concepts in novel and interesting ways, says Raymond Huey, an emeritus ecologist at the University of Washington, Seattle. "Most people focus on 'A' or 'B,' a few can add A plus B, but Martha can multiply them."

**IN 2019**, Muñoz landed her current job at Yale, where ecology and evolutionary biology department chair Thomas Near has been working to recruit faculty from underrepresented groups and provide a welcoming environment. In his interview with Muñoz, Near acknowledged the challenges she'd face if she took the job. "He understood that I would have to battle the diversity dimension as well as the academic dimension," she says.

These were challenges she knew well, having previously experienced the "imposter syndrome" common among scientists from underrepresented groups, who feel (however unjustly) that they don't deserve to be where they are. She'd endured slights and insults as well, such as being told she'd have to work hard even though she was a diversity hire. The reality, Muñoz says, is that scientists from underrepresented groups feel tremendous pressure to work even harder than their

peers. "We know that we have undue visibility due to our sparse numbers and correspondingly, we have a responsibility to be the best role models possible."

At Yale, Muñoz signed up to be a resident fellow in one of the colleges, where undergraduates are housed, so she and Vigo, her German shepherd, would be embedded in the community. Seven months after arriving in New Haven, Connecticut, COVID-19 grounded her—and gave her time to write a proposal for the grant that now supports the salamander work.

The dozens of woodland-dwelling *Plethodon* species in the southern Appalachians posed irresistible evolutionary questions. These salamanders look so much alike, and the environment they live in seems so uniform, that researchers have considered them an example of "nonadaptive" radiation, in which organisms split into multiple species through the accumulation of random mutations and the slow march of geographic isolation, not because they have evolved different traits. Based on her work on lizards, Muñoz suspected there might be more to that story. Perhaps these salamanders have evolved behavioral or physiological differences that make each species distinctive, or perhaps their environment isn't as uniform as it appears, creating subtle selective pressure to diversify.

Like about two-thirds of the 700 or so species of salamanders, *Plethodon* species lack lungs, breathing instead through their skin. Lungless salamanders have limited oxygen to fuel their activities and must make sure their skin stays moist enough to absorb as much oxygen from the air as possible. They've adapted by hiding and resting during the day, and by having a simplified nervous system to reduce their energy needs. As the evening cools down, they emerge from burrows, leaf litter, or rock crevices to sit, wait, and nab any insects or other prey that wander by. Most salamanders spend their lives within just a few square meters.

In the past few years Muñoz and her colleagues have collected thousands of observations of these animals, carefully recording the temperature and humidity at the exact spot where each salamander was spotted and at many other spots nearby. Already, they have documented diverse "microhabitats" in their study area—at the base of trees, under rhododendron leaves, on rocky ledges, and elsewhere—each with a specific range of temperature and humidity.

In a 2020 study of 26 species led by her postdoc Vincent Farallo, now at the University of Scranton, Muñoz and colleagues found that each prefers a slightly different combination of temperature and humidity. By choosing certain spots, each species is hydro- and thermoregulating, Muñoz says. Overall, the species mostly fall into two groups. One chooses warm, wet surroundings, where the moisture helps their skin absorb oxygen. "If their environment is wet, then they can capitalize on warmer temperatures," which allows them to be more active, Muñoz explains. A second group can tolerate drier environments—but must opt for shade or other cooler places to keep from dying out.

Muñoz hosts hundreds of salamanders from dozens of species in her lab, where she and colleagues are measuring metabolic rates, water loss rates, preferred temperatures, heat tolerance, cold tolerance, and other traits. They hope to learn whether the animals' preferences for specific spots, combined with physiological adaptations, may be contributing to the formation of new species.

So far they've found that resistance to water loss varies considerably among species, suggesting this physiological trait is evolving rapidly. Species that are less tolerant of water loss prefer wetter environments in the wild, whereas species that are more resistant to desiccation can use drier environments. If salamanders have chosen different microhabitats to suit their different moisture requirements, some populations could be becoming isolated from others, potentially setting the stage for them to become a new species.

**Comfort zone**
Habitat details for each animal could hint at how lungless salamanders have evolved.

| | |
|---|---|
| AIR TEMP. (°C) | 23.1 |
| WIND SPEED (KM) | 0 |
| DEW POINT (°C) | 17.7 |
| HUMIDITY (%) | 72.2 |
| SURFACE TEMP. (°C) | 16.2 |
| SOIL TEMP. (°C) | 18 |
| BEHAVIOR | Sit |
| SUBSTRATE | Dirt |

PHOTO: MIKE BELLEME

**AFTER WORKING IN THE LAB** all summer, Muñoz's students are eager to see the salamanders in their native habitat. The first night out is challenging, as the species they're seeking proves elusive. But by the second night the students know the routine better, and they've set their sights on a different species that proves to be more plentiful. Anderson, with the aid of the new walking stick, catches a few to help the group meet its goal. And Buenrostro, who as a youth worked alongside his mom packing fruit, shows no fear as he digs into the dirt. They finish up before midnight, far earlier than expected. "You guys are awesome," Muñoz says. "In one day, you figured it all out."

Such encouragement is quintessential Muñoz, says Jessica Coutee, one of the students on the trip. Coutee, an Army veteran, admits she wasn't sure what to make of Muñoz when they first met. Muñoz was wearing an elegant red dress as she led a group of veterans on a tour of Yale's natural history museum. But she didn't hesitate to don a pair of long yellow gloves and plunge her hands into a tub of chemicals to pull out a preserved giant iguana to show the group. "When you look at her, you might think she's a girly girl, but she's not," Coutee says. Coutee, who calls herself Louisiana Creole as she's a mix of Black, French, and Native American, is part of the first generation in her family to go to college. She, too, has wrestled with imposter syndrome, but not in Muñoz's lab. "I feel I belong," she says. "It's an unbelievable feeling that I just don't want to let go of."

Providing a nurturing community for students of all backgrounds is Muñoz's goal. "The first step into science is the hardest, so I try to make it as easy as possible," she explains. Meanwhile, she's still trying to figure some things out for herself. She is thinking about starting a family, but she has yet to receive tenure and still feels pressure to be perfect. "It feels as if I'm barely above water."

Those closest to Muñoz say she works too hard, and she doesn't deny it. But she says her work keeps her optimistic. "What nature is teaching us is that—like the lizards and salamanders I study—we are not passive vessels at the whim and mercy of our environments," she says. "While we cannot extract ourselves from existing in a certain environmental context, I see hope and possibility in our future." ■

# INSIGHTS

PERSPECTIVES

PHOTO: PAU BARRENA/AFP/GETTY IMAGES

OPTICS

# Creating the ultimate virtual reality display

## Scientists are exploring new material designs to make smaller and denser pixel displays

*By* **Won-Jae Joo**[1] *and* **Mark L. Brongersma**[2]

After decades of relative dormancy, augmented and virtual reality (AR and VR) are among the fastest developing consumer product technologies today. The growing market demand for these headset devices is being driven by a wide range of applications, from social networking to education, medicine, and, of course, gaming. However, a larger-scale adoption of the technology by the general public will require the headsets to be smaller, lighter, and cheaper and to have more data-processing power. The competing demand for their displays to be smaller and to have higher resolution is particularly challenging because of the inherent trade-off between the two. Recent advances on the display technology for VR may help to achieve that goal.

Today's commercially available VR headsets are still bulky and unwieldy when worn on the head. They are usually several inches thick and have most of the volume occupied by optical elements with big empty spaces in between. The adoption of thinner Fresnel

[1]Samsung Advanced Institute of Technology, Samsung Electronics, Suwon, 16678, Korea. [2]Geballe Laboratory for Advanced Materials, Stanford University, Stanford, CA 94305, USA. Email: brongersma@stanford.edu

**Virtual reality headsets can benefit from microdisplays with denser pixels by reducing the size of the devices and improving the perceived resolution of their large fields of view.**

lenses and/or advanced nanostructured films known as metasurfaces has enabled some degree of miniaturization (*1*–*3*). However, the spaces between the various optical elements and the display are still quite large after these improvements, and the spaces cannot be easily compressed much further by these methods. New types of metasurfaces that can manipulate light waves for different incident angles may be needed to achieve further miniaturization (*4*, *5*). For now, many of these newly proposed components do not work across the visible color spectrum without producing severe aberrations. There are also proposed optical designs that, for example, use holographic lenses or microlens arrays with a shorter focal length, which could reduce the lens-to-display distance by a factor of 2 (*6*). Another proposed design approach to further miniaturize the headset size involves the use of so-called folded optics, such as the "pancake lens." Here, folded optics bounce the light from a display back and forth to enable a size reduction of a factor of 3. However, the benefits of folded optics come at the cost of lowering the brightness because of the presence of polarizers and waveplates.

This kind of trade-off between size and efficient light management is commonplace in the design of AR and VR systems. For example, an AR design using a so-called "waveguide optical combiner" has the advantage of being ultrathin but the disadvantage of low light transmission efficiencies. This means that more than 95% of the light generated by the display is lost by the time it reaches the eye (*7*). If such trade-offs are proven to be unavoidable, the development of highly efficient microdisplays will be critical to the miniaturization of these headsets.

In addition to the size of a headset, its image quality is very important. A person with 20/20 vision can distinguish ~60 pixels per degree (PPD) near the center of their field of vision. To put this in context, for a 75-inch ultrahigh-definition TV with 8000 pixels across (an 8K resolution), the resolution as seen by a viewer from 10 feet away is >200 PPD. However, because of the small distance between the display and the user's eyes for a VR headset today, the best resolution for that experience is only ~15 PPD. To reach the resolution limit of the human eye and a field of view of 160° horizontally and 175° vertically, ~200 million pixels are needed in a microdisplay. This translates to ~7100 pixels per inch for a 2-inch square panel (*8*). To further reduce a VR headset to the size of a regular pair of glasses, the display panel size will need to be further reduced and may require up to 10,000 pixels per inch. For comparison, the pixel density for the latest generation smartphones, e.g., the Apple iPhone 13, is ~460 pixels per inch.

Various microdisplays have been used in VR headsets, including, most notably, the organic light-emitting diode (OLED) and the liquid crystal display (LCD). Each technology brings specific strengths and limitations. OLEDs offer greater contrast and a wider range of colors, as well as faster responses and a thinner profile. However, current OLED displays tend to degrade faster when operating at high brightness. By comparison, LCDs modulate the light emission from a separate back-illumination unit that makes these display systems bulkier. Although, as a benefit, they can rely on very bright and robust inorganic LEDs for the backlight, which makes the wwvvvdisplay more durable overall. The ability to support high brightness over 30,000 cd/m$^2$ (or "nits," a unit for measuring the amount of light emitted per area) is also a great advantage for VR headsets. However, with the rapid development of better optics and more–power-efficient and durable OLEDs, OLEDs will likely become the technology of choice for future VR systems over LCDs because of their inherently smaller size and lower weight.

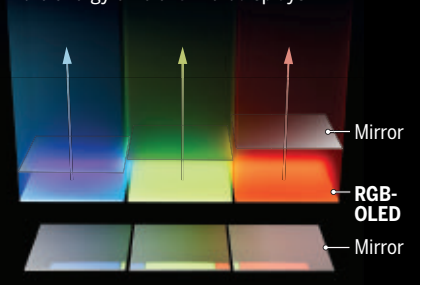
The commercialization of OLED displays began with the development of smartphones around 2010. The displays have a red, green, blue (RGB) side-by-side pixel structure, which is fabricated by depositing materials through a fine metal mask. For these RGB-OLEDs, the independent primary color subpixels enable a superior energy efficiency, brightness, and color purity. Unfortunately, the need for the metal mask brings about serious manufacturing challenges that have hampered the adoption of RGB-OLEDs for microdisplays with a few micrometer pixels as well as large-format TVs (*9*, *10*).

This prompted the development of color-filtered white OLEDs, whose fabrication does not rely on masking. The colors are generated by placing primary color filters on top of the white OLEDs—similar to the aforementioned LCD design. Because they do not require masking, the color-filtered OLEDs are much more scalable than RGB-OLEDs in their fabrication. They have already been used for a wide range of products, from few-micrometer pixelated displays to large, square meter–scale panels (*11*). Compared with the RGB-OLEDs, which is the technology currently used in VR headsets with an angular resolution of ~15 PPD, the color-filtered white OLEDs may enable a resolution of up to 60 PPD.

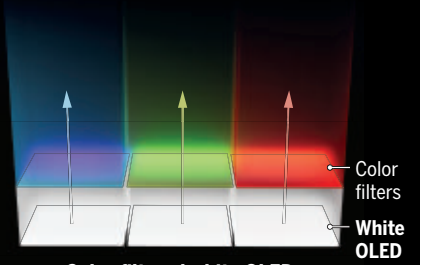
The color-defining method used by color-filtered white OLEDs is, however, inherently inefficient because the filters operate by removing valuable photons that the OLED just generated. Thus, for every red photon coming out of the red pixel, the filter would have absorbed a green photon and a blue photon and would have there-

## The near-future evolution of pixel displays

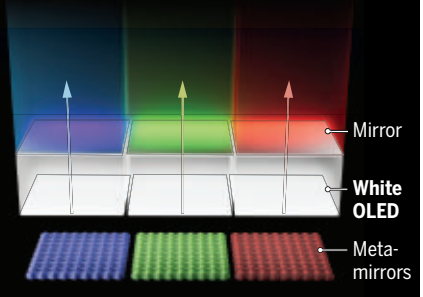
After decades of pixel miniaturization driven by smartphone and TV manufacturers, virtual reality headset makers have joined the forefront in developing denser and more energy efficient microdisplays.




**RGB-OLED**
The requirement to use fine metal masks to pixelate the red, green, blue organic light-emitting diode (RGB-OLED) limits the pixel density to ~1000 pixels per inch (PPI).




**Color-filtered white OLED**
The use of color filters and white OLEDs instead of the pixelated RGB-OLED can enable pixel densities up to 5000 PPI.




**Meta-OLED**
Nanopatterned metasurface mirrors can enable pixel density up to 10,000 PPI, near the physical limit of how small a light pixel in the visible spectrum can be.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

fore "wasted" two-thirds of the output. This inefficiency means either a higher power consumption or a lower brightness. This inherent design flaw can be improved by using a structure where two or more electroluminescent units are stacked with a "charge generation layer" in between. This is promising in concept because the layer can act as an interface electrode to deliver power to the lights, but it also creates serious electrical cross-talk between the densely packed color pixels.

For higher luminescent efficiencies and pixel density, a new type of OLED known as a "meta-OLED" was recently proposed (see the figure) (*12*). Conventional microcavity OLEDs have shown that it is possible to exploit optical resonances to enhance the light extraction for a specific color. This is similar to how an acoustic cavity, say that of a flute, only resonates and emits sound waves of a certain frequency. In meta-OLEDs, a nanopattern on the mirrors determines the color of the emitted light. This architecture does not require the use of color filters or fine metal masks. The resulting meta-OLED display can be implemented at an ultrahigh pixel density of >10,000 pixels per inch—approaching the fundamental physical limit set by the wavelength of visible light. Meta-OLEDs can also achieve higher efficiency and better color definition than their predecessors. Because the reflection phase of patterned metamirrors can be designed to be frequency dependent, meta-OLEDs can also circumvent the optical cross-talk seen in conventional microcavity OLEDs. Meta-OLED technology is currently gaining interest from companies and may become one of the first commercial metamaterials technologies.

Besides creating smaller and better displays, there are still other avenues for improving AR and VR devices. For example, the use of a curved screen will allow for an expansion of the field of view without relying as heavily on lenses. To power each individual pixel, a single crystalline silicon backplane may be most suitable from an electronic materials perspective, but it is challenging to make this curved. For ultrathin silicon membranes with high performance and flexibility, several sustainable approaches using silicon-on-insulator technology and mechanical grinding have been demonstrated, but cost and mass production concerns remain (*13*). Another avenue for improvement is to find hacks in human biology. For instance, the ultrahigh pixel density is not needed over the entire field of view all the time. The aforementioned 60-PPD sensitivity of the human eye is only true near a small part at the center of the retina, known as the fovea centralis. A foveated display, where an image is only rendered at the highest resolution when the user is staring at it, can greatly lighten the workload for the graphics processor by reducing the image quality in the peripheral vision (*13*, *14*). It can also reduce motion sickness caused by a nonoptimized VR experience. Foveated rendering can automatically blur the peripheral vision of the user, which has been shown to alleviate VR-induced motion sickness (*15*). However, an effective foveated display will require accurate tracking of the user's eye and head movements, which adds another dimension of complexity to the overall design. That being said, a foveated rendering system may be the most immediately achievable next step to improve VR experiences, if it can reduce the response time for the motion sensor, process rendering, and display to create a smooth enough experience from the user's point of view.

The introduction of the various display concepts is bound to be gradual and sequential. For many practical applications, VR hardware does not need to be perfect. Many new use cases are emerging that will trigger the transition from mobile to wearable devices. If and when this transition occurs, it will cause a major paradigm shift in the way humans interact with the digital world. This is perhaps the reason that almost all of the major electronics companies worldwide are currently investing substantial resources in AR and VR technology. In the coming years—if the commercial interest in VR headsets remains high—the various cutting-edge technologies mentioned here should begin to find their way into consumer products and help to further popularize the use of VR and AR technologies. ■

## REFERENCES AND NOTES


1. P. Lalanne, P. Chavel, *Laser Photonics Rev.* **11**, 1600295 (2017).
2. G.-Y. Lee *et al.*, *Nat. Commun.* **9**, 4562 (2018).
3. Z. Li *et al.*, *Sci. Adv.* **7**, eabe4458 (2021).
4. C. Guo, H. Wang, S. Fan, *Optica* **7**, 1133 (2020).
5. O. Reshef *et al.*, *Nat. Commun.* **12**, 3512 (2021).
6. J. Ratcliff, A. Supikov, S. Alfaro, R. Azuma, *IEEE Trans. Vis. Comput. Graph.* **26**, 1981 (2020).
7. J. D. Waldern, A. J. Grant, M. M. Popovich, *SID Symp. Dig. Tech. Pap.* **51**, 49 (2020).
8. A. K. Bhowmik, "Recent Developments in Virtual-Reality and Augmented-Reality Technologies," *Information Display 6/17*, 2017, p. 20.
9. C. Xiang *et al.*, *Light Sci. Appl.* **2**, e74 (2013).
10. A. Ghosh *et al.*, *SID Symp. Dig. Tech. Pap.* **47**, 837 (2016).
11. H.-J. Kim *et al.*, *SID Symp. Dig. Tech. Pap.* **12**, 526 (2016).
12. W.-J. Joo *et al.*, *Science* **370**, 459 (2020).
13. R. Hussain *et al.*, *Sensors* **21**, 4006 (2021).
14. G. Tan *et al.*, *Opt. Express* **26**, 25076 (2018).
15. H. Buhler *et al.*, in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)* (IEEE, 2018), pp. 517–519.


10.1126/science.abq7011

PHYSIOLOGY

# Recasting the whale's wonderful net

## Modification of cerebral vasculature helps to cushion the brains of whales and dolphins against injury

*By* **Terrie M. Williams**

More than 50 million years ago, terrestrial ancestors of dolphins and whales reinvaded the oceans in one of the most revolutionary events in mammalian history. The transition from land to sea required marked remodeling of the terrestrial mammalian form to withstand high hydrostatic pressures at depth, exponential increases in drag forces when moving locomotor appendages through water, and extreme breath-hold durations when diving (exceeding 3 hours in Cuvier's beaked whales, *Ziphius cavirostris*) (*1*). The changes were so radical that evolutionary selection pressures seem insurmountable. Yet, the transitions did occur, resulting in 47 extant cetacean (dolphin and whale) family lineages that radiated throughout the global oceans (*2*). How this evolutionary leap was accomplished has been the subject of much speculation. On page 1452 of this issue, Lillie *et al.* (*3*) continue this multimillennial investigation on aquatic adaptations in cetaceans, detailing how specialized vascular networks provide protection for their brains during submergence.

With a proportionately large cerebral cortex (*4*) and the capacity for high-speed performance while breath-holding (*5*), diving dolphins and whales face an unusual physiological challenge in terms of managing the steady delivery of oxygenated blood to the brain. Powerful undulatory movements of the tail required for propelling their large bodies through water can disrupt this process by causing differential pressures in arterial and venous vessels that rise and fall with each propulsive stroke (*6*). As noted by Lillie *et al.*, such repetitive excursions in vascular pressures have the potential to injure

Department of Ecology and Evolutionary Biology, University of California - Santa Cruz, Santa Cruz, CA, USA. Email: williams@biology.ucsc.edu

the comparatively thin microvasculature that perfuses the cetacean brain. Vascular impairment or rupture, especially while the animals are at depth, could quickly turn fatal. The protective solution appears to lie in the repurposing of blood vessels that originally served to efficiently distribute blood from the beating heart to the brain of terrestrial ancestors of cetaceans.

A dense plexus of intertwined arteries and veins, the rete mirabile (translated "extraordinary, wonderful net"), is interposed between the aorta exiting the heart and the cerebral circulation in cetaceans. Rather than the relatively simple vascular pathways of many terrestrial mammals, arterial blood of dolphins and whales is dispersed through a series of sieve-like retia found in thoracic, intravertebral, and cranial regions before reaching the brain (see the figure). The structures are both anatomically conspicuous and functionally mysterious. Decades of research have been conducted on these unusual vascular bundles, resulting in a wide variety of potential functional roles. Arterial carotid retia have been described as radiators assisting with brain cooling and consequent oxygen preservation in such diverse species as hunting dogs (*7*), African ungulates (*8*), and diving ducks (*9*). Cranial retia associated with the whale's brain may also act as a filter that prevents harmful nitrogen bubbles from instigating decompression syndromes that lead to stranding (*10*). Furthermore, the complex net structure and large surface area characteristic of the rete mirabile may also assist in modulation of cerebral blood pressure, a function proposed previously for resting dolphins (*11*) but not fully understood for active divers.

The study by Lillie *et al.* focuses on the role of the thoracic and spino-cranial retia complex in tempering pulsatile swings in arterial and venous blood pressures in the brain as cetaceans exercise while breath-holding. Like running terrestrial mammals, the locomotor movements of swimming dolphins and whales may act to generate abdominal and thoracic pressures as the abdominal organs (viscera) push against the diaphragm. These pressure pulses are subsequently translated into the vasculature. Running mammals alleviate large pressure fluctuations by exhaling. However, such a strategy is unavailable for breath-holding cetaceans.

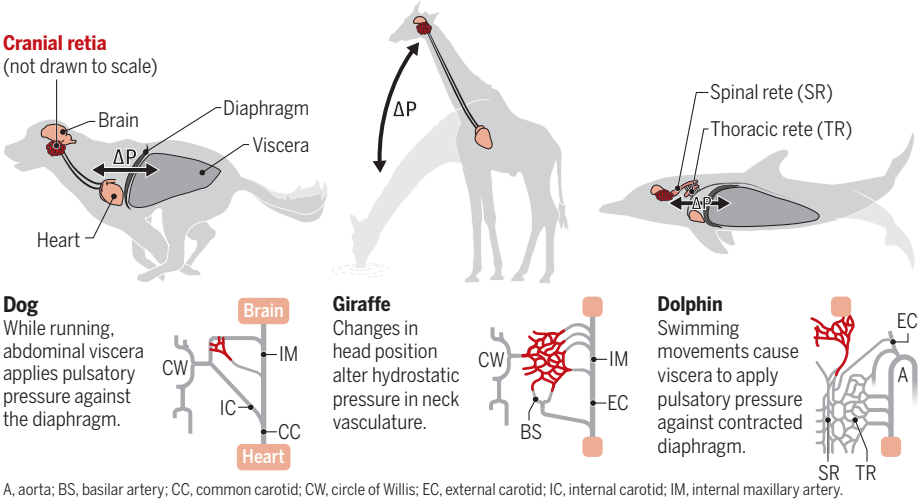
By developing a computational hemodynamic model incorporating the intricate balance of pulsatile pressures in the arteries, veins, and cerebrospinal fluid, Lillie *et al.* hypothesize that the retial complex found in cetaceans assists in circumventing this potential problem. Through a distinct "pulse-transfer" mechanism, the net-like structure of the retia spreads locomotor-generated pressure pulses across a large surface area. Compensatory flux by the cerebrospinal fluid helps to minimize pulsatility in the pressure differentials of cerebral arteries and veins that could induce vascular injury.

Notably, the role of cranial retia as blood pressure regulators is not without precedence, especially for mammals representing morphological extremes. Giraffes (*Giraffa camelopardalis*), which share an artiodactyl evolutionary lineage with cetaceans (*2*), face their own hydrostatic vascular challenges as a result of their elongated necks. Like dolphins and whales, these tall mammals may rely in part on cranial retia to avoid marked cerebral pressure changes and the possibility of "fainting" as the animals stoop and then lift their heads when drinking (see the figure) (*12*).

Cetaceans have evolved into the most diverse of all mammalian lineages in terms of body size, ranging across nearly four orders of magnitude in mass from the 27-kg highly endangered vaquita porpoise (*Phocoena sinus*) to the largest mammal to exist, the 200,000-kg blue whale (*Balaenoptera musculus*). It is unknown whether the same retial structures will be equally effective across this diversity of body and brain sizes, although the current study, by examining the vascular anatomy of small porpoises to giant baleen whales, suggests that they will. Such a universal safeguard is especially important in modern oceans, with whales and dolphins increasingly exposed to unabated environmental threats from climate change to anthropogenic noise (*13*). Clearly, the retial complexes of cetaceans meet their physiological demands under routine conditions. But with each unanticipated threat comes new forms of escape, which require rapid responses that challenge cetacean physiology. Marked reductions in heart rate (bradycardia), faster and more forceful propulsive stroking, as well as altered diving and breathing patterns characteristic of escape responses (*14*) will intensify vascular pressure pulses experienced by the animals. The question is, will the whale's ancient "wonderful net" be resilient enough to protect the brains of these remarkable marine mammals when pushed to their physiological limits in today's unpredictable oceans. ■

## Cranial retia in dogs, giraffes, and dolphins

Changes in blood pressure pulses (ΔP) in the brain vasculature can occur in concert with body movements, such as running in dogs, changes in head position in giraffes, and swimming and breath-holding in dolphins. Complex vascular connections, called retia, between the heart and brain of these animals can preserve requisite blood pressures while avoiding cerebral vascular rupture. Based on *(12, 15)*.




A, aorta; BS, basilar artery; CC, common carotid; CW, circle of Willis; EC, external carotid; IC, internal carotid; IM, internal maxillary artery.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

### REFERENCES AND NOTES


1. N. J. Quick *et al.*, *J. Exp. Biol.* **223**, jeb222109 (2020).
2. R. R. Reeves *et al.*, *Guide to Marine Mammals of the World* (Alfred A. Knopf, 2002).
3. M. A. Lillie *et al.*, *Science* **377**, 1452 (2022).
4. L. Marino, D. Sol, K. Toren, L. Lefebvre, *Mar. Mamm. Sci.* **22**, 413 (2006).
5. N. Aguilar Soto *et al.*, *J. Anim. Ecol.* **77**, 936 (2008).
6. M. Bonato *et al.*, *J. Exp. Biol.* **222**, jeb198457 (2019).
7. M. A. Baker, *Annu. Rev. Physiol.* **44**, 85 (1982).
8. G. Mitchell, A. Lust, *Biol. Lett.* **5**, 99 (2009).
9. M. Caputa, L. Folkow, A. S. Blix, *Am. J. Physiol.* **275**, R363 (1998).
10. A. S. Blix, L. Walløe, E. B. Messelt, *J. Exp. Biol.* **216**, 3385 (2013).
11. E. L. Nagel, P. J. Morgane, W. L. McFarlane, R. E. Galliano, *Science* **161**, 898 (1968).
12. G. Mitchell, J. D. Skinner, *Trans. R. Soc. S. Afr.* **48**, 207 (1993).
13. C. M. Duarte *et al.*, *Science* **371**, eaba4658 (2021).
14. T. M. Williams *et al.*, *Funct. Ecol.* **36**, 2251 (2022).
15. G. D. Castrillo, *An. de la Real Acad. Nac. de Farm.* **82**, 372 (2017).


### ACKNOWLEDGMENTS


T.M.W. is supported by the Office of Naval Research (grants N00014-17-1-2737 and N00014-20-1-2762).

10.1126/science.ade3117

SYNTHETIC BIOLOGY

# Plants to mine metals and remediate land

## Engineered plants can clean up pollution and recover technology-critical metals

*By* **Elizabeth L. Rylott** *and* **Neil C. Bruce**

Anthropogenic activities have released vast amounts of metals, metalloids, and persistent organic pollutants (POPs) into the ecosystem. The cost to human health is acute (*1*), and reserves of valuable technology-critical metals are running out (*2*). Phytoremediation, by using synthetically engineered plants to clean up polluted environments, and phytomining, a method to recover valuable metals, offer a solution to help alleviate these problems.

POPs such as polychlorinated biphenyls and dioxins are toxic carcinogens produced predominantly from industrial processes and chemical manufacturing that resist biodegradation in the environment for decades. These xenobiotic compounds are often diluted into ecosystems, making their remediation logistically challenging. The use of plant root networks can help address these issues. Furthermore, planting in polluted land can restore biodiversity and soil health, reduce pollution exposure, and is aesthetically pleasing with high public acceptance. Plants lack the ability to degrade most POPs and instead follow a classic three-step detoxification pathway in planta that results in sequestration or incorporation into macromolecular structures such as lignin. In this state, the original pollutant is considered biologically unavailable, with any further catabolism or mineralization occurring at the end of the plant's life through microbial decomposition.

There is a wealth of biochemical diversity within the plant kingdom to detoxify xenobiotics, but where plants lack sufficient enzymatic activities, there is a seemingly bottomless pool of microbial enzymes that can tackle even the most ubiquitous and challenging pollutants. Thus, genetically modified (GM) plants can be engineered to express microbial xenobiotic-degrading enzymes to extend their physiology to degrade or detoxify organic pollutants. Rice (*Oryza sativa*) engineered with an apoplastic-targeted thiocyanate hydrolase from *Thiobacillus thioparus* can effectively degrade thiocyanate from industrial mine wastes (*3*). For the military explosive and toxic pollutant hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX), expression of a bacterial cytochrome P450 was shown to mineralize the pollutant in planta—a technology that has now been successfully tested on military ranges with native switchgrass (*Panicum virgatum*) (*4*).

Industry-critical metals and finite resources—including platinum group metals (PGMs), noble metals such as gold and silver, and rare earth elements (REEs; lanthanide metals, as well as yttrium and scandium) (*2*)—are often geopolitically controlled. These elements are increasingly used in developing technologies with no current substitutes and are vulnerable to supply chain disruptions. Moreover, the mining process has a high carbon footprint and causes considerable environmental damage (*2*). Indeed, metal pollution has arisen from wastes from commercial mining, metal smelting and processing, dispersion along roadside verges from catalytic exhaust emissions, and wastes from artisanal gold mining activities. These wastes contain high-value metals but at levels that are too low for current mining practices to be economically viable. Moreover, the number of catastrophic dam failures from mine tailings has increased globally. Phytoremediation could be combined with phytomining to revegetate and restore often denuded environments, reduce dam breeching, and recover valuable metals. The process can be carried out by smallholders in high-metal sites, sometimes in difficult-to-access locations, and increasingly, the "eco-metal" recovered can attract a premium price.

Plant metal uptake starts in the rhizosphere, where plant roots and associated microbes exude compounds that solubilize metals in the soil. These metal cations enter the plant through an assortment of transporters. At uncontrolled concentrations, metals become toxic, interacting with electron transport activities and generating reactive oxygen species (ROS) that indiscriminately damage cellular components. Metals can also disrupt sulfhydryl, carboxyl, amino, and phosphoryl groups in proteins, inhibiting critical cellular enzyme activities. Thus, cellular levels of all metals need to be tightly regulated. The first point of controlling metal homeostasis is regulating influx transporters to reduce uptake and efflux transporters to pump metals out of the plant roots. This enables some "excluder" species to successfully inhabit metalliferous environments. Once taken up, metals are detoxified by complexing with chelators such as phytochelatins, metallothioneins, and amino acids, then further transported through the xylem, and compartmentalized into the vacuole or cell wall (see the figure).

In contrast to excluder species, "hyperaccumulator" plants have the ability to grow on naturally occurring metalliferous environments and take up metals with exquisite specificity to many thousandfold above background environmental levels. Their remarkable biochemistry is hypothesized to confer antiherbivory properties but could also be the key to successful phytomining (*5*). More than 700 hyperaccumulator species, including many in the *Brassicaceae* and tropical tree species of *Phyllanthaceae*, have been identified. They store a range of metals and metalloids, including arsenic (As); cadmium (Cd); copper; cobalt (Co); manganese; nickel (Ni); selenium; thallium; zinc (Zn); and the REEs lanthanum (La), cerium (Ce), and neodymium (Nd) (*6*). The efficacy of Ni phytomining was demonstrated with a 7-year field study in Pojska, Albania, where the hyperaccumulator *Odontarrhena chalcidica* growing on ultramafic (Ni-rich) soils extracted ~100 kg Ni $ha^{-1}$ annually (*5*). After harvest, the plant biomass can be ashed, producing heat as a by-product, then high-purity Ni salts are recovered by using acid-based chemical treatments. In South Africa, *Berkheya coddii* has been successfully used to phytomine and remediate Ni and Co from contaminated soils around a Ni-smelting plant, with the metals recovered from the biomass by the smelting plant (*7*).

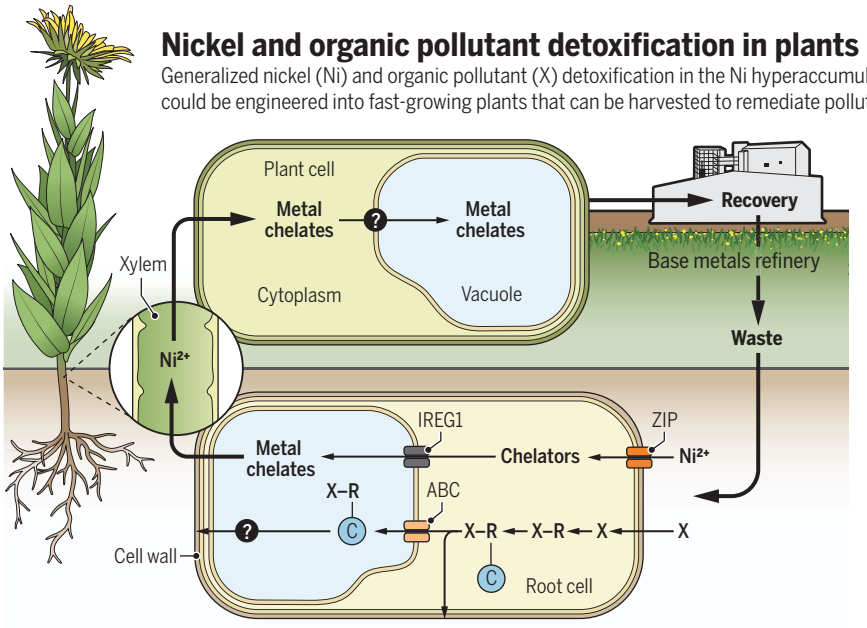
Geobotanical prospecting—identifying metalliferous soils by observing the presence of indicator species known to grow in those conditions—has revealed species containing >1000 μg $g^{-1}$ of REEs (*8*). The fern *Dicranopteris linearis*, which grows on tailings at a former REE mine in southeastern China (*9*), accumulates predominantly La, Nd, and Ce, producing biomass at commercially harvestable levels. Although relatively little is known of the mechanisms behind REE hyperaccumulation, potassium, sodium, calcium, and aluminium channels are involved in their uptake (*9*).

Gene expression studies on the most-studied Ni-hyperaccumulator species in the *Noccaea* and *Odontarrhena* genera indicate that Ni accumulation is achieved predominantly by increased expression of the genes that exist in nonhyperaccumulators, rather than increased metal-specificity per se (*10*).

Centre for Novel Agricultural Products, Department of Biology, University of York, York, UK.
Email: liz.rylott@york.ac.uk; neil.bruce@york.ac.uk

## Nickel and organic pollutant detoxification in plants

Generalized nickel (Ni) and organic pollutant (X) detoxification in the Ni hyperaccumulator *Berkheya codii* is shown. These pathways could be engineered into fast-growing plants that can be harvested to remediate polluted soil and extract valuable metals.




**Metals**

Soil microbiota solubilize metals, allowing them to be taken up in plant roots by transporters, such as ZIP family members. Metal cations are detoxified by complexing with chelators, such as phytochelatins, metallothioneins, and amino acids, then transported to the aerial tissue through the xylem and sequestered into the vacuole or cell wall.

**Organic pollutants**

Upon entering the plant, organic pollutants (X) are activated to convert them into more polar forms (R); glutathione or chemical groups such as glucosyls (C) are added to increase solubility; then these conjugated compounds can be transported into the vacuole, by ABC transporters, where they are transiently stored before incorporation into macromolecules, such as lignin in the cell wall, throughout the plant.

ABC, ATP-binding cassette; IREG1, iron-regulated transporter 1; ZIP, zinc regulated and iron regulated-like protein.

GRAPHIC: K. HOLOSKI/*SCIENCE*

This up-regulation includes low-affinity transporters such as the Zn-regulated transporter and iron-regulated transporter—like protein (ZIP) family and the metal chelators histidine and nicotianamine. How Ni is loaded into the xylem and unloaded into the aerial tissue has not yet been conclusively determined, but tonoplast iron-regulated transporters (IREGs) are implicated in transport into the vacuoles of aerial tissues in three families of hyperaccumulator species (*10*). Within the vacuole, Ni-citrate and Ni-malate are the predominant forms. In addition to enhanced detoxification pathways, hyperaccumulators often have up-regulated antioxidant systems, presumably to ameliorate metal-produced ROS.

There has been little evidence of successful commercialization of hyperaccumulator plants to phytomine metals. Key limiting factors include the availability of agronomically suitable varieties that will grow in broad geographical locations and soil types; the absence of hyperaccumulator species for PGMs and noble metals; the volatile prices of technology-critical metals; and the presence of phytotoxic levels of co-contaminating metals and, on industrial sites, organic pollutants. Increasing demand is likely to further increase metal prices and thus the viability of phytomining, but gaining additional value would improve financial stability. Toward this goal, the catalytic activity of in planta metals such as Pd or Ni can be used to enhance production of value-added platform chemicals and biofuels from biomass (*11*).

Although the agronomic traits of some species, such as *O. chalcidica*, have been developed into commercial crops, there are arguably no hyperaccumulator species that can produce the same amount of biomass as that of nonhyperaccumulator, fast-growing species such as willow (*Salix spp*) and poplar (*Populus spp*), which are currently widely cultivated across temperate regions. There are genetic transformation protocols available for these biomass species, and thus, enormous potential exists to use synthetic biology tools to create artificial hyperaccumulators, combined if necessary with degrading abilities for POPs. AlfaFold, which can predict protein structures (*12*), could be used to design proteins with metal-binding abilities or activities to specific xenobiotics. Together with gene editing techniques such as CRISPR-Cas (*13*), these tools could be used to confer traits into biomass species. For example, expression of metal-binding proteins such as lanmodulin, a La-binding protein from *Methylobacterium extorquens* (*14*), or up-regulating target metal transporters and chelators could boost metal specificity and accumulation. Alternatively, artificial chromosomes containing synthetic promoters, transcriptional activators, and repressors could be assembled into a gene stack encoding complete metabolic pathways (*13*) for metal accumulation and xenobiotic degradation.

The rhizosphere is increasingly recognized as an essential part of the metal uptake process and one that could be manipulated—for example, by using cyanogenic bacteria to solubilize metals in soils. A step further is the design of artificial organelles that bind and concentrate specific metals. Additionally, organ-level compartmentalization could be engineered to promote metal accumulation in harvestable tissues, such as aboveground, woody tissues.

Although there is still controversy over the use of GM crops, their use is established in the United States, Brazil, Argentina, and India, and there is evidence of increasing public support (*15*). Furthermore, there is higher public acceptance for nonfood GM crops, and environmental pollution is understandably an issue that many would like addressed. But what of the relatively low-value, toxic metals such as Cd, Zn, lead, and chromium, and the metalloid As? Often found together, these elements compose the majority of inorganic pollutants; they cause adverse health effects in millions globally, and remediating them is a substantial challenge. There are many plant species that can be used to take up these elements and genetic modifications to further increase plant tolerance and uptake. However, they have little current commercial value. Focusing on high-value metals could provide the knowledge and financial incentive to develop cost-effective technologies to phytoremediate lower-value metals and metalloids. Furthermore, the answer will involve a multipronged approach, with phytotechnologies filling a valuable niche. ■

### REFERENCES AND NOTES


1. Food and Agriculture Organization (FAO) and United Nations Environment, (UNEP), *Global Assessment of Soil Pollution* (FAO and UNEP, 2021).
2. T. E. Graedel, *Nat. Resour. Res.* **27**, 181 (2018).
3. J.-J. Gao *et al.*, *Sci. Total Environ.* **820**, 153283 (2022).
4. T. J. Cary *et al.*, *Nat. Biotechnol.* **39**, 1216 (2021).
5. A. Bani *et al.*, in *Agromining: Farming for Metals: Extracting Unconventional Resources Using Plants*, A. Van der Ent, G. Echevarria, A. J. M. Baker, J. L. Morel, Eds. (Springer, 2018), pp. 221–232.
6. R. D. Reeves *et al.*, *New Phytol.* **218**, 407 (2018).
7. M. Rue *et al.*, *Metallomics* **12**, 1278 (2020).
8. P. N. Nkrumah *et al.*, *Plant Soil* **464**, 375 (2021).
9. Z. Chour *et al.*, *J. Environ. Chem. Eng.* **8**, 103961 (2020).
10. L. Van der Pas, R. A. Ingle, *Plants* **8**, 11 (2019).
11. V.-N. Edgar *et al.*, *Appl. Sci.* **11**, 2982 (2021).
12. J. Jumper *et al.*, *Nature* **596**, 583 (2021).
13. W. Liu *et al.*, *Nat. Rev. Genet.* **14**, 781 (2013).
14. C. Deane, *Nat. Chem. Biol.* **15**, 2 (2019).
15. S. Evanega *et al.*, *GM Crops Food* **13**, 38 (2022).


10.1126/science.abn6337

MEDICINE

# A coagulation factor moonlights in the heart

## Liver-derived coagulation factor XI protects the heart from failure

By **Dan Tong**[1] and **Joseph A. Hill**[1,2]

Heart failure and liver diseases are among the most common causes of morbidity and mortality worldwide, and they often coexist (*1*, *2*). Heart failure is frequently associated with liver dysfunction, and liver diseases can promote heart failure, pointing to an intimate connection between these two vital organs (*3*). However, little is known regarding how these organs communicate with each other. On page 1399 of this issue, Cao *et al.* (*4*) demonstrate that coagulation factor XI (FXI), a protein exclusively produced by the liver, plays an unexpected role in the heart and protects it in response to triggers of pathological remodeling. This previously unrecognized liver-heart interaction not only enriches our understanding of interorgan communication but also suggests a promising therapeutic strategy for heart failure.

Consistent with globally increasing rates of obesity, heart failure with preserved ejection fraction (HFpEF) and nonalcoholic fatty liver disease (NAFLD), which are both strongly associated with obesity, are emerging in high-income countries as the most common subtypes of heart failure and chronic liver disease, respectively (*2*). Patients with NAFLD have an increased risk of developing heart failure, particularly HFpEF, even after adjusting for common cardiometabolic risk factors, suggesting that there are additional direct links between the two conditions (*5*). By using a systemic genetics approach, an experimental strategy that involves analysis of intermediate molecular phenotypes to bridge DNA variation with a trait of interest, Cao *et al.* provide compelling evidence in mice that liver-derived FXI cleaves bone morphogenetic protein 7 (BMP7) precursor in the myocardium, subsequently activating a BMP7-SMAD1-SMAD5 pathway. This, in turn, ultimately results in attenuation of cardiac inflammation, fibrosis, and diastolic dysfunction in the context of HFpEF (see the figure).

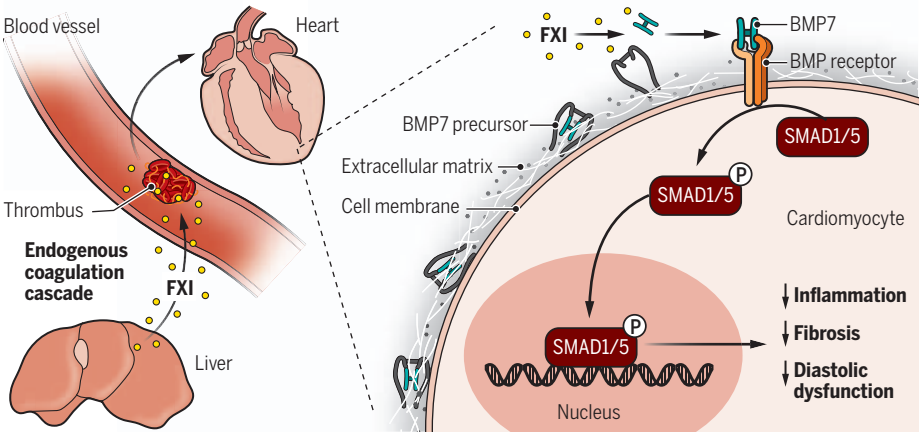
**"...coagulation factor XI (FXI), a protein exclusively produced by the liver, plays an unexpected role in the heart..."**

Factors produced by the liver, called hepatokines, regulate the function of remote organs or tissues and play important physiological and pathological roles (*6*, *7*). Most known hepatokines, such as fibroblast growth factor 21 (FGF21), exert their biological function by binding to specific receptors and co-receptors in target organs (*7*). Cao *et al.* unveil a previously unknown aspect of liver-heart interaction. The canonical function of liver-derived FXI is to cleave and thereby activate factor IX in the coagulation cascade (*8*). The authors find that the protease activity of FXI remains essential for its cardioprotective role but targets a different substrate through BMP7. BMP7 is expressed in a variety of tissues but seems enriched in cardiomyocytes, which likely underlies the privileged effect of FXI on the heart. The observation that mice harboring FXI overexpression manifest other systemic metabolic benefits, including less weight gain and a better lipid profile, suggests additional actions of FXI beyond the heart. Whether these additional effects are mediated by the BMP7-SMAD1-SMAD5 pathway remains to be determined.

Defining how this liver-heart FXI-BMP7 axis is regulated in different physiological and pathological conditions will help to elucidate its function. Cao *et al.* report that concentrations of circulating FXI are decreased in a "two-hit" HFpEF mouse model in which concurrent systemic metabolic and hypertensive stresses successfully recapitulate the myriad clinical features of human HFpEF (*9*). Although hepatic features in these mice remain to be fully characterized, they likely fall into the spectrum of NAFLD. Whether decreased circulating FXI is a clinical feature of human NAFLD and, more intriguingly, whether reduced FXI could serve as a biomarker to identify NAFLD patients with increased risk of developing HFpEF are worthy of further investigation. Conversely, approaches that are beneficial in NAFLD and HFpEF, including exercise, caloric restriction, and sodium-glucose cotransporter 2 (SGLT2) inhibition (*2*), might exert synergistic effects on both the myocardium and liver by normalizing FXI concentrations and restoring the protective effect of the FXI-BMP7 axis.

### Protecting the heart with a coagulation factor

Coagulation factor XI (FXI) is produced by the liver and functions in the coagulation cascade to promote thrombus formation. Cao *et al.* show that circulating FXI can travel to the heart and cleave bone morphogenetic protein 7 (BMP7) precursor in the extracellular matrix of cardiomyocytes, thereby activating BMP7. Activated BMP7 binds to the BMP receptor, which promotes phosphorylation (P) of SMAD1 and SMAD5 and translocation to the nucleus, where gene expression is induced. This results in reduced inflammation and fibrosis and attenuated diastolic dysfunction in the context of heart failure with preserved ejection fraction in mice.




GRAPHIC: K. HOLOSKI/*SCIENCE*

[1]Department of Internal Medicine (Cardiology), University of Texas Southwestern Medical Center, Dallas, TX, USA. [2]Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX, USA. Email: joseph.hill@utsouthwestern.edu

Increased concentrations of plasma FXI are associated with increased risk of thrombosis and ischemic stroke (*8*). However, the protective effect of the downstream BMP7-SMAD1-SMAD5 pathway observed in mouse models of HFpEF highlights its potential as a therapeutic target. Therapies that enhance myocardial BMP7-SMAD signaling have yielded beneficial effects in attenuating the pathological remodeling observed in preclinical models of pressure overload and diabetic cardiomyopathy (*10*, *11*). Further exploration of their potential therapeutic effects in HFpEF is warranted.

FXI has emerged as a major target for the development of next-generation anticoagulants with reduced risk of bleeding, and a variety of FXI or FXIa (activated FXI) inhibitors are now in clinical trials (*12*, *13*). One of the major targeted populations for these inhibitors is patients with nonvalvular atrial fibrillation, a condition that often coexists with HFpEF (*14*). The role of FXI in the heart reported by Cao *et al.* raises the possibility that inhibiting FXI, especially chronically, might lead to untoward effects in the myocardium and other organs. Careful clinical evaluation is warranted.

Cao *et al.* unveil an important mechanism of interorgan cross-talk through which the liver communicates with the heart. It is well recognized that HFpEF is a systemic, multiorgan clinical syndrome. Indeed, interconnections between the heart and other organs, including skeletal muscle, kidney, liver, lung, and nervous system, have been implicated in the pathogenesis and progression of HFpEF (*15*). Therefore, further detailed dissection of these mechanisms will be required to facilitate the identification of new therapies for this complex clinical syndrome. ■

## REFERENCES AND NOTES


1. S. Møller, M. Bernardi, *Eur. Heart J.* **34**, 2804 (2013).
2. H. M. Salah *et al.*, *JACC Basic Transl. Sci.* **6**, 918 (2021).
3. A. Mantovani *et al.*, *J. Am. Coll. Cardiol.* **79**, 180 (2022).
4. Y. Cao *et al.*, *Science* **377**, 1399 (2022).
5. M. Fudim *et al.*, *J. Am. Heart Assoc.* **10**, e021654 (2021).
6. H. J. Yoo, K. M. Choi, *Diabetes Metab. J.* **39**, 10 (2015).
7. D. Y. Seo *et al.*, *J. Clin. Med.* **10**, 385 (2021).
8. D. Gailani, A. Gruber, *Arterioscler. Thromb. Vasc. Biol.* **36**, 1316 (2016).
9. G. G. Schiattarella *et al.*, *Nature* **568**, 351 (2019).
10. A. B. Salido-Medina *et al.*, *Biomed. Pharmacother.* **149**, 112910 (2022).
11. M. Tate *et al.*, *Front. Pharmacol.* **12**, 719290 (2021).
12. J. P. Piccini *et al.*, *Lancet* **399**, 1383 (2022).
13. G. Poenou *et al.*, *Vasc. Health Risk Manag.* **18**, 359 (2022).
14. D. Kotecha *et al.*, *J. Am. Coll. Cardiol.* **68**, 2217 (2016).
15. S. Mishra, D. A. Kass, *Nat. Rev. Cardiol.* **18**, 400 (2021).


## ACKNOWLEDGMENTS


The authors are supported by National Institutes of Health grants HL128215, HL147933, HL155765, and HL164586 to J.A.H. and HL157697 to D.T. and by American Heart Association grant CDA851313 to D.T.

10.1126/science.ade2538


NEURODEGENERATION

# Balancing neuronal circuits

## Correcting synaptic defects in development delays Huntington's disease symptoms in older mice


*By* **Sonja Blumenstock**[1,2,3] *and* **Irina Dudanova**[1,2,4]


Neurodegenerative diseases are associated with old age. However, by the time clinical symptoms manifest, many pathological changes have already occurred in the brain. Owing to the brain's powerful ability to adapt and compensate, these changes go unnoticed until the damage becomes overwhelming. But how early do these changes start? Could neurodegeneration be the result of neuronal defects that occur in prenatal development and childhood? On page 1398 of this issue, Braz *et al.* (*1*) show that subtle synaptic impairments are already present in newborn mice carrying a mutation that causes Huntington's disease (HD).

HD is a hereditary neurodegenerative movement disorder caused by an abnormal expansion of a repetitive stretch of DNA in one allele of the Huntingtin (*HTT*) gene. This results in the presence of mutant HTT protein with an abnormally long stretch of glutamine residues, in addition to normal HTT. HD leads to severe neurodegeneration, especially in the basal ganglia and cerebral cortex. The disease typically manifests in midlife and is inevitably lethal. Recent research exploring the brain structure of mouse embryos and human fetuses found several developmental defects caused by expression of mutant HTT (*2*–*4*). These unexpected findings raised the possibility that HD has a profound impact on the brain much earlier than previously thought.

Building on those results, Braz *et al.* probed the function of cortical circuits in newborn HD mice. Using electrophysiological recordings, the authors discovered reduced neurotransmission at excitatory synapses. In addition, they observed underdeveloped dendrites and a reduction in synaptic glutamate receptors. Intriguingly, several days later, these differences disappeared and synaptic activity became indistinguishable from that of control mice, suggesting that the brain can quickly adapt to and overcome these defects. However, transient developmental disturbances caused by mutant HTT do not remain without consequences: Even temporally restricted mutant HTT expression during the first three postnatal weeks is enough for mice to develop HD-related symptoms later in life (*5*).

Braz *et al.* attempted to fix the early synaptic impairments by treating newborn mice with CX516, a drug that binds to and modulates glutamate receptors, enhancing excitatory neurotransmission. Not only did this short-term treatment (daily for the first week of life) restore the shape of the dendrites in young mice, but it was also sufficient to prevent brain atrophy and sensorimotor behavioral deficits in older animals. A related compound, CX929, has already been used in adult HD mice, where it improved some disease phenotypes but not motor defects (*6*). The current study therefore suggests that the right timing is crucial for such a treatment.

Although CX516 administration was beneficial for HD mice, it worsened the behavioral performance of healthy control animals. This observation emphasizes the subtle balance that underlies the operation of the cortical network, where either too much or too little activity has detrimental consequences. It also highlights the complexity of designing treatments for dysfunctional neural circuits, because interventions have to be fine-tuned to restore the normal level of activity without tipping the balance in the opposite direction. Achieving this balance would be very important and at the same time particularly challenging in humans.

The role of neurodevelopmental synaptic defects, and the promise of early interventions, has recently gained broad recognition for psychiatric disorders (*7*, *8*). Furthermore, it is becoming increasingly evident that neurodevelopmental and neurodegenerative diseases share converging mechanisms. Thus, synaptic impairments that dysregulate brain circuits early in life and cause clinical deficits in the long term have also been implicated in Parkinson's disease (*9*) and Alzheimer's disease (*10*). HD, being a monogenic disorder, provides an ideal experimental basis for investigating these complex pathological links, which extend almost over the entire life span


[1]Molecular Neurodegeneration Group, Max Planck Institute for Biological Intelligence, i.f., Martinsried, Germany. [2]Department of Molecules–Signaling–Development, Max Planck Institute for Biological Intelligence, i.f., Martinsried, Germany. [3]Department of Neurobiology, Center for Neural Circuits and Behavior, Department of Neurosciences, University of California, San Diego, CA, USA. [4]Center for Anatomy, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. Email: irina.dudanova@uk-koeln.de

and are much more difficult to characterize in the case of sporadic neurodegenerative diseases.

The study by Braz *et al.* raises several questions for future research. It will be important to determine the early circuit defects that occur in humans with the HD mutation, in addition to the structural neurodevelopmental alterations (*3*). Moreover, it will be crucial to investigate other brain circuits involved in HD (*11*), because pathological changes in subcortical regions, including the basal ganglia, play a key role in the clinical symptoms. At the molecular level, the cascade of events linking mutant HTT to changes in the synaptic machinery also remains to be explored. One interesting hint comes from the experiments performed by Braz *et al.* with mice lacking *Htt*. These mice exhibited neonatal circuit impairments similar to those of the HD mice but did not return to the normal state. The similarity of phenotypes caused by mutant HTT and by lack of HTT suggests that the impairments in HD are at least partially due to insufficient amounts of the normal version of the protein, and not solely to the toxic mutated version. This finding is important, because not all approaches to treat HD by lowering HTT expression distinguish between the mutant and normal forms of the protein (*12*). The study of Braz *et al.* reinforces the idea that decreased levels of the normal version can be deleterious. HTT has roles in multiple physiological processes, including intracellular trafficking, autophagy, and synaptic transmission (*13*, *14*), which could be disturbed when normal HTT is not present in adequate amounts. The exact nature of the compensatory mechanisms that counteract the early defects and prevent disease onset is another exciting question to be addressed.

The findings by Braz *et al.* have crucial translational implications, too. Because many HD mutation carriers are not identified until adult age, it will be necessary to determine the duration of the "window of opportunity" for potential interventions that ensure benefit later in life. Is perinatal treatment the ideal option, or would a treatment at a later presymptomatic stage also be sufficient? In addition, a lot of attention is currently focused on HTT-lowering treatments (*12*). Although these are promising, it is important to keep looking for alternative or complementary options, such as those targeting early synaptic deficits. ■


### REFERENCES AND NOTES

1. B. Y. Braz *et al.*, *Science* **377**, eabq5011 (2022).
2. M. Barnat, J. Le Friec, C. Benstaali, S. Humbert, *Neuron* **93**, 99 (2017).
3. M. Barnat *et al.*, *Science* **369**, 787 (2020).
4. M. Capizzi *et al.*, *Neuron* **110**, 36 (2022).
5. A. E. Molero *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5736 (2016).
6. D. A. Simmons *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4906 (2009).
7. Y. Bernardinelli, I. Nikonenko, D. Muller, *Front. Neuroanat.* **8**, 123 (2014).
8. O. Marín, *Nat. Med.* **22**, 1229 (2016).
9. C. A. Morato Torres *et al.*, *Int. J. Mol. Sci.* **21**, 5724 (2020).
10. T. Arendt, J. Stieler, U. Ueberham, *J. Neurochem.* **143**, 396 (2017).
11. S. Blumenstock, I. Dudanova, *Front. Neurosci.* **14**, 82 (2020).
12. S. J. Tabrizi, R. Ghosh, B. R. Leavitt, *Neuron* **101**, 801 (2019).
13. F. Saudou, S. Humbert, *Neuron* **89**, 910 (2016).
14. J. C. Barron, E. P. Hurley, M. P. Parsons, *Front. Cell. Neurosci.* **15**, 689332 (2021).



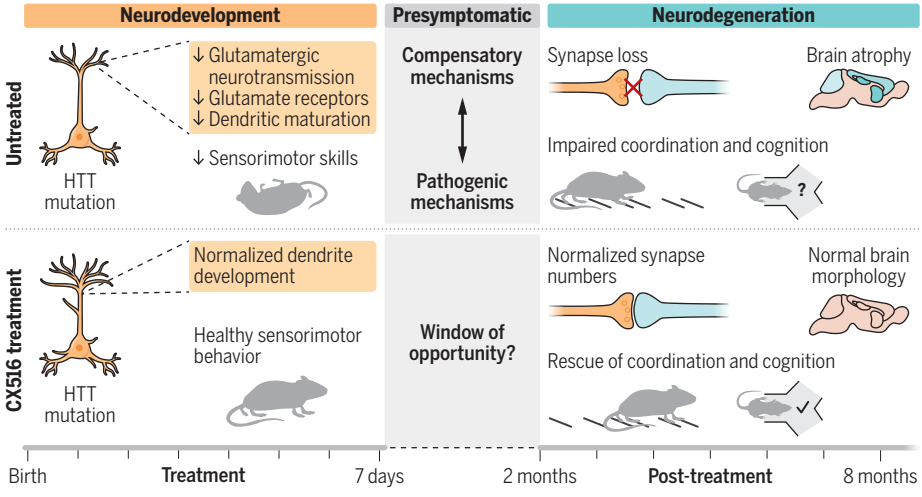
### ACKNOWLEDGMENTS

Thanks to K. Voelkl for critical reading.

10.1126/science.ade3116


## Preventing Huntington's disease in mice

Huntington's disease is caused by mutation of the Huntingtin (HTT) protein, leading to neurodegeneration in the basal ganglia and cerebral cortex. In mice with mutant HTT, excitatory synaptic activity is reduced in the cortex after birth. Boosting excitatory neurotransmission with CX516 in the first week of life prevented neurodegeneration and behavioral deficits in adult animals.




BIODIVERSITY

# Estimating global genetic diversity loss

## A mathematical framework may help inform conservation efforts

*By* **Kristen Ruegg** *and* **Sheela Turbek**

Preservation of genetic diversity is critical to the resilience of species in the face of global change. To meet international calls to preserve at least 90% of species' genetic diversity, researchers and conservationists need a way to reliably predict genetic diversity loss resulting from human activities (*1*). On page 1431 of this issue, Exposito-Alonso *et al.* present a mathematical framework that elegantly bridges biodiversity and population genetics theory to model the relationship between genetic diversity and habitat loss (*2*). This approach builds on methods already used by biodiversity policy experts for predicting species extinctions based on habitat loss (*3*) and should be useful to those tasked with setting goals for preserving genetic diversity.

The species-area relationship (SAR), one of the oldest and most well-documented relationships in ecology (*4*), describes the positive correlation between the abundance of species and the size of a habitat. The SAR has been observed to follow a power law, where the number of species is proportional to the habitat area to the power of $z$. For example, a $z$ value of two implies that the number of species is multiplied by four when the area is doubled, but for a $z$ value of one, the number of species and the habitat size would correlate linearly. Higher $z$ values are typically found in more species-rich or spatially structured ecosystems, such as rainforests. Despite its simplicity, the SAR has been very useful for predicting species extinctions as a function of habitat loss.

In addition to species loss because of habitat destruction, understanding the loss in genetic diversity within individual species is also important. However, a straightforward framework for calculating genetic

Biology Department, Colorado State University, Fort Collins, CO 80521, USA. Email: kristen.ruegg@colostate.edu

GRAPHIC: A. MASTIN/*SCIENCE*

Exposito-Alonso *et al*. claim that even species whose conservation status is considered "least concern," such as the willow flycatcher shown here, may have already lost a substantial amount of genetic diversity.

PHOTO: ALAN MURPHY/MINDEN PICTURES

diversity loss concerning reductions in habitat does not yet exist. Exposito-Alonso *et al*. adopt the mathematical framework of the SAR to demonstrate that the accumulation of genetic variants (i.e., mutations) follows a similar trend as species diversity, with the number of mutations within a species being proportional to the habitat area to the power $z_{MAR}$. The MAR subscript is used by the authors to represent what they call the mutations-area relationship (MAR).

Exposito-Alonso *et al*. evaluated the MAR in the small flowering plant thale cress (*Arabidopsis thaliana*), a model species in plant biology and genetics, and found that their model can accurately predict genetic diversity loss from habitat contractions. Because the power of the MAR lies in its potential to inform genetic diversity loss in species with or without genetic data, the authors expanded their testing dataset. In all, they tested the generality of the MAR using publicly available genomic data from 20 plant and animal species, including humans, several plants, and fruit flies. They found that the critical component of the MAR calculation, $z_{MAR}$, was quite consistent across the 20 species. Thus, this relationship may be applicable for approximating $z_{MAR}$ in species where genomic data is absent.

By combining the average $z_{MAR}$ and estimates of pre–21st century land transformations, Exposito-Alonso *et al*. conclude that an average of 10 to 16% of genetic diversity has been lost globally since the industrial revolution. This number is already greater than 10%, which is the permissible percentage of genetic diversity loss as recommended by biodiversity policy experts for healthy ecosystems moving forward (*1, 5*). Further, when this model is used to estimate the rate of genetic diversity loss for individual species, it is clear that even species classified as "least concern," such as the willow flycatcher (*Empidonax traillii*), a North American songbird, have already lost a substantial amount of genetic diversity.

Although compelling, indiscriminately applying the average $z_{MAR}$ to all taxa comes with some caveats. For instance, the confidence intervals surrounding the estimate of global genetic diversity loss are very wide (ranging from 0 to 100%). This raises some questions as to its broad applicability. For instance, different ecosystems have been affected by human activities to different degrees, and there is a need for a more granular approach to estimating genetic diversity loss. As Exposito-Alonso *et al*. themselves have pointed out, high-altitude ecosystems have only lost 0.3% of their area, whereas highly managed forests have lost a whopping 67% of theirs. Additional variations may also arise from biological differences between species (i.e., ability to disperse, mating systems, geographic ranges), which will, in turn, influence patterns of gene flow across space. Although the authors investigated the potential influence of between-species variation in these traits on $z_{MAR}$ and found no statistically significant associations, their sample size was limited to 20 species and more work in this area is needed for this to be conclusive. The ability to further test the robustness of the global genetic diversity loss calculations should improve as more landscape and genomic data become available.

Despite its potential limitations, this framework for calculating genetic diversity loss as a function of habitat loss holds promise for conservation biologists and policy experts charged with species preservation in the face of rapid environmental change. Further testing of the MAR with empirical and simulated datasets will reveal additional insights into the broad-scale utility of global genetic diversity loss estimates for conservation efforts. In a time when habitat loss and climate change are altering ecosystems faster than scientists can study them, the MAR will be a vital tool for scientists and policy-makers who are attempting to understand the magnitude of past genetic diversity losses and plan for the future. ■

## REFERENCES AND NOTES


1. Convention on Biological Diversity, "First draft of the post-2020 Global Biodiversity Framework" (CBD/WG2020/3/3, United Nations Environment Programme, 2021).
2. M. Exposito-Alonso *et al*., *Science* **377**, 1431 (2022).
3. S. G. Potts, V. L. Imperatriz-Fonseca, H. T. Ngo, Eds., "The assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production" (Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, 2016).
4. O. Arrhenius, *J. Ecol*. **9**, 95 (1921).
5. S. Díaz *et al*., *Science* **370**, 411 (2020).


## ACKNOWLEDGMENTS


This work was supported by a National Science Foundation (NSF) CAREER award to K.R. (008933-00002) and an NSF Postdoctoral Research Fellowship in Biology (2208881) to S.T.

10.1126/science.add0007

POLICY FORUM

ENERGY AND CLIMATE

# Toward a European carbon footprint rule for batteries

## Focus on upstream production, not downstream use

*By* **Leopold Peiseler[1,2,3], Christian Bauer[4], Martin Beuse[1,5], Vanessa Wood[2,3], Tobias S. Schmidt[1,3]**

Lithium-ion batteries (LIBs) are a key decarbonization technology for transport and electricity sectors (*1*). Governments, including the European Commission (EC), stress LIBs' relevance from a climate and "green" industrial policy standpoint (*2*). However, producing LIBs causes substantive greenhouse gas (GHG) emissions—for example, from fossil fuel use in mining raw materials or energy consumption during battery production (*3*, *4*). Hence, the EC aims to address this issue through a new regulation proposed in late 2020 (*5*, *6*). Article 7 of the proposal mandates a carbon footprint (CF) declaration from mid-2024 and sets upper CF limits for European markets, which will be applicable from 2027 for electric vehicle (EV) batteries and stationary batteries with more than 2 kilowatt-hours (kWh) of storage capacity. Here, we aim to inform the debate on the European CF regulation by discussing regulatory design options, along with their advantages and disadvantages, including implementation barriers and potential adverse incentives.

Batteries having more than 2 kWh of storage capacity would be used largely for EVs (20 to 120 kWh), residential storage (5 to 20 kWh), commercial and industrial storage (50 to 5000 kWh), and grid-scale storage (250 to 1000 MWh). To calculate the CF of LIBs, the EC recommends the use of their Product Environmental Footprint Category Rules (PEFCR), a set of guidelines to calculate products' environmental impacts. But regulating the CF of batteries is not straightforward for two reasons. First, a battery's overall GHG emissions strongly depend on its useful lifetime. Batteries degrade depending on any one of hundreds of combinations of battery design and application. Second, the industry's supply chain complexity and secrecy make data access and governance difficult. An ill-designed directive might backfire and have adverse effects on actors along the LIB supply chain. This includes actors in non-European countries who currently manufacture the majority of LIBs and export (components of) LIBs into Europe, the second-largest battery market. On the basis of an in-depth understanding of battery degradation in different applications, we suggest that the use and end-of-life phases should be excluded from the CF calculation and instead be addressed through separate instruments.

[1]Energy and Technology Policy Group, ETH Zurich, Zurich, Switzerland. [2]Materials and Device Engineering Group, ETH Zurich, Zurich, Switzerland. [3]Institute for Science, Technology and Policy, ETH Zurich, Zurich, Switzerland. [4]Laboratory for Energy Systems Analysis, Paul Scherrer Institute, Villigen, Switzerland. [5]HagerEnergy GmbH, Osnabrück, Germany. Email: tobiasschmidt@ethz.ch

### EMISSIONS ACROSS THE VALUE CHAIN

A key regulatory design question relates to the locus of emissions in the value chain of LIBs, which comprises multiple heterogeneous sectors (see the figure). Material suppliers extract and convert natural resources into high-purity substances and precursors. Battery producers use materials, metals, and other chemicals to manufacture LIB cells that are further integrated into application-specific battery modules or packs. System integrators procure these batteries and produce final products, such as EVs or residential storage systems. The integrators are typically the distributors within the EU and as such are the actors directly targeted by the regulation. Once the batteries are distributed, users operate them in varying forms and settings. For example, the battery of an EV can also be used for vehicle-to-grid applications, and—once unfit for mobile applications—can later be converted into a stationary battery (second life). Once fully retired, it goes to the recycling industry, where valuable materials are extracted, refined, and ideally fed back into the commodity market. Emissions occur at every stage of the value chain, often because of heat and electricity requirements (*7*).

To determine the CF of a battery, two fundamentally different scopes of analysis can be used (*7*). The first scope includes all upstream emissions to produce 1 kWh of battery storage capacity. It only requires information about the material, production, and integration phases. In the second scope, all occurring emissions (upstream and downstream) are related to 1 kWh of energy throughput. Because throughput is determined by the energy stored and discharged throughout a battery's life span, the battery's use phase must be considered.

Because production emissions are "spread out" over longer lifetimes under a throughput-based approach, increasing battery longevity and use is a key lever for reducing the CF (*3*, *4*). Particular attention must be paid in the case of LIBs because the products' throughput and thus lifetimes are heavily dependent on different cell chemistries, designs, and applications (*8*–*13*). Policy-makers need to account for this diversity and complexity in the use phase (see the figure) and integrate technical understanding into the regulatory design process.

### REGULATORY DESIGN OPTIONS

Although the EC proposes a throughput-based scope, how this should be implemented in detail is unclear. Because the CF must be declared when the batteries enter the European market (before they are used), a regulatory design must be developed to estimate the throughput a priori. Assuming that all batteries last equally long does not do justice to the complexities in battery design and applications. Treating batteries uniformly would defeat the purpose of the regulation. Thus, battery-specific lifetime assessments would be necessary.

#### Test-based option

To determine the throughput of individual batteries, the regulator can define realistic, application-specific test cycles for all batteries designated for the European market. Each battery must demonstrate how long—the number of cycles—it can be operated under these test conditions. The resulting throughputs of the batteries could then be used to calculate their CF.

Once a cell is retired, its actual energy throughput can be determined and reported, for example, by using the foreseen European "battery passport" that could facilitate the logging of usage data for each battery. If the last actors in the value chain, recyclers, are responsible for evaluating the usage data stored on the passport, the CF of the recycling phase could also be incorporated (see the figure). If recyclers report to the EC that a battery delivered more (or less) energy in reality than anticipated

from initial testing, the EC could allocate a carbon credit to (or impose a malus on) the battery's distributor. This design would incentivize product design for longer life.

However, there are considerable drawbacks to this option. First, the regulator would be required to find consensus on standardized yet application-specific and realistic test cycles, a lengthy task prone to lobbying and gaming by industrial players. To determine the throughput, Article 7 and the Proposal's annex build on the battery-specific PEFCR, which makes use of a normed charge-discharge procedure. This appears problematic given its unrealistic testing conditions—for example, using favorable temperature ranges at uniform and substantially lower charging currents as compared with, for example, that of fast-charging an EV. This does not do justice to the widely varying application profiles and environments that determine degradation dynamics of batteries. The PEFCR may also undermine the comparability of the CF by opening the door for company-specific, nonstandardized cycle tests.

Second, cell manufacturers and system integrators would need to test every battery type intended for the European market. Even moderate design or manufacturing modifications require retesting because these can influence lifetime characteristics (*9*, *12*). Because of quality fluctuations, testing would require representative batches of cells with one given configuration at a time (*12*). All of this would result in high testing costs, penalizing smaller and newer cell producers and thus potentially constraining innovation.

Third, testing can only simulate battery degradation to a limited degree. Battery degradation is driven by interaction of use-based (cycle) and time-based (calendar) aging (*13*). The time-based degradation process is especially hard to simulate in accelerated tests, and despite existing research in this field, there is not a universally agreed-on degradation model. As such, the accuracy of the degradation estimates and test procedures will only be determined many years if not decades later.

### Warranty-based option

As an alternative, LIB throughput, and thus CF, may be derived from existing warranties that typically define the minimum lifetime and permissible operating conditions for specific applications. Warranties exist on the business-to-business level, such as a minimum cycle life for stationary peak shaving operation, and on the business-to-consumer level, such as a minimum kilometer range for an EV (see the figure).

With this design, battery producers are spared expensive degradation tests beyond existing testing. Instead, they can align the lifetime estimate with their customers' individual operational conditions. Implementing this option uses existing documents negotiated between stakeholders in the value chain. This option could incentivize battery producers to target longevity but not overpromise the performance and durability of their devices.

However, this approach has drawbacks. First, although it covers the use phase without causing additional testing costs, further regulatory instruments might be necessary to address aspects that are not contained in warranties, such as second-life applications or recycling. Second, many warranties are negotiated on a business-to-business level and are confidential. Revealing them, even partly, would mark a noticeable intervention in the market, requiring mechanisms that guarantee confidentiality. Business-to-consumer warranties are public but hardly standardized, even within one application field, severely limiting their practical usefulness. Last, warranties are always part of a business model, sometimes justifying deliberate decisions against technical realities. For example, warranties for combustion cars are usually far shorter than their actual service life, making them ill-suited proxies for their actual lifetimes. Such short warranties would artificially inflate the CF of LIBs.

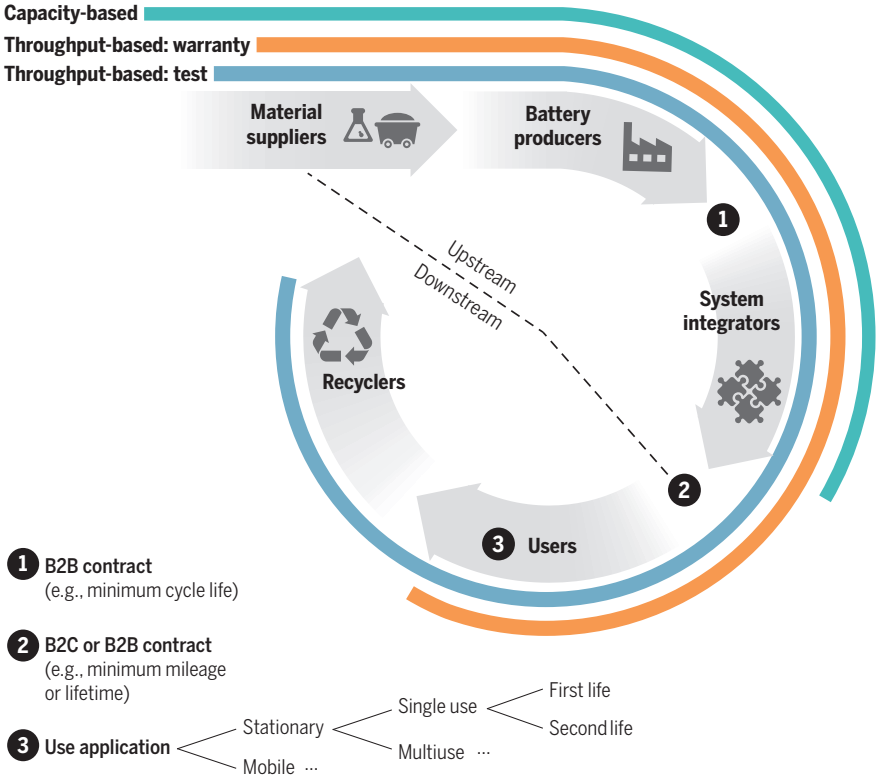
### Capacity-based option

Although throughput-based design options can incentivize long-lasting batteries, their practical implication barriers described above are not to be underestimated. Thus, the EC should consider an alternative approach: limiting the calculation scope to the simpler capacity-based option.

The capacity-based option, the default scope in life-cycle assessment literature, only considers upstream material, manufacturing, and integration emissions, thus massively reducing bureaucratic efforts for both the regulator and industry (see the figure). At the same time, such a narrow scope deliberately foregoes various incentive opportunities along the value chain—most importantly, long-lasting batteries. However,

## Three regulatory design options

This simplified value chain illustrates capacity-based and throughput-based (through warranty and testing) approaches to lithium-ion battery (LIB) carbon footprint regulation. Upstream and downstream reflect actors or processes before and after, respectively, the battery enters the European market for the first time. Opportunities for business-to-business (B2B) and business-to-consumer (B2C) warranties are illustrated, as is complexity in the use phase, where heterogenous LIBs can be deployed in heterogeneous and evolving combinations and applications.




GRAPHIC: N. CARY/*SCIENCE*

we argue in favor of the classic economic policy rule that each policy aim should be addressed by its own instrument (*14*) instead of overburdening the CF regulation. For example, industrial recycling will be stimulated by mandating recycling quotas (Article 57) and the minimum recycled content for new batteries (Article 8), as is currently planned by the EC. In addition, the GHG impact of recycling processes can be included in the CF of recycled material. To incentivize throughput-increasing activities such as second-life applications, a credit system such as the "eco-innovation" from the European Union (EU) emission standard (Regulation 2019/631) can be introduced. Stakeholders with innovative low-carbon activities or inventions can apply for credits accredited, for example, by the scientific arm of the EC, the Joint Research Centre (JRC).

Even if the use-phase is excluded from the CF calculation, there is an inherent market incentive for sufficiently long-lasting batteries. The already existing alignment of economic and ecological interests of both regulator and markets raises doubts about the marginal effects of longevity incentives under a throughput-based approach. Even if the EC would overcome the barriers and strike a balance between meaningful data and costs, the additional policy efforts and room for gaming seem difficult to justify in light of already existing market incentives for sufficient longevity.

## DISCLOSURE AND GOVERNANCE

Regardless of the regulatory design, emissions must be accounted for in the upstream supply chain where—particularly for mining, refining, and the production of cell components—there are substantial differences in GHG emissions between batteries from different manufacturers. However, because of limited data availability, use of proxies and industry averages is common in life-cycle assessment (*4*, *7*). Because a regulator needs to balance data quality with feasibility and compliance costs, it is unrealistic to expect complete data transparency throughout the entire LIB value chain. Yet relying on secondary data sourced from existing life-cycle inventories or reports rather than on primary data directly collected bears the risk of being too generic and obtaining inaccurate estimates, undermining the effectiveness of the regulation.

To strike a balance between practical implementation and effectiveness, we suggest that the EC provide standardized CF default values for specific upstream processes or battery components if a supplier cannot provide credible and transparent calculations. The JRC should regularly update these default values, which should be informed by industry self-declarations and independent sources. To incentivize corporate engagement and disclosure, the default CF values of the components should be high enough that if a battery producer does not disclose any primary data but instead only relies on default values for processes and procured components, the overall CF of its batteries is above the maximum CF threshold defined in the regulation. Batteries with no underlying data disclosure would thus be assigned the lowest CF performance class (in 2026) or not allowed on the European market (from 2027 onward).

Companies that aim to reduce the CF of their batteries would obtain a strong incentive to source from suppliers with low-carbon processes and high data transparency. Whereas the diverse and fragmented upstream supply chain currently impedes a transparent flow of information, a system of sufficiently high default values stimulates market dynamics that incentivize upstream value chain actors to be more transparent about and reduce their carbon emissions.

Battery producers have voiced concerns that disclosure of primary data, such as energy usage or used materials, can be misused to reverse-engineer company secrets. This highlights the need for independent auditors to collect and verify individual CF calculations confidentially. Cell producers would provide auditors with detailed CF calculations that are based on a granular breakdown of components, processes, and associated emissions. Once verified, auditors can aggregate the underlying data into an overall battery CF and report this value to the JRC. This confidentiality should also apply to future battery technologies and manufacturing practices. Auditors need to liaise with the regulator to harmonize methodological guidelines about upcoming technologies without compromising company-specific details.

Exchange between industry and regulator is paramount for ensuring the regulation's effectiveness. It is also in everybody's interest because it allows an up-to-date and realistic assessment of the industry's CF practices. On the basis of this information, the JRC can decide on sensible maximum thresholds for CF beyond 2027. The maximum CF threshold system should follow a dynamic worst-in-class exclusion mechanism, in which the CF of the worst-performing LIBs during one period will be excluded in determining the threshold for the next period. Through this, an industry-wide incentive for continuous innovation is created.

## INVALUABLE EXPERIENCE

The proposal's details are being negotiated in a "trilogue" between the EC, the Parliament, and the Council, and an agreement is expected in the course of 2022. Because of the advanced state of the proposal, major unannounced changes are unlikely, and a throughput-based approach is expected to be "locked" into European law. However, by mid-2023 the EC is mandated to adopt a Delegated Act, specifying methodological details in accordance with essential elements of the Proposal's annex. If by then a consensus in favor of the capacity-based design has emerged, we believe that a de facto capacity-based method could still be implemented through the delegated act and the PEFCR, without needing to change the regulation.

We anticipate that any throughput-based design option for quantifying the CFs of batteries may fall short of its expectations and do more harm than good. This proposed regulation departs from the production-based carbon-accounting approach of international climate agreements (*15*) and instead deploys a consumption-based approach in which the EU regulates emissions that (partly) occur overseas. Because there is no public regulation setting similar standards on other products, the proposal is a first of its kind and can play a pioneering role in climate change policy. Experiences with this regulation are invaluable for non-European regulators and for the regulatory design of products that share characteristics with batteries. Thus, the European take on addressing the CF of complex, multipurpose technologies will be closely observed internationally. ■

### REFERENCES AND NOTES


1. M. S. Ziegler, J. Song, J. E. Trancik, *Energy Environ. Sci.* **14**, 6074 (2021).
2. M. Beuse, T. S. Schmidt, V. Wood, *Science* **361**, 1075 (2018).
3. T. S. Schmidt *et al.*, *Environ. Sci. Technol.* **53**, 3379 (2019).
4. J. F. Peters, M. Baumann, B. Zimmermann, J. Braun, M. Weil, *Renew. Sustain. Energy Rev.* **67**, 491 (2017).
5. EC, "Proposal for a Regulation on batteries and waste batteries" (EC, 2020).
6. H. E. Melin *et al.*, *Science* **373**, 384 (2021).
7. E. Crenna, M. Gauch, R. Widmer, P. Wäger, R. Hischier, *Resour. Conserv. Recycling* **170**, 105619 (2021).
8. M. Ebner, F. Marone, M. Stampanoni, V. Wood, *Science* **342**, 716 (2013).
9. Y. Preger *et al.*, *J. Electrochem. Soc.* **167**, 120532 (2020).
10. H. Wang, J. F. Whitacre, *J. Energy Storage* **35**, 102244 (2021).
11. M. Elliott, L. G. Swan, M. Dubarry, G. Baure, *J. Energy Storage* **32**, 101873 (2020).
12. T. Baumhöfer, M. Brühl, S. Rothgang, D. U. Sauer, *J. Power Sources* **247**, 332 (2014).
13. P. S. Sabet *et al.*, *J. Power Sources* **449**, 227369 (2020).
14. W. Knudson, *Bull. Sci. Technol. Soc.* **29**, 308 (2009).
15. S. Afionis, M. Sakai, K. Scott, J. Barrett, A. Gouldson, *Wiley Interdiscip. Rev. Clim. Change* **8**, e438 (2017).


10.1126/science.abp9505

BOOKS *et al.*

Flooding threatens military assets at Offutt Air Force Base in Bellevue, Nebraska, in 2019.

PHOTO: Z LONG/OMAHA WORLD-HERALD/AP IMAGES

CLIMATE CHANGE

# Greening national security

## The US military's eye-watering carbon footprint must be mitigated, argues a political scientist

*By* **Oliver Belcher**

The United States' reliance on Middle Eastern oil featured prominently in arguments leading up to the 2003 invasion of Iraq, but the potential effects of that war on climate change were not a major consideration at the time. In her new book, *The Pentagon, Climate Change, and War*, political scientist Neta Crawford argues that they probably should have been. She notes that between 2001 and 2017—the height of the United States' wars in Iraq and Afghanistan—the US military emitted 1.2 billion metric tons of greenhouse gases, including 458 million metric tons of carbon dioxide that can be directly attributed to post-9/11 war-related fuel consumption.

Crawford's wide-ranging, meticulously researched volume focuses on the US defense establishment's carbon-based energy dependency and the military's outsized institutional role in global environmental change. The book offers the best analysis to date of the US military's contribution to climate change, including emissions from the Iraq and Afghanistan conflicts.

Crawford goes beyond the usual historical overviews of the military's reliance on fossil fuels in the 20th and 21st centuries, although she deftly covers that territory as well, given that the US military is a climate actor whose consumption of hydrocarbon-based fuels contributes directly to climate change. While the military accounts for a relatively small portion of the United States' total annual emissions, it is still the single largest institutional consumer of energy in the world. If the US military were a country, it would fall between Peru and Portugal in the global league tables of fuel purchasing (*1*).

**The Pentagon, Climate Change, and War**
*Neta C. Crawford*
MIT Press, 2022. 392 pp.

The book is divided into four parts, the first of which explores the relationship between economic growth, fossil fuel use, American wars, and foreign oil dependency. Crawford calls these entanglements "the deep cycle," to underscore how the US economy's reliance on foreign energy resources, especially in times of crisis (e.g., the 1970s oil embargo or Russia's ongoing attack on Ukraine), is a key consideration in ever-evolving US strategic and military doctrines. This is familiar terrain for conventional histories on geopolitics and oil (*2*, *3*).

Crawford pivots in part 2 to the fraught contemporary relationship between the US military and mainstream climate science. Here, she notes that the US military was a formative influence in the development of modern climate science. At the height of this relationship early in the Cold War, the Department of Defense invested heavily in university-based and private climate science to calculate the possible effects of a nuclear war on the atmosphere (*4*). By the late 1990s, however, the military's enthusiasm for such considerations had begun to wane, culminating with US negotiators' successful exclusion of US military emissions from the 1997 Kyoto treaty—a glaring omission made for "national security" reasons.

In part 3, Crawford considers the possible consequences of climate change on US national security. While some politicians persist in sowing doubt about climate science, the Department of Defense has been quietly preparing for a geopolitical world radically altered by climate change. Crawford argues, however, that it is not enough for the United States to adapt to a changing geopolitical landscape. Mitigation measures, such as dedicating land at military installations to carbon sequestration, must be part and parcel of any viable security strategy, she maintains, and reducing military emissions is itself an important security measure that must be undertaken.

In the book's final chapter, "The Path to Climate Security," Crawford makes the compelling case that if climate change poses more of a long-term national security threat than many other threats, then the Department of Defense needs to rethink its global force posture "beyond adaptation…to true climate change-related conflict prevention by further reducing fuel use and greenhouse gas emissions." This would entail drastically reducing the United States' reliance on Persian Gulf oil.

Crawford is not naïve about the extent to which the Biden administration, which recently rekindled relations with Saudi Arabia, or future administrations will seek to effectively mitigate the military's carbon emissions. Nevertheless, her book is a brilliant guide for how we might effectively confront global warming and is filled with practical measures we can implement along the way. ■

The reviewer is at the School of Government and International Affairs, Durham University, Durham DH1 3TU, UK. Email: oliver.belcher@durham.ac.uk

### REFERENCES AND NOTES

1. O. Belcher, P. Bigger, B. Neimark, C. Kennelly, *Trans. Inst. Br. Geogr.* **45**, 65 (2020).
2. D. Yergin, *The Prize: The Epic Quest for Oil, Money & Power* (Simon & Schuster, 1990).
3. A. J. Bacevich, *America's War for the Greater Middle East: A Military History* (Random House, 2016).
4. P. N. Edwards, *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming* (MIT Press, 2010).

10.1126/science.add9472

SCIENCE AND SOCIETY

# Digital spaces and their perils

## A pair of books confront the dark sides of virtual communities

By **Jennifer Golbeck**

Readers with an interest in the internet ecosystem have two new books that explore complementary themes from which to choose this fall: *Like, Comment, Subscribe*, which goes deep on a platform, and *Meme Wars*, which goes deep on a movement.

*Like, Comment, Subscribe*, written by journalist Mark Bergen, takes readers on an inside journey through the advent and meteoric rise of the video-sharing platform YouTube. When the average person is asked to list social media platforms, YouTube is usually an afterthought. Yet it has more than 2 billion users who watch more than 3 billion hours of content each month. Creators also appreciate it as a platform that has reliable revenue sharing that can support a business.

Beginning with the birth of the platform in a literal garage, *Like, Comment, Subscribe* follows YouTube's early challenges and its evolution. Moderation was a problem from the start, we learn, as the company's small team of 10 toiled away in 2005 in a rat-infested office trying to stem the tide of pornography, animal abuse, and shock videos posted by early users. This challenge never fully faded, and questions about how to handle hate speech and anti-science conspiracy theories continue to plague moderators today.

But Bergen's book is not just about managing YouTube's content. It also details the difficult and evolving interplay among content creators, their communities, the platform's recommendation algorithms, and its need to generate profit.

YouTube thrives because of the good content that creators post and the people who engage with it. Yet users, especially creators, often struggle in the face of corporate decisions and algorithmic tweaks that drastically affect their ability to make money, that subject them to harassment, and that create copyright battles that privilege rich and powerful actors.

For those who follow the tech ecosystem, *Like, Comment, Subscribe* will not bring dramatic new insights. However, it captures critical snapshots from YouTube's history with great texture, and Bergen places those moments insightfully into a broader context. For anyone with an interest in the growth and impact of platform policies, the book offers a wealth of information to consider.


**Like, Comment, Subscribe**
*Mark Bergen*
Viking, 2022. 464 pp.


**Meme Wars**
*Joan Donovan, Emily Dreyfuss, Brian Friedberg*
Bloomsbury, 2022. 432 pp.

To shift perspective and dive into a problem that exists across platforms, pick up *Meme Wars*, which explores the 6 January 2021 attack on the US Capitol and the forces that helped build it. Harvard researchers Joan Donovan, Emily Dreyfuss, and Brian Friedberg, the authors of this book, write of the attack: "For those who had been watching these communities…the events of that day were entirely foreseeable. They were tragic. And sad. But they were not unexpected." The dramatic story of how dozens of different right-wing groups joined forces to perpetrate a unified, violent attack on American democracy, and why such an event had been anticipated by researchers who read online forums, finds commonalities among online movements such as Gamergate, a misogynistic social media–based harassment campaign; the conspiracies that inspired mass shooter Dylann Roof; and the political rise of Donald Trump.

"Meme"—an image, video, or catchphrase that spreads online—is broadly interpreted in this text, and the authors make a convincing argument that the political right has successfully "meme-ified" their arguments, uniting supporters with messages that exclude outsiders, make followers feel smart and included, and simplify complex issues in appealing ways. For people who feel forgotten and minimized or like they are losing power, a meme-filled community can be very appealing.

There are deep social issues and conditions that give rise to any extremist movement, but these are not the focus of *Meme Wars*. Those who want a comprehensive picture of the broader buildup to the 6 January attack will find important details missing from this book. But for readers curious about the roles played by social media personalities who exploited online ecosystems and fueled existing social divisions prior to the attack, *Meme Wars* offers a great deal of insight.

The online world has long ("long" being relative) been considered separate from the so-called real world. As these two titles show, however, such a distinction is not accurate. Together, *Like, Comment, Subscribe* and *Meme Wars* give readers new ways to understand the many connections that exist between our virtual and corporeal lives. ■

The reviewer is at the Social Intelligence Lab, University of Maryland, College Park, MD 20742, USA.
Email: golbeck@cs.umd.edu


Simple slogans that make followers feel smart and included can appeal to those who feel forgotten.

PHOTO: THOMAS O'NEILL/NURPHOTO/GETTY IMAGES

10.1126/science.add5970

Water rushes down stairs in a New York City subway station during flash floods caused by Hurricane Ida in 2021.

PHOTO: ANTHONY BEHAR/AP IMAGES

*Edited by* **Jennifer Sills**

# Education can improve response to flash floods

As climate change intensifies and torrential rainfall and extreme floods increase (*1*), flash floods, one of the deadliest hazards because of their sudden onset (*2*), are becoming more frequent and widespread. Accurate flash flood early warning systems can mitigate casualties, but even the best warnings cannot protect people who disregard them. Public awareness is a crucial aspect of flash flood safety.

Scientific advances have led to extensive use of flash flood monitoring systems and accurate early warnings (*3*), and efforts to mitigate harm continue to focus on hydrometeorological monitoring, flood forecasting, risk analysis, and countermeasures. The contribution of human factors such as perception and behavior to reducing harm has received less attention (*4*).

Social behavior has undermined the success of flash flood early warning systems across the world. In 2021, flash flood warnings were broadcast widely hours in advance of hazardous conditions in China, central Europe, and the United States (*5*–*7*). Yet the predicted floods led to more than 300 casualties in Zhengzhou, China and 200 in central Europe, mostly local residents. In New York, despite warnings, people were caught by floods in unexpected locations, such as subway stations. On 13 August, flood warnings were issued 1 hour in advance of a flash flood in Longcao Stream, Sichuan Province, China (*8*), where people had come from cities to escape a heat wave. About 100 people refused to leave and 7 tourists died (*9*). This failure of social response and organization should be addressed through education and clearer dissemination of information.

Social education, based on social science data about the most effective way to convey information, is essential to enhancing the public's perception of and response to natural hazards (*10*). Such education is well established in regions vulnerable to traditionally abrupt natural hazards, such as earthquakes and tsunamis. In hazard-prone countries such as Japan and New Zealand, drills are conducted regularly and local inhabitants are familiar with the hazards and evacuation routes (*11*, *12*). Unfortunately, awareness campaigns for flash floods are less common (*5*).

People in traditionally high-risk areas, those in places where flash floods have been rare until recently, and those who live in low-risk regions but might travel to locations where flooding is possible all need easy access to details about disaster identification, emergency response, and self-rescue. Local governments in disaster-prone countries or regions should regularly disseminate disaster response information through school education, social media, and social networks. Regular drills, similar to those in Japan and New Zealand, should also be conducted, not only in people's daily living and working environments but also in places that are usually ignored, such as tourist attractions. Education must include clear instructions, targeting people who do not understand what action to take as well as those who might be tempted to dismiss directives to vacate the area. To maximize the effect of early warnings, we must increase the likelihood that all people will heed them.

**Lu Wang[1], Ruihua Nie[1], Louise J. Slater[2], Zhonghou Xu[3], Dawei Guan[4], Yifan Yang[5]***

[1]State Key Laboratory of Hydraulics and Mountain River Engineering, College of Water Resource and Hydropower, Sichuan University, Chengdu, China. [2]School of Geography and the Environment, University of Oxford, Oxford, UK. [3]National Institute of Water and Atmospheric Research, Hamilton, New Zealand. [4]Key Laboratory of Ministry of Education for Coastal Disaster and Protection, Hohai University, Nanjing, China. [5]School of Engineering, The University of Waikato, Hamilton, New Zealand.
*Corresponding author.
Email: yifan.yang@waikato.ac.nz

## REFERENCES AND NOTES

1. L. Slater *et al.*, *Geophys. Res. Lett.* **48**, e2020GL091824 (2021).
2. M. Borga *et al.*, *WIREs Water* **6**, e1338 (2019).
3. D. Sun *et al.*, *Chin. Flood Drought Manag.* **32**, 24 (2022) [in Chinese].
4. J. M. Bodoque *et al.*, *J. Hydrol.* **568**, 769 (2019).
5. W. Cornwall, *Science*, 10.1126/science.abl5271 (2021).
6. R. E. Morss *et al.*, *J. Hydrol.* **541**, 649 (2016).
7. X. Guo, A. Zhu, Q. Li, R. Chen, *Science* **374**, 831 (2021).
8. H. Wan, "7 killed, 8 injured in flash floods in SW China's 'new Instagram-worthy' valley" *Global Times* (2022).
9. N. Chung, "China 'wild trip' deaths put social media

influencers under spotlight," *The Guardian* (2022).
10. M. Siudak, in Coping with Flash Floods, E. Gruntfest and J. Handmer, Eds. (Springer, Dordrecht, 2001), pp. 15–18.
11. A. Komen, *Austral. J. Emerg. Manag.* **27**, 21 (2012).
12. Tauranga City Council, "Tsunami awareness programme" (2022); www.tauranga.govt.nz/community/civil-defence/tsunami/tsunami-awareness-programme.

10.1126/science.ade6616

# Time to improve disaster preparedness in Brazil

Extreme events are growing in frequency and intensity worldwide (*1*). In Brazil, national policy exacerbates natural disasters by failing to consider risk reduction. Real estate speculation and construction encourage urban development in risky locations, such as areas close to water bodies and high slopes (*2, 3*), and efforts to raise public awareness of the risks are rare. The next administration should take steps to improve the country's disaster prevention and response.

In the past 2 years, Brazil has faced landslides, floods, and wildfires across the country that have affected millions of people and led to hundreds of deaths (*4*–*6*). In the country's northeast region alone, floods and landslides caused an estimated economic loss of US$625 million between December 2021 to May 2022 (*7*). During this time, difficult access to funding and precarious infrastructure hindered the ability of Brazil's municipal civil defense agencies to offer support. For instance, 59% of municipal agencies have only one or two members, 67% do not have vehicles, and 30% do not have a computer (*8*). Yet, Brazil spent only 50% (US$154 million) of funds budgeted for civil protection in 2020 and only 41% (US$90 million) in 2021 (*9*).

Brazil's presidential election will take place on 2 October. Despite the country's poor record on disaster response, disaster risk reduction strategies have not been addressed by the candidates. Brazil's next administration could strengthen disaster preparedness by improving land use regulation to prevent development in risky areas and by amending the forest code to prevent the removal of vegetation that, for example, keeps the soil stable in slopes. The government could also prioritize prevention measures such as early warning systems and improve communications used by local responders. Climate change will only accelerate these events, and Brazil's next administration must be prepared to deal with them.

**Maria Lucia F. Barbosa[1]*, Isadora Haddad[1], Liana O. Anderson[2]**
[1]Tropical Ecosystems and Environmental Sciences Laboratory (TREES), National Institute for Space Research (INPE), São José dos Campos, SP 12227-010, Brazil. [2]National Centre for Monitoring and Early Warning of Natural Disasters (CEMADEN), São Jose dos Campos, Brazil.
*Corresponding author.
Email: malucsp@gmail.com

## REFERENCES AND NOTES

1. A. AghaKouchak *et al.*, *Annu. Rev. Earth Planet. Sci.* **48**, 519 (2020).
2. M. Lahsen, G. D. A. Couto, I. Lorenzoni, *Clim. Change* **158**, 213 (2020).
3. R. Coates, *Int. J. Disast. Risk Reduct.* **74**, 102918 (2022).
4. E. Alcântara *et al.*, *Nat. Hazards Earth Syst. Sci. Discuss.* [preprint], 10.5194/nhess-2022-163 (2022).
5. R. Dalagnol, *Clim. Resil. Sustain.* **1**, e15 (2022).
6. Confederação Nacional dos Municípios, "Danos E Prejuízos Causados por Desastres no Brasil entre 2013 A 2022" (2022); www.desastres.cnm.org.br/principal/publicacoes [in Portuguese].
7. Confederação Nacional dos Municípios, "Danos e prejuízos causados por excesso de chuvas na região nordeste entre dezembro de 2021 a 30 de maio de 2022" (2022); www.desastres.cnm.org.br/principal/publicacoes [in Portuguese].
8. Brazil, Ministério do Desenvolvimento Regional, Secretaria Nacional de Proteção e Defesa Civil, "Diagnóstico de capacidades e necessidades municipais em proteção e defesa civil" (2021) [in Portuguese].
9. Presidência da República (Brazil); Controladoria-Geral da União, "Portal da Transparência do Governo Federal, Ações de Proteção e Defesa Civil" (2022); https://portaltransparencia.gov.br/programas-e-acoes/acao/22BO-acoes-de-protecao-e-defesa-civil [in Portuguese].

10.1126/science.ade6396

# Mercury threatens human health in Brazil

Brazil exports more than 100 tons of gold annually, about 20% of which is produced illegally (*1*). Illegal gold mining in the Amazon region of Brazil has increased substantially in the past few years, as inspections and penalties for violators have diminished (*2, 3*). The gold mining process has led to mercury contamination in Brazil's waterways (*4, 5*). When consumed by humans, mercury damages the central nervous system and disrupts the metabolic function of organs (*6*). Brazil's government must step in to protect the country's population by enforcing mining laws and limiting mercury use.

Illegal gold is often extracted from the riverbeds in the Amazon region, especially the Madeira River, which has huge gold reserves (*4*). Miners use mercury to purify and separate the gold from the riverbed stones. Although Brazil has health and industrial standards that limit the amount of mercury that can be released into the environment (*1*), penalties for noncompliance are rare. As a result, large amounts of toxic mercury have polluted Brazil's water resources.

Once in the water, mercury causes severe toxicity in the food chain, starting with chromosomal abnormalities, changes in the proper functioning of enzymes, and other physiological problems in fish, leading to reduced fish populations (*7, 8*). The people most affected by mercury pollution are the Indigenous communities along the rivers of the Amazon region, who consume fish that have been contaminated (*9*). According to the Federal Police of Brazil, the rivers of the Indigenous Yanomami region of the Amazon contain 86 times the concentration of mercury considered safe for human consumption (*10*). A toxic concentration of mercury was found in 40% of children under 5 years old from areas affected by illegal gold mining, a tragedy with permanent developmental effects (*10*).

Brazil's government should rigorously monitor the volume of mercury imports. Providing a report reconciling all imported mercury with official use could minimize the likelihood of unauthorized use by illegal mining entities. The government should also forbid the sale of illegally obtained gold. Finally, police patrols should strictly monitor illegal mining sites, and miners in those locations should receive heavy fines. Limiting access to mercury and removing the incentives to illegally mine would effectively reduce illegal gold mining. The health of Brazil's population must be prioritized over profits.

**Masoud Negahdary, Berlane Gomes Santos, Lúcio Angnes***
Department of Fundamental Chemistry, Institute of Chemistry, University of São Paulo, São Paulo 05508-000, Brazil.
*Corresponding author. Email: luangnes@iq.usp.br

## REFERENCES AND NOTES

1. L. Prazeres, "Mining in the Amazon: What is behind the Madeira River invasion," BBC (2021); https://www.bbc.com/portuguese/brasil-59425015 [in Portuguese].
2. J. Siqueira-Gay, L. E. Sánchez, *Region. Environ. Change* **21**, 1 (2021).
3. C. Mestanza-Ramón *et al.*, *Land* **11**, 221 (2022).
4. I. A. Pestana, C. E. de Rezende, R. Almeida, L. D. de Lacerda, W. R. Bastos, *Extract. Indust. Societ.* **11**, 101122 (2022).
5. J. R. Gerson *et al.*, *Nat. Commun.* **13**, 1 (2022).
6. G. Guzzi, A. Ronchi, P. Pigatto, *Chemosphere* **263**, 127990 (2021).
7. C. J. Sergeant *et al.*, *Sci. Adv.* **8**, eabn0929 (2022).
8. M. M. Al-Sulaiti, L. Soubra, M. A. Al-Ghouti, *Curr. Pollut. Rep.* **8**, 249 (2022).
9. F. Ilha, "Explosion of illegal mining in the Amazon dumps 100 tons of mercury in the region," *El Pais* (2021); www.brasil.elpais.com/brasil/2021-07-20/explosao-do-garimpo-ilegal-na-amazonia-despeja-100-toneladas-de-mercurio-na-regiao.html [in Portuguese].
10. Y. Ramalho, V. Oliveira, M. Marques, L. Abreu, "Rivers in Yanomami Land have 8600% mercury contamination, reveals PF report" (2022); https://g1.globo.com/rr/roraima/noticia/2022/06/06/rios-na-terra-yanomami-tem-8600percent-de-contaminacao-por-mercurio-revela-laudo-da-pf.ghtml [in Portuguese].

10.1126/science.ade7525

# RESEARCH

## IN *SCIENCE* JOURNALS

*Edited by* **Michael Funk**

DRYLAND FORESTATION

### Just a little help

Forestation of global drylands has been suggested to be a way to decrease global warming, but how much promise does it actually have? Rohatyn *et al*. found that the climatic benefits are minor. Although drylands have considerable carbon sequestration potential, which could be used to lower the amount of carbon dioxide in the atmosphere and thereby slow warming, the reduction of albedo caused by forestation would counteract most of that effect. So, although forestation is clearly important, it cannot substitute for reducing emissions. —HJS

*Science*, abm9684, this issue, p. 1436

**Reforestation efforts in dry areas, such as this project in the High Atlas mountains of Morocco, may have limited climate change mitigation potential due to the decreased albedo of forests.**

PHOTO: GAERTNER/ALAMY STOCK PHOTO

PEROVSKITES

### Slowing nanoparticle growth

Inorganic materials with more covalent bonding, such as cadmium selenide, form uniform nanoparticles under fast growth conditions, but perovskites such as cesium lead bromide ($CsPbBr_3$) are more ionic and grow rapidly to form larger nanoparticles. Akkerman *et al*. controlled the nanoparticles' growth kinetics by using trioctylphosphine oxide, which solubilized the $PbBr_2$ precursor, bound to the cation-$[PbBr_3]$ monomer (solute), and weakly coordinated to the crystal nuclei surfaces. Nanoparticles with diameters from 3 to 13 nanometers were stabilized and isolated in high yield with lecithin, a long-chain zwitterion. Four well-resolved excitonic transitions with size-dependent confinement energies were seen for cesium as well as organic cations. —PDS

*Science*, abq3616, this issue p. 1406

DEVELOPMENTAL BIOLOGY

### Stem cell quality control

Blood stem cells that emerge in embryonic development migrate to specialized supportive niches for blood cell formation. Studying zebrafish embryos, Wattrus *et al*. discovered that macrophages contact stem cells upon their arrival to the niche and probe their cell surface to measure stress. Stem cells exhibiting high levels of a stress-activated protein are engulfed and killed by macrophages, whereas those showing low levels of stress have cellular material removed and selectively proliferate. By regulating death and division, macrophages quality control which individual stem cells establish lifelong blood production. —DJ

*Science*, abo4837, this issue p. 1413

ATMOSPHERE

### Up in the air

The eruption of the submarine volcano Hunga Tonga-Hunga Ha'apai in January of 2022 was so violent that its plume penetrated into the stratosphere. Vömel *et al*. studied in situ measurements by radiosondes (weather balloons), which showed that the event injected at least 50 teragrams of water vapor into the stratosphere. Because the volcano was underwater, the amount of water vapor in the developing stratospheric plume was high, and, unlike other large eruptions, it may have increased the amount of global stratospheric water vapor by more than 5%. —HJS

*Science*, abq2299, this issue p. 1444

NEUROSCIENCE

### Anatomy determines network membership

A hallmark of neuronal networks is the selective recruitment of

neurons into active ensembles, which form transiently stable patterns of activity. How are individual neurons selected to participate in these patterns of coactivity? Hodapp *et al*. examined the recruitment of pyramidal neurons into sharp-wave ripples, specific oscillations that support the consolidation of memories. Neurons with axons emerging from dendrites rather than from soma were more likely to be recruited into sharp-wave ripples. Membership in population ensembles are thus constrained or facilitated not only by excitatory input strength, but also by the axodendritic morphology of participating neurons. —PRS

*Science*, abj1861, this issue p. 1448

ARCHAEOLOGY

## Modeling early rice farming in Japan

The spread of domesticated plants in prehistory has often been modeled as a wave of advance across the landscape—one characterized by little variation in the rates and locations of dispersal. Using Bayesian statistical modeling of direct radiocarbon dates of rice remains at archaeological sites in Japan, Crema *et al*. found that instead of a uniform wave, there was substantial geographic and temporal variation in the appearance of rice farming. Local ecological suitability, regional differences in population density, and existing social networks may be responsible for the observed patterns. —MSA

*Sci. Adv.* 10.1126/sciadv.adc9171 (2022).

CELL BIOLOGY

## Purging the crypt in the small intestine

Crypts in the small intestine maintain sterility because of the continuous secretion of mucus and antimicrobial factors from the epithelium. Dolan *et al*. found that cholinergic stimulation elicited a distinctive form of secretion from goblet cells characterized by loss of mucus granule integrity, intracellular mucus swelling, and rupture of the plasma membrane of goblet cells (see the Focus by Weigert). Adjacent enterocytes coordinated the release of fluid with that of mucus into the crypt, leading to further expansion of the mucus and flushing of the crypt. —AMV

*Sci. Signal.* 15, eabl5848, add5975 (2022).

PLANT SCIENCE

## Transcription-activating nitrate sensor

Plants depend on nitrogen, responding with changes in growth and metabolism when nitrogen supplies change. Indeed, nitrogen fertilizer underlies a good deal of agricultural crop productivity. Studying a family of seven similar genes, Liu *et al*. identified the nitrate sensor in the small mustard plant *Arabidopsis thaliana*. The protein's nitrate-binding pocket resembles that found in bacteria nitrate sensors. Conformation change upon nitrate binding allows the protein to then function as a transcriptional activator, triggering the plant's responses to nitrogen availability. —PJH

*Science*, add1104, this issue p. 1419

DECOMPOSITION

## Heat-dependent wood decay

Decomposition rates vary with temperature and precipitation, in part because of the effects of climate on decomposer organisms. Although microbes are widely recognized as decomposers, animals such as insects also play a key role in tropical systems. Zanne *et al*. replicated an experiment at 133 global sites to quantify climate-related variation in wood decomposition by both microbes and termites. Climate influenced both microbial and termite decomposition, but termite presence and activity were more sensitive to temperature. Termites may thus play a larger role in global wood decomposition as the climate warms. —BEL

*Science*, abo3856, this issue p. 1440

# IN OTHER JOURNALS

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

CANCER

## Detecting cancer through a platelet lens

The early detection of cancer is highly desirable because early-stage cancers are usually easier to treat and more likely to be curable. As many as 18 different cancer types were detected in a single blood test using a method developed by In 't Veld *et al.*, which is based on analyzing tumor-derived RNA in platelets. The authors identified specific characteristics of platelet RNA in patients with cancer. Although the accuracy of this method varied by tumor type and was better for late-stage cancers, it did pick up on many early-stage tumors. For a subset of tumor types, it was also possible to classify site of origin. —YN

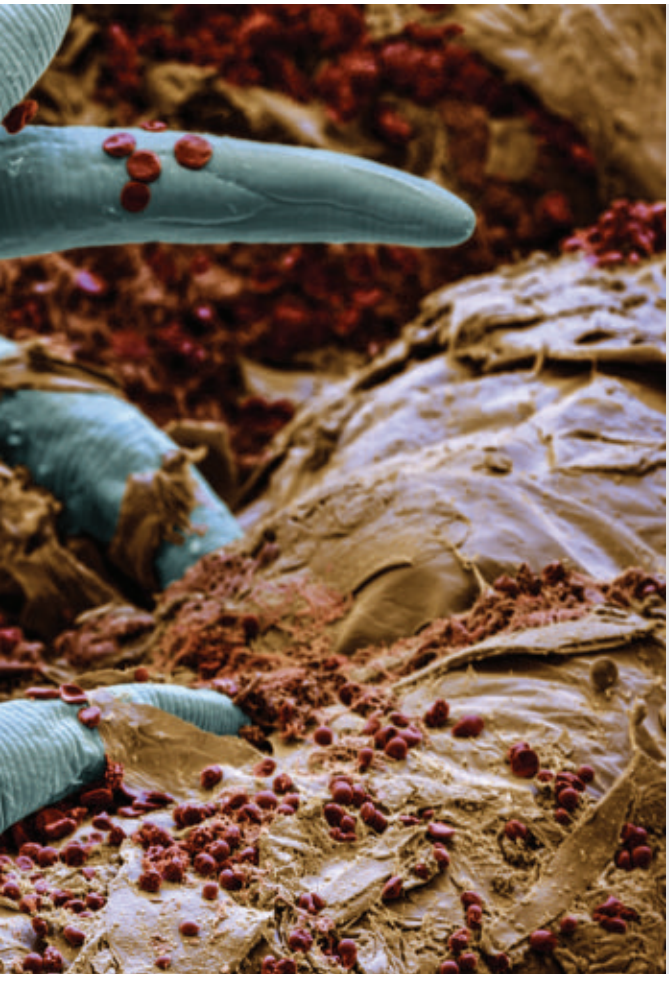
*Cancer Cell* **9**, 999 (2022).

EPITRANSCRIPTOME

## METTL muddle

Modifications in mRNA affect RNA splicing, translation, and stability and play critical roles in cellular processes such as cancer and differentiation. $N^6$-methyladenosine ($m^6A$) is the most abundant internal mRNA modification. A methyltransferase called METTL3 is the only enzyme identified during m6A formation. Paradoxically, 30 to 80% of $m^6A$ remains in cells depleted of METTL3 by CRISPR/Cas9, a finding that has motivated the search for additional methyltransferases that catalyze $m^6A$ formation. Poh *et al.* analyzed mouse embryonic stem cell lines with $m^6A$ in their mRNA after Mettl3 knockout and found that these cells expressed alternatively spliced *Mettl3* transcript isoforms that evade CRISPR/Cas9 mutations

PHOTO: EYE OF SCIENCE/SCIENCE SOURCE



IMMUNOLOGY

## ILCs are mite-ty good in the skin

Type 2 innate lymphoid cells (ILC2s) help to initiate, coordinate, and sustain allergic responses and antiparasite immunity partly through the secretion of type 2 cytokines such as interleukin-4 (IL-4), IL-5, and IL-13. Ricardo-Gonzalez *et al.* found that activated ILC2s in the skin of mice produce IL-13 that slows hair regrowth and reduces hair follicle stem cell proliferation. When ILC2s are absent, growth of commensal *Demodex* mites in hair follicles is unrestrained, resulting in inflammation, hair follicle exhaustion, skin barrier defects, and dermatitis. Topical anti-mite treatments, rather than antibacterial therapy, was able to reverse hair loss, pointing to a possible approach to treating similar types of dermatitis in humans. —STS *Immunity* 10.1016/j.immuni.2022.08.001 (2022).

***Demodex* mites, imaged here with scanning electron microscopy, are countered by type 2 innate lymphoid cells.**

PHOTO: TORRIPHOTO/GETTY IMAGES

and produce active methyltransferases. Therefore, although METTL3 is responsible for the vast majority of m6A, this work supplies a cautionary tale on how to interpret CRISPR/Cas9 knockout experiments. —DJ

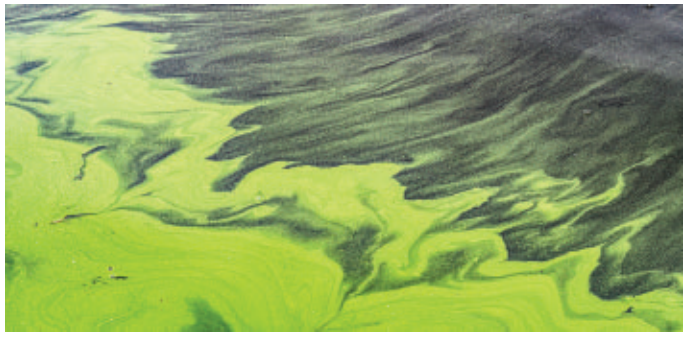
*PLoS Biol*. **20**, e3001683 (2022).

AQUATIC ECOLOGY

## Lake regime shifts

Environmental changes such as nutrient loading and warming are causing major changes in lakes, inducing profuse algal growth and "greening" (eutrophication). Gilarranz *et al*. identified possible regime shifts in large lakes using what is known as the trophic state index, a proxy for productivity derived from multispectral satellite imagery. Although fewer than one-third of lakes surveyed showed signs of a regime shift from 2002 to 2012, since then, the rates of regime shift seem to have increased. Lakes with less stable ecosystems are more likely to experience regime shifts, and these tend to be found in catchments within the tropics with high human populations having low average incomes. This study highlights inequalities in access to clean water and other ecosystem services. —BEL

*Proc. Natl. Acad. Sci. U.S.A.* **119**, e2116413119 (2022).



**Ecosystem-level regime shifts in lakes are leading to more algal blooms.**

WILDFIRES

## Cooling by fire

How does the smoke injected into the stratosphere by extreme wildfires affect climate? Liu *et al.* used a climate model to show that the smoke from two large wildfires, the Pacific Northwest Event of 2017 and the Australian New Year wildfire event of 2019–2020, cooled the global climate significantly. They also determined that the climate effects of wildfire smoke can cause nearly three times as much cooling per unit mass as sulfate aerosols ejected by volcanos. —HJS

*Geophys. Res. Lett.* 10.1029/2022GL100175 (2022).

BIOMATERIALS

## More efficient bone grafting

Although the body can repair and replace small bone defects, it struggles with larger ones caused by disease, trauma, or tumors. Grafting can be used in some cases, but that technique is limited by the supply of source material, which has led to the development of ceramics, polymers, and composites to enhance the process. Kageyama *et al.* encapsulated human mesenchymal stem cells in collagen microgels that spontaneously contract due to attractive forces between the cells. This led to an enrichment of collagen molecules and cell density, creating a more favorable environment for osteogenic differentiation compared with conventional spheroid culture and subsequent in vivo bone regeneration in cranial defects in mice. —MSL

*Acta Biomater.* 10.1016/j.actbio.2022.08.044 (2022).

NUCLEAR PHYSICS

## Searching for a rare event

Physicists have been on a quest to observe a hypothetical nuclear transition known as the neutrinoless double-beta decay. The observation would shed light on the nature of the neutrino, as well as matter-antimatter asymmetry in the universe. If it does exist, the decay is expected to be exceedingly rare, and experiments have been placing ever increasing bounds on its half-life for several nuclei. Azzolini *et al.* report improved half-life limits on this decay for the $^{82}$Se nucleus, called CUPID-0, which in their experiment was embedded in the detector. Increasing the detector mass is expected to lead to further improvements in half-life limits. —JS

*Phys. Rev. Lett.* **129**, 111801 (2022).

## ALSO IN *SCIENCE* JOURNALS

*Edited by* **Michael Funk**

SYNTHETIC BIOLOGY

### Cleaning polluted soil with plants

Anthropogenic activities have caused the contamination of soils with metals, metalloids, and various organic pollutants. The diluted nature of these contaminants means that they are challenging and not economically viable to remove from the soil with normal mining practices, but plants may offer a solution. In a Perspective, Rylott and Bruce discuss the use of plants to remediate soils and extract useful metals. Understanding the detoxification pathways, especially those in plants that naturally grow on metalliferous soils, could allow the engineering of fast-growing plants to remediate polluted soils and mine technology-critical metals. Many challenges remain, including developing approaches to grow plants in soils with complex contaminants and to degrade inorganic pollutants. —GKA

*Science*, abn6337, this issue p. 1380

DEVELOPMENT

### Forcing a twist in the gut

The transcription factor Pitx2 helps to establish asymmetry of the left and right sides of the tissues that give rise to the vertebrate gut. Sanketi *et al*. used studies of chick and mouse embryos to show that Pitx2 provides another key regulatory role later in development to generate the characteristic shape of intestines. When cells on the right side expand, mechanical stress extends to the left side, causing the morphogen transforming growth factor-β to be released. This process enhances *Pitx2* expression and consequent stiffening of the left side, which helps to control the overall degree of tilting of the developing intestines. —LBR

*Science*, abl3921, this issue p. 1396

QUANTUM PHYSICS

### Learning many-body behavior

Predicting the properties of strongly interacting many-body quantum systems is notoriously difficult. One approach is to use quantum computers, but at the current stage of the technology, the most interesting problems are still out of reach. Huang *et al*. explored a different technique: using classical machine learning to learn from experimental data and then applying that knowledge to predict physical properties or classify phases of matter for specific types of many-body problems. The authors show that under certain conditions, the algorithm is computationally efficient. —JS

*Science*, abk3333, this issue p. 1397

NEURODEVELOPMENT

### Preserving with early intervention

Symptoms of Huntington's disease only become evident during adulthood even though the underlying gene mutation has been present throughout development. Braz *et al*. show that during the first neonatal week, brains of mice carrying the disease-driving mutation show disrupted neural circuit physiology (see the Perspective by Blumenstock and Dudanova). By the second week, their brains had normalized these physiological disruptions, but the mice went on to develop disease symptoms in adulthood. Conversely, early pharmacologic enhancement of glutamatergic transmission rescued the disruptions and forestalled disease symptoms in these mice. —PJH

*Science*, abq5011, this issue p. 1396; see also ade3116, p. 1383

HEART FAILURE

### A liver-to-heart chat

It is thought that the liver and heart share physiological communication, helping to explain why, for example, nonalcoholic fatty liver disease increases the risk of heart failure. By examining a large collection of mouse strains, Cao *et al*. identified coagulation factor XI as a mediator of such liver-heart cross-talk (see the Perspective by Tong and Hill). In mouse models of diet-induced heart failure with preserved ejection fraction, factor XI was inversely correlated with the extent of diastolic dysfunction, with greater expression of factor XI reducing cardiac fibrosis and inflammation. Factor XI itself was only expressed in the liver, but it had reproducible effects on the heart, which the authors connected to the activity of the SMAD pathway. —YN

*Science*, abn0910, this issue p. 1399; see also ade3528, p. 1382

SOLAR CELLS

### Pure perovskite topcoats

Two-dimensional (2D) halide perovskite passivation layers grown on three-dimensional (3D) perovskite can boost the power conversion efficiency (PCE) of solar cells, but spin-coating of these layers usually forms heterogeneous 2D phases or only ultrathin layers. Sidhik *et al*. found that solvents with the appropriate dielectric constant and donor strength could grow phase-pure 2D phases of controlled thickness and composition on 3D substrates without dissolving them. Solar cells maintained a peak PCE of 24.5% for 2000 hours with less than 1% degradation under continuous light at 55°C and 65% relative humidity. —PDS

*Science*, abq7652, this issue p. 1425

BIODIVERSITY

### Declining genetic diversity

Habitat loss is one of the major drivers of species extinctions and declines of species richness at local scales. Smaller areas of remnant habitat also harbor smaller populations and lower genetic diversity, which may limit potential adaptation to environmental change. Exposito-Alonso *et al*. developed a framework to predict decreases in naturally occurring mutations, and thus genetic diversity, with habitat loss (see the Perspective by Ruegg and Turbek). Georeferenced genomic data from across the native ranges of the small mustard plant *Arabidopsis thaliana* and 20 other species suggest that the mutation-area relationship follows a power law. This relationship predicts that many species have already experienced substantial genetic diversity loss. —BEL

*Science*, abn5642, this issue p. 1431; see also add0007, p. 1384

MAMMALIAN PHYSIOLOGY

### Protected from pressure

Marine mammals are highly adapted to living underwater. One of the most challenging aspects of this environment is the extreme pressure that animals experience at increasing depth. This condition increases the need for protection of the brain from pulsatile blood flow, something experienced by all mammals. Furthermore, the movement of marine mammal flukes during dives exerts even greater pulsatility. Lillie *et al*. modeled the extensive array of blood vessels, or retia mirabilia, found in cetacean brains across 11 species and concluded that this array minimizes blood pressure differentials, thus protecting the brain without reducing the pressure pulses and facilitating fluking locomotion (see the Perspective by Williams). —SNV

*Science*, abn3315, this issue p. 1452; see also ade3117, p. 1378

ELECTRONIC MATERIALS

## Molecules are ideal dopants for 2D

Two-dimensional (2D) materials show great promise as semiconductor materials, yet the conventional strategy for controlling their doping level by substituting one element for another inherently introduces defects that impede charge transport. Jang *et al.* report that the doping level of atomically thin molybdenum sulfide ($MoS_2$) materials can be effectively modulated by capping with a thin insulating boron nitride layer and depositing molecules on top that transfer electrons into the semiconductor. In this "remote" doping method, the spatial separation of the molecular dopants from the $MoS_2$ semiconductor leads to superior transistor properties by minimizing any deleterious interactions between the dopant and the mobile charges, which may enable their use in high-performance 2D electronic devices. —JEG

*Sci. Adv.* 10.1126/sciadv.abn3181 (2022).

IMMUNODEFICIENCY

## NLRP1 knockout to the rescue

The NLRP1 inflammasome is a protein complex crucial for various immune responses, and its dysregulation can lead to severe immune disorders. Harapas *et al.* studied four children with dipeptidyl peptidase 9 (DPP9) loss-of-function mutations and found that these patients had various immune disorders and spontaneous activation of the NLRP1 inflammasome in their isolated keratinocytes. Using a mouse model with a catalytically inactive DPP9 mutation, the authors found that neonate lethality could be rescued by also knocking out NLRP1. In a zebrafish model, knockout of a molecule involved in NLRP1 inflammasome formation also rescued the survival of zebrafish with a DPP9 deletion. Thus, DPP9 mutations seem to result in immune disorders caused by NLRP1 inflammasome overactivation. —DAE

*Sci. Immunol.* **7**, eabi4611 (2022).

CANCER

## Engaging ESR1 mutant tumors

Breast cancers often express estrogen receptor α (ERα), and targeted therapies to take advantage of this feature can often lead to acquired resistance, requiring the development of next-generation ERα-targeted therapies such as giredestrant. Liang *et al*. leveraged the mouse mammary gland to examine altered ERα signaling caused by mutations in estrogen receptor-1 (ESR1) and evaluated the effectiveness of giredestrant against ESR1 mutant patient–derived xenografts. They found that mutant ERα and progesterone both drove tumorigenicity, but tumors with these alterations remained sensitive to giredestrant, representing a promising therapy to overcome acquired resistance. —DLH

*Sci. Transl. Med.* **14**, eabo5959 (2022).

RESEARCH ARTICLE SUMMARY

DEVELOPMENT

# *Pitx2* patterns an accelerator-brake mechanical feedback through latent TGFβ to rotate the gut

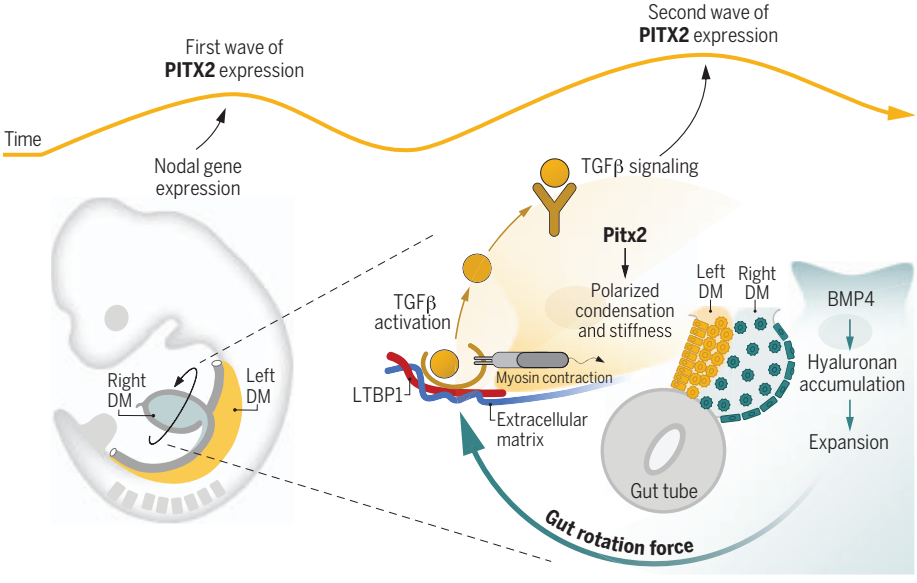
**Bhargav D. Sanketi, Noam Zuela-Sopilniak, Elizabeth Bundschuh, Sharada Gopal, Shing Hu, Joseph Long, Jan Lammerding, Sevan Hopyan, Natasza A. Kurpios***

**INTRODUCTION:** Nearly all vertebrate animals appear symmetrical on the outside, but internal organs such as the heart, liver, and stomach are carefully arranged in a left-right asymmetric pattern. This packs the organs into the limited space of the body while preserving their function. An important model to study left-right organ asymmetry is the process of gut rotation, during which the intestine achieves its familiar twisted form. Gut rotation is predictable in healthy embryos: always counterclockwise and timed perfectly. Gut rotation hinges on the neighboring dorsal mesentery (DM), a mesodermal tissue bridge anchoring the gut tube through which intestinal blood and lymphatic vessels traverse. The left and right sides of the DM are physically continuous but exhibit discrete and asymmetric tissue changes, causing the mesentery to deform and tilt the attached gut tube leftward. This leftward tilt provides a critical bias to initiate asymmetric gut rotation that is conserved throughout evolution. Gut rotation is orchestrated by the transcription factor *Pitx2*, which is expressed on the left side of embryos. *Pitx2*-driven asymmetry is also evolutionarily conserved, and altered *Pitx2* activity disrupts the growth of lateralized organs.

**RATIONALE:** In early embryos, *Pitx2* expression is induced during gastrulation by the highly conserved morphogen *Nodal*, a member of the transforming growth factor–β (TGFβ) superfamily. This asymmetric *Nodal* expression is transient and stops before asymmetric organ morphogenesis. However, *Pitx2* expression is present on the left side in most asymmetric organs, including the DM, where it orchestrates gut rotation. This has left unresolved the question of how *Pitx2* directs organ development when Nodal is gone. We reasoned that *Pix2* expression during gut rotation must be locally regulated by undescribed mechanisms that are required to correctly shape asymmetric organs.

**RESULTS:** We found that deletion of Nodal from the mouse lateral plate mesoderm, a precursor of the DM, preserved normal gut laterality and *Pitx2* expression, supporting the presence of additional mechanisms regulating *Pitx2*. Indeed, *Pitx2* expression in the left DM was not continuous and required a second wave of induction during gut rotation. This second wave of *Pitx2* expression was tuned by a local positive feedback loop through the latent TGFβ mechanosensor, a new player here that links *Pitx2* expression to the mechanical forces driving rotation. TGFβ-Pitx2 activity repressed bone morphogenetic protein 4 (*Bmp4*) expression in the left mesentery, marking the first molecular asymmetry of gut rotation. BMP4 signaling persisted to drive the right-side program, which acted through the extracellular matrix (ECM) component hyaluronan to expand and deform the right side of the mesentery, initiating gut rotation. Tilting forces from this right-sided expansion were then sensed and mechanically transduced into TGFβ-dependent *Pitx2* expression changes in the neighboring left DM, resulting in polarized mesenchymal condensation and increased tissue stiffness. These findings can be explained by an accelerator on the right (BMP4) and a brake on the left (TGFβ-*Pitx2*), which cooperate through mechanical feedback to tune the conserved counterclockwise gut rotation. Thus, vertebrate gut rotation combines biochemical and biomechanical inputs to break gut symmetry and direct the evolutionarily conserved intestinal rotation.

**CONCLUSION:** We uncovered a mechanosensitive TGFβ feedback loop driving the *Pitx2*-governed left-sided transcriptional program that reproducibly rotates the vertebrate intestine. These findings address a long-standing question of how asymmetric organs interpret the embryonic body plan to execute local programs of laterality. Uncovering the basic mechanisms for how organs form could aid efforts to diagnose and prevent birth defects, including intestinal malrotation and volvulus, which may cause large stretches of intestinal tissue to die, a surgical emergency in neonates. Though focused on the intestine, our studies are relevant to other asymmetric organs where laterality errors also cause lethal birth defects in human babies. ■




**Gut rotation requires a second wave of *Pitx2* expression tuned by the latent TGFβ mechanosensor.** The second *Pitx2* wave represses BMP4 on the left, but BMP4 persists on the right side to drive tissue expansion, initiating gut rotation. Tilting forces from expansion (accelerator) deform the ECM-resident Ltbp1, liberating TGFβ from the latent complex to initiate its downstream signaling. TGFβ-dependent *Pitx2* expression then causes polarized condensation and increased tissue stiffness (brake), a mechanical feedback loop with the right side.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: natasza.kurpios@cornell.edu
Cite this article as B. D. Sanketi *et al.*, *Science* **377**, eabl3921 (2022). DOI: 10.1126/science.abl3921

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abl3921

RESEARCH ARTICLE

DEVELOPMENT

# *Pitx2* patterns an accelerator-brake mechanical feedback through latent TGFβ to rotate the gut

Bhargav D. Sanketi[1], Noam Zuela-Sopilniak[2], Elizabeth Bundschuh[1], Sharada Gopal[1], Shing Hu[1], Joseph Long[2], Jan Lammerding[2], Sevan Hopyan[3,4], Natasza A. Kurpios[1]*

The vertebrate intestine forms by asymmetric gut rotation and elongation, and errors cause lethal obstructions in human infants. Rotation begins with tissue deformation of the dorsal mesentery, which is dependent on left-sided expression of the Paired-like transcription factor *Pitx2*. The conserved morphogen *Nodal* induces asymmetric *Pitx2* to govern embryonic laterality, but organ-level regulation of *Pitx2* during gut asymmetry remains unknown. We found *Nodal* to be dispensable for *Pitx2* expression during mesentery deformation. Intestinal rotation instead required a mechanosensitive latent transforming growth factor–β (TGFβ), tuning a second wave of *Pitx2* that induced reciprocal tissue stiffness in the left mesentery as mechanical feedback with the right side. This signaling regulator, an accelerator (right) and brake (left), combines biochemical and biomechanical inputs to break gut morphological symmetry and direct intestinal rotation.

Evolutionarily conserved left-right (LR) asymmetries of vertebrate internal organs require a coordinated set of sharply defined gene expression events that drive changes in extracellular matrix (ECM) biomechanics and cell behavior. For example, the vertebrate embryonic midgut, which gives rise to most of the large and small intestine (Fig. 1A), undergoes a stereotypical counterclockwise rotation that is necessary for subsequent gut looping (*1*, *2*). Failure to initiate midgut rotation leads to midgut volvulus, a catastrophic blockage of the gut tube and mesenteric blood vessels (*3*, *4*). In birds and mice, the direction of midgut rotation is driven by asymmetric cellular behaviors within the dorsal mesentery (DM), the adjacent connective organ that suspends the gut tube from the dorsal body wall (Fig. 1B, colored regions), and through which intestinal blood and lymphatic vessels traverse (*1*, *2*, *5*).

The DM forms during embryogenesis by apposition and fusion of the left and right flat sheets of splanchnic mesoderm (derived from lateral plate mesoderm). This starts on day 3 in the chicken, which corresponds to Hamburger-Hamilton stage 17 (HH17) (*6*), or on embryonic day 9.5 (E9.5) in mice (*1*). Initially, the mesenchyme of the left and right splanchnic mesoderm appears homogeneous (*1*), but with DM formation, it develops distinct LR asymmetry where the left side condenses and the right side expands (Fig. 1B, bottom panel) (*1*, *7*, *8*). The resulting deformation of the DM tilts the attached gut tube leftward, providing a critical bias for the initiation of counterclockwise gut rotation (Fig. 1B, HH21) (*1*, *5*).

Gut rotation is directed by the transcription factor *Pitx2* (*1*, *9*–*11*), which is responsible for the transfer of LR patterning information from early gastrulation to downstream organ morphogenesis (*1*, *5*, *8*, *11*). *Pitx2* is expressed on the left side of the DM (Fig. 1C), where it activates pathways that regulate actin cytoskeleton organization, cell adhesion, and ECM compaction to induce polarized condensation of the left DM (*5*, *8*), suggesting key roles in DM morphogenesis. However, cell mechanistic studies downstream of *Pitx2* have been hampered because loss of *Pitx2* on the left causes a double right-side identity (double-right isomerism), which causes a cell fate change preventing analysis of *Pitx2*-directed morphogenesis (*10*).

In early embryos, *Pitx2* expression is induced in the left lateral plate mesoderm by the highly conserved morphogen *Nodal*, a member of the transforming growth factor–β (TGFβ) superfamily (*10*, *12*). This asymmetric *Nodal* expression is transient and stops before asymmetric organ morphogenesis (*13*, *14*). However, *Pitx2* expression is present on the left side in all derivatives of the lateral plate, including the DM (Fig. 1C), and long after *Nodal* expression has ceased, leaving unresolved the question of how *Pitx2* regulates asymmetric organogenesis when *Nodal* is gone.

We found that *Pitx2* expression in the left DM required a second signaling regulator, a mechanosensitive latent transforming growth factor β (TGFβ), to drive gut rotation. This second wave of *Pitx2* expression repressed the initially bilateral bone morphogenetic protein 4 (*Bmp4*) expression, marking the first molecular asymmetry in the DM. Persistent BMP4 signaling within the right DM promoted mesenchymal expansion on the right. This expansion was sensed and mechanically transduced into TGFβ-dependent *Pitx2* expression changes in the left DM, resulting in polarized mesenchymal condensation and increased tissue stiffness. We further showed that the BMP antagonist Noggin simultaneously inhibits right-derived BMP4 and left-derived TGFβ-*Pitx2* activity to spatiotemporally restrict tissue deformation of the DM. This finding allowed us to uncouple the mechanosensitive role of TGFβ-*Pitx2* in asymmetric gut morphogenesis from *Pitx2*-driven laterality, a phenotype that is masked by the double right-side identity in *Pitx2*-null mice.

We propose that TGFβ-dependent control of *Pitx2* dosage on the left provides transcriptionally patterned resistance to expansion from the right. These findings can be explained by an accelerator on the right (BMP4) and a brake on the left (TGFβ-*Pitx2*), which cooperate through mechanical feedback to drive the conserved counterclockwise gut rotation.

## Repression of symmetrical *Bmp4* by *Pitx2* initiates DM asymmetry

Gut rotation is first initiated by an expansion of the ECM in the right DM (Fig. 1B, bottom panel, teal) (*7*). This expansion and subsequent gut tilting depends on matrices of hyaluronan (HA) modified by the enzyme *Tsg6* (tumor necrosis factor-alpha–stimulated gene 6) (*7*), but the upstream transcriptional control of this process is unknown. Our prior transcriptomics studies performed in the chicken DM at the time of tilting (HH21) (*8*) revealed enrichment of *Bmp4* on the right side of the chicken DM. This raised the possibility that *Tsg6* expression and the resulting ECM expansion may be regulated by BMPs, members of the TGFβ superfamily. RNA in situ hybridization in the chicken revealed that *Bmp4* expression is initially bilateral in the DM but decreases on the left as tilting begins (Fig. 1D). We reasoned that this pattern might be governed by *Pitx2*, because loss of left-sided *Bmp4* expression coincided in space and time with the onset of *Pitx2* expression in the left DM (Fig. 1C). This was confirmed by right-sided electroporation (fig. S1B) of *Pitx2* before chicken DM formation, which resulted in a reduction of both *Bmp4* expression (Fig. 2A) and HA accumulation on the right (Fig. 2, B and C). This double-left isomerism caused a loss of gut tilting (Fig. 2G, net-zero tilting angle from the midline measured at HH21). Co-electroporating *Pitx2* and *Bmp4* on the right rescued the right-sided program, as determined by restored HA accumulation (Fig. 2, B and C) and normal gut tilting (Fig. 2G). Thus, *Pitx2* contributes to *Bmp4* transcriptional suppression.

[1]Department of Molecular Medicine, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA. [2]Weill Institute for Cell and Molecular Biology and Department of Biomedical Engineering, Cornell University, Ithaca, NY 14850, USA. [3]Program in Developmental and Stem Cell Biology, Research Institute, The Hospital for Sick Children, Toronto, Ontario M5G 0A4, Canada. [4]Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada.
*Corresponding author. Email: natasza.kurpios@cornell.edu

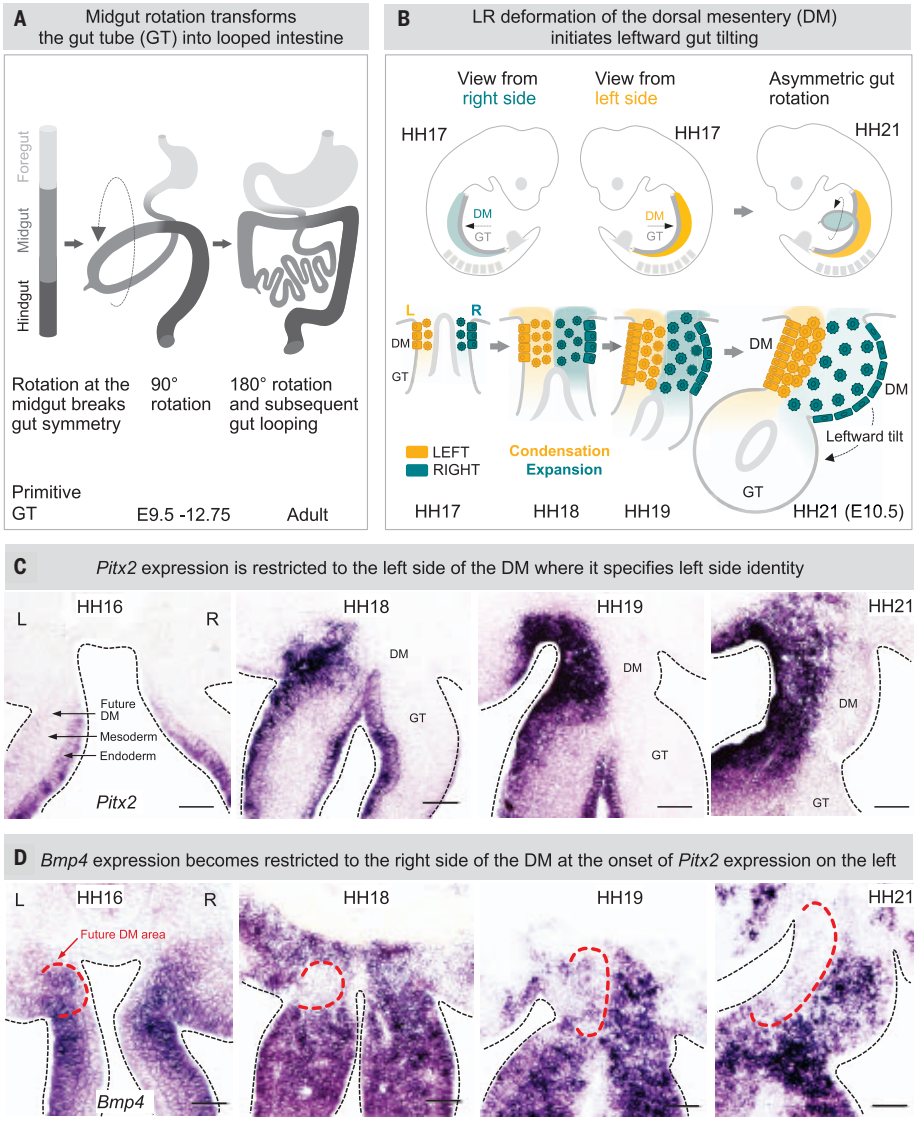



**Fig. 1. BMP4 becomes restricted to the right at the onset of *Pitx2* expression on the left.** (**A**) Midgut rotation transforms the gut tube into looped intestine. (**B**) Tissue changes across the LR DM initiate gut tube (GT) rotation. (**C** and **D**) RNA in situ hybridization (ISH) for *Bmp4* and *Pitx2* during DM formation and GT rotation. Scale bars, 50 μm.

To more clearly define the role of right-sided BMP signaling, we electroporated the BMP antagonist *Noggin* (*15*–*17*) into the right DM. This resulted in a loss of HA accumulation (Fig. 2, C and D), *Tsg6* expression (Fig. 2D), and gut tilting (Fig. 2G). Because *Noggin* is a broad antagonist of several members of the TGFβ superfamily (*15*, *18*), we also electroporated a dominant-negative BMP receptor 1 (*BmpR1-DN*, truncated kinase domain) (*17*) into the right DM. This produced a similar loss of HA (Fig. 2, C and D), *Tsg6* (Fig. 2D), and gut tilting (Fig. 2G). Perturbation of BMP signaling did not produce ectopic *Pitx2* expression in the right DM (Fig. 2E), indicating that BMP signaling independently drives the right-sided ECM expansion and not does repress the left-lateralizing *Pitx2* program.

To assess the role of *Bmp4* in the mouse embryonic midgut, *Bmp4* was conditionally deleted (fig. S2, A and B) using an inducible *Hoxb6* driver that is specifically active in the posterior lateral plate mesoderm (*19*). This produced a loss of HA accumulation (fig. S2C) and led to gut rotation defects evident at E12.75 (fig. S2D) without ectopic *Pitx2* expression on the right (Fig. 2H). Thus, BMP signaling independently induces *Tsg6* expression and ECM expansion of the right DM.

In mice lacking *Pitx2* (*Pitx2*$^{hd/hd}$), the left DM mesenchyme becomes expanded like the right side and gut tilting is arrested (*1*). To test whether *Pitx2* expression in the left DM is necessary to repress drivers of right-side identity, we examined *Bmp4* expression and HA accumulation in *Pitx2*$^{hd/hd}$ embryos at E10.5 (Fig. 2I). Both HA and *Bmp4* were present bilaterally (a double-right isomerism), indicating that *Pitx2* expression on the left represses gene expression and cell behavior normally associated with the right side of the DM. This contrasts with the loss of *Bmp4* on the right, which does not lead to bilateral *Pitx2* expression (Fig. 2H). In summary, *Pitx2* is both necessary and sufficient to inhibit BMP signaling in the DM and to specify the left-side identity of the DM.

### *Noggin* restricts asymmetric tissue deformation to the DM

The mesenchymal cells within the DM and those within the gut tube arise from adjacent precursor populations in the left and right splanchnic mesoderm (*17*, *20*) (Fig. 3A). LR asymmetric cell behaviors that specifically deform the DM are limited only to the DM cells and never extend ventrally into the gut tube (Fig. 3A). For example, *Smad6*, a canonical regulatory feedback gene downstream of *Bmp4* expression (*21*), was found strictly within the right DM and only at the onset of tilting, but never in the gut tube despite the persistence of bilateral *Bmp4* mRNA expression in the gut (fig. S1A). We reasoned that BMP signaling must be actively repressed at the protein level in the gut tube. Consistent with this idea, we initially found *Noggin* expression throughout the early left and right splanchnic mesoderm (fig. S3A; HH15), but at the onset of DM formation, *Noggin* expression became progressively restricted to the cells of the gut tube in both chicken and mouse embryos (fig. S3, A and B). Furthermore, right-sided electroporation of constitutively active *BmpR1* (Q233D, *BmpR1-CA*) (*22*) into the gut splanchnic mesoderm bypassed Noggin activity and induced DM-like accumulation of HA in the gut tube (fig. S4). These data suggest that gut tube-derived Noggin limits BMP signaling–driven ECM expansion to the right DM.

The displacement of *Noggin* expression away from the forming DM was also commensurate with the onset of *Pitx2* expression on the left, suggesting a previously undescribed, antagonistic relationship between *Noggin* and *Pitx2* (fig. S3A). Indeed, electroporation of *Noggin* in the left DM caused a loss of *Pitx2* expression (Fig. 4, A and B). Moreover, mouse embryos lacking *Noggin* displayed an abnormally extended domain of *Pitx2* expression into the gut tube (fig. S3C), premature gut tilting at E9.5 (fig. S3D), and aberrant gut rotation patterns at E12.75 (fig. S3, E and F). These results indicate that gut tube–derived *Noggin* limits LR gene expression and cell behavior to the DM, shedding light on the local mechanisms that pattern molecular and morphological boundaries between the DM and the adjoining gut tube that allow timely asymmetric deformation.

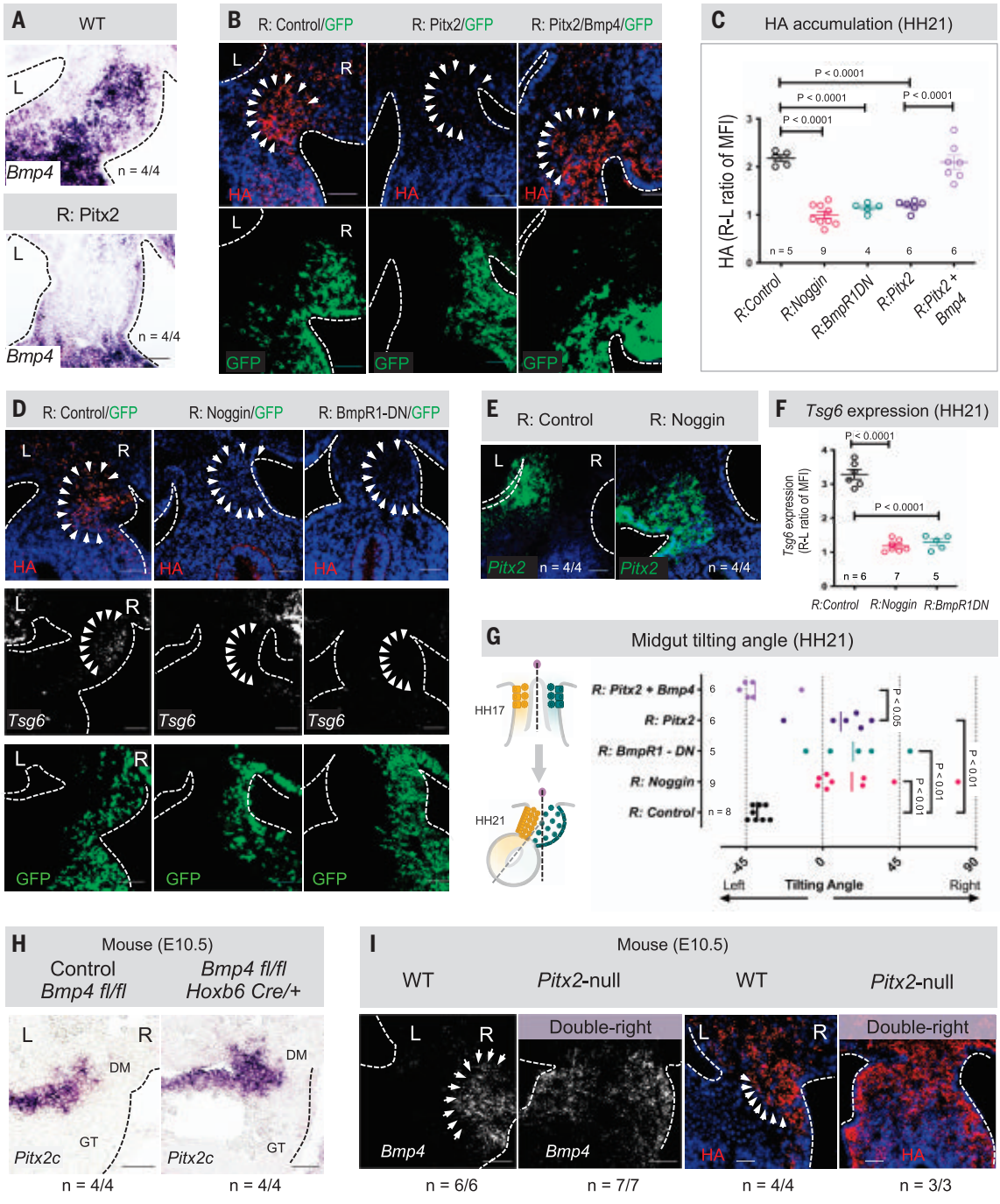



**Fig. 2. BMP4 drives morphological DM asymmetry through the *Tsg6*/HA pathway.** Right-sided *Pitx2* electroporation in the chicken DM inhibits *Bmp4* expression [(**A**), ISH], HA accumulation [(**B**) in red, and quantified in (C)], and gut tilting [quantified in (G)], whereas right-sided Pitx2 with Bmp4 co-electroporation restored the right-sided program (HH21). Effect of right-sided *Noggin* or *BmpR1-DN* electroporation (HH21) on HA [(**D**) in red, and quantified in (C)] and *Tsg6* [RNAScope, (D) in white and quantified in (F)] and *Pitx2* [RNAScope, (**E**) in green]. Data were analyzed with one-way ANOVA and multiple comparisons with Tukey's correction for HA (**C**) and *Tsg6* (**F**) (presented as mean ± SEM), and Watson's test for tilting angles (**G**) (presented as circular mean) upon specified electroporations marked by GFP. (**H** and **I**) Mouse *Pitx2c* [(H), ISH] was not affected upon *Bmp4* deletion, whereas *Pitx2* loss (I) caused a "double-right" phenotype, including ectopic HA and *Bmp4* (RNAScope). MFI, mean fluorescence intensity. All scale bars, 50 μm.

### Gut rotation requires a second wave of *Pitx2* expression independent of mesodermal *Nodal*

We examined *Pitx2* expression during the pre- and early DM formation periods (Fig. 3A). Early *Pitx2* expression was detected in the left splanchnic mesoderm at E8.0 (mouse) (*23*) and HH12 (chicken) (*10*, *23*, *24*) (Fig. 3, B and D), but was unexpectedly absent in this region shortly afterward (E9.0 mouse, Fig. 3C; HH15 chicken, Fig. 3, B and D, and fig. S3A). *Pitx2* expression was detected again as the DM began to form, both in the mouse (E9.5, Fig. 3C) and chicken (HH17, Fig. 3D and fig. S3A).

We next analyzed conditional *Nodal* mouse mutants generated with a Hoxb1-Cre driver, which ablates all *Nodal* transcription and subsequent *Pitx2* expression in the lateral plate

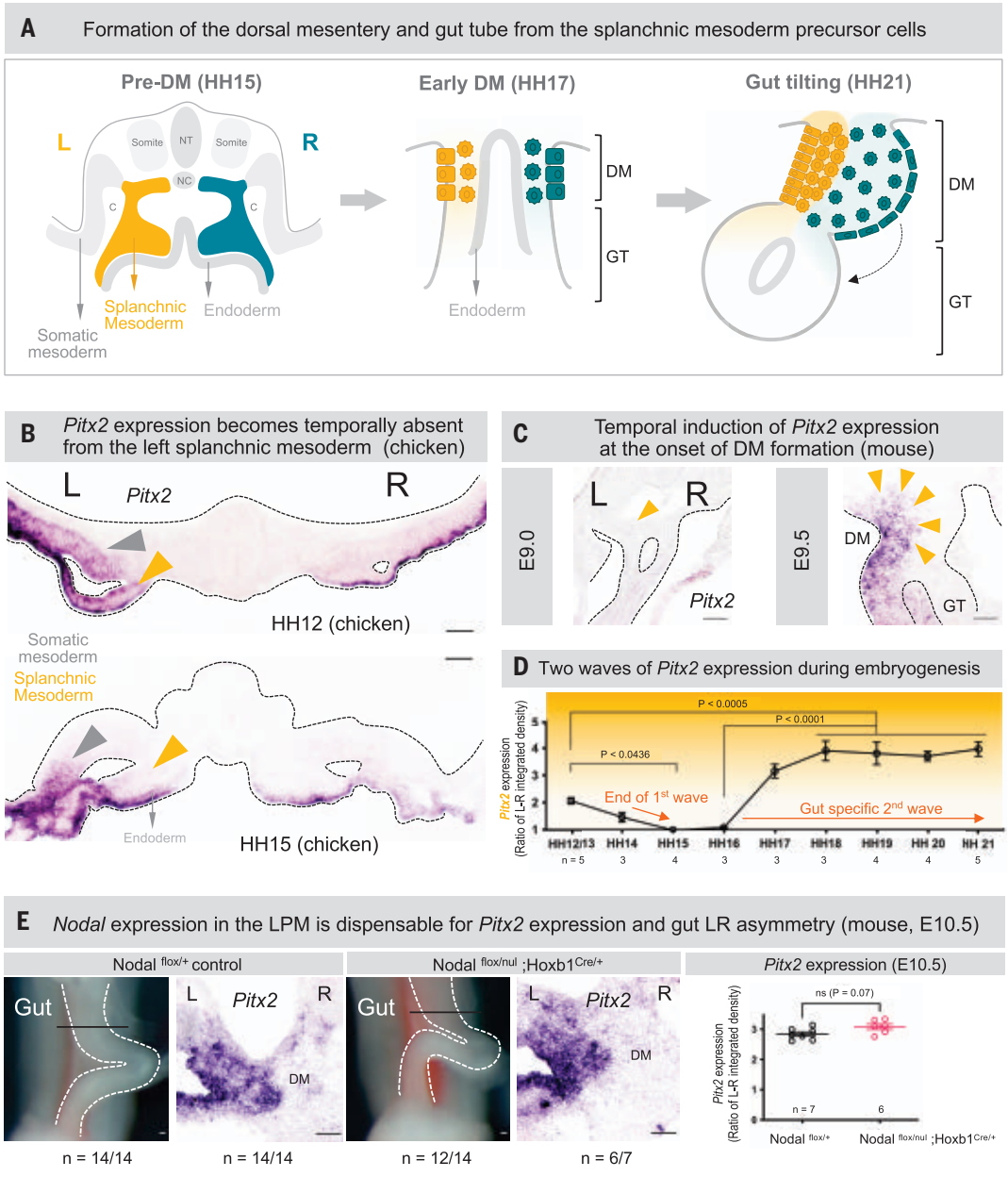



**Fig. 3. Gut rotation requires a second wave of *Pitx2* independently of mesodermal *Nodal*.** (**A**) DM forms by fusion of the left and right splanchnic mesoderm; fusion of the somatic mesoderm forms the body wall. Subsequent LR tissue changes are specific to the DM and never extend ventrally into the GT. NT, neural tube; NC, notochord, C, coelom. (**B**) *Pitx2* (ISH) is expressed in the splanchnic mesoderm at HH12 but is absent at HH15 before DM formation. (**C**) In mice, *Pitx2c* is lost by E9.0 but reemerges with DM formation at E9.5. (**D**) Two waves of *Pitx2* expression quantified by integrated density normalized to somatic mesoderm, using one-way ANOVA followed by multiple comparisons with Tukey's correction (presented as mean ± SEM). (**E**) Mouse gut tilting and *Pitx2c* expression (compared by unpaired *t* test, presented as mean ± SEM) are not perturbed by conditional *Nodal* deletion (E10.5). Scale bars, 50 μm.

mesoderm (*25*), causing heart and lung laterality defects (*26*). Neither gut tilting (Fig. 3E; n = 12/14) nor *Pitx2* expression in the DM (Fig. 3E; *n* = 6/7) was affected by the deletion of mesodermal *Nodal*. Thus, whereas *Nodal* drives the first wave of asymmetric *Pitx2* to govern embryonic laterality, subsequent organ-level regulation of *Pitx2* expression (second wave) during the establishment of gut-specific asymmetry may be regulated by an additional pathway.

### Latent TGFβ directs gut-specific asymmetry through *Pitx2* expression

We first tested whether the second wave of *Pitx2* expression is BMP dependent (*27*–*29*), which would be consistent with the loss of *Pitx2* expression observed upon *Noggin* electroporation on the left (Fig. 4, A and B). However, electroporating *BmpR1-DN* on the left did not alter *Pitx2* expression or HA accumulation (Fig. 4, A and B), nor did it affect gut tilting (Fig. 4C), indicative of a BMP-independent mechanism to modulate *Pitx2* expression during gut rotation.

Noggin can antagonize several non-BMP TGFβ ligands (*15*, *18*), and our DM transcriptomics data revealed that the latent TGFβ binding protein 1 (*Ltbp1*), a major regulator of TGFβ pathway activation (*30*, *31*), is expressed strictly within the left chicken and mouse DM, akin to the spatiotemporal kinetics of *Pitx2* expression (Fig. 4D). Similarly, TGFβ-induced (*TGFβ-i*), an effector downstream of TGFβ signaling (*32*),

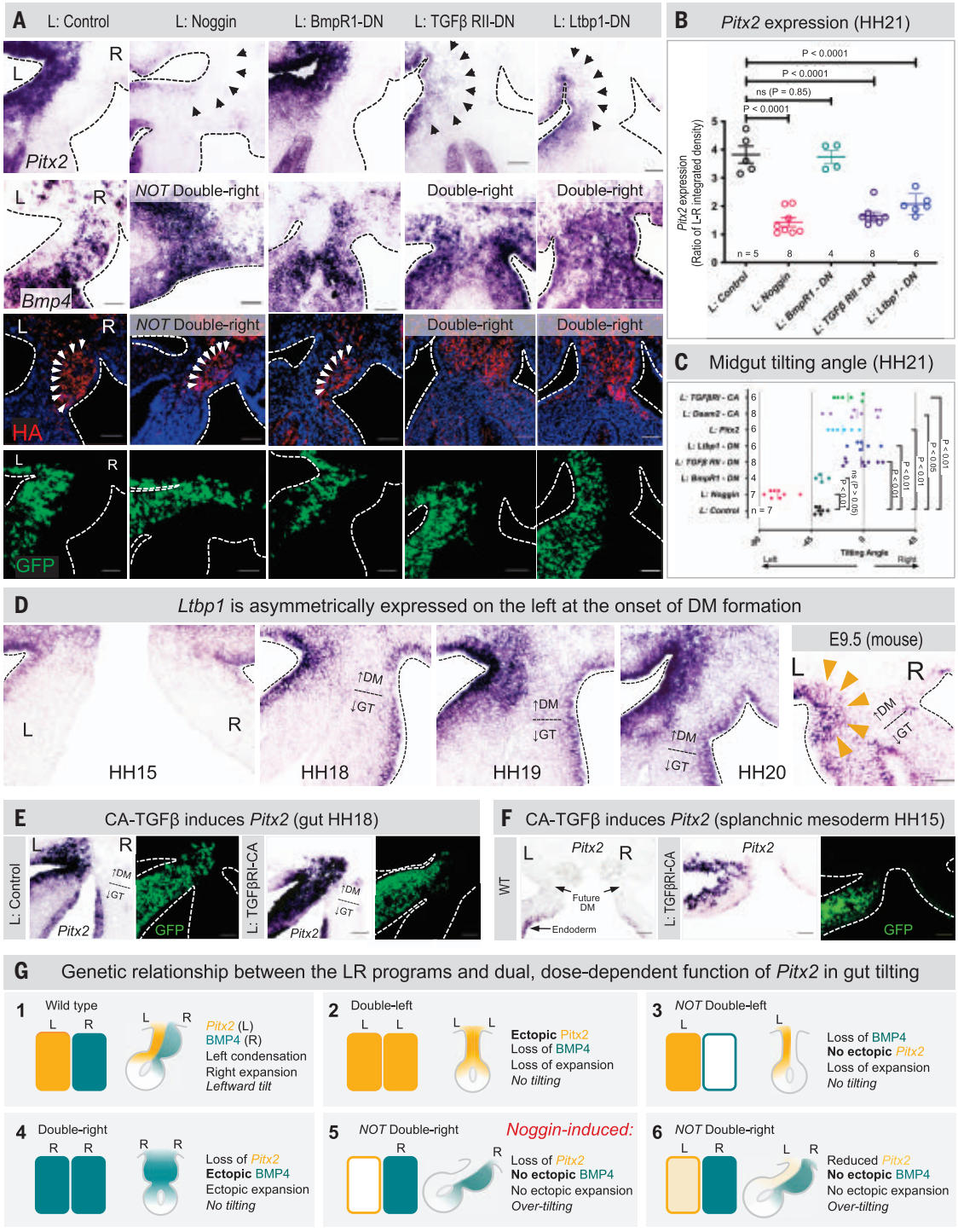



**Fig. 4. Latent TGFβ directs gut-specific asymmetry through *Pitx2* expression.** (**A**) *Pitx2* and *Bmp4* expression (ISH), HA (red), and gut tilting upon specified electroporations in the chicken DM cells marked by GFP. (**B** and **C**) *Pitx2* expression (B) was compared with one-way ANOVA followed by multiple comparisons with Tukey's correction (presented as mean ± SEM), and gut tilting (C) was compared using Watson's test (presented as a circular mean). (**D**) *Ltbp1* in chicken and mouse DM (ISH). (**E** and **F**) TGFβ RI-CA electroporation in chicken DM cells marked by GFP (E) drives *Pitx2* expression in the gut and splanchnic mesoderm (F). (**G**) *Pitx2* specifies the left side by suppressing *Bmp4* (**1** to **4**); *Pitx2* also directs gut-tilting morphogenesis, a function of *Pitx2* unmasked only in the absence of the double-right phenotype (**4** to **6**). Scale bars, 50 μm.

was also asymmetrically expressed in the left chicken DM (fig. S5A). Thus, TGFβ activation may regulate gut-specific *Pitx2* expression and be a target for Noggin inhibition. To test this, we electroporated a kinase-defective dominant-negative TGFβRII (TGFβRII-DN) (*33*) on the left side. This reduced *Pitx2* expression in the DM (Fig. 4, A and B). *Pitx2* expression was similarly decreased when a dominant-negative *Ltbp1* (*Ltbp1-DN*) was electroporated into the left DM to block TGFβ activation (*34*) (Fig. 4, A and B). By contrast, electroporation of

constitutively active TGFβRI (*TGFβ RI-CA*) (*35*) into the left gut splanchnic mesoderm expanded *Pitx2* expression along the gut tube primordium (Fig. 4E; HH18) and prematurely induced it at HH15 (Fig. 4F), bypassing the local presence of *Noggin* (fig. S4A). We also inserted resin beads into the left coelomic cavity (fig. S1C) (*7*, *11*) that were coated with a TGFβ receptor type 1 (TGFβ RI) inhibitor (SB431542) (*36*) or a more specific TGFβ RI and RII dual inhibitor (LY2109761) (*37*). In both cases, *Pitx2* expression was significantly reduced in response to TGFβ inhibition (Fig. 5, A and B). Finally, explants isolated from mouse intestine and cultured with LY2109761 had

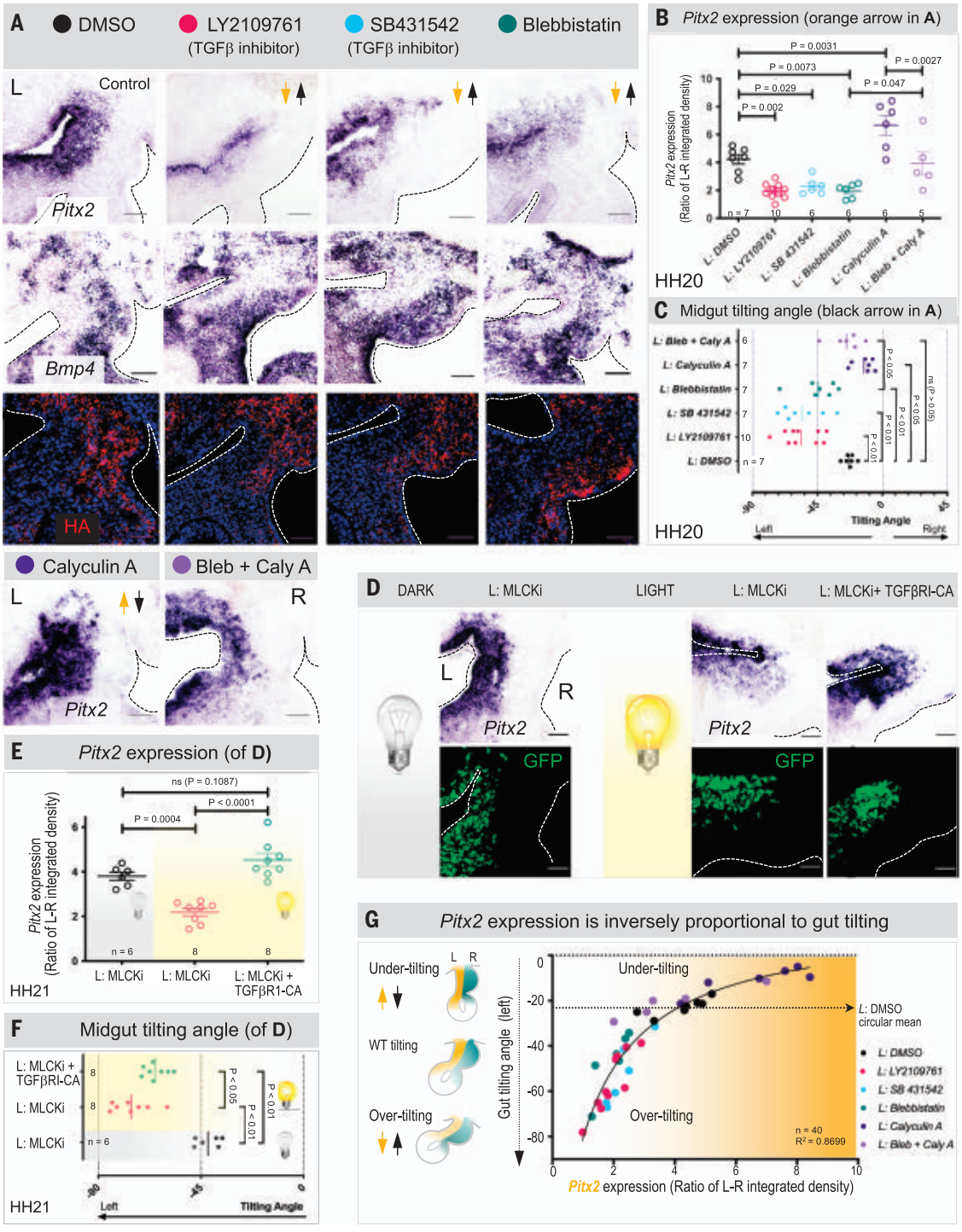



**Fig. 5. The contractile status of the DM is mechanically sensed by TGFβ-*Pitx2*.** (**A**) *Pitx2* and *Bmp4* expression (ISH) and HA (red) upon resin beads coated with specified drugs. (**B** and **C**) *Pitx2* expression (B) was compared with one-way ANOVA, followed by multiple comparisons with Tukey's correction (presented as mean ± SEM), and gut tilting (C) was compared using Watson's test (presented as circular mean). (**D** to **F**) *Pitx2* expression [quantified in (E)] and gut tilting [quantified in (F)] upon electroporating photoactivatable myosin light chain kinase (MLCK) or MLCK/TGFβ RI-CA in chicken DM cells marked by GFP. (**G**) Negative correlation between *Pitx2* expression and gut tilting, fitting a nonlinear regression curve ($R^2$ = 0.8699, HH20). Scale bars, 50 μm.

reduced *Pitx2* expression and perturbed gut rotation (fig. S5B).

### Gain of function of BMP4 on the left masks morphogenetic phenotypes of *Pitx2* loss

Genetic loss of all *Pitx2* expression in mice results in a DM with a double right-side isomerism and loss of gut tilting because BMP4 inhibition is lost on the left (Fig. 4G, panels 1 to 4). Similarly, electroporating *TGFβ RII-DN* or *Ltbp1-DN* caused a nearly complete loss of *Pitx2* expression (Fig. 4, A and B), resulting in bilateral activation of BMP4-mediated HA accumulation (Fig. 4A) and loss of gut tilting (Fig. 4C), identical to *Pitx2*-null mice. However, inhibiting *Pitx2* expression by electroporating Noggin in chicken embryos does not result in a double-right DM because Noggin also inhibits BMP4 and HA accumulation (no double-right isomerism; Fig. 4A). In such Noggin-expressing *Pitx2*-lacking embryos, the left-sided rotation (tilting angle) was increased ("over-tilting"; Fig. 4, A, C, and G, panels 1, 4, and 5). Similarly, perturbing TGFβ activation with drugs that reduced *Pitx2* expression caused the same over-tilting (Figs. 5, A to C, and 4G, panel 6). Thus, the over-tilting phenotype appears to result from an absence or reduction of *Pitx2* expression on the left, but can be revealed only when there is no accompanying double-right conversion. We conclude that *Pitx2* has two distinct roles in gut asymmetry: (i) it specifies the left side by suppressing *Bmp4* expression and (ii) it directs morphogenesis of gut tilting downstream of TGFβ activation (Fig. 4G).

### The contractile status of the DM is mechanically sensed by TGFβ-*Pitx2*

TGFβ becomes active when liberated from the latent complex that is covalently linked to ECM-resident Ltbp1 (*30*, *38*, *39*). Several mechanisms induce conformational changes of the latent complex, including the contractile force of cells (*38*–*41*) and mechanical tissue stretch in vitro and in vivo (*42*–*45*), leading to the release of TGFβ from the ECM (*46*). To test whether *Pitx2* expression in the DM responds to mechanical stress through mechanosensitive TGFβ release, we subjected slices of the DM and gut tissue to physical stretches of up to 20% (fig. S6A) (*47*). Stretching these slices resulted in increased (free) TGFβ (pan-TGFβ) protein (fig. S6B). This was accompanied by increased *Pitx2* expression on the left side of the DM that was proportional to the stretch applied (fig. S6C).

To test Pitx2 mechanosensitivity in vivo, we targeted the left coelomic cavity with resin beads soaked in blebbistatin, an inhibitor of nonmuscle myosin II ATPase (*48*). This decreased *Pitx2* expression in the left DM by >50% (Fig. 5, A and B) and increased the tilting angle (Fig. 5, A and C), a phenotype similar to that obtained after pharmacologic TGFβ inhibition (Fig. 5, A to C). By contrast, beads soaked with calyculin A, which activates myosin II and increases contractile force (*49*), increased *Pitx2* expression and decreased gut tilting (Fig. 5, A to C). A combination of blebbistatin and calyculin A restored both *Pitx2* expression and gut tilting (Fig. 5, A to C). We also electroporated a photoactivatable myosin light chain kinase (MLCK) inhibitor (PA-MKI), a more specific inhibitor of myosin II (*50*), allowing precise temporal disruption of contractility in the left DM after the establishment of molecular LR asymmetry. This decreased *Pitx2* expression (Fig. 5, D and E) and increased the tilting angle (Fig. 5, D and F). Co-electroporating PA-MKI and TGFβ RI-CA rescued *Pitx2* expression and partially restored gut tilting (Fig. 5, D to F). Thus, direct modulation of contractility in the DM is sufficient to alter *Pitx2* mRNA expression through latent release of TGFβ. Quantitative analyses of the above perturbations confirmed the mechanosensitivity of *Pitx2* abundance and the inverse relationship between *Pitx2* expression levels and the degree of gut tilting (Fig. 5G).

### *Pitx2* induces polarized condensation on the left by mechanical feedback with the right

A second wave of *Pitx2* expression is observed at the onset of DM formation, but this expression is further increased commensurate with expansion on the right, which might reflect mechanical feedback (Fig. 1C). Moreover, *Pitx2* drives polarized condensation on the left shortly after expansion on the right (*7*). Thus, we tested whether forces deriving initially from the right-side expansion might drive mechanical TGFβ ligand release and *Pitx2* dosage on the left, tightly regulating and polarizing mesenchymal condensation in response to expansion. This model would explain how *Pitx2* reduction on the left would impair the (active) responses to expansion on the right, causing deregulated (passive) over-tilting.

To test this model in vivo, we inhibited expansion and gut tilting by electroporating chicken embryos with hyaluronidase 2 (*Hyal2*) to degrade extracellular HA (*7*). To allow precise temporal manipulation of expansion after the establishment of molecular LR asymmetry, we used the tetracycline (Tet)-on inducible chicken expression construct TRE-EGFP for these experiments (*51*). Loss of expansion on the right decreased *Pitx2* expression on the left (Fig. 6, A and B). Conversely, we exaggerated ECM expansion by electroporating *Tsg6* with HA synthase 2 (*Has2*) into the right DM (*7*) (Fig. 6, A and E). Ectopic *Tsg6* plus *Has2* increased expansion on the right and significantly increased condensation of the neighboring left DM (Fig. 6E). Increased expansion also increased accumulation of free TGFβ protein (Fig. 6D) and *Pitx2* expression (Fig. 6 A and B). Thus, we propose an "accelerator-brake" model in which right-sided ECM expansion acts as an accelerator of gut rotation, whereas TGFβ-*Pitx2*–directed mesenchymal condensation on the left acts as a brake to inhibit rotation. Consistent with this model, direct overexpression of *TGFβRI-CA*, *Pitx2*, or the constitutively active form of the Pitx2 effector Daam2 (Daam2-CA) (*8*) on the left resulted in reduced gut tilting (Fig. 4C). Moreover, over-tilting caused by Noggin-induced loss of *Pitx2* expression (absence of brake) was associated with a loss of mesenchymal cell polarity within the left DM, as measured by the orientation of the Golgi apparatus (*8*) relative to the nucleus (fig. S7A). Disorganized condensation in the absence of polarity also resulted in the clumping of actin fibers (fig. S7B) in crowded areas, which was never observed during the wild-type *Pitx2*-driven condensation program (fig. S7B) (*8*) or during the mechanically exacerbated wild-type *Pitx2* program induced by the right-sided *Tsg6* plus *Has2* co-electroporation (fig. S7B). We interpret this to mean that mesenchymal cell crowding in the absence of *Pitx2* is a deregulated consequence of right-sided expansion (passive left-side compaction) as opposed to *Pitx2*-patterned (active) polarized condensation in response to expansion.

### *Pitx2* patterns DM stiffness on the left to resist right-side generated forces

Tissue deformation reflects a balance of forces and stiffness (*52*). We hypothesized that polarized condensation stiffens the left side, which resists right side–generated forces. To test whether gut tilting results from a combination of unequal forces and stiffness across the LR axis of the DM, we performed direct measurements of DM tissue stiffness during gut tilting using a microindentation system (Chiaro nanoindenter) (Fig. 6F). These studies revealed that the condensed left side of the wild-type DM was significantly stiffer than the hydrated nature of the HA matrix on the right side, whereas the gut tube had uniform intermediate stiffness (Fig. 6F and fig. S8). Ectopic expression of *Pitx2* on the right side stiffened the ECM, approximating the stiffness of the left side of the wild-type DM (Fig. 6F). Moreover, Noggin-induced loss of *Pitx2* expression on the left caused a loss in tissue stiffness of that side, resulting in tissue stiffness approximating that of the right side (Fig. 6F). Thus, *Pitx2* is necessary and sufficient to regulate DM stiffness during gut tilting.

After exaggerating expansion on the right (with *Tsg6* plus *Has2*), the soft hydrated HA-rich ECM became even softer and the left side became stiffer (Fig. 6F), reflecting the consequential increase in TGFβ (Fig. 6D) and *Pitx2* expression on the left (Fig. 6, A and B). These changes did not significantly alter the overall gut tilting (Fig. 6C), reinforcing the importance

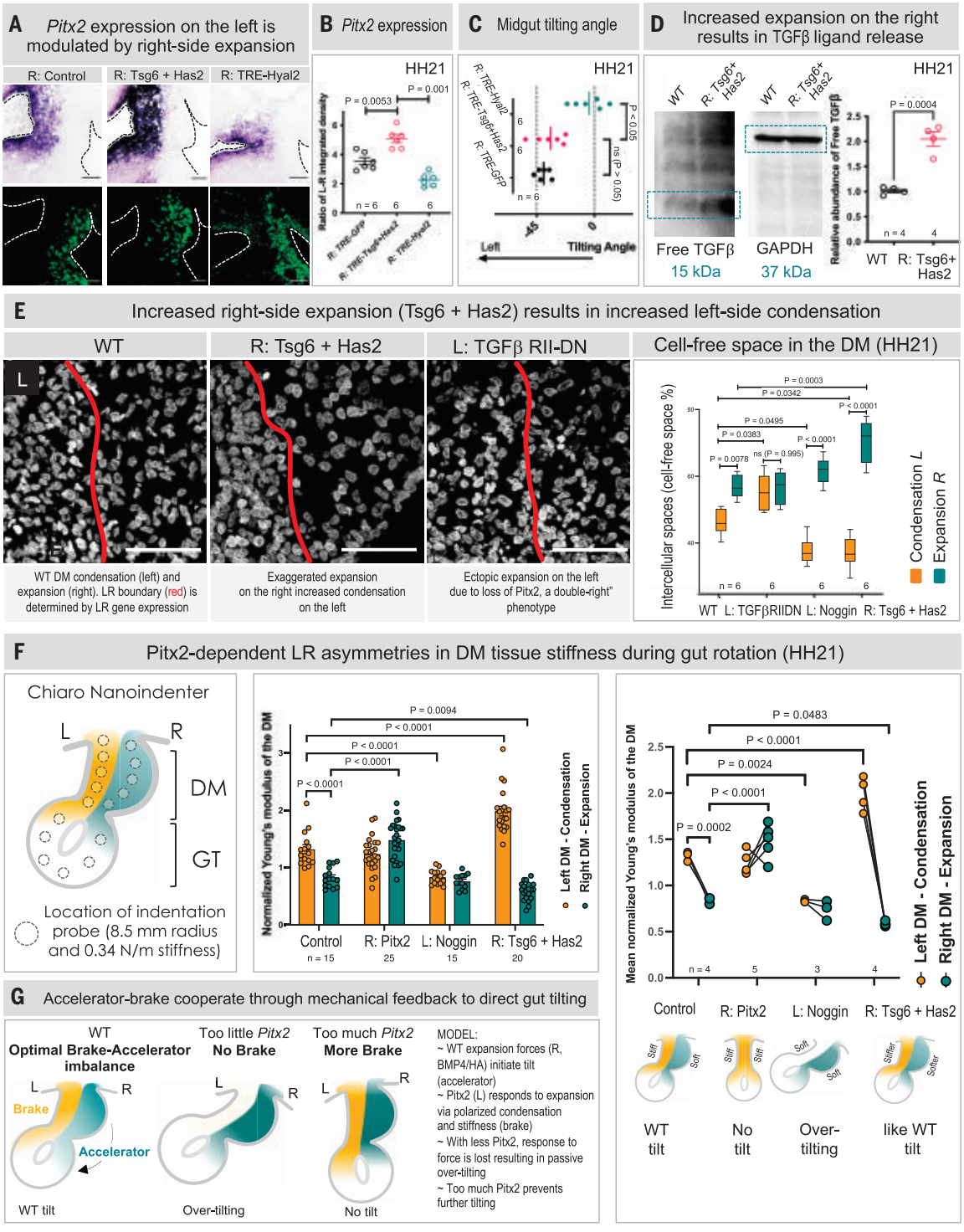



**Fig. 6. *Pitx2* patterns DM stiffness on the left by mechanical feedback with the right.** (**A**) *Pitx2* expression (ISH) and gut tilting upon loss of expansion (Hyal2) or exaggerated expansion (*Tsg6* plus *Has2*). (**B** and **C**) *Pitx2* expression (B) was compared with one-way ANOVA followed by multiple comparisons with Tukey's correction (presented as mean ± SEM), and gut tilting (C) was compared using Watson's test (presented as circular mean). (**D**) Exaggerated expansion increased free TGFβ on Western blot with control GAPDH and densitometric quantification by unpaired *t* test (presented as mean ± SEM). (**E**) DM mesenchymal cell compaction measured by percentage of internuclear space (DAPI) in wild-type and electroporated left and right sides of the chicken DM. Statistics were analyzed by one-way ANOVA followed by multiple comparisons with Tukey's correction (presented as mean ± SEM). (**F**) Indentation measures chicken DM stiffness (HH21). Quantifications were obtained by analyzing independent measurements (graph on the left) or by biological replicates (right) and normalized to the average stiffness of the gut tube. Statistics for left versus right comparison for conditions were analyzed by unpaired *t* test and across conditions by one-way ANOVA followed by multiple comparisons with Tukey's correction (presented as mean ± SEM). (**G**) Model for accelerator-brake mechanical feedback to rotate the gut. Scale bars, 50 μm.

of the mechanical feedback loop in balancing gut tilting. Our results show that the left-sided latent TGFβ complex responds to mechanical stress from the expanding right side and amplifies *Pitx2* expression, which in turn stiffens the left side through polarized condensation of the mesenchyme, tuning a second wave of *Pitx2* expression. Thus, our accelerator-brake model incorporates both biochemical and biomechanical inputs to cooperatively drive and steer the crucial counterclockwise rotation of the embryonic midgut (Fig. 6G).

As an additional test of our model, we predicted that we could reverse the conserved gut tilting by misexpression of the wild-type *Pitx2* on the right and TGFβRII-DN on the left to inhibit *Pitx2* expression. Indeed, gain of function of *Pitx2* on the right and loss of the left-sided TGFβ signaling reversed gut tilting (fig. S9).

## Discussion

*Pitx2* functions in both establishing LR asymmetry and translating this laterality to morphogenesis of the internal organs. However, it remains unresolved how *Pitx2* expression modifies cell behavior and how this is regulated within local, organ-specific microenvironments. In the DM, these events are characterized by an expansion of the right-side mesenchyme, which is driven by BMP-induced accumulation of HA; this expansion is accompanied by mesenchymal compaction on the left side. These asymmetries produce the characteristic counterclockwise gut rotation found in most vertebrates.

*Pitx2* expression during broad specification of axial chirality and the later execution of asymmetric tissue changes appear to be controlled by different regulatory pathways (Fig. 7, A and B). In the lateral plate mesoderm, *Nodal* induces the first wave of *Pitx2* expression on the left to govern embryonic laterality (Fig. 7A). In the DM, a second wave of *Pitx2* expression is controlled by the mechanosensitive TGFβ latency complex on the left (Fig. 7B). Deletion of *Nodal* from the lateral plate mesoderm preserved wild-type *Pitx2* expression ($n$ = 6/7) and gut laterality ($n$ = 12/14). Thus, whereas early *Nodal* expression at the node, the mouse signaling center where symmetry is first broken, is essential to specifying the embryonic LR body plan (*12*, *53*, *54*), further changes in local intestinal asymmetry can be uncoupled from *Nodal* transcription in the lateral plate mesoderm. The incomplete penetrance of correct gut rotation (two of 14 embryos rotated to the right with right-sided *Pitx2*) suggests that *Nodal* may be important to ensure the robust alignment of organ-level laterality with the LR body axis. Neither electroporation of *Nodal* nor TGFβRI-CA in the right splanchnic mesoderm was sufficient to induce bilateral *Pitx2* expression in the DM (fig. S10), reinforcing that local amplification of gut-specific *Pitx2* expression requires prior laterality information established during early embryogenesis. Thus, our study supports a two-step model of asymmetric organogenesis (*55*), and our findings support the presence of additional mechanisms through which laterality information is delivered to individual organ primordia.

We observed that Noggin is produced by splanchnic mesoderm surrounding the gut tube, and progressive distancing of the Noggin$^+$ gut tube primordium from the Noggin$^-$ cells of the DM may affect the timely activation of DM-specific LR programs (Fig. 7, A and B). Indeed, Noggin not only antagonizes BMP4 signaling within the right DM but also inhibits TGFβ-*Pitx2* activity on the left. This revealed a separate role for *Pitx2* in governing the mechanical properties of the DM, in which the forces generated by an adjoining tissue (right DM) are converted into a transcriptional response (*Pitx2* expression)

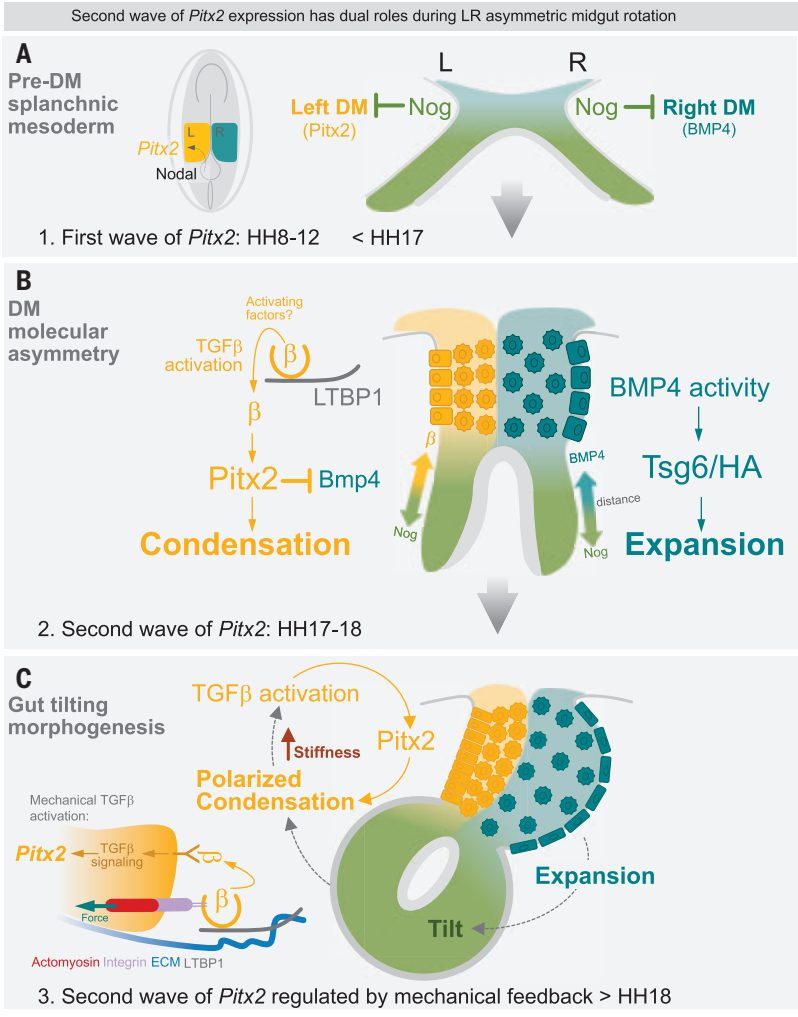



**Fig. 7. Second wave of *Pitx2* expression has dual roles during gut rotation.** (**A**) *Pitx2* expression is first initiated by *Nodal* at the node (first wave, HH8) but diminishes in the splanchnic mesoderm after HH12. Noggin in the splanchnic mesoderm ensures TGFβ and BMP signals are sequestered, preventing premature development of asymmetric LR programs. (**B**) First role of *Pitx2* to establish DM left-side identity: As the DM forms (HH17), the Noggin-rich gut tube primordium shifts ventrally, and the second wave of *Pitx2* expression appears, enabling the initiation of DM asymmetries. Regulated downstream of TGFβ, *Pitx2* expression on the left represses *Bmp4* to specify the DM left side. Confined to the right, BMP4 activity drives Tsg6/HA-mediated ECM expansion to initiate the leftward tilt (HH18). (**C**) Second role of *Pitx2* regulating gut-tilting morphogenesis. Tilting forces from expansion on the right (accelerator) potentiate TGFβ activation on the left (>HH18). The consequential *Pitx2* accumulation induced by TGFβ inhibits gut tilting through polarized mesenchymal cell condensation and tissue stiffness. Inset on the left depicts the release of free TGFβ cytokine from the latent complex. TGFβ pro-peptide (orange open circle) is bound to LTBP1 (gray). Cytoskeletal contraction forces (red, actomyosin filaments) engage integrins (purple) to release active TGFβ free to bind its receptor and activate TGFβ signaling.

through mechanical activation of the latent TGFβ complex (Fig. 7C), a phenomenon previously observed in cancer and fibrosis (*46*). Our data reveal that material properties across the LR axis of the DM, including tissue stiffness, determine the extent of deformation during gut rotation, a process patterned by *Pitx2* and executed by TGFβ and BMP4.

*Ltbp1*, the major regulator of latent TGFβ activation, is transcribed asymmetrically on the left side commensurate with *Pitx2* expression, which presumably enhances TGFβ release and activation on the left (Fig. 7, B and C). The LTBP family of proteins are expressed in restricted tissue types to provide specificity to TGFβ ligands that have widespread expression (*46*). This diversity is further expanded by the variety of stresses that can cause TGFβ release, including mechanical forces, pH, reactive oxygen species, and matrix metalloproteinases (*46*). Our findings in the DM reveal that mechanical forces propagate TGFβ-*Pitx2* signaling in the mechanically condensed left DM, but other stimuli may also participate. TGFβ and BMP signaling also function in LR asymmetric gut bending in sea urchin embryos, suggesting an evolutionarily ancient mechanism for gut-specific asymmetry (*56*).

## REFERENCES AND NOTES


1. N. M. Davis *et al*., The chirality of gut rotation derives from left-right asymmetric changes in the architecture of the dorsal mesentery. *Dev. Cell* **15**, 134–145 (2008). doi: 10.1016/j.devcel.2008.05.001; pmid: 18606147
2. T. Savin *et al*., On the growth and form of the gut. *Nature* **476**, 57–62 (2011). doi: 10.1038/nature10277; pmid: 21814276
3. M. S. Shalaby, K. Kuti, G. Walker, Intestinal malrotation and volvulus in infants and children. *BMJ* **347** (nov26 2), f6949 (2013). doi: 10.1136/bmj.f6949; pmid: 24285798
4. K. E. Applegate, Evidence-based diagnosis of malrotation and volvulus. *Pediatr. Radiol.* **39** (Suppl 2), S161–S163 (2009). doi: 10.1007/s00247-009-1177-x; pmid: 19308378
5. N. A. Kurpios *et al*., The direction of gut looping is established by changes in the extracellular matrix and in cell:cell adhesion. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8499–8506 (2008). doi: 10.1073/pnas.0803578105; pmid: 18574143
6. V. Hamburger, H. L. Hamilton, A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951). pmid: 24539719
7. A. Sivakumar *et al*., Midgut laterality is driven by hyaluronan on the right. *Dev. Cell* **46**, 533–551.e5 (2018). doi: 10.1016/j.devcel.2018.08.002; pmid: 30174180
8. I. C. Welsh *et al*., Integration of left-right Pitx2 transcription and Wnt signaling drives asymmetric gut morphogenesis via Daam2. *Dev. Cell* **26**, 629–644 (2013). doi: 10.1016/j.devcel.2013.07.019; pmid: 24091014
9. M. F. Lu, C. Pressman, R. Dyer, R. L. Johnson, J. F. Martin, Function of Rieger syndrome gene in left-right asymmetry and craniofacial development. *Nature* **401**, 276–278 (1999). doi: 10.1038/45797; pmid: 10499585
10. M. Logan, S. M. Pagán-Westphal, D. M. Smith, L. Paganessi, C. J. Tabin, The transcription factor Pitx2 mediates situs-specific morphogenesis in response to left-right asymmetric signals. *Cell* **94**, 307–317 (1998). doi: 10.1016/S0092-8674(00)81474-9; pmid: 9708733
11. A. Mahadevan *et al*., The left-right Pitx2 pathway drives organ-specific arterial and lymphatic development in the intestine. *Dev. Cell* **31**, 690–706 (2014). doi: 10.1016/j.devcel.2014.11.002; pmid: 25482882
12. M. Levin, R. L. Johnson, C. D. Stern, M. Kuehn, C. Tabin, A molecular pathway determining left-right asymmetry in chick embryogenesis. *Cell* **82**, 803–814 (1995). doi: 10.1016/0092-8674(95)90477-8; pmid: 7671308
13. M. Logan, C. Tabin, Targeted gene misexpression in chick limb buds using avian replication-competent retroviruses. *Methods* **14**, 407–420 (1998). doi: 10.1006/meth.1998.0595; pmid: 9608511
14. A. K. Ryan *et al*., Pitx2 determines left-right asymmetry of internal organs in vertebrates. *Nature* **394**, 545–551 (1998). doi: 10.1038/29004; pmid: 9707115
15. L. B. Zimmerman, J. M. De Jesús-Escobar, R. M. Harland, The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell* **86**, 599–606 (1996). doi: 10.1016/S0092-8674(00)80133-6; pmid: 8752214
16. J. Groppe *et al*., Structural basis of BMP signalling inhibition by the cystine knot protein Noggin. *Nature* **420**, 636–642 (2002). doi: 10.1038/nature01245; pmid: 12478285
17. A. A. Arraf, R. Yelin, I. Reshef, A. Kispert, T. M. Schultheiss, Establishment of the visceral embryonic midline is a dynamic process that requires bilaterally symmetric BMP signaling. *Dev. Cell* **37**, 571–580 (2016). doi: 10.1016/j.devcel.2016.05.018; pmid: 27326934
18. A. V. Bayramov *et al*., Novel functions of Noggin proteins: Inhibition of Activin/Nodal and Wnt signaling. *Development* **138**, 5345–5356 (2011). doi: 10.1242/jcs.105031; pmid: 22071106
19. M. T. Nguyen, J. Zhu, E. Nakamura, X. Bao, S. Mackem, Tamoxifen-dependent, inducible Hoxb6CreERT recombinase function in lateral plate and limb mesoderm, CNS isthmic organizer, posterior trunk neural crest, hindgut, and tailbud. *Dev. Dyn.* **238**, 467–474 (2009). doi: 10.1002/dvdy.21846; pmid: 19161221
20. R. T. Thomason, D. M. Bader, N. I. Winters, Comprehensive timeline of mesodermal development in the quail small intestine. *Dev. Dyn.* **241**, 1678–1694 (2012). doi: 10.1002/dvdy.23855; pmid: 22930586
21. K. Goto, Y. Kamiya, T. Imamura, K. Miyazono, K. Miyazawa, Selective inhibitory effects of Smad6 on bone morphogenetic protein type I receptors. *J. Biol. Chem.* **282**, 20603–20611 (2007). doi: 10.1074/jbc.M702100200; pmid: 17493940
22. E. Raja *et al*., The protein kinase LKB1 negatively regulates bone morphogenetic protein receptor signaling. *Oncotarget* **7**, 1120–1143 (2016). doi: 10.18632/oncotarget.6683; pmid: 26701726
23. H. Shiratori *et al*., Two-step regulation of left-right asymmetric expression of Pitx2: Initiation by nodal signaling and maintenance by Nkx2. *Mol. Cell* **7**, 137–149 (2001). doi: 10.1016/S1097-2765(01)00162-9; pmid: 11172719
24. Y. Saijoh *et al*., Left-right asymmetric expression of lefty2 and nodal is induced by a signaling pathway that includes the transcription factor FAST2. *Mol. Cell* **5**, 35–47 (2000). doi: 10.1016/S1097-2765(00)80401-3; pmid: 10678167
25. S. Forlani, K. A. Lawson, J. Deschamps, Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. *Development* **130**, 3807–3819 (2003). doi: 10.1242/dev.00573; pmid: 12835396
26. A. Desgrange, J. F. Le Garrec, S. Bernheim, T. H. Bønnelykke, S. M. Meilhac, Transient nodal signaling in left precursors coordinates opposed asymmetries shaping the heart loop. *Dev. Cell* **55**, 413–431.e6 (2020). doi: 10.1016/j.devcel.2020.10.008; pmid: 33171097
27. A. Monsoro-Burq, N. M. Le Douarin, BMP4 plays a key role in left-right patterning in chick embryos by maintaining Sonic Hedgehog asymmetry. *Mol. Cell* **7**, 789–799 (2001). doi: 10.1016/S1097-2765(01)00223-4; pmid: 11336702
28. J. Veerkamp *et al*., Unilateral dampening of Bmp activity by nodal generates cardiac left-right asymmetry. *Dev. Cell* **24**, 660–667 (2013). doi: 10.1016/j.devcel.2013.01.026; pmid: 23499359
29. O. H. Ocaña *et al*., A right-handed signalling pathway drives heart looping in vertebrates. *Nature* **549**, 86–90 (2017). doi: 10.1038/nature23454; pmid: 28880281
30. K. Miyazono, A. Olofsson, P. Colosetti, C. H. Heldin, A role of the latent TGF-β 1-binding protein in the assembly and secretion of TGF-β 1. *EMBO J.* **10**, 1091–1101 (1991). doi: 10.1002/j.1460-2075.1991.tb08049.x; pmid: 2022183
31. J. P. Annes, J. S. Munger, D. B. Rifkin, Making sense of latent TGFbeta activation. *J. Cell Sci.* **116**, 217–224 (2003). doi: 10.1242/jcs.00229; pmid: 12482908
32. J. Skonier *et al*., β ig-h3: A transforming growth factor-β-responsive gene encoding a secreted protein that inhibits cell attachment in vitro and suppresses the growth of CHO cells in nude mice. *DNA Cell Biol.* **13**, 571–584 (1994). doi: 10.1089/dna.1994.13.571; pmid: 8024701
33. J. L. Wrana *et al*., TGF β signals through a heteromeric protein kinase receptor complex. *Cell* **71**, 1003–1014 (1992). doi: 10.1016/0092-8674(92)90395-S; pmid: 1333888
34. J. P. Annes, Y. Chen, J. S. Munger, D. B. Rifkin, Integrin alphaVbeta6-mediated activation of latent TGF-β requires the latent TGF-β binding protein-1. *J. Cell Biol.* **165**, 723–734 (2004). doi: 10.1083/jcb.200312172; pmid: 15184403
35. A. Morén, T. Imamura, K. Miyazono, C. H. Heldin, A. Moustakas, Degradation of the tumor suppressor Smad4 by WW and HECT domain ubiquitin ligases. *J. Biol. Chem.* **280**, 22115–22123 (2005). doi: 10.1074/jbc.M414027200; pmid: 15817471
36. G. J. Inman *et al*., SB-431542 is a potent and specific inhibitor of transforming growth factor-β superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. *Mol. Pharmacol.* **62**, 65–74 (2002). doi: 10.1124/mol.62.1.65; pmid: 12065756
37. D. Melisi *et al*., LY2109761, a novel transforming growth factor β receptor type I and type II dual inhibitor, as a therapeutic approach to suppressing pancreatic cancer metastasis. *Mol. Cancer Ther.* **7**, 829–840 (2008). doi: 10.1158/1535-7163.MCT-07-0337; pmid: 18413796
38. P. J. Wipff, D. B. Rifkin, J. J. Meister, B. Hinz, Myofibroblast contraction activates latent TGF-β1 from the extracellular matrix. *J. Cell Biol.* **179**, 1311–1323 (2007). doi: 10.1083/jcb.200704042; pmid: 18086923
39. L. Buscemi *et al*., The single-molecule mechanics of the latent TGF-β1 complex. *Curr. Biol.* **21**, 2046–2054 (2011). doi: 10.1016/j.cub.2011.11.037; pmid: 22169532
40. M. M. Giacomini, M. A. Travis, M. Kudo, D. Sheppard, Epithelial cells utilize cortical actin/myosin to activate latent TGF-β through integrin α(v)β(6)-dependent physical force. *Exp. Cell Res.* **318**, 716–722 (2012). doi: 10.1016/j.yexcr.2012.01.020; pmid: 22309779
41. N. C. Henderson *et al*., Targeting of αv integrin identifies a core molecular pathway that regulates fibrosis in several organs. *Nat. Med.* **19**, 1617–1624 (2013). doi: 10.1038/nm.3282; pmid: 24216753
42. R. Sakata, T. Ueno, T. Nakamura, H. Ueno, M. Sata, Mechanical stretch induces TGF-beta synthesis in hepatic stellate cells. *Eur. J. Clin. Invest.* **34**, 129–136 (2004). doi: 10.1111/j.1365-2362.2004.01302.x; pmid: 14764076
43. M. T. Hamzeh, R. Sridhara, L. D. Alexander, Cyclic stretch-induced TGF-β1 and fibronectin expression is mediated by β1-integrin through c-Src- and STAT3-dependent pathways in renal epithelial cells. *Am. J. Physiol. Renal Physiol.* **308**, F425–F436 (2015). doi: 10.1152/ajprenal.00589.2014; pmid: 25477471
44. A. R. Froese *et al*., Stretch-induced activation of transforming growth factor-β1 in pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **194**, 84–96 (2016). doi: 10.1164/rccm.201508-1638OC; pmid: 26771871
45. M. Walker, M. Godin, A. E. Pelling, Mechanical stretch sustains myofibroblast phenotype and function in microtissues through latent TGF-β1 activation. *Integr. Biol.* **12**, 199–210 (2020). doi: 10.1093/intbio/zyaa015; pmid: 32877929
46. I. B. Robertson, D. B. Rifkin, Regulation of the Bioavailability of TGF-β and TGF-β-Related Proteins. *Cold Spring Harb. Perspect. Biol.* **8**, a021907 (2016). doi: 10.1101/cshperspect.a021907; pmid: 27252363
47. M. L. Lombardi, M. Zwerger, J. Lammerding, Biophysical assays to probe the mechanical properties of the interphase cell nucleus: Substrate strain application and microneedle manipulation. *J. Vis. Exp.* (55): 3087 (2011). doi: 10.3791/3087; pmid: 21946671
48. M. Zhang, P. V. Rao, Blebbistatin, a novel inhibitor of myosin II ATPase activity, increases aqueous humor outflow facility in perfused enucleated porcine eyes. *Invest. Ophthalmol. Vis. Sci.* **46**, 4130–4138 (2005). doi: 10.1167/iovs.05-0164; pmid: 16249490
49. A. Suzuki, T. Itoh, Effects of calyculin A on tension and myosin phosphorylation in skinned smooth muscle of the rabbit mesenteric artery. *Br. J. Pharmacol.* **109**, 703–712 (1993). doi: 10.1111/j.1476-5381.1993.tb13631.x; pmid: 8395295
50. J. J. Yi, H. Wang, M. Vilela, G. Danuser, K. M. Hahn, Manipulation of endogenous kinase activity in living cells using photoswitchable inhibitory peptides. *ACS Synth. Biol.* **3**, 788–795 (2014). doi: 10.1021/sb5001356; pmid: 24905630
51. T. Watanabe *et al*., Tet-on inducible system combined with in ovo electroporation dissects multiple roles of genes in somitogenesis of chicken embryos. *Dev. Biol.* **305**, 625–636 (2007). doi: 10.1016/j.ydbio.2007.01.042; pmid: 17359965
52. N. I. Petridou, C.-P. Heisenberg, Tissue rheology in embryonic organization. *EMBO J.* **38**, e102497 (2019). doi: 10.15252/embj.2019102497; pmid: 31512749
53. J. Brennan, D. P. Norris, E. J. Robertson, Nodal activity in the node governs left-right asymmetry. *Genes Dev.* **16**, 2339–2344 (2002). doi: 10.1101/gad.1016202; pmid: 12231623
54. Y. Saijoh, S. Oki, S. Ohishi, H. Hamada, Left-right patterning of the mouse lateral plate requires nodal produced in the node.

*Dev. Biol.* **256**, 160–172 (2003). doi: 10.1016/S0012-1606(02)00121-5; pmid: 12654299
55. N. A. Brown, L. Wolpert, The development of handedness in left/right asymmetry. *Development* **109**, 1–9 (1990). doi: 10.1242/dev.109.1.1; pmid: 2209459
56. H. Suzuki, S. Yaguchi, Transforming growth factor-β signal regulates gut bending in the sea urchin embryo. *Dev. Growth Differ.* **60**, 216–225 (2018). doi: 10.1111/dgd.12434; pmid: 29878318

## ACKNOWLEDGMENTS

We thank J. Liu, D. Gludish, D. Noden, and members of the Kurpios lab for suggestions on the manuscript; A. Desgrange and S. Meilhac for the production of *Nodal* mutant mouse embryos and critical feedback; R. Harland for the Noggin-null mouse; S. Mackem for the Hoxb6 CreERT mice; T. Schultheiss, A. Bandyopadyay, C. Tabin, C. Cepko, C. Krull, L. Niswander, P. ten Dijke, J. Massague, J. Wrana, A. Moustakas, and P. Brickell for plasmids; and C. Demler, R. Slater, F. Lee, B. Laslow, A. Sivakumar, C. Harris, M. Simoes-Costa, and the Cornell Imaging Facility for technical assistance. **Funding:** This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (grants R01 DK092776 and R01 DK107634 to N.A.K.); the March of Dimes (grant 1-FY11-520 to N.A.K.); the National Heart, Lung and Blood Institute (grant R01 HL082792 to J.L.); the Volkswagen Foundation (J.L.); the Cornell Center for Vertebrate Genomics Scholarship (B.D.S.); and the National Institutes of Health (grant 1S10RR025502 to the Cornell Institute of Biotechnology). **Author contributions:** Conceptualization: B.D.S., N.A.K.; Funding acquisition: N.A.K.; Investigation: B.D.S.; Methodology: B.D.S., E.B., S.G., S.H., J.L., N.Z.S., J.L.; Supervision: N.A.K.; Visualization: B.D.S., E.B.; Writing: B.D.S., N.A.K. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. 

## SUPPLEMENTARY MATERIALS



Submitted 12 July 2021; accepted 25 August 2022
10.1126/science.abl3921

RESEARCH ARTICLE SUMMARY

QUANTUM PHYSICS

# Provably efficient machine learning for quantum many-body problems

Hsin-Yuan Huang*, Richard Kueng, Giacomo Torlai, Victor V. Albert, John Preskill

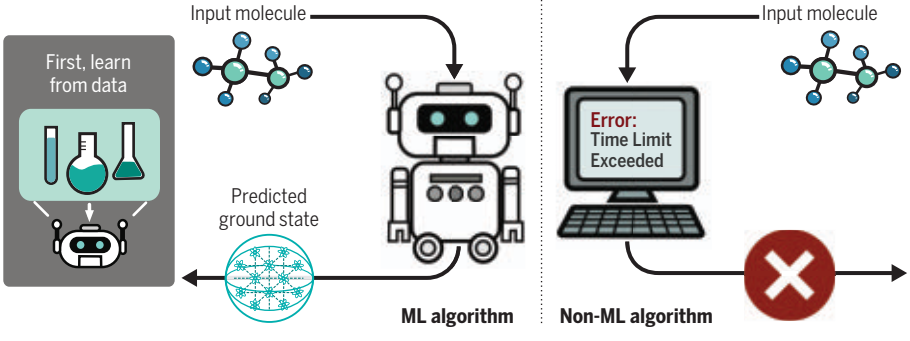
**INTRODUCTION:** Solving quantum many-body problems, such as finding ground states of quantum systems, has far-reaching consequences for physics, materials science, and chemistry. Classical computers have facilitated many profound advances in science and technology, but they often struggle to solve such problems. Scalable, fault-tolerant quantum computers will be able to solve a broad array of quantum problems but are unlikely to be available for years to come. Meanwhile, how can we best exploit our powerful classical computers to advance our understanding of complex quantum systems? Recently, classical machine learning (ML) techniques have been adapted to investigate problems in quantum many-body physics. So far, these approaches are mostly heuristic, reflecting the general paucity of rigorous theory in ML. Although they have been shown to be effective in some intermediate-size experiments, these methods are generally not backed by convincing theoretical arguments to ensure good performance.

**RATIONALE:** A central question is whether classical ML algorithms can provably outperform non-ML algorithms in challenging quantum many-body problems. We provide a concrete answer by devising and analyzing classical ML algorithms for predicting the properties of ground states of quantum systems. We prove that these ML algorithms can efficiently and accurately predict ground-state properties of gapped local Hamiltonians, after learning from data obtained by measuring other ground states in the same quantum phase of matter. Furthermore, under a widely accepted complexity-theoretic conjecture, we prove that no efficient classical algorithm that does not learn from data can achieve the same prediction guarantee. By generalizing from experimental data, ML algorithms can solve quantum many-body problems that could not be solved efficiently without access to experimental data.

**RESULTS:** We consider a family of gapped local quantum Hamiltonians, where the Hamiltonian $H(x)$ depends smoothly on $m$ parameters (denoted by $x$). The ML algorithm learns from a set of training data consisting of sampled values of $x$, each accompanied by a classical representation of the ground state of $H(x)$. These training data could be obtained from either classical simulations or quantum experiments. During the prediction phase, the ML algorithm predicts a classical representation of ground states for Hamiltonians different from those in the training data; ground-state properties can then be estimated using the predicted classical representation. Specifically, our classical ML algorithm predicts expectation values of products of local observables in the ground state, with a small error when averaged over the value of $x$. The run time of the algorithm and the amount of training data required both scale polynomially in $m$ and linearly in the size of the quantum system. Our proof of this result builds on recent developments in quantum information theory, computational learning theory, and condensed matter theory. Furthermore, under the widely accepted conjecture that nondeterministic polynomial-time (NP)–complete problems cannot be solved in randomized polynomial time, we prove that no polynomial-time classical algorithm that does not learn from data can match the prediction performance achieved by the ML algorithm.

In a related contribution using similar proof techniques, we show that classical ML algorithms can efficiently learn how to classify quantum phases of matter. In this scenario, the training data consist of classical representations of quantum states, where each state carries a label indicating whether it belongs to phase $A$ or phase $B$. The ML algorithm then predicts the phase label for quantum states that were not encountered during training. The classical ML algorithm not only classifies phases accurately, but also constructs an explicit classifying function. Numerical experiments verify that our proposed ML algorithms work well in a variety of scenarios, including Rydberg atom systems, two-dimensional random Heisenberg models, symmetry-protected topological phases, and topologically ordered phases.

**CONCLUSION:** We have rigorously established that classical ML algorithms, informed by data collected in physical experiments, can effectively address some quantum many-body problems. These rigorous results boost our hopes that classical ML trained on experimental data can solve practical problems in chemistry and materials science that would be too hard to solve using classical processing alone. Our arguments build on the concept of a succinct classical representation of quantum states derived from randomized Pauli measurements. Although some quantum devices lack the local control needed to perform such measurements, we expect that other classical representations could be exploited by classical ML with similarly powerful results. How can we make use of accessible measurement data to predict properties reliably? Answering such questions will expand the reach of near-term quantum platforms. ■




**Classical algorithms for quantum many-body problems.** Classical ML algorithms learn from training data, obtained from either classical simulations or quantum experiments. Then, the ML algorithm produces a classical representation for the ground state of a physical system that was not encountered during training. Classical algorithms that do not learn from data may require substantially longer computation time to achieve the same task.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: hsinyuan@caltech.edu
Cite this article as H.-Y. Huang *et al.*, *Science* 377, eabk3333 (2022). DOI: 10.1126/science.abk3333

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abk3333

RESEARCH ARTICLE

QUANTUM PHYSICS

# Provably efficient machine learning for quantum many-body problems

Hsin-Yuan Huang[1]*, Richard Kueng[2], Giacomo Torlai[3], Victor V. Albert[4], John Preskill[1,3]

Classical machine learning (ML) provides a potentially powerful approach to solving challenging quantum many-body problems in physics and chemistry. However, the advantages of ML over traditional methods have not been firmly established. In this work, we prove that classical ML algorithms can efficiently predict ground-state properties of gapped Hamiltonians after learning from other Hamiltonians in the same quantum phase of matter. By contrast, under a widely accepted conjecture, classical algorithms that do not learn from data cannot achieve the same guarantee. We also prove that classical ML algorithms can efficiently classify a wide range of quantum phases. Extensive numerical experiments corroborate our theoretical results in a variety of scenarios, including Rydberg atom systems, two-dimensional random Heisenberg models, symmetry-protected topological phases, and topologically ordered phases.

Solving quantum many-body problems, such as finding ground states of quantum systems, has far-reaching consequences for physics, materials science, and chemistry. Although classical computers have facilitated many profound advances in science and technology, they often struggle to solve such problems. Powerful methods, such as density functional theory (*1*, *2*), quantum Monte Carlo (*3*–*5*), and density-matrix renormalization group (*6*, *7*), have enabled solutions to certain restricted instances of many-body problems, but many general classes of problems remain outside the reach of even the most advanced classical algorithms.

Scalable, fault-tolerant quantum computers will be able to solve a broad array of quantum problems but are unlikely to be available for years to come. Meanwhile, how can we best exploit our powerful classical computers to advance our understanding of complex quantum systems? Recently, classical machine learning (ML) techniques have been adapted to investigate problems in quantum many-body physics (*8*, *9*) with promising results (*10*–*27*). So far, these approaches are mostly heuristic, reflecting the general paucity of rigorous theory in ML. Although they were shown to be effective in some intermediate-size experiments (*28*–*30*), these methods are generally not backed by convincing theoretical arguments to ensure good performance, particularly for problem instances where traditional classical algorithms falter.

[1]Institute for Quantum Information and Matter and Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA, USA. [2]Institute for Integrated Circuits, Johannes Kepler University, Linz, Austria. [3]AWS Center for Quantum Computing, Pasadena, CA, USA. [4]Joint Center for Quantum Information and Computer Science, National Institute of Standards and Technology and University of Maryland, College Park, MD, USA.
*Corresponding author. Email: hsinyuan@caltech.edu

In general, simulating quantum many-body physics is hard for classical computers because accurately describing an $n$-qubit quantum system may require an amount of classical data that is exponential in $n$. In prior work, this bottleneck has been addressed using classical shadows—succinct classical descriptions of quantum many-body states that can be used to accurately predict a wide range of properties with rigorous performance guarantees (*31*, *32*). Furthermore, this quantum-to-classical conversion technique can be readily implemented in various existing quantum experiments (*33*–*35*). Classical shadows create opportunities for addressing quantum problems using classical methods, such as ML. In this paper, we build on the classical shadow formalism and devise polynomial-time classical ML algorithms for quantum many-body problems that are supported by rigorous theory.

We consider two applications of classical ML, indicated in Fig. 1. The first application we examine is learning to predict classical representations of quantum many-body ground states. We consider a family of Hamiltonians, where the Hamiltonian $H(x)$ depends smoothly on $m$ real parameters (denoted by $x$). The ML algorithm is trained on a set of training data consisting of sampled values of $x$, each accompanied by the corresponding classical shadow for the ground state $\rho(x)$ of $H(x)$. These training data could be obtained from either classical simulations or quantum experiments. During the prediction phase, the ML algorithm predicts a classical representation of $\rho(x)$ for values of $x$ different from those in the training data. Ground-state properties can then be estimated using the predicted classical representation.

This learning algorithm is efficient, provided that the ground-state properties to be predicted do not vary too rapidly as a function of $x$. Sufficient upper bounds on the gradient can be derived for any family of gapped, geometrically local Hamiltonians in any finite spatial dimension, as long as the property of interest is the expectation value of a sum of few-body observables. The conclusion is that any such property can be predicted with a small average error, where the amount of training data and the classical computation time are polynomial in $m$ and linear in the system size. Furthermore, we show that classical algorithms that do not learn from data cannot make accurate predictions in polynomial time without violating widely accepted complexity-theoretic conjectures. Together, we rigorously establish the advantage of ML algorithms with data over those without data (*36*) in a physically relevant task.

The classical ML algorithm could generalize from training data that are obtained either through quantum experiments or classical simulations; the same rigorous performance guarantees apply in either case. If the training data are obtained from quantum experiments, the rigorous result shows that classical ML can explore and predict properties of new physical systems that are challenging to prepare and measure in the laboratory. Even if the experimentalists only have limited measurement capability, such as being able to measure a specific property of $\rho(x)$, the theorem established in this work immediately implies that a classical ML model can predict that specific property accurately. If the training data are generated classically, it could be more efficient and more accurate to use the ML model to predict properties for new values of the input $x$ rather than doing new simulations, which could be computationally very demanding. Promising insights into quantum many-body physics are already being obtained using classical ML based on classical simulation data (*10*, *12*, *14*, *17*, *19*, *20*, *23*–*25*, *37*, *38*). Our rigorous analysis identifies general conditions that guarantee the success of classical ML models and elucidate the advantages of classical ML models over non-ML algorithms, which do not learn from data. These results enhance the prospects for interpretable ML techniques (*38*–*40*) to further shed light on quantum many-body physics.

In the second application we examine, the goal is to classify quantum states of matter into phases (*41*) in a supervised learning scenario. Suppose that during training we are provided with sample quantum states that carry labels indicating whether each state belongs to phase $A$ or phase $B$. Our goal is to classify the phase for new quantum states that were not encountered during training. We assume that, during both the training and classification stages, each quantum state is represented by its classical shadow, which could be obtained either from a classical computation or from an experiment on a quantum device. The classical ML model, then, trains on labeled classical shadows

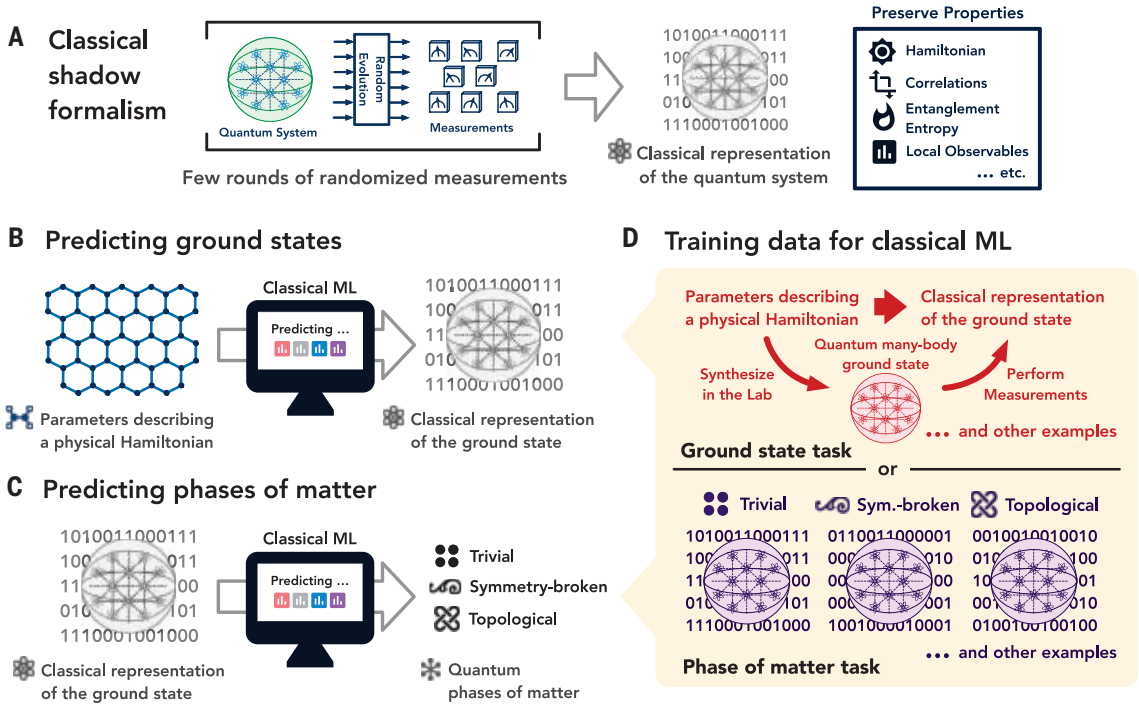



**Fig. 1. Central concepts.** (**A**) Efficient quantum-to-classical conversion. The classical shadow of a quantum state, constructed by measuring very few copies of the state, can be used to predict many properties of the state with a rigorous performance guarantee. (**B**) Predicting ground-state properties. After training on data obtained in quantum experiments, a classical ML model predicts a classical representation of the ground state $\rho(x)$ of the Hamiltonian $H(x)$ for parameters $x$ spanning the entire phase. This representation yields estimates of the properties of $\rho(x)$, avoiding the need to run exhaustive classical computations or quantum experiments. (**C**) Classifying quantum phases. After training, a classical ML model receives a classical representation of a quantum state and classifies the phase from which the state was drawn. (**D**) Training data. For predicting ground states, the classical ML model receives a classical representation of $\rho(x)$ for each value of $x$ sampled during training. For predicting quantum phases of matter, the training data consist of classical representations of quantum states accompanied by labels identifying the phase to which each state belongs.

and learns to predict labels for new classical shadows.

We assume that the $A$ and $B$ phases can be distinguished by a nonlinear function of marginal density operators of subsystems of constant size. This assumption is reasonable because we expect the phase to be revealed in subsystems that are larger than the correlation length but do not depend on the total system size. We show that if such a function exists, a classical ML model can learn to distinguish the phases using an amount of training data and classical processing that are polynomial in the system size. We do not need to know anything about this nonlinear function in advance, apart from its existence.

Here, we review the classical shadow formalism (*31*) and use this formalism to derive rigorous guarantees for ML algorithms in predicting ground-state properties and classifying quantum phases of matter. We also describe numerical experiments in a wide range of physical systems to support our theoretical results.

### Constructing efficient classical representations of quantum systems

We begin with an overview of the randomized measurement toolbox (*31*, *32*, *42*–*45*), relegating further details to section S1 (*46*). We approximate an $n$-qubit quantum state $\rho$ by performing randomized single-qubit Pauli measurements on $T$ copies of $\rho$. That is, we measure every qubit of the unknown quantum state $\rho$ in a random Pauli basis $X$, $Y$, or $Z$ to yield a measurement outcome of ±1. Collapse of the wave function implies that this measurement procedure transforms $\rho$ into a random pure product state $|s^{(t)}\rangle = \otimes_{i=1}^{n} |s_i^{(t)}\rangle$, where $|s_i^{(t)}\rangle \in \{|0\rangle, |1\rangle, |+\rangle, |-\rangle, |i+\rangle, |i-\rangle\}$ are eigenstates of the selected Pauli matrices. Performing one randomized measurement grants us classical access to one such snapshot. Performing a total of $T$ randomized measurements grants us access to an entire collection $S_T(\rho) = \left\{ |s_i^{(t)}\rangle : i \in \{1, ..., n\}, \ , t, \in \{1, ..., T\} \right\}$. Each element is a highly structured single-qubit pure state, and there are $nT$ of them in total. So, $3nT$ bits suffice to store the entire collection in classical memory. The randomized measurements can be performed in actual physical experiments or through classical simulations. Resulting data can then be used to approximate the underlying $n$-qubit state $\rho$

$$\rho \approx \sigma_T(\rho) = \frac{1}{T} \sum_{t=1}^{T} \sigma_1^{(t)} \otimes ... \otimes \sigma_n^{(t)}$$
$$\text{where } \sigma_i^{(t)} = 3 |s_i^{(t)}\rangle \langle s_i^{(t)}| - I \qquad (1)$$

and $I$ denotes the $2 \times 2$ identity matrix. This classical shadow representation (*31*, *32*) exactly reproduces the global density matrix in the limit $T \to \infty$, but $T = O[\text{const}^r \log(n)/\varepsilon^2]$ already provides an $\varepsilon$-accurate approximation of all reduced-$r$-body density matrices (in trace distance). This, in turn, implies that we can use $\sigma_T(\rho)$ to predict any function that depends on only reduced-density matrices, such as expectation values of (sums of) local observables and (sums of) entanglement entropies of small subsystems. Classical storage and postprocessing costs also remain tractable in this regime. To summarize, the classical shadow formalism equips us with an efficient quantum-to-classical converter that allows classical machines to efficiently and reliably estimate subsystem properties of any quantum state $\rho$.

### Predicting ground states of quantum many-body systems

We consider the task of predicting ground-state representations of quantum many-body Hamiltonians in finite spatial dimensions. Suppose that a family of geometrically local, $n$-qubit Hamiltonians $\{H(x) : x \in [-1, 1]^m\}$ is parameterized by a classical variable $x$. That is, $H(x)$ smoothly maps a bounded $m$-dimensional

vector $x$ (parametrization) to a Hermitian matrix of size $2^n \times 2^n$ ($n$-qubit Hamiltonian). We do not impose any additional structure on this mapping; in particular, we do not assume knowledge about how the physical Hamiltonian depends on the parameterization. The goal is to learn a model $\hat{\sigma}(x)$ that can predict properties of the ground state $\rho(x)$ associated with the Hamiltonian. This problem arises in many practical scenarios. Suppose diligent experimental effort has produced experimental data for ground-state properties of various physical systems. We would like to use these data to train an ML model that predicts ground-state representations of hitherto unexplored physical systems.

### An ML algorithm with rigorous guarantee

We will prove that a classical ML algorithm can predict classical representations of ground states after training on data belonging to the same quantum phase of matter. Formally, we consider a smooth family of Hamiltonians $H(x)$ with a constant spectral gap. During the training phase of the ML algorithm, many values of $x$ are randomly sampled, and for each sampled $x$, the classical shadow of the corresponding ground state $\rho(x)$ of $H(x)$ is provided, either by classical simulations or quantum experiments. The full training data of size $N$ are given by $\{x_l \to \sigma_T[\rho(x_l)]\}_{l=1}^{N}$, where $T$ is the number of randomized measurements in the construction of the classical shadows at each value of $x_l$.

We train classical ML models using the size-$N$ training data, such that when given the input $x_l$, the ML model can produce a classical representation $\hat{\sigma}(x)$ that approximates $\sigma_T[\rho(x_l)]$. During prediction, the classical ML model produces $\hat{\sigma}(x)$ for values of $x$ different from those in the training data. Although $\hat{\sigma}(x)$ and $\sigma_T[\rho(x_l)]$ classically represent exponentially large density matrices, the training and prediction can be done efficiently on a classical computer using various existing classical ML models, such as neural networks with large hidden layers (*47*–*50*) and kernel methods (*51*, *52*). In particular, the predicted output of the trained classical ML models can be written as the extrapolation of the training data using a learned metric $\kappa(x, x_l) \in R$

$$\hat{\sigma}(x) = \frac{1}{N}\sum_{l=1}^{N} \kappa(x, x_l)\sigma_T[\rho(x_l)] \qquad (2)$$

For example, prediction using a trained neural network with large hidden layers (*46*) is equivalent to using the metric $\kappa(x, x_l) = \sum_{l'=1}^{N} f^{(NTK)}(x, x_{l'})\left(F^{-1}\right)_{l'l}$, where $f^{(NTK)}(x, x')$ is the neural tangent kernel (*47*) corresponding to the neural network and $F_{l'l} = f^{(NTK)}(x_{l'}, x_l)$ [see section S3 (*46*) for more discussion]. The ground-state properties are then estimated using these predicted classical representations $\hat{\sigma}(x)$. Specifically, $f_O(x) = tr[O\rho(x)]$ can be predicted efficiently whenever $O$ is a sum of few-body operators.

To derive a provable guarantee, we consider the simple metric $\kappa(x, x_l) = \sum_{k \in Z^m, \|k\|_2 \le \Lambda} \cos[\pi k \cdot (x - x_l)]$ with cutoff $\Lambda$, which we refer to as the $l_2$-Dirichlet kernel. We prove that the prediction will be accurate and efficient if the function $f_O(x)$ does not vary too rapidly when $x$ changes in any direction. Sufficient upper bounds on the gradient magnitude of $f_O(x)$ can be derived using quasi-adiabatic continuation (*53*, *54*).

Under the $l_2$-Dirichlet kernel, the classical ML model is equivalent to learning a truncated Fourier series to approximate the function $f_O(x)$. The parameter $\Lambda$ is a cutoff for the wave number $k$ that depends on (upper bounds on) the gradient of $f_O(x)$. Using statistical analysis, one can guarantee that $E_x|tr[O\hat{\sigma}(x)] - f_O(x)|^2 \le \varepsilon$ as long as the amount of training data obeys $N = m^{O(1/\varepsilon)}$ in the $m \to \infty$ limit. The conclusion is that any such $f_O(x)$ can be predicted with a small constant average error, where the amount of training data and the classical computation time are polynomial in $m$ and at most linear in the system size $n$. Moreover, the training data need only contain a single classical shadow snapshot at each point $x_l$ in the parameter space (i.e., $T = 1$). An informal statement of the theorem is given below; we explain the proof strategy in section S5 and provide more details in section S6 (*46*). We also discuss how one could generalize the proof to long-range interacting systems, electronic Hamiltonians, and other settings, including when one cannot perform classical shadow tomography (*31*), in section S6.2 (*46*).

*Theorem 1 (learning to predict ground-state representations; informal)*

For any smooth family of Hamiltonians $\{H(x) : x \in [-1, 1]^m\}$ in a finite spatial dimension with a constant spectral gap, the classical ML algorithm can learn to predict a classical representation of the ground state $\rho(x)$ of $H(x)$ that approximates few-body reduced-density matrices up to a constant error $\varepsilon$ when averaged over $x$. The required training data size $N$ and computation time are polynomial in $m$ and linear in the system size $n$.

Though formally efficient in the sense that $N$ scales polynomially with $m$ for any fixed approximation error $\varepsilon$, the required amount of training data scales badly with $\varepsilon$. This unfortunate scaling is not a shortcoming of the considered ML algorithm, but a necessary feature. In section S7 (*46*), we show that the data size and time complexity cannot be improved further without making stronger assumptions about the class of gapped local Hamiltonians. However, in cases of practical interest, the Hamiltonian may obey restrictions such as translational invariance or graph structure that can be exploited to obtain better results. Incorporating these restrictions can be achieved by using a suitable $\kappa(x, x_l)$, such as one that corresponds to a large-width convolutional neural network (CNN) (*48*) or a graph neural network (*49*). Rigorously establishing that neural network–based ML algorithms can achieve improved prediction performance and efficiency for particular classes of Hamiltonians requires further investigation.

### Computational hardness for non-ML algorithms

In the following proposition, we show that a classical polynomial-time algorithm that does not learn from data cannot achieve the same guarantee in estimating ground-state properties without violating the widely believed conjecture that nondeterministic polynomial-time (NP)–complete problems cannot be solved in randomized polynomial time. This proposition is a corollary of standard complexity-theoretic results (*55*, *56*). See section S8 (*46*) for the detailed statement and proof.

*Proposition 1 (informal)*

Consider a randomized polynomial-time classical algorithm $A$ that does not learn from data. Suppose for any smooth family of two-dimensional (2D) Hamiltonians $\{H(x) : x \in [-1, 1]^m\}$ with a constant spectral gap, $A$ can efficiently compute expectation values of one-body observables in the ground state $\rho(x)$ of $H(x)$ up to a constant error when averaged over $x$. Then, there is a randomized classical algorithm that can solve NP-complete problems in polynomial time.

It is instructive to observe that a classical ML algorithm with access to data can perform tasks that cannot be achieved by classical algorithms that do not have access to data. This phenomenon is studied in (*36*), where it is shown that the complexity class defined by classical algorithms that can learn from data is strictly larger than the class of classical algorithms that do not learn from data. (The data can be regarded as a restricted form of randomized advice string.) We caution that obtaining the data to train the classical ML model could be challenging. However, if we focus only on data that could be efficiently generated by quantum-mechanical processes, it is still possible that a classical ML algorithm that learns from data could be more powerful than classical computers. In section S8 (*46*), we present a contrived family of Hamiltonians that establishes this claim based on the (classical) computational hardness of factoring.

## Classifying quantum phases of matter

Classifying quantum phases of matter is another important application of ML to physics.

We will consider this classification problem in the case where quantum states are succinctly represented by their classical shadows. For simplicity, we consider the classification of two phases (denoted $A$ and $B$), but the analysis naturally generalizes to classifying any number of phases.

### ML algorithms

We envision training a classical ML algorithm with classical shadows, where each classical shadow carries a label $y$ indicating whether it represents a quantum state $\rho$ from phase $A$ [$y(\rho) = 1$] or phase $B$ [$y(\rho) = -1$]. We want to show that a suitably chosen classical ML algorithm can learn to efficiently classify the phase for new classical shadows beyond those encountered during training. Following a strategy standard in learning theory, we consider a classical ML algorithm that maps each classical shadow to a corresponding feature vector in a high-dimensional feature space and then attempts to find a hyperplane that separates feature vectors in the $A$ phase from feature vectors in the $B$ phase. The learning is efficient if the geometry of the feature space is efficiently computable and if the feature map is sufficiently expressive. Thus, our task is to construct a feature map with the desired properties.

In the simpler task of classifying symmetry-breaking phases, there is typically a local order parameter $O = \sum_i O_i$ given as a sum of $r$-body observables for some $r > 0$ that satisfies

$$tr(O\rho) \geq 1, \forall \rho \in \text{phase } A, tr(O\rho) \leq -1, \forall \rho \in \text{phase } B \quad (3)$$

Under this criterion, the classification function may be chosen to be $y(\rho) = sign[tr(O\rho)]$. Hence, classifying symmetry-breaking phases can be achieved by finding a hyperplane that separates the two phases in the high-dimensional feature space that subsumes all $r$-body reduced-density matrices of the quantum state $\rho$. The feature vector consisting of all $r$-body reduced-density matrices of the quantum state $\rho$ can be accurately reconstructed from the classical shadow representation $S_T(\rho)$ when $T$ is sufficiently large.

Finding a suitable choice of hyperplane in the feature space can be cast as a convex optimization problem known as the soft-margin support vector machine (SVM), discussed in more detail in section S10.1 (*46*). With a sufficient amount of training data, the hyperplane found by the classical ML model will generalize so that the phase $y(\rho)$ can be predicted accurately for a previously unseen quantum state $\rho$. The classical ML model is not merely a black box; it also discovers the order parameter (encoded by the hyperplane), guiding physicists toward a deeper understanding of the phase structure.

For more exotic quantum phases of matter, such as topologically ordered phases, the above classical ML model no longer suffices. The topological phase of a state is invariant under a constant-depth quantum circuit, and a phase containing the product state $|0\rangle^{\otimes n}$ is called the trivial phase. Using these notions, we can prove that no observable—not even one that acts on the entire system—can be used to distinguish between two topological phases. The proof, given in section S9 (*46*), uses the observation that random single-qubit unitaries can confuse any global or local order parameter.

*Proposition 2*

Consider two distinct topological phases $A$ and $B$ (one of the phases could be the trivial phase). No observable $O$ exists such that

$$tr(O\rho) > 0, \forall \rho \in \text{phase } A, tr(O\rho) \leq 0, \forall \rho \in \text{phase } B \quad (4)$$

Although this proposition implies that no linear function $tr(O\rho)$ can be used to classify topologically ordered phases, it does not exclude nonlinear functions, such as quadratic functions $tr(O\rho \otimes \rho)$, degree-$d$ polynomials $tr(O\rho^{\otimes d})$, and more general analytic functions. For example, it is known that the topological entanglement entropy (*57*, *58*), a nonlinear function of $\rho$, can be used to classify a wide variety of topologically ordered phases. For this purpose, it suffices to consider a subsystem whose size is large compared with the correlation length of the state but is independent of the total size of the system. The correlation length in the ground state of a local Hamiltonian increases when the spectral gap between the ground state and the first excited state becomes smaller (*59*). On the other hand, a linear function on the full system will fail even with constant correlation length.

To learn nonlinear functions, we need a more expressive ML model. For this purpose, we devise a powerful feature map that takes the classical shadow $S_T(\rho)$ of the quantum state $\rho$ to a feature vector that includes arbitrarily-large $r$-body reduced-density matrices, as well as an arbitrarily-high–degree polynomial expansion

$$\varphi^{(\text{shadow})}[S_T(\rho)] = \lim_{D,R\to\infty} \oplus_{d=0}^{D} \sqrt{\frac{\tau^d}{d!}} \left( \oplus_{r=0}^{R} \sqrt{\frac{1}{r!}\left(\frac{\gamma}{n}\right)^r} \oplus_{i_1=1}^{n} \cdots \oplus_{i_r=1}^{n} vec\left[\frac{1}{T}\sum_{t=1}^{T} \otimes_{l=1}^{r} \sigma i_l^{(t)}\right] \right)^{\otimes d} \quad (5)$$

where $\tau,\gamma > 0$ are hyperparameters. The direct sum $\oplus_{r=0}^{R}$ is a concatenation of all $r$-body reduced-density matrices, and the other direct sum $\oplus_{d=0}^{D}$ subsumes all degree-$d$ polynomial expansions. The computational cost of finding a hyperplane in feature space that separates the training data into two classes is dominated by the cost of computing inner products between feature vectors. The inner product $\langle \varphi^{(\text{shadow})}[S_T(\rho)], \varphi^{(\text{shadow})}[S_T(\tilde{\rho})] \rangle$ can be analytically computed by reorganizing the direct sums, writing it as a double series, and wrapping both series into an exponential, which gives

$$k^{(\text{shadow})}[S_T(\rho), S_T(\tilde{\rho})] = \exp\left\{ \frac{\tau}{T^2} \sum_{t,t'=1}^{T} \exp\left[ \frac{\gamma}{n} \sum_{i=1}^{n} tr\left(\sigma_i^{(t)} \tilde{\sigma}_i^{(t)}\right) \right] \right\} \quad (6)$$

where $S_T(\rho)$ and $S_T(\tilde{\rho})$ are classical shadow representations of $\rho$ and $\tilde{\rho}$, respectively. The computation time for the inner product is $O(nT^2)$, linear in the system size $n$ and quadratic in $T$, the number of copies of each quantum state that are measured to construct the classical shadow.

### Rigorous guarantee

By statistical analysis, we can establish a rigorous guarantee for the classical ML model $\langle \alpha, \varphi^{(\text{shadow})}[S_T(\rho)] \rangle$, where $\alpha$ is the trainable vector defining the classifying hyperplane. The result is the following theorem, proven in section S10 (*46*).

*Theorem 2 (classifying quantum phases of matter; informal)*

If there is a nonlinear function of few-body reduced-density matrices that classifies phases, then the classical algorithm can learn to classify these phases accurately. The required amount of training data and computation time scale polynomially in system size.

If there is an efficient procedure based on few-body reduced-density matrices for classifying phases, the proposed ML algorithm is guaranteed to find the procedure efficiently. This includes local order parameters for classifying symmetry-breaking phases and topological entanglement entropy in a sufficiently large local region for partially classifying topological phases (*57*, *58*). We expect that, to classify topological phases accurately, the classical ML model will need access to local regions that are sufficiently large compared with the correlation length, and as we approach the phase boundary, the correlation length increases. As a result, the classifying function for topological phases may depend on $r$-body subsystems with a larger $r$, and the amount of training data and computation time required would increase accordingly. The classical ML model not only classifies phases accurately but also constructs a classifying function explicitly.

Our classical ML model may also be useful for classifying and understanding symmetry-protected topological (SPT) phases. SPT phases are characterized much like topological phases but with the additional constraint that all structures involved (states, Hamiltonians, and quantum circuits) respect a particular symmetry. It

is reasonable to expect that an SPT phase can be identified by examining reduced-density matrices on constant-size regions (*60*–*65*), where the size of the region is large compared with the correlation length. The existence of classifying functions based on reduced matrices has been rigorously established in some cases (*66*–*73*). In section S12 (*46*), we prove that the ML algorithm is guaranteed to efficiently classify a class of gapped spin-1 chains in one dimension. For more general SPT phases, the ML algorithm should be able to corroborate known classification schemes, determine new and potentially more-compact classifiers, and shed light on interacting SPT phases in two or more dimensions for which complete classification schemes have not yet been firmly established.

The hypothesis of theorem 2, stating that phases can be recognized by inspecting regions of constant size independent of the total system size, is particularly plausible for gapped phases, but it might apply to some gapless phases as well. Our classical ML model would be able to efficiently classify such gapless phases. On the other hand, the contrapositive of theorem 2 asserts that if the classical ML model is not able to distinguish between two distinct gapless phases, then nonlocal data are required to characterize at least one of those phases.

## Numerical experiments

We have conducted numerical experiments assessing the performance of classical ML algorithms in some practical settings. The results demonstrate that our theoretical claims carry over to practice, with the results sometimes turning out even better than our guarantees suggest.

### *Predicting ground-state properties*

For predicting ground states, we consider classical ML models encompassed by Eq. 2. We examine various metrics $\kappa(x,x_l)$ equivalent to training neural networks with large hidden layers (*47*, *50*) or training kernel methods (*51*, *74*). We find the best ML model and the hyperparameters using a validation set to minimize root mean square error (RMSE) and report the predictions on a test set. The full details of the models and hyperparameters, as well as their comparisons, are given in sections S4.2 and S4.3 (*46*).

### *Rydberg atom chain*

Our first example is a system of trapped Rydberg atoms (*75*, *76*), a programmable and highly controlled platform for Ising-type quantum simulations (*77*–*82*). Following (*77*), we consider a 1D array of $n = 51$ atoms, with each atom effectively described as a two-level system composed of a ground state $|g\rangle$ and a highly excited Rydberg state $|r\rangle$. The atomic chain is characterized by a Hamiltonian $H(x)$ (given in Fig. 2A) whose parameters are the laser detuning $x_1 = \Delta/\Omega$ and the interaction range $x_2 = R_b/a$. The phase diagram (Fig. 2B) features a disordered phase and several broken-symmetry phases, stemming from the competition between the detuning and the Rydberg blockade (arising from the repulsive van der Waals interactions).

We trained a classical ML model using 20 randomly chosen values of the parameter $x = (x_1,x_2)$; these values are indicated by gray circles in Fig. 2B. For each such $x$, an approximation to the exact ground state was found using density matrix renormalization group (DMRG) (*6*) based on the formalism of matrix product states (MPSs) (*83*). For each MPS, we performed $T = 500$ randomized Pauli measurements to construct a classical shadow. The classical ML model then predicted classical representations at the testing points in the parameter space, and these predicted classical representations were used to estimate expectation values of local observables at the testing points.

Predictions for expectation values of Pauli operators $Z_i$ and $X_i$ at the testing points are shown in Fig. 2C and were found to agree well with exact values obtained from the DMRG computation of the ground state at the testing points. Additional predictions can be found in section S4.1 (*46*). Also shown are results from a more-naïve procedure, in which properties are predicted using only the data at the point in the training set that is closest to the testing point. The naïve procedure predicts poorly, illustrating that the considered classical ML model effectively leverages the data from multiple points in the training set.

This example corroborates our expectation that classical machines can learn to efficiently predict ground-state representations. An important caveat is that the rigorous guarantee in theorem 1 applies only when the training points and the testing points are sampled from the same phase, whereas in this example, the training data include values of $x$ from

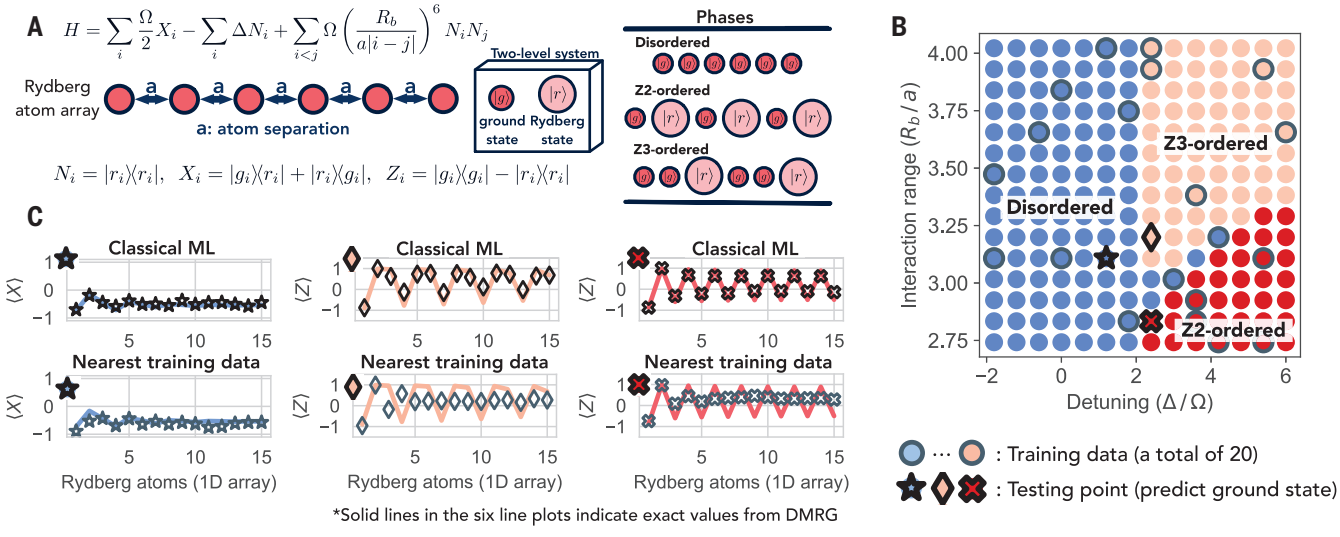



**Fig. 2. Numerical experiment for predicting ground-state properties in a 1D Rydberg atom system with 51 atoms.** (**A**) Hamiltonian and the illustrations of the Rydberg array geometry and the three resulting phases. (**B**) Phase diagram. The system's three distinct phases (*77*) are characterized by two order parameters (for $Z_2$ and $Z_3$ orders). Training data are enclosed by gray circles, and three specific testing points are indicated by the star, diamond, and cross symbols, respectively. (**C**) Local expectation values. We use classical ML (the best model is selected from a set of ML models) to predict the expectation values of Pauli operators $X_i$ and $Z_i$ for each atom at the three testing points. (Top) Results for the first 15 atoms. (Bottom) Predictions obtained from the training data nearest to the testing points. The markers denote predicted values, whereas the solid lines denote exact values obtained from DMRG. Additional predictions are shown in section S4.1 (*46*).

three different phases. Nevertheless, our numerics show that classical machines can still learn to predict well.

### 2D antiferromagnetic Heisenberg model

Our next example is the 2D antiferromagnetic Heisenberg model. Spin-½ particles (i.e., qubits) occupy sites on a square lattice, and for each pair $(ij)$ of neighboring sites, the Hamiltonian contains a term $J_{ij}(X_iX_j + Y_iY_j + Z_iZ_j)$, where the couplings $\{J_{ij}\}$ are uniformly sampled from the interval [0, 2]. The parameter $x$ is a list of all $J_{ij}$ couplings; hence, in this case, the dimension of the parameter space is $m = O(n)$, where $n$ is the number of qubits. The Hamiltonian $H(x)$ on a $5 \times 5$ lattice is shown in Fig. 3A.

We trained a classical ML model using 90 randomly chosen values of the parameter $x = \{J_{ij}\}$. For each such $x$, the exact ground state was found using DMRG, and we simulated $T = 500$ randomized Pauli measurements to construct a classical shadow. The classical ML model predicted the classical representation at new values of $x$, and we used the predicted classical representation to estimate a two-body correlation function, the expectation value of $C_{ij} = \frac{1}{3}(X_iX_j + Y_iY_j + Z_iZ_j)$, for each pair of qubits $(ij)$. In Fig. 3B, the predicted and actual values of the correlation function are displayed for a particular value of $x$, showing reasonable agreement.

Figure 3C shows the prediction performance for all pairs of spins and for variable system sizes. Each red point in the plot represents the RMSE in the correlation function estimated using our predicted classical representation for a particular pair of spins and averaged over sampled values of $x$. For comparison, each blue point is the RMSE when the correlation function is predicted using the classical shadow obtained by measuring the actual ground state $T = 500$ times. For most correlation functions, the prediction error achieved by the best classical ML model is comparable to the error achieved by measuring the actual ground state.

### Classifying quantum phases of matter

For classifying quantum phases of matter, we consider an unsupervised classical ML model that constructs an infinite-dimensional nonlinear feature vector for each quantum state $\rho$ by applying the map $\varphi^{(\mathrm{shadow})}$ in Eq. 5 with $\tau,\gamma = 1$ to the classical shadow $S_T(\rho)$ of the quantum state $\rho$. We then perform a principal components analysis (PCA) (*84*) in the infinite-dimensional nonlinear feature space. The low-dimensional subspace found by PCA in the nonlinear feature space corresponds to a nonlinear low-dimensional manifold in the original quantum state space. This method is efficient using the shadow kernel $k^{(\mathrm{shadow})}$ given in Eq. 6 and the kernel PCA procedure (*85*). Details are given in sections S4.4. and S4.5 (*46*).

### Bond-alternating XXZ model

We begin by considering the bond-alternating XXZ model with $n = 300$ spins. The Hamiltonian is given in Fig. 4A; it encompasses the bond-alternating Heisenberg model ($\delta = 1$) and the bosonic version of the Su-Schrieffer-Heeger model ($\delta = 0$) (*86*). The phase diagram in Fig. 4B is obtained by evaluating the partial reflection many-body topological invariant (*62*, *87*). There are three distinct phases: trivial, SPT, and symmetry broken.

For each value of $J$ and $\delta$ considered, we construct the exact ground state using DMRG and find its classical shadow by performing randomized Pauli measurement $T = 500$ times. We then consider a 2D principal subspace of the infinite-dimensional nonlinear feature space found by the unsupervised ML based on the shadow kernel, which is visualized in Fig. 4, C and D. We can clearly see that the different phases are well separated in the principal subspace. This shows that even without any phase labels on the training data, the ML model can classify the phases accurately. Hence, when trained with only a small amount of labeled data, the ML model will be able to correctly classify the phases as guaranteed by theorem 2.

### Distinguishing a topological phase from a trivial phase

We consider the task of distinguishing the toric code topological phase from the trivial phase in a system of $n = 200$ qubits. Figure 5A illustrates the sampled topological and trivial states. We generate representatives of the nontrivial topological phase by applying low-depth geometrically local random quantum circuits to Kitaev's toric code state (*88*) with code distance 10, and we generate representatives of the trivial phase by applying random circuits to a product state.

Randomized Pauli measurements are performed $T = 500$ times to convert the states to their classical shadows, and these classical shadows are mapped to feature vectors in the high-dimensional feature space using the feature map $\varphi^{(\mathrm{shadow})}$. Figure 5B displays a 1D projection of the feature space using the unsupervised classical ML model for various values of the circuit depth, indicating that the phases become harder to distinguish as the circuit depth increases. In Fig. 5C, we show the classification accuracy of the unsupervised classical ML model. We also compare with training CNNs that use measurement outcomes

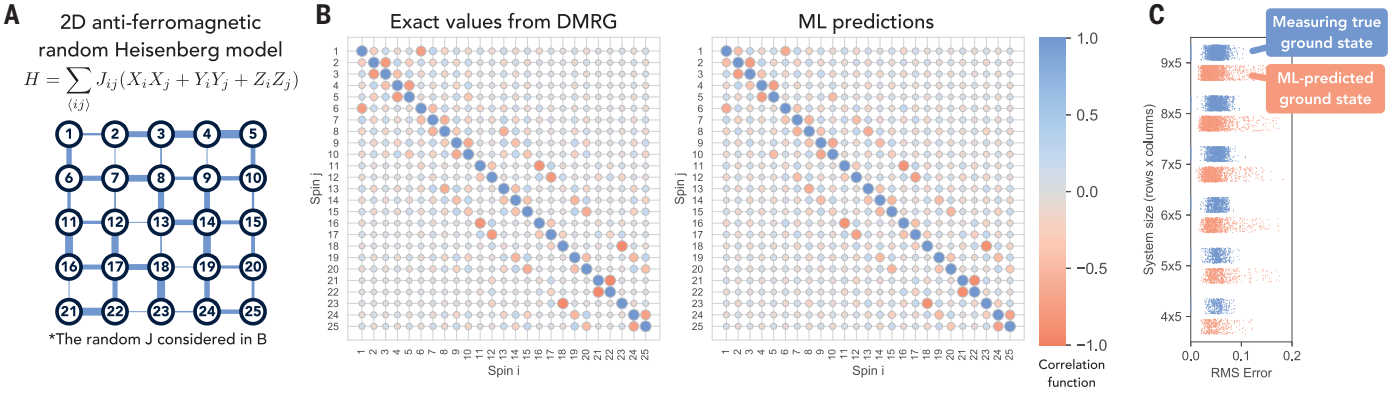



**Fig. 3. Numerical experiment for predicting ground-state properties in the 2D antiferromagnetic Heisenberg model with 25 atoms.** (**A**) Hamiltonian and the illustration of the Heisenberg model geometry. We consider random couplings $J_{ij}$, sampled uniformly from [0, 2]. A particular instance is shown with coupling strength indicated by the thickness of the edges connecting lattice points. (**B**) Two-point correlator. Exact values and ML predictions of the expectation value of the correlation function $C_{ij} = \frac{1}{3}(X_iX_j + Y_iY_j + Z_iZ_j)$ for all spin pairs $(ij)$ in the lattice, for the Hamiltonian instance shown in (A). The absolute value of $C_{ij}$ is represented by the size of each circle, and the circle's color indicates the actual value. (**C**) Prediction error. Each blue point indicates the RMSE (averaged over Heisenberg model instances) of the correlation function for a particular pair $(ij)$, where the estimate of $C_{ij}$ is obtained using a classical shadow with $T = 500$ randomized Pauli measurements of the true ground state. Red points indicate errors in ML predictions for $C_{ij}$.

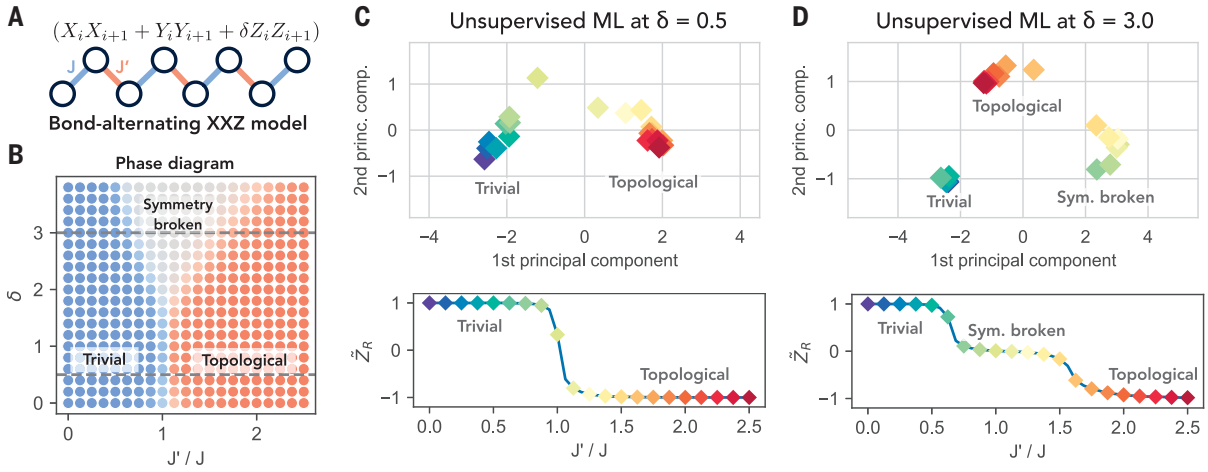



**Fig. 4. Numerical experiments for classifying quantum phases in the bond-alternating XXZ model.** (**A**) Illustration of the model—a 1D qubit chain, where the coefficient of $(X_iX_{i+1} + Y_iY_{i+1} + \delta Z_iZ_{i+1})$ alternates between $J$ and $J'$. (**B**) Phase diagram. The system's three distinct phases are characterized by the many-body topological invariant $\tilde{Z}_R$, discussed in (*62*, *87*). Blue denotes $\tilde{Z}_R = 1$, red denotes $\tilde{Z}_R = -1$, and gray denotes $\tilde{Z}_R \approx 0$. (**C** and **D**) Unsupervised phase classification. (Bottom) $\tilde{Z}_R$ versus $J/J'$ at cross sections $\delta = 0.5$ (C) and $\delta = 3.0$ (D) of the phase diagram. (Top) Visualization of the quantum states projected to two dimensions using the unsupervised ML (nonlinear PCA with shadow kernel). In all panels, the colors of the points indicate the value of $J/J'$; the upper panels suggest that the phases naturally cluster in the expressive feature space.

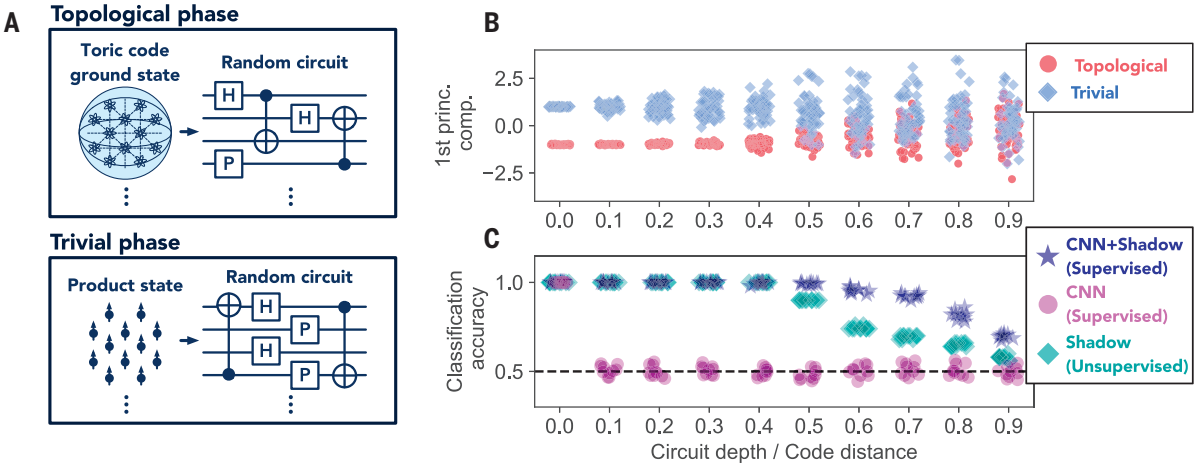



**Fig. 5. Numerical experiments for distinguishing between trivial and topological phases.** (**A**) State generation. Trivial or topological states are generated by applying local random quantum circuits of some circuit depth to a product state or exactly solved topological state, respectively. (**B**) Unsupervised phase classification. Visualization of the quantum states projected to one dimension using the unsupervised ML (nonlinear PCA with shadow kernel), shown for varying circuit depth (divided by the code distance 10, which quantifies the depth at which the topological properties are washed out). The feature space is sufficiently expressive to resolve the phases for a small enough depth without training, with classification becoming more difficult as the depth increases. (**C**) Classification accuracy for three ML algorithms described in the text.

from the Pauli-6 positive operator-valued measure (POVM) (*89*) as input to learn an observable for classifying the phases. Because proposition 2 establishes that no observable (even a global one) can classify topological phases, this CNN approach is doomed to fail. On the other hand, if the CNN takes classical shadow representations as input, then it can learn nonlinear functions and successfully classify the phases.

## Outlook

We have rigorously established that classical ML algorithms, informed by data collected in physical experiments or using classical calculations, can effectively address some quantum many-body problems. These results boost our hopes that classical ML trained on experimental data can solve practical problems in chemistry and materials science that would be too hard to solve using classical processing alone.

Our arguments build on the concept of a classical shadow derived from randomized Pauli measurements. We expect, though, that other succinct classical representations of quantum states could be exploited by classical ML with similarly powerful results. For example, some currently available quantum simulators are highly programmable but lack the local control needed to perform arbitrary single-qubit Pauli measurements. Instead, after preparing a many-body quantum state of interest, one might switch rapidly to a different Hamiltonian and then allow the state to evolve for a short time before performing a computational basis measurement. How can we make use of that measurement data to predict properties reliably (*90*, *91*)? For that matter, might we be able to generalize from experimental

data that are already routinely available to predict properties of chemical compounds and materials that have not yet been synthesized? Answering such questions will be important goals for future research.

### Materials and methods summary

Here, we provide the key ideas for designing ML algorithms to predict ground states and to classify quantum phases of matter. We refer the readers to the supplementary materials (*46*) for algorithmic details and the proofs of the main theorems.

#### *Predicting ground states*

To understand why the ML algorithm works, we begin by considering a simpler task: training an ML model to predict a single ground-state property $tr(O\ \rho)$, where $O$ is an observable and $\rho$ is the ground state. In this simpler task, the training data are $\{x_l \to tr[O\rho(x_l)]\}_{l=1}^{N}$, where $x_l \in [-1,1]^m$ is a classical description of the Hamiltonian $H(x_l)$ and $\rho(x_l)$ is the ground state of $H(x_l)$. Intuitively, in a quantum phase of matter, the ground-state property $tr[O\ \rho(x)]$ changes smoothly as a function of the input parameter $x$. The smoothness condition can be rigorously established as an upper bound on the average magnitude of the gradient of $tr[O\ \rho(x)]$ using quasi-adiabatic evolution (*53*, *54*), assuming that the spectral gap of $H(x)$ is bounded below by a nonzero constant throughout the parameter space. The upper bound on the average gradient magnitude enables us to design a simple classical ML model based on an $l_2$-Dirichlet kernel for generalizing from the training set to a new input $x \in [-1,1]^m$

$$\hat{O}(x) = \frac{1}{N}\sum_{l=1}^{N} \kappa(x, x_l)\, tr[O\rho(x_l)] \qquad (7)$$

where $\kappa(x, x_l) = \sum_{k \in Z^m, \|k\|_2 \le \Lambda} \cos[\pi k \cdot (x - x_l)]$ is the $l_2$-Dirichlet kernel with cutoff $\Lambda$. Using statistical analysis, we can guarantee that the prediction error is small given a number of training data $N$ polynomial in the number of parameters $m$.

The main idea of the statistical analysis is to bound the model complexity. In particular, the model complexity depends on the number of wave vectors $k$ in the $l_2$-Dirichlet kernel. The more wave vectors $k$ that we include, the higher the model complexity and the more data needed in the ML model to achieve good prediction performance. We show that the number of $m$-dimensional wave vectors with a Euclidean norm bounded by $\Lambda$ is $m^{O(\Lambda)}$, and we only need to consider $\Lambda$ to be of order $\sqrt{1/\varepsilon}$ to achieve prediction error at most $\varepsilon$.

We then generalize this idea to the task of predicting the ground-state representation. We consider a training data $\{x_l \to \sigma_T[\rho(x_l)]\}_{l=1}^{N}$, where $\sigma_T[\rho(x_l)]$ is the classical shadow representation of the quantum state $\rho(x_l)$ obtained from performing randomized Pauli measurement on the state $\rho(x_l)$. Following the expression for predicting a fixed property, the predicted ground-state representation is given by

$$\hat{\sigma}(x) = \frac{1}{N}\sum_{l=1}^{N} \kappa(x, x_l)\sigma_T[\rho(x_l)] \qquad (8)$$

Using the property of classical shadows, we have $tr\{O\sigma_T[\rho(x_l)]\} \approx tr[O\rho(x_l)]$ for a wide range of observables $O$. By moving the sum outside of the trace, we can reduce the problem to predicting a fixed ground-state property. Hence, if the classical ML model based on an $l_2$-Dirichlet kernel can predict ground-state properties accurately, then it can predict the ground-state representation accurately.

#### *Classifying quantum phases of matter*

The ML algorithm is based on the SVM model. The underlying idea of SVM is simple and intuitive. Suppose that we have $N$ data points that form two well-separated clusters. We may try to separate these training clusters with a linear hyperplane. When we get a new data point, we simply check which half space it belongs to and assign the label accordingly. However, there could be many hyperplanes that separate these two training clusters. SVM considers the hyperplane that yields the largest margin, which is equivalent to maximizing the distance from each cluster to the hyperplane. Intuitively, maximizing the margin allows the hyperplane to be most robust to the sampling errors of the training data. Using statistical analysis, one can rigorously show that the bigger the margin, the better the generalization performance would be.

SVM can be enhanced using the kernel trick. When the $N$ data points cannot be separated using a linear hyperplane, we need to separate them using a more complex surface. This is achieved by mapping each data point to a high-dimensional vector space through a nonlinear mapping and looking for a linear hyperplane in the high-dimensional space. One can perform the training and prediction in the high-dimensional space by only computing inner products between two points in the high-dimensional space. The inner product is often referred to as the kernel function, and this technique of mapping to a much larger space is known as the kernel trick. In many situations, one considers the high-dimensional space to be infinite dimensional. The shadow kernel that we defined in Eq. 6 also corresponds to an infinite-dimensional vector space.

For the task of classifying quantum phases of matter, we assume that there exists a classifying function $f(\rho)$ based on a nonlinear function of the reduced-density matrices of the quantum state. More precisely, we assume that states $\rho_A$ in phase $A$ satisfy $f(\rho_A) > 1$ and states in phase $B$ satisfy $f(\rho_A) < -1$. This assumption is often satisfied when we focus on states not too close to the phase boundary. We show in the supplementary materials (*46*) that various SPT phases and topologically ordered phases do satisfy this assumption. Because the shadow kernel corresponds to an inner product in an infinite-dimensional space containing all possible nonlinear combinations of the reduced-density matrices, SVM based on the shadow kernel is able to learn the classifying function. The amount of data required to learn this classifying function depends on the margin of the hyperplane in the infinite-dimensional space, which can be shown to scale polynomially in system size.

#### *Numerical experiments*

For experiments on predicting ground-state properties, we consider the supervised ML algorithm described in Eq. 2. We examine metrics $\kappa(x, x_l) \in R$ based on Gaussian kernel, Dirichlet kernel, and neural tangent kernel (*50*). Depending on different training data sizes and the number of measurements per quantum state, we found that different kernels perform better than others. For classifying quantum phases of matter, we consider an unsupervised ML algorithm, where no labeled training data are provided. The kernel trick described above can also be applied to unsupervised ML algorithms. A standard example is kernel PCA. PCA tries to find a direction, known as the principal component, such that the data points along this direction are most separated. If the points are not well separated in any direction, then we can consider mapping all points to an infinite-dimensional space. Similar to the supervised setting, we only need to consider inner products between pairs of points in the infinite-dimensional space (kernel function) to find the principal component. Hence, we can also apply the shadow kernel to classify quantum phases of matter in an unsupervised fashion. This is what we considered in the numerical experiments shown in Fig. 4 and Fig. 5.

### REFERENCES AND NOTES


1. P. Hohenberg, W. Kohn, Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964). doi: 10.1103/PhysRev.136.B864
2. W. Kohn, Nobel lecture: Electronic structure of matter—wave functions and density functionals. *Rev. Mod. Phys.* **71**, 1253–1266 (1999). doi: 10.1103/RevModPhys.71.1253
3. D. Ceperley, B. Alder, Quantum monte carlo. *Science* **231**, 555–560 (1986). doi: 10.1126/science.231.4738.555; pmid: 17750966
4. A. W. Sandvik, Stochastic series expansion method with operator-loop update. *Phys. Rev. B* **59**, R14157–R14160 (1999). doi: 10.1103/PhysRevB.59.R14157
5. F. Becca, S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge Univ. Press, 2017).
6. S. R. White, Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.* **69**, 2863–2866 (1992). doi: 10.1103/PhysRevLett.69.2863; pmid: 10046608
7. S. R. White, Density-matrix algorithms for quantum renormalization groups. *Phys. Rev. B* **48**, 10345–10356 (1993). doi: 10.1103/PhysRevB.48.10345; pmid: 10007313

8. G. Carleo *et al*., Machine learning and the physical sciences. *Rev. Mod. Phys.* **91**, 045002 (2019). doi: 10.1103/RevModPhys.91.045002
9. J. Carrasquilla, Machine learning for quantum matter. *Adv. Phys. X* **5**, 1797528 (2020). doi: 10.1080/23746149.2020.1797528
10. D.-L. Deng, X. Li, S. Das Sarma, Machine learning topological states. *Phys. Rev. B* **96**, 195145 (2017). doi: 10.1103/PhysRevB.96.195145
11. J. Carrasquilla, R. G. Melko, Machine learning phases of matter. *Nat. Phys.* **13**, 431–434 (2017). doi: 10.1038/nphys4035
12. G. Carleo, M. Troyer, Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017). doi: 10.1126/science.aag2302; pmid: 28183973
13. G. Torlai, R. G. Melko, Learning thermodynamics with Boltzmann machines. *Phys. Rev. B* **94**, 165134 (2016). doi: 10.1103/PhysRevB.94.165134
14. Y. Nomura, A. S. Darmawan, Y. Yamaji, M. Imada, Restricted boltzmann machine learning for solving strongly correlated quantum systems. *Phys. Rev. B* **96**, 205152 (2017). doi: 10.1103/PhysRevB.96.205152
15. E. P. L. van Nieuwenburg, Y.-H. Liu, S. D. Huber, Learning phase transitions by confusion. *Nat. Phys.* **13**, 435–439 (2017). doi: 10.1038/nphys4037
16. L. Wang, Discovering phase transitions with unsupervised learning. *Phys. Rev. B* **94**, 195105 (2016). doi: 10.1103/PhysRevB.94.195105
17. J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, G. E. Dahl, in *Proceedings of the 34th International Conference on Machine Learning*, D. Precup, Y. W. Teh, Eds. (PMLR, 2017), pp. 1263–1272.
18. G. Torlai *et al*., Neural-network quantum state tomography. *Nat. Phys.* **14**, 447–450 (2018). doi: 10.1038/s41567-018-0048-5
19. R. A. Vargas-Hernández, J. Sous, M. Berciu, R. V. Krems, Extrapolating quantum observables with machine learning: Inferring multiple phase transitions from properties of a single phase. *Phys. Rev. Lett.* **121**, 255702 (2018). doi: 10.1103/PhysRevLett.121.255702; pmid: 30608785
20. K. T. Schütt, M. Gastegger, A. Tkatchenko, K.-R. Müller, R. J. Maurer, Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **10**, 5024 (2019). doi: 10.1038/s41467-019-12875-2; pmid: 31729373
21. I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, J. I. Cirac, Neural-network quantum states, string-bond states, and chiral topological states. *Phys. Rev. X* **8**, 011006 (2018). doi: 10.1103/PhysRevX.8.011006
22. J. F. Rodriguez-Nieva, M. S. Scheurer, Identifying topological order through unsupervised machine learning. *Nat. Phys.* **15**, 790–795 (2019). doi: 10.1038/s41567-019-0512-x
23. Z. Qiao, M. Welborn, A. Anandkumar, F. R. Manby, T. F. Miller 3rd, OrbNet: Deep learning for quantum chemistry using symmetry-adapted atomic-orbital features. *J. Chem. Phys.* **153**, 124111 (2020). doi: 10.1063/5.0021955; pmid: 33003742
24. K. Choo, A. Mezzacapo, G. Carleo, Fermionic neural-network states for ab-initio electronic structure. *Nat. Commun.* **11**, 2368 (2020). doi: 10.1038/s41467-020-15724-9; pmid: 32398658
25. H. Kawai, Y. O. Nakagawa, Predicting excited states from ground state wavefunction by supervised quantum machine learning. *Mach. Learn. Sci. Technol.* **1**, 045027 (2020). doi: 10.1088/2632-2153/aba183
26. J. R. Moreno, G. Carleo, A. Georges, Deep learning the Hohenberg-Kohn maps of density functional theory. *Phys. Rev. Lett.* **125**, 076402 (2020). doi: 10.1103/PhysRevLett.125.076402; pmid: 32857556
27. K. Kottmann, P. Corboz, M. Lewenstein, A. Acín, Unsupervised mapping of phase diagrams of 2D systems from infinite projected entangled-pair states via deep anomaly detection. *SciPost Phys.* **11**, 025 (2021). doi: 10.21468/SciPostPhys.11.2.025
28. A. Bohrdt *et al*., Classifying snapshots of the doped Hubbard model with machine learning. *Nat. Phys.* **15**, 921–924 (2019). doi: 10.1038/s41567-019-0565-x
29. B. S. Rem *et al*., Identifying quantum phase transitions using artificial neural networks on experimental data. *Nat. Phys.* **15**, 917–920 (2019). doi: 10.1038/s41567-019-0554-0
30. G. Torlai *et al*., Integrating neural networks with a quantum simulator for state reconstruction. *Phys. Rev. Lett.* **123**, 230504 (2019). doi: 10.1103/PhysRevLett.123.230504; pmid: 31868463
31. H.-Y. Huang, R. Kueng, J. Preskill, Predicting many properties of a quantum system from very few measurements. *Nat. Phys.* **16**, 1050–1057 (2020). doi: 10.1038/s41567-020-0932-7
32. M. Paini, A. Kalev, D. Padilha, B. Ruck, Estimating expectation values using approximate quantum states. *Quantum* **5**, 413 (2021). doi: 10.22331/q-2021-03-16-413
33. G. Struchalin, Y. A. Zagorovskii, E. Kovlakov, S. Straupe, S. Kulik, Experimental estimation of quantum state properties from classical shadows. *PRX Quantum* **2**, 010307 (2021). doi: 10.1103/PRXQuantum.2.010307
34. A. Elben *et al*., Mixed-state entanglement from local randomized measurements. *Phys. Rev. Lett.* **125**, 200501 (2020). doi: 10.1103/PhysRevLett.125.200501; pmid: 33258654
35. J. Choi *et al*., Emergent quantum randomness and benchmarking from Hamiltonian many-body dynamics. arXiv:2103.03535 [quant-ph] (2021).
36. H.-Y. Huang *et al*., Power of data in quantum machine learning. *Nat. Commun.* **12**, 2631 (2021). doi: 10.1038/s41467-021-22539-9; pmid: 33976136
37. Y. Zhang, R. G. Melko, E.-A. Kim, Machine learning $\mathbb{Z}_2$ quantum spin liquids with quasiparticle statistics. *Phys. Rev. B* **96**, 245119 (2017). doi: 10.1103/PhysRevB.96.245119
38. Y. Zhang, P. Ginsparg, E.-A. Kim, Interpreting machine learning of topological quantum phase transitions. *Phys. Rev. Res.* **2**, 023283 (2020). doi: 10.1103/PhysRevResearch.2.023283
39. M. T. Ribeiro, S. Singh, C. Guestrin, in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, 2016), pp. 1135–1144.
40. A. Barredo Arrieta *et al*., Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf. Fusion* **58**, 82–115 (2020). doi: 10.1016/j.inffus.2019.12.012
41. N. Read, Topological phases and quasiparticle braiding. *Phys. Today* **65**, 38–43 (2012). doi: 10.1063/PT.3.1641
42. M. Ohliger, V. Nesme, J. Eisert, Efficient and feasible state tomography of quantum many-body systems. *New J. Phys.* **15**, 015024 (2013). doi: 10.1088/1367-2630/15/1/015024
43. S. J. van Enk, C. W. J. Beenakker, Measuring $\mathrm{Tr}\rho^n$ on single copies of $\rho$ using random measurements. *Phys. Rev. Lett.* **108**, 110503 (2012). doi: 10.1103/PhysRevLett.108.110503; pmid: 22540446
44. A. Elben, B. Vermersch, C. F. Roos, P. Zoller, Statistical correlations between locally randomized measurements: A toolbox for probing entanglement in many-body quantum states. *Phys. Rev. A* **99**, 052323 (2019). doi: 10.1103/PhysRevA.99.052323
45. T. J. Evans, R. Harper, S. T. Flammia, Scalable Bayesian Hamiltonian learning. arXiv:1912.07636 [quant-ph] (2019).
46. See the supplementary materials.
47. A. Jacot, F. Gabriel, C. Hongler, in *Advances in Neural Information Processing Systems*, S. Bengio *et al*, Eds. (Curran Associates, Inc., 2018), pp. 8571–8580.
48. Z. Li *et al*., Enhanced convolutional neural tangent kernels. arXiv:1911.00809 [cs.LG] (2019).
49. S. S. Du, K. Hou, B. Póczos, R. Salakhutdinov, R. Wang, K. Xu, in *Proceedings of the 33rd International Conference on Neural Information Processing Systems* (Curran Associates, Inc., 2019), pp. 5723–5733.
50. R. Novak, L. Xiao, J. Hron, J. Lee, A. A. Alemi, J. Sohl-Dickstein, S. S. Schoenholz, Neural Tangents: Fast and Easy Infinite Neural Networks in Python. arXiv:1912.02803 [stat.ML] (2020).
51. C. Cortes, V. Vapnik, Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995). doi: 10.1007/BF00994018
52. C.-C. Chang, C.-J. Lin, LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**, 1–27 (2011). doi: 10.1145/1961189.1961199
53. M. B. Hastings, X.-G. Wen, Quasiadiabatic continuation of quantum states: The stability of topological ground-state degeneracy and emergent gauge invariance. *Phys. Rev. B* **72**, 045141 (2005). doi: 10.1103/PhysRevB.72.045141
54. S. Bachmann, S. Michalakis, B. Nachtergaele, R. Sims, Automorphic equivalence within gapped phases of quantum lattice systems. *Commun. Math. Phys.* **309**, 835–871 (2012). doi: 10.1007/s00220-011-1380-0
55. D. Lichtenstein, Planar formulae and their uses. *SIAM J. Comput.* **11**, 329–343 (1982). doi: 10.1137/0211025
56. L. Valiant, V. Vazirani, NP is as easy as detecting unique solutions. *Theor. Comput. Sci.* **47**, 85–93 (1986). doi: 10.1016/0304-3975(86)90135-0
57. A. Kitaev, J. Preskill, Topological entanglement entropy. *Phys. Rev. Lett.* **96**, 110404 (2006). doi: 10.1103/PhysRevLett.96.110404; pmid: 16605802
58. M. Levin, X.-G. Wen, Detecting topological order in a ground state wave function. *Phys. Rev. Lett.* **96**, 110405 (2006). doi: 10.1103/PhysRevLett.96.110405; pmid: 16605803
59. M. B. Hastings, T. Koma, Spectral gap and exponential decay of correlations. *Commun. Math. Phys.* **265**, 781–804 (2006). doi: 10.1007/s00220-006-0030-4
60. H. Li, F. D. M. Haldane, Entanglement spectrum as a generalization of entanglement entropy: Identification of topological order in non-Abelian fractional quantum Hall effect states. *Phys. Rev. Lett.* **101**, 010504 (2008). doi: 10.1103/PhysRevLett.101.010504; pmid: 18764098
61. F. Pollmann, A. M. Turner, E. Berg, M. Oshikawa, Entanglement spectrum of a topological phase in one dimension. *Phys. Rev. B* **81**, 064439 (2010). doi: 10.1103/PhysRevB.81.064439
62. F. Pollmann, A. M. Turner, Detection of symmetry-protected topological phases in one dimension. *Phys. Rev. B* **86**, 125441 (2012). doi: 10.1103/PhysRevB.86.125441
63. J. Haegeman, D. Pérez-García, I. Cirac, N. Schuch, Order parameter for symmetry-protected phases in one dimension. *Phys. Rev. Lett.* **109**, 050402 (2012). doi: 10.1103/PhysRevLett.109.050402; pmid: 23006149
64. H. Shapourian, K. Shiozaki, S. Ryu, Many-Body Topological Invariants for Fermionic Symmetry-Protected Topological Phases. *Phys. Rev. Lett.* **118**, 216402 (2017). doi: 10.1103/PhysRevLett.118.216402; pmid: 28598638
65. H. Dehghani, Z.-P. Cian, M. Hafezi, M. Barkeshli, Extraction of the many-body Chern number from a single wave function. *Phys. Rev. B* **103**, 075102 (2021). doi: 10.1103/PhysRevB.103.075102
66. A. Y. Kitaev, Anyons in an exactly solved model and beyond. *Ann. Phys.* **321**, 2–111 (2006). doi: 10.1016/j.aop.2005.10.005
67. Y. Zhang, E.-A. Kim, Quantum loop topography for machine learning. *Phys. Rev. Lett.* **118**, 216401 (2017). doi: 10.1103/PhysRevLett.118.216401; pmid: 28598670
68. M. B. Hastings, S. Michalakis, Quantization of Hall Conductance for Interacting Electrons on a Torus. *Commun. Math. Phys.* **334**, 433–471 (2015). doi: 10.1007/s00220-014-2167-x
69. A. Kapustin, N. Sopenko, Hall conductance and the statistics of flux insertions in gapped interacting lattice systems. *J. Math. Phys.* **61**, 101901 (2020). doi: 10.1063/5.0022944
70. S. Bachmann, A. Bols, W. De Roeck, M. Fraas, A Many-Body Index for Quantum Charge Transport. *Commun. Math. Phys.* **375**, 1249–1272 (2020). doi: 10.1007/s00220-019-03537-x
71. S. Bachmann, B. Nachtergaele, On Gapped Phases with a Continuous Symmetry and Boundary Operators. *J. Stat. Phys.* **154**, 91–112 (2014). doi: 10.1007/s10955-013-0850-5
72. H. Tasaki, Topological phase transition and $\mathbb{Z}_2$ index for $S = 1$ quantum spin chains. *Phys. Rev. Lett.* **121**, 140604 (2018). doi: 10.1103/PhysRevLett.121.140604; pmid: 30339444
73. H. Tasaki, *Physics and Mathematics of Quantum Many-Body Systems*, (Springer, 2020).
74. K. P. Murphy, *Machine Learning: A Probabilistic Perspective* (MIT Press, 2012).
75. P. Fendley, K. Sengupta, S. Sachdev, Competing density-wave orders in a one-dimensional hard-boson model. *Phys. Rev. B* **69**, 075106 (2004). doi: 10.1103/PhysRevB.69.075106
76. A. Browaeys, T. Lahaye, Many-body physics with individually controlled Rydberg atoms. *Nat. Phys.* **16**, 132–142 (2020). doi: 10.1038/s41567-019-0733-z
77. H. Bernien *et al*., Probing many-body dynamics on a 51-atom quantum simulator. *Nature* **551**, 579–584 (2017). doi: 10.1038/nature24622; pmid: 29189778
78. P. Schauß *et al*., Crystallization in Ising quantum magnets. *Science* **347**, 1455–1458 (2015). doi: 10.1126/science.1258351; pmid: 25814579
79. M. Endres *et al*., Atom-by-atom assembly of defect-free one-dimensional cold atom arrays. *Science* **354**, 1024–1027 (2016). doi: 10.1126/science.aah3752; pmid: 27811284
80. H. Labuhn *et al*., Tunable two-dimensional arrays of single Rydberg atoms for realizing quantum Ising models. *Nature* **534**, 667–670 (2016). doi: 10.1038/nature18274
81. S. Ebadi *et al*., Quantum phases of matter on a 256-atom programmable quantum simulator. *Nature* **595**, 227–232 (2021). doi: 10.1038/s41586-021-03582-4; pmid: 34234334
82. P. Scholl *et al*., Quantum simulation of 2D antiferromagnets with hundreds of Rydberg atoms. *Nature* **595**, 233–238 (2021). doi: 10.1038/s41586-021-03585-1; pmid: 34234335
83. U. Schollwöck, The density-matrix renormalization group in the age of matrix product states. *Ann. Phys.* **326**, 96–192 (2011). doi: 10.1016/j.aop.2010.09.012
84. K. Pearson, LIII. On lines and planes of closest fit to systems of points in space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **2**, 559–572 (1901). doi: 10.1080/14786440109462720
85. B. Schölkopf, A. Smola, K.-R. Müller, Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput.* **10**, 1299–1319 (1998). doi: 10.1162/089976698300017467

86. W. Su, J. Schrieffer, A. J. Heeger, Solitons in polyacetylene. *Phys. Rev. Lett.* **42**, 1698–1701 (1979). doi: 10.1103/PhysRevLett.42.1698
87. A. Elben *et al*., Many-body topological invariants from randomized measurements in synthetic quantum matter. *Sci. Adv.* **6**, eaaz3666 (2020). doi: 10.1126/sciadv.aaz3666; pmid: 32300654
88. A. Y. Kitaev, Fault-tolerant quantum computation by anyons. *Ann. Phys.* **303**, 2–30 (2003). doi: 10.1016/S0003-4916(02)00018-0
89. J. Carrasquilla, G. Torlai, R. G. Melko, L. Aolita, Reconstructing quantum states with generative models. *Nat. Mach. Intell.* **1**, 155–161 (2019). doi: 10.1038/s42256-019-0028-1
90. J. S. Cotler *et al*., Emergent quantum state designs from individual many-body wavefunctions. arXiv:2103.03536 [quant-ph] (2021).
91. H.-Y. Hu, Y.-Z. You, Hamiltonian-driven shadow tomography of quantum states. *Phys. Rev. Res.* **4**, 013054 (2022). doi: 10.1103/PhysRevResearch.4.013054
92. M. Fishman, S. R. White, E. M. Stoudenmire, The Itensor Software Library for Tensor Network Calculations. arXiv:2007.14822 [cs.MS] (2020).
93. G. Torlai, M. Fishman, PastaQ.jl: A Package for Simulation, Tomography and Analysis of Quantum Computers, Github (2020); https://github.com/GTorlai/PastaQ.jl.
94. H.-Y. Huang, hsinyuan-huang/provable-ml-quantum: (Science), *Zenodo* (2022); .doi: 10.5281/zenodo.6942183
95. H.-Y. Huang, Provably efficient machine learning for quantum many-body problems, data set, *Zenodo* (2022). doi: 10.5281/zenodo.6525856

## ACKNOWLEDGMENTS

The authors thank N. Bar-Gill, J. Carrasquilla, S. Chen, Y. Chen, A. Elben, M. Fishman, M. Fraas, S. Glancy, J. Haah, F. Kueng, J. McClean, S. Michalakis, J. Taylor, Y. Su, and T. Vidick for valuable input and inspiring discussions. H.-Y.H. thanks A. Elben for providing the code on the bond-alternating XXZ model. The numerical simulations were performed on AWS EC2 computing infrastructure using the software packages Itensors (*92*) and PastaQ (*93*). V.V.A. thanks O. Albert, H. Kandratsenia and R. Kandratsenia, as well as Ta. Albert and Th. Albert for providing daycare support throughout this work. Contributions to this work by NIST, an agency of the US government, are not subject to US copyright. Any mention of commercial products does not indicate endorsement by NIST. **Funding:** H.-Y.H. is supported by the J. Yang & Family Foundation and a Google PhD fellowship. V.V.A. acknowledges funding from NSF QLCI award no. OMA-2120757. J.P. acknowledges funding from the US Department of Energy Office of Science, Office of Advanced Scientific Computing Research (DE-NA0003525 and DE-SC0020290), and the National Science Foundation (PHY-1733907). The Institute for Quantum Information and Matter is an NSF Physics Frontiers Center. **Author contributions:** H.-Y.H. and J.P. were involved in the conceptualization and planning of the project. H.-Y.H., R.K., and V.V.A. contributed to the theorems and proofs. H.-Y.H. and G.T. contributed to the numerical experiments. All authors were involved in the writing and presentation of the work. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** In addition to the data in the paper and supplemental materials, code related to this experiment is hosted by Zenodo (*94*). The data needed to reproduce figures are also hosted by Zenodo (*95*). 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.abk3333
Materials and Methods
Supplementary Text
Figs. S1 to S8
References (*96–209*)

Submitted 7 July 2021; accepted 16 August 2022
10.1126/science.abk3333

RESEARCH ARTICLE SUMMARY

NEURODEVELOPMENT

# Treating early postnatal circuit defect delays Huntington's disease onset and pathology in mice

**Barbara Yael Braz, Doris Wennagel, Leslie Ratié, Diego Alves Rodrigues de Souza, Jean Christophe Deloulme, Emmanuel L. Barbier, Alain Buisson, Fabien Lanté, Sandrine Humbert***

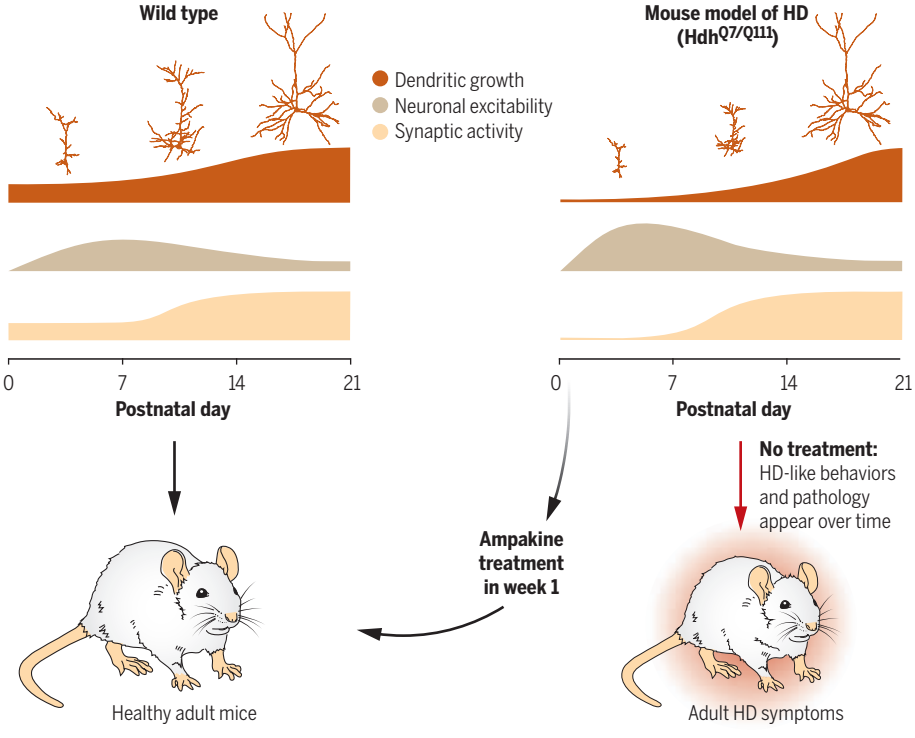
**INTRODUCTION:** Neurodegenerative diseases usually appear in middle age or later, even when caused by a genetic mutation present from conception. Huntington's disease (HD) is a case in point: it is caused by mutations in the gene encoding Huntingtin (HTT), a scaffolding protein that is particularly known for its role in transporting various molecules within cells and along neuronal axons. Although mutant HTT (mHTT) affects the earliest stages of brain development—influencing everything from the division of neuronal progenitor cells to the migration of neurons as they form the layers of the cortex—the brain is so adept at compensation that overt signs of disease do not develop for several decades. Neuropsychological testing and neuroimaging can distinguish presymptomatic HD mutation carriers, but it has remained difficult to relate the early molecular defects to the dysfunctions that develop later in life. Nonetheless, mouse studies show that either expressing mHTT or depleting wild-type (WT) HTT for only 2 to 3 weeks after birth is sufficient for the mice to later develop the hallmark features of HD pathology, suggesting that developmental anomalies contribute to the evolution of the disease. This led us to ask whether counteracting these dysfunctions when they first appear could prevent later development of the disease.

**RATIONALE:** Recognizing that the changes observed in embryos expressing mHTT affect processes that are governed by neuronal activity, we hypothesized that such changes should be accompanied by changes in cortical circuit physiology. We then sought to correct these early functional abnormalities to see if this would delay the onset of HD-related pathologies in a knock-in mouse model of HD.

**RESULTS:** We recorded the electrical activity of cortical neurons in a mouse model of HD during the first 3 weeks after birth. We found that excitatory neurons showed a transient reduction in excitatory synaptic transmission during the first postnatal week; they were more excitable and had shorter and less complex dendritic arbors than normal. The HD brain was able to rectify these defects on its own by the end of the second postnatal week. We then asked whether depletion of WT HTT contributed to these defects (HD mice bear one mutant allele and one normal allele, so the dosage of WT HTT is half of what it should be). Cortical neurons in mice depleted of WT HTT showed similar but longer-lasting defects, suggesting that loss of WT HTT is detrimental and that compensatory mechanisms in the HD context require HTT.

That HD mice go on to develop the disease even though their brains restore excitatory transmission to normal levels during the second week of life suggested that critical groundwork for adult neurophysiology is set in place within a few days after birth. We therefore sought to correct the glutamatergic defect during the first week of life. We treated HD pups with CX516—an ampakine that increases the responsiveness of AMPA receptors, which bind glutamate—and found that it restored dendritic arborization and sensorimotor function in HD pups. This neonatal CX516 treatment prevented the HD mice from developing the sensory, motor, and cognitive behavioral deficits that we observed in adult untreated mice. On the other hand, CX516 had deleterious effects on WT mice: The developing brain appears to be sensitive to too much or too little synaptic activity. Magnetic resonance imaging showed that adult HD mice had abnormal relative volumes of the striatum, cortex, hippocampus, and cerebellum, but brain morphology was normalized in CX516-treated HD mice.

**CONCLUSION:** Transient decreases or increases in perinatal circuit activity affect brain structure and function in ways that may not become apparent until the animal reaches adulthood. Our results suggest that early treatment in HD mutation carriers could change the course of the disease. ■




**Treating neonatal deficits delays adult HD.** Huntington's disease (HD) transiently alters cortical circuit function by decreasing synaptic activity, increasing excitability, and reducing the complexity of dendritic arborization. These alterations normalize in the second postnatal week but the HD mice nonetheless go on to develop HD-related behaviors and pathology unless they are treated with the ampakine CX516.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: sandrine.humbert@inserm.fr
Cite this article as B. Y. Braz *et al.*, *Science* 377, eabq5011 (2022). DOI: 10.1126/science.abq5011

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abq5011

RESEARCH ARTICLE

NEURODEVELOPMENT

# Treating early postnatal circuit defect delays Huntington's disease onset and pathology in mice

Barbara Yael Braz[1], Doris Wennagel[1], Leslie Ratié[1], Diego Alves Rodrigues de Souza[1], Jean Christophe Deloulme[1], Emmanuel L. Barbier[1], Alain Buisson[1], Fabien Lanté[1], Sandrine Humbert[1,2]*

Recent evidence has shown that even mild mutations in the Huntingtin gene that are associated with late-onset Huntington's disease (HD) disrupt various aspects of human neurodevelopment. To determine whether these seemingly subtle early defects affect adult neural function, we investigated neural circuit physiology in newborn HD mice. During the first postnatal week, HD mice have less cortical layer 2/3 excitatory synaptic activity than wild-type mice, express fewer glutamatergic receptors, and show sensorimotor deficits. The circuit self-normalizes in the second postnatal week but the mice nonetheless develop HD. Pharmacologically enhancing glutamatergic transmission during the neonatal period, however, rescues these deficits and preserves sensorimotor function, cognition, and spine and synapse density as well as brain region volume in HD adult mice.

Neural activity shapes neural development and ongoing plasticity. Electrical stimulation is transduced into a complex cascade of responses ranging from neurotransmitter release to gene transcription, axonal growth, and other functions that depend, to varying degrees, on the movement of numerous molecules along the cytoskeletal network (*1*). It is therefore not surprising that Huntingtin (HTT)—a scaffolding protein essential for intracellular transport (*2*)—is crucial not only for healthy neurophysiology (*3*–*5*) but also for neurodevelopment in both mice and humans (*6*, *7*). The importance of developmental abnormalities caused by the loss of HTT or by mutant HTT (mHTT) has been slow to be recognized, however, because this protein is so strongly associated with Huntington's disease (HD), a prototypical "late-onset" neurodegenerative disease. Embryonic defects noted in HD mouse models have often been considered curiosities with no real connection to the disease. There is compelling evidence, however, that development does influence later pathology: expressing mHTT or temporarily depleting HTT in mice up to postnatal day 21 (P21) is sufficient to ensure that the animals develop features of HD (*8*, *9*), and suggests the existence of a developmental window during which disruption of normal physiology primes the brain to eventually succumb to HD pathogenesis. Given the evidence that other apparently late-onset neurodegenerative conditions such as Alzheimer's and Parkinson's disease also have roots in early life (*10*), mapping the connections between development and degeneration seems imperative.

We previously found that cortical progenitor cell division, neural migration, dendritic maturation, and axonal growth are impaired very early in HD models (*5*, *11*–*13*). Realizing that these processes are all regulated during development by neural activity, we hypothesized that changes in synaptic transmission and excitability could be the underlying cause. We studied the electrophysiological properties of pyramidal neurons of cortical layer 2/3 in both HD knock-in pups and pups with HTT depletion, attempted to correct the deficits observed, and assessed the behavioral consequences in older mice.

## HD neurons show transient circuit physiology alterations

We first examined glutamatergic (excitatory) activity in HD knock-in mice, which bear one wild-type (WT) allele and one polyglutamine-expanded allele ($Hdh^{Q7/Q111}$). $Hdh^{Q7/Q111}$ (hereafter HD) mice express mHTT at the endogenous level, recapitulate the genetics of human HD, and provide a model of neurodevelopment that parallels processes taking place in HD human embryos (*7*). We performed whole-cell recordings of somatosensory cortex layer 2/3 neurons, which are affected by HTT dysfunction (*5*, *12*). We labeled this population with green fluorescent protein (GFP) at embryonic day 15.5, when layer 2/3 neurons are born (Fig. 1A), and recorded GFP-positive neurons in HD pups and their WT littermates (Fig. 1B) at P1 to P3 and P4 to P6 (as neurons migrate to their destined location in the brain), P7 to P10 (when migration is complete), and P21 to P26 (when neuronal morphology and electrophysiological characteristics reach a plateau) (*14*, *15*).

To estimate network activity, we measured the frequency and amplitude of spontaneous excitatory postsynaptic currents (sEPSC). We also recorded miniature EPSC (mEPSC) by bath application of tetrodotoxin (to block action potential generation) and of the $GABA_A$ antagonist Bicuculline (to isolate the contribution of glutamatergic current events). In WT neurons, sEPSC and mEPSC frequencies started low but increased by P21 (Fig. 1, C and D). Amplitudes peaked earlier, at P7 to P10 (Fig. 1, E and F). HD neurons exhibited lower sEPSC and mEPSC frequencies at early timepoints and lower mEPSC amplitude at P1 to P3. At P7 to P10, mEPSC and sEPSC frequencies in HD neurons were normalized to control levels, though with higher amplitude, possibly reflecting a compensatory response. By P21 to P26, the frequencies and amplitudes in HD neurons were indistinguishable from those of the wild type.

These transient synaptic transmission deficits suggested some alteration in the expression of the AMPA glutamate receptor, which mediates fast excitatory transmission. We analyzed cortical extracts of P2, P5, and P8 pups for the AMPA receptor subunit GluA1 and found that its expression was down-regulated in HD cortices at P2 but normalized by P8 (Fig. 1G). An additional postsynaptic marker, postsynaptic density protein 95 (PSD95), was increased at P2 and P8 (fig. S1A), perhaps reflecting a compensatory response favoring the clustering of glutamatergic receptors and increasing EPSC amplitude at P7 to P10. The presynaptic markers vesicular glutamate transporter 1 (VGLUT1) and synaptophysin had different responses: VGLUT1 was down-regulated only at P2 whereas synaptophysin was down-regulated at both P2 and P8 (fig. S1, B and C). The transient circuit alterations could thus be related to both pre- and postsynaptic mechanisms.

Next, we studied the excitability of layer 2/3 by analyzing the development of active and passive membrane properties. At P1 to P3, only ~30% of WT or HD neurons were capable of firing an action potential upon current injection (Fig. 2A), but this percentage increased with age (Fig. 2, B and D). The neurons of both genotypes became less excitable as they matured, as greater current intensities were needed to elicit spiking responses (Fig. 2, A to D). Accordingly, the rheobase (the minimal current needed to elicit an action potential) increased with age (Fig. 2E), the resting potential became more negative (Fig. 2F), and the input resistance (the change in membrane potential with a determined current pulse) fell (Fig. 2G). Concomitantly, the action potential increased in amplitude (Fig. 2H), diminished in half duration, and had a slightly more negative threshold (table S1). Despite exhibiting the same general

[1]Univ. Grenoble Alpes, Inserm, U1216, Grenoble Institut Neurosciences, 38000 Grenoble, France. [2]Institut du Cerveau-Paris Brain Institute, Sorbonne Université, Inserm, CNRS, Hôpital Pitié-Salpêtrière, Paris, France.
*Corresponding author. Email: sandrine.humbert@inserm.fr

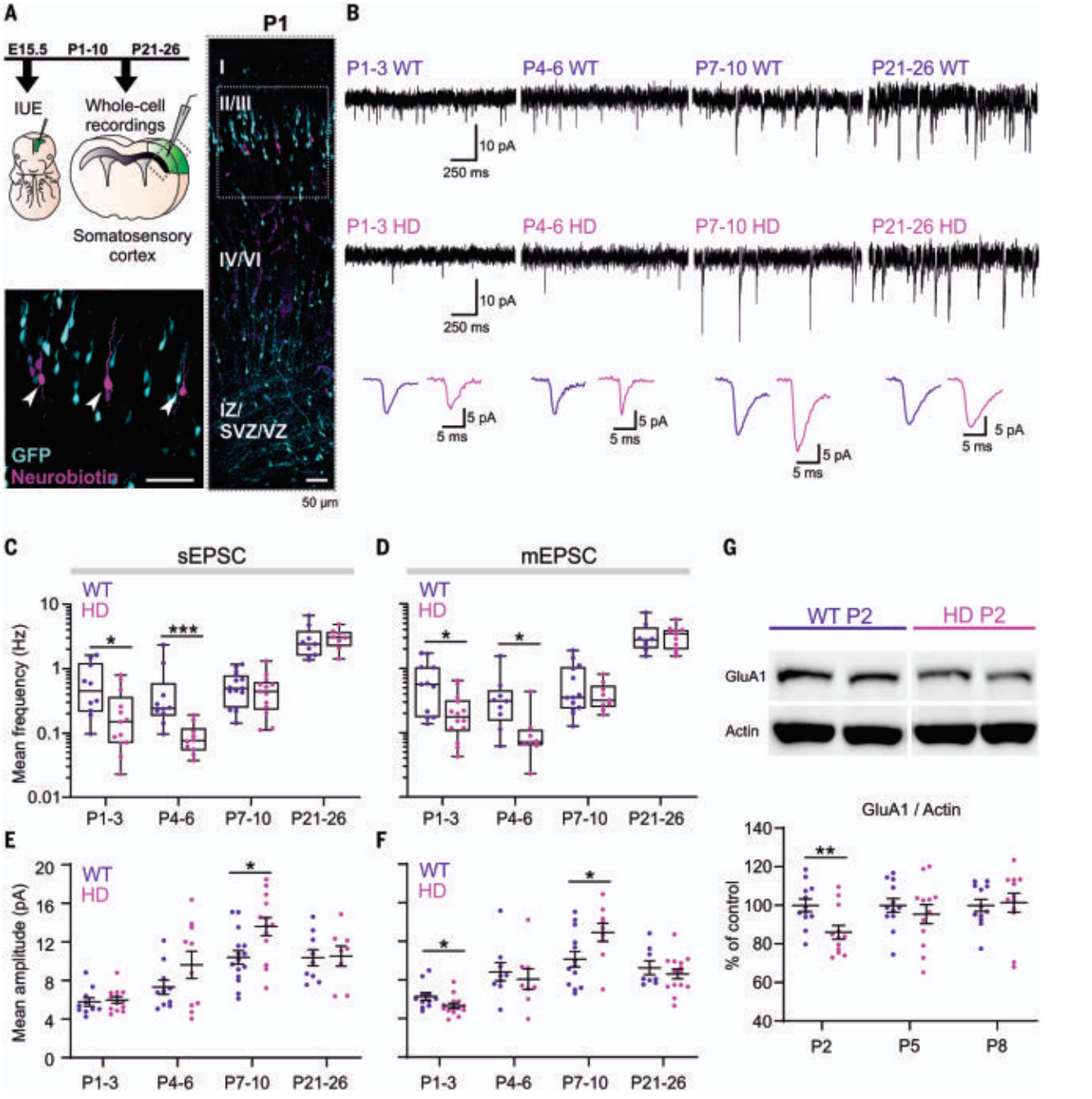



**Fig. 1. Excitatory transmission is decreased in the neonatal cortex of HD mice.** (**A**) We labeled layer 2/3 neurons by in-utero electroporation at E15.5 and performed whole-cell recordings at different postnatal stages. A representative histological section at P1 of neurobiotin-filled neurons is shown (IZ, intermediate zone; SVZ, subventricular zone; VZ, ventricular zone). Scale bars, 50 µm. (**B**) Representative recordings of miniature mEPSC (black traces) and average mEPSC (blue and magenta traces). (**C** and **D**) Mean frequency of sEPSC and mEPSC. Mann-Whitney test, $*P < 0.05$, $***P < 0.001$. (**E** and **F**) Mean amplitude of sEPSC and mEPSC. Unpaired t-test, $*P < 0.05$. (A) to (F) Per condition: at least eight neurons, from three to six animals, from at least three litters. (**G**) Quantification of GluA1/Actin intensities from immunoblotting of whole cortical protein extracts. Two independent samples from the same blot are shown per condition. Unpaired t-test, $**P < 0.01$. Per condition: 12 animals from three litters. Results are mean ± SEM or median with interquartile range.

trends as the WT, HD neurons were more excitable at P4 to P6, showing a greater spiking response (Fig. 2B) and lower rheobase (Fig. 2E). This could be related to the combination of slightly increased input resistance and hyperpolarized threshold. The transient increase in excitability in HD layer 2/3 neurons may compensate for their decreased synaptic input, in line with previous findings showing that cortical neurons adjust their excitability in an activity-dependent manner to preserve network function (*16*).

### HTT depletion creates long-lasting functional circuit changes

Although HD is usually said to be caused by a toxic gain-of-function from the mutant (expanded) allele, experiments that deplete HTT during development suggest that loss of normal

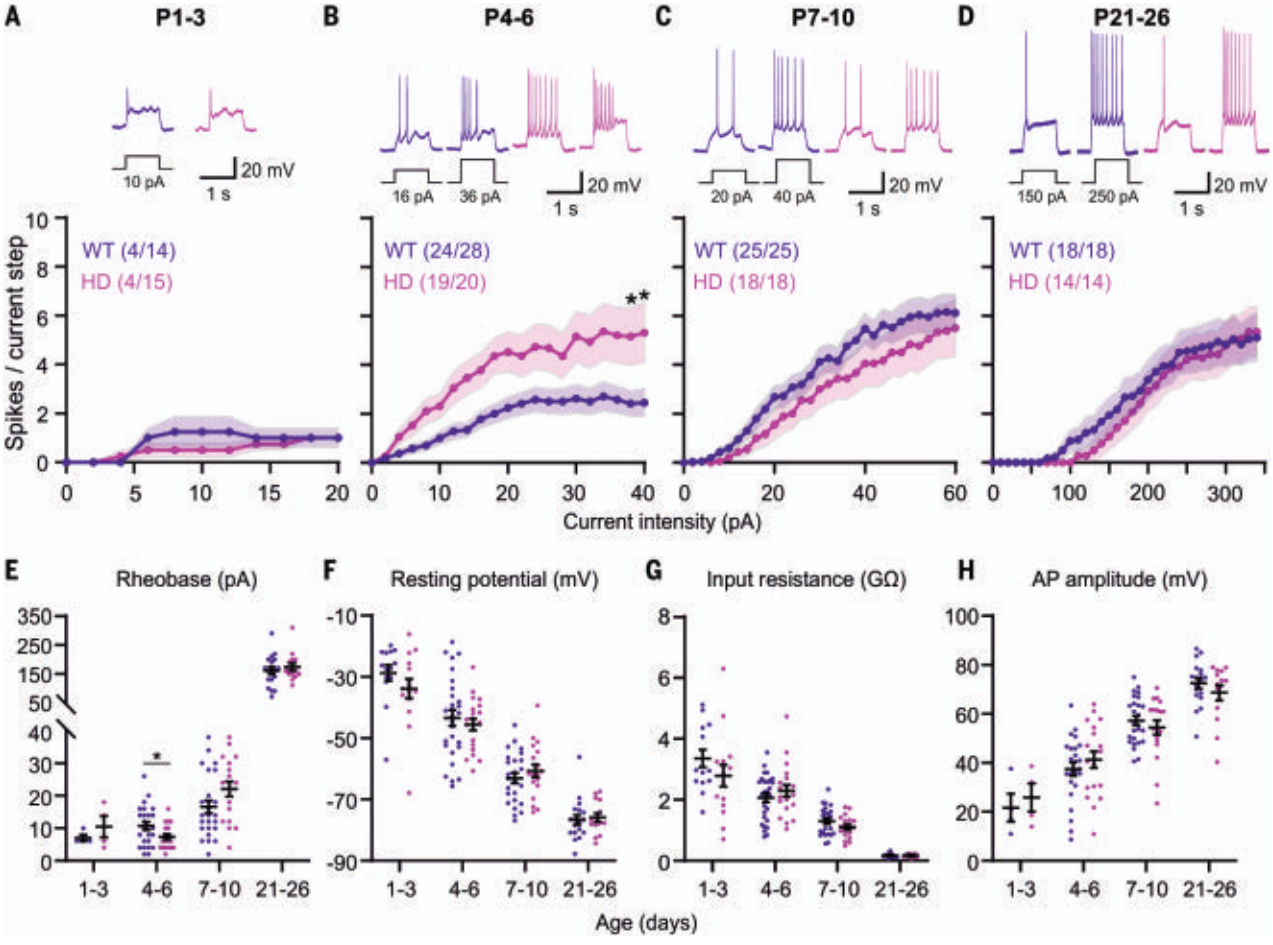



**Fig. 2. Excitability is transiently increased in HD neonatal cortical neurons.** (**A** to **D**) Action potential response to current injections of different intensity in current clamp configuration. Representative traces are shown. The number of neurons (with active response/total) is shown. Two-way repeated measures analysis of variance (ANOVA), *$P$ < 0.05 Sidak's comparisons after significant interaction. (**E** to **H**) Rheobase, resting potential, input resistance, and action potential (AP) amplitude. Unpaired t-test, *$P$ < 0.05. Results are mean ± SEM. Per condition: at least 10 neurons, from three to six animals, from at least three litters.

HTT function also contributes to HD pathophysiology (*8*). To test this possibility in the context of postnatal circuit development, we repeated the above experiments in $HTT^{lox/lox}$ mice depleted of HTT (fig. S2) with a plasmid targeting CRE-recombinase expression to postmitotic neurons by a NeuroD (ND) promoter (ND:CRE-GFP) (*12*). As controls, we used a plasmid expressing only GFP (pCAG-GFP).

HTT-depleted neurons showed abnormalities similar to those of HD neurons except that they did not normalize by the end of the first postnatal month. Their low sEPSC frequencies persisted (fig. S3, A and B); their sEPSC amplitudes were lower than controls from P7 onward (fig. S3C), and they were hyperexcitable at P7 to P10 and P21 to P26 (fig. S4, A to D), with lower rheobase, higher input resistance, lower action potential amplitude at P21 to P26, and no difference in resting potential, half duration, or threshold (fig. S4, E to H, and table S2). These results suggest that loss of WT HTT function contributes to the early circuit deficits in HD, and that the presence of some normal HTT function—through the half-dosage of WT protein, with or without the help of the partial function of the mutant (*2*)—is necessary for the circuit to be able to compensate for a time.

### Enhancing glutamatergic signaling corrects delayed dendritic maturation

Because glutamatergic transmission is involved in dendritic maturation of cortical neurons (*17*, *18*), we would expect diminished activity to affect dendritic morphology. We therefore performed three-dimensional reconstructions of neurobiotin-filled recorded neurons and traced their postnatal development (Fig. 3A). As expected, WT dendritic arbors became longer and more elaborate with maturation. At early time points HD neurons showed the same number of primary dendrites as the wild type except shorter and simpler; by P21, however, they had become indistinguishable from the controls (Fig. 3, A to C, and fig. S5, A and B). HTT-depleted neurons showed the same deficits, without the recovery seen in HD (fig. S5, C to F).

We then tested the effects of enhancing endogenous glutamatergic transmission in HD mice and their WT littermates by using the ampakine CX516, a positive AMPA allosteric modulator that increases the amplitude of excitatory synaptic responses (fig. S6, A to D). We used in-utero electroporation to label layer 2/3 neurons with GFP, then administered saline or CX516 (2.5 mg/kg, injected either once or twice per day) from P0 to P7. At P8, we fixed brains and performed three-dimensional reconstructions (Fig. 4A). CX516 restored dendritic length and complexity in HD mice (Fig. 4, B to C, and fig. S6, E to G). CX516 also increased the complexity of the distal dendritic arbor in WT animals (Fig. 4B) and increased total dendritic length in both genotypes (Fig. 4C).

### Enhancing glutamatergic transmission rescues HD sensorimotor deficits

We were interested in whether the physiological deficits observed affect behaviors that depend on the sensorimotor cortex. One of the earliest such behaviors is huddling, which

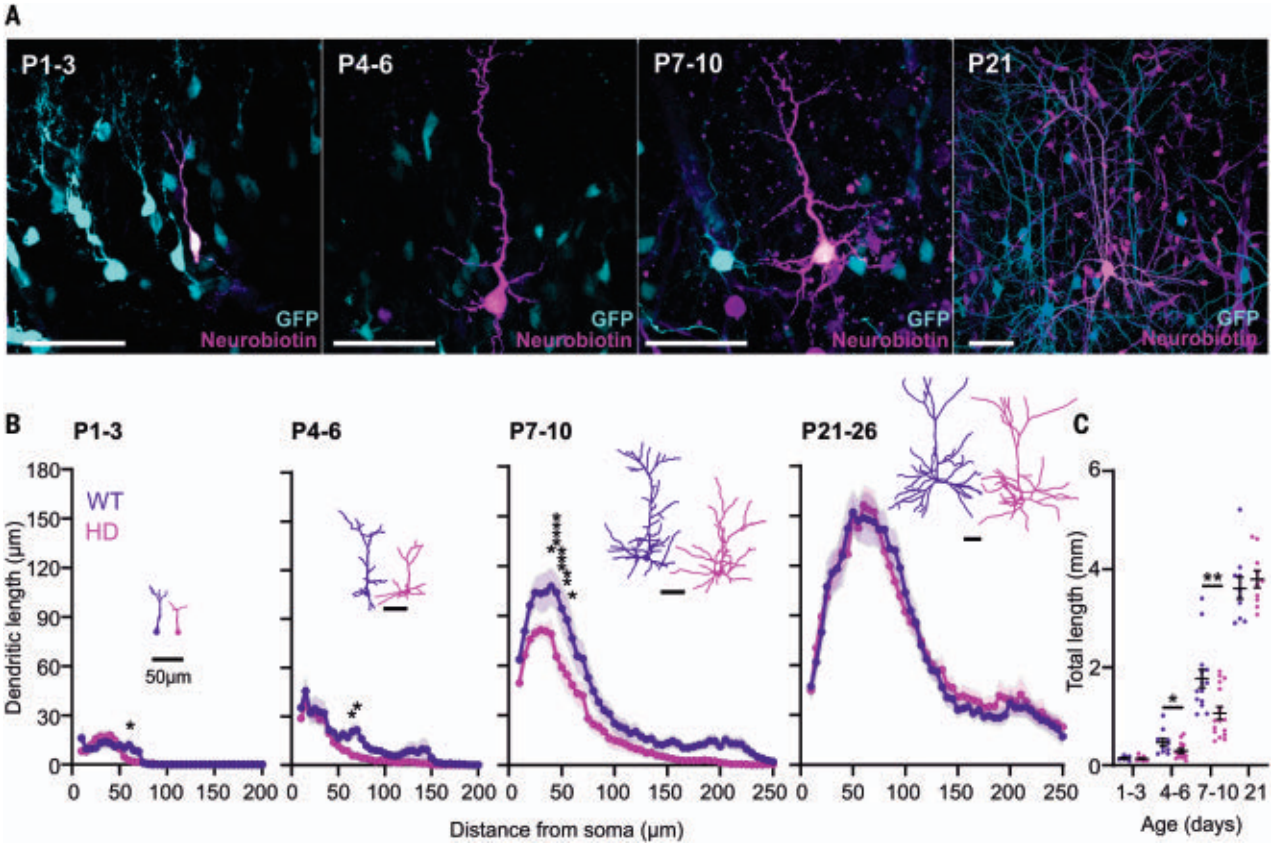



**Fig. 3. Dendritic maturation is delayed in HD neurons.** (**A**) Representative images showing GFP-positive layer 2/3 neurons filled with Neurobiotin. Scale bars, 50 µm. (**B**) Sholl analysis at different postnatal stages for WT and HD layer 2/3 neurons. Two-way repeated measures ANOVA, *$P$ < 0.05, **$P$ < 0.01, ***$P$ < 0.001, ****$P$ < 0.0001 Sidak's comparisons. (**C**) Total dendritic length. Unpaired t-test, *$P$ < 0.05, **$P$ < 0.01. Results are mean ± SEM. Per condition: at least eight neurons, from three to six animals, from at least three litters.

conserves heat and fosters affiliation (*19*). We separated a litter of 10 pups from their mother and videorecorded them in an empty arena on consecutive days (P4 to P9) (Fig. 4D). HD pups were less able than the wild type to establish huddling (Fig. 4E), particularly at P8 and P9 (fig. S7A). This deficit was not related to any motor impairment (fig. S7B). However, restoring glutamatergic transmission with CX516 treatment (2.5 mg/kg, twice daily) restored huddling behavior (Fig. 4E and fig. S7A). CX516 also increased the weight of the HD pups at P5 to P9 (fig. S7C).

We next assessed the righting reflex, which involves vestibular sensation, proprioception, touch sensation, and motor coordination (*20*). We placed each mouse on its back and recorded the time it took them to roll over and place all four paws on the surface (Fig. 4F). Because skill in the righting reflex improves as mice develop, we tested mice during three periods when they normally make noticeable gains (fig. S7, D to F). HD pups took longer to right themselves than the controls at P1 to P3 and P4 to P6, but CX516 treatment decreased this latency (Fig. 4F). Control pups treated with CX516 displayed longer righting times at all stages, suggesting that the postnatal circuit is harmed by too much or too little activity.

### Neonatal CX516 treatment delays HD onset

Next, we investigated whether restoration of glutamatergic transmission during the first postnatal week would alter the course of HD. We first assessed haptic skills and fine sensorimotor coordination in 5-week-old mice through the gap-crossing test (Fig. 5A). We used an elevated runway with a gap in the middle that is slowly widened until mice can no longer reach across it and measured the maximum distance crossed (*21*). Saline-treated HD mice and CX516-treated WT mice were impaired in this task, but CX516-treated HD mice performed similarly to saline-treated WT mice (Fig. 5B).

We then tested 2-month-old mice on two versions of the horizontal ladder test (*22*). Animals were trained to cross a ladder with a regular rung pattern and then placed on a ladder with irregularly spaced rungs (Fig. 5C). We measured the latency to cross each ladder and the number of hindlimb and forelimb misplacements (errors) (Fig. 5C and fig. S8, A and B). The irregular ladder was more challenging, with a latency ratio (irregular or regular) greater than one and the difference in the number of errors (irregular or regular) >0 in all groups. The latency ratio and the number of errors were greater in HD mice than control mice but their performance was improved by neonatal CX516 treatment.

To assess features that change over time, we considered weight and activity level. HD mice tend to weigh less than WT mice, and our HD males weighed less than the wild type at 5 weeks in the saline condition; both sexes weighed less than the wild type at 6 months (fig. S8, C and D). CX516-treated mice, however, weighed more at all ages. HD mice also tend to become hypoactive as they age (*23*), so we analyzed horizontal locomotion in an open field (Fig. 5D). At P21 there were no differences between groups in the total distance traveled or in the percentage of distance traveled in the center. As expected, a hypoactive phenotype emerged at six months in the saline-treated HD mice but not in the CX516-treated HD group. There were no differences in the distance traveled in the center, indicating the absence of an anxiety-like phenotype.

Because HD circuit physiology could be altered in cortical regions besides somatosensory

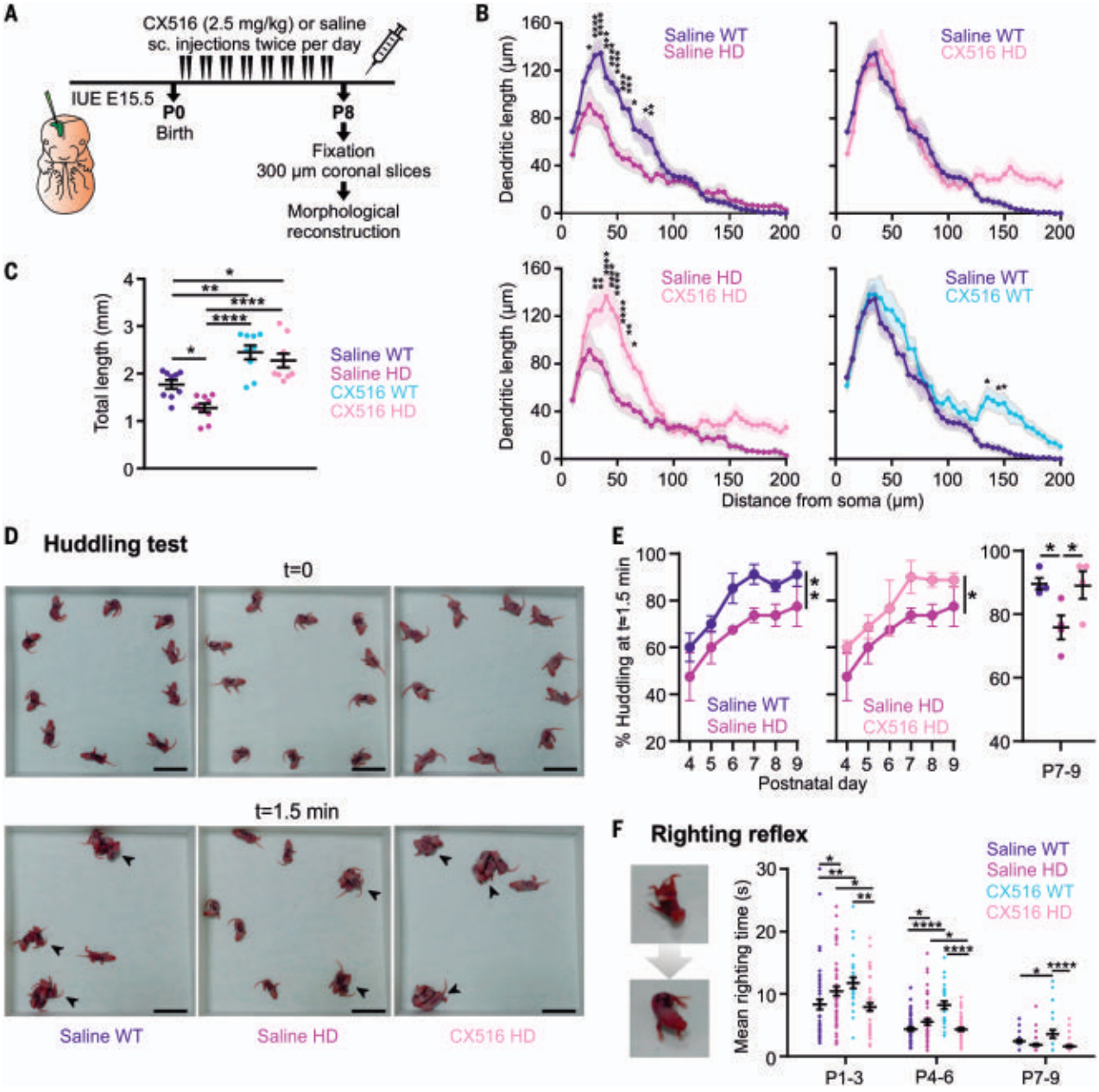



**Fig. 4. Enhancing glutamatergic transmission restores dendritic arborization and sensorimotor function in HD pups.** (**A**) Mice were electroporated at E15.5 and treated daily (P0 to P7) with CX516 (2.5 mg/kg, two subcutaneous injections per day) or saline (fig. S6 shows the results of CX516, one injection per day). At P8 brains were fixed, coronal slices were cut, and morphology reconstructed. (**B**) Sholl analysis at P8. Two-way repeated measures ANOVA, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. Sidak's comparisons. (**C**) Total dendritic length. Two-way ANOVA, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{****}P < 0.0001$. Tukey's comparisons. (B) to (C) Per condition: at least eight neurons, from three to six animals, from at least three litters. (**D**) Representative photos of P6 huddling behavior at $t = 0$ and $t = 1.5$ min. Scale bars, 5 cm. (**E**) Mean huddling percentage at $t \approx 1.5$ min at different postnatal stages. Two-way repeated measures ANOVA, $^{*}P < 0.05$ $^{**}P < 0.01$ significant effect of treatment. Right graph: One-way ANOVA, $^{*}P < 0.05$ Dunnett comparisons. (**F**) Mean righting time at the indicated periods. Two-way ANOVA $^{*}P < 0.05$, $^{**}P < 0.01$, $^{****}P < 0.0001$ Tukey's comparisons. Four litters were analyzed (except for CX516 WT in which we have two litters) with a mean size of 12 pups per litter. Results are mean ± SEM.

and motor areas, we analyzed spontaneous alternation in a Y maze (Fig. 5E). In this test, which assesses working memory and depends on prefrontal cortex function, HD mice made a lower percentage of alternations than controls at P21, and this deficit persisted at six months of age. CX516-treated WT mice exhibited a decrease in alternation at six months whereas CX516-treated HD mice did not develop this deficit even by six months of age, the last time point tested.

To determine whether these behavioral improvements in CX516-treated HD mice correlated with improvements in spine and excitatory synapse density (*24*, *25*), we stained cortical sections with the dendritic spine marker spinophilin (Fig. 6A), and the synaptic markers PSD95 and VGLUT1 (Fig. 6B) at the end of the behavioral experiments at eight months. In the saline-treated condition, HD mice showed a lower density of smaller spinophilin-immunoreactive puncta and coupled PSD95-VGLUT1 puncta.

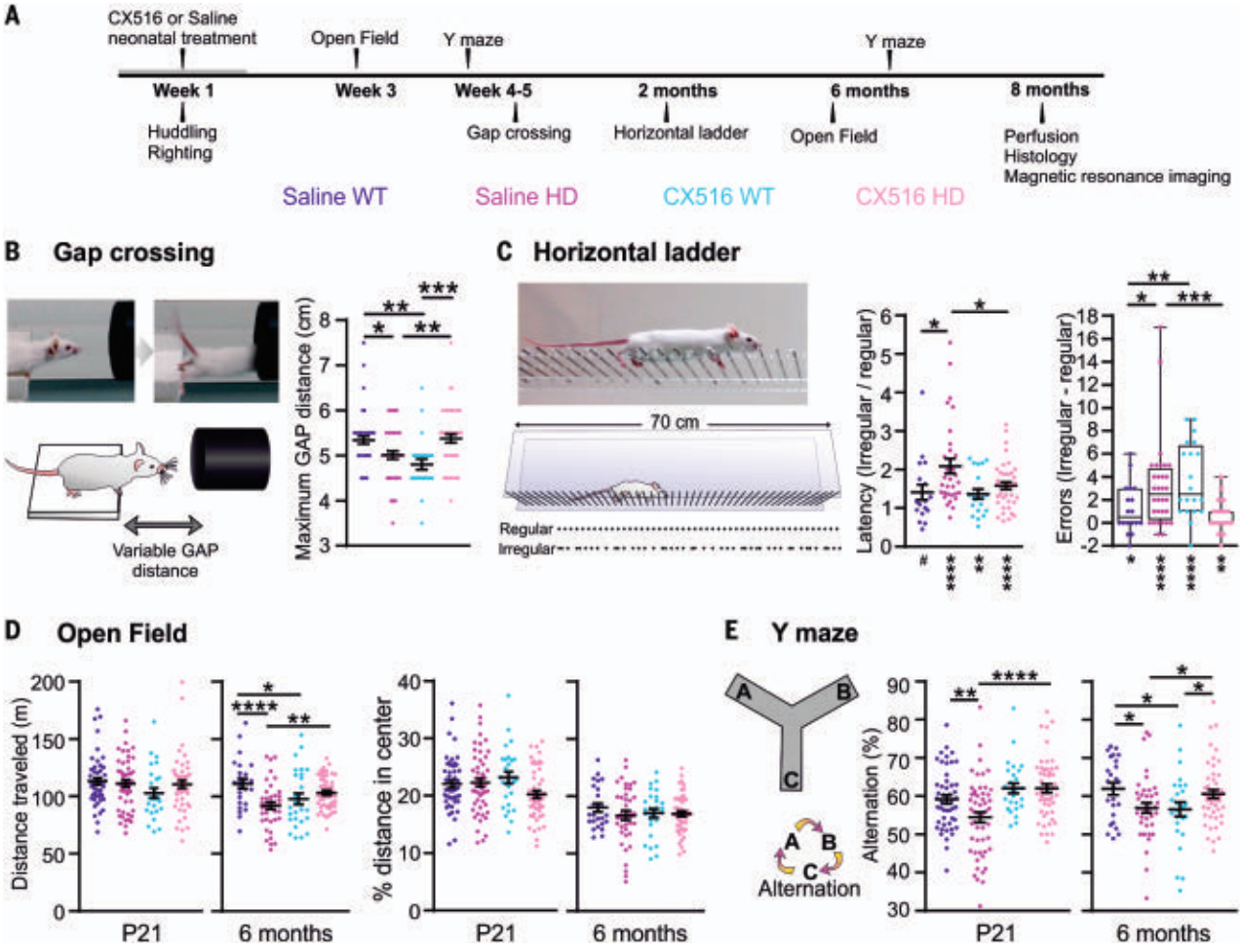



**Fig. 5. Neonatal CX516 treatment delays the emergence of HD behavioral deficits.** (**A**) Timeline of treatments with saline and CX516 (2.5 mg/kg, two injections per day), behavioral tests and histological analyses. (**B**) Gap crossing: maximum distance crossed. Two-way ANOVA, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$ Sidak's comparisons. (**C**) Horizontal ladder: ratio of latencies (time to cross) for irregular and regular ladders [(left) higher than one, one-sample t-test, $\#P = 0.05$, $**P < 0.01$, $****P < 0.0001$], and difference in number of errors made in these two conditions [(right) higher than 0, Wilcoxon signed rank test, $*P < 0.05$, $**P < 0.01$, $****P < 0.0001$]. Latency: Two-way ANOVA, $*P < 0.05$ Sidak's comparisons; Errors: Kruskal-Wallis test, followed by the Mann-Whitney test $*P < 0.05$, $**P < 0.01$ $***P < 0.0001$. (**D**) Open field: total distance traveled (left) and percentage of distance traveled in the center (right). Two-way ANOVA, $*P < 0.05$, $**P < 0.01$, $****P < 0.0001$ Tukey's comparisons. (**E**) Y maze: percentage of alternation. Two-way ANOVA, $*P < 0.05$, $**P < 0.01$, $****P < 0.0001$, Tukey's comparisons.

Neonatal CX516 treatment rescued these deficits in HD mice but caused deficits in spine and synapse density in WT mice. Next, we used immunohistochemistry to analyze the levels of the dopamine and cAMP-regulated 32-kDa phosphoprotein DARPP-32, a marker of striatal neuronal dysfunction in HD mice (*26*). HD mice had low DARPP-32 levels that were restored by CX516 treatment (Fig. 6C). CX516-treated WT mice also had low DARPP-32 levels.

To determine whether CX516 treatment improved HD brain morphology (*26*), we performed magnetic resonance imaging (MRI) at eight months (Fig. 6D). Even though HD mice showed greater total brain volume at this stage, the volumes of the striatum, cortex, and hippocampus were relatively low (as a percentage of the total volume) whereas the cerebellum was enlarged. CX516 treatment normalized all of these alterations in HD mice.

## Discussion

Transient decreases (*27*) or increases (*28*) in perinatal circuit activity affect adult functional connectivity and behavior and have been proposed to underlie the subsequent manifestation of early-onset neurological diseases such as schizophrenia, epilepsy, and autism spectrum disorders (*29*). The possibility that developmental abnormalities set the stage for diseases that declare themselves late in life has been met with some resistance even though many of the genes and risk factors associated with these diseases regulate brain development. It is difficult to capture early changes for sporadic diseases such as Alzheimer's or Parkinson's disease, though some associations—such as educational attainment and toxic exposures—have been identified (*10*). Our results establish that disturbances in early development in HD are central to the eventual development of HD in adulthood. The observation that ampakine (CX516) treatment benefitted HD mice while impairing WT mice confirms how delicate the postnatal circuit is and how finely tuned its activity must be.

Although abnormalities in glutamatergic transmission and intrinsic excitability have been found in the cortex and striatum in various HD mouse models (*23*, *26*, *30*–*32*), there has been no direct evidence that perinatal circuit development is delayed. Our finding that GluA1 and VGLUT1 expression is reduced further supports glutamatergic dysregulation in HD (*23*). The efficacy of the ampakine treatment in restoring dendritic tree maturation and sensorimotor behavior in the HD knock-in pups supports the notion that low glutamatergic activity contributes to defective development of dendritic arbors and physiological function of HD cortical neurons. Nonetheless, mHTT

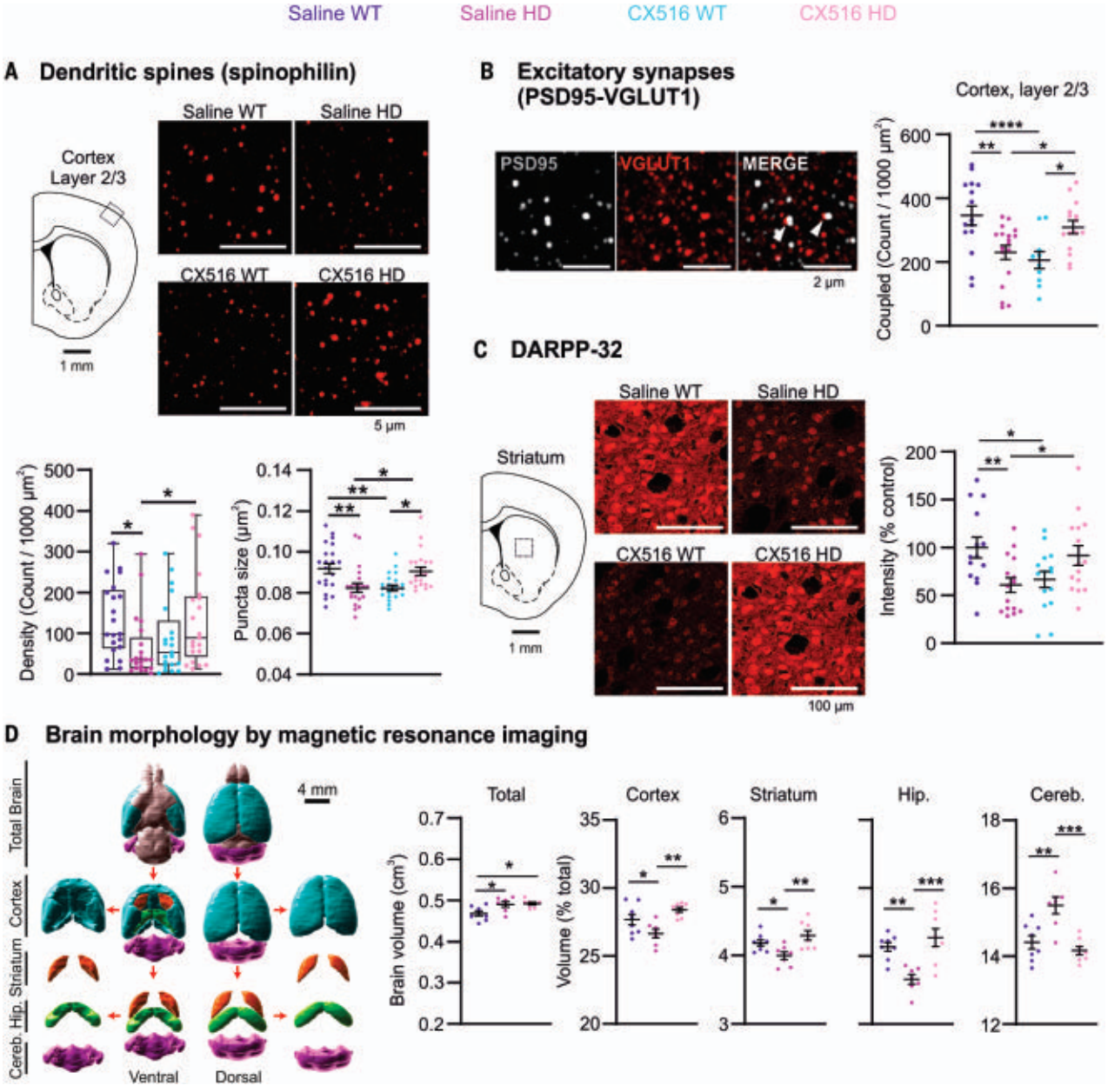



**Fig. 6. Neonatal CX516 treatment delays HD histopathological features and restores brain region volumetrics in HD mice.** (**A**) Spine density: Spinophilin-immunoreactive punta density (Kruskal-Wallis test, *$P$ < 0.05 Dunn's comparisons) and size (Two-way ANOVA, *$P$ < 0.05, **$P$ < 0.01 Sidak's comparisons). (**B**) Density of excitatory synapses: coupled PSD95-VGLUT1 dots (Two-way ANOVA, *$P$ < 0.05, **$P$ < 0.01, ****$P$ < 0.0001 Tukey's comparisons). (**C**) DARPP-32 striatal staining (Two-way ANOVA, *$P$ < 0.05, **$P$ < 0.01 Tukey's comparisons). (A) to (C) Per condition: five to seven animals per group, one to three sections per animal. (**D**) Total brain volume and volume (percentage of total) of the cortex, striatum, hippocampus, and cerebellum (one-way ANOVA, *$P$ < 0.05, **$P$ < 0.01, ***$P$ < 0.001 Holm-Sidak's comparisons, seven to eight animals per group). Representative images of all labeling and brain reconstruction are shown.

might impair dendritic maturation by other mechanisms, as it also disrupts AMPA receptor trafficking and surface diffusion (*33*, *34*).

The literature on ampakine effects on synaptic transmission and memory (*35*) encompasses both animal and clinical studies for the treatment of neurological conditions such as schizophrenia and autism spectrum disorders (*36*, *37*). By facilitating long-term potentiation, ampakines promote long-term reference memory as well as short-term and working memory in control (*36*, *38*) and HD animals (*39*). Alterations in synaptic plasticity have been documented in HD mice in corticostriatal and hippocampal circuits (*25*, *39*). In adult HD animals, ampakine treatment up-regulates brain-derived neurotrophic factor (BDNF), mitigating some deficits in memory and plasticity (*39*). All of these studies focused on the acute actions of ampakines in adult animals, but we provide evidence that ampakine treatment in the first postnatal week exerts long-lasting benefits on locomotion, sensorimotor skills, cognition, neuropathology, and brain morphology in adult HD animals. This reinforces the notion that in HD, targeting the right period during development can alter the course of the disease.

Our results have three major implications for approaches to therapy in human HD. Most research has focused on reducing the accumulation of mHTT without regard for the fact that stopping the expression of mHTT by P21 is not enough to prevent later-onset disease (*9*). The

current study supports the notion that rearguard efforts to manage late disease stages may prove too little, too late. On the other hand, optimism about enhancing compensatory changes may be equally unjustified: The circuit abnormality corrected by ampakine treatment in week 1 was corrected by the HD brain on its own by week 2 in untreated mice, but these untreated mice went on to develop the features of HD. This seems to indicate that endogenous compensatory mechanisms are also rearguard efforts by the diseased brain and that, as such, even if they succeed temporarily, they may either come too late or create further pathology. A much deeper understanding is needed regarding the effects of HTT and mHTT in the developing brain as well as the interplay between compensatory and pathogenic mechanisms taking place during the prodromal phases of HD. Finally, our observation that depleting WT HTT produced even more enduring deficits than mHTT suggests that treatments that reduce levels of both mHTT and WT HTT would be deleterious. Nevertheless, this work raises the possibility that treating human HD gene carriers during postnatal development could change the course of the disease.

## Methods summary

Experimental models: We used $Hdh^{Q111/Q7}$ knock-in mice to model HD (*7*). For studies involving HTT depletion we used $HTT^{flox/flox}$ mice (*12*).

In-utero electroporation (IUE): Performed at E15.5 according to published protocols (*12*). We used ND:CRE-GFP and Lox-GFP plasmids (*12*) to label layer 2/3 cortical neurons in HD and mice and their WT littermates, and to deplete HTT in $HTT^{flox/flox}$ mice. As an additional control, we used $HTT^{flox/flox}$ mice electroporated with pCAG-GFP.

Acute slice preparation and whole-cell recordings: At P1 to P10, 300 μm-thick coronal slices were obtained in artificial cerebrospinal fluid (ACSF) and allowed to recover 60 minutes at room temperature (RT). ACSF consisted of (in mM) 119 NaCl, 2.5 KCl, 26 $NaHCO_3$, 1.25 $NaH_2PO_4$, 1.3 $MgSO_4$, 2.5 $CaCl_2$, 11 D-glucose. At P21-26, 300 μm-thick sagittal slices cut in cutting solution 2.5 KCl, 26 $NaHCO_3$, 1.25 $NaH2PO_4$, 10 $MgSO_4$, 0.5 $CaCl_2$, 11 D-glucose, and 234 Sucrose, and allowed to recover in ACSF for 30 to 40 min at 37°C and then 30 min at RT. All solutions were equilibrated with $O_2/CO_2$ (95:5 mixture). Spontaneous (ACSF) and miniature excitatory postsynaptic currents (50 μM Bicuculline, 1 μM TTX) were recorded at −60 mV, filtered at 1 kHz and analyzed with Clampfit11.1. Active and passive properties were recorded in ACSF and calculated for the first action potential at the rheobase.

Pharmacological experiments with CX516: We administered CX516 (2.5 mg/kg) or saline solution subcutaneously to $Hdh^{Q7/Q111}$ pups and their WT littermates from P0 during eight days. Pups were used for dendritic tree reconstructions (with IUE) and behavioral analyses (without IUE).

Histology, immunohistochemistry, and dendritic tree reconstructions: Pups and adult mice were anesthetized with sodium pentobarbital (360 mg/kg) and perfused with PBS followed by 4% paraformaldehyde (PFA). Brains were dissected and postfixed overnight. 30 μm-thick coronal sections were obtained in a cryostat and stained for spinophilin, PSD95-VGLUT1 and DARPP-32. 300 μm-thick sections were obtained from pups (treated with CX516 or saline) using a vibratome. At the end of electrophysiological recordings, the slices were post-fixed overnight in PFA 4% and stained for Neurobiotin. Dendritic trees were acquired on a Zeiss LSM710 confocal microscope and 3D reconstructions were made in NeuronStudio.

Magnetic resonance imaging: adult mice were perfused with 4% PFA containing 6.25 mm of Gd-DOTA (contrast agent) the skull was exposed by removing the skin and muscles, postfixed and transferred to a Fomblin. Magnetic resonance imaging was performed at 9.4 T (IRMaGe facility, Grenoble). Brain volume was analyzed with Fiji software.

Behavioral experiments: All tests were video-recorded, and some tests used automated tracking (EthoVision). Our behavioral battery in pups included the huddling test (*19*), righting reflex (placed each pup on its back and measured the time to right itself) and locomotor activity in an open field (total distance traveled). In young and adult mice, we performed the gap-crossing test (*21*) analyzing the maximum gap distance crossed, the horizontal ladder (*22*) with a regular and irregular pattern (latency to cross and number of errors), open field (total distance traveled, distance in the center) and Y-maze (percentage of alternation).

## REFERENCES AND NOTES


1. Y. Bando, M. Ishibashi, S. Yamagishi, A. Fukuda, K. Sato, Orchestration of Ion Channels and Transporters in Neocortical Development and Neurological Disorders. *Front. Neurosci.* **16**, 827284 (2022). doi: 10.3389/fnins.2022.827284; pmid: 35237124
2. F. Saudou, S. Humbert, The Biology of Huntingtin. *Neuron* **89**, 910–926 (2016). doi: 10.1016/j.neuron.2016.02.003; pmid: 26938440
3. A. D. Lawrence *et al*., Evidence for specific cognitive deficits in preclinical Huntington's disease. *Brain* **121**, 1329–1341 (1998). doi: 10.1093/brain/121.7.1329; pmid: 9679784
4. S. Schippling *et al*., Abnormal motor cortex excitability in preclinical and very early Huntington's disease. *Biol. Psychiatry* **65**, 959–965 (2009). doi: 10.1016/j.biopsych.2008.12.026; pmid: 19200948
5. S. U. McKinstry *et al*., Huntingtin is required for normal excitatory synapse development in cortical and striatal circuits. *J. Neurosci.* **34**, 9455–9472 (2014). doi: 10.1523/JNEUROSCI.4699-13.2014; pmid: 25009276
6. A. Reiner, I. Dragatsis, S. Zeitlin, D. Goldowitz, Wild-type huntingtin plays a role in brain development and neuronal survival. *Mol. Neurobiol.* **28**, 259–275 (2003). doi: 10.1385/MN:28:3:259; pmid: 14709789
7. M. Barnat *et al*., Huntington's disease alters human neurodevelopment. *Science* **369**, 787–793 (2020). doi: 10.1126/science.aax3338; pmid: 32675289
8. E. E. Arteaga-Bracho *et al*., Postnatal and adult consequences of loss of huntingtin during development: Implications for Huntington's disease. *Neurobiol. Dis.* **96**, 144–155 (2016). doi: 10.1016/j.nbd.2016.09.006; pmid: 27623015
9. A. E. Molero *et al*., Selective expression of mutant huntingtin during development recapitulates characteristic features of Huntington's disease. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5736–5741 (2016). doi: 10.1073/pnas.1603871113; pmid: 27140644
10. M. C. Stephens, V. Brandt, J. Botas, The developmental roots of neurodegeneration. *Neuron* **110**, 1–3 (2022). doi: 10.1016/j.neuron.2021.12.004; pmid: 34990574
11. M. Molina-Calavita *et al*., Mutant huntingtin affects cortical progenitor cell division and development of the mouse neocortex. *J. Neurosci.* **34**, 10034–10040 (2014). doi: 10.1523/JNEUROSCI.0715-14.2014; pmid: 25057205
12. M. Barnat, J. Le Friec, C. Benstaali, S. Humbert, Huntingtin-Mediated Multipolar-Bipolar Transition of Newborn Cortical Neurons Is Critical for Their Postnatal Neuronal Morphology. *Neuron* **93**, 99–114 (2017). doi: 10.1016/j.neuron.2016.11.035; pmid: 28017473
13. M. Capizzi *et al*., Developmental defects in Huntington's disease show that axonal growth and microtubule reorganization require NUMA1. *Neuron* **110**, 36–50.e5 (2022). doi: 10.1016/j.neuron.2021.10.033; pmid: 34793694
14. T. Kroon, E. van Hugte, L. van Linge, H. D. Mansvelder, R. M. Meredith, Early postnatal development of pyramidal neurons across layers of the mouse medial prefrontal cortex. *Sci. Rep.* **9**, 5037 (2019). doi: 10.1038/s41598-019-41661-9; pmid: 30911152
15. H. L. Picken Bahrey, W. J. Moody, Early development of voltage-gated ion currents and firing properties in neurons of the mouse cerebral cortex. *J. Neurophysiol.* **89**, 1761–1773 (2003). doi: 10.1152/jn.00972.2002; pmid: 12611962
16. A. P. Y. Brown, L. Cossell, T. W. Margrie, Visual Experience Regulates the Intrinsic Excitability of Visual Cortical Neurons to Maintain Sensory Function. *Cell Rep.* **27**, 685–689.e4 (2019). doi: 10.1016/j.celrep.2019.03.073; pmid: 30995467
17. W. Chen *et al*., AMPA glutamate receptor subunits 1 and 2 regulate dendrite complexity and spine motility in neurons of the developing neocortex. *Neuroscience* **159**, 172–182 (2009). doi: 10.1016/j.neuroscience.2008.11.038; pmid: 19110036
18. H. Jiang *et al*., The GluN2B subunit of N-methy-D-asparate receptor regulates the radial migration of cortical neurons in vivo. *Brain Res.* **1610**, 20–32 (2015). doi: 10.1016/j.brainres.2015.03.031; pmid: 25838242
19. S. Naskar *et al*., The development of synaptic transmission is time-locked to early social behaviors in rats. *Nat. Commun.* **10**, 1195 (2019). doi: 10.1038/s41467-019-09156-3; pmid: 30867422
20. H. Arakawa, R. S. Erzurumlu, Role of whiskers in sensorimotor development of C57BL/6 mice. *Behav. Brain Res.* **287**, 146–155 (2015). doi: 10.1016/j.bbr.2015.03.040; pmid: 25823761
21. E. Troncoso *et al*., Recovery of evoked potentials, metabolic activity and behavior in a mouse model of somatosensory cortex lesion: Role of the neural cell adhesion molecule (NCAM). *Cereb. Cortex* **14**, 332–341 (2004). doi: 10.1093/cercor/bhg131; pmid: 14754871
22. G. A. Metz, I. Q. Whishaw, The Ladder Rung Walking Task: A Scoring System and Its Practical Application. Journal of Visualized Experiments. *J. Vis. Exp.* **28**, e1204 (2009). doi: 10.3791/1204; pmid: 19525918


## ACKNOWLEDGMENTS

We kindly thank the staff of animal facilities (GIN) and the GIN imaging facility platform (PIC-GIN), as well as M. Bartolomucci for excellent technical assistance, E. Denarier for image analysis advice, V. Brandt for critical editing of the manuscript, and J. Belforte, A. Durr, and F. Saudou for helpful discussions. **Funding:** This work was funded by the following: Agence Nationale pour la Recherche ANR-15-IDEX-02 NeuroCoG in the framework of the "Investissements d'avenir" program (to S.H. and E.B.); AXYON: ANR-18-CE16-0009-01 (to S.H.); ANR-11-INBS-0006: France Life Imaging network (to E.B.); Fondation pour la Recherche Médicale FRM, équipe labellisée DEQ20170336752 (to S.H.); AGEMED program from INSERM (to S.H.). **Author contributions:** B.Y.B. and S.H. designed the study and wrote the manuscript. B.Y.B. performed most of the experiments. D.W. performed Western Blot

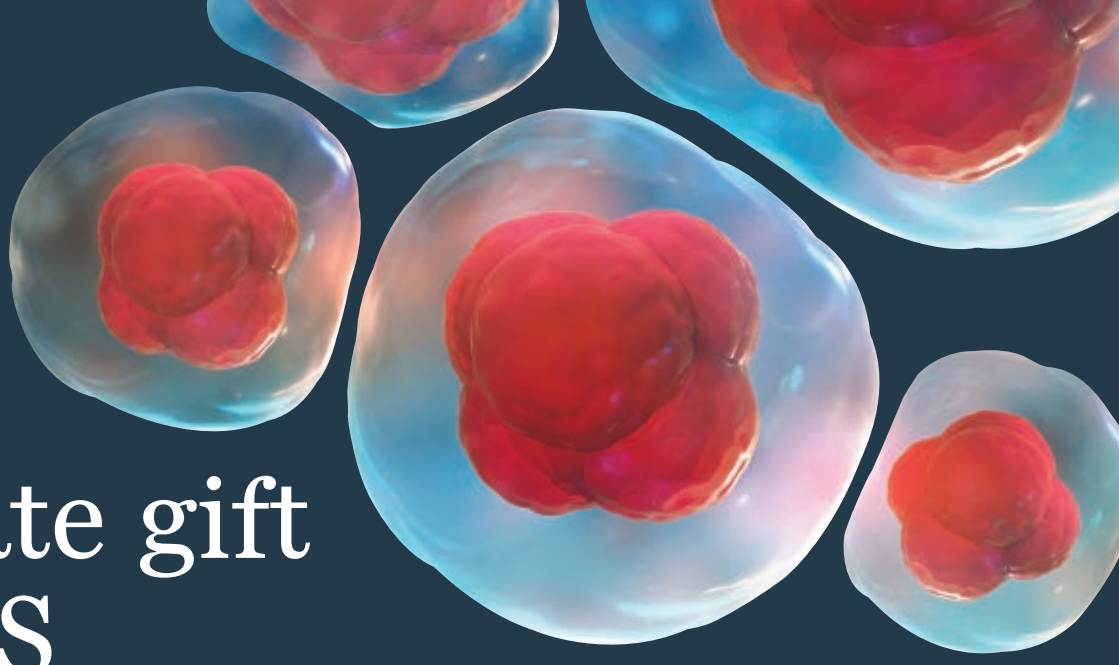
experiments. D.W. and L.R. performed immunolabelings. J.C.D. and D.A.R.S. performed magnetic resonance imaging experiments. F.L., A.B., and E.L.B. provided expertise and reagents. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. 

**SUPPLEMENTARY MATERIALS**

science.org/doi/10.1126/science.abq5011
Materials and Methods
Figs. S1 to S8
Tables S1 to S47
References (*23*–*39*)
MDAR Reproducibility Checklist

Submitted 13 April 2022; accepted 5 August 2022
10.1126/science.abq5011

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 ARTICLES

HEART FAILURE

# Liver-heart cross-talk mediated by coagulation factor XI protects against heart failure

Yang Cao[1], Yuchen Wang[1], Zhenqi Zhou[2], Calvin Pan[1], Ling Jiang[3], Zhiqiang Zhou[1], Yonghong Meng[1], Sarada Charugundla[1], Tao Li[3], Hooman Allayee[4], Marcus M. Seldin[5], Aldons J. Lusis[1,6,7]*

**Tissue-tissue communication by endocrine factors is a vital mechanism for physiologic homeostasis. A systems genetics analysis of transcriptomic and functional data from a cohort of diverse, inbred strains of mice predicted that coagulation factor XI (FXI), a liver-derived protein, protects against diastolic dysfunction, a key trait of heart failure with preserved ejection fraction. This was confirmed using gain- and loss-of-function studies, and FXI was found to activate the bone morphogenetic protein (BMP)–SMAD1/5 pathway in the heart. The proteolytic activity of FXI is required for the cleavage and activation of extracellular matrix–associated BMP7 in the heart, thus inhibiting genes involved in inflammation and fibrosis. Our results reveal a protective role of FXI in heart injury that is distinct from its role in coagulation.**

Tissue-tissue cross-talk by endocrine factors, including secreted proteins (*1*), is a vital mechanism to maintain proper physiologic homeostasis. The heart and the liver display multifaceted interactions (*2*), and in clinical practice it is common to observe heart diseases affecting the liver and vice versa (*3*). For instance, nonalcoholic fatty liver disease increases the risk for heart failure with diastolic and systolic dysfunction (*4*, *5*). On the basis of these observations, we hypothesized that secreted proteins may mediate communication between liver and heart. We screened for such endocrine factors using a "systems genetics" approach that integrates natural variation for physiological and clinical traits with global transcriptomics data in cohorts of genetically diverse mice. In our studies, we used a resource consisting of ~100 diverse, inbred strains of mice called the Hybrid Mouse Diversity Panel (HMDP) (*6*). Among several candidates that were identified was coagulation factor XI (FXI), a protein produced exclusively by liver.

[1]Department of Medicine, Division of Cardiology, University of California, Los Angeles, CA 90095, USA. [2]Division of Endocrinology, Diabetes and Hypertension, Department of Medicine, University of California, Los Angeles, CA 90095, USA. [3]Department of Anesthesiology, Laboratory of Mitochondria and Metabolism, West China Hospital of Sichuan University, Chengdu 610041, China. [4]Departments of Population and Public Health Sciences and Biochemistry and Molecular Medicine, University of Southern California Keck School of Medicine, Los Angeles, CA 90089, USA. [5]Department of Biological Chemistry and Center for Epigenetics and Metabolism, University of California, Irvine School of Medicine, Irvine, CA 92697, USA. [6]Department of Human Genetics, University of California, Los Angeles, CA 90095, USA. [7]Department of Microbiology, Immunology and Molecular Genetics, University of California, Los Angeles, CA 90095, USA.
*Corresponding author. Email: jlusis@mednet.ucla.edu

We validated several of these factors in a mouse model of a common form of heart failure, heart failure with preserved ejection fraction (HFpEF). We reasoned that we were more likely to see an effect if the heart was stressed. HFpEF is characterized by diastolic dysfunction and preserved ejection fraction, which is distinct from heart failure with reduced ejection fraction (HFrEF) (*7*). HFpEF accounts for half of all cases of heart failure and is associated with multiple comorbidities, including diabetes, hypertension, and restrictive cardiomyopathies (*8*, *9*). In HFpEF, chronic systemic inflammation and metabolic disorders affect not only the myocardium, but also other organs such as the kidneys, lungs, and skeletal muscles. However, little is known about the molecular mechanisms underlying impaired cardiac relaxation and how other organs interact with the heart to regulate the pathophysiology of HFpEF.

We chose to follow up on FXI, which was particularly interesting because not only did it perturb gene expression in the heart, but it also affected diastolic function. In the mouse model of HFpEF, mice overexpressing FXI in liver showed improved diastolic function, whereas FXI-knockout mice were sensitized for diastolic dysfunction. We identified potential pathways by which FX1 affects diastolic function by examining differential gene expression in response to changes in FXI levels. FXI overexpression activated the bone morphogenetic protein (BMP)–SMAD1/5 pathway in the heart. The action of FXI on the heart requires proteolytic activity, because point mutations in its catalytic domain eliminated the effects on BMP signaling and heart function. BMP7 is secreted as an inactive precursor that binds to the extracellular matrix, and our results indicate that it is cleaved by FXI, releasing the active growth factor from the prodomain. We also provide evidence that FXI has a similar function in humans. Our results identify FXI as an endocrine factor that influences heart function, and this is distinct from its role in coagulation.

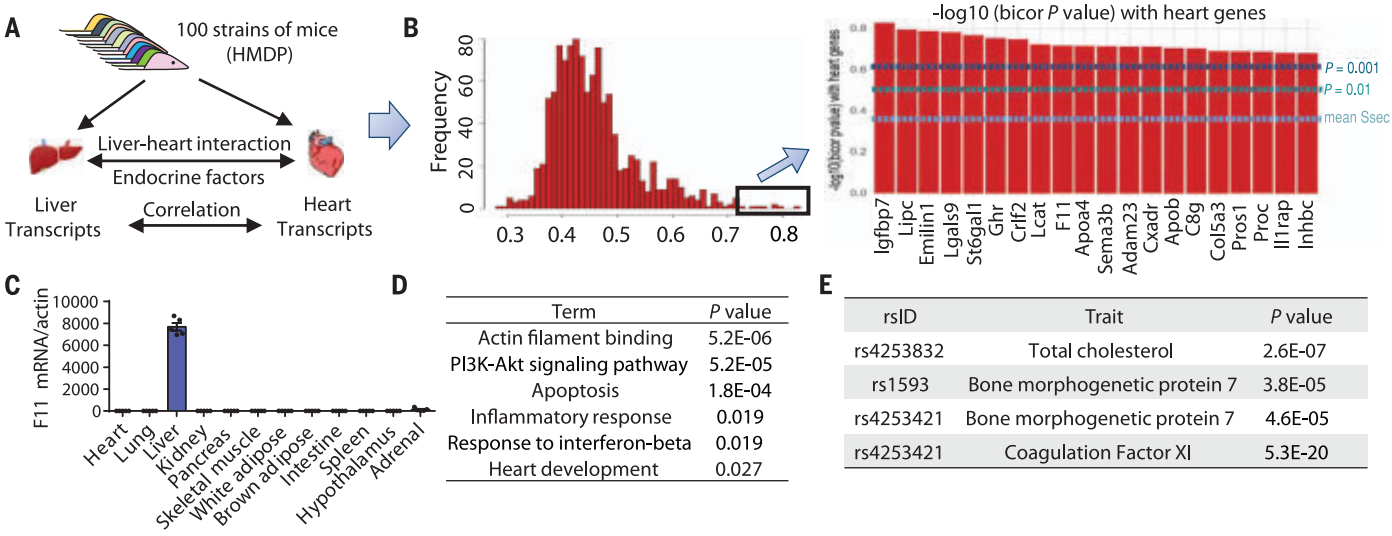
## Results

### *Systems genetics screening for potential regulators of liver-heart cross-talk*

To identify endocrine circuits mediating liver-heart cross-talk (*10*), we took advantage of a recently developed bioinformatics approach that uses natural variation to identify correlations between tissues. For this, we used a panel of 100 diverse inbred mouse strains in the HMDP (*6*, *11*). Global transcriptomic data from the heart and the liver were generated across all 100 inbred strains and used to detect the correlations between secreted proteins from the liver and their downstream effects on the heart (Fig. 1A). By assessing the strength of cross-tissue predictions, we generated a list of potential liver-heart mediators (Fig. 1B and table S1). The top-ranked candidates included *Igfbp7*, *Lipc*, *Emilin1*, *Lgals9*, *St6gal1*, *Ghr*, *Crlf2*, *Lcat*, and *F11*. This list revealed several previously described mediators with consistent functions. For instance, insulin-like growth factor-binding protein-7 (*Igfbp7*) has been reported to be correlated with diastolic function in HFrEF and HFpEF patients (*12*).

### *FXI protects against diastolic dysfunction, fibrosis, and inflammation in a mouse model of HFpEF*

On the basis of their specific expression in the liver, data from the literature, and functional annotation, we selected *Hgfac*, *C8g*, and *F11* as candidate mediators of liver-heart communication (Fig. 1B). To determine whether these factors have clinically relevant effects on the heart, we examined them in a mouse model of HFpEF, which is characterized by diastolic dysfunction. We overexpressed these genes individually in the livers of C57BL/6J male mice with an adeno-associated virus serotype 8 (AAV8) vector carrying target genes or green fluorescent protein (GFP) control and directed by the liver-specific thyroid hormone-binding globulin promoter (fig. S1A). After AAV8 injection, mice were subjected to a "two-hit" HFpEF model induced by a combination of high-fat diet (HFD) and inhibition of nitric oxide synthase using Nω-nitrol-arginine methyl ester (l-NAME) (*13*), followed by assessment of cardiac functions (fig. S1B). After 7 weeks of HFD + l-NAME feeding, mice developed heart failure phenotypes that recapitulated the clinical symptoms of HFpEF, including diastolic dysfunction [increased E/A ratio, E/e′ ratio, left ventricular (LV) mass, heart weight, and lung weight], metabolic disorders (increased body weight, fat mass, plasma lipids, and glucose




D

| Term | *P* value |
|---|---|
| Actin filament binding | 5.2E-06 |
| PI3K-Akt signaling pathway | 5.2E-05 |
| Apoptosis | 1.8E-04 |
| Inflammatory response | 0.019 |
| Response to interferon-beta | 0.019 |
| Heart development | 0.027 |

E

| rsID | Trait | *P* value |
|---|---|---|
| rs4253832 | Total cholesterol | 2.6E-07 |
| rs1593 | Bone morphogenetic protein 7 | 3.8E-05 |
| rs4253421 | Bone morphogenetic protein 7 | 4.6E-05 |
| rs4253421 | Coagulation Factor XI | 5.3E-20 |

**Fig. 1. Systems genetics analysis of cross-tissue correlations identifies proteins mediating liver-heart cross-talk.** (**A**) Schematic illustrating the identification of the liver-heart interaction using 100 inbred strains of mice (HMDP). The correlation between the secreted factors (from the liver) and cardiac gene expression (RNA-Seq) was used for liver-heart predictions. This framework identified peptides secreted by the liver and strongly associated with the cardiac gene network. *n* = 4 to 20 mice for each strain. (**B**) Distribution of significance score for all liver genes across all heart gene expression in 100 strains (left). List shows the top 20 genes potentially mediating liver-heart communication (right). (**C**) Quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis of *F11* expression across indicated tissues in C57BL/6J mice (*n* = 4). All data are presented as means ± SEM. (**D**) Pathway enrichment derived from heart genes correlated with liver *F11* expression. (**E**) GWAS loci for indicated clinical traits in human populations. The GWAS catalog and PhenoScanner databases consist of human genotype-phenotype associations from publicly available genetic association studies.

intolerance), exercise intolerance (reduced running distance), and preserved LV ejection fraction (LVEF) (fig. S1, C to N).

We observed that when overexpressed, liver-derived hepatocyte growth factor activator (HGFAC) increased LV mass and complement C8 gamma chain (C8G) decreased heart weight in the model of HFpEF (figs. S2 and S3). However, we focused on the other top candidate, FXI, because it had additional effects on several HFpEF traits, including diastolic function. FXI acts downstream of FXII (*14*, *15*) and triggers the middle phase of the intrinsic pathway of blood coagulation by activating FIX. Like HGFAC and C8G, FXI is also exclusively expressed in the liver (Fig. 1C and fig. S4, A and B). Furthermore, on the basis of associations with heart transcript levels in the HMDP, FXI was predicted to be strongly correlated with critical pathways in the heart and a number of clinical traits related to HFpEF (Fig. 1D and fig. S4C). In addition, human genome-wide association studies (GWAS) revealed that genetic loci encompassing the *F11* gene were associated with the total cholesterol and BMP7 levels (Fig. 1E and tables S2 and S3) (*16*). These data suggested a potential role of FXI in heart failure.

We observed reduced plasma FXI in the mice on HFD + l-NAME relative to those given a chow diet (fig. S4, D and E). We then induced the HFpEF model in 30 genetically diverse, inbred strains of mice, a subset of HMDP, to examine the association between plasma FXI and diastolic dysfunction in the context of naturally occurring variation. We found that FXI levels were inversely correlated with diastolic dysfunction after feeding the HFpEF diet (Fig. 2A). Taken together, these results suggest that FXI may protect against diastolic dysfunction.

We then directly validated the function of FXI in the HFpEF model using overexpression. C57BL/6J male mice injected with AAV8-*GFP* or AAV8-*F11* were subjected to chow diet or HFD + l-NAME for 7 weeks (fig. S1B). After AAV8 injection, *F11* expression was elevated in the liver and FXI protein was increased in the plasma (Fig. 2, B and C, and fig. S5, A and B). We injected sufficient virus to increase FXI protein levels only modestly (about 1.4-fold) to avoid nonphysiological artifacts. Plasma alanine transaminase levels were not significantly changed ($P = 0.29$) by FXI overexpression, suggesting no deleterious effects on the liver from overexpression (fig. S5C). FXI protein was not detected in the heart, confirming the specificity of AAV8 to target the liver and supporting the concept that FXI is an endocrine factor produced by the liver that affects the heart (fig. S5D). Mice receiving AAV8-*F11* had lower body weight and fat mass after HFpEF compared with those receiving AAV8-*GFP* (fig. S5E). Blood pressure was not affected by FXI (fig. S5F). Consistent with our genetic results in the HMDP, FXI overexpression decreased E/A ratio, E/e′ ratio, heart weight, and lung weight in the HFpEF model while preserving LVEF, indicating an improvement in diastolic function (Fig. 2, D to I, and fig. S5G). Running distance was also improved by FXI overexpression, indicating that FXI ameliorates exercise intolerance in HFpEF (Fig. 2J). FXI overexpression also had beneficial metabolic effects on fat mass and plasma lipid levels (fig. S6, A to D) but not on glucose tolerance (fig. S6, E to G).

To test whether FXI overexpression affects blood coagulation, we measured blood thrombin-antithrombin complexes in mice with GFP or FXI overexpression and found that they were not significantly changed ($P = 0.79$) in mice receiving AAV8-*F11* versus AAV8-*GFP* (Fig. 2K), suggesting that the coagulation system was not affected by FXI overexpression. It has been found that mean platelet volume, reflecting the size and activity of platelets, is increased in decompensated heart failure patients and correlates with disease severity, serving as an independent predictor of 6-month mortality after decompensation (*17*). We observed a small but significant increase of mean platelet volume upon HFpEF development ($P < 0.05$), and FXI overexpression reversed it ($P < 0.01$) (fig. S6H). FXI overexpression significantly reduced circulating inflammatory cells ($P < 0.05$) and cytokine levels [interleukin β (IL-β) and IL-6, $P < 0.05$; interferon γ, $P < 0.01$] in the HFpEF model (fig. S6, I and J). Moreover, the expression of inflammatory genes in the heart was also reduced by FXI overexpression (Fig. 2L and fig. S6K). When mice were maintained on a chow diet, the number of blood immune

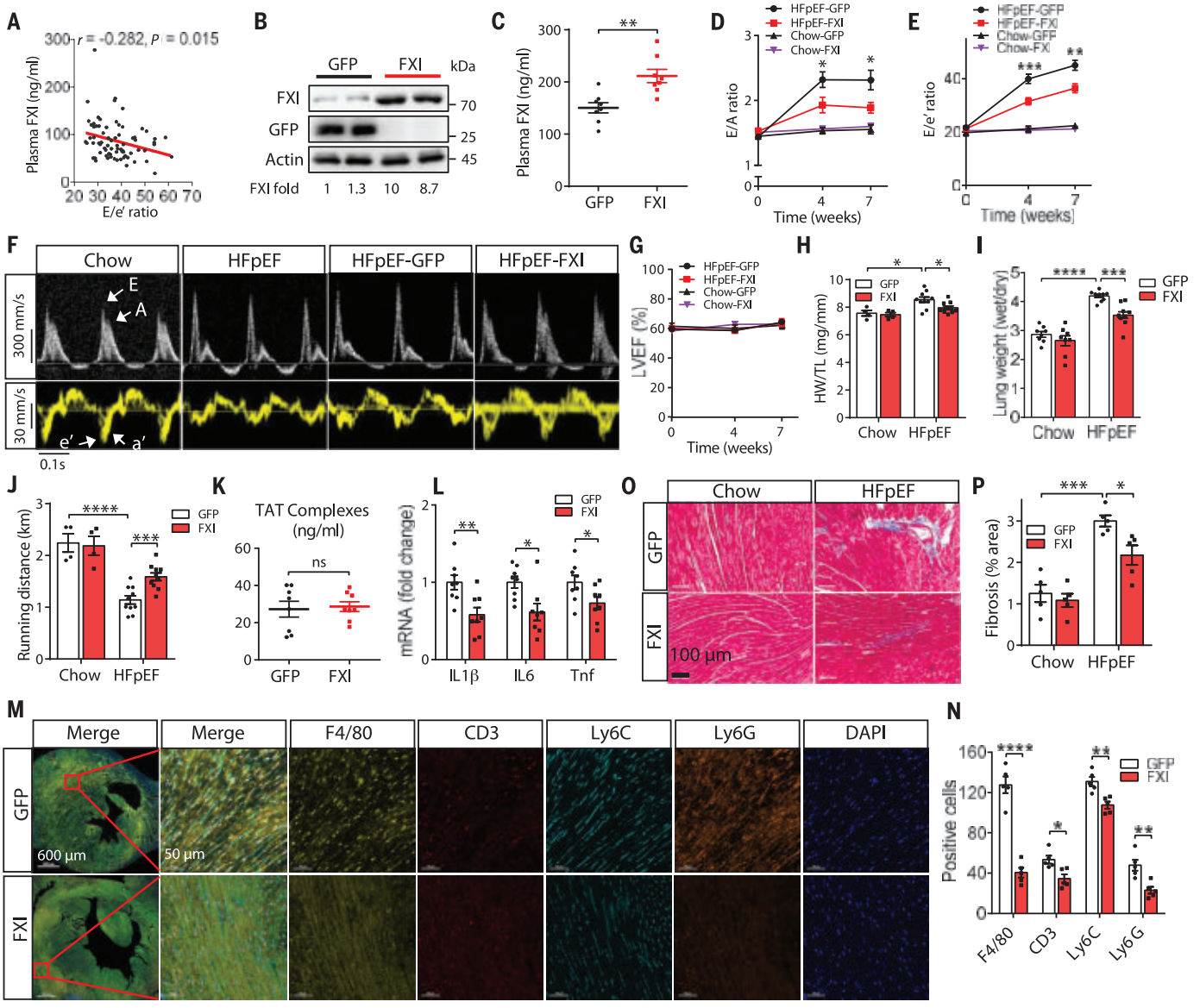



**Fig. 2. FXI overexpression reverses HFpEF-induced diastolic dysfunction, inflammation, and fibrosis.** (**A**) Thirty inbred strains of male mice were subjected to HFD + l-NAME to induce HFpEF. Plasma FXI concentrations and diastolic function (E/e′ ratio) were assessed after 7 weeks of feeding. Plasma FXI concentrations were inversely correlated with diastolic dysfunction. (**B** to **L**) C57BL/6J male mice were injected with AAV8 containing the cDNA sequence for *GFP* or *F11* and then fed with HFD + l-NAME for 7 weeks. Western blotting shows liver FXI protein (B), plasma FXI concentrations (C), E/A ratio (D), E/e′ ratio (E), representative images of echocardiography (F), LVEF (G), heart weight/tibia length ratio (H), lung weight [wet/dry ratio (I)], running distance (J), thrombin-antithrombin complexes [TAT (K)], and relative mRNA expression of indicated genes in the heart (L). $n$ = 4 for chow in (D) to (H); in other panels, $n$ = 8 to 10. (**M** and **N**) C57BL/6J male mice injected with AAV8-*GFP* or AAV8-*F11* were on HFD + l-NAME for 7 weeks ($n$ = 5). Representative images of immunohistochemistry staining (M) and quantification of positive cells (N) showing inflammatory cell infiltration in the heart tissue. (**O** and **P**) C57BL/6J male mice injected with AAV8-*GFP* or AAV8-*F11* were given a chow diet or HFD + l-NAME for 7 weeks ($n$ = 5). Representative images of Masson's trichrome staining (O) and quantification (P) show fibrosis in the heart tissue. Each point represents a mouse. All data are presented as means ± SEM. ns, not significant. *$P$ < 0.05, **$P$ < 0.01, ***$P$ < 0.001, and ****$P$ < 0.0001 by two-way ANOVA [(D) to (J) and (P)] or by Student's $t$ test [(B) and (C) and (K) to (N)]. For (A) to (C) and (K) to (N), all mice were on HFD + l-NAME. LYM, lymphocytes; MONO, monocytes; GRAN, granulocytes.

cells was not changed by FXI overexpression (fig. S6L).

To further test whether the cardiac infiltration of inflammatory cells was attenuated by FXI, we performed multiplex immunohistochemistry using antibodies against macrophages (F4/80), T cells (CD3), monocytes (Ly6C), and granulocytes (Ly6G). We observed significantly decreased inflammatory cells (F4/80, $P$ < 0.0001; CD3, $P$ < 0.05; Ly6C and Ly6G, $P$ < 0.01) in heart tissue from FXI-overexpressing mice versus GFP-overexpressing mice (Fig. 2, M and N), suggesting that FXI overexpression reduced inflammation in heart tissue in the HFpEF model. In addition, FXI overexpression also decreased fibrosis in the heart (Fig. 2, O and P).

### *FXI activates the BMP-SMAD1/5* pathway in cardiomyocytes

To investigate the molecular mechanism underlying the impact of FXI on the heart,

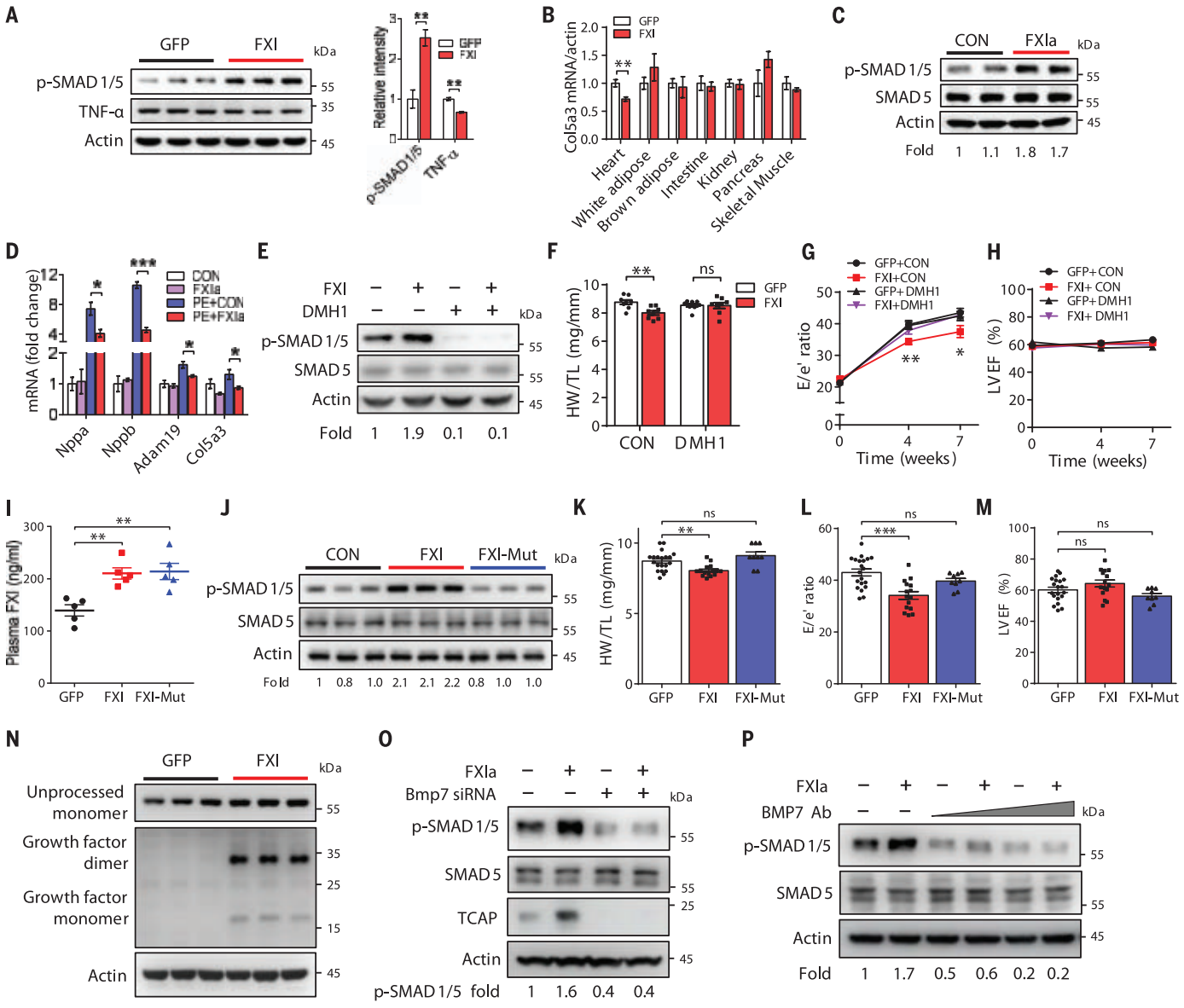



**Fig. 3. FXI activates BMP-SMAD1/5 pathway in the heart.** (**A**) Western blotting and quantification showing protein levels in heart tissue from C57BL/6J male mice injected with AAV8-*GFP* or AAV8-*F11* and fed with 7 weeks of HFD + l-NAME. Actin served as the loading control. $n = 5$. (**B**) qRT-PCR analysis showing the mRNA levels of *Col5a3* in the indicated tissue from C57BL/6J male mice injected with AAV8-*GFP* or AAV8-*F11* and fed HFD + l-NAME for 7 weeks. Heart, $P < 0.001$; others, not significant. $n = 8$. (**C** and **D**) NRVMs were treated with control or human FXIa protein (1 μg/ml) with medium containing control or phenylephrine (PE, 100 μM) for 24 hours. p-SMAD1/5 (C) and the indicated genes (D) were examined. Actin served as the loading control. $n = 6$. (**E** to **H**) C57BL/6J male mice were injected with AAV8-*GFP* or AAV8-*F11* with DMH1 and fed with HFD + l-NAME for 7 weeks. Heart p-SMAD1/5 level (E), heart weight/tibia length ratio (F), E/e′ ratio (G), and LVEF (H) were determined. $n = 3$ for (E) and $n = 8$ for (F) to (H). (**I** to **M**) C57BL/6J male mice were injected with AAV8-*GFP*, AAV8-*F11*, or AAV8-*F11*-Mut (m*F11*-Mut2) and fed with HFD + l-NAME for 7 weeks. Plasma FXI levels (I), heart p-SMAD1/5 protein level (J), heart weight/tibia length ratio (K), E/e′ ratio (L), and LVEF (M) were measured. $n = 5$ for (I), $n = 6$ for (J), and $n = 10$ to 20 for (K) to (M). (**N**) C57BL/6J male mice were injected with either AAV8-*GFP* or AAV8-*F11* and then fed with HFD + l-NAME for 7 weeks. BMP7 proteins in unprocessed monomer, growth factor dimer, and monomer under nonreducing condition were determined. (**O**) NRVMs were treated with control or human FXIa protein (1 μg/ml) plus negative control or Bmp7 siRNA with medium containing PE (100 μM) for 24 hours. p-SMAD1/5 and Tcap proteins were examined. Actin served as the loading control. $n = 3$. (**P**) NRVMs were treated with control or human FXIa protein (1 μg/ml), BMP7 antibody (no antibody control, 1:100 and 1:50), with medium containing PE (100 μM) for 2 hours, and the p-SMAD1/5 level was determined. Actin served as the loading control. $n = 4$. Each point represents a mouse. All data are presented as means ± SEM. ns, not significant. $^*P < 0.05$, $^{**}P < 0.01$, and $^{***}P < 0.001$ by two-way ANOVA [(D) to (H)], one-way ANOVA [(I) to (M)], or Student's *t* test [(A) to (C)].

we performed RNA sequencing (RNA-Seq) of the mice with FXI versus GFP overexpression in heart and adipose. Compared with GFP controls, 124 genes in the heart were significantly changed (adjusted $P < 0.05$) by FXI overexpression (fig. S7, A to C). Differentially expressed genes were enriched in pathways related to circadian rhythm, cardiac muscle contraction, inflammation, focal adhesion, the phosphatidylinositol 3-kinase (PI3K)–Akt pathway, and insulin signaling (fig. S7D). In contrast to the heart, only six genes in white adipose tissue were significantly changed (adjusted $P < 0.05$) by FXI overexpression (fig. S7E). *Tcap* (Titin-cap), and *Lrrc10* (Leucine-rich

repeat-containing 10), two genes involved in cardiac myofibril assembly and cardiac muscle tissue morphogenesis (2), were increased in the heart tissue of FXI-overexpressing mice (fig. S7, F and G).

To identify pathways perturbed by FXI, we again turned to the HMDP. Because we had performed global transcriptomics in the heart as well as the liver in all 100 strains, we could identify heart genes in which expression was correlated with the expression of FXI in the liver. On the basis of this, we examined the protein or RNA levels of the predicted pathways (Fig. 1E) and RNA-Seq (fig. S7), including the PI3K-Akt, nuclear factor κB, SMAD, and tumor necrosis factor-α (TNF-α) pathways (Fig. 3A and fig. S8A). We observed that members of the BMP pathway were correlated with FXI expression, and the link to BMP was supported by data from human GWAS (discussed below). Consistent with this, our overexpression studies showed that FXI induced an increase in SMAD1/5 phosphorylation and a decrease in TNF-α in the heart but not in other tissues (Fig. 3A and fig. S8, A to H), suggesting activation of the BMP-SMAD1/5 pathway and a decrease of inflammation in the heart. To test whether nuclear p-SMAD1/5 was also increased, we isolated the nuclear fraction from the same heart tissue and observed that it was significantly induced in the FXI overexpression group relative to GFP controls ($P < 0.0001$) (fig. S8I). We injected C57BL/6J male mice with saline control or mouse FXI protein and, after 2 hours, observed the phosphorylation of SMAD1/5 in the heart but not in other tissues, supporting the tissue-specific activation of the BMP-SMAD1/5 pathway by FXI (fig. S8, J to M). Plasminogen activator inhibitor-1 (PAI-1) was comparable in the hearts receiving AAV8-*F11* relative to those receiving AAV8-*GFP* (fig. S8A). However, FXI overexpression reversed the expression of the fibrotic and inflammatory genes *Col5a1*, *Col5a3*, *Adam19*, *IL1β*, *IL6*, and *Tnf* in the heart but not in other tissues examined, consistent with the observed decrease in fibrosis and inflammation in the heart (Fig. 3B and fig. S8, N and O).

To determine the localization of p-SMAD1/5, we stained p-SMAD1/5 and the markers of cardiomyocytes (troponin I), fibroblasts (vimentin), macrophages (CD68), and endothelial cells (CD31) in the heart after FXI overexpression. p-SMAD1/5 was colocalized with troponin I, but not with other markers (fig. S9), suggesting that p-SMAD1/5 was mainly activated in cardiomyocytes. To directly test whether FXI protein activates the BMP-SMAD1/5 pathway in cardiomyocytes, we incubated neonatal rat ventricular myocytes (NRVMs), human embryonic stem cell–induced cardiomyocytes, and other cell lines with control medium or medium containing human activated FXI protein (FXIa) in the presence of phenylephrine for 24 hours. We observed that FXIa increased the phosphorylation of SMAD1/5 and decreased the expression of *Nppa*, *Nppb*, *Col5a3*, and *Adam19* in NRVMs and human embryonic stem cell–induced cardiomyocytes but not in other cell types (Fig. 3, C and D, and fig. S10).

The above experiments were performed with male mice and we were interested in determining whether the results were similar with females. C57BL/6J female mice injected with AAV8-*GFP* or AAV8-*F11* were subjected to HFD + l-NAME for 7 weeks (fig. S11A). After AAV8 injection, *F11* expression was increased in the liver (fig. S11B). We observed similar effects of FXI in female mice, including decreased body mass, decreased inflammatory cells in the blood, reduced plasma lipids, increased SMAD1/5 phosphorylation, improved diastolic function, and reduced heart weight and lung weight (fig. S11, C to N). By contrast, blood pressure was comparable between the FXI and GFP groups (fig. S11O).

To determine whether the effects of FXI that we observed were specific to the HFpEF model, we examined a "multi-hit" HFpEF model induced by the combination of aging, HFD, and angiotensin II (*18*). Aged C57BL/6J male mice were injected with AAV8-*GFP* or AAV8-*F11* and then fed a HFD for 12 weeks. After 8 weeks of HFD, mice were infused with angiotensin II for 4 weeks (fig. S12A). *F11* mRNA was increased in the liver by AAV8-*F11* compared with AAV8-*GFP* controls (fig. S12B). Similar to the beneficial effects in the "two-hit" HFpEF model, we observed significant improvement in diastolic function and related traits in FXI-overexpressing mice relative to GFP-overexpressing mice, including reduced body mass ($P < 0.05$) (fig. S12C) and improved diastolic function as measured by lower E/A ratio ($P < 0.05$), lower E/e′ ratio ($P < 0.001$), and lower LV mass ($P < 0.05$) (fig. S12, D to H). In addition, FXI-overexpressing mice exhibited reduced heart weight and plasma total cholesterol, as well as increased p-SMAD1/5 relative to GFP controls (fig. S12, I to K).

To confirm that FXI overexpression activates BMP signaling to protect against diastolic dysfunction, we blocked the BMP receptor with the dorsomorphin homolog 1 (DMH1) (*19*). DMH1 is a selective inhibitor of activin receptor-like kinase 3, a type 1 BMP receptor. In NRVMs, DMH1 treatment suppressed SMAD1/5 phosphorylation induction by FXIa (fig. S13A). We also examined the effect of DMH1 in vivo. C57BL/6J mice were injected with AAV8-*F11* and fed with HFD + l-NAME for 7 weeks. Injection of DMH1 every other day to block SMAD1/5 phosphorylation (*20*) suppressed the change of body mass induced by FXI (fig. S13B), as well as the effect of FXI on p-SMAD1/5 levels, diastolic function, adipose weight, blood cell numbers, and plasma cholesterol (Fig. 3, E to H, and fig. S13, C to G). These results confirmed that FXI protects against HFpEF by activating the BMP pathway.

### *FXI protease activity is required for the activation of BMP signaling*

The FXI protein is conserved in human, mouse, rat, and other species and consists of four apple domains and one catalytic domain (fig. S14A). It is present in that plasma as a zymogen, which exists as a homodimer consisting of two identical polypeptide chains linked by disulfide bonds (fig. S14B) (*21*). During FXI activation, an internal peptide bond is cleaved by FXIIa (or XII) in each of the two chains, resulting in activated FXIa, a serine protease composed of two heavy and two light chains held together by disulfide bonds (fig. S14B). To test whether the catalytic domain is required for the function of FXI on the heart, we introduced two missense mutations in human and mouse FXI catalytic domains (fig. S14, A to E). These mutations were predicted to be exposed at the surface of the FXI molecule and to cause functional defects (type II mutation) (*22*). Next, we tested their function in vitro using a co-culture system. Huh7 human liver cells and AML12 mouse liver cells were transfected with respective human or mouse plasmids containing GFP control, wild-type (WT) *F11*, or *F11* with mutations. Then, cells were placed in co-cultures with NRVMs or 3T3-L1 adipocytes (fig. S15A). Twenty-four hours after transfection, FXI was highly induced in both Huh7 cells and AML12 cells (fig. S15B). In NRVMs, phosphorylation of SMAD1/5 was induced by WT FXI overexpression from both human and mouse liver cells, whereas mutant FXI did not exhibit a comparable effect (fig. S15C). By contrast, SMAD1/5 phosphorylation was not significantly ($P = 0.84$) induced by FXI in 3T3-L1 adipocytes, suggesting a heart-specific effect (fig. S15D). Consistent with phosphorylated SMAD1/5, *Col5a3* was decreased by WT FXI but not mutant FXI in NRVMs, indicating that the catalytic activity is required for its effect (fig. S15E).

To test the effect of missense mutation in vivo, we produced AAV8 with the mouse WT and mutant *F11* coding sequences. AAV8 containing *GFP* control, WT *F11*, and mutant *F11* (mF11-Mut2) was injected into C57BL/6J male mice followed by HFD + l-NAME for 7 weeks, after which plasma FXI was increased in FXI group and was comparable to the mutant FXI group (Fig. 3I and fig. S15F). Body weight and fat mass were decreased by WT FXI overexpression, but there was no significant difference ($P > 0.05$) between groups of mutant FXI and GFP controls (fig. S15, G to I), suggesting functional defects of mutant FXI. Consistently, the effects of FXI on p-SMAD1/5, heart weight, E/A ratio, E/e′ ratio, adipose weight, plasma cholesterol, and blood immune cells were not observed in mice carrying mutant FXI, demonstrating that catalytic activity is essential for

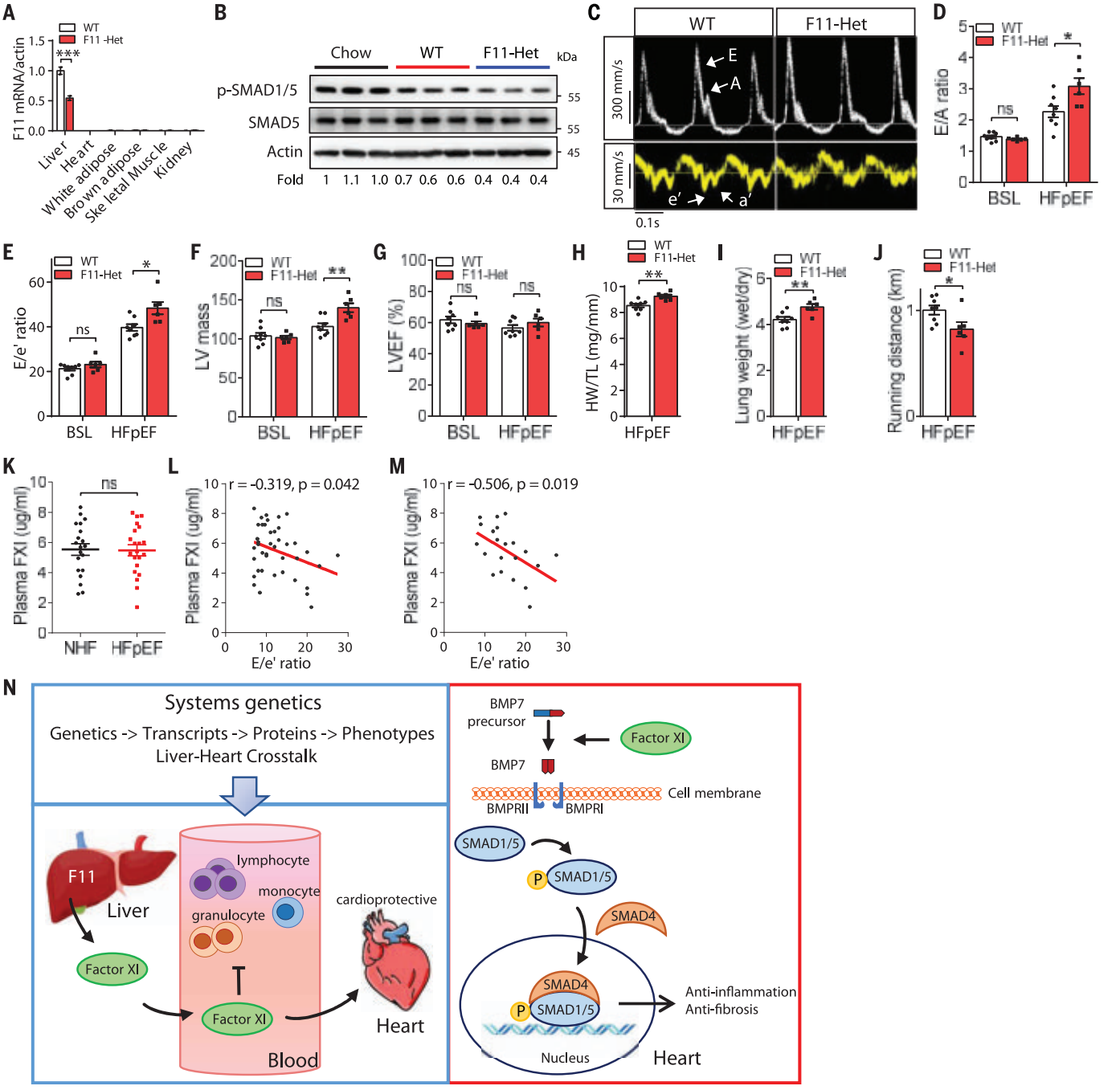



**Fig. 4. Reduced FXI concentrations are associated with diastolic dysfunction in mice and humans.** Heterozygous B6.129X1-F11$^{tm1Gjb}$/J (*F11*-Het) mice and WT littermates at 8 weeks of age were subjected to HFD + l-NAME for 7 weeks. $n$ = 8 for WT and $n$ = 6 for *F11*-Het. (**A**) qRT-PCR showing *F11* mRNA in the indicated tissues from WT and *F11*-Het mice. $n$ = 4. (**B**) Western blotting showing p-SMAD1/5 in the hearts of WT mice fed with chow diet (Chow), WT, and *F11*-Het mice fed with HFD + l-NAME for 7 weeks. $n$ = 5. (**C** to **G**) Representative images of echocardiography (C), E/A ratio (D), E/e' ratio (E), LV mass (F), and LVEF (G) were examined at baseline (BSL) and after 7 weeks of HFD + l-NAME feeding (HFpEF). $n$ = 8 for WT and $n$ = 6 for *F11*-Het. (**H** to **J**) Heart weight/tibia length ratio (H), lung weight [wet/dry ratio (I)], and running distance (J) were examined after 7 weeks of HFD + l-NAME feeding (HFpEF). $n$ = 8 for WT and $n$ = 6 for *F11*-Het. (**K**) Plasma FXI protein in non-HFpEF controls (NHF, $n$ = 20) and HFpEF patients ($n$ = 21). (**L** and **M**) Plasma FXI protein was inversely correlated with E/e' ratio in all participants (L), including HFpEF patients (M). (**N**) Illustration summarizing FXI-mediated liver-heart cross-talk in protecting against heart failure. Using a bioinformatic framework that integrates global liver-heart transcriptome and cardiometabolic trait data from the HMDP, we found that coagulation FXI, secreted by the liver, exhibits cardioprotective effects on the progression of HFpEF. FXI overexpression in the liver mitigates the diastolic dysfunction, inflammation, and fibrosis induced by HFpEF. FXIa cleaves the BMP7 precursor and activates the BMP7-SMAD1/5 pathway in the heart to mediate the anti-inflammatory and anti-fibrotic effects. Each point represents a mouse. All data are presented as means ± SEM. ns, not significant. *$P$ < 0.05, **$P$ < 0.01, and ***$P$ < 0.001 by two-way ANOVA [(D) to (G)], one-way ANOVA (B), or Student's $t$ test [(A) and (H) to (K)].

the function of FXI in protecting against deleterious phenotypes in HFpEF (Fig. 3, J to M, and fig. S15, J to P).

### *FXI cleaves the BMP7 proprotein, activating the resulting growth factor fragment*

As a serine protease, FXIa catalyzes the proteolysis of its substrates. BMP7 is synthesized as a large precursor molecule (inactive) that is cleaved to growth factor dimer or monomer (active) by proteolytic enzymes (*23*). We found that the cleavage site of the full-length BMP7 protein, at an arginine, is a common FXIa cleavage site (fig. S16, A and B) (*24*). We therefore hypothesized that BMP7 is a substrate of FXIa that mediates SMAD1/5 activation. FXI overexpression increased BMP7 growth factor dimer and monomer in the heart (Fig. 3N). Moreover, incubation of FXIa with BMP7 protein resulted in the cleavage of BMP7 (fig. S16C). Knocking down BMP7 or treatment with BMP7 antibody in NRVMs greatly reduced the activation of SMAD1/5 by FXIa (Fig. 3, O and P). The BMP7 protein is considerably enriched in heart tissue and cardiomyocytes (fig. S17), which may explain the preferential effect of FXI on the heart. The prodomain of BMP7 appears to bind to the extracellular matrix (*23*), suggesting that FXI cleaves the precursor BMP7 bound to the extracellular matrix in the heart; this then releases the dimer and monomer growth factors from the matrix to bind to the BMP receptor and activate SMAD1/5.

### *FXI knockout mice have reduced p-SMAD1/5 levels and increased diastolic dysfunction*

We sought to further examine the cardioprotective effect of FXI using FXI knockout male mice in which the *F11* gene was disrupted by a PGK-neo cassette (*25*). *F11* transcripts in the liver of heterozygous null mice (*F11*-Het) were reduced by ~50% compared with WT littermates (Fig. 4A). FXI was either absent or barely detectable in other tissues (Fig. 4A and fig. S4, A and B). Adult WT and *F11*-Het mice were then subjected to HFD + l-NAME for 7 weeks to induce HFpEF phenotypes. Compared with WT littermates on the HFpEF diet, p-SMAD1/5 was reduced in the hearts of *F11*-Het mice (Fig. 4B). Consistent with reduced p-SMAD1/5, *F11*-Het mice exhibited more severe diastolic dysfunction, as evidenced by the increased E/A ratio, E/e′ ratio, and LV mass but preserved ejection fraction (Fig. 4, C to G). Moreover, heart weight and lung weight were higher in *F11*-Het mice relative to WT controls, suggesting cardiac hypertrophy and lung congestion in FXI-deficient mice (Fig. 4, H and I). Exercise tolerance was also decreased in *F11*-Het mice compared with WT mice (Fig. 4J). By contrast, blood pressure was not significantly changed ($P > 0.05$) by FXI deficiency (fig. S18), indicating that FXI does not influence heart function through effects on blood pressure. These results collectively demonstrated the increased severity of diastolic dysfunction in FXI-deficient mice. We observed consistent effects of FXI in female mice with FXI heterozygous knockout (fig. S19).

### *FXI levels correlate with diastolic function in human cohorts*

To determine the clinical relevance of FXI, we quantified plasma FXI in human patients with HFpEF and in normal participants. Plasma FXI protein was not significantly different ($P = 0.91$) between non–heart failure controls and HFpEF patients (Fig. 4K). However, plasma FXI was inversely correlated with E/e′ ratio in all participants (Fig. 4L), including HFpEF patients (Fig. 4M), supporting the conclusion that FXI protects against diastolic dysfunction in HFpEF.

## Discussion

Our results indicate that liver-derived FXI specifically regulates cardiomyocytes through the BMP-SMAD1/5 pathway, resulting in attenuation of fibrosis, inflammation, and diastolic dysfunction in the context of an HFpEF model (Fig. 4N). Our analysis of diastolic function in a cohort of heart failure patients indicates the relevance of the pathway in humans and in mouse models, and the human GWAS results are consistent with that conclusion.

Prior studies have implicated the BMP and SMAD pathways in traits relevant to heart failure. It has been reported that the BMP pathway is enriched for HFpEF but not HFrEF (*26*). BMP2 has been found to alleviate heart failure with type 2 diabetes by inhibiting inflammasome formation (*27*), and is inversely correlated with the concentrations of atrial natriuretic peptide and brain natriuretic peptide in chronic heart failure patients with diabetes. Another study observed increased BMP6 in chronic heart failure patients, suggesting that BMP6 may be involved in the pathophysiology of systolic heart failure (*28*). In addition, SMAD1 protein was differentially expressed in a high-salt diet-induced HFpEF model (*29*).

FXI is a component of the intrinsic pathway of blood coagulation, acting downstream of FXII and functioning as a protease to activate FIX (*14*, *15*, *21*, *30*, *31*). Our data indicate that FXI overexpression also influences various systemic aspects of metabolism, and we cannot rule out the possibility that it may also affect organs other than the heart. FXI-deficient patients generally do not have spontaneous bleeding, because FXI is not required for the initial thrombin generation step (*32*), consistent with the possibility that it exhibits other previously unknown functions. Inactivating mutations of *F11* are relatively common among Ashkenazi Jews (*33*). A number of studies investigated the relationship between FXI and incident coronary heart disease, stroke, and ischemic cardiomyopathy (*34*, *35*). FXI was reported to improve or protect against the inflammatory responses and cytokine responses to infections independently of its intrinsic coagulation activity (*36*–*40*).

The fact that FXI is a direct mediator of liver-heart communication suggests the possibility of using it in therapeutic applications for heart failure. It is important to note that elevated FXI is associated with various thromboses and ischemic stroke (*41*–*43*), so elevating FXI would be problematic as a therapeutic goal. However, the downstream BMP pathway could provide potential therapeutic targets.

## REFERENCES AND NOTES


1. J. M. Friedman, J. L. Halaas, *Nature* **395**, 763–770 (1998).
2. S. Møller, M. Bernardi, *Eur. Heart J.* **34**, 2804–2811 (2013).
3. K. K. Baskin, A. L. Bookout, E. N. Olson, *EMBO Mol. Med.* **6**, 436–438 (2014).
4. M. Packer, *Am. J. Med.* **133**, 170–177 (2020).
5. L. Pacifico *et al*., *Hepatology* **59**, 461–470 (2014).
6. Y. Cao *et al*., *Circulation* **145**, 1102–1104 (2022).
7. S. M. Dunlay, V. L. Roger, M. M. Redfield, *Nat. Rev. Cardiol.* **14**, 591–602 (2017).
8. K. Hogg, K. Swedberg, J. McMurray, *J. Am. Coll. Cardiol.* **43**, 317–327 (2004).
9. T. E. Owan *et al*., *N. Engl. J. Med.* **355**, 251–259 (2006).
10. M. M. Seldin *et al*., *Cell Metab.* **27**, 1138–1155.e6 (2018).
11. A. J. Lusis *et al*., *J. Lipid Res.* **57**, 925–942 (2016).
12. P. U. Gandhi *et al*., *JACC Heart Fail.* **4**, 860–869 (2016).
13. G. G. Schiattarella *et al*., *Nature* **568**, 351–356 (2019).
14. P. N. Walsh, *Thromb. Haemost.* **86**, 75–82 (2001).
15. J. Emsley, P. A. McEwan, D. Gailani, *Blood* **115**, 2569–2577 (2010).
16. E. Ferkingstad *et al*., *Nat. Genet.* **53**, 1712–1721 (2021).
17. H. Kandis *et al*., *Emerg. Med. J.* **28**, 575–578 (2011).
18. C. Withaar *et al*., *Cardiovasc. Res.* **117**, 2108–2124 (2021).
19. M. A. Petersen *et al*., *Neuron* **96**, 1003–1012.e7 (2017).
20. Y. Shan *et al*., *Front. Neurosci.* **12**, 964 (2018).
21. K. Fujikawa, D. W. Chung, L. E. Hendrickson, E. W. Davie, *Biochemistry* **25**, 2417–2424 (1986).
22. R. E. Saunders *et al*., *Thromb. Haemost.* **102**, 287–301 (2009).
23. K. E. Gregory *et al*., *J. Biol. Chem.* **280**, 27970–27980 (2005).
24. X. Ge *et al*., *Blood* **131**, 353–364 (2018).
25. D. Gailani, N. M. Lasky, G. J. Broze Jr., *Blood Coagul. Fibrinolysis* **8**, 134–144 (1997).
26. L. Adamo *et al*., *J. Am. Coll. Cardiol.* **76**, 1982–1994 (2020).
27. J. M. Zhang *et al*., *Exp. Ther. Med.* **22**, 897 (2021).
28. J. Banach *et al*., *Clin. Exp. Pharmacol. Physiol.* **43**, 1247–1250 (2016).
29. G. Zhou *et al*., *Front. Physiol.* **12**, 607089 (2021).
30. W. Wu *et al*., *J. Biol. Chem.* **283**, 18655–18664 (2008).
31. P. H. Bolton-Maggs, *Baillieres Clin. Haematol.* **9**, 355–368 (1996).
32. A. P. Wheeler, D. Gailani, *Expert Rev. Hematol.* **9**, 629–637 (2016).
33. R. Asakai, D. W. Chung, E. W. Davie, U. Seligsohn, *N. Engl. J. Med.* **325**, 153–158 (1991).
34. D. Appiah *et al*., *Blood Coagul. Fibrinolysis* **28**, 389–392 (2017).
35. M. Zabczyk, S. Butenas, I. Palka, J. Nessler, A. Undas, *Pol. Arch. Med. Wewn.* **120**, 334–340 (2010).
36. I. Stroo *et al*., *Thromb. Haemost.* **117**, 1601–1614 (2017).
37. A. T. P. Ngo *et al*., *J. Thromb. Haemost.* **19**, 1001–1017 (2021).
38. C. Puy *et al*., *J. Immunol.* **206**, 1784–1792 (2021).
39. E. I. Tucker *et al*., *Blood* **119**, 4762–4768 (2012).
40. C. E. Bane Jr. *et al*., *PLOS ONE* **11**, e0152968 (2016).
41. V. Rovite *et al*., *Thromb. Res.* **134**, 659–663 (2014).
42. Y. Li *et al*., *J. Thromb. Haemost.* **7**, 1802–1808 (2009).
43. M. Chong *et al*., *Circulation* **140**, 819–830 (2019).

44. C. Pan, R script used to perform the volcano plots and PCA plot (F11_SF7) for: Y. Cao *et al*., Liver-heart cross-talk mediated by coagulation factor XI protects against heart failure, Zenodo (2022); https://doi.org/10.5281/zenodo.6961041.

**ACKNOWLEDGMENTS**

**Funding:** This work was supported by the National Institutes of Health (grants DK120342 and HL147883; grants HL138193, DK130640, and DK097771 to M.M.S.; grants R01HL133169 and R01HL148110 to H.A.; and grant DK125354 to Z.Z.). **Author contributions:** Y.C., M.M.S., and A.J.L. designed the experiments. Y.C., Y.W., Z.Z., L.J., Z.Z., Y.M., S.C., and T.L. performed the experiments. Y.C., L.J., M.M.S., and C.P. analyzed raw data. Z.Z., H.A., M.M.S., and A.J.L. reviewed the data and made substantial contributions to improving the studies. Y.C. and A.J.L. wrote the manuscript, which was reviewed by all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data supporting the conclusions in this manuscript can be found in the main text or the supplementary materials. RNA-Seq data were deposited to the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE200496. The R script used to perform the volcano plots and PCA plot are available at Zenodo (*44*). 

**SUPPLEMENTARY MATERIALS**

science.org/doi/10.1126/science.abn0910
Materials and Methods
Figs. S1 to S20
Tables S1 to S8
References (*45*–*50*)
MDAR Reproducibility Checklist

Submitted 2 November 2021; resubmitted 10 May 2022
Accepted 5 August 2022
10.1126/science.abn0910

PEROVSKITES

# Controlling the nucleation and growth kinetics of lead halide perovskite quantum dots

**Quinten A. Akkerman[1,2]†, Tan P. T. Nguyen[3], Simon C. Boehme[1,2], Federico Montanarella[1,2]‡, Dmitry N. Dirin[1,2], Philipp Wechsler[1], Finn Beiglböck[2], Gabriele Rainò[1,2], Rolf Erni[4], Claudine Katan[3], Jacky Even[5], Maksym V. Kovalenko[1,2]***

**Colloidal lead halide perovskite nanocrystals are of interest as photoluminescent quantum dots (QDs) whose properties depend on the size and shape. They are normally synthesized on subsecond time scales through hard-to-control ionic metathesis reactions. We report a room-temperature synthesis of monodisperse, isolable, spheroidal $APbBr_3$ QDs ("A" indicates cesium, formamidinium, and methylammonium) that are size tunable from 3 to >13 nanometers. The kinetics of both nucleation and growth are temporally separated and substantially slowed down by the intricate equilibrium between the precursor ($PbBr_2$) and the $A[PbBr_3]$ solute, with the latter serving as a monomer. QDs of all these compositions exhibit up to four excitonic transitions in their linear absorption spectra, and we demonstrate that the size-dependent confinement energy for all transitions is independent of the A-site cation.**

Colloidal lead halide perovskite (LHP) nanocrystals (NCs) are light-emissive materials (*1*, *2*) of practical interest for light-emitting diodes (*3*, *4*), liquid crystal displays (*5*), lasers (*6*), scintillators (*7*, *8*), and luminescent solar concentrators (*9*, *10*). These materials exhibit several favorable photoluminescent (PL) properties, including near-unity PL quantum yields (QYs), a PL peak tailorable across the spectral range of 410 to 800 nm, small PL full width at half maximum (<100 meV), large absorption cross sections, long exciton coherence times, and fast (subnanosecond) radiative rates at low temperatures (*11*–*13*). Of rising scientific interest are single-photon emission from LHP NCs (*13*–*17*) and collective phenomena in LHP NC assemblies such as superfluorescence (*18*–*21*).

Given their synthetic availability, most studies on colloidal LHP NCs focused on relatively large NC sizes with edge lengths of or exceeding ~10 nm. In this weak excitonic confinement regime, tunability of the bandgap energy is predominantly achieved through mixed-halide compositions. Being structurally soft with mainly ionic chemical bonding, LHP NCs differ rather orthogonally from more conventional quantum dot (QD) materials such as CdSe and InP that have covalent bonding and rigid crystal structures. This difference manifests in the difficulty in producing small LHP QDs (only a few nanometers in diameter) and has hindered studies into the strong size quantization of excitons in LHPs (and possible practical use), as well as understanding of the mechanism of LHP QD formation.

The high lattice ionicity and low lattice formation energy of LHP NCs lead to challenges for the synthesis of small QDs. They form too quickly during ionic coprecipitation, with subsecond formation kinetics. In this regard, the surface capping ligands, which are crucially important for the controlled nucleation and growth, as well as for the structural integrity of the resulting QDs and their stability against solvents and environment, bind in a noncovalent and dynamic manner (*22*), which also adds to the difficulty of arresting QD formation and stabilizing them at small sizes.

Monodispersity and shape uniformity, as well as precise size tunability in the sub–10-nm regime, require a scalable synthesis that yields isolable, pure, and robust QD samples. Limited success had been reached for $CsPbX_3$ ("X" indicates Cl, Br, and I) compositions (*21*, *23*, *24*), wherein QD fractions down to 3 to 4 nm were obtained postsynthetically. However, the synthesis yield was small for a specific QD size fraction, mainly because of the initial broad size dispersion or very lossy isolation. Long-chain zwitterionic capping ligands exhibited improved surface adhesion relative to more conventional single–binding-head molecules and rendered the small $CsPbBr_3$ QDs sufficiently robust to sustain multiple steps of size-selective precipitation, leading to monodisperse fractions (*21*). Small colloidal $FAPbX_3$ and $MAPbX_3$ NCs (in which FA indicates formamidinium and MA indicates methylammonium) remained elusive because they are even more labile structurally.

We reasoned that the slower formation of small monodisperse LHP QDs at full reaction yield cannot be accomplished by only mere adjustments of typical reaction parameters, such as lowering of the reaction temperature and concentration of reagents. We hypothesized that the precursor–monomer–QD nuclei conversion path must be characterized by the mutual chemical equilibrium between these moieties governed by the common complexing agent, thus having a self-limiting character. Notably, this equilibrium should not be obscured by any competing process, such as binding of a stronger capping ligand to the QD surface or constituting ions. The latter condition somewhat goes against the previous efforts on stabilizing the perovskite NC surfaces by stronger binding ligands.

We found that trioctylphosphine oxide (TOPO) served this threefold requirement as it complexes (solubilizes) the $PbBr_2$ precursor, binds to the $Cs[PbBr_3]$ monomer (solute), and then weakly coordinates to the QD nucleus surface. The issue of isolation and subsequent

[1]Department of Chemistry and Applied Biosciences, ETH Zürich, CH-8093 Zürich, Switzerland. [2]Empa-Swiss Federal Laboratories for Materials Science and Technology, CH-8600 Dübendorf, Switzerland. [3]Univ Rennes, ENSCR, INSA Rennes, CNRS, ISCR - UMR 6226, F-35000 Rennes, France. [4]Electron Microscopy Center, Empa – Swiss Federal Laboratories for Materials Science and Technology, Dübendorf, Switzerland. [5]Univ Rennes, INSA Rennes, CNRS, Institut FOTON - UMR 6082, F-35000 Rennes, France.
***Corresponding author. Email: mvkovalenko@ethz.ch**
†Present address: Lehrstuhl für Photonik und Optoelektronik, Nano-Institut München und Fakultät für Physik, Ludwig-Maximilians-Universität (LMU), Königinstr. 10, D-80539 München, Germany.
‡Present address: Seaborough BV, Matrix VII Innovation Center, Science Park 106, 1098 XG Amsterdam, Netherlands.

robustness of the obtained QDs was then mitigated by adding a stronger binding ligand—lecithin as a long-chain zwitterion—at the end of the QD formation step (not during their synthesis). We thus present a room-temperature synthesis, in which the overall QD formation occurred on a time scale of up to 30 min, which also allowed the in situ observation of the reaction by optical absorption and emission spectroscopy. After isolation, monodisperse rhombicuboctahedral $CsPbBr_3$ QDs, with mean size tunable between 3 and 13 nm, were obtained at 100% precursor-to-QD conversion rate. These $CsPbBr_3$ QDs, as well as $FAPbBr_3$ and $MAPbBr_3$ QDs obtained analogously, exhibited

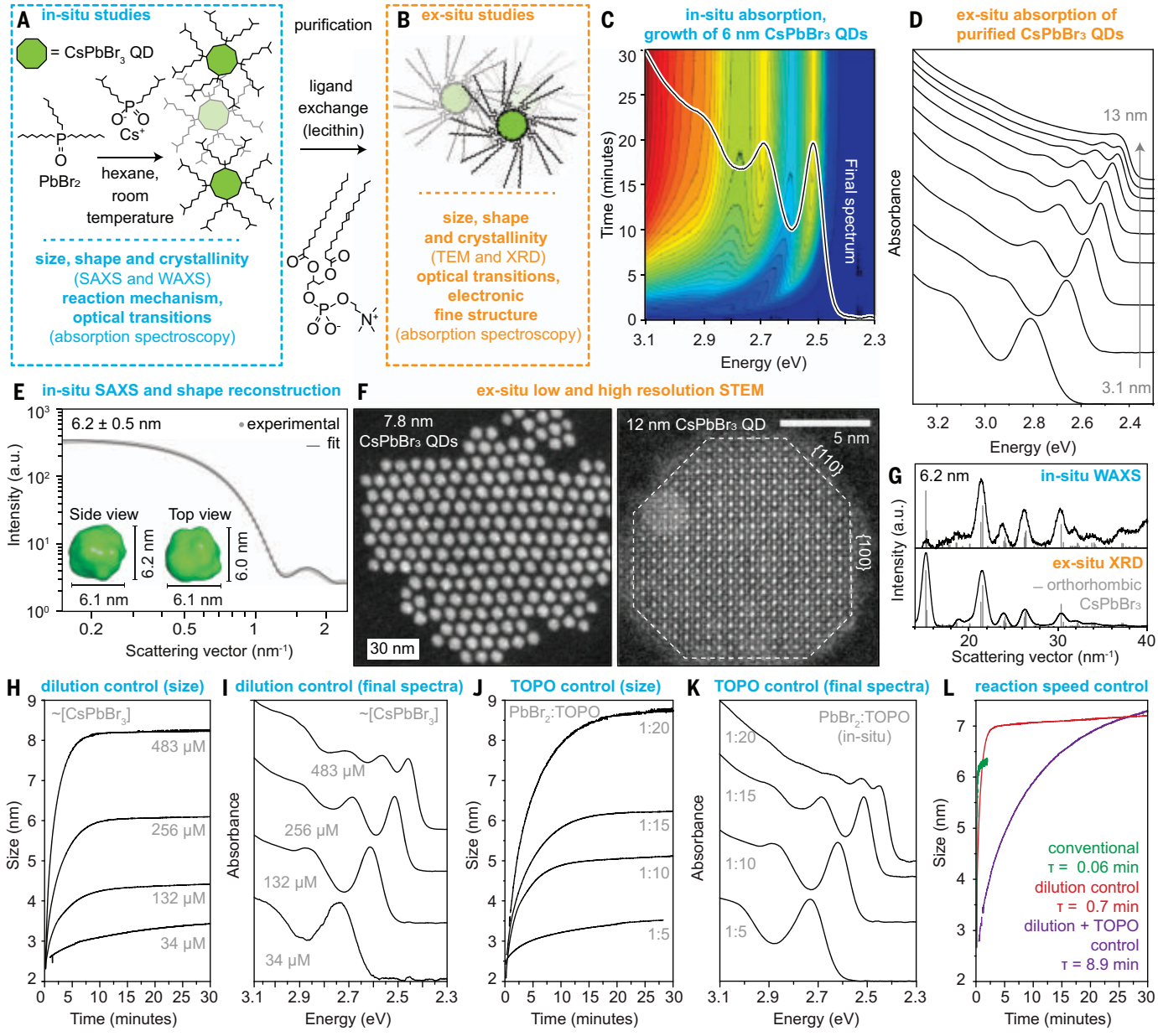



**Fig. 1. Room-temperature, controlled-rate synthesis of monodisperse $CsPbBr_3$ QDs.** (**A**) Reaction scheme and overview of in situ monitoring techniques. (**B**) Overview of used ex situ techniques on ligand-exchanged and washed QDs, which are complementary with the used in situ techniques. (**C**) Example of in situ recorded absorption spectra of 6-nm QDs during 30-min reaction, with the solid line being the final recorded absorption spectrum, demonstrating clear and sharp first and higher-order absorption peaks. (**D**) Absorption spectra of series of purified $CsPbBr_3$ QDs ranging from ~3 to 13 nm. (**E**) SAXS scattering curve of 6-nm QDs recorded at the end of a reaction, resulting in a size dispersion of ~8%, and a fitted pseudospherical shape with one axis being slightly shorter than the other two. a.u., arbitrary units. (**F**) STEM image of washed 7.8-nm QDs and HR-STEM image of a single QD showing the pseudospherical shape arising from truncation. (**G**) Diffractograms of 6.2-nm QDs recorded both in situ in a crude solution by using WAXS and from a washed dispersion by using XRD on a film, both exhibiting a match with the orthorhombic crystal structure of $CsPbBr_3$. (**H**) In situ recorded size of $CsPbBr_3$ QDs at various concentrations (concentration of Cs-DOPA) showing both the decrease in size and reaction speed upon dilution. (**I**) Respective final absorption spectra of reactions shown in (H), indicating that all final samples are highly monodisperse. (**J**) In situ recorded size of $CsPbBr_3$ QDs at various $PbBr_2$:TOPO ratios (with a fixed $PbBr_2$ concentration) showing the increase in size with increasing TOPO concentration, whereas the reaction rate remained relatively unchanged. (**K**) Respective final absorption spectra of reactions shown in (C), indicating that all final samples are highly monodisperse. (**L**) Comparison of reaction rates of three different reactions resulting in roughly the same QDs size, showing how controlling both the overall dilution and the TOPO concentration can slow down the reaction by two orders of magnitude.

up to four well-resolved excitonic transitions, and the size-dependent confinement energies for all transitions were independent of the A-site cations. The energies of these transitions, obtained by using the statistics from >25,000 individual spectra, agreed with those calculated by using an effective an mass/**k·p** model. We anticipate that analogous "low ionic strength" reactions will be developed for precision synthesis of NCs of other structurally soft metal halides, as well as for engineering complex metal halide NC heterostructures.

### Synthesis of monodisperse $CsPbBr_3$ QDs

We hypothesized that conventional approaches, which we refer to as oleylamine–oleic acid (OLAM/OA) path (*25*), suffer from the rapid conversion of the $PbBr_2$ into high concentrations of haloplumbate ionic solute ($PbBr_3^-$) upon the action of OLAM/OA (*26*, *27*). The subsequent reaction with injected Cs ions was thus quantitative and fast, and hence, nucleation and growth were not temporally separated. We aimed at establishing the mechanism, in which the formation of $PbBr_3^-$ anions was activated by the introduction of Cs cations, by having the latter as the only available cation in the system throughout the synthesis. The equilibrium was retained with the nonreactive $PbBr_2$ precursor reservoir through a common coordinating agent, in this case the neutral molecule TOPO as a sole coordinating agent for solubilizing $PbBr_2$ precursor (Fig. 1, A and B) (*28*). We injected cesium-diisooctylphosphinate (Cs-DOPA) solution into this $PbBr_2$:TOPO solution at room temperature (both in hexane; materials and methods, supplementary text, and figs. S1 to S5). Both TOPO and DOPA are weakly binding ligands for LHP QDs (*28*, *29*) and hence were readily displaced with lecithin, a strongly binding zwitterionic ligand (*30*), for subsequent isolation and purification steps as well as ex situ characterization (detailed comparison of in situ and ex situ data analysis, as well as the in situ setup, is shown in figs. S6 to S13). The reaction volume can be readily scaled to yield gram-scale quantities of QDs with identical optical properties (table S3 and fig. S14). The stability of the obtained QDs of all sizes for at least 1 year is seen as a retention of identical absorption spectra (fig. S15).

In situ optical absorption spectra (6-nm $CsPbBr_3$ QDs; Fig. 1C and movie S1) recorded the slow growth over 30 min that led to a narrow size dispersion, as evidenced by sharp excitonic absorption peaks. Purified lecithin-coated QDs exhibited the same sharp absorption features when measured ex situ and could be prepared from 3 to 13 nm in diameter (Fig. 1D and fig. S16), with an average PL QY of 87 ± 3% for QD sizes down to 4 nm and 72 ± 7% for 3-to-4-nm QDs. In situ synchrotron small-angle x-ray scattering (SAXS) measurements (6.2-nm QDs; Fig. 1E) revealed a narrow size dispersion (8.1%) and isotropic spheroidal particle shapes. The latter was in stark contrast to previous ex situ SAXS measurements on cuboidal $CsPbBr_3$ QDs, for which on average a 20% offset between the shortest and longest edge length of the cuboid was reported (*21*, *30*). These findings also held for 12-nm QDs (fig. S17).

Transmission electron microscopy (TEM) and scanning TEM (STEM) images of purified 8-nm QDs (Fig. 1F and fig. S18) revealed a preference for hexagonal packing, indicating spheroidal QDs, rather than the commonly observed nanocubes (*20*, *21*, *23*, *25*). High-resolution STEM (HR-STEM) of 12-nm QDs showed truncation along the {110} and {111} facets, hence the rhombicuboctahedral shape of QDs (fig. S19), in good agreement with the spheroidal shape retrieved from in situ SAXS. $CsPbBr_3$ QDs crystallize in a phase-pure orthorhombic phase, as seen from in situ wide-angle x-ray scattering (WAXS) of crude solutions as well as powder x-ray diffraction (XRD) (Fig. 1G). $CsPb(Cl:Br)_3$ QDs, which exhibit larger bandgaps than $CsPbBr_3$ QDs and are of interest for blue light-emitting diodes, could be obtained with this method by applying an in situ anion exchange, using a $ZnCl_2$-TOPO solution added before the lecithin ligand (fig. S20).

In situ optical measurements allowed for thorough and fast parametric screening. The overall dilution and the concentration of the TOPO ligands were most effective for controlling the QD size in our room-temperature synthesis, contrary to conventional size control by varying the reaction temperature (*21*, *23*, *25*, *30*).

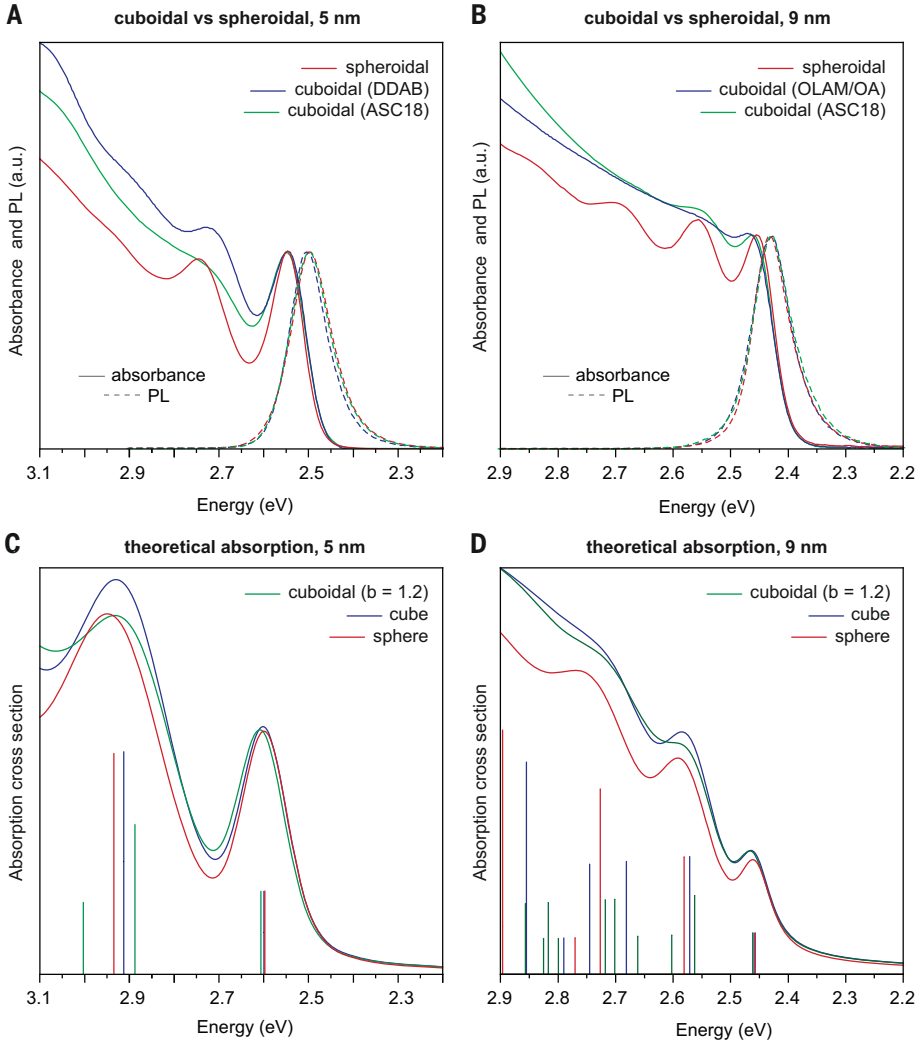



**Fig. 2. Effect of QD shape on the absorption spectral features.** (**A** and **B**) Experimental absorption and PL spectra of 5- and 9-nm cuboidal and spheroidal QDs. Monodisperse cuboidal QDs coated with DDAB (*20*), OLAM/OA (*20*), or ACS18 (*21*). (**C** and **D**) Absorption spectra of both 5- and 9-nm QDs with different shapes calculated by using the EMA with the inclusion of electron-hole Coulomb interaction in a single-shot calculation. For the cuboidal QD, an oblate shape was chosen with 20% shortening along one of the cube's edges (parameter $b$ = 1.2).

For example, dilution of the reaction by an order of magnitude led to a decrease in size from 9 to 3 nm (Fig. 1, H and I, and fig. S21). Similar size changes resulted by adjusting the concentrations of $PbBr_2$ or Cs-DOPA precursors individually (figs. S1 to S4). Alternatively, a fourfold increase in TOPO concentration increased the QD size from ~3 to 9 nm (Fig. 1, J and K). A combined effect of dilution and larger TOPO concentration translated into substantially slower QD growth (Fig. 1L, fig. S22, and movie S2).

### Spheroidal versus cuboidal QDs

Thus far, reported colloidal perovskite NCs were mostly of a {001}-bound, cuboidal shape (*25*), which can be ascribed to commonly used alkylammonium ligands stabilizing the planes of fully built $PbX_6$-octahedra (AX as terminating surface planes) with ammonium moieties incorporated by substituting the surface A-cations. The absence of cationic and other strongly binding surfactants in the present study may explain the tendency to form spheroidal NCs, exposing several crystal facets. We elucidated the origin of the exceptionally well-resolved excitonic transitions in the spheroidal QDs. The difference could not be attributed merely to the size dispersion. For example, Fig. 2, A and B, compares our spheroidal $CsPbBr_3$ QDs with several of our own examples of cuboidal $CsPbBr_3$ QDs with same or better size dispersion (fig. S23). The latter include NCs capped with didodecyldimethylammonium bromide (DDAB) or OLAM/OA, as used recently for NC superlattices (*20*), as well as those capped with long-chain sulfobetaine zwitterionic ligands (3-(*N*,*N*-dimethyloctadecylammonio) propanesulfonate, or ASC18) after size-selective precipitation (*21*, *31*). The similarity of PL full width at half maximum across these different shapes also evidences that the inhomogeneous broadening was not the cause for the "blurred" higher-order absorption transitions in cuboidal QDs either.

We analyzed the effect of the shape itself by computing a one-photon absorption cross section (Fig. 2, C and D, and fig. S24). For this, the effective mass approximation (EMA) was used with an energy-dependent effective mass to emulate the effect of full two-band coupling (*32*). The electron-hole Coulomb interaction was included in a single-shot calculation (that is, without self-consistency iteration; see details in supplementary text). In comparison with a perfect sphere, the cubic symmetry introduced a perturbation that coupled the spherical states of different angular momenta. This coupling led to mixing between and splitting of the higher-order absorption states, which resulted in the experimentally observed smoothing out of the absorption spectra. Furthermore, recent SAXS studies on cuboidal $CsPbBr_3$ NCs revealed up to 20% reduction in length along one of the cuboid edges (oblation), which further contributed to the splitting of higher-order absorption states (*21*, *30*). In particular, the transition from a spherical shape to an oblate cuboidal shape led to the nearly complete flattening of the second and third absorption transitions.

### Reaction mechanism

We used in situ optical absorption spectra to understand how the use of a solely neutral coordinating agent (TOPO) prevents premature and uncontrolled conversion of $PbBr_2$ into the $PbBr_3^-$ solute. $PbBr_2$ dissolved in TOPO to form a $PbBr_2$[TOPO] complex (*28*) that had a broad absorption peak around 4.4 eV (Fig. 3A). Its intensity change allowed us to monitor the

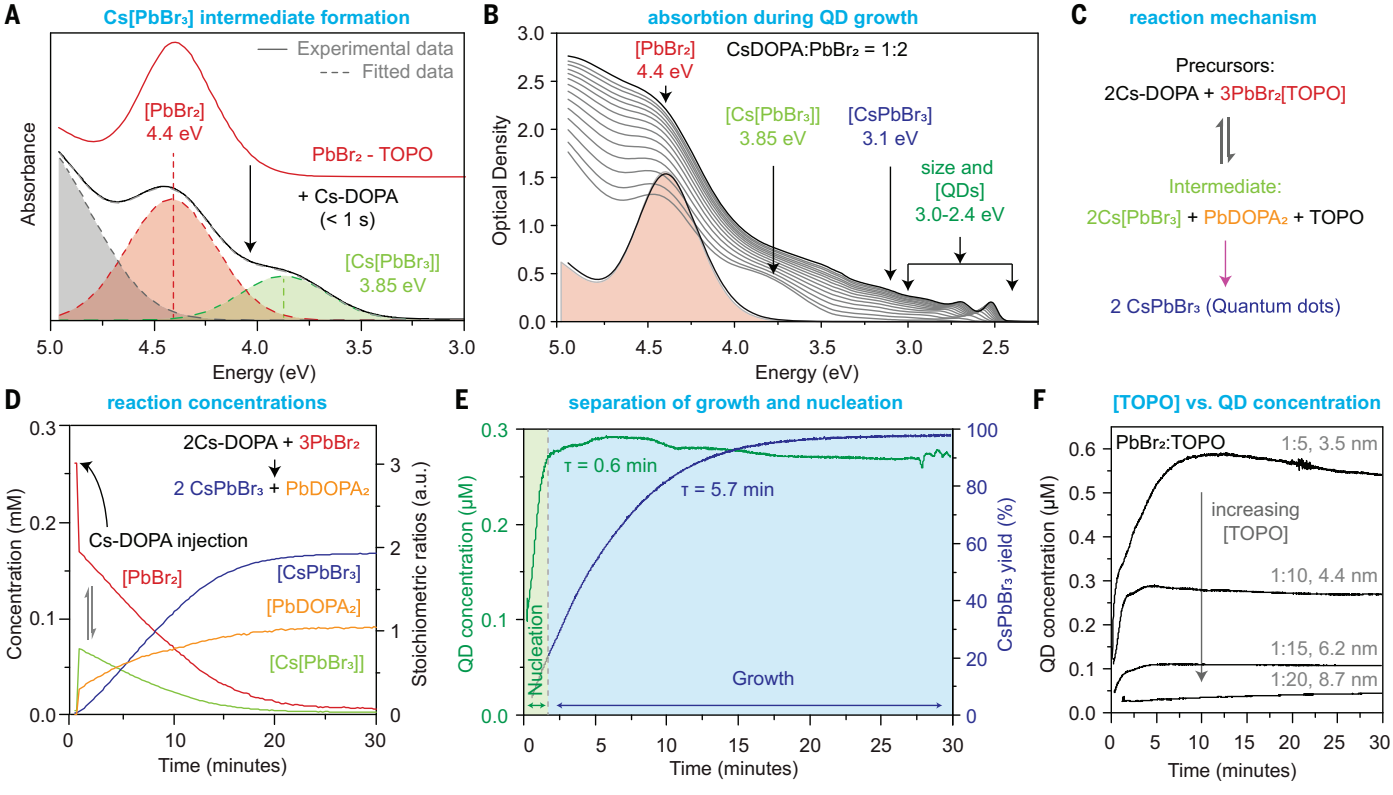



**Fig. 3. In situ observation of the conversion of precursors to $CsPbBr_3$ QDs.** (**A**) $PbBr_2$-TOPO solution before and within 5 s after the injection of Cs-DOPA. (**B** and **C**) Typical evolution of the absorption spectra from $PbBr_2$ through $Cs[PbBr_3]$ monomer and $CsPbBr_3$ QDs and the corresponding reaction path. (**D**) Temporal evolution of precursors' and reaction products' concentrations for a stochiometric reaction (Cs-DOPA:$PbBr_2$ = 2:3), wherein both the $PbBr_2$ precursor and the $Cs[PbBr_3]$ intermediate are consumed upon the formation of $CsPbBr_3$ QDs. $PbBr_2$ and $Cs[PbBr_3]$ exist in equilibrium, as both are present until $PbBr_2$ is consumed. (**E**) Effective separation of nucleation and growth, as well as absence of QD coalescence or Ostwald ripening, is apparent from the QD concentration evolution and the respective reaction yield. (**F**) QD concentration versus time dependence at different TOPO concentrations. Higher TOPO quantities suppress the available $Cs[PbBr_3]$ monomer and hence the number of nuclei.

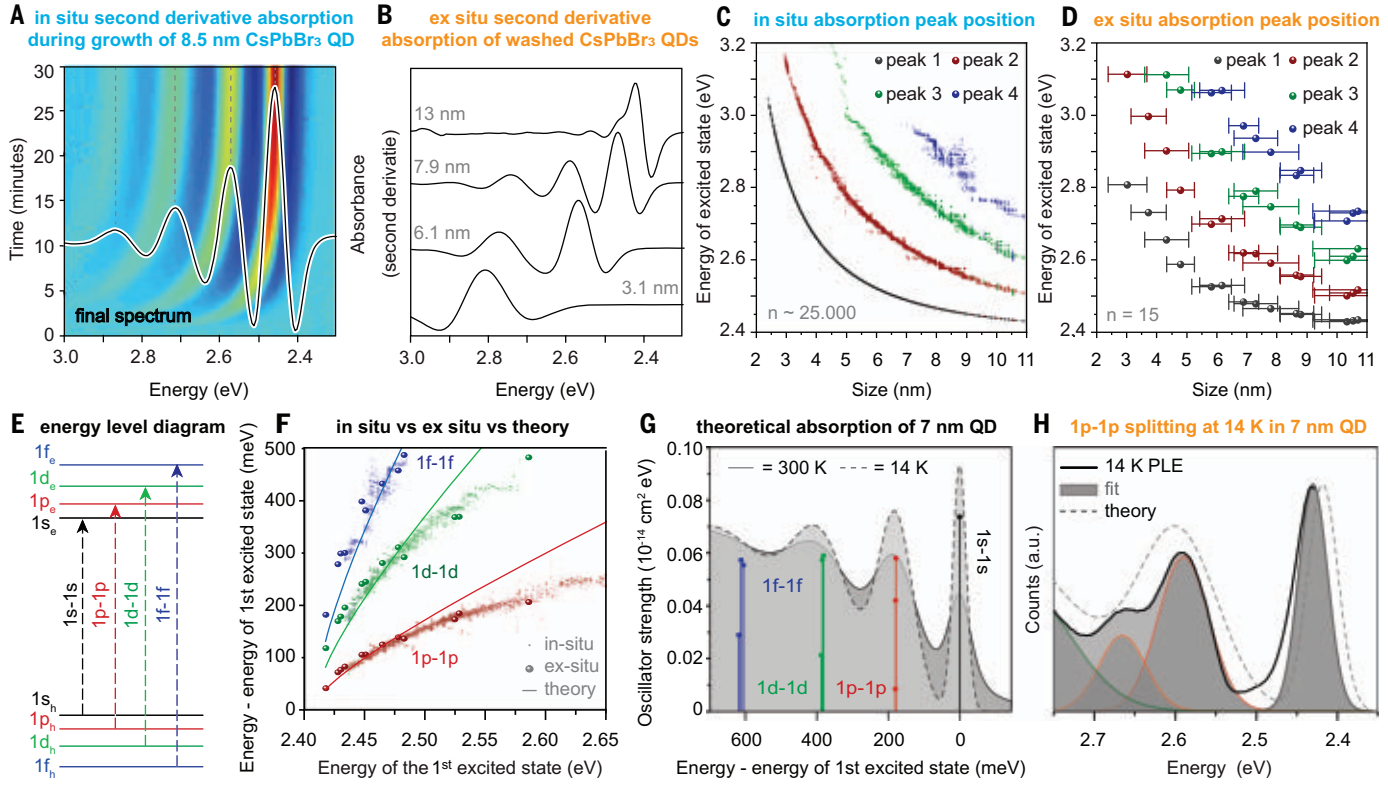



**Fig. 4. Excitonic transitions in $CsPbBr_3$ QDs seen with in and ex situ spectroscopy corroborated with theoretical calculations.** (**A**) Typical second derivative of the absorption during the growth of $CsPbBr_3$ QDs, featuring up to four transitions that can be extracted by using a peak-finding algorithm. (**B**) Second derivative absorption spectra for purified QDs. (**C**) Compilation of the extracted in situ transition energy values across several syntheses ($n$ ~ 25,000 individual absorption spectra). (**D**) Similar dataset obtained ex situ from 15 different QD sizes. (**E**) Energy-level diagram with the relevant electron (e) and hole (h) states used for the two-band **k·p** theory with Hartree-Fock–level description of the Coulomb interaction (*32*, *38*). (**F**) Comparison of experimental relative absorption transitions and theoretical calculated states. (**G**) Theoretical absorption spectrum in which the electron-photon interaction is considered within the framework of second-order many-body perturbation theory with the corresponding peak assignment. (**H**) PLE spectrum at 14 K revealing a splitting of the 1p-1p transition caused by a slight anisotropy in effective masses.

depletion of $PbBr_2$ during the reaction. Only upon the injection of Cs-DOPA did $PbBr_3^-$ appear in the course of 100 ms, as evidenced as an additional peak emerging at lower energies (3.85 eV) (fig. S25), from the formation of a $PbBr_3^-$ complex (*27*, *33*). Once the accumulation of $PbBr_3^-$ surpassed a nucleation threshold (supersaturation), $CsPbBr_3$ QDs formed (on the order of seconds; Fig. 3B). More conventional ligands for $CsPbBr_3$ QDs, such as OA or OLAM, reacted with the $PbBr_2$[TOPO] complex (*28*), but free DOPA on its own did not (fig. S26).

As the only counter-ion available to $PbBr_3^-$ is $Cs^+$, we assumed the formation of a Cs[$PbBr_3$] complex (tight ionic pair) as the concentration of $PbBr_3^-$ species scaled with Cs-DOPA concentration (fig. S27A). This Cs[$PbBr_3$] complex might be additionally coordinated by TOPO. Notably, a higher TOPO concentration suppressed the formation of Cs[$PbBr_3$], indicating that it formed in equilibrium (fig. S27B) with the more stable $PbBr_2$[TOPO] species. Particularly, an injection of additional TOPO after the initial Cs-DOPA injection decomposed the Cs[$PbBr_3$] complex back into $PbBr_2$ species (fig. S28).

An overall precursor-to-QDs path is depicted in Fig. 3C. We illustrate the effect of TOPO-driven $PbBr_2$:Cs[$PbBr_3$] equilibrium on the formation of $CsPbBr_3$ QDs for a stoichiometric $PbBr_2$:Cs-DOPA = 3:2 reaction by plotting in situ–monitored $PbBr_2$, Cs[$PbBr_3$], and $CsPbBr_3$ concentrations (Fig. 3D). An immediate decrease in $PbBr_2$ concentration upon injection of Cs-DOPA was accompanied by the formation of Cs[$PbBr_3$]. As the reaction progressed, Cs[$PbBr_3$] complexes were converted into $CsPbBr_3$ QD nuclei, and both $PbBr_2$ and Cs[$PbBr_3$] species were proportionally consumed but remained present, consistent with the $PbBr_2$:Cs[$PbBr_3$] equilibrium. The stoichiometric conversion $3PbBr_2 \rightarrow 2PbBr_3^- + Pb^{2+}$ was confirmed by subtracting Cs[$PbBr_3$] and $CsPbBr_3$ concentrations to yield the $PbDOPA_2$ concentration. The reaction eventually reached near 100% yield (fig. S29).

The rate-limiting step in this reaction was the conversion of the Cs[$PbBr_3$] intermediate into $CsPbBr_3$. A reaction with a fourfold excess of $PbBr_2$ compared with Cs-DOPA (fig. S30) showed that the Cs[$PbBr_3$] complex solely acted as an intermediate for the $CsPbBr_3$ QDs and was not stable on its own. The $PbBr_2$:Cs[$PbBr_3$] equilibrium self-limited the available amount of Cs[$PbBr_3$] for both the nucleation and the subsequent growth of $CsPbBr_3$ QDs. The separation of nucleation and growth, a prerequisite for the narrow size dispersion, was evident from a comparison of the QD concentration versus the $CsPbBr_3$ molar concentration evolution (Fig. 3E and movie S3). The self-limiting equilibrium mechanism also explained the larger QD size and lower reaction rate with increasing TOPO concentration, which shifted the equilibrium toward $PbBr_2$, reducing the available quantity of Cs[$PbBr_3$] for the nucleation and hence decreasing the number of nuclei (and the final QD concentration; Fig. 3F), whereas the overall yield remained relatively unchanged (fig. S31). For example, a 4-fold increase in TOPO concentration increased the particle size from 3.5 to 8.7 nm, corresponding to a 14-fold increase in QD volume. The QD concentration correspondingly showed a 15-fold decrease. Additional experiments confirming the self-limiting equilibrium mechanism are described in the supplementary text. Also, other approaches to

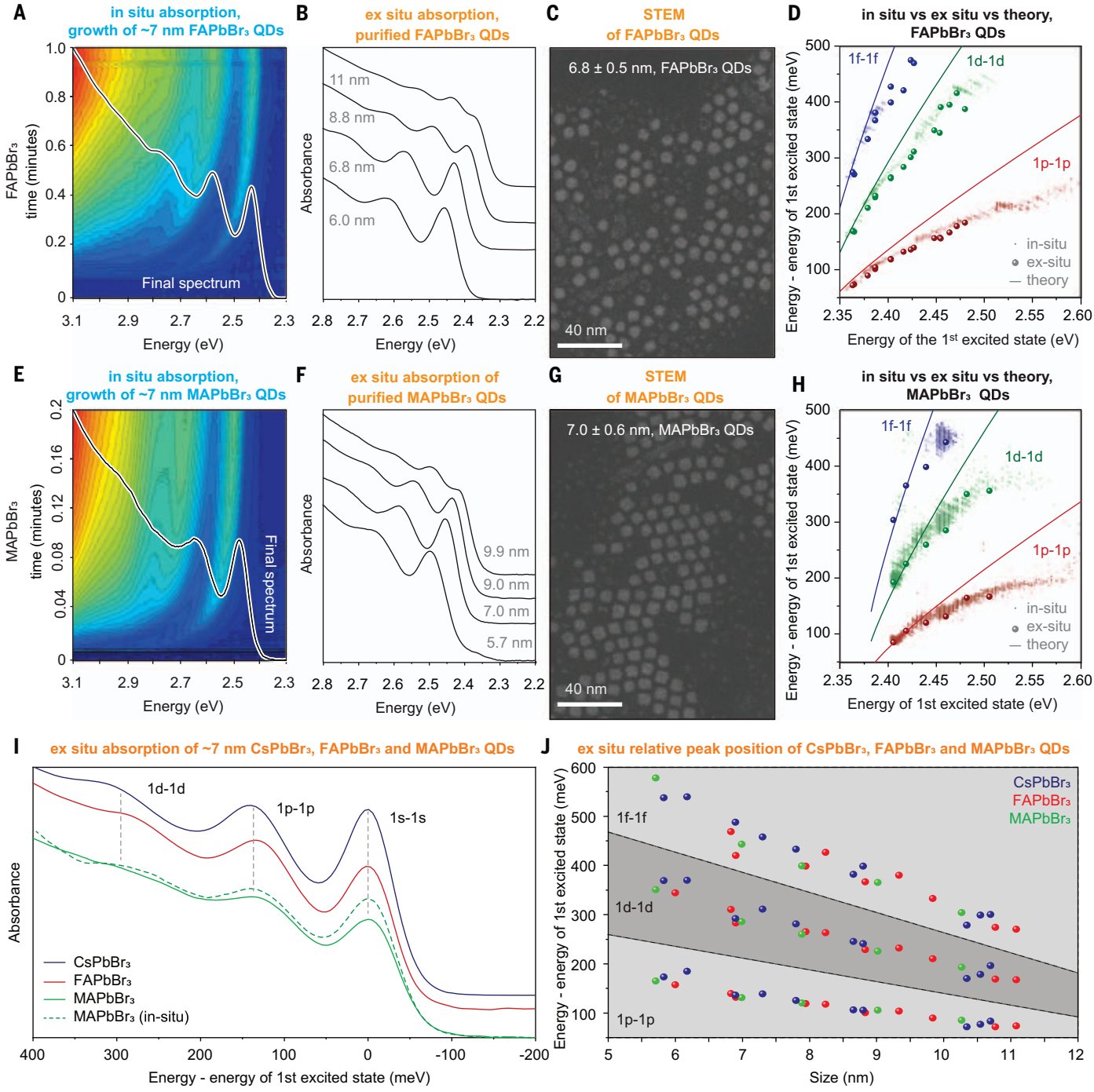



**Fig. 5. Monodisperse $FAPbBr_3$ and $MAPbBr_3$ QDs and comparison with $CsPbBr_3$ QDs.** (**A** and **E**) In situ spectra during the formation of 7-nm QDs. (**B** and **F**) Absorption spectra of isolated series. (**C** and **G**) STEM images of ~7-nm QDs. (**D** and **H**) In situ, ex situ, and calculated excitonic transition energies plotted with respect to their first excitonic transition energy. (**I**) Absorption spectra of ~7-nm $CsPbBr_3$, $FAPbBr_3$, and $MAPbBr_3$ QDs with matched 1s-1s transition energy. (**J**) Experimental size-dependent relative peak position of $CsPbBr_3$, $FAPbBr_3$, and $MAPbBr_3$ QDs.

finely controlled in situ generation of monomers—for instance, by molecular engineering of the library of precursors—have been successful for metal chalcogenide NCs (*34*).

### Optical properties of $CsPbBr_3$ QDs

Higher-order excitonic transitions, along with their size dependency and shape dependency, have been assigned in CdSe and PbS QDs (*35*, *36*), but the corresponding datasets for $CsPbBr_3$ QDs have been very limited, and typically, only the lowest two transitions (1s-1s and 1p-1p) are reported (*21*, *37*). We were able to resolve the first four excitonic transitions readily resolved both in situ (various final QD sizes are shown in Fig. 4A and movie S4) and ex situ linear absorption spectra (Fig. 4B). Size dependencies captured by 25,000 in situ experimental spectra recorded for 2.5-to-11-nm QDs (examples in Fig. 4C) were retained in the purified QDs (Fig. 4D).

To assign the absorption features, the specific excitonic transitions were calculated with a two-band **k·p** model (*21*, *37*). The Coulomb

interaction was included under the self-consistent Hartree-Fock approximation. For the 1s-1s, 1p-1p, 1d-1d, and 1f-1f transitions, we found good agreement of the theoretical and experimental transition energies, plotted relative to the first excited state (1s-1s) as a function of their first excited state energies (Fig. 4F) (*35*). The spheroidal $CsPbBr_3$ QDs optically behaved as confined model systems with symmetrical electron and hole manifolds, which had approximately equal effective masses. The theoretical absorption spectrum was then obtained from calculations of the oscillator strength for all allowed transitions by using second-order many-body perturbation theory, which allowed for partial incorporation of correlation effects in the electron-photon interaction (*32*, *38*). Each transition was broadened with a normalized Voigt line shape, and the results for all transitions were added to obtain the final absorption spectra shown, for instance, for 7.0 nm (Fig. 4G and fig. S32).

This model was further extended for a broad range of temperatures between 14 and 300 K, as shown in fig. S33. More discussion on the broadening mechanisms is found in the supplementary text and fig. S34. The calculated low-temperature absorption spectrum was compared to the PL excitation (PLE) spectrum at 14 K (Fig. 4H). Whereas only one 1p-1p state transition was expected and observed at room temperature, the narrowing of the absorption features upon cooling to 14 K revealed two 1p-1p contributions, separated by ~70 meV. This splitting was observed in temperature-dependent absorption spectra for various QD sizes (fig S35) and was previously ascribed to the shape asymmetry in cuboidal QDs (oblation in one direction) (*21*, *37*).

A likewise observation in isotropic spheroidal QDs called for other considerations, such as anisotropies of the effective mass of electrons or holes that arose from the orthorhombic crystal structure of the $CsPbBr_3$ (*39*, *40*). Such anisotropic effects were already inferred from electrical characteristics (*41*). Inclusion of 10 to 15% of anisotropy in the mass of the electron and hole gave rise to the splitting of the 1p-1p transition with the magnitude found experimentally (figs. S36 to S38). The splitting of the 1p states potentially had a fairly complex dependence on the QD size and temperature; the supplementary text contains further discussion.

Simultaneous in situ acquisition of both absorption and PL spectra allowed us to study the inherent size-dependent Stokes shift (fig. S39 and movies S1 and S5) with excellent energy accuracy (standard deviation of only ~2% in the range of 4 to 10 nm; fig. S40). The measured Stokes shift agreed with recent calculations that assume an intrinsic confined hole state with size-dependent alignment above the valence band maximum (*42*).

### Extension to $FAPbBr_3$ and $MAPbBr_3$ QDs

We extended the synthesis methodology to hybrid small organic-inorganic LHP QDs (Fig. 5). $FAPbBr_3$ QDs were synthesized by replacing $Cs_2CO_3$ with FA-acetate and DOPA with a mixture of DOPA and OA. $FAPbBr_3$ QDs were observed to form an order of magnitude faster than $CsPbBr_3$ QDs (minutes time scale, see Fig. 5A for in situ absorption in the synthesis of 7-nm $FAPbBr_3$ QDs). Analogous experiments with MA-acetate gave rise to a size series of monodisperse $MAPbBr_3$ NCs, albeit with further reduction of the synthesis time to ~10 s (Fig. 5E). Such a pronounced effect of the A-site cation on the reaction rate agreed with the proposed reaction mechanism (Fig. 3C), wherein coordination of A-cation to $PbBr_3^-$ was hypothesized to govern the stability and hence reactivity of these transient species. Size series of 6-to-11-nm lecithin-capped $FAPbBr_3$ and $MAPbBr_3$ QDs (Fig. 5, B and F), which were found to be quasi-spherical in shape (Fig. 5, C and G), were analogous to $CsPbBr_3$ QDs. These hybrid LHP NCs had narrow size dispersion of <10% and well-resolved excitonic peaks in both in situ and ex situ measurements (Fig. 5, D and H, and fig. S41). Notably, the faster formation kinetics did not jeopardize the utility of the synthesis, as the QD size control is accomplished at a full reaction yield by adjusting the concentrations, not the growth time. We compared all three LHP homologs, as they featured an analogous Pb-halide octahedral framework. As illustrated in Fig. 5I, the 7-nm $CsPbBr_3$, $FAPbBr_3$, and $MAPbBr_3$ QDs exhibited nearly identical absorption spectra when matching the positions of their first excitonic transition. Moreover, across the experimental and theoretical size series, all three materials exhibited the same effect of quantum confinement on higher-order excitonic transitions (Fig. 5J and fig. S42). High PL QY (80 to 100%) at room temperature, along with sharp and tunable excitonic features and facile processing into diverse photonic structures, makes small LHP QDs an attractive emitter of choice for studying strong light-matter interaction, such as exciton-polariton condensates and polaritonic lasing.

### REFERENCES AND NOTES


1. Q. A. Akkerman, G. Rainò, M. V. Kovalenko, L. Manna, *Nat. Mater.* **17**, 394–405 (2018).
2. A. Dey *et al.*, *ACS Nano* **15**, 10775–10981 (2021).
3. Y.-H. Kim *et al.*, *Nat. Photonics* **15**, 148–155 (2021).
4. Y. Hassan *et al.*, *Nature* **591**, 72–77 (2021).
5. Avantama; www.avantama.com.
6. Q. Zhang, Q. Shang, R. Su, T. T. H. Do, Q. Xiong, *Nano Lett.* **21**, 1903–1914 (2021).
7. Q. Chen *et al.*, *Nature* **561**, 88–93 (2018).
8. M. Gandini *et al.*, *Nat. Nanotechnol.* **15**, 462–468 (2020).
9. M. Wei *et al.*, *Nat. Energy* **4**, 197–205 (2019).
10. J. Wu *et al.*, *Angew. Chem. Int. Ed.* **59**, 7738–7742 (2020).
11. M. A. Becker *et al.*, *Nano Lett.* **18**, 7546–7551 (2018).
12. G. Rainò *et al.*, *ACS Nano* **10**, 2485–2490 (2016).
13. H. Utzat *et al.*, *Science* **363**, 1068–1072 (2019).
14. P. Tamarat *et al.*, *Nat. Mater.* **18**, 717–724 (2019).
15. Y.-S. Park, S. Guo, N. S. Makarov, V. I. Klimov, *ACS Nano* **9**, 10386–10393 (2015).
16. C. Yin *et al.*, *Phys. Rev. Lett.* **119**, 026401 (2017).
17. K. Cho *et al.*, *Nano Lett.* **21**, 7206–7212 (2021).
18. G. Rainò *et al.*, *Nature* **563**, 671–675 (2018).
19. F. Mattiotti, M. Kuno, F. Borgonovi, B. Jankó, G. L. Celardo, *Nano Lett.* **20**, 7382–7388 (2020).
20. I. Cherniukh *et al.*, *Nature* **593**, 535–542 (2021).
21. F. Krieg *et al.*, *ACS Cent. Sci.* **7**, 135–144 (2021).
22. J. De Roo *et al.*, *ACS Nano* **10**, 2071–2081 (2016).
23. Y. Dong *et al.*, *Nano Lett.* **18**, 3716–3722 (2018).
24. M. Imran *et al.*, *Nano Lett.* **18**, 7822–7831 (2018).
25. L. Protesescu *et al.*, *Nano Lett.* **15**, 3692–3696 (2015).
26. G. Almeida *et al.*, *ACS Nano* **12**, 1704–1711 (2018).
27. J. Hui *et al.*, *Chem. Mater.* **32**, 4574–4583 (2020).
28. G. Almeida *et al.*, *J. Am. Chem. Soc.* **140**, 14878–14886 (2018).
29. Y. Shynkarenko *et al.*, *ACS Energy Lett.* **4**, 2703–2711 (2019).
30. F. Krieg *et al.*, *J. Am. Chem. Soc.* **141**, 19839–19849 (2019).
31. F. Krieg *et al.*, *ACS Energy Lett.* **3**, 641–646 (2018).
32. P. C. Sercel, J. L. Lyons, N. Bernstein, A. L. Efros, *J. Chem. Phys.* **151**, 234106 (2019).
33. S. J. Yoon, K. G. Stamplecoskie, P. V. Kamat, *J. Phys. Chem. Lett.* **7**, 1368–1373 (2016).
34. M. P. Hendricks, M. P. Campos, G. T. Cleveland, I. Jen-La Plante, J. S. Owen, *Science* **348**, 1226–1230 (2015).
35. D. J. Norris, M. G. Bawendi, *Phys. Rev. B Condens. Matter* **53**, 16338–16346 (1996).
36. A. D. Andreev, A. A. Lipovskii, *Phys. Rev. B Condens. Matter* **59**, 15402–15404 (1999).
37. W. Shcherbakov-Wu, P. C. Sercel, F. Krieg, M. V. Kovalenko, W. A. Tisdale, *J. Phys. Chem. Lett.* **12**, 8088–8095 (2021).
38. T. P. T. Nguyen, S. A. Blundell, C. Guet, *Phys. Rev. B* **101**, 195414 (2020).
39. T. Sakuma *et al.*, *Solid State Ion.* **154-155**, 237–242 (2002).
40. B. Traoré *et al.*, *Phys. Rev. B* **99**, 035139 (2019).
41. P. Zhang *et al.*, *J. Phys. Chem. Lett.* **9**, 5040–5046 (2018).
42. M. C. Brennan *et al.*, *J. Am. Chem. Soc.* **139**, 12201–12208 (2017).


### ACKNOWLEDGMENTS


We thank M. Blankenburg and U. Lienert for assistance during the SAXS and WAXS measurements. We thank F. Krumeich and I. Cherniukh for HAADF-STEM measurements and F. Krieg and I. Cherniukh for providing reference cuboidal $CsPbBr_3$ NCs. **Funding:** This work was financially supported by the European Union through the Horizon 2020 Research and Innovation Programme through the European Research Commission (ERC-CoG, agreement no. 819740, project SCALE-HALO) and FET Open Research and Innovation Action (grant agreement no. 899141, project PoLLoC) and by the Air Force Office of Scientific Research and the Office of Naval Research (award no. FA8655-21-1-7013), and the Institute Universitaire de France. SAXS and WAXS measurements were carried out at beamline P21.2 at PETRA III at DESY, a member of the Helmholtz Association (HGF); the research leading to this result has been supported by the project CALIPSOplus (grant agreement 730872) from the EU Framework Programme for Research and Innovation Horizon 2020. **Author contributions:** Q.A.A. performed the synthesis and optical characterization, with contributions from P.W., F.B., and D.N.D.; T.P.T.N, C.K., and J.E. carried out theoretical calculations. S.C.B. performed the low-temperature PLE measurements. F.M. carried out the WAXS and SAXS experiments. R.E. conducted electron microscopy. Q.A.A. and M.V.K. wrote the manuscript, with input from all authors. M.V.K, G.R., C.K., and J.E. supervised the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary material.



### SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq3616
Materials and Methods
Supplementary Text
Figs. S1 to S42
Tables S1 and S2
References (*43*–*64*)
Movies S1 to S5


Submitted 3 April 2022; accepted 24 August 2022
10.1126/science.abq3616

DEVELOPMENTAL BIOLOGY

# Quality assurance of hematopoietic stem cells by macrophages determines stem cell clonality

Samuel J. Wattrus[1,2], Mackenzie L. Smith[1,2], Cecilia Pessoa Rodrigues[1,2], Elliott J. Hagedorn[1,2]†, Ji Wook Kim[1,2], Bogdan Budnik[3], Leonard I. Zon[1,2]*

Tissue-specific stem cells persist for a lifetime and can differentiate to maintain homeostasis or transform to initiate cancer. Despite their importance, there are no described quality assurance mechanisms for newly formed stem cells. We observed intimate and specific interactions between macrophages and nascent blood stem cells in zebrafish embryos. Macrophage interactions frequently led to either removal of cytoplasmic material and stem cell division or complete engulfment and stem cell death. Stressed stem cells were marked by surface Calreticulin, which stimulated macrophage interactions. Using cellular barcoding, we found that Calreticulin knock-down or embryonic macrophage depletion reduced the number of stem cell clones that established adult hematopoiesis. Our work supports a model in which embryonic macrophages determine hematopoietic clonality by monitoring stem cell quality.

Tissue stem cells born during embryogenesis support homeostasis for life. Despite the importance of these cells for proper tissue function, there are no described quality assurance mechanisms for newly formed stem cells. To explore this possibility, we studied zebrafish embryonic blood development. Hematopoietic stem and progenitor cells (HSPCs) emerge from the ventral wall of the dorsal aorta (VDA), enter circulation, and lodge in the embryonic niche—a vascular plexus called the caudal hematopoietic tissue (CHT) (*1*, *2*). HSPCs rapidly expand in the CHT for 3 to 4 days before migrating to the kidney marrow, the adult hematopoietic niche. In both niches, HSPCs interact with a variety of cell types, including vascular endothelial cells, mesenchymal stromal cells, and macrophages (MΦs) (*3*–*6*). In vivo clonal labeling shows that 20 to 30 of the hematopoietic stem cell (HSC) clones born in the VDA ultimately give rise to the adult blood system (*7*). It remains unclear whether nascent HSCs from the VDA undergo quality assurance before establishing adult hematopoiesis. Here, using live imaging and cellular barcoding, we found discrete interactions between stem cells and embryonic macrophages that regulated the number of long-lived HSC clones that produce blood in adulthood.

## Results

### *Macrophages interact with nascent HSPCs in the CHT*

Macrophages help maintain homeostasis by modulating inflammation, producing cytokines, and patrolling to clear dead, stressed, or aged cells (*8*–*10*). Given these roles in somatic tissue, the enrichment of macrophages in the CHT, and previous observations between macrophages and hematopoietic cells (*6*), we investigated macrophage function in the niche. We undertook high-resolution live imaging using *mpeg1:mCherry;runx1+23:EGFP* zebrafish embryos with mCherry$^+$ macrophages and enhanced green fluorescent protein–positive (EGFP$^+$) HSPCs (*4*, *11*). Shortly after lodgment in the CHT, HSPCs were contacted by a nearby macrophage and their surfaces were scanned. These interactions sometimes resulted in the uptake of fluorescent HSPC material by the macrophage (Fig. 1A and movie S1). From 56 to 106 hours postfertilization (hpf), ~20 to 30% of HSPCs were engaged by a macrophage at any time point (Fig. 1B). These interactions were specific to HSPCs; macrophage engagement with erythrocytes and endothelial cells was considerably lower (0.6 to 3.9% of erythrocytes and 0.5 to 6.7% of endothelial cells) (fig. S1A). Macrophages contacted HSPCs for up to 45 min, sometimes taking up fluorescent HSPC material. We classified interactions into three types: prolonged cell-cell contact; "grooming," during which the HSPC was left intact but had a small portion of cellular material taken up by the macrophage; or "dooming," during which the HSPC was fully engulfed and destroyed by the macrophage (movie S2). We also found similar interactions with other HSPC reporters *cd41:GFP* and *cmyb:GFP* (fig. S1, D and E). To examine whether macrophage-HSPC interactions occurred in mammals, we studied embryonic day 14.5 (E14.5) murine fetal liver sections by immunofluorescence and found that 33% of c-Kit$^+$ hematopoietic cells were in contact with F4/80$^+$ macrophages. This included c-Kit$^+$ cells being pinched or fully engulfed by macrophages, similar to our observations in zebrafish (fig. S1F). Overall, these data identify previously uncharacterized macrophage-HSPC interactions in the embryonic hematopoietic niche.

To better characterize macrophage-HSPC interactions, we tracked individual HSPCs in the CHT at 2 or 3 days postfertilization (dpf) and recorded macrophage interactions. We found that 70% of HSPCs experienced prolonged macrophage contact over a 3-hour imaging period (Fig. 1C). Within this time frame, 13% of these HSPCs were groomed and 13% were doomed. Some HSPCs were contacted by macrophages multiple times and underwent grooming or dooming after repeated interaction, suggesting that most HSPCs may eventually undergo grooming or dooming at some point during the 3 to 4 days that they occupy the CHT. Of HSPC divisions, 81% occurred within 30 min of grooming or prolonged contact (Fig. 1D). Using the *Tg[EF1a:mAG-zGem(1/100)]*$^{rw0410h}$ (Fucci) transgene (*12*) to label cells in S, $G_2$, and M phases of the cell cycle, we found that ~65% of Fucci$^+$ HSPCs contacted macrophages, compared with only 16% of Fucci$^-$ HSPCs (Fig. 1E). We next assessed the viability of HSPCs engulfed by macrophages. Staining for cell death with acridine orange or an annexin V–yellow fluorescent protein (YFP) construct (*13*) showed almost no apoptotic HSPCs in the CHT that were not already engulfed by macrophages (figs. S1, B and C). Only after full engulfment did HSPCs exhibit apoptosis (movie S3). Together, these data identify a set of macrophage-HSPC interactions that precede either HSPC division or death.

### *A subset of primitive macrophages regulates stem cell clone number*

Because we saw proliferation after macrophage-HSPC interactions, we next sought to determine whether this might influence the number of stem cell clones that contribute to adult hematopoiesis. We used TWISTR (tissue editing with inducible stem cell tagging via recombination) (*14*) to combine morpholino-mediated gene knock-down with Zebrabow HSC color labeling. *Zebrabow-M;draculin:CreER*$^{T2}$ embryos enable unique lineage labeling of individual HSC clones at 24 hpf (Fig. 2A) (*7*, *15*). To deplete embryonic macrophages, we injected the *irf8* morpholino to block macrophage formation (*16*) or delivered clodronate liposomes to ablate macrophages at various time points: 28 hpf, before HSPC emergence in the VDA; 48 hpf, before HSPC lodgment in the CHT; 72 hpf, after HSPC lodgment in the CHT; 96 hpf, after HSPCs have doubled (*4*); or 120 hpf, as HSPCs start to colonize the marrow. Zebrabow analysis of adult marrow myelomonocytes revealed a consistent reduction in hematopoietic clonality compared with sibling controls when macrophages were depleted before 96 hpf (Fig. 2B). These results demonstrate that embryonic macrophages regulate HSC clone number

[1]Howard Hughes Medical Institute, Boston Children's Hospital Boston, MA, USA. [2]Harvard Stem Cell Institute, Stem Cell and Regenerative Biology Department, Harvard University, Cambridge, MA, USA. [3]Mass Spectrometry and Proteomics Resource Laboratory, Faculty of Arts and Sciences Division of Science, Harvard University, Cambridge, MA, USA.
*Corresponding author. Email: zon@enders.tch.harvard.edu
†Present address: Section of Hematology and Medical Oncology and Center for Regenerative Medicine, Boston University School of Medicine and Boston Medical Center, Boston, MA, USA.

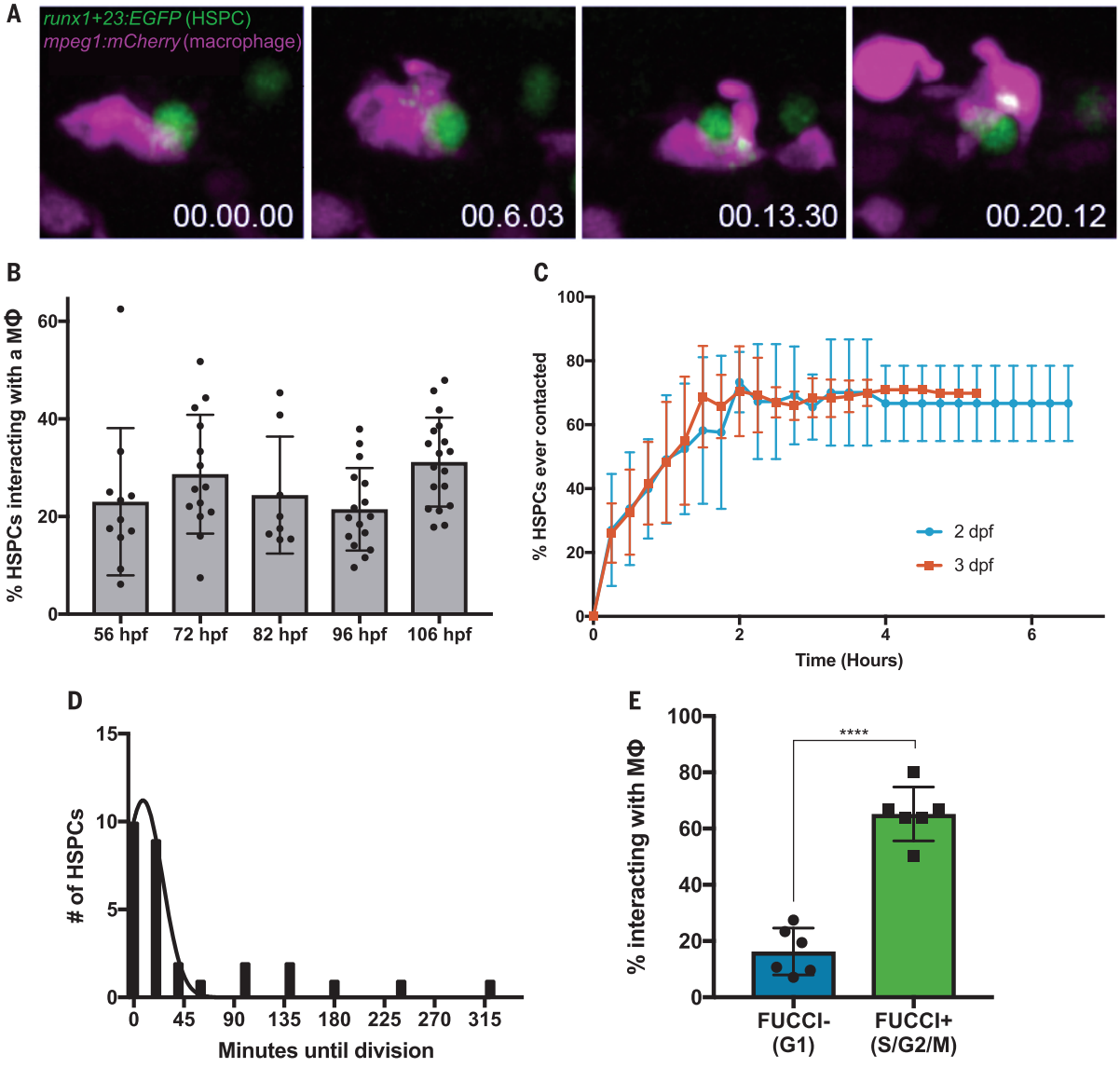



**Fig. 1. Macrophages make intimate interactions with newly formed HSPCs.** (**A**) Time-lapse live imaging identifies prolonged cell-cell contacts between *runx1+23:EGFP*$^+$ HSPCs and *mpeg1:mCherry*$^+$ primitive macrophages that involve the exchange of fluorescent material. (**B**) About 20 to 30% of HSPCs interact with macrophages in the CHT at any one time from 56 to 106 hpf. Data are means ± SD. (**C**) High-resolution tracking of individual *runx1+23:mCherry*$^+$ cells over several hours in the CHT reveals that most HSPCs eventually make sustained contact with macrophages (>5 min). Data are means ± SD. (**D**) HSPCs frequently complete a cell division shortly after macrophage interactions. About 81% of HSPC divisions occur within 30 min of a macrophage interaction. (**E**) About 65% of Fucci$^+$ HSPCs in S, $G_2$, and M phases interact with macrophages at any one time, as compared to less than 20% of Fucci$^-$ HSPCs. Data are means ± SD. Data were analyzed by unpaired Student's *t* test; ****$P < 0.0001$.

after VDA emergence and niche colonization until at least one round of amplification has completed.

To better understand the mechanism and cellular consequences of macrophage-HSPC interactions, we pursued transcriptomic analysis of niche macrophages. Because macrophages can take up fluorescent material from HSPCs, we reasoned that it would be possible to identify interacting macrophages by their fluorescence profile. Indeed, flow cytometry of dissociated *mpeg1:EGFP;runx1+23:mCherry* embryos revealed a rare population of EGFP$^+$mCherry$^+$ cells that was morphologically consistent with macrophages containing HSPC fragments (Fig. 2C). We dissected embryonic zebrafish tails and purified interacting macrophages (EGFP$^+$mCherry$^+$) and noninteracting macrophages (EGFP$^+$mCherry$^-$) for single-cell RNA sequencing (scRNA-seq). We identified a single population of macrophages that segregated by both gene expression and mCherry fluorescence (Fig. 2D). These cells were enriched for genes associated with engulfment, lysosomal degradation, and cholesterol transport and were marked by genes including *hmox1a*, *ctsl.1*, *slc40a1*, *lrp1ab*, and *c1qa* (Fig. 2D and fig. S2A). We validated these data with a fluorescent cholesterol mimic, LysoTracker dye, and in situ hybridization (fig. S2, B to D). Together, these data show that a transcriptionally distinct and relatively homogeneous subset of macrophages engage HSPCs in the CHT.

### *Surface Calreticulin drives macrophage-HSPC interactions*

To gain insight into the proteinaceous material taken up by macrophages, we pursued a modified form of single-cell proteomics called few-cell proteomics (*17*) to compare interacting macrophages with noninteracting macrophages.

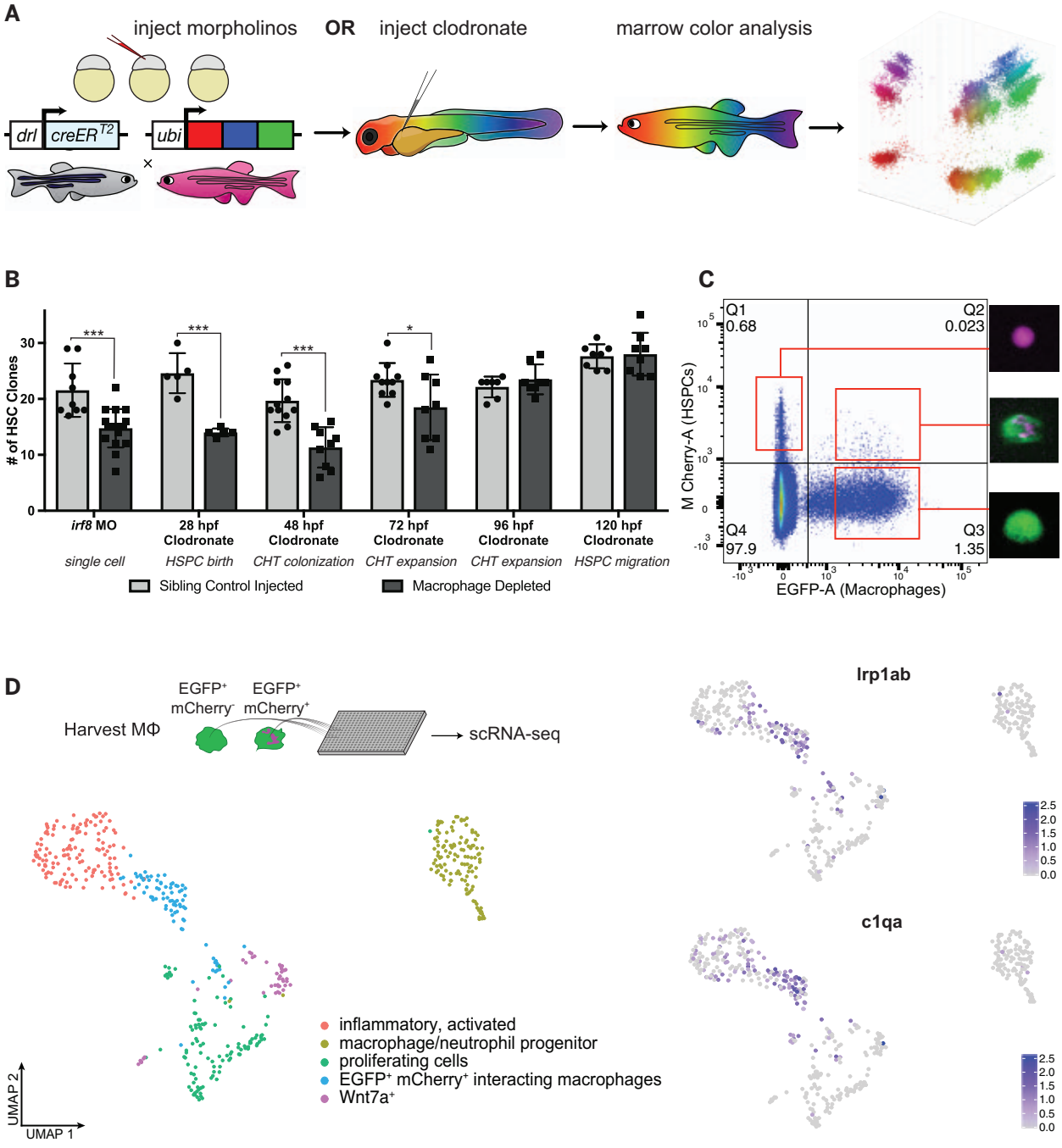



**Fig. 2. Macrophages in the CHT regulate stem cell clonality.** (**A**) A schematic overview of the *Zebrabow-M* system: Animals with 15 to 20 insertions of a multicolor fluorescent cassette are crossed to the *draculin:CreER*$^{T2}$ line to enable clonal labeling of lateral plate mesoderm lineages. By treating with 4-hydroxytamoxifen (4-OHT) at 24 hpf just after HSC specification, individual stem cell lineages express specific fluorescent hues that can be quantified in the adult marrow. (**B**) Families of *Zebrabow-M;draculin:CreER*$^{T2}$ animals injected with either clodronate liposomes or the *irf8* morpholino exhibit reduced numbers of HSC clones in the adult marrow, even when macrophages are not depleted until after emergence from the VDA. Data are means ± SD. Data were analyzed by unpaired Student's *t* test; **P* < 0.05 and ****P* < 0.001. (**C**) Macrophages (*mpeg1:EGFP*$^{+}$) that have interacted with HSPCs (*runx1+23:mCherry*$^{+}$) and removed fluorescent material can be harvested by fluorescence-activated cell sorting (FACS). (**D**) Macrophages that engage HSPCs are marked by *lrp1ab* and *c1qa*. The spectral scale reports *z*-scores. UMAP, uniform manifold approximation and projection.

We identified 203 peptides enriched in interacting macrophages, potentially representing a repertoire of proteins either involved in the process of macrophage-HSPC interaction or taken directly from HSPCs. To identify molecular patterns recognized on HSPCs, we excluded peptides with enriched transcripts in interacting macrophages and compared the remaining peptides with the Cell Surface Protein Atlas (*18*). Notably, surface peptides enriched in interacting macrophages included three Calreticulin paralogs: *calr*, *calr3a*, and *calr3b* (Fig. 3A). Though Calreticulin is widely expressed and typically functions as a chaperone protein in the endoplasmic reticulum (ER),

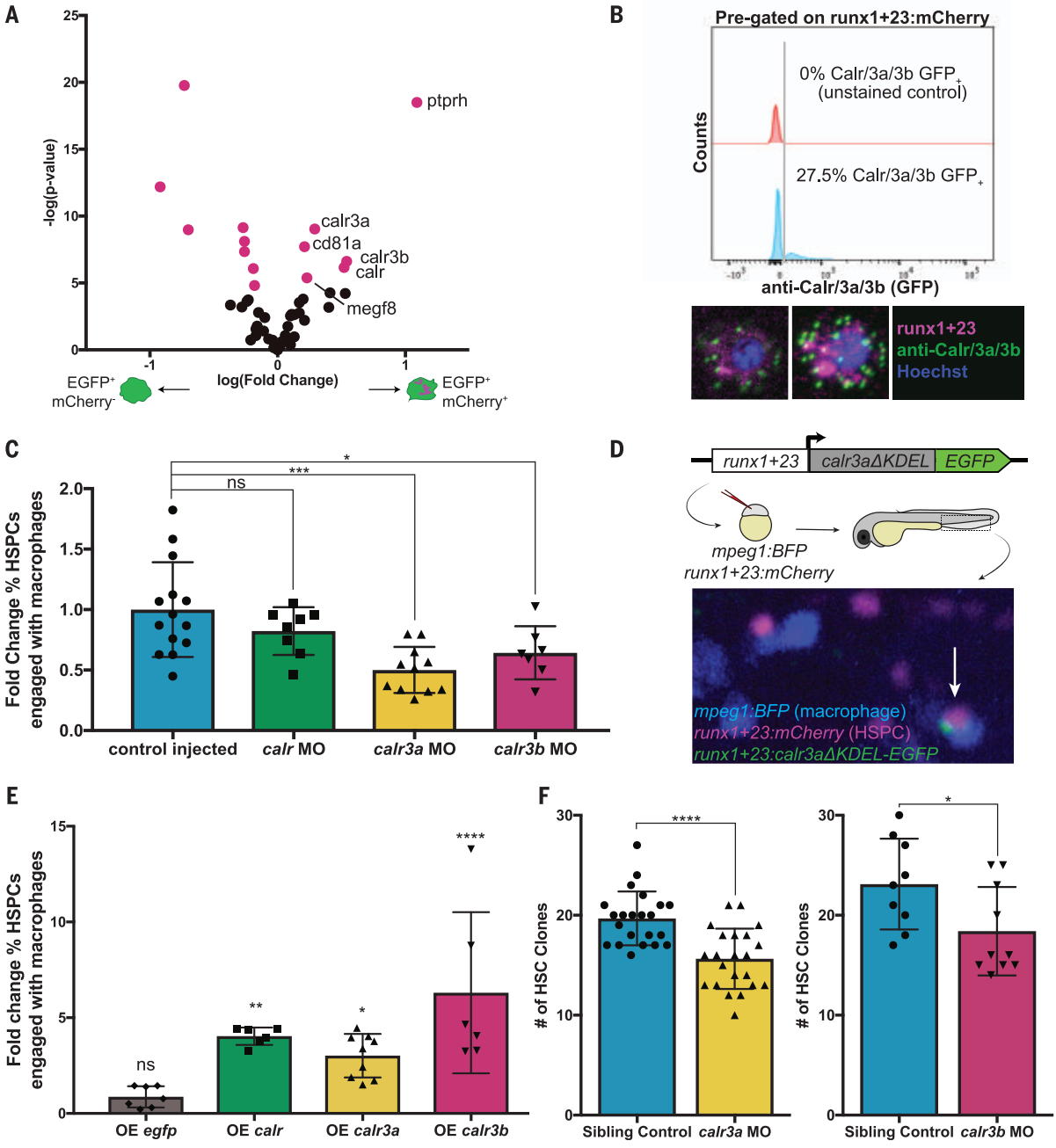



**Fig. 3. Calreticulin drives HSPC-macrophage interactions to regulate clonality.** (**A**) Analysis of differentially enriched potential surface proteins from interacting macrophages identifies three paralogs of Calreticulin. (**B**) Flow cytometry shows that ~30% of *runx1+23:mCherry*$^{+}$ HSPCs stain for surface Calreticulin. (**C**) Morpholino knock-down of *calr3a* or *calr3b* significantly reduces the fraction of HSPCs interacting with macrophages at any one time. Data are means ± SD. Data were analyzed by one-way ANOVA with Dunnett's multiple comparisons test; ns is not significant, **P* < 0.05, and ****P* < 0.001. MO, morpholino. (**D**) Calreticulin paralogs without the ER-retention KDEL sequence were fused to EGFP, driven by the HSPC-specific *runx1+23* enhancer, and injected into stable *runx1+23:mCherry;mpeg1:BFP* embryos. Mosaic animals overexpress Calreticulin in a random subset of HSPCs. The arrow indicates an HSPC overexpressing *calr3a* engaged by a macrophage. (**E**) HSPCs overexpressing *calr*, *calr3a*, or *calr3b* are more frequently engaged by macrophages compared with non-overexpressing HSPCs in the same embryos. Overexpressing *egfp* alone has no effect. Data are means ± SD. Data were analyzed by one-way ANOVA with Dunnett's multiple comparisons test; ns is not significant, **P* < 0.05, ***P* < 0.01, and *****P* < 0.0001. (**F**) Knock-down of *calr3a* or *calr3b* reduces the number of HSC clones that contribute to adult hematopoiesis. Data are means ± SD. Data were analyzed by unpaired Student's *t* test; **P* < 0.05 and *****P* < 0.0001.

it can also sometimes be displayed on the cell surface as an "eat-me" signal (*9*, *10*, *19*). Based on our proteomic results, we hypothesized that HSPCs could display surface Calreticulin, stimulating macrophage interactions. We found that 30% of HSPCs at 72 hpf exhibited classic punctate surface Calreticulin staining (*20*) (Fig. 3B), similar to the percentage of HSPCs interacting with macrophages in vivo (Fig. 1B). Additionally, the canonical surface Calreticulin binding partners, *lrp1ab* and *c1qa*, were transcriptionally enriched in interacting macrophages (Fig. 2D). Together, Lrp1ab and C1qa contact Calreticulin and form a bridging complex to initiate phagocytic activity (*10*, *20*, *21*).

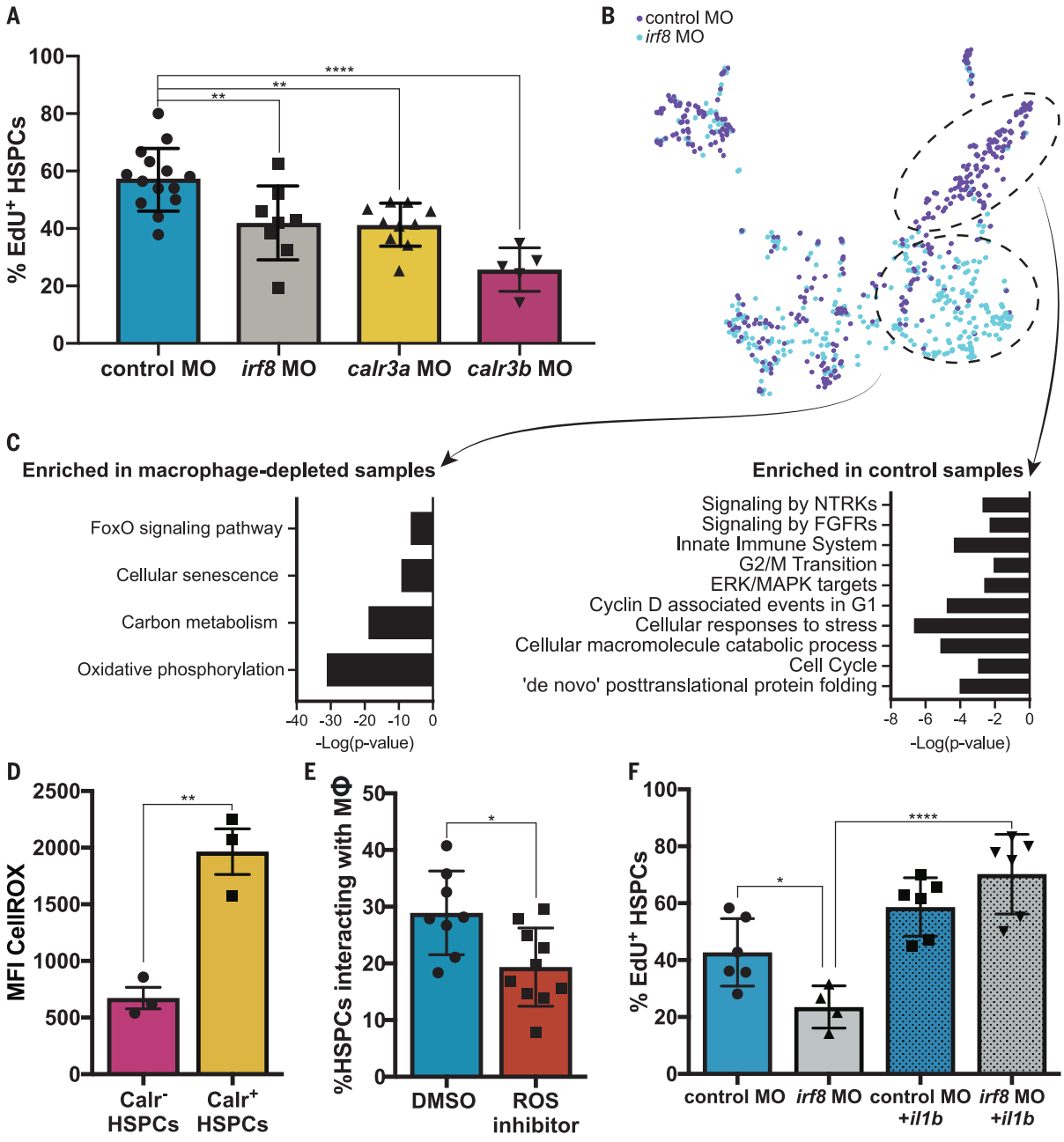



**Fig. 4. Macrophages buffer HSPC stress and regulate HSPC expansion.** (**A**) EdU staining of *runx1+23:mCherry* embryos injected with either the *calr3a*, *calr3b*, or *irf8* morpholinos identifies a significant reduction in proliferating HSPCs at 3 dpf. Data are means ± SD. Data were analyzed by one-way ANOVA with Dunnett's multiple comparisons test; $^{**}P < 0.01$ and $^{****}P < 0.0001$. (**B** and **C**) scRNA-seq analysis of *runx1+23*$^{+}$ FACS-purified cells from *irf8* or control morphants reveals a population of stressed HSPCs that persist in the absence of macrophages and a population of cycling cells enriched in the control sample. (**D**) Embryonic HSPCs marked by surface Calreticulin exhibit higher levels of ROS. Data are means ± SD. Data were analyzed by unpaired Student's *t* test; $^{**}P < 0.01$. MFI, median fluorescence intensity. (**E**) ROS inhibition with diphenylene iodonium significantly reduces macrophage-HSPC interactions. Data are means ± SD. Data were analyzed by unpaired Student's *t* test; $^{*}P < 0.05$. DMSO, dimethyl sulfoxide. (**F**) Expression of *il1b* by heat shock rescues the effect of macrophage depletion on HSPC proliferation. Data are means ± SD. Data were analyzed by one-way ANOVA with Sidak's multiple comparisons test; $^{*}P < 0.05$ and $^{****}P < 0.001$.

These results show that Calreticulin decorates the surface of HSPCs and may promote macrophage interaction.

To study the role of Calreticulin in macrophage-HSPC interactions, we used morpholinos to knock down Calreticulin gene expression. Knockdown of *calr3a* or *calr3b* significantly reduced the percentage of HSPCs engaged by macrophages (Fig. 3C). This effect was reversed in genetic rescue experiments (fig. S3, A and B). We then generated parabiotic fusions of embryos with or without Calreticulin knock-down and found that knock-down HSPCs had reduced interactions with control macrophages. By contrast, control HSPCs had normal levels of interaction with Calreticulin knock-down macrophages, indicating that Calreticulin presentation is HSPC autonomous (fig. S3, C and D). Next, we tested the effect of constitutively surface-translocated Calreticulin expressed under the HSPC-specific *runx1+23* enhancer

(*4*) (Fig. 3D). Injecting this construct into early embryos generated mosaic animals, which permitted direct comparison of HSPCs with or without Calreticulin overexpression. HSPCs overexpressing *calr*, *calr3a*, or *calr3b* were 3- to 5-fold more likely to interact with macrophages compared with non-overexpressing HSPCs in the same embryo (Fig. 3E). When *calr3a* or *calr3b* were knocked down, prolonged contact, grooming, and dooming interactions all decreased, with a more severe decrease in dooming (fig. S3E). By contrast, nearly all cells overexpressing Calreticulin were doomed (fig. S3F). Taken together, these data show that surface Calreticulin promotes macrophage-HSPC interactions and suggest that differing levels of Calreticulin determine whether an HSPC experiences prolonged contact, is groomed, or is doomed.

To determine whether Calreticulin-dependent interactions during development were responsible for regulating HSC clonality into adulthood, we knocked down *calr3a* or *calr3b* and color labeled HSCs at 24 hpf. Adult Zebrabow analysis of morphants showed a significant reduction in the number of HSC clones compared with sibling controls (Fig. 3F). These data show that Calreticulin-dependent interactions in development support a greater number of long-lived HSC clones.

### *Macrophages buffer HSPC stress and promote divisions*

We next assessed Calreticulin function in HSPC development. To analyze changes to HSPC emergence, we injected the *irf8, calr3a*, or *calr3b* morpholino into *cd41:GFP;kdrl:mCherry* embryos to visualize the endothelial-to-hematopoietic transition (*22*). Quantification of $EGFP^{+}mCherry^{+}$ cells in the VDA revealed no significant difference in HSPC budding (fig. S4A). Serial imaging of *cd41:GFP*$^{+}$ cells over early development revealed that knockdown of *calr3a* or *calr3b* did not affect HSPC numbers through 60 hpf but later reduced HSPCs at 72 and 84 hpf (fig. S4, B and C). This was not due to apoptosis or altered trafficking to the kidney marrow (fig. S4, D and E). Rather, depletion of macrophages or knockdown of *calr3a* or *calr3b* significantly reduced the fraction of proliferative HSPCs in the CHT at 72 hpf, as measured by 5-ethynyl-2′-deoxyuridine (EdU) incorporation (Fig. 4A). This corroborates the association of macrophage-HSPC interactions with HSPC division that was identified by live imaging (Fig. 1, D and E). These data show that Calreticulin-dependent macrophage-HSPC interactions serve to expand and maintain HSPCs during early development by promoting proliferation in the CHT.

To molecularly evaluate the effect of macrophage interactions on HSPCs and the qualities that lead to surface Calreticulin, we injected *runx1+23:mCherry* embryos with the *irf8* or control morpholino and performed scRNA-seq on HSPCs at 72 hpf. This analysis identified a population of HSPCs enriched in *irf8* morphants marked by genes associated with FoxO activity and cellular senescence (Fig. 4, B and C). FoxO activity initiates in response to increased reactive oxygen species (ROS) and mediates detoxification of ROS and repair of ROS-induced damage (*23*, *24*). In murine HSCs, FoxO deletion and ROS accumulation results in dysregulation of apoptosis, cell cycling, and colony formation (*25*). The enrichment of HSPCs with FoxO activity in *irf8* morphants suggested potential ROS accumulation, which is ordinarily resolved by macrophages. In agreement with this, flow cytometric analysis showed higher ROS levels in HSPCs marked by surface Calreticulin, with significant correlation between ROS levels and surface Calreticulin intensity (Spearman's correlation; $^{****}P < 2.2 \times 10^{-16}$) (Fig. 4D). Inhibiting ROS with diphenylene iodonium reduced macrophage-HSPC interactions (Fig. 4E), whereas increasing ROS with hydrogen peroxide or D-glucose (*26*) increased macrophage-HSPC interactions (fig. S5A). Consistent with prior work linking ROS, ER stress, and surface Calreticulin (*27*), ER stress inhibition by *perk* knock-down also decreased macrophage-HSPC interactions (fig. S5B). These data show that without macrophages, a population of HSPCs with increased ROS accumulates and that higher ROS levels correlate with surface Calreticulin.

Embryos injected with the *irf8* morpholino also had fewer HSPCs marked by genes associated with cell cycling and ERK/MAPK signaling (Fig. 4, B and C). In accordance with this, depleting macrophages or inhibiting ERK/MAPK without reducing macrophage interactions decreased HSPC proliferation. ERK/MAPK inhibition in the context of macrophage depletion did not further reduce proliferation, indicating that macrophages likely stimulate division through this pathway (fig. S5C). Because inflammation is a critical developmental determinant of HSPC proliferation (*28*, *29*), we reasoned that cytokines expressed by the macrophages, such as *il1b*, could be responsible for HSPC divisions (fig. S5D) (*29*, *30*). To investigate this possibility, we generated parabiotic fusions of control-injected embryos to *il1b* morpholino-injected embryos (*31*) and evaluated interactions between HSPCs and macrophages from both parabionts (fig. S5E). Interactions with control macrophages led to significantly more HSPC divisions than interactions with *il1b* knock-down macrophages, showing that macrophage-produced Il1b promotes HSPC division (fig. S5F). Additionally, heat-shock overexpression of *il1b* rescued HSPC proliferation after macrophage depletion (*29*) (Fig. 4F). ERK/MAPK inhibition abolished the effect of *il1b* (fig. S5G), indicating that Il1b-mediated HSPC divisions act through ERK/MAPK, matching the proliferation signature identified by scRNA-seq (Fig. 4, B and C). These results show that HSPC cycling in the CHT is mediated through ERK/MAPK activity induced by macrophage-derived Il1b.

## Discussion

Our data support a model in which macrophages of the embryonic niche vet the quality of newly formed HSPCs through prolonged physical contact, leading to either expansion or engulfment. This process is mediated by the display of cell surface Calreticulin, which is associated with increased ROS. It has previously been reported that metabolic shifts during HSPC generation in the VDA increase ROS to mediate HIF1α stabilization (*26*). Cells with high ROS are also at increased risk for DNA damage and dysfunction (*23*, *24*). Our work suggests that although ROS promotes stem cell emergence in the VDA, titration of ROS is ultimately required for normal hematopoiesis. Although we see no evidence for *vcam1* expression in embryonic macrophages, previous studies have indicated a role for macrophages in HSPC homing (*6*). Macrophages are involved in HSPC mobilization in the VDA (*32*), and murine macrophage subpopulations facilitate HSPC engraftment (*33*). By contrast, our study finds that macrophages in the CHT remove clones with high surface Calreticulin that have not down-regulated ROS. Healthy HSPCs with low-to-moderate ROS and Calreticulin experience prolonged macrophage contact and grooming, avoid complete engulfment, and respond to pro-proliferative Il1b enabling competition for marrow colonization.

Our work establishes that stem cells are quality assured for stress levels during development, and this affects the clones that contribute to blood formation in adulthood. Calreticulin functions as an "eat-me" molecule that initiates macrophage-HSPC interaction and leads to programmed cell clearance or stem cell expansion. Orthologs of CD47 and SIRPα, "don't-eat-me" signals, have not been identified in zebrafish, but other primitive signals could influence macrophage behaviors. This quality assurance mechanism may also operate in adulthood in response to environmental stress, such as during marrow transplantation or in clonal stem cell disorders, including myelodysplasia and leukemia. Macrophages may selectively expand or remove clones of tissue-specific stem cells in other systems similar to our findings. Other tissue stem cells rely on macrophages to assure adequate tissue regeneration (*34*), which could occur through selective proliferation of certain clones. Manipulating this quality assurance mechanism may have important therapeutic implications for stem cell disorders and tissue regeneration.

## REFERENCES AND NOTES

1. K. Kissa, P. Herbomel, *Nature* **464**, 112–115 (2010).
2. J. Y. Bertrand *et al*., *Nature* **464**, 108–111 (2010).
3. E. Murayama *et al*., *Immunity* **25**, 963–975 (2006).
4. O. J. Tamplin *et al*., *Cell* **160**, 241–252 (2015).
5. E. J. Hagedorn *et al*., bioRxiv 2021.11.03.467105 [Preprint] (2021); https://doi.org/10.1101/2021.11.03.467105.
6. D. Li *et al*., *Nature* **564**, 119–124 (2018).
7. J. Henninger *et al*., *Nat. Cell Biol.* **19**, 17–27 (2017).
8. P. Herbomel, B. Thisse, C. Thisse, *Development* **126**, 3735–3745 (1999).
9. M. Feng *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2145–2150 (2015).
10. M. Feng *et al*., *Nat. Commun.* **9**, 3194 (2018).
11. F. Ellett, L. Pase, J. W. Hayman, A. Andrianopoulos, G. J. Lieschke, *Blood* **117**, e49–e56 (2011).
12. M. Sugiyama *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20812–20817 (2009).
13. T. J. van Ham, J. Mapes, D. Kokel, R. T. Peterson, *FASEB J.* **24**, 4336–4342 (2010).
14. S. Avagyan *et al*., *Science* **374**, 768–772 (2021).
15. Y. A. Pan *et al*., *Development* **140**, 2835–2846 (2013).
16. L. Li, H. Jin, J. Xu, Y. Shi, Z. Wen, *Blood* **117**, 1359–1369 (2011).
17. B. Budnik, E. Levy, G. Harmange, N. Slavov, *Genome Biol.* **19**, 161 (2018).
18. D. Bausch-Fluck *et al*., *PLOS ONE* **10**, e0121314 (2015).
19. M. P. Chao *et al*., *Sci. Transl. Med.* **2**, 63ra94 (2010).
20. S. J. Gardai *et al*., *Cell* **123**, 321–334 (2005).
21. C. A. Ogden *et al*., *J. Exp. Med.* **194**, 781–796 (2001).
22. K. Kissa *et al*., *Blood* **111**, 1147–1156 (2008).
23. G. J. Kops *et al*., *Nature* **419**, 316–321 (2002).
24. H. Tran *et al*., *Science* **296**, 530–534 (2002).
25. Z. Tothova *et al*., *Cell* **128**, 325–339 (2007).
26. J. M. Harris *et al*., *Blood* **121**, 2483–2493 (2013).
27. T. Panaretakis *et al*., *EMBO J.* **28**, 578–590 (2009).
28. R. Espín-Palazón *et al*., *Cell* **159**, 1070–1085 (2014).
29. J. M. Frame *et al*., *Dev. Cell* **55**, 133–149.e6 (2020).
30. E. M. Pietras *et al*., *Nat. Cell Biol.* **18**, 607–618 (2016).
31. A. López-Muñoz *et al*., *Mol. Immunol.* **48**, 1073–1083 (2011).
32. J. Travnickova *et al*., *Nat. Commun.* **6**, 6227 (2015).
33. A. Chow *et al*., *J. Exp. Med.* **208**, 261–271 (2011).
34. D. Ratnayake *et al*., *Nature* **591**, 281–287 (2021).



## ACKNOWLEDGMENTS

We thank the Boston Children's Hospital (BCH) veterinary staff, the BCH flow cytometry core, Single Cell Discoveries, and the Harvard Center for Mass Spectrometry. We also thank A. Han for her assistance with fetal mouse dissection, B. Miller for her help preparing cryosections, and colleagues for critical reading of the manuscript. S.J.W. thanks P. Chen for her continued support in all things. **Funding:** This work was supported by National Institutes of Health grants 1F31HL149154-01 (S.J.W.); 5T32HL007574-40 (C.P.R.); K01DK111790 (E.J.H.); and P01HL131477, P01HL032262, U54DK110805, R24DK092760, R24OD017870, U01HL134812, and R01HL144780-01 (L.I.Z.). This work was also supported by the Edward P. Evans Foundation (L.I.Z.) and Alex's Lemonade Stand Fund (L.I.Z.). L.I.Z. is a Howard Hughes Medical Institute investigator. **Author contributions:** Conceptualization: S.J.W., L.I.Z.; Methodology: S.J.W., M.L.S., B.B.; Investigation: S.J.W., M.L.S., C.P.R., J.W.K., E.J.H., B.B.; Visualization: S.J.W., M.L.S., C.P.R.; Funding acquisition: S.J.W., C.P.R., E.J.H., L.I.Z.; Project administration: S.J.W.; Supervision: S.J.W., L.I.Z.; Writing – original draft: S.J.W.; Writing – review and editing: S.J.W., M.L.S., C.P.R., J.K., E.J.H., B.B., L.I.Z. **Competing interests:** L.I.Z. is a founder and stockholder of Fate Therapeutics, CAMP4 Therapeutics, Amagma Therapeutics, and Scholar Rock and is a consultant for Celularity. B.B. is on the scientific advisory board of Preverna, Inc. All other authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. The scRNA-seq data are available in the NCBI Gene Expression Omnibus (GSE196553). Proteomic data are available in the MassIVE database (MSV000088780) and PRIDE repository (PXD031434).



## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo4837
Materials and Methods
Figs. S1 to S5
Tables S1 and S2
References (*35*–*56*)
MDAR Reproducibility Checklist
Movies S1 to S3

Submitted 6 February 2022; resubmitted 11 July 2022
Accepted 19 August 2022
10.1126/science.abo4837

PLANT SCIENCE

# NIN-like protein 7 transcription factor is a plant nitrate sensor

**Kun-Hsiang Liu[1,2,3]*†, Menghong Liu[1]†, Ziwei Lin[1]†, Zi-Fu Wang[4]†, Binqing Chen[1]†, Cong Liu[1], Aping Guo[1], Mineko Konishi[5], Shuichi Yanagisawa[5], Gerhard Wagner[4], Jen Sheen[3]***

**Nitrate is an essential nutrient and signaling molecule for plant growth. Plants sense intracellular nitrate to adjust their metabolic and growth responses. Here we identify the primary nitrate sensor in plants. We found that mutation of all seven *Arabidopsis* NIN-like protein (NLP) transcription factors abolished plants' primary nitrate responses and developmental programs. Analyses of NIN-NLP7 chimeras and nitrate binding revealed that NLP7 is derepressed upon nitrate perception via its amino terminus. A genetically encoded fluorescent split biosensor, mCitrine-NLP7, enabled visualization of single-cell nitrate dynamics in planta. The nitrate sensor domain of NLP7 resembles the bacterial nitrate sensor NreA. Substitutions of conserved residues in the ligand-binding pocket impaired the ability of nitrate-triggered NLP7 to control transcription, transport, metabolism, development, and biomass. We propose that NLP7 represents a nitrate sensor in land plants.**

Nitrogen, the main limiting factor for plant growth, is fundamental to agricultural productivity, animal and human nutrition, and sustainable ecosystems (*1*, *2*). Photosynthetic plants drive the terrestrial nitrogen cycle by assimilating inorganic nitrogen into biomolecules (DNA, RNA, proteins, chlorophyll, and vitamins) that sustain the plants and the food webs that depend on them (*1*–*4*). To compete with microbes that prefer organic nitrogen or ammonium in the soil, most plants have evolved regulatory pathways that respond to fluctuating nitrate availability (*3*–*5*). Plants that sense available nitrate will, within minutes, orchestrate transcriptome, metabolism, hormone, system-wide shoot and root growth, and reproduction responses (*5*–*8*). Here we identify the transcription factor NIN-like protein 7 (NLP7) as the primary nitrate sensor. We show that NLP7 acts as an intracellular nitrate sensor distinct from the plasma membrane extracellular nitrate transporter-sensor (transceptor) NRT1.1 (also known as CHL1 or NPF6.3) (*9*–*13*).

[1]State Key Laboratory of Crop Stress Biology for Arid Areas and College of Life Sciences, Northwest Agriculture and Forestry University, Yangling, Shaanxi 712100, China. [2]Institute of Future Agriculture, Northwest Agriculture and Forestry University, Yangling, Shaanxi 712100, China. [3]Department of Molecular Biology and Centre for Computational and Integrative Biology, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, MA 02114, USA. [4]Department of Biological Chemistry and Molecular Pharmacology, Blavatnik Institute, Harvard Medical School, Boston, MA 02115, USA. [5]Agro-Biotechnology Research Center, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657, Japan.
***Corresponding author. Email: khliu@molbio.mgh.harvard.edu (K.-H. L.); sheen@molbio.mgh.harvard.edu (J.S.)**
†These authors contributed equally to this work.

NIN-like proteins 6 and 7 (NLP6/7), which were identified by homology to the regulator NODULE INCEPTION (NIN) that controls root nodule initiation and nitrogen fixation (*14*), are master transcription factors in nitrate signaling and widely conserved in land plants and major crops (*9*, *15*–*19*). There are nine NLPs in *Arabidopsis thaliana* featuring an N-terminal nitrate response region, a conserved RWP-RK DNA binding domain for the nitrate response cis-element (*NRE*), and a C-terminal PB1 protein-protein interaction domain (*13*, *15*, *17*, *20*). The N-terminal nitrate-response region of NLP1 to NLP9 (NLP1-9) encompasses an evolutionarily conserved regulatory phosphorylation site critical for nitrate-induced nuclear retention and transcription activation (*18*, *20*). Functions of putative GAF (cGMP-specific phosphodiesterase, adenylyl cyclases, and FhlA)–like domains and other conserved motifs remain unclear (*17*). Nitrate activates group III calcium-sensor protein kinases (CPKs), which phosphorylate NLP7 at S205 to retain NLP7 in the nucleus, where it activates primary nitrate responses that shape organ biomass and architecture (*18*). Nitrate-dependent transactivation activity of NLP2 requires the conserved S202 residue (*20*). Nitrate is required to activate the primary reporter gene despite coexpression of NLP7 and constitutively active CPK10ac in nitrate-free leaf cells (*18*), suggesting a missing nitrate sensor action. Here, we demonstrate that NLP7 is an intracellular nitrate sensor. The nitrate-binding domain seems evolutionarily ancient and is conserved in plant NLPs and the bacteria nitrate sensors NreA

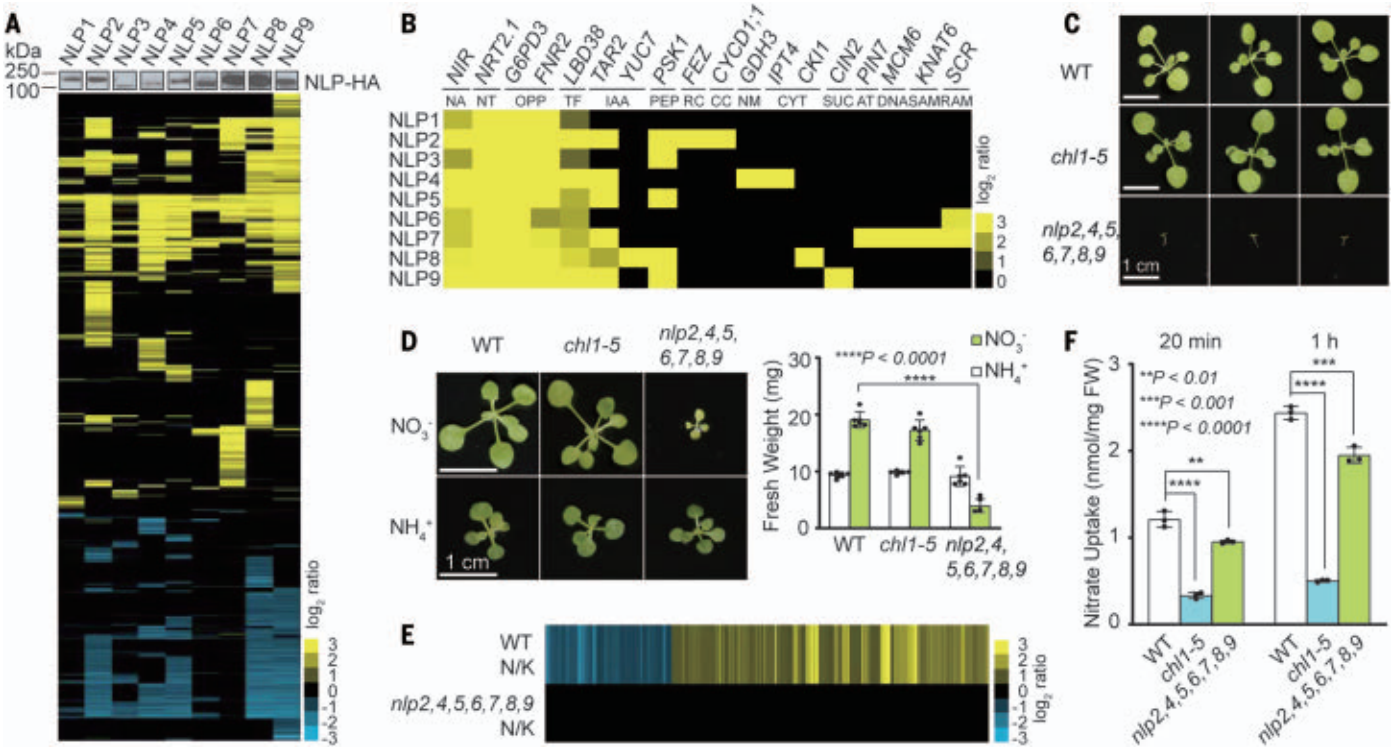



**Fig. 1. Combinatorial NLP transcription factors are central to primary nitrate responses and developmental programs.** (**A**) NLPs regulate both common and distinct target genes. Hierarchical clustering of RNA-seq analysis defines putative target genes of NLP1-9 by transient expression in leaf cells. NLP protein expression was determined by immunoblot analyses. (**B**) Heatmap of nitrate-responsive core genes and distinct target genes activated by NLP1-9. NA, nitrate assimilation; NT, nitrate transporter; OPP, oxidative pentose phosphate; TF, transcription factor; IAA, indole-3-acetic acid biosynthesis; PEP, peptide hormone; RC, root cap; CC, cell cycle; NM, nitrogen metabolism; CYT, cytokinin biosynthesis and signaling; SUC, sucrose invertase; AT, auxin transport; DNA, DNA synthesis; SAM, shoot apical meristem; RAM, root apical meristem. (**C**) The *nlp2,4,5,6,7,8,9* mutant abolishes shoot growth in soil. (**D**) The *nlp2,4,5,6,7,8,9* mutant displays nitrate-specific reduction in shoot and biomass. Error bars, SD; $n = 6$ biological replicates. (**E**) Primary nitrate-responsive transcriptome is abolished in *nlp2,4,5,6,7,8,9*. N, $KNO_3$; K, KCl. (**F**) Nitrate uptake in *nlp2,4,5,6,7,8,9*. FW, fresh weight. Error bars, SD; $n = 3$ biological replicates. Scale bars, 1 cm.

(*21*). Nitrate directly binds to NLP7 through residues evolutionarily conserved in other NLPs and NLP7 orthologs, triggers a conformational change, and derepresses NLP7 as a transcription activator.

### Combinatorial NLPs control primary nitrate responses

All nine *NLP* genes are expressed in *Arabidopsis* shoots. Analyses of individual *nlp1-9* single mutants in nitrate-mediated shoot growth revealed statistically significant defects only in *nlp2* and *nlp7* (*9*, *18*, *20*). To circumvent NLP redundancy and better define the overlapping or distinctive functions of NLP1-9, we conducted a genome-wide target gene survey. Each NLP was transiently expressed for 4.5 hours in transfected leaf cells from soil-grown plants for RNA sequencing (RNA-seq) analyses (*15*, *18*, *21*). Hierarchical clustering analysis of putative NLP target genes ($\log_2 \geq 1$ or $\leq -1$; $P \leq 0.05$) revealed the capability of all NLPs to activate universal primary nitrate-responsive marker genes previously identified by microarray, RNA-seq, chromatin immunoprecipitation on chip (ChIP-chip), ChIP-seq, and promoter analyses (*15*, *16*, *18*, *21*, *22*) (Fig. 1, A and B, and tables S1 and S2). NLP2, NLP4, NLP7, NLP8, or NLP9 specifically activated some target genes with known functions in modulating auxin and cytokinin hormone functions, cell cycle, metabolism, peptide signaling, and shoot and root meristem activities (Fig. 1, A and B, and tables S1 and S2). NLP2 and NLP7 regulated broader nonredundant target genes with diverse functions, which might manifest as growth defects observed in *nlp2* or *nlp7* after seed germination (*9*, *18*, *20*). NLP6/7 acted predominantly as transcription activators, whereas NLP2,4,5,8,9 could activate or repress target genes. NLP1,3,6 modulated fewer target genes than other NLPs. For instance, the auxin biosynthesis gene *TAR2* was only activated by NLP2,4,5,7,8,9 (*23*). These results are consistent with NLP variants functioning combinatorially in controlling the nitrate response network and causing retarded shoot and root development in the *nlp2,4,5,6,7,8,9* septuple mutant plants grown in soil (Fig. 1C).

NRT1.1(CHL1/NPF6.3) acts as a plasma membrane nitrate transceptor that senses external nitrate in *Arabidopsis* (*10*–*12*). However, the null *chl1-5* mutant resembled wild-type (WT) plants when grown in the soil under light, whereas the *nlp2,4,5,6,7,8,9* septuple mutant failed to develop further after germination (Fig. 1C). When grown on a sterile culture medium with 0.5% sucrose and ammonium succinate, WT, *chl1-5*, and *nlp2,4,5,6,7,8,9* plants were able to develop similar shoots, albeit ones smaller than those of WT plants grown in nitrate medium (*1*, *19*, *21*). In contrast, shoot and root development of *nlp2,4,5,6,7,8,9* remained constrained in nitrate medium (Fig. 1D and fig. S1, A to F) (*20*). Transcriptome reprogramming in primary nitrate responses triggered by nitrate at 20 min ($\log_2 \geq 1$ or $\leq -1$; $q \leq 0.05$) was abolished in *nlp2,4,5,6,7,8,9* (Fig. 1E). The activation of primary nitrate response genes are only partially reduced in *chl1-5* (fig. S2 and table S3). However, the uptake of nitrate showed a stronger reduction in chlorate-resistant *chl1-5* but was only moderately reduced in *nlp2,4,5,6,7,8,9* (Fig. 1F) (*6*, *24*). Chlorate could not activate primary nitrate response genes (fig. S3). These findings further supported the essential role of NLPs in regulating primary nitrate responses (Fig. 1E, fig. S2, and tables S1 to S3) and plant development, as well as the existence of nitrate

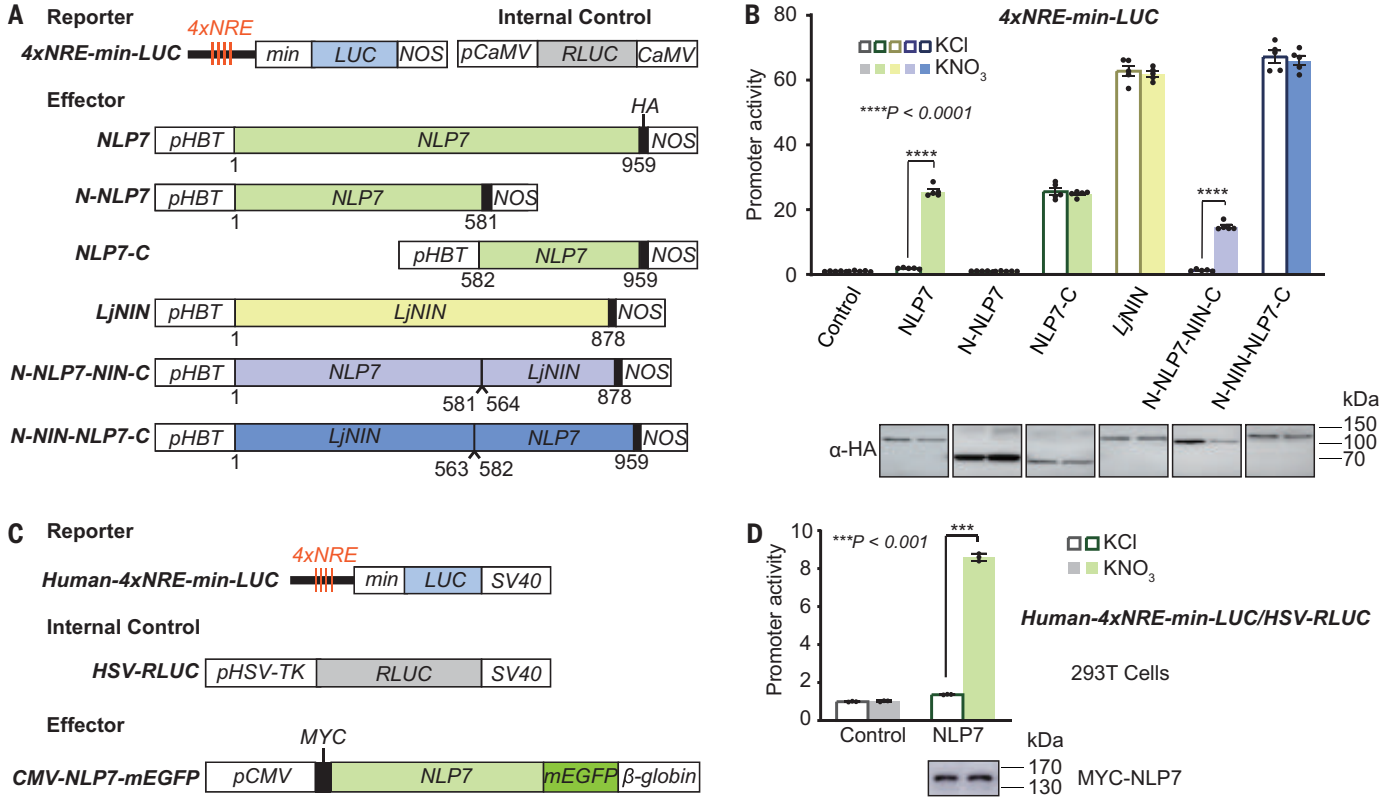



**Fig. 2. Nitrate derepresses NLP7.** (**A**) Reporters and effectors for transient assays in leaf cells. The nitrate-responsive reporter (*4xNRE-min-LUC*) contains four copies of nitrate-responsive element (*NRE*) and the *35S* minimal promoter (*min*) fused to firefly luciferase gene (*LUC*) and *NOS* terminator (*NOS*). The *35S* cauliflower mosaic virus promoter controls the *Renilla* luciferase gene (*pCaMV-RLUC*) as the internal control. The effector expression is controlled by a constitutive *HBT* promoter (*19*). Numbering shows the amino acid positions. (**B**) Analyses of NLP-NIN chimeras reveal derepression of NLP7 by nitrate. Transactivation assays of *NRE-LUC* were carried out in *nlp2,4,5,6,7,8,9* leaf cells by expressing effectors for 4 hours before induction (0.5 mM $KNO_3$ for 2 hours). Luciferase activity was normalized to *pCaMV-RLUC* activity. Error bars, SD; *n* = 5 biological replicates. Effector protein expression was determined by immunoblot analyses. (**C**) Reporters and effector for transient assays in 293T human cells. The nitrate-responsive *Human-4xNRE-min-LUC* reporter contains the *SV40* terminator. The herpes simplex virus (*HSV*) promoter–driven *RLUC* (*pHSV-RLUC*) is an internal control. The *MYC-NLP7-mGFP* gene is controlled by a human cytomegalovirus (*CMV*) promoter with the β-*globin* gene terminator. (**D**) *NLP7* activates *Human-4xNRE-min-LUC* in response to nitrate in heterologous 293T human cells. Luciferase activity was normalized to *pHSV-RLUC* activity (10 mM $KNO_3$ for 2 hours). Error bars, SD; *n* = 3 biological replicates.

sensors presumably distinct from NRT1.1 (CHL1/NPF6.3) (*9*–*13*).

### Nitrate derepresses NLP7 via the N terminus

The N terminus of NLPs contain the GAF-like domain and multiple conserved motifs with unknown functions (*17*). When fused to the *LexA* DNA binding domain and a nuclear localization signal, the N terminus of NLP6(1-546) is sufficient to confer nitrate response to the *8xLexA-min:GUS* reporter in nitrate-free seedlings (*15*). Single-mutant studies showed that *nlp7* but not *nlp6* exhibited overt shoot growth defects (fig. S1A). NLP6 also activated fewer target genes than did the closely related NLP7 transcription activator in transfected leaf cells (Fig. 1A and tables S1 and S2). Therefore, we examined the role of NLP7 and the functional domains of NLP7 for the activation of a synthetic nitrate-responsive reporter *4xNRE-min-LUC* in nitrate-free *Arabidopsis* leaf cells (*13*, *15*, *18*). Although *4xNRE-min-LUC* was activated by nitrate (10 mM, 2 hours) in transfected WT and *nlp6* leaf cells, its activation was diminished in *nlp7* or *nlp6,7* and abolished in *nlp2,4,5,6,7,8,9* (fig. S1G). *NRE* is evolutionarily conserved in the primary nitrate-responsive *NIR* gene promoter from *Arabidopsis*, spinach, bean, birch, and maize (*25*). These data suggested a role of NLP7 in activating *NRE*.

Full-length NLP7(1-959) conferred nitrate-specific *4xNRE-min-LUC* activation at low (0.5 mM) nitrate for 2 hours in nitrate-free leaf cells. However, the N-NLP7(1-581) domain was inactive, whereas the NLP7-C(582-959) domain was constitutively activated in the same reporter without nitrate (Fig. 2, A and B). The *Lj*NIN transcription factor from *Lotus japonicus*, which controls root nodule initiation and nitrogen fixation and shares RWP-RK DNA binding and PB1 domains with NLPs but not the N terminus (*15*, *26*), activated the *NRE*-based reporter in nitrate-free leaf cells. Analyses of the chimeric transcription factors N-NLP7-NIN-C and N-NIN-NLP7-C suggested that N-NLP7 acted as a repressor domain in full-length NLP7 and is derepressed by nitrate (Fig. 2, A and B). As the NLP7 orthologs in diverse plant species were more ancient and arose after the green algal lineage, NIN in legume species likely evolved later, having lost autorepression and nitrate responsiveness (fig. S4) (*9*, *14*, *17*, *27*).

RWP-RK transcription factors are specific to vascular plants, green algae, and slime molds (*17*). Although human cells can transport and reduce nitrate to support an alternative plant diet–based NO pathway (*28*), humans have no orthologs for NRT1.1 transceptor or NLPs. We asked whether NLP7 could activate a synthetic version of the nitrate-responsive reporter *Human-4xNRE-min-LUC* with a mammalian minimal promoter and *SV40* terminator in heterologous 293T human cells (Fig. 2C). As in nitrate-free *Arabidopsis* leaf cells, *Human-4xNRE-min-LUC* was activated by NLP7 only in the presence of nitrate (Fig. 2D). Thus, NLP7 alone may function both as a nitrate sensor and a transcription activator in the heterologous human cell system.

### Nitrate directly binds to NLP7

To test the hypothesis that NLP7 is both a ligand-dependent transcription activator and

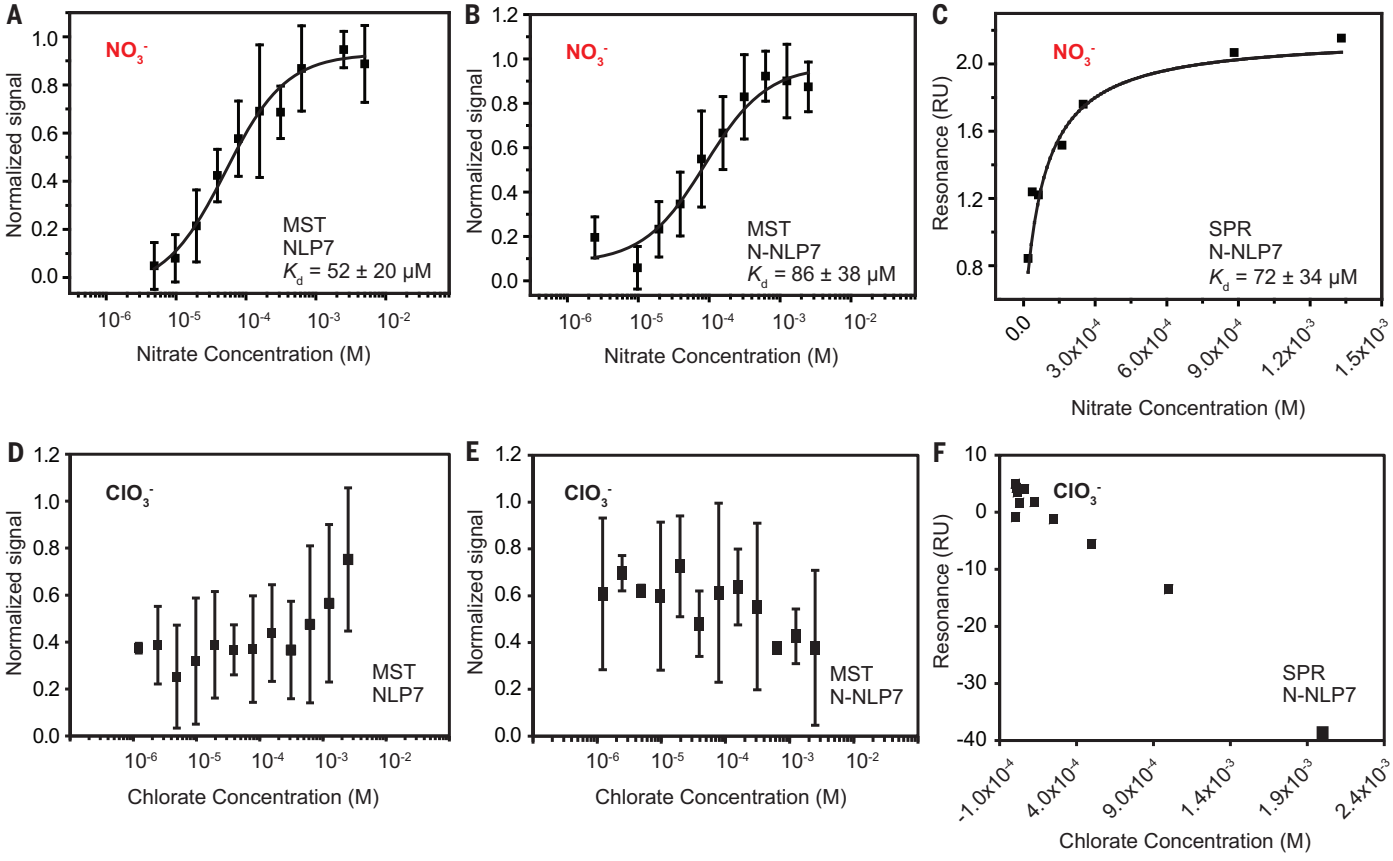



**Fig. 3. Nitrate directly binds to NLP7.** (**A**) $NO_3^-$ binding to full-length NLP7, as measured by microscale thermophoresis (MST). Dissociation constant ($K_d$) = 52 ± 20 μM. Error bar, SD; $n$ = 3. (**B**) $NO_3^-$ binding to N-NLP7. $K_d$ = 86 ± 38 μM. Error bar, SD; $n$ = 3 independent thermophoresis measurements. (**C**) Analysis of $NO_3^-$–N-NLP7 interaction, as measured by surface plasmon resonance (SPR). The result is a representative of three independent experiments. The $K_d$ is the average of three independent experiments. RU, resonance units. (**D**) $ClO_3^-$ does not bind to full-length NLP7. Error bars, SD; $n$ = 3 independent thermophoresis measurements. (**E**) $ClO_3^-$ does not bind to N-NLP7. Error bars, SD; $n$ = 3 independent thermophoresis measurements. (**F**) SPR analysis of $ClO_3^-$–N-NLP7 interaction. The result is a representative of three independent experiments.

an intracellular nitrate sensor, we measured nitrate binding to NLP7 using microscale thermophoresis (MST) and surface plasmon resonance (SPR) (*29*, *30*). We expressed and purified GST-NLP7-8xHIS fusion protein from *Escherichia coli* (fig. S5A) and performed binding experiments by MST at different nitrate concentrations (Fig. 3A). Similar to NreA, which can bind nitrate with a dissociation constant ($K_d$) of 22 μM, NLP7 (not HIS-GST) selectively recognized nitrate (Fig. 3A) with a $K_d$ of 52 ± 20 μM but not its structural analog chlorate or anion phosphate (Fig. 3, A and D, and fig. S5, B to D and F). Because the N terminus of NLP7 is crucial for nitrate responsiveness, we also expressed and purified the N terminus NLP7(1-581)–based GST-N-NLP7-8xHIS fusion protein and carried out the MST assay. Nitrate could bind to N-NLP7 with a $K_d$ of 86 ± 38 μM but not chlorate or phosphate (Fig. 3, B and E, and fig. S5E). To confirm the interaction between NLP7 and nitrate, we performed alternative binding experiments using SPR. Purified GST-N-NLP7-8xHIS protein was immobilized on a sensor chip and different concentrations of nitrate, chlorate, or phosphate were monitored by SPR with an average $K_d$ of 72 ± 34 μM only for nitrate (Fig. 3, C and F, and fig. S5, G to I).

### A genetically encoded fluorescent biosensor visualizes nitrate in plants

Ligand–sensor interaction may trigger a conformational change in the sensor protein. We generated a genetically encoded fluorescent biosensor, mCitrine-NLP7, similar to the single fluorescent protein–based glucose biosensor Green Glifon (*31*). We hypothesized that the nitrate-bound split mCitrine-NLP7 nitrate biosensor (sCiNiS) would reconstitute mCitrine to emit fluorescent signals (*31*). The predicted nuclear localization signal ($_{630}$RRKKK$_{638}$) of NLP7 was mutated to AAAAA to avoid competition with the endogenous NLP7, which is retained in the nucleus after nitrate induction for transcriptional activation (Fig. 4A and fig. S6, A and B) (*16*, *18*). Quantitative confocal imaging of the cytoplasmic nitrate by sCiNiS was carried out in mesophyll cells of cotyledons and columella cells of root tips in transgenic plants (Fig. 4, B and C). The reconstituted mCitrine fluorescent signal was detected within 5 min after nitrate (10 mM), but not KCl, induction in both mesophyll cells and primary root tip cells at single-cell resolution in intact sCiNiS transgenic seedlings that developed normally (Fig. 4, B and C, and fig. S6C). Soil nitrate concentrations can vary from the micromolar to the millimolar range (*6*). We tested different nitrate concentrations using nitrate-free transgenic seedlings and showed that the sCiNiS biosensor detected a range of nitrate concentrations from 100 μM to 10 mM in single mesophyll cells in intact plants, consistent with a sensitive and specific nitrate binding $K_d$ of 52 μM for NLP7 in vitro (Fig. 3A and fig. S6D).

### Evolutionary conservation of nitrate sensors in plants

Architectures of nitrate-binding sites in bacteria and cyanobacteria proteins have been solved by crystal structure analyses of cytoplasmic NreA of *Staphylococci carnosus*, periplasmic NarX of *E. coli*, and periplasmic NrtA of *Synechocystis* sp. These nitrate-binding proteins support anaerobic respiration and nitrate transport. Their crystal structures are

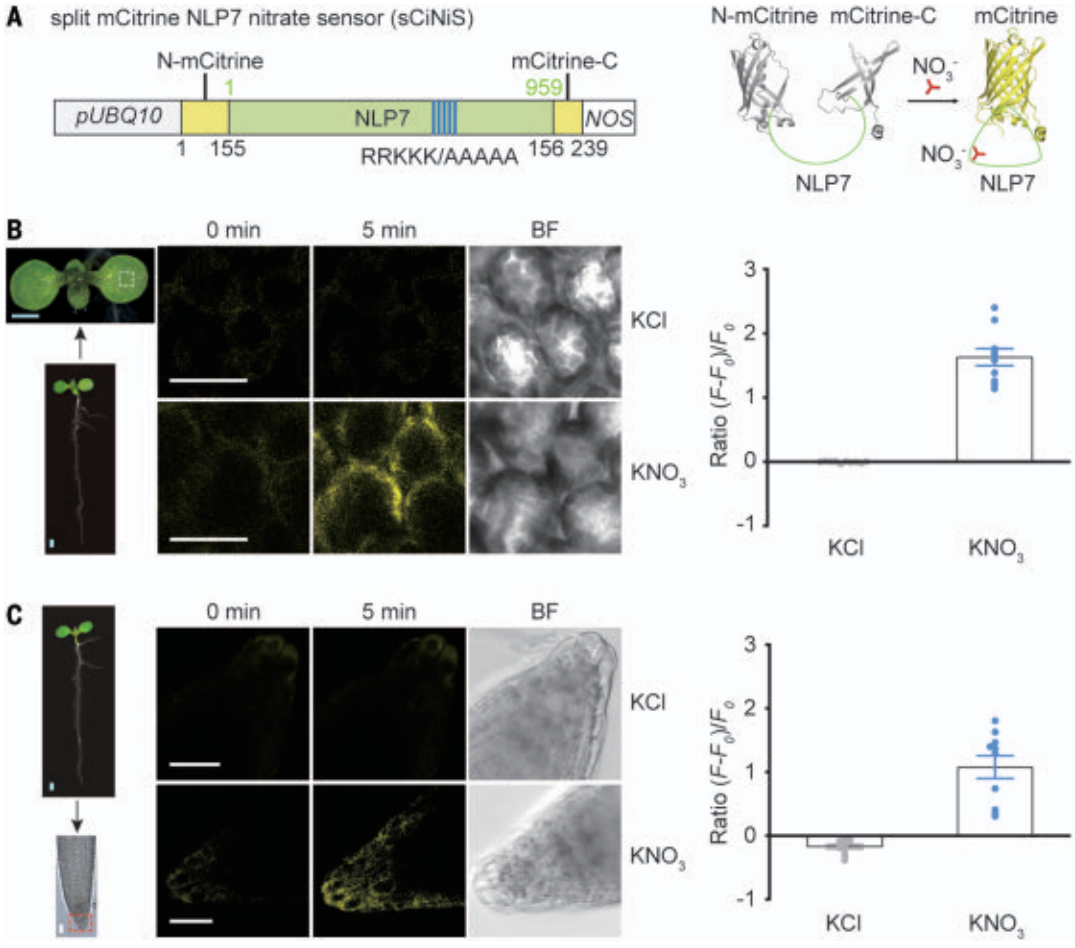



**Fig. 4. The genetically encoded biosensor detects intracellular nitrate in transgenic shoots and roots.** (**A**) A schematic representation of domain structure and a model of the nitrate biosensor. Nitrate triggers a conformational change of the split mCitrine-NLP7 nitrate biosensor (sCiNiS) and reconstitutes mCitrine to emit fluorescent signals. The predicted nuclear localization signal ($_{630}$RRKKK$_{638}$) of NLP7 was mutated to AAAAA to avoid competition with the nitrate-induced endogenous NLP7 in the nucleus. (**B**) Imaging the cytoplasmic nitrate by sCiNiS in mesophyll cells of cotyledons. Dashed white line indicates the imaging site. Images are representative of 10 cotyledons. Blue scale bar, 1 mm. White scale bar, 30 μm. (**C**) Imaging the cytoplasmic nitrate by sCiNiS in root tips. Dashed red line indicates the imaging site. Images are representative of 10 root tips, KCl or $KNO_3$ (10 mM). $(F - F_0)/F_0$, the relative fluorescence intensity. Error bars, SD; $n$ = 10 biological replicates.

available in the Protein Data Bank (PDB IDs 4IUK, 3EZH, and 2G29, respectively) (*21*). We found that seven of eight critical residues in the nitrate-binding pocket of the NreA nitrate sensor could be aligned to the conserved motifs next to the putative GAF-like domain in NLP1-9 (*17*, *21*) (Fig. 5, A and B, and fig. S7A). The protein structure of the nitrate-binding domain in NLP7 predicted by the artificial intelligence programs AlphaFold2 and RoseTTAfold (*32*, *33*) also resembled that in NreA (Fig. 5C).

To functionally define the essential residues for nitrate binding in NLP7, we carried out alanine (A) scanning mutagenesis of eight putative nitrate-binding residues defined in the nitrate-NreA crystal form (*21*) and examined the nitrate response of mutated NLP7 in nitrate-free leaf cells. NLP7 mutations of four residues, $Trp^{395}$→Ala (W395A), H404A, L406A, or Y436A, significantly decreased nitrate-induced *4xNRE-min-LUC* activity at low (0.5 mM) nitrate for 2 hours. Because H404, L406, and Y436 are conserved in NLP2,4,5,6,8,9 with similar structures in the nitrate-binding domain (fig. S7, A and B), we next generated and analyzed double (HL/AA) and triple (HLY/AAA) mutants of NLP7, which abolished nitrate-induced *4xNRE-min-LUC* activity (Fig. 5D). The HLY nitrate-binding residues are also conserved among NLP7 orthologs within the structurally similar nitrate sensor domain in crop plants including rapeseed *Bna*NLP7, soybean *Gm*NLP6, maize *Zm*NLP6, wheat *Ta*NLP7, and rice *Os*NLP3 (Fig. 5E and fig. S7, C and D). We propose that NLP7 and its orthologs may serve as nitrate sensors conserved in photosynthetic plants preserved from charophytes to angiosperms, including eudicots and monocots but not chlorophytes (fig. S7, C and D).

To demonstrate the function of the nitrate-binding domain in NLP7, we examined the ability of NLP7 and NLP7(HLY/AAA) to activate the nitrate-responsive reporter *Human-4xNRE-min-LUC* in the heterologous 293T cell system. Compared with NLP7, NLP7(HLY/AAA) abrogated the LUC activity induced by nitrate-bound and activated NLP7 (Fig. 5F). Similar to NLP7, *Os*NLP3 as the rice NLP7 ortholog activated the nitrate-dependent *Human-4xNRE-min-LUC* activity, which was eliminated with *Os*NLP3(HLY/AAA, mutations in H392A, L394A, and Y424A) (Fig. 5F). To verify that the residues H404, L406, and Y436 are required for nitrate binding in the NLP7 sensor domain, we performed MST assays and SPR analyses using purified NLP7 (HLY/AAA) or N-NLP7(HLY/AAA, 1-581) protein (Fig. 5, G to I, and fig. S5A). Mutation of these three amino acids impeded nitrate binding in vitro. Furthermore, NLP7-HA but not NLP7(HLY/AAA)-HA complemented shoot growth and biomass phenotypes (Fig. 5J and fig. S8A) or activation of nitrate-responsive genes for nitrate transport, assimilation, and transcription in *nlp7* (fig. S8B). The data support the role of these residues in NLP7 to function as a dual master transcription activator and nitrate sensor central to transcriptional reprogramming and plant development.

## Discussion

Nitrate is an essential nutrient for photosynthetic plant growth in most soils and controls metabolic and developmental processes pivotal to plant vegetative and reproductive development. However, nitrogen fertilizer is energy intensive to produce and causes pollution; furthermore, its overuse in agriculture to boost crop yields has led to environmentally disastrous eutrophication worldwide. Global and regional studies indicate declining nitrogen availability on Earth (*1*–*7*). Improved plant nitrogen use efficiency could contribute to sustainable agriculture and ecosystem preservation.

Our data identified NLP2,4,5,6,7,8,9 as transcription factors that initiate signaling for

nitrate-mediated transcriptome reprogramming and coordinate transport, metabolism, hormone, signaling, transcription, and root-shoot developmental programs (*5*–*7*, *18*). With genetic, genomic, cellular, imaging, biochemical, and structural analyses, we discovered that NLP7 has dual regulatory modes as a transcription activator and an intracellular nitrate sensor. Nitrate binding triggers a conformational change and transcriptional derepression of NLP7, acting simultaneously and synergistically with CPK-dependent NLP7

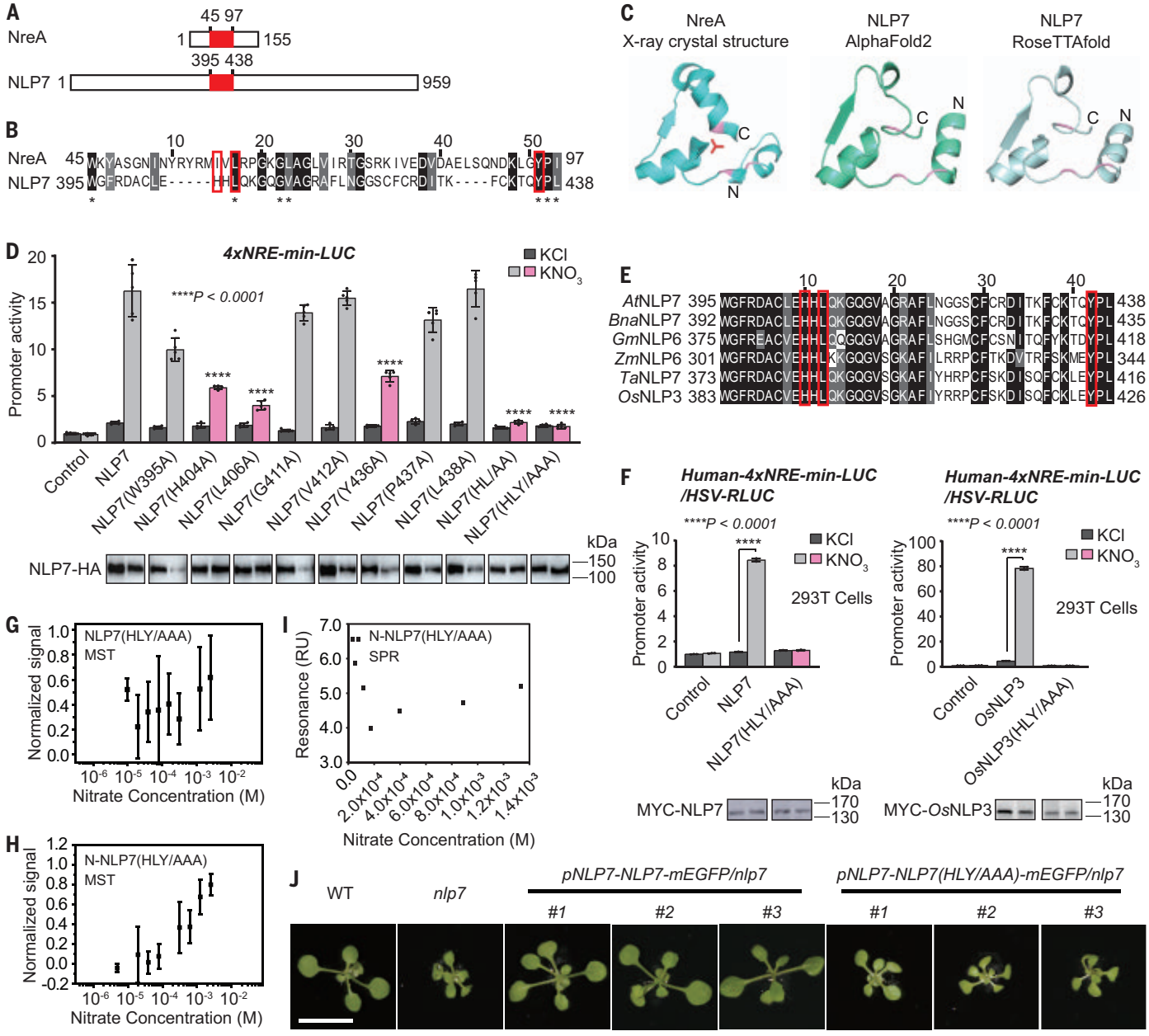



**Fig. 5. The NLP7 sensor domain resembles NreA with conserved residues for nitrate perception and signaling.** (**A**) The nitrate-binding domain of the bacterial nitrate sensor NreA shares homology to a region of NLP7. The homologies are in red. Numbering refers to amino acid positions. (**B**) Sequence alignment of the nitrate-binding domain of NreA and NLP7. Black box indicates conserved residue. Gray box indicates semiconserved residue. Three essential nitrate-binding residues are outlined in red. Asterisks mark the conserved residues in the nitrate-binding pocket of NreA. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (**C**) Comparison of the crystal structure of the NreA nitrate-binding domain and the predicted structure of NLP7. Red stick: nitrate; Pink: NreA(I59, L61, Y95) or NLP7(H404, L406, Y436). (**D**) Nitrate-binding mutant screens by transient assays in leaf cells. Transactivation of *4xNRE-min-LUC* by NLP7 or nitrate-binding mutants was analyzed in *nlp2,4,5,6,7,8,9* leaf protoplasts (0.5 mM $KNO_3$ for 2 hours). Error bars, SD; *n* = 5 biological replicates. (**E**) The key nitrate-binding residues are conserved in NLPs from major crops. Critical nitrate-binding residues are outlined in red. (**F**) Functional conservation of NLP7 and rice *Os*NLP3 for nitrate perception and transcription activation in heterologous 293T human cells. The nitrate-binding residues of NLP7(H404, L406, Y436) or *Os*NLP3(H392, L394, Y424) are essential for nitrate activation of *Human-4xNRE-min-LUC*. Error bars, SD; *n* = 3 biological replicates. (**G** to **I**) Mutant full-length NLP7(HLY/AAA) and N-NLP7(HYL/AAA) abolish nitrate binding. Error bars, SD; *n* = 3 independent thermophoresis measurements. The result is a representative of three independent experiments in SPR assays. (**J**) *pNLP7-NLP7-GFP* but not *pNLP7-NLP7(HLY/AAA)-GFP* complements the *nlp7-1* mutant. Scale bar, 1 cm.

(S205) phosphorylation for nuclear retention (*16*, *18*).

The sequence and structure of the nitrate-binding pocket of NLP7 is evolutionarily conserved and found in bacteria nitrate sensor NreA proteins (*34*), as well as in other *Arabidopsis* NLPs and the NLP7 orthologs of land plants and freshwater green algae in charophytes. We speculate that NLP7 orthologs and other NLPs, but not NINs, in diverse plant species also act as sensor-transcription factors to initiate primary nitrate responses regulating overlapping and distinctive target genes for a range of physiological functions in the nutrient-regulatory networks (*5*–*8*, *18*). The functions of NRT1.1(CHL1/NPF6.3) as an extracellular nitrate transceptor and transporter (*10*–*12*) seem distinct when plants are grown in soil or $NO_3^-$ medium. The development of the genetically encoded fluorescent nitrate biosensor sCiNiS enables sensitive and specific monitoring of single-cell nitrate signals in whole seedlings, which can now support real-time live imaging and elucidate how plants dynamically translocate nitrate between tissues or cell types.

Our results reveal a regulatory mechanism that photosynthetic plants use to sense inorganic nitrogen, which then activates plant signaling networks and growth responses. Our insights may suggest avenues through which to enhance nitrogen utilization efficiency in crops, reduce fertilizer and energy inputs, and mitigate climate change arising from the emission of greenhouse gases in the interest of supporting more-sustainable agriculture (*1*, *5*, *6*, *35*).

## REFERENCES AND NOTES


1. A. J. Bloom, *Curr. Opin. Plant Biol.* **25**, 10–16 (2015).
2. R. E. Mason *et al.*, *Science* **376**, eabh3767 (2022).
3. M. Stitt, *Curr. Opin. Plant Biol.* **2**, 178–186 (1999).
4. N. M. Crawford, B. G. Forde, *Arabidopsis Book* **2002**, e0011 (2002).
5. E. A. Vidal *et al.*, *Plant Cell* **32**, 2094–2119 (2020).
6. Y. Y. Wang, Y. H. Cheng, K. E. Chen, Y. F. Tsay, *Annu. Rev. Plant Biol.* **69**, 85–122 (2018).
7. L. Li, K. H. Liu, J. Sheen, *Annu. Rev. Cell Dev. Biol.* **37**, 341–367 (2021).
8. A. Gaudinier *et al.*, *Nature* **563**, 259–264 (2018).
9. L. Castaings *et al.*, *Plant J.* **57**, 426–435 (2009).
10. P. Walch-Liu, B. G. Forde, *Plant J.* **54**, 820–828 (2008).
11. C. H. Ho, S. H. Lin, H. C. Hu, Y. F. Tsay, *Cell* **138**, 1184–1194 (2009).
12. R. Wang, X. Xing, Y. Wang, A. Tran, N. M. Crawford, *Plant Physiol.* **151**, 472–478 (2009).
13. M. Konishi, S. Yanagisawa, *Plant J.* **63**, 269–282 (2010).
14. L. Schauser, W. Wieloch, J. Stougaard, *J. Mol. Evol.* **60**, 229–237 (2005).
15. M. Konishi, S. Yanagisawa, *Nat. Commun.* **4**, 1617 (2013).
16. C. Marchive *et al.*, *Nat. Commun.* **4**, 1713 (2013).
17. C. Chardin, T. Girin, F. Roudier, C. Meyer, A. Krapp, *J. Exp. Bot.* **65**, 5577–5587 (2014).
18. K. H. Liu *et al.*, *Nature* **545**, 311–316 (2017).
19. J. M. Alvarez *et al.*, *Nat. Commun.* **11**, 1157 (2020).
20. M. Konishi, T. Okitsu, S. Yanagisawa, *J. Exp. Bot.* **72**, 5735–5750 (2021).
21. V. Niemann *et al.*, *J. Mol. Biol.* **426**, 1539–1553 (2014).
22. J. Canales, T. C. Moyano, E. Villarroel, R. A. Gutiérrez, *Front. Plant Sci.* **5**, 22 (2014).
23. W. Ma *et al.*, *Plant J.* **78**, 70–79 (2014).
24. K. H. Liu, C. Y. Huang, Y. F. Tsay, *Plant Cell* **11**, 865–874 (1999).
25. M. Konishi, S. Yanagisawa, *Biochem. Biophys. Res. Commun.* **411**, 708–713 (2011).
26. W. Suzuki, M. Konishi, S. Yanagisawa, *Plant Signal. Behav.* **8**, e25975 (2013).
27. F. Fichtner, I. M. Dissanayake, B. Lacombe, F. Barbier, *Trends Plant Sci.* **26**, 352–374 (2021).
28. J. O. Lundberg, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 13144–13145 (2012).
29. J. L. Parker, S. Newstead, *Nature* **507**, 68–72 (2014).
30. S. Q. Hutsell, R. J. Kimple, D. P. Siderovski, F. S. Willard, A. J. Kimple, *Methods Mol. Biol.* **627**, 75–90 (2010).
31. M. Mita *et al.*, *Anal. Chem.* **91**, 4821–4830 (2019).
32. M. Baek *et al.*, *Science* **373**, 871–876 (2021).
33. J. Jumper *et al.*, *Nature* **596**, 583–589 (2021).
34. G. Unden, S. Nilkens, M. Singenstreu, *Dalton Trans.* **42**, 3082–3087 (2013).
35. D. Coskun, D. T. Britto, W. Shi, H. J. Kronzucker, *Nat. Plants* **3**, 17074 (2017).


## ACKNOWLEDGMENTS


Our research was inspired by the pioneer works on NLP7 and CHL1 by A. Krapp, Y. F. Tsay, and N. Crawford. We thank J. Shin and M. McCormack for their advice on RNA-seq data analysis and L. Li, F. Yu, X. Liu, and A. Diener for comments. **Funding:** Supported by National Natural Science Foundation of China grants 32170270 and 31870227 (K.-H.L.); National Institutes of Health grants GM060493 and GM129093 (J.S.); Northwest A&F University for scientific research start-ups grant Z1090121029 (B.C.); and JSPS KAKENHI grant 22H04977 (S.Y.). **Author contributions:** K.-H.L., Z.-F.W., and J.S., designed the experiments. K.-H.L., Z.-F.W., M.L., Z.L., B.C., C.L., and A.G. performed the experiments. K.-H.L., M.L., and J.S. wrote the manuscript with input from all authors. **Competing interests:** The authors declare no competing interests. G.W. is cofounder and has equity in Enanta Pharmaceuticals, PIC Therapeutics, Eutropics, Olaris Therapeutics, Skinap Therapeutics, Cellmig Biolabs, NOW Scientific, Virtual Discovery, and QuantumTx. These companies did not participate in any aspect of the study's experimental design or analyses or in the preparation of the manuscript. **Data and materials availability:** All data are available in the main text or supplementary materials. The plasmids and the transgenic *Arabidopsis* seeds generated in this study are available upon request and subject to a material transfer agreement. The RNA-seq raw data generated in this study have been deposited at the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE198475.



## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.add1104
Materials and Methods
Figs. S1 to S8
Tables S1 to S6
References (*36*–*45*)
MDAR Reproducibility Checklist


Submitted 21 May 2022; accepted 16 August 2022
10.1126/science.add1104


# REPORTS

SOLAR CELLS

# Deterministic fabrication of 3D/2D perovskite bilayer stacks for durable and efficient solar cells


**Siraj Sidhik[1,2], Yafei Wang[2,3], Michael De Siena[4], Reza Asadpour[5], Andrew J. Torma[6], Tanguy Terlier[7], Kevin Ho[8], Wenbin Li[2,6], Anand B. Puthirath[1], Xinting Shuai[1], Ayush Agrawal[2], Boubacar Traore[9], Matthew Jones[1,10], Rajiv Giridharagopal[8], Pulickel M. Ajayan[1], Joseph Strzalka[11], David S. Ginger[8], Claudine Katan[9], Muhammad Ashraful Alam[5], Jacky Even[12]*, Mercouri G. Kanatzidis[4], Aditya D. Mohite[1,2]***



**Realizing solution-processed heterostructures is a long-enduring challenge in halide perovskites because of solvent incompatibilities that disrupt the underlying layer. By leveraging the solvent dielectric constant and Gutmann donor number, we could grow phase-pure two-dimensional (2D) halide perovskite stacks of the desired composition, thickness, and bandgap onto 3D perovskites without dissolving the underlying substrate. Characterization reveals a 3D–2D transition region of 20 nanometers mainly determined by the roughness of the bottom 3D layer. Thickness dependence of the 2D perovskite layer reveals the anticipated trends for n-i-p and p-i-n architectures, which is consistent with band alignment and carrier transport limits for 2D perovskites. We measured a photovoltaic efficiency of 24.5%, with exceptional stability of $T_{99}$ (time required to preserve 99% of initial photovoltaic efficiency) of >2000 hours, implying that the 3D/2D bilayer inherits the intrinsic durability of 2D perovskite without compromising efficiency.**


The progressive increase in the power conversion efficiency (PCE) of solution-processed perovskite solar cells (PSCs) (*1*, *2*) has been enabled in part by strategies to passivate the grain boundaries and interfaces between the perovskite absorber and the charge transport layers (*3*–*9*). Two-dimensional (2D) halide perovskite (HaP) passivation layers, which have been the most effective in improving the open-circuit voltage ($V_{OC}$) and fill factor (FF) (*10*–*13*), are commonly grown by spin-coating an organic cation dispersed in isopropyl alcohol or chloroform on top of 3D HaPs (*14*, *15*). This coating removes some excess lead iodide ($PbI_2$) from the 3D perovskite layer to then form heterogeneous 2D phases or ultrathin layers of wide-bandgap 2D HaP (*16*–*18*).

These advances have enhanced durability, as demonstrated recently by Azmi *et al.*, using damp-heat tests (*19*); however, the lack of control over the phase purity, film thickness, orientation, and structural phase of the 2D HaP has limited their use as an interfacial passivation layer (*20*). A solvent-free growth of the 2D $BA_2PbI_4$ perovskite on the 3D film by controlling the pressure, temperature, and time was demonstrated by Jang *et al.*, indicating the importance of a high-quality 3D/2D interface (*21*). However, such solid-state in-plane growth is difficult to scale to large areas. Thus, the fabrication of solution-processed heterostructures of 3D/2D HaP with the desired energy levels, thickness, and orientation has been lacking.

We report a solvent design principle for fabricating solution-processed 3D/2D HaP bilayer structures with the desired film thickness and phase purity of any 2D HaP—including Ruddlesden-Popper (RP), Dion-Jacobson (DJ), or alternating cation interlayer (ACI)—described by the general formula of $L'A_{n-1}B_nX_{3n+1}$ (DJ) where L′ is a long-chain organic cation, A is a small monovalent cation, B is a divalent metal, X is a monovalent anion, and $n$ is the number of $PbI_6$ bonded octahedra along the stacking axis. Our approach leverages two essential properties of the processing solvents, the dielectric constant ($\varepsilon_r$) and the Gutmann donor number ($D_N$), which controls the coordination between the precursor ions and the solvent (*22*). Processing solvents with dielectric constant $\varepsilon_r > 30$ and Gutmann number, $5 < D_N < 18$ kcal/mol could effectively dissolve the 2D HaP powders without dissolving or degrading the underlying 3D perovskite film during processing by using spin coating, doctor blading, or slot die coating.

Control over the different $n$ value and film thicknesses allowed us to progressively tune the heterostructure from a type I to a type II,

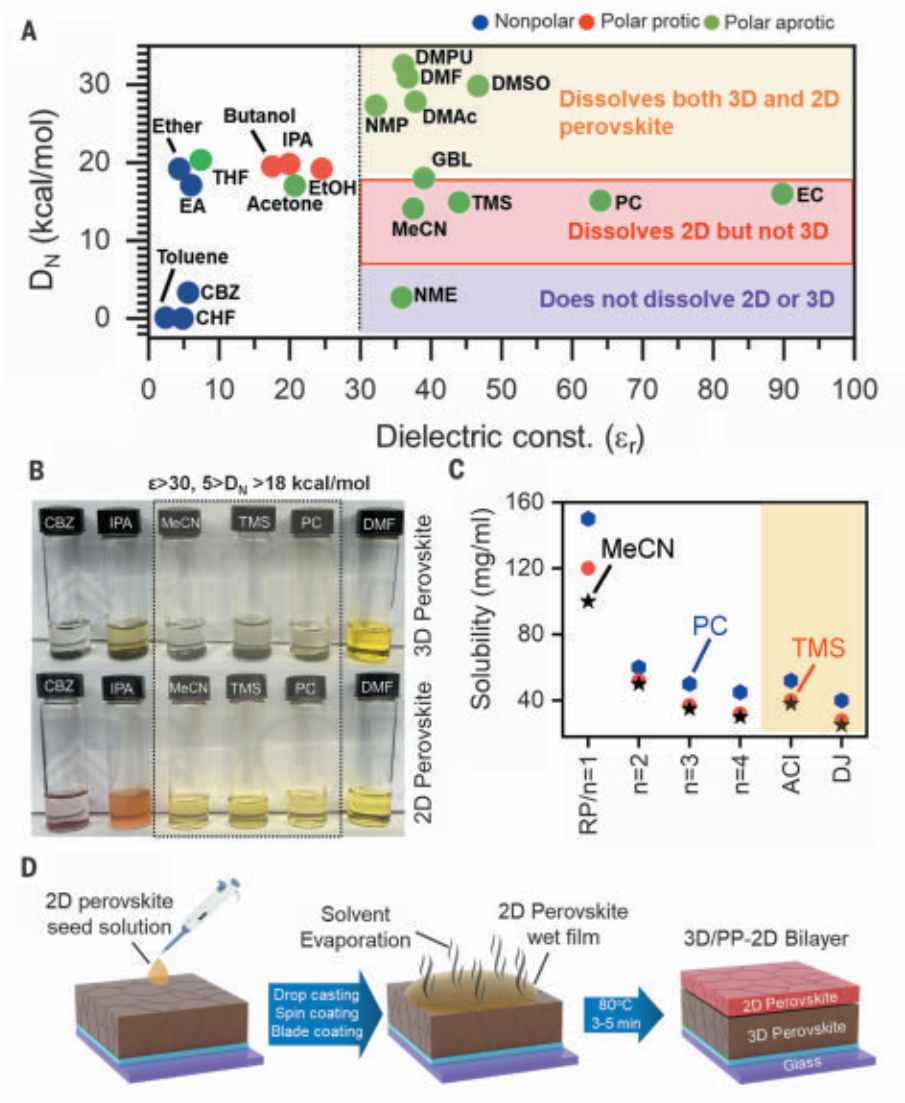



**Fig. 1. Design principle for fabricating a solution-processed 3D/PP 2D HaP bilayer stack.** (**A**) Plot showing different solvents based on the dielectric constant ($\varepsilon_r$) and the Gutmann number ($D_N$) to identify the differences in solubility of the 3D and 2D perovskite powders for making a 3D/2D bilayer stack. (**B**) Optical images of 2D (RP $BA_2MAPb_2I_7$) and 3D perovskite powders in different solvents. (**C**) Solubility of RP $BA_2MA_{n-1}Pb_nI_{3n+1}$ ($n$ = 1 to 4), DJ, (4AMP)$MAPb_2I_7$, and ACI, (GA)$MA_2Pb_2I_7$–based 2D perovskites (where BA is butylamine, MA is methylammonium, AMP is aminomethyl piperidine, and GA is guanidinium) in the solvents MeCN, TMS, and PC as shown. (**D**) Fabrication steps of a 3D/2D HaP bilayer stack. CHF, chloroform; CBZ, chlorobenzene; EA, ethyl acetate; IPA, isopropyl alcohol; NMP, *N*-methyl pyrrolidone; NME, nitromethane; DMF, dimethylformamide; MeCN, acetonitrile; DMAc, dimethylacetamide; GBL, γ-butyrolactone; TMS, tetramethylene sulfone; THF, tetrahydrofuran; DMSO, dimethyl sulfoxide; PC, propylene carbonate; EC, ethylene carbonate; DMPU, *N*,*N*′-Dimethylpropyleneurea.

with the 2D perovskites acting as a transport layer. We achieved a PCE of 24.5% with a high $V_{OC}$ of 1.20 V in a regular n-i-p device using a RP 2D $BA_2MA_2Pb_3I_{10}$ perovskite with a thickness of 50 nm. A comparison of the International Summit on Organic Photostability ISOS-L-1 protocol [maximum power point tracking under ambient conditions] stability of 3D/phase-pure 2D (PP-2D) HaP bilayer PSC with the 2D-passivated 3D, control-3D, and control-2D PSCs showed that the 3D/PP-2D HaP bilayer device exhibited exceptional stability with $T_{99}$ (time required to preserve 99% of the initial PCE) > 2000 hours. Thus, these structures had the durability of the 2D perovskite films without compromising PCE (*23*).

The comprehensive selection criteria for solvents that could selectively dissolve either the 2D or 3D HaP without disrupting the underlying 3D or 2D layer, respectively, were based on the dielectric constant and the Gutmann donor number (Fig. 1A). These two distinct

[1]Department of Material Science and Nanoengineering, Rice University, Houston, TX 77005, USA. [2]Department of Chemical and Biomolecular Engineering, Rice University, Houston, TX 77005, USA. [3]School of Mechanical and Electric Engineering, Guangzhou University, Guangzhou, Guangdong 510006, China. [4]Department of Chemistry and Department of Materials Science and Engineering, Northwestern University, Evanston, IL 60208, USA. [5]School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA. [6]Applied Physics Graduate Program, Smalley-Curl Institute, Rice University, Houston, TX 77005, USA. [7]Shared Equipment Authority, Secure and Intelligent Micro-Systems (SIMS) Laboratory, Rice University, Houston, TX 77005, USA. [8]Department of Chemistry, University of Washington, Seattle, WA 98195, USA. [9]École Nationale Supérieure de Chimie de Rennes (ENSCR), Univ Rennes, CNRS, Institut des Sciences Chimiques de Rennes (ISCR)–UMR 6226, F-35000 Rennes, France. [10]Department of Chemistry, Rice University, Houston, TX 77005, USA. [11]X-Ray Science Division, Argonne National Laboratory, Argonne, IL 60439, USA. [12]Institut National des Sciences Appliquées (INSA) Rennes, Univ Rennes, CNRS, Institut Fonctions Optiques pour les Technologies de l'Information (FOTON)–UMR 6082, F-35000 Rennes, France.
***Corresponding author. Email: adm4@rice.edu (A.D.M.); jacky.even@insa-rennes.fr (J.E.)**

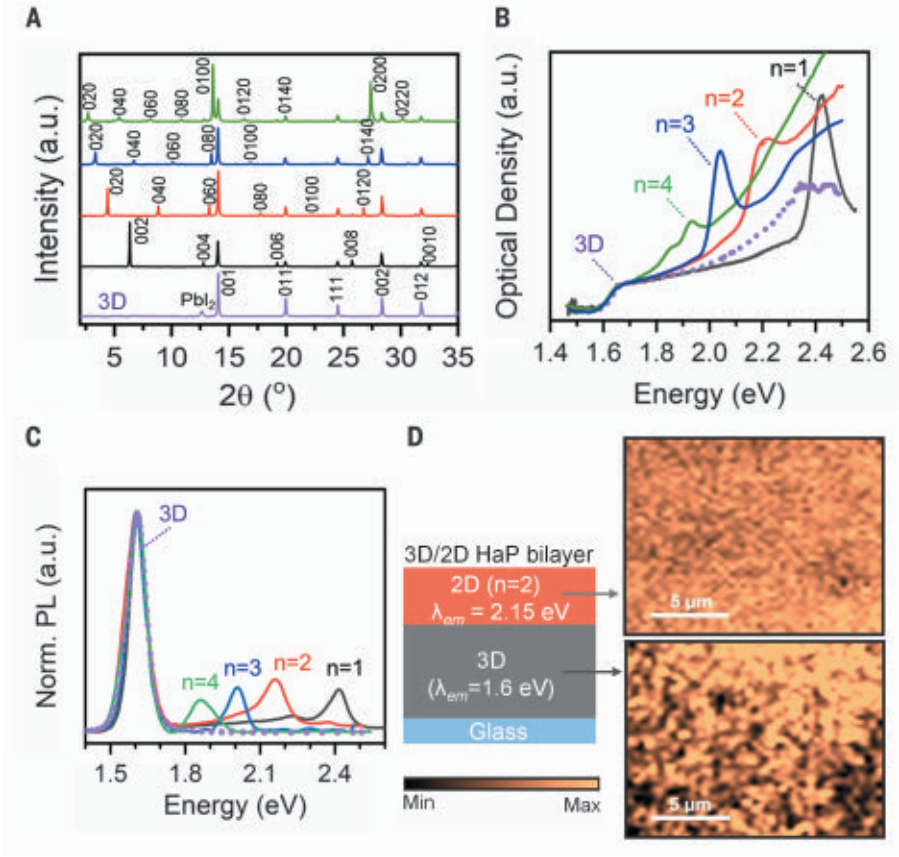



**Fig. 2. Structural and optical spectroscopic characterization of 3D and 3D/PP-2D HaP bilayers with $n$ = 1 to 4.** (**A**) X-ray diffraction pattern from control 3D $Cs_{0.05}(MA_{0.10}FA_{0.85})Pb(I_{0.90}Br_{0.10})_3$ and 3D/PP-2D ($BA_2MA_{n-1}Pb_nI_{3n+1}$; $n$ = 1 to 4, 100 nm) stack. (**B**) Optical absorbance spectra of 3D HaP film and 3D/PP-2D HaP bilayers for RP $BA_2MA_{n-1}Pb_nI_{3n+1}$ ($n$ = 1 to 4) identified through the excitonic peak. (**C**) PL spectra of 3D and 3D/PP-2D ($BA_2MA_{n-1}Pb_nI_{3n+1}$; $n$ = 1 to 4, 100 nm) HaP bilayer measured with excitation at 480 nm and 360 Watts/cm$^2$ with emission from 3D band edge and 2D HaP from $n$ = 1 to 4. (**D**) Schematic diagram depicting the 3D/2D ($BA_2MAPb_2I_7$) bilayer stack with the measured confocal PL map centered at the 3D emission peak (1.6 eV), and at 2D emission peak (2.15 eV) showing uniform coverage of the 2D layer and the underlying 3D layer.

attributes are correlated. The dielectric constant determines the power of a solvent to dissolve any ionic compound by screening the Coulomb attraction between the ions, whereas the Gutmann donor number describes the Lewis basicity of the solvent and measures the extent to which coordination compounds may form between solvent and cations (high donor number) or between the precursors themselves (low donor number) in the absence of competitive binding of the solvent (*24*, *25*). For example, in the precursor solution composed of methylammonium iodide (MAI), formamidinium iodide (FAI), and $PbI_2$, a solvent with a high Gutmann number would strongly coordinate with divalent metal centers ($Pb^{2+}$) and suppress the formation of molecular clusters {for example, iodoplumbates such as $[PbI_{6-x}(solvent)_x]^{-(4-x)}$, where $x \leq 6$} that would otherwise form in a solvent with low Gutmann number. The strength of these interactions determined the differences in the solubility of the 3D and 2D HaP powders in various solvents (Fig. 1A).

Typical vials of RP 2D ($BA_2MAPb_2I_7$) and the 3D HaP powders (Fig. 1B) illustrate their solubility in different solvents. Polar aprotic solvents (Fig. 1A, green dots) with a dielectric constant > 30, such as *N*, *N*-dimethylformamide (DMF) and dimethyl sulfoxide (DMSO), fully dissolved both 3D and 2D HaP. Solvents such as acetonitrile (MeCN), tetramethyl sulfone (TMS), propylene carbonate (PC), and ethylene carbonate (EC) also have a dielectric constant > 30 but did not dissolve the 3D HaP powders because their weak Lewis basicity ($D_N$ < 18 kcal/mol) made the formation of $Pb^{2+}$ solvent coordination complexes unlikely. However, these solvents completely dissolved the 2D perovskite powders. This difference is consistent with 3D perovskite lattices being more stable and difficult to disrupt with solvents of intermediate $D_N$ than the 2D perovskites and implied the presence of additional favorable interactions of these solvents with the organic spacer cations (readily accessible from the edges of the slabs) that were absent in the 3D perovskites.

We observed that both nonpolar (Fig. 1A, blue dots) and polar protic solvents (Fig. 1A, red dots) did not completely dissolve the 2D HaP powders (Fig. 1B). As controls, we also tested solvents with a high Gutmann number and low dielectric constant, such as tetrahydrofuran ($D_N$ = 20 kcal/mol, $\varepsilon_r$ = 7.6), and vice versa, such as nitromethane (NME) ($D_N$ = 2.7 kcal/mol, $\varepsilon_r$ = 35.9), both of which did not dissolve the 2D perovskite powders (figs. S1 and S2). Taken together, these results implied that solvents with a dielectric constant > 30 and the Gutmann number 5 < $D_N$ < 18 kcal/mol should enable the fabrication of 3D/2D HaP bilayers without disrupting or degrading the underlying 3D HaP film.

To fabricate 3D/2D HaP bilayers with different phases of 2D perovskite and having desired $n$ values and thicknesses, we tested the solubility of the archetypical RP 2D perovskites ($BA_2MA_{n-1}Pb_nI_{3n+1}$, with $n$ = 1 to 4) and other crystal phases such as DJ [(4AMP)$MAPb_2I_7$] and ACI [(GA)$MA_2Pb_2I_7$] in the identified 2D perovskite-selective solvents (Fig. 1C). Because of its high polarity, propylene carbonate dissolved the 2D RP phases better than tetramethyl sulfone or acetonitrile. In general, we observed an overall decrease in the solubility of RP 2D HaPs with increasing $n$ value, from >100 mg/ml for $n$ = 1 to 30 to 40 mg/ml for $n$ = 4. The decrease in the solubility as a function of increasing $n$ value was consistent with the increase in inorganic lattice fraction of the 2D HaP as the $n$ value approached 3D composition.

In addition, the $n$ = 2 ACI and DJ perovskites exhibited low solubilities of 20 to 40 mg/ml and 10 to 25 mg/ml, respectively, which was consistent with the structure of ACI and DJ 2D perovskites, which was near that of 3D with short interlayer cations that reflected the role of organic spacer cations in the dissolution process. Of all the target solvents, the high volatility of MeCN [boiling point (b.p.) ≈ 82°C] compared with those of the others—such as TMS (b.p. ≈ 285°C), PC (b.p. ≈ 242°C), and EC (b.p. ≈ 248°C)—made it attractive for low-temperature processing without affecting the stability of the entire stack (figs. S3 and S4 and movies S1 and S2). We focused on the solvent MeCN for fabricating the targeted 3D/2D HaP bilayer.

The protocols for fabricating the 3D/2D HaP bilayer by use of spin coating, drop-casting, blade coating, or slot-die coating (Fig. 1D) followed our recent work on obtaining PP-2D HaP films. We created a stable dispersion of 2D perovskite seed solution by dissolving the parent crystal powders in MeCN (fig. S5 and supplementary text) (*26*, *27*). In general, other than

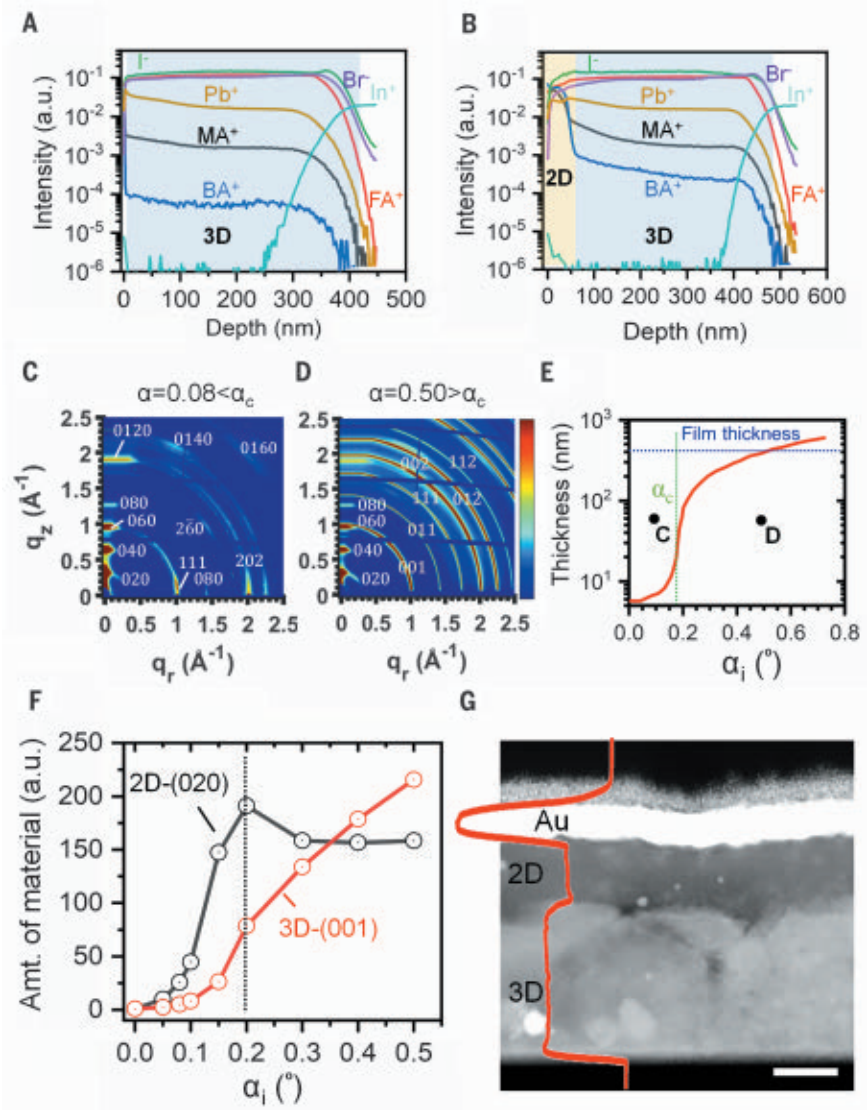



**Fig. 3. 3D/PP-2D HaP interface characterization.** (**A** and **B**) ToF-SIMS depth profiles of the (A) control $Cs_{0.05}(MA_{0.10}FA_{0.85})Pb(I_{0.90}Br_{0.10})_3$ 3D perovskite film and (B) 3D/PP-2D ($BA_2MAPb_2I_7$) HaP stack deposited on top of indium tin oxide show the distribution of different ions across the thickness and the sharpness of the heterointerface. (**C** and **D**) Angle-dependent GIWAXS pattern of the 3D/PP-2D (RP $BA_2MAPb_2I_7$) HaP stack shows evolution of 2D and 3D perovskites for increasing incident angle. The most efficient 3D/PP-2D HaP solar cell was obtained for $BA_2MA_2Pb_3I_{10}$ RP perovskite, which exhibited mixed vertical orientation (Fig. 4). (**E**) Simulated x-ray penetration depth curve for the RP-2D HaP at various incident angles showing the critical angle and the thickness of the probed film. (**F**) Amount of 2D and 3D HaP materials extracted from the angular integrated diffraction peaks. (**G**) Cross-sectional dark-field HR-TEM image of the 3D/PP-2D ($BA_2MAPb_2I_7$) HaP stack with the overlaying intensity profile (red) shows transition width from 3D to 2D of 20 nm. Scale bar, 100 nm.

the solvents with Gutmann number $5 < D_N <$ 18 kcal/mol, any $Pb^{2+}$-weakly coordinating solvents such as DMF, GBL, and DMAc compared with DMSO can produce a dispersion of 2D perovskite seed solution. The precursor solution with 2D perovskite seeds was coated on top of a 3D HaP layer [$Cs_{0.05}(MA_{0.10}FA_{0.85})Pb(I_{0.90}Br_{0.10})_3$] and then annealed at 80°C for 5 min. We also used other scalable techniques such as doctor blading, drop-casting, and blade-coating (fig. S6) to demonstrate the industrial viability of the process. By controlling the concentration and deposition technique, we tuned the thickness of the 2D perovskite layer on top of the 3D film ranging from sub–10 nm to submicrometer scales (fig. S7).

Using the design approach described in Fig. 1, we fabricated a 3D/2D HaP [RP $BA_2MA_{n-1}Pb_nI_{3n+1}$ ($n$ = 1 to 4)] bilayer stack and characterized its phase purity using x-ray diffraction (XRD), optical absorbance, and photoluminescence (PL) techniques. The XRD pattern for the control-3D and 3D/2D HaP bilayers with different 2D RP $BA_2MA_{n-1}Pb_nI_{3n+1}$ perovskite ($n$ = 1 to 4) (Fig. 2A), matching the low-angle Bragg peaks with the corresponding simulated patterns suggested that the overlying 2D HaP layer had high phase purity (fig. S8). The phase purity was further confirmed with absorbance (Fig. 2B) and PL (Fig. 2C) measurements. The optical absorbance revealed the presence of a 3D HaP band-edge at 1.6 eV for the control film, which was accompanied by a sharp excitonic peak varying from 2.4 eV ($n$ = 1) to 1.9 eV ($n$ = 4) corresponding to the different $n$ values in the 3D/2D bilayer. The steady-state PL measurements further confirmed the phase purity of the 3D/2D HaP bilayers. We did not observe any change in the bandgap of the underlying 3D HaP, indicating that the presence of the 2D HaP layer did not disrupt the 3D perovskite.

To further assess the spatial homogeneity of the 3D/2D HaP bilayer, we monitored the PL emission centered at the 3D peak (1.6 eV) and the 2D peak ($n$ = 2, 2.15 eV) using confocal microscopy (Fig. 2D). We observed a uniform emission over a large area of 20 by 20 μm from the 2D as well as the underlying 3D HaP layer, confirming the homogeneity of the 3D/2D HaP bilayer stack. These measurements indicated that we had grown highly crystalline PP-2D HaP of varying $n$ values on the 3D surface. This fabrication process also worked well with the pure $FAPbI_3$ and $MAPbI_3$ 3D HaP, demonstrating its broad applicability (fig. S9).

We used several techniques to characterize the interface between the 3D and PP-2D layers. The time-of-flight secondary ion mass spectrometry (TOF-SIMS) depth profile of the control 3D (Fig. 3A) and 3D/PP-2D HaP bilayer (Fig. 3B) revealed the distribution of ions as a function of film thickness. Upon introduction of $BA_2MAPb_2I_7$ 2D perovskite layer on top of the 3D film, in addition to the $Cs_2I^+$ and $Cs_2Br^+$ ions and $CH_5N_2^+$ ($FA^+$) cations, we observed higher intensity of the $C_4H_{12}N^+$ ($BA^+$) compared with the background $BA^+$ intensity (supplementary text) and $CH_6N^+$ ($MA^+$) ions. These ions were uniformly distributed for a thickness of 50 nm, after which there was a sharp decrease in the concentration. Fitting the change in intensity of the $BA^+$ to the background (representing no BA) yielded an interface transition of 20 nm from the top 2D layer to the underlying 3D HaP. The TOF-SIMS measurements of the 3D|PP-2D (RP $BA_2MA_2Pb_3I_{10}$) bilayer stack show similar results (fig. S10).

The grazing incidence wide-angle x-ray scattering (GIWAXS) spectra of the 3D/PP-2D HaP bilayer for different incident angles (or depth) (Fig. 3, C and D) shows the presence of horizontal oriented $BA_2MAPb_2I_7$ 2D HaP on the surface ($\alpha$ = 0.08°) and the 3D

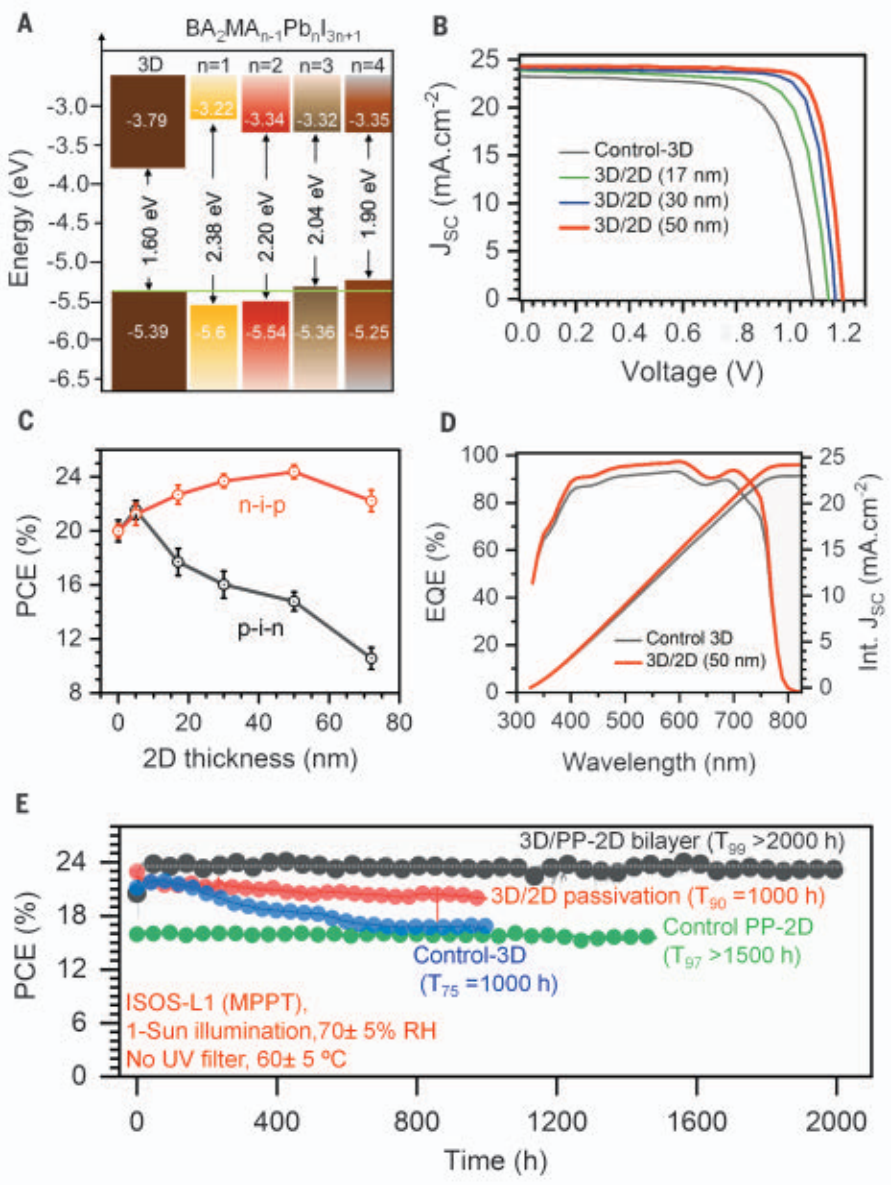



**Fig. 4. Photovoltaic performance and long-term stability of the 3D/PP-2D ($BA_2MA_2Pb_3I_{10}$) HaP bilayer solar cells.** (**A**) Energy-level alignment for different $n$ values ($n \leq 4$) of 2D perovskite with the 3D perovskite layer with an error of ±0.05eV. (**B**) Current-voltage (*I-V*) curves of the champion 3D/PP-2D n-i-p PSCs as a function of the 2D layer thickness obtained by spin coating different concentration of the 2D perovskite solution in MeCN. (**C**) Variation in PCE of the n-i-p and p-i-n planar 3D/PP-2D PSCs as a function of 2D perovskite layer thickness. (**D**) External quantum efficiency of the device with and without the 2D layer, showing the absorption and current generation ability of the stack. (**E**) ISOS-L-1 stability measured at maximum power point tracking in ambient condition under continuous 1-sun illumination (55°C) for an epoxy encapsulated PSC. The initial PCE of the control device is 21%; that of the 3D/2D passivated device is 22.93%; that of the 3D/PP-2D bilayer PSC is 23.75%; and that of the PP-2D perovskite device is 16.3%.

HaP [$Cs_{0.05}(MA_{0.10}FA_{0.85})Pb(I_{0.90}Br_{0.10})_3$] in the form of rings at the bottom ($\alpha = 0.5°$) (fig. S11 and supplementary text). This result is well correlated with the calculated x-ray penetration depth curve for the RP-2D HaP, showing a critical angle ($\alpha_c$) of 0.181° (Fig. 3E) (*28*, *29*). The integrated area under the (020) peak (2D signature feature) and the (111)/(001) peak (3D signature feature) from the GIWAXS spectra for various incident angles (0° to 0.5°) (Fig. 3F) revealed the 3D/PP-2D interface at an incident angle of 0.2°. The integrated area under the (020) peak became invariant, whereas the (111)/(001) increased, which confirmed the presence of a 2D HaP layer with a thickness of 50 to 60 nm stacked on top of the 3D HaP film as estimated from the penetration depth curve. The $BA_2MA_2Pb_3I_{10}$ 2D HaP shows a mixed orientation in the 3D/PP-2D ($n = 3$, RP) HaP bilayer stack with similar interfacial characteristics (detailed GIWAXS analysis is available in fig. S12). The presence of some texture in the same orientation as in the $n = 3$, RP perovskite is required for efficient charge transfer through the 2D perovskite, resulting in the efficient 3D/PP-2D bilayer HaP solar cells (Fig. 4) (*26*, *27*). The dark-field cross-sectional high-resolution transmission electron microscopy (HR-TEM) image of the 3D/PP-2D (RP $BA_2MAPb_2I_7$) HaP bilayer grown on the silicon substrate (with gold and carbon as the top protective layer) (Fig. 3G) verified the homogeneity of the 2D HaP layer on top of the 3D HaP thin film. The average intensity profile (Fig. 3G, red) showed an interfacial 3D/PP-2D sharpness of 20 nm, which was consistent with the TOF-SIMS measurements and matched the roughness of the control-3D film (fig. S13). The HR-TEM image of the 3D/PP-2D (RP $BA_2MA_2Pb_3I_{10}$) HaP bilayer stack also showed similar interfacial characteristics (fig. S14). All of these results demonstrated that our solvent design strategy allowed for the fabrication of 3D/PP-2D bilayers with a high-quality interface without destroying or altering the crystallinity of the underlying 3D perovskite.

We fabricated solar cells with the 3D/PP-2D HaP bilayer perovskites and measured their performance and durability. We first created an energy-level diagram for the 2D HaP, RP $BA_2MA_{n-1}Pb_nI_{3n+1}$ ($n$ = 1 to 4), with the 3D perovskite $Cs_{0.05}(MA_{0.10}FA_{0.85})Pb(I_{0.90}Br_{0.10})_3$ (Fig. 4A) according to values predicted from first-principles calculations (fig. S15) and corroborated with photoemission yield spectroscopy (PES) and absorption measurements (*30*) (figs. S16 to S18). We observed a type II band alignment between the 3D and the $n = 3$ 2D HaP with a near-perfect alignment of valence band edges for 3D and 2D HaP, which is ideal for extracting holes but presents a large energy barrier for electrons (fig. S19). The 3D/PP-2D ($n$ = 1, $n$ = 2) HaP bilayer stack shows a type I alignment, which was further confirmed by fabricating solar cells showing a barrier for charge extraction using higher thicknesses of the corresponding 2D layers (supplementary text and fig. S20). As a result, we fabricated n-i-p PSCs with the architecture FTO/$SnO_2$/3D/PP-2D ($n = 3$)/spiro-OMeTAD/Au, controlling the PP-2D ($n = 3$) HaP thickness by spin-coating different concentrations of 2D perovskite solution in MeCN (fig. S21).

The current-voltage (*I-V*) characteristics for three selected thicknesses of 2D HaP layers are shown in Fig. 4B, with 3D HaP as the control. For n-i-p cells, increased 2D HaP thickness

from 0 to 50 nm increased $V_{OC}$ from 1.09 to 1.20 V. This increase in the $V_{OC}$ was accompanied by a slight increase in the FF from 0.80 to 0.84, and even a small increase in $J_{SC}$ from 23.54 to 24.34 mA/cm$^2$ resulting in a peak PCE of 24.5% for a 2D HaP thickness of 50 nm with no hysteresis (figs. S22 and S23). The PCE as a function of 2D HaP thickness shows that increasing the 2D HaP thickness beyond 50 nm decreased the overall PCE (Fig. 4C). We attributed this decrease in PCE to the reduced transport of the free charge carriers from 3D to the 2D HaP limited by the <100 nm diffusion length for a polycrystalline 2D $n$ = 3, RP HaP film with mixed orientation (*31*, *32*) (PV statistics are provided in fig. S23). However, the p-i-n devices with ITO/PTAA/3D/PP-2D($n$ = 3)/C60/BCP/Cu exhibited an increase in PCE for a 2D HaP thickness of 5 nm, which is consistent with recent studies of 2D/3D interfaces with an ultrathin 2D layer passivation followed by a drastic decrease for higher thicknesses of 2D HaP (Fig. 4C, black curve, and figs. S24 and S25) (*33*, *34*). The reduction in PCE was consistent with the energy band diagram (Fig. 4A and fig. S19B), which showed a large barrier to electron collection in the p-i-n geometry.

To understand the increase in PV parameters ($J_{SC}$, $V_{OC}$, and FF) of the 3D/PP-2D HaP bilayer solar cells, we performed optical and self-consistent transport modeling to simulate the PV characteristics that are supported by the surface photovoltage (SPV), steady-state PL, and time-resolved PL measurements. The optical modeling revealed that the 2D HaP layer increased the photogeneration of the bilayer stack and improved $J_{SC}$ (fig. S26). This result was consistent with the observed increase in the $J_{SC}$ by 1.5 mA/cm$^2$ as we went from the 3D control to the champion 3D/PP-2D HaP bilayer device, validated by external quantum efficiency (EQE) measurements shown in Fig. 4D and the change in EQE (ΔEQE) between them (figs. S27 and S28 and supplementary text).

The self-consistent transport simulation predicted reduced recombination at the 3D/HTL interface after the introduction of a 2D HaP layer, improving the $V_{OC}$ and FF (fig. S29). To corroborate these results, we measured the surface photovoltage (SPV) using scanning Kelvin probe microscopy (SKPM) on the ITO/$SnO_2$/3D/PP-2D (RP $BA_2MA_2Pb_3I_{10}$) stack, which showed an increase in the SPV as a function of the 2D layer thickness (figs. S30 and S31). This result suggested an increase in the quasi-Fermi level separation related to the $V_{OC}$ of the device (*35*–*38*). The dark *I*-*V* curve traces of the solar cells further confirm the increase in $V_{OC}$ (supplementary text and fig. S32). Additionally, the steady-state PL and time-resolved photoluminescence (TRPL) measurements on the ITO/$SnO_2$/3D/PP-2D (RP $BA_2MA_2Pb_3I_{10}$) stack showed enhanced PL emission and increased charge carrier lifetime up to a 2D HaP thickness of 50 nm (fig. S33) but decreases for the higher thickness of 2D HaPs. The overlaying 2D HaP minimized the nonradiative recombination pathways between the 3D perovskite/electrode interface as confirmed by the transport modeling measurements, improving both the FF and the $V_{OC}$ of the 3D/PP-2D devices (*15*, *39*).

We tested the long-term operational stability of our 3D/PP-2D bilayer encapsulated device following the ISOS-L-1 protocol (Fig. 4G) (*23*). After 2000 hours of continuous illumination, the 3D/PP-2D HaP bilayer device showed negligible degradation with $T_{99}$ > 2000 hours, whereas the control 3D device lost 25% of its initial PCE. As controls, we also measured the stability of the 3D PSC passivated with a spin-coated organic cation, butylammonium iodide, and compared it with our 3D/PP-2D PSCs using the same conditions. The 2D HaP passivated 3D PSCs show a 10% loss of efficiency after 1000 hours of continuous operation, which is consistent with other recent reports (*4*, *19*). We also measured a pure 2D HaP control device, which showed a $T_{97}$ > 1500 hours, implying that the 3D/PP-2D bilayer perovskite has acquired the inherent stability of the 2D perovskite material.

## REFERENCES AND NOTES


1. J. Jeong *et al.*, *Nature* **592**, 381–385 (2021).
2. H. Min *et al.*, *Nature* **598**, 444–450 (2021).
3. Q. Jiang *et al.*, *Nat. Photonics* **13**, 460–466 (2019).
4. F. Zhang *et al.*, *Science* **375**, 71–76 (2022).
5. X. Zheng *et al.*, *Nat. Energy* **5**, 131–140 (2020).
6. Y. Liu *et al.*, *Sci. Adv.* **5**, eaaw2543 (2019).
7. N. K. Noel *et al.*, *ACS Nano* **8**, 9815–9821 (2014).
8. D. W. deQuilettes *et al.*, *ACS Energy Lett.* **1**, 438–444 (2016).
9. I. L. Braly *et al.*, *Nat. Photonics* **12**, 355–361 (2018).
10. A. A. Sutanto *et al.*, *Chem* **7**, 1903–1916 (2021).
11. G. Wu *et al.*, *Adv. Mater.* **34**, e2105635 (2022).
12. G. Li *et al.*, *ACS Energy Lett.* **6**, 3614–3623 (2021).
13. M. A. Mahmud *et al.*, *Adv. Funct. Mater.* **32**, 2009164 (2022).
14. E. Jokar *et al.*, *ACS Energy Lett.* **6**, 485–492 (2021).
15. J. J. Yoo *et al.*, *Energy Environ. Sci.* **12**, 2192–2199 (2019).
16. T. Zhang *et al.*, *Joule* **2**, 2706–2721 (2018).
17. P. Li *et al.*, *Adv. Mater.* **30**, e1805323 (2018).
18. F. Wang *et al.*, *Joule* **2**, 2732–2743 (2018).
19. R. Azmi *et al.*, *Science* **376**, 73–77 (2022).
20. M. Xiong *et al.*, *ACS Energy Lett.* **7**, 550–559 (2022).
21. Y.-W. Jang *et al.*, *Nat. Energy* **6**, 63–71 (2021).
22. V. Gutmann, *Coord. Chem. Rev.* **18**, 225–255 (1976).
23. M. V. Khenkin *et al.*, *Nat. Energy* **5**, 35–49 (2020).
24. J. C. Hamill Jr., J. Schwartz, Y.-L. Loo, *ACS Energy Lett.* **3**, 92–97 (2017).
25. J. C. Hamill Jr. *et al.*, *J. Phys. Chem. C* **124**, 14496–14502 (2020).
26. S. Sidhik *et al.*, *Adv. Mater.* **33**, e2007176 (2021).
27. S. Sidhik *et al.*, *Cell Rep. Phys. Sci.* **2**, 100601 (2021).
28. M. Tolan, M. Tolan, *X-Ray Scattering From Soft-Matter Thin Films: Materials Science and Basic Research* (Springer, 1999), vol. 148.
29. R. M. Kennard *et al.*, *Chem. Mater.* **33**, 7290–7300 (2021).
30. I. Spanopoulos *et al.*, *J. Am. Chem. Soc.* **141**, 5518–5534 (2019).
31. M. Seitz *et al.*, *Nat. Commun.* **11**, 2035 (2020).
32. E. D. Kinigstein *et al.*, *Mater. Lett.* **2**, 1360–1367 (2020).
33. M. Degani *et al.*, *Sci. Adv.* **7**, eabj7930 (2021).
34. X. Wang *et al.*, *Nat. Commun.* **12**, 52 (2021).
35. S. Kavadiya *et al.*, in *2020 47th IEEE Photovoltaic Specialists Conference (PVSC)* (IEEE, 2020), pp. 1439–1440.
36. I. Levine *et al.*, *Joule* **5**, 2915–2933 (2021).
37. L. Kronik, Y. Shapira, *Surf. Sci. Rep.* **37**, 1–206 (1999).
38. R. Giridharagopal *et al.*, *ACS Nano* **13**, 2812–2821 (2019).
39. J. Wang *et al.*, *ACS Energy Lett.* **4**, 222–227 (2018).


## ACKNOWLEDGMENTS

We thank M. H. K. Samani and A. B. Marciel for useful discussions on the DLS measurements. **Funding:** The work at Rice University was supported by the DOE-EERE 0008843 program. J.E. acknowledges the financial support from the Institut Universitaire de France and GENCI national computational resources. The work at ISCR and Institut FOTON was performed with funding from the European Union's Horizon 2020 research and innovation program under grant agreement 861985 (Pero CUBE). At Northwestern, the work was supported by the Office of Naval Research (ONR) under grant N00014-20-1-2725. This research used facilities of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract DE-AC02-06CH11357. ToF-SIMS analysis were carried out with support provided by the National Science Foundation CBET-1626418. This work was conducted in part using resources of the Shared Equipment Authority at Rice University. The work at the University of Washington by K.H., R.G., and D.S.G. is supported by DOE BES under award DE-SC0013957. This research used facilities and instrumentation supported by the US National Science Foundation through the UW Molecular Engineering Materials Center (MEM-C), a Materials Research Science and Engineering Center (DMR-1719797). **Author contributions:** A.D.M., S.S., and J.E. conceived the idea, designed the experiments, analyzed the data, and cowrote the manuscript. S.S. and Y.W. studied different solvents. Y.W. performed 2D/3D passivation experiments with S.S.; W.L. and J.S. performed GIWAXS measurements, A.J.T. performed PL maps, X.S. performed SEM and AFM imaging, and A.A. helped with solar cell fabrication and characterization. M.D. performed photoluminescence measurements, M.G.K. and M.J. analyzed the solvation chemistry along with synthesis of 2D HaP crystals. C.K. and B.T. helped understand the 3D/2D interface with DFT modeling. A.B.P. and P.M.A. performed focused ion beam milling (FIB) and electron microscopy (TEM and STEM). K.H., R.G., and D.S.G. performed SKPM and TRPL measurements, R.A. and M.A.A. performed device modeling and optical simulations. T.T. helped perform TOF-SIMS and analyze data. All authors contributed to the manuscript. **Competing interests:** Rice University have filed patent for method of fabricating the 3D/PP-2D bilayer stack. **Data and materials availability:** All data are available in the main text or the supplementary materials. 

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq7652
Materials and Methods
Supplementary Text
Figs. S1 to S33
Tables S1 to S8
References (*40*–*70*)
Movies S1 and S2

Submitted 2 May 2022; accepted 18 August 2022
10.1126/science.abq7652

BIODIVERSITY

# Genetic diversity loss in the Anthropocene

Moises Exposito-Alonso[1,2,3]*, Tom R. Booker[4,5], Lucas Czech[1], Lauren Gillespie[1,6], Shannon Hateley[1], Christopher C. Kyriazis[7], Patricia L. M. Lang[2], Laura Leventhal[1,2], David Nogues-Bravo[8], Veronica Pagowski[2], Megan Ruffley[1], Jeffrey P. Spence[9], Sebastian E. Toro Arana[1,2], Clemens L. Weiß[9], Erin Zess[1]

Anthropogenic habitat loss and climate change are reducing species' geographic ranges, increasing extinction risk and losses of species' genetic diversity. Although preserving genetic diversity is key to maintaining species' adaptability, we lack predictive tools and global estimates of genetic diversity loss across ecosystems. We introduce a mathematical framework that bridges biodiversity theory and population genetics to understand the loss of naturally occurring DNA mutations with decreasing habitat. By analyzing genomic variation of 10,095 georeferenced individuals from 20 plant and animal species, we show that genome-wide diversity follows a mutations-area relationship power law with geographic area, which can predict genetic diversity loss from local population extinctions. We estimate that more than 10% of genetic diversity may already be lost for many threatened and nonthreatened species, surpassing the United Nations' post-2020 targets for genetic preservation.

Anthropogenic habitat loss and climate change (*1*) have led to the extinction of hundreds of species over the past centuries, and ~1 million more species (~25% of all known species) are at risk of extinction (*2*). Studies of species' ranges, however, have detected geographic range reductions in at least 47% of surveyed plant and animal species, likely in response to the past centuries of anthropogenic activities (*3*) [see (*4*) and table S17]. Though this loss might seem inconsequential compared with losing an entire species, this range contraction reduces genetic diversity, which dictates species' ability to adapt to new environmental conditions (*5*). The loss of geographic range can spiral into a feedback loop, where genetic diversity loss further increases the risk of species extinction (*6*, *7*).

Although genetic diversity is a key dimension of biodiversity (*8*), it has been overlooked in international conservation initiatives (*9*). Only in 2021 did the United Nations (UN) Convention of Biological Diversity propose to preserve at least 90% of all species' genetic diversity (*10*, *11*). Recent meta-analyses of animal populations with genetic marker samples have been used as proxies to quantify recent genetic changes (*12*, *13*). However, theory and scalable approaches to estimate genome-wide diversity loss across species do not yet exist, impairing prioritization and evaluation of conservation targets. Here, we introduce a framework to estimate global genetic diversity loss by bridging biodiversity theory with population genetics and by combining data on global ecosystem transformations with newly available genomic datasets.

The first studies that predicted biodiversity reductions in response to habitat loss and climate change in the 1990s and the 2000s projected species extinctions using the relationship of biodiversity with geographic area, termed the species-area relationship (SAR) (*14*) (figs. S1 to S3). In this framework, ecosystems with a larger area ($A$) harbor a larger number of species ($S$) resulting from an equilibrium among limited species dispersal, habitat heterogeneity, and colonization-extinction-speciation dynamics. Thus, the more a study area is extended, the more species are found. The SAR has been empirically shown to follow a power law, $S = A^z$. It scales consistently across continents and ecosystems (*15*), with a higher $z$ characterizing ecosystems that are species rich and highly spatially structured. Given estimates of decreasing ecosystem area over time ($A_{t-1} > A_t$) due to human activities and climate change, Thomas *et al.* (*16*) proposed rough estimates of the percentage of species extinctions in the 21st century ranging from 15 to 37%. Though this may be an oversimplification, the SAR has become a common practical tool for policy groups, including the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) (*2*).

Akin to species richness, within-species variation can be quantitatively described by the number of genetic mutations within a species, defined here as DNA nucleotide variants that appear in individuals of a species. Although population genetics theory has long established that larger populations have higher genetic diversity (*17*), and geographic isolation between populations of the same species results in geographically separated accumulation of different mutations, there have been no attempts to describe the extent of genetic diversity loss driven by species' geographic range reduction using an analogous "mutations-area relationship" (MAR).

We suspected that such a MAR must exist given that another well-known assumption is shared between SARs and population genetics, namely that both species in ecosystems and mutations in populations are typically found in low frequencies, whereas relatively few occur at high frequencies [those that prevail through stochastic genetic drift or are favored by natural selection (*18*)]. This principle of "commonness of rarity" is well known for species and, together with limited spatial dispersal of organisms in the landscape, is a key statistical condition for the power-law SAR (when a study area is expanded, mostly rare local species are newly identified). We then quantified the rarity of mutations using 11,769,920 biallelic genetic variants of the *Arabidopsis thaliana* 1001 Genomes dataset (Fig. 1A) (*19*) by fitting several common models of species abundances (*20*) to the distribution of mutation frequencies ($q$), termed the site frequency spectrum in population genetics (Fig. 1B and figs. S3, S4, S12, and S13). The canonical L-shaped probability distribution ($1/q$) of this spectrum, which is expected under population-equilibrium and the absence of natural selection processes, fits these data well (Fig. 1B), although the more parameter-rich Preston's species abundance log-normal model achieved the best Akaike information criterion (AIC) value (tables S3 and S10). Despite some differences in fit, these models all showcase the similarities of abundance distributions between mutations within species and species within ecosystems, suggesting that they may also behave similarly in their relationship to geographic area (*18*, *20*).

To finally quantify how genetic diversity within a species increases with geographic area, we constructed the MAR by subsampling regions of different sizes across *A. thaliana*'s native range using more than 1000 georeferenced genomes (Fig. 1, A and C). As a metric of genetic diversity, we modeled the number of mutations in space ($M$, also referred to as the number of segregating sites or allelic richness) consistent with the species-centric approach of SAR, which uses species richness as the metric of biodiversity. The MAR also followed the power law relationship $M = cA^z$, where $c$ is a

[1]Department of Plant Biology, Carnegie Institution for Science, Stanford, CA 94305, USA. [2]Department of Biology, Stanford University, Stanford, CA 94305, USA. [3]Department of Global Ecology, Carnegie Institution for Science, Stanford, CA 94305, USA. [4]Department of Zoology, University of British Columbia, Vancouver, Canada. [5]Biodiversity Research Centre, University of British Columbia, Vancouver, Canada. [6]Department of Computer Science, Stanford University, Stanford, CA 94305, USA. [7]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA. [8]Center for Macroecology, Evolution and Climate, GLOBE Institute, University of Copenhagen, Copenhagen, Denmark. [9]Department of Genetics, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. Email: moisesexpositoalonso@gmail.com

constant and the scaling value is $z_{MAR}$ = 0.324 [95% confidence interval (95% CI) = 0.238 to 0.41] (Fig. 1C and tables S4 to S6). The discovered power law is robust to different methods of area quantification, subsampling effects (fully nested, outward or inward), and raster resolution (~10 to 1000 km) and is adjusted for limited sample sizes (figs. S14 to S18 and tables S7 to S9). The expected genetic diversity increase that results only from more individuals being sampled only accounts for $M \approx c\log(A) \approx cA^{z \to 0}$ [see theoretical derivation (*4*); figs. S5 and S6 and tables S1 and S2]; thus, the MAR law is attributed to fundamental evolutionary and ecological forces of population genetic drift and spatial natural selection that cause structuring of genetic diversity across populations (with a maximum $z_{MAR} \to 1$; fig. S5). MAR thus emerges in population genetics coalescent and individual-based simulations in two dimensions (figs. S6 to S8 and S10) and continuous space (fig. S9), as well as in mainland-island community assembly simulations according to the unified neutral theory of biodiversity (fig. S24).

We then asked whether MAR can predict the loss of genetic diversity due to species' range contractions. We explored several scenarios of range contraction in *A. thaliana* by removing in silico grid cells in a map representing populations (Fig. 2B). Our simulations included random local population extinction as if habitat destruction was scattered across large continents, radial expansion of an extinction front due to intense localized mortality, or local extinction in the warmest regions within a species range (*3*, *21*), among others (fig. S18). The MAR-based predictions of genetic loss, using $1 - (1 - A_t/A_{t-1})^{z_{MAR}}$ and assuming $z_{MAR}$ = 0.3, conservatively followed the simulated local loss in *A. thaliana* [pseudo–coefficient of determination ($R^2$) = 0.87 across all simulations] (Fig. 2A and fig. S18).

To test the generality of the MAR, we searched in public nucleotide repositories for datasets of hundreds to thousands of whole-genome-sequenced individuals for the same species sampled across geographic areas within their native ranges (Table 1). In total, we identified 20 wild plant and animal species with such published resources and assembled a dataset amassing a total of 10,095 individuals of these species, with 1522 to 88,332,015 naturally occurring genome-wide mutations per species, covering a geographic area ranging from 0.03 million to 115 million km$^2$. Fitting the MAR for these diverse species, we recovered $z_{MAR}$ values centered around that of *A. thaliana*, with many species overlapping in confidence intervals, and a number of outliers [mean (±SE) $z_{MAR}$ = 0.31 (±0.038), median = 0.26, interquartile range (IQR) = ±0.15, range = 0.10 to 0.82, mean (±SE) $z_{MAR}$ scaled ($z^*_{MAR}$) = 0.26 (±0.048); see Table 1, figs. S5 and S22, table S10, and (*4*)].

**Table 1. MAR across diverse species.** Summary statistics of the number of individuals sampled broadly per species (with the final number of samples analyzed after quality filters in parentheses), the number of naturally occurring mutations discovered through various DNA sequencing methods (see table acronyms), and the total area covered by all the samples within a species (as a convex hull of coordinates). We also report the MAR parameter $z_{MAR}$ and its scaled version for low-sampling genomic effort per species and the percent area that needs to be kept for a species to maintain 90% of its genetic diversity calculated using $z^*_{MAR}$. Protected area predictions are not provided for threatened species because these have likely already lost substantial genetic diversity and require protection of their full geographic range (indicated with "–"). CA, included in the California Endangered Species Act; CR, Red List critically endangered; D, population decline reported in the Red List; GBS, genotyping by sequencing of biallelic SNP markers; GC, genotyping chip; Herb., herbaceous plant; VU, Red List vulnerable; W, whole-genome resequencing or discovery SNP calling.

| Species | Group | Number of samples | Number of mutations | Area (km$^2$ × 10$^6$) | $z_{MAR}$ (95% CI) | $z^*_{MAR}$ (95% CI) | Minimum area$_{90\%}$ (%) |
|---|---|---|---|---|---|---|---|
| *Arabidopsis thaliana* | Herb. | 1,135 (1,001)* | 11,769,920 (W) | 27.34 | 0.324 (0.238–0.41) | 0.312 (0.305–0.32) | 71–78 |
| *Arabidopsis lyrata* | Herb. | 108 | 17,813,817 (W) | 2.79 | 0.236 (0.218–0.254) | 0.151 (0.137–0.165) | 50–66 |
| *Amaranthus tuberculatus* | Herb. | 162 (155)† | 1,033,443 (W) | 0.80 | 0.109 (0.081–0.136) | 0.142 (0.136–0.149) | 48–65 |
| *Eucalyptus melliodora* (VU) | Tree | 275 | 9,378 (GBS) | 0.95 | 0.466 (0.394–0.538) | 0.403 (0.398–0.407) | 77–82 |
| *Yucca brevifolia* (CA) | Yucca palm | 290 | 10,695 (GBS) | NA¶ | 0.128# (0.109–0.147) | 0.049 (0.037–0.062) | – |
| *Mimulus guttatus* | Herb. or shrub | 521 (286)* | 1,522 (GBS) | 25.14 | 0.274 (0.259–0.29) | 0.231 (0.221–0.241) | 63–73 |
| *Panicum virgatum* | Grass | 732 (576)‡ | 33,905,044 (W) | 6.29 | 0.232 (0.211–0.252) | 0.126 (0.116–0.136) | 43–63 |
| *Panicum hallii* | Grass | 591 | 45,589 (W) | 2.19 | 0.824 (0.719–0.928) | 0.814 (0.745–0.883) | 88–90 |
| *Pinus contorta* | Tree | 929 | 32,449 (GC) | 0.89 | 0.015# (0.014–0.016) | –0.061 (–0.062–0.060) | – |
| *Pinus torreyana* (CR) | Tree | 242 | 478,238 (GBS) | NA¶ | 0.142# (0.142–0.142) | 0.015 (0.015–0.015) | – |
| *Populus trichocarpa* | Tree | 882 | 28,342,826 (W) | 1.12 | 0.275 (0.218–0.332) | 0.165 (0.155–0.176) | 53–67 |
| *Anopheles gambiae* | Mosquito | 1,142 | 52,525,957 (W) | 19.96 | 0.214 (0.164–0.264) | 0.122 (0.111–0.132) | 42–62 |
| *Acropora millepora* (NT) | Coral | 253 | 17,931,448 (W) | 0.03 | 0.246 (0.209–0.283) | 0.287 (0.28–0.294) | 69–77 |
| *Drosophila melanogaster* | Fly | 271§ | 5,019 (W) | 115.21 | 0.437 (0.397–0.477) | 0.325 (0.314–0.336) | 72–79 |
| *Empidonax traillii* (D) | Bird | 219 (199)† | 349,014 (GBS, GC) | 7.03 | 0.214 (0.174–0.254) | 0.074 (0.047–0.102) | 24–54 |
| *Setophaga petechia* (D) | Bird | 199 | 104,711 (GBS) | 15.17 | 0.178 (0.134–0.223) | 0.149 (0.135–0.163) | 49–66 |
| *Peromyscus maniculatus* | Mammal | 80 (78)† | 14,076 (GBS) | 22.61 | 0.488 (0.264–0.713) | 0.683 (0.615–0.751) | 86–88 |
| *Dicerorhinus sumatrensis* (CR) | Mammal | 16 | 8,870,513 (W) | NA¶ | 0.412# (0.369–0.456) | 0.127 (0.11–0.144) | – |
| *Canis lupus* | Mammal | 349 (230)‡ | 1,517,226 (W) | 19.10 | 0.256 (0.232–0.28) | 0.184 (0.175–0.193) | 56–70 |
| *Homo sapiens* | Mammal | 2,504 | 88,332,015 (W) | 80.76 | 0.431# (0.347–0.514) | 0.281 (0.23–0.332) | – |

*Only individuals in the native range were used for the analyses. †Only individuals with available coordinates or matching IDs were used for the analyses. ‡Only natural populations were used, excluding breeds, landraces, and cultivars. §Numbers indicate pools of flies used for pool sequencing rather than individuals. ¶Not applicable. Area was not reported for species with unknown locations or where two or fewer populations were sampled. #Values excluded from global averages were used for conservation applications owing to uncertain estimates, suboptimal genomic data type, or because estimates should not be applied for conservation (i.e., humans or nearly extinct Sumatran rhinoceros) [see (*4*)].

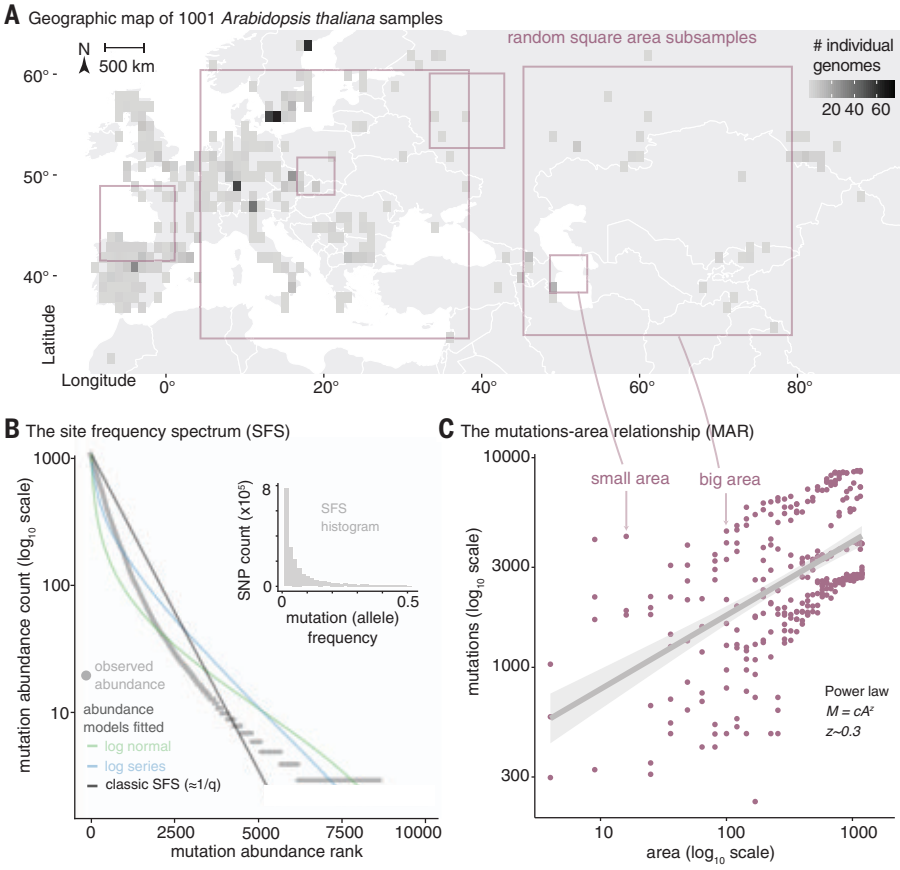



**Fig. 1. Mutations across populations fit models of species abundance distributions and a power law with species range area.** (**A**) Density of individual genomes projected in a 1°-by-1° latitude-longitude map of Eurasia and exemplary subsample areas of different sizes. (**B**) Distribution of mutation [single-nucleotide polymorphism (SNP)] frequencies in 1001 *A. thaliana* plants using a site frequency spectrum (SFS) histogram (gray inset) and a Whittaker's rank abundance curve plot. Also shown are the fitted models of common species abundance functions in *A. thaliana* using a dataset random sample of 10,000 mutations that are also used in (C). (**C**) The MAR in log-log space built from 340 random square subsamples (red dots) of different areas of increasing size within *A. thaliana*'s geographic range along with the number of mutations discovered within each area subset.

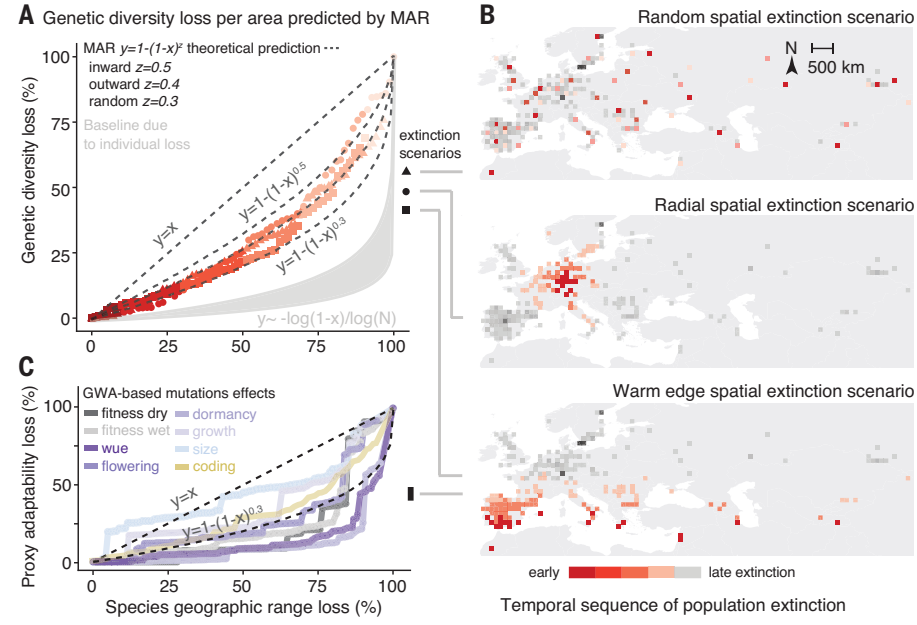



**Fig. 2. The power law of genetic diversity loss with range area loss.** (**A**) Percentage of loss of total genetic diversity in *A. thaliana* from several stochastic simulations (red) of local extinction in (B) and theoretical model projections of genetic diversity loss using the MAR (dashed lines). The expectation for genetic diversity loss based only on individuals is in gray [using starting populations of $N = 10^4$ to $10^9$; see derivation (4)]. (**B**) Illustrated example of several possible range contractions simulated by progressively removing grid cells across the map of Eurasia (red and gray boxes) after different hypothesized spatial extinction patterns. (**C**) A metric of adaptive capacity loss during warm edge extinction in (B) using GWA to estimate effects of mutation on fitness in different rainfall conditions, water-use efficiency (wue), flowering time, seed dormancy, plant growth rate, and plant size. Plotted are the fraction loss of the summed squared effects ($\sum a^2$) of 10,000 mutations from the top 1% tails of effects. Also plotted is the fraction of protein-coding alleles lost (nonsynonymous, stop codon loss or gain, and frameshift mutations; yellow line).

Although we expect species-specific traits related to dispersibility or gene flow to affect $z_{MAR}$ (e.g., migration rate and environmental selection in population genetic simulations significantly influence $z_{MAR}$; table S2), no significant association was found in an analysis of variance (ANOVA) between $z_{MAR}$ and different traits, mating systems, home continents, and so on for the 20 species analyzed (tables S12 to S13). There may be too few species with large population genomic data to find such a signal (Table 1 and tables S12 and S13). For conservation purposes, an average $z_{MAR}$ ~ 0.3 (IQR = ±0.15; Table 1 and table S11) could be predictive of large-scale trends of genetic diversity loss in many range-reduced species that lack genomic information. Further, although species will naturally have different starting levels of total genetic diversity before range reductions due to, for instance, genome size, structure, or mating system differences (*17*), the application of $z_{MAR}$ will provide relative estimates of genetic diversity loss. For instance, assuming $z_{MAR}$ ~ 0.3, we would predict that an area reduction of ~50% creates a loss of ~20% of genetic diversity relative to the total genetic diversity of a given species.

Finally, we used MAR to estimate the average global genetic diversity loss caused by pre–21st century land transformations. Although accurate species-specific geographic area reduction data in the past centuries are scarce, we leveraged global land cover transformations from primary ecosystems to urban or cropland systems (*2*, *22*) (tables S14 to S16). Using the average $z^*_{MAR}$ and several global averages of Earth's land and coastal transformations for present day [38% global area transformation from (*22*), 34% from (*2*), and 43 to 50% from (*23*)], we estimate a 10 to 16% global genetic diversity loss, on average, across species (Fig. 3, A and B). Although these estimates may approximate central values across species in an ecosystem, we expect a substantial variation in the extent of loss across species, ranging theoretically from 0 to 100% (Fig. 3, fig. S26, and table S18), and expect that species extinction (100% area and genetic diversity lost) will substantially contribute to ecosystem-wide genetic diversity losses (figs. S25 and S27). One cause of this variation of losses across species is the heterogeneity in land-cover transformations across ecosystems; for example, more-pristine high-altitude systems have only lost 0.3% of their area, whereas highly managed temperate forests and woodlands have lost 67% (Fig. 3B and tables S14 and S15).

Another cause for the variability in genetic loss among species (even within the same ecosystem) may be their differential geographic ranges and abundances, life histories, tolerance to transformed habitats, or species-specific threats. We gathered data from species red-listed by the International Union for Conservation of Nature (IUCN) (*24*), which evaluates recent population or geographic range area reduction over ±10 years or ±3 generations to place assessed species in different threat categories using several criteria and thresholds (*24*). Assuming that the average $z^*_{MAR}$ can capture general patterns, we translate criterion A2-4c thresholds, which document geographic range loss, into genetic diversity loss [Fig. 3C and table S17; see (*4*)]. "Vulnerable" species, having lost at least 30% of their geographic distribution, may have experienced >9% of genetic diversity loss; "endangered" species, which have lost more than 50% of their geographic distribution, should have incurred >16% of genetic diversity loss; and "critically endangered" species, with more than 80% area reduction, likely suffered >33% of genetic diversity loss (Fig. 3C). This showcases that even species at no imminent risk of extinction (e.g., least concern, near threatened, vulnerable), such as most species for which population genomic data exists, may already be losing genetic diversity (Fig. 3A).

If future habitat losses are not prevented, genetic diversity will continue eroding. In support of this claim, habitat projections to 2070 of 19,365 threatened and nonthreatened mammal, bird, and amphibian species from the Map of Life (*25*) show an average 2% genetic diversity decline (Fig. 3D), and a total of 1843 species (3%) are projected to

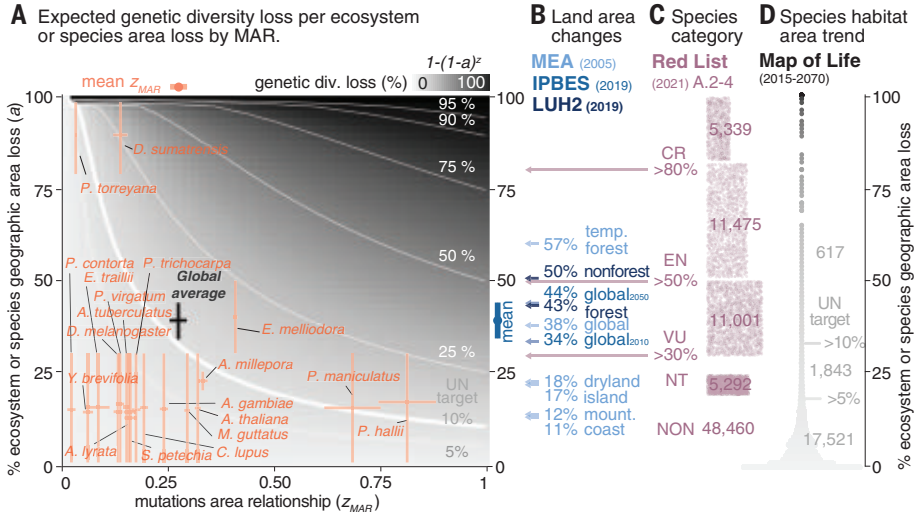



**Fig. 3. The parameter space of genetic diversity loss mapping before 21st century ecosystem transformations and species threat categories against possible values of the MAR.** (**A**) Possible values of two key parameters, the MAR scaling parameter ($z_{MAR}$) and percentage of area reduction (*a*) of a species geographic range (as a proxy of entire ecosystem transformation). The theoretical percentage of genetic diversity loss is represented as filled gray color, with isolines in white. Estimates of $z^*_{MAR}$ from Table 1 per species are in orange with their 95% CIs (see fig. S23 for the unscaled $z_{MAR}$). Past area losses for these species are unknown. As a proxy with much uncertainty, species are plotted based on their IUCN Red List status, which under criterion A2-4 can be used to determine the minimum range area decline (C). The global average is calculated with the average $z^*_{MAR}$ across species and the total percentage of Earth transformed as published by IPBES. (**B**) Percentages of transformed ecosystem areas from the Millennium Ecosystem Assessment (MEA) (table S14) are represented by light blue arrows, from IPBES (*2*) for 2010 and 2050 by dark blue arrows (table S15), and from the Land Use Harmonization 2 (LUH2) dataset (*23*) by deep blue arrows (table S16). (**C**) Lilac-colored arrows indicate the minimum value of geographic area loss under criteria A2-4c to be classified in each category of the IUCN Red List guidelines (*24*): near threatened (NT), vulnerable (VU), endangered (EN), critically endangered (CR), and nonthreatened (NON). The dotplots with pseudo-random numbers within the established thresholds represent plant, mammal, bird, and amphibian species categorized using criteria A2-4 (table S17). (**D**) Projected area loss from 2015 to 2070 by the Map of Life (mol.org) for mammal, bird, and amphibian species (*25*) (data are not available for plants). Color scale is the same as in (A) and indicates genetic diversity loss assuming the global average $z^*_{MAR}$.

lose at least 5% genetic diversity in the next decades.

Once lost, the recovery of genetic diversity through natural mutagenesis is extremely slow (*26*), especially for positive mutations that contribute to adaptation. Simulating a species undergoing only a 5 to 10% reduction in area, it would take at least ≈140 to 520 generations to recover its original genetic diversity (2100 to 7800 years for a fast-growing tree or medium–life span mammal of 15-year generation length); although for most simulations, recovery virtually never happened over millennia (fig. S11).

The ultimate challenge is to understand how genetic diversity loss relates to loss of adaptive capacity of a species. To this end, we leveraged the extensive knowledge of the effect of mutations in ecologically relevant traits in *A. thaliana* from genome-wide associations (GWAs). We again conducted spatial warm-edge extinction simulations, this time tracking metrics of adaptive capacity, including the total sum of effects estimated from GWA of remaining mutations ($\sum_i a^2{}_i$ for $i$ = 1...10,000 variants of putative $a_i$ effect), the additive genetic variance [$Va = \sum_i p_i(1 - p_i)a_i{}^2$, which accounts for each variant's population frequency $p_i$], and the loss of nonsynonymous mutations (Fig. 2C and figs. S19 to S21). Although determining the effect of mutations through GWA is technically challenging even in model species (*27*), and variants may even be either deleterious or advantageous depending on genomic backgrounds (*28*) or environments (*29*), our analyses suggest that putatively functional mutations may be more slowly lost as area is lost ($z < 0.3$; Fig. 2C) than neutral genetic diversity (Fig. 2A and table S9). Indeed, the additive variance $Va$ parameter, often equated to the rate of adaptation, appears rather stable (*30*) until just before the extinction event when it sharply collapses (fig. S21; for simulations that replicate this pattern see fig. S9). This is analogous to the famous "rivet popper" metaphor where ecosystem structure and function may suddenly collapse as species are inadvertently lost (*31*). Projections of the MAR using genome-wide variation thus may crucially serve as an early conservation tool in nonthreatened species (*32*), before species reach accelerating collapsing extinction dynamics—an acceleration that we expect to be even more dramatic owing to increased drift and accumulation of deleterious mutations of small critically endangered populations (*6*).

To achieve the recently proposed UN target to protect "at least 90% of genetic diversity within all species" (*11*), it will be necessary to aggressively protect as many populations as possible for each species. Here, we have discovered the existence of a MAR and provided a mathematical framework to forecast genetic diversity loss with shrinking geographic species ranges. The MAR contrasts with existing studies on the risk of losing entire species by focusing on quantifying the magnitude and dynamics of genetic diversity loss that is likely ongoing in most species. This framework demonstrates that even with conservative estimates, substantial area protection will be needed to meet the UN Sustainable Development Goals. For vulnerable or endangered species, we have likely already failed.

## REFERENCES AND NOTES


1. S. Díaz *et al*., *Science* **366**, eaax3100 (2019).
2. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES), *The Global Assessment Report on Biodiversity and Ecosystem Services*, E. S. Brondizio, J. Settele, S. Diaz, H. T. Ngo, Eds. (IPBES Secretariat, Bonn, 2019).
3. J. J. Wiens, *PLOS Biol.* **14**, e2001104 (2016).
4. See supplementary materials and methods.
5. C. Parmesan, *Annu. Rev. Ecol. Evol. Syst.* **37**, 637–669 (2006).
6. M. Lynch, J. Conery, R. Burger, *Am. Nat.* **146**, 489–518 (1995).
7. D. Spielman, B. W. Brook, R. Frankham, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15261–15264 (2004).
8. W. Steffen *et al*., *Science* **347**, 1259855 (2015).
9. L. Laikre *et al*., *Science* **367**, 1083–1085 (2020).
10. S. Díaz *et al*., *Science* **370**, 411–413 (2020).
11. Convention on Biological Diversity (CBD), "First draft of the post-2020 Global Biodiversity Framework" (CBD/WG2020/3/3, United Nations Environment Programme, 2021); https://www.cbd.int/doc/c/abb5/591f/2e46096d3f0330b08ce87a45/wg2020-03-03-en.pdf.
12. D. M. Leigh, A. P. Hendry, E. Vázquez-Domínguez, V. L. Friesen, *Evol. Appl.* **12**, 1505–1512 (2019).
13. K. L. Millette *et al*., *Ecol. Lett.* **23**, 55–67 (2020).
14. O. Arrhenius, *J. Ecol.* **9**, 95–99 (1921).
15. D. Storch, P. Keil, W. Jetz, *Nature* **488**, 78–81 (2012).
16. C. D. Thomas *et al*., *Nature* **427**, 145–148 (2004).
17. V. Buffalo, *eLife* **10**, e67509 (2021).
18. R. A. Fisher, *Proc. R. Soc. Edinb.* **50**, 204–219 (1931).
19. 1001 Genomes Consortium, *Cell* **166**, 481–491 (2016).
20. F. W. Preston, *Ecology* **43**, 185 (1962).
21. A. Hampe, R. J. Petit, *Ecol. Lett.* **8**, 461–467 (2005).
22. Millennium Ecosystem Assessment, "Ecosystems and human well-being: Biodiversity synthesis" (World Resources Institute, 2005); https://www.millenniumassessment.org/en/index.html.
23. G. C. Hurtt *et al*., *Geosci. Model Dev.* **13**, 5425–5464 (2020).
24. International Union for Conservation of Nature (IUCN), "Guidelines for using the IUCN Red List categories and criteria," version 15 (IUCN, 2022); https://www.iucnredlist.org/resources/redlistguidelines.
25. R. P. Powers, W. Jetz, *Nat. Clim. Chang.* **9**, 323–329 (2019).
26. D. L. Halligan, P. D. Keightley, *Annu. Rev. Ecol. Evol. Syst.* **40**, 151–172 (2009).
27. M. V. Rockman, *Evolution* **66**, 1–17 (2012).
28. M. J. Harms, J. W. Thornton, *Nat. Rev. Genet.* **14**, 559–571 (2013).
29. M. Exposito-Alonso, H. A. Burbano, O. Bossdorf, R. Nielsen, D. Weigel, 500 Genomes Field Experiment Team, *Nature* **573**, 126–129 (2019).
30. H. R. Taft, D. A. Roff, *Conserv. Genet.* **13**, 333–342 (2012).
31. P. Ehrlich, B. Walker, *Bioscience* **48**, 387 (1998).
32. M. Kardos *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104642118 (2021).
33. M. Exposito-Alonso, Scripts to build the mutations area relationship. Zenodo (2022); https://doi.org/10.5281/zenodo.6408624.


## ACKNOWLEDGMENTS


We are grateful to the openness of many researchers who make genomic data publicly available and this research possible: the 1001 *Arabidopsis* Genomes Consortium, the *Anopheles gambiae* 1000 Genomes Consortium, the 1000 (Human) Genome Consortium, the European Drosophila Evolution over Space and Time (DEST) Consortium, MacLachlan *et al*., Fuller *et al*., Ruegg *et al*., Kingsley *et al*., Schweizer *et al*., and Royer *et al*. In addition, we are grateful to all the authors who shared intermediate genome variant files: S. Mamidi, J. Lovell, D. Jacobson, M. Shah, J. Kreiner, K. Lucek, Y. Willi, J. D. Palacio Mejia, J. Borevitz, M. Supple, M. Vallejo-Marin, N. Dussex, L. Di Santo, and J. Hamilton. We thank W. Jetz for fruitful discussions and for sharing range data from Map of Life (mol.org). We thank J. Wiens, S. Rhee, D. Weigel, E. Armstrong, M. Kardos, D. Petrov, R. Colwell, R. Nielsen, A. Michalak, S. Hoban, J. Pritchard, T. Fukami, and members of the Moi and Mordecai labs for comments, discussion, or references. **Funding:** M.E.-A. and this research are supported by the Office of the Director of the National Institutes of Health's Early Investigator Award (1DP5OD029506-01); by the U.S. Department of Energy, Office of Biological and Environmental Research (DE-SC0021286); and by the Carnegie Institution for Science. J.P.S. is supported by an NIH training grant (5T32HG000044-23), and S.T.A by the National Institute of General Medical Sciences center of the NIH (T32GM007276). S.H. and C.L.W. are supported by Stanford's Center for Computational, Evolutionary, and Human Genomics. P.L.M.L. is supported by a Human Frontier Science Program Long-Term Fellowship (LT000330/2019-L). M.R. is supported by the NSF Postdoctoral Research Fellowships in Biology program (2109868). L.L. and L.G. are supported by NSF's Graduate Research Fellowship Program. Computational analyses were done on the high-performance computing clusters *Memex*, *Calc*, and *MoiNode* supported by the Carnegie Institution for Science. **Author contributions:** After the first author, authors are listed in alphabetical order. M.E.-A. conceived and led the project. M.E.-A., J.P.S., M.R., S.H., L.G., L.C., L.L., S.E.T.A., V.P., E.Z., P.L.M.L., C.C.K., T.R.B., and C.L.W. conducted research; all authors interpreted the results and wrote the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** The analyzed datasets are publicly available (see supplementary materials for data links). Code is available at Github (https://github.com/moiexpositoalonsolab/mar) and Zenodo (*33*). **License information:** Copyright © 2022 the authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original US government works. https://www.science.org/about/science-licenses-journal-article-reuse


## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn5642
Materials and Methods
Figs. S1 to S27
Tables S1 to S18
References (*34*–*90*)
MDAR Reproducibility Checklist

Submitted 6 December 2021; accepted 8 August 2022
10.1126/science.abn5642

DRYLAND FORESTATION

# Limited climate change mitigation potential through forestation of the vast dryland regions

Shani Rohatyn[1]*, Dan Yakir[2]*, Eyal Rotenberg[2], Yohay Carmel[1]

Forestation of the vast global drylands has been considered a promising climate change mitigation strategy. However, its actual climatic benefits are uncertain because the forests' reduced albedo can produce large warming effects. Using high-resolution spatial analysis of global drylands, we found 448 million hectares suitable for afforestation. This area's carbon sequestration potential until 2100 is 32.3 billion tons of carbon (Gt C), but 22.6 Gt C of that is required to balance albedo effects. The net carbon equivalent would offset ~1% of projected medium-emissions and business-as-usual scenarios over the same period. Focusing forestation only on areas with net cooling effects would use half the area and double the emissions offset. Although such smart forestation is clearly important, its limited climatic benefits reinforce the need to reduce emissions rapidly.

Leveraging the ability of forests to sequester carbon is considered a promising approach to mitigating global climate change (*1*–*3*). Forestation (including afforestation to create new forests and reforestation to restore depleted forests) is also known to cool the local climate by increasing evaporation and inducing increased cloud formation (*4*, *5*). A rich body of scientific research supports tree planting as an effective approach to mitigating global warming. Griscom *et al.* (*2*) calculate that reforestation of ~700 Mha in temperate and tropical zones would result in sequestration of almost three billion tons of carbon per year (Gt C $year^{-1}$). Bastin *et al.* (*3*) refer to tree restoration as "among the most effective strategies for climate change mitigation." They estimate that reforesting 1700 Mha could potentially sequester 205.7 Gt C (133.2 to 276.2 Gt C) over the lifetime of the forests (*6*).

Trees sequester atmospheric $CO_2$, and thus planting has a cooling effect by lowering its atmospheric concentration (*7*). Forestation also reduces the reflectance of shortwave radiation (albedo) more than most other forms of land coverage and thus increases net radiation and sensible heat flux, creating local and, potentially, global warming effects (*8*). These contrasting effects have long been recognized (*9*–*11*). However, this warming effect is largely confined to boreal regions. Recognition of this phenomenon is evident in recent publications supporting reforestation as a climate mitigation tool (*2*, *12*), wherein the albedo effect was avoided by excluding the boreal biome from the analysis to obtain maximal climatic benefits. However, there are recent indications that albedo warming effects are also substantial in temperate zones and hot drylands (*13*, *14*). In some dryland regions, the albedo warming effect of afforestation may strongly outweigh the cooling effect of carbon sequestration owing to the change from bright desert land to darker dense forest cover (*15*).

Drylands are defined as having an aridity index (or AI, the ratio between mean annual precipitation and mean annual potential evapotranspiration) of <0.65 (*16*). Drylands cover 40% of the global land area (*17*), with much of their area available for forestation actions. Drylands are also considered potential carbon sinks because of their soil properties and their long turnover time, which suggests that foresting drylands may result in carbon being transferred efficiently from the forest to the underlying dryland soils (*18*, *19*). An analysis of two global restoration opportunities indicated that 50% of global restoration potential is located in drylands (*3*, *20*). Afforestation and reforestation projects in drylands are ongoing around the world, and recently some large-scale projects were initiated or are planned to commence soon in places such as China, the Sahel, and Saudi Arabia (*21*, *22*). Together, these initiatives aim to convert >500 Mha of dryland from nonforested to forested land. However, given that in some regions the net effect of forestation is warming, these large projects may produce unintended climate warming outcomes. A fine-scale, spatially explicit analysis of the contrasting effects of forestation (*23*, *24*) is thus imperative to correctly assess the expected climate-related outcomes of such projects and their cost-effectiveness (*25*). Large-scale afforestation may eliminate rare species that depend on nonforested drylands and may thus have serious consequences for biodiversity (*26*–*29*). Such extinctions may be avoided by limiting afforestation to specific areas within a region, rather than covering the whole available area with forests (*30*, *31*). In any case, biodiversity conservation considerations impose additional constraints that further limit the amount of land available for afforestation.

Given the costs of large-scale forestation, as well as the possible consequences for biodiversity arising from changes in land coverage, it is of utmost importance to produce (i) precise site-specific estimates of the climatic benefits of dryland forestation and (ii) a robust global estimate of the maximum potential contribution of large-scale dryland forestation as a tool to mitigate climatic warming. Consequently, the overarching goal of this study is to conduct a high-resolution spatial analysis to identify drylands with afforestation potential and to evaluate the actual climatic benefits of undertaking global afforestation actions in those areas, including carbon sequestration and albedo effects.

We used suitability analysis based on land-cover and biological criteria to identify potential dryland for afforestation actions involving the conversion of low vegetation to dense forest cover. We examined the potential contribution of afforestation as a climate mitigation approach, including both carbon sequestration and albedo effects, using a combination of remote sensing tools and data-based estimations (for more information, see methods section in the supplementary materials). To widen our purview beyond afforestation, we also used two previously published forest restoration datasets (*3*, *20*) that applied different criteria than those of our study. Both studies allowed tree planting in areas already covered by woody vegetation (densification) and proposed tree planting in areas that were once covered by forests (reforestation). In contrast, our study focused on semiarid areas that were not previously forested (afforestation). We then simulated carbon sequestration and albedo effects for the restoration maps using the same method as for our afforestation map. Finally, we combined the three forestation maps to simulate the maximal climate change mitigation potential attainable from the forestation of global drylands.

The results provide a quantitative assessment of the published suggestions that climate change may be mitigated by foresting the reportedly large nonforested dryland areas. Our high-resolution spatial analysis of the global semiarid and dry subhumid land areas and associated afforestation suitability analysis identified 448 Mha of global drylands as potentially afforestable (Fig. 1A; ~6% of the global dryland area). The remaining dryland area (~94%) was excluded as lacking suitability for afforestation. The excluded areas were urban areas (<1%), water bodies and wetlands (2%), cropland (17%), areas above the tree line (3%), shrubland and forested areas with woody-vegetation coverage

[1]Faculty of Civil and Environmental Engineering, The Technion–Israel Institute of Technology, Haifa 3200003, Israel. [2]Earth and Planetary Sciences, Weizmann Institute of Science, Rehovot 7610001, Israel.
*Corresponding author. Email: shani.rohatyn@gmail.com (S.R.); dan.yakir@weizmann.ac.il (D.Y.)

above our 15% threshold (25%), and arid and hyper-arid land (AI < 0.2) incompatible with tree survival (47%).

We then simulated the effects of afforesting these 448 Mha over a period of 80 years (2020 to 2100; as a conservative forest life span in these regions). For this period, we estimated the net cumulative carbon sequestration potential (ΔSP) of afforestation as 32.3 Gt C. However, the estimated emissions equivalent of shortwave forcing (EESF) associated with the reduced albedo after forestation of previously unforested drylands greatly reduced this potential in climatic terms. Our analysis indicated that 22.6 Gt C should be sequestered over this period to compensate for the EESF arising from albedo effects (relying on productivity and albedo change from nearby forests relative to the state of current vegetation; see methods). Consequently, the net climatic change (calculated as the net equivalent carbon stock change, NESC = ΔSP − EESF) resulting in cooling was equivalent to the sequestration of only 9.7 Gt C until 2100 (Table 1).

The spatial distribution of the climatic effects of our potential dryland afforestation scheme is presented in Fig. 1. We found the effects of afforestation to follow a clear spatial pattern, with negative NESC (i.e., warming effects) at high latitudes and positive NESC (i.e., cooling effects) at lower latitudes. These patterns indicate that afforestation in countries such as South Africa and Australia would result in positive NESC values (Fig. 1, F and G), whereas afforestation in Kazakhstan and Mongolia is likely to result in large negative NESC values (Fig. 1, B and C). Intermediate results are indicated for afforestation in China and the US (Fig. 1, D and E).

We compared the climatic effects of afforestation (in terms of NESC) using data from previous studies to assess the range of the potential effects from afforestation and reforestation schemes. This comparison expands forestation actions from a narrow focus on afforestation to include both reforestation with diverse forest cover and densification of existing forests, based on the reforestation scenarios of Potapov *et al.* (*20*) and Bastin *et al.* (*3*). Considering the full 448 Mha afforestation area proposed in our study (6% of total drylands), the reforestation scenarios of Potapov *et al.* (*20*) and Bastin *et al.* (*3*) covered forestation opportunity areas that were three and four times larger, respectively, than our afforestation area (15 and 25% of total drylands, respectively; Table 1 and figs. S1 to S3). We simulated the estimated cooling and warming effects in the dryland areas proposed in each of those scenarios using the same protocol and over the same 80-year forest life-span period that we applied to our afforestation map (Table 1 and figs. S1 to S3). For comparison between the different scenarios, we calculated the climate change mitigation efficiency as the normalized rate of NESC per unit of forested area. Climate change mitigation efficiency was highest for the Potapov *et al.* (*20*) reforestation scenario [40.2 tons of carbon per hectare (t C $ha^{-1}$)] and lowest for that of Bastin *et al.* (*3*) (16.0 t C $ha^{-1}$), with our afforestation scenario showing an intermediate efficiency (21.6 t C $ha^{-1}$) (Table 1).

We also simulated applying a "smart forestation" approach to both scenarios over the 80-year period. The smart forestation analysis excluded locations where our simulations predicted net warming effects (Table 1, NESC < 0; i.e., the red-colored areas in Fig. 1). In our afforestation scheme, application of smart forestation approximately halved the potential afforestation area while nearly doubling the total NESC values and more than tripling the climate change mitigation efficiency as measured by average NESC rates per hectare (Table 1). A large increase in climatic change mitigation efficiency was also found for both the Bastin *et al.* (threefold increase) and Potapov *et al.* (nearly twofold increase) scenarios. Application of smart forestation increased total NESC by factors of 1.8, 1.3, and 1.9 for the present study, Potapov *et al.* (*20*), and Bastin *et al.* (*3*), respectively (Table 1). We used the results from all three smart forestation scenarios to obtain a first approximation of the upper limit to which forestation can potentially mitigate climate change by increasing net carbon sequestration. To that end, we combined the three available mitigation potential scenarios [current study, Potapov *et al.* (*20*), and Bastin *et al.* (*3*)], selecting the maximum per-pixel NESC value over the three scenarios. As expected, this produced the highest total NESC value over the 80-year simulation period (113.6 Gt C), although not the highest climate change mitigation efficiency (Table 1).

We then used the maximum mitigation potential scenario to estimate the maximum potential of forestation to mitigate climatic warming. An examination of forestation initiatives in northern China, the Sahel region of Africa, and the northern Middle East indicated that 25, 44, and 40%, respectively, of

**Table 1. Potential climatic effects of three dryland forestation scenarios and their combined application.** Values for the net equivalent carbon stock change (NESC) and its components, the net carbon sequestration potential (ΔSP) and the emissions equivalent of shortwave forcing (EESF), where NESC = ΔSP − EESF, are presented for an 80-year forest lifetime, summed (first three columns) and averaged (last three columns) for the entire forested area. For each scenario, results are presented for the total area of potential forestation ("Total") and solely for forestation in areas where it has a cooling effect ("NESC > 0"). The maximum mitigation potential selects the maximum per-pixel NESC over all three forestation scenarios. Carbon sequestration estimates for ΔSP are based on remote sensing and actual flux measurements in the relevant areas (see materials and methods in the supplementary materials).

| Forestation scenario | Sum over entire area (Gt C) | | | Average rates (t C $ha^{-1}$) | | |
|---|---|---|---|---|---|---|
| | ΔSP | EESF | NESC | ΔSP | EESF | NESC |
| Afforestation (current study) | | | | | | |
| Total (448 Mha) | 32.3 | 22.6 | 9.7 | 72.1 | 50.5 | 21.6 |
| NESC > 0 (251 Mha) | 27.8 | 10.1 | 17.7 | 110.7 | 40.1 | 70.6 |
| Reforestation [Popatov *et al.* (*20*)] | | | | | | |
| Total (1134 Mha) | 75.9 | 30.3 | 45.6 | 66.9 | 26.7 | 40.2 |
| NESC > 0 (836 Mha) | 76.5 | 19.1 | 57.4 | 91.5 | 22.9 | 68.6 |
| Reforestation [Bastin *et al.* (*3*)] | | | | | | |
| Total (1882 Mha) | 57.1 | 27.1 | 30.0 | 30.3 | 14.4 | 16.0 |
| NESC > 0 (1148 Mha) | 65.9 | 8.0 | 57.9 | 57.4 | 6.9 | 50.4 |
| Maximum mitigation potential | | | | | | |
| NESC maximum (1804 Mha) | 143.5 | 29.9 | 113.6 | 79.6 | 16.6 | 63.0 |

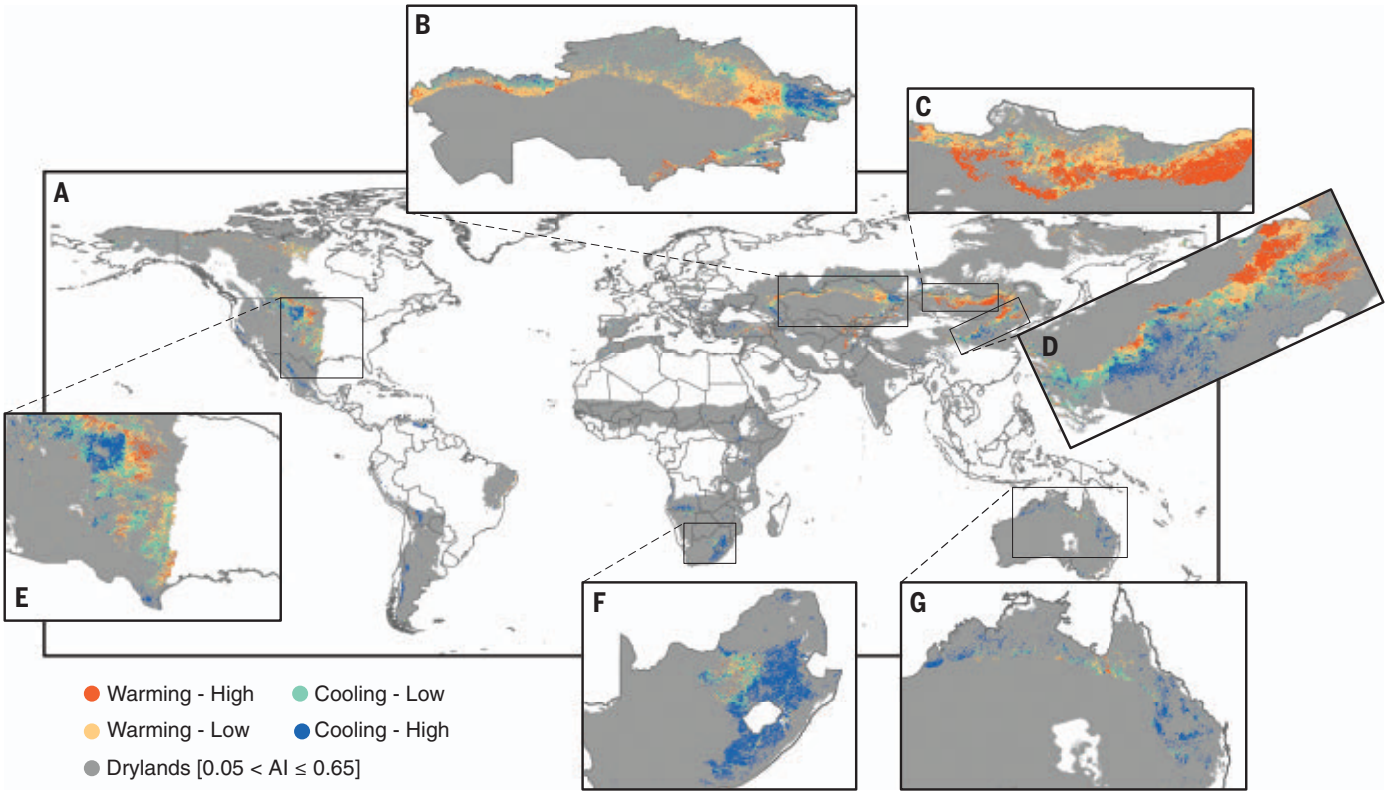



**Fig. 1. Net equivalent carbon stock change obtainable from the afforestation of suitable nonforested drylands.** (**A** to **G**) NESC outcomes calculated as the net difference between the carbon sequestration potential (ΔSP) and the emissions equivalent of shortwave forcing (EESF) arising from forestation-induced changes in albedo. Colors represent the NESC effect range, where NESC was calculated in units of tons of carbon per hectare over a forest lifetime of 80 years (2020–2100): high warming, NESC ≤ −50; low warming −50 < NESC ≤ 0 (represents a near-neutral climatic effect); low cooling, 0 < NESC ≤ 50; and high cooling, NESC > 50 (represents the largest potential climate cooling effect). The dark gray background indicates the full extent of global drylands [defined as semiarid and dry-subhumid lands within the aridity index (AI) range of 0.05 < AI ≤ 0.65]. (A) Global map. Zoom-ins of drylands in (B) Kazakhstan, (C) Mongolia, (D) northeastern China (Inner Mongolia), (E) USA, (F) South Africa, and (G) Australia. An interactive map of the results can be found here: https://tinyurl.com/mrt4ycha.

**Table 2. Potential contribution of dryland forestation to mitigating global $CO_2$ equivalents of the greenhouse gas emissions by 2100.** Estimated global $CO_2$ equivalents of the greenhouse gas emissions and the proportion of those emissions potentially mitigable by dryland forestation are shown for three possible emissions pathways. The four dryland forestation scenarios are as defined in Table 1 (limited to areas in which they have cooling effects; NESC > 0). The global greenhouse gas emissions predicted for each climate change response are based on the C-ROADS world climate simulator (*34*), accumulated over a forest lifetime of 80 years (2020–2100). The climate change responses considered were: business-as-usual (BAU); intended nationally determined contribution to reducing greenhouse gas emissions, as of September 2015 (INDC); and pledges to control greenhouse gas emissions to limit global warming to 2°C above preindustrial values (2C).

| Response to climate change | Global $CO_2$ equivalent emissions (Gt C) | Proportion (%) of global $CO_2$ equivalent emissions potentially mitigable by dryland forestation achieved through: | | | |
|---|---|---|---|---|---|
| | | Afforestation (current study) | Reforestation [Potapov *et al.* (*20*)] | Reforestation [Bastin *et al.* (*3*)] | Maximum mitigation potential |
| BAU | 2390 | 0.7 | 2 | 2 | 5 |
| INDC | 1592 | 1.1 | 4 | 4 | 7 |
| 2C | 608 | 2.9 | 9 | 10 | 19 |

the potential forestation lands will still have net climatic warming effects after 80 years of forestation efforts (fig. S4 and table S1). Clearly, forestation planners and decision-makers should consider climatic warming potential when selecting areas for forestation initiatives.

Finally, the contribution that forestation of drylands can potentially make to offsetting $CO_2$ equivalents of the greenhouse gas emissions by 2100 was estimated for all scenarios, as summarized in Table 2. We used the $CO_2$ equivalents emissions predicted by the World Climate Simulator (C-ROADS 2015) for a high-emissions scenario [business as usual (BAU)];

a medium-emissions scenario [intended nationally determined contribution (INDC) to reducing greenhouse gas emissions, as of September 2015]; and a low-emissions scenario to limit the global temperature increase to 2°C above preindustrial averages (2C). Surprisingly, given the vast area involved and its considerable carbon uptake potential, the global potential of large-scale dryland forestation to mitigate climate change is relatively poor, which reflects the large EESF in these regions. When considering BAU projections, even the maximum mitigation potential scenario could compensate for just 5% of cumulative emissions over the next 80 years (Table 2). Only under the highly optimistic 2C response (which assumes a reduction in emissions to only 25% of their BAU values) does the proportion of emissions potentially mitigatable by global dryland forestation (involving >20% of the drylands area) rise to nearly a fifth (19%).

It is important to note that forestation, if carefully planned and implemented, may provide local benefits, including soil erosion prevention, recreation, local evaporative cooling, and possibly increased precipitation (*4*, *5*, *32*). Moreover, although our study simulates the net climatic cooling benefits of forestation over an 80-year period, dryland forests may sustain a large carbon sink for a longer time, owing to their large potential soil carbon stock (*18*, *19*), thus providing long-term mitigation of climatic warming.

Previous estimates of the potential to mitigate climatic warming through large-scale forest restoration projects predicted a mitigation effect much larger than the results of this study. Using the restoration opportunities map of Potapov *et al.* (*20*), Griscom *et al.* (*2*) estimated that over an 80-year forest lifetime, the global reforestation of 700 Mha globally (~30% in drylands) could mitigate climatic warming to a maximum of 200 Gt C, which is nearly twice the value we obtained. This translates to a forestation sequestration potential per unit area of ~300 t C $ha^{-1}$ over that period. Similarly, Bastin *et al.* (*3*) estimated a potential carbon stock density of ~200 t C $ha^{-1}$ for the restoration of deserts, xeric shrublands, and Mediterranean forests. Both estimates are considerably higher than those of the present study. These differences likely arise from the additional consideration in the present study of two main factors: (i) the potential sequestration of current vegetation cover before reforestation; and (ii) the warming effect arising from the reduced albedo of forested drylands.

Our results demonstrate the importance of assessments of climatic warming mitigation plans including the warming effect arising from the reduced albedo of global dryland forestation. Accounting for albedo and avoiding foresting drylands where forestation would have a net warming effect (NESC < 0, Table 1) almost doubles the overall expected effect on climate. In contrast, forestation actions over negative-NESC areas would risk exacerbating, rather than ameliorating, global warming. Our analysis does not include additional effects that can further complicate a climate mitigation assessment of forestation, such as climate change–related effects on atmospheric temperature, clouds, or the extent of radiative cooling (from upwelling of long-wave radiation). Such effects influence both productivity and albedo and can move the aridity of some land areas to values outside the forestation suitability range considered here ($0.2 < AI \leq 0.65$) [e.g., (*33*)]. A detailed climate change impact analysis is well beyond the scope of this Report, but for a first approximation, we performed a cross-analysis by superimposing maps of the expected AI in 2100, considering a BAU scenario [+4°C (*33*)] over our forestation map. We found that ~3% of the potential forestation land (~10 Mha) will shift to a drier aridity value, below our minimum AI threshold of 0.2, by 2100. This analysis indicates that future climate change has only minor effects on our estimates of the land available for forestation and does not alter our conclusions.

Here we demonstrate, therefore, that it is critical that forestation opportunities be assessed with respect to their potential to mitigate climatic warming, and that doing so can greatly improve the cooling effect of forestation opportunities (both per-hectare and in terms of total land area used) of forestation opportunities. Forestation efforts, focusing on the limited areas with the potential for net climatic cooling, could benefit from high-resolution (1-km) maps, such as those developed in the present study. Overall, we estimate the total contribution toward offsetting $CO_2$ emissions obtainable from all dryland forestation actions to be limited, emphasizing the need to reduce emissions rapidly to meet climate targets.

## REFERENCES AND NOTES


1. S. L. Lewis, C. E. Wheeler, E. T. A. Mitchard, A. Koch, *Nature* **568**, 25–28 (2019).
2. B. W. Griscom *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11645–11650 (2017).
3. J.-F. Bastin *et al.*, *Science* **365**, 76–79 (2019).
4. J. I. Syktus, C. A. McAlpine, *Sci. Rep.* **6**, 29194 (2016).
5. G. Yosef *et al.*, *Sci. Rep.* **8**, 996 (2018).
6. *Science* **368**, eabc8905 (2020).
7. D. A. N. Ussiri, R. Lal, *Carbon Sequestration for Climate Change Mitigation and Adaptation* (Springer International Publishing, 2017).
8. G. B. Bonan, *Science* **320**, 1444–1449 (2008).
9. R. A. Betts, *Nature* **408**, 187–190 (2000).
10. G. Bala *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6550–6555 (2007).
11. R. G. Anderson *et al.*, *Front. Ecol. Environ.* **9**, 174–182 (2011).
12. S. C. Cook-Patton *et al.*, *Nature* **585**, 545–550 (2020).
13. Q. Zhang *et al.*, *Global Change Biol.* **26**, 3384–3401 (2020).
14. S. Rohatyn, E. Rotenberg, D. Yakir, Y. Carmel, *Environ. Res. Lett.* **16**, 104039 (2021).
15. E. Rotenberg, D. Yakir, *Science* **327**, 451–454 (2010).
16. R. J. Zomer, A. Trabucco, D. A. Bossio, L. V. Verchot, *Agric. Ecosyst. Environ.* **126**, 67–80 (2008).
17. U. Safriel, Z. Adeel, in *Ecosystems and Human Well-Being: Current State and Trends: Findings of the Condition and Trends Working Group*, R. Hassan, R. Scholes, N. Ash, Eds. (Island Press, 2005), pp. 623–662.
18. R. Qubaja, J. M. Grünzweig, E. Rotenberg, D. Yakir, *Global Change Biol.* **26**, 1626–1637 (2020).
19. R. Lal, *Environ. Manage.* **33**, 528–544 (2004).
20. P. Potapov, L. Laestadius, S. Minnemeyer, Global map of forest landscape restoration opportunities (World Resources Institute, 2011); www.wri.org/forest-restoration-atlas.
21. D. Goffner, H. Sinare, L. J. Gordon, *Reg. Environ. Change* **19**, 1417–1428 (2019).
22. L. Laestadius, K. Buckingham, S. Maginnis, C. Saint-Laurent, *Unasylva* **66**, 11–18 (2015).
23. R. A. Betts, *Nat. Geosci.* **4**, 504–505 (2011).
24. V. K. Arora, A. Montenegro, *Nat. Geosci.* **4**, 514–518 (2011).
25. S. Cao *et al.*, *Earth Sci. Rev.* **104**, 240–245 (2011).
26. K. D. Holl, P. H. S. Brancalion, *Science* **368**, 580–581 (2020).
27. J. W. Veldman *et al.*, *Bioscience* **65**, 1011–1018 (2015).
28. W. J. Bond, N. Stevens, G. F. Midgley, C. E. R. Lehmann, *Trends Ecol. Evol.* **34**, 963–965 (2019).
29. A. F. Souza, N. P. Ramos, M. A. Pizo, I. Hübel, L. O. Crossetti, *J. Nat. Conserv.* **21**, 56–62 (2013).
30. S. Gómez-González, R. Ochoa-Hueso, J. G. Pausas, *Science* **368**, 1439 (2020).
31. Y. Choi *et al.*, *Biol. Conserv.* **267**, 109474 (2022).
32. I. Stavi, *Forests* **10**, 737 (2019).
33. A. G. Koutroulis, *Sci. Total Environ.* **655**, 482–511 (2019).
34. T. Fiddaman, L. S. Siegel, E. Sawin, A. P. Jones, J. Sterman, C-ROADS Simulator Reference Guide (2019).


## ACKNOWLEDGMENTS


S.R. was supported by scholarships from the Technion–Israel Institute of Technology and the Weizmann Institute of Science. The authors thank A. Smadi for help with data collection and processing, R. Chemke for code optimization, and A. Singer-Lavi for important comments on an earlier draft. **Funding:** This project was supported in part by the Blumenstein Family Information Center at GWRI, the Technion–Israel Institute of Technology (ISF 1976/17 to D.Y.); a Minerva Foundation grant (to D.Y.); and the Yotam project and the Weizmann Institute Sustainability and Energy Research Initiative (SAERI P134470 to D.Y.). **Author contributions:** All authors jointly planned and designed the research. S.R. analyzed the data, and all authors interpreted the results. S.R. wrote the original draft, and all authors contributed substantially to revisions. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data that support the findings of this study were all derived from resources available in the public domain (see table S2 for descriptive information on the datasets). Data corresponding to the results reported here are available in the supplementary materials (data S1 to S3).



## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm9684
Materials and Methods
Figs. S1 to S5
Tables S1 and S2
References (*35*–*53*)
Data S1 to S3


Submitted 31 January 2022; accepted 23 August 2022
10.1126/science.abm9684

DECOMPOSITION

# Termite sensitivity to temperature affects global wood decay rates

Amy E. Zanne[1,2]*, Habacuc Flores-Moreno[3], Jeff R. Powell[4], William K. Cornwell[5], James W. Dalling[6,7], Amy T. Austin[8], Aimée T. Classen[9], Paul Eggleton[10], Kei-ichi Okada[11], Catherine L. Parr[12,13,14], E. Carol Adair[15], Stephen Adu-Bredu[16,17], Md Azharul Alam[18], Carolina Alvarez-Garzón[19], Deborah Apgaua[20], Roxana Aragón[21], Marcelo Ardon[22], Stefan K. Arndt[23], Louise A. Ashton[24], Nicholas A. Barber[25]†, Jacques Beauchêne[26], Matty P. Berg[27,28], Jason Beringer[29], Matthias M. Boer[4], José Antonio Bonet[30], Katherine Bunney[13], Tynan J. Burkhardt[31], Dulcinéia Carvalho[32], Dennis Castillo-Figueroa[33,34], Lucas A. Cernusak[35], Alexander W. Cheesman[35]‡, Tainá M. Cirne-Silva[32], Jamie R. Cleverly[35], Johannes H. C. Cornelissen[36], Timothy J. Curran[18], André M. D'Angioli[37], Caroline Dallstream[38], Nico Eisenhauer[39,40], Fidele Evouna Ondo[41], Alex Fajardo[42], Romina D. Fernandez[21], Astrid Ferrer[6], Marco A. L. Fontes[32], Mark L. Galatowitsch[43], Grizelle González[44], Felix Gottschall[45], Peter R. Grace[46], Elena Granda[47], Hannah M. Griffiths[12]§, Mariana Guerra Lara[8]¶, Motohiro Hasegawa[48], Mariet M. Hefting[49], Nina Hinko-Najera[50], Lindsay B. Hutley[51], Jennifer Jones[6]#, Anja Kahl[52], Mirko Karan[35,53], Joost A. Keuskamp[54,55], Tim Lardner[29], Michael Liddell[35], Craig Macfarlane[56], Cate Macinnis-Ng[31], Ravi F. Mariano[32], M. Soledad Méndez[8], Wayne S. Meyer[57], Akira S. Mori[58], Aloysio S. Moura[32], Matthew Northwood[51], Romà Ogaya[59], Rafael S. Oliveira[60], Alberto Orgiazzi[61], Juliana Pardo[62], Guille Peguero[63]**, Josep Penuelas[64,65], Luis I. Perez[8], Juan M. Posada[66], Cecilia M. Prada[6]††, Tomáš Přívětivý[67], Suzanne M. Prober[56,68], Jonathan Prunier[69]‡‡, Gabriel W. Quansah[70], Víctor Resco de Dios[71,72], Ronny Richter[52,73,74], Mark P. Robertson[13], Lucas F. Rocha[32], Megan A. Rúa[75], Carolina Sarmiento[7,76], Richard P. Silberstein[77,78], Mateus C. Silva[79]§§, Flávia Freire Siqueira[32], Matthew Glenn Stillwagon[22], Jacqui Stol[80], Melanie K. Taylor[81,82], François P. Teste[83]¶¶, David Y. P. Tng[20], David Tucker[46], Manfred Türke[39,40], Michael D. Ulyshen[81], Oscar J. Valverde-Barrantes[84], Eduardo van den Berg[79], Richard S. P. van Logtestijn[85], G. F. (Ciska) Veen[86], Jason G. Vogel[87], Timothy J. Wardlaw[88], Georg Wiehl[56], Christian Wirth[52,73], Michaela J. Woods[75], Paul-Camilo Zalamea[7,76]

Deadwood is a large global carbon store with its store size partially determined by biotic decay. Microbial wood decay rates are known to respond to changing temperature and precipitation. Termites are also important decomposers in the tropics but are less well studied. An understanding of their climate sensitivities is needed to estimate climate change effects on wood carbon pools. Using data from 133 sites spanning six continents, we found that termite wood discovery and consumption were highly sensitive to temperature (with decay increasing >6.8 times per 10°C increase in temperature)—even more so than microbes. Termite decay effects were greatest in tropical seasonal forests, tropical savannas, and subtropical deserts. With tropicalization (i.e., warming shifts to tropical climates), termite wood decay will likely increase as termites access more of Earth's surface.

Forested systems contain ~676 billion metric tons (Gt) of biomass (*1*), with a large fraction of their carbon immobilized for centuries in living wood and deadwood (*2*, *3*). Carbon storage depends partly on decay rates of deadwood pools by organisms, which vary across climatic gradients (*4*, *5*). Regional studies have suggested that wood decay by microbes approximately doubles with a 10°C temperature increase (decay effective $Q_{10}$ = ~2, where $Q_{10}$ is the increase in rate of a chemical reaction or biological process for each 10°C increase in temperature) (*2*, *6*) driven, in part, by enzyme kinetics. Further, microbial decay occurs through extracellular enzymes, whose delivery is dependent on moisture (*7*, *8*), which means that microbial wood decay should increase with humidity. Less is known about the climate sensitivities of important animal decayers, which also influence how climate change affects deadwood carbon stores.

Increasing evidence shows that termites are important decayers at local to regional scales (*7*, *9*, *10*). The abundance of wood-feeding termites across biomes is poorly understood (*11*), but decay by termites should be temperature sensitive. Termites increasingly contribute to wood decay in warm locations (*12*–*14*), with distributions set in part by ectothermic temperature tolerances (*15*). Termite wood decay depends on both discovery and consumption of wood by searching animals, followed by chemical decay through a cultivated set of microbial symbionts. Therefore, this symbiont chemical decay will also be shaped by temperature-dependent enzyme kinetics. In contrast to microbes, termites are likely less sensitive to moisture. Termites have a diversity of adaptations to conserve moisture, which presumably buffers their sensitivities to low precipitation (*16*–*18*). In other words, termite discovery and decay should continue with increasing aridity.

To test climate sensitivities of termite and microbial wood decay, we conducted a replicated experiment at 133 sites across extensive temperature and precipitation gradients representing most of the global bioregions (Fig. 1). At each site, researchers monitored decay of wood blocks for a common substrate, *Pinus radiata* [or, in a few cases, closely related *Pinus* species; (*19*)], for up to 48 months. All sites had harvests at ~12 months and most at ~24 months, with some sites including ~6-month, ~36-month, and/or ~48-month harvests. We allowed microbial access to all samples and manipulated termite access ("microbes" versus "microbes + termites" treatments); wood blocks were wrapped in fine mesh with or without larger holes to allow or exclude termites. At each site, researchers placed pairs of treatment blocks with the number of pairs equal to the number of harvests planned at each of 20 stations (a few sites placed fewer stations), which meant that each harvest from a site had 40 wood blocks [mean = 33.6 ± 14.2 (1 SD)] harvested at a given time point across both treatments. Stations were spaced at least 5 m apart (*19*) (table S11). A total of 8922 blocks were collected across all sites. Our focal species, *P. radiata*, was nonnative at all locations, which meant that no site decay agents evolved with it as a substrate.

Termite discovery (i.e., the estimated percentage of wood blocks with evidence of termites per year at a site) was greatest, but also highly variable, at low latitudes and elevations and where temperature and precipitation were high (Fig. 1, A and B; fig. S1; and table S1); low latitudes and elevations represent these warmer climates. High wood block discovery (>50%) occurred at temperatures above 21.33°C. In multivariate models, wood block discovery by termites rapidly increased with increasing temperatures (Fig. 2A and table S3), and temperature and precipitation significantly interacted (Fig. 1B, Fig. 2A, and table S3). Termite discovery was higher in warm tropical biomes in arid and semiarid sites (despite small sample sizes) compared with mesic and humid sites (at 25°C, discovery estimates at 250 mm were 1.4 times as high as those at 2000 mm and 1.9 times as high as those at 2700 mm), whereas in cool temperate biomes, the reverse patterns were observed (at 7°C, discovery estimates at 2700 mm were 4 times as high as those at 2000 mm and 150 times as high as those at 250 mm).

Microbial wood decay was fastest at low latitudes and elevations and where temperature and precipitation were high, although latitude and precipitation were weaker predictors than elevation and temperature (Fig. 1C, fig. S2, and table S2). Microbial temperature sensitivity was similar to that observed in regional studies [decay effective $Q_{10}$ of 1.73; 95% confidence interval (CI), 1.44 to 2.09] (*2*, *6*). In multivariate

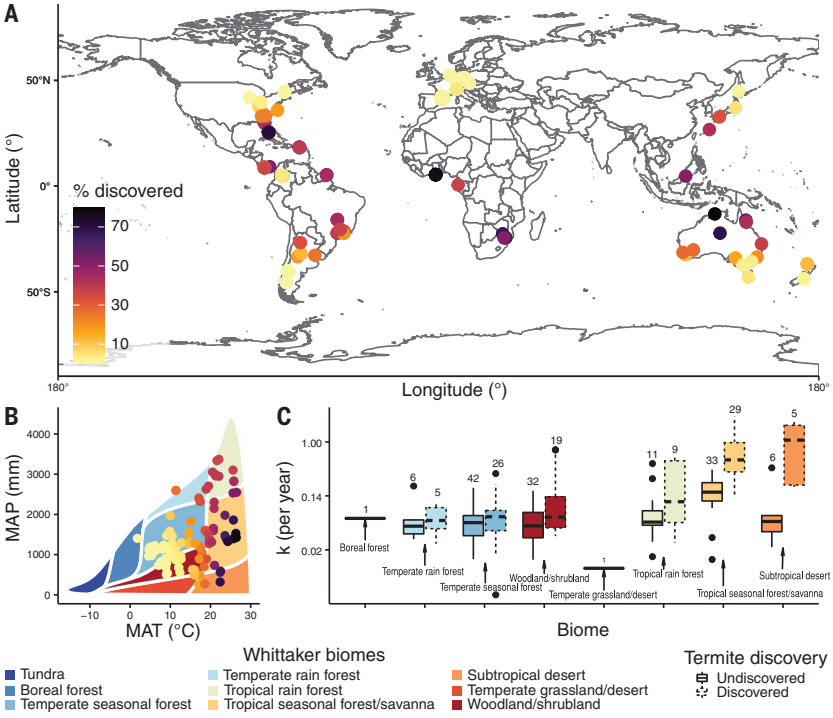



**Fig. 1. Geographic, biome, and climatic distribution of experimental sites.** (**A**) Dots denote the 133 study site locations. (**B**) Study site distribution across mean annual temperatures (MATs), mean annual precipitations (MAPs), and Whittaker biomes (*24*). In (A) and (B), the color of the dots represents termite discovery rate (i.e., estimated percentage of wood blocks with evidence of termites per year at a site). (**C**) Decay rate (k) estimates across Whittaker biomes [shown by arrows and colors matching the legend for (B)], with boxplots for each biome representing blocks discovered by termites (dashed boxplots on the right side of each pair) and blocks undiscovered by termites (solid boxplots on the left side of each pair) (examples of discovered blocks are given in fig. S3). The *y* axis is ln-transformed, but tick labels represent untransformed values for decay. For the boxplots, center line indicates the median, box limits indicate upper and lower quartiles, whiskers indicate the 1.5× interquartile range, and points indicate outliers. Numbers on top of the solid boxplots indicate the total number of sites per biome, and numbers on top of the dashed boxplots indicate the number of sites where termite discovery occurred.

[1]Department of Biology, University of Miami, Miami, FL, USA. [2]Department of Biological Sciences, George Washington University, Washington, DC, USA. [3]Terrestrial Ecosystem Research Network, University of Queensland, St Lucia, QLD, Australia. [4]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW, Australia. [5]School of Biological, Earth & Environmental Sciences, University of New South Wales, Sydney, NSW, Australia. [6]Department of Plant Biology, University of Illinois, Urbana-Champaign, Urbana, IL, USA. [7]Smithsonian Tropical Research Institute, Panama City, Panama. [8]Instituto de Investigaciones Fisiológicas y Ecológicas Vinculadas a la Agricultura (IFEVA), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Facultad de Agronomía, Universidad de Buenos Aires, Buenos Aires, Argentina. [9]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI, USA. [10]The Soil Biodiversity Group, Entomology Department, The Natural History Museum, London, UK. [11]Department of Northern Biosphere Agriculture, Tokyo University of Agriculture, Abashiri, Japan. [12]School of Environmental Sciences, University of Liverpool, Liverpool, UK. [13]Department of Zoology & Entomology, University of Pretoria, Pretoria, South Africa. [14]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, Wits, South Africa. [15]Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT, USA. [16]Biodiversity Conservation and Ecosystem Services Division, Forestry Research Institute of Ghana, Council for Scientific and Industrial Research, Kumasi Ashanti Region, Ghana. [17]Department of Natural Resources Management, CSIR College of Science and Technology, Kumasi Ashanti Region, Ghana. [18]Department of Pest-management and Conservation, Lincoln University, Lincoln, New Zealand. [19]Departamento de Biología/Ecología/Laboratorio de Ecología Funcional y Ecosistémica, Universidad del Rosario, Bogotá DC, Colombia. [20]Centre for Rainforest Studies, The School for Field Studies, Yungaburra, QLD, Australia. [21]Instituto de Ecología Regional, Universidad Nacional de Tucumán-CONICET, Tucumán, Argentina. [22]Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, USA. [23]School of Ecosystem and Forest Sciences, The University of Melbourne, Melbourne, VIC, Australia. [24]School of Biological Sciences, The University of Hong Kong, Hong Kong, Hong Kong SAR, China. [25]Department of Biological Sciences, Northern Illinois University, DeKalb, IL, USA. [26]UMR Ecologie des Forêts de Guyane (EcoFoG), AgroParisTech, CNRS, INRA, Universite des Antilles, Universite de Guyane, CIRAD, Kourou, France. [27]Department of Ecology and Evolution, Amsterdam Institute of Life and Environment, Vrije Universiteit, Amsterdam, Netherlands. [28]Community and Conservation Ecology, Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, Netherlands. [29]School of Agriculture and Environment, The University of Western Australia, Perth, WA, Australia. [30]Joint Research Unit, CTFC-AGROTECNIO-CERCA Center, Lleida, Spain. [31]School of Biological Sciences, University of Auckland, Auckland, New Zealand. [32]Departamento de Ciências Florestais, Universidade Federal de Lavras, Lavras, MG, Brazil. [33]Biology Department/Faculty of Natural Sciences, Universidad del Rosario, Bogotá, Colombia. [34]Biology Department/Faculty of Natural Sciences/Functional and Ecosystem Ecology Lab, Universidad del Rosario, Bogotá, Colombia. [35]College of Science and Engineering, James Cook University, Cairns, QLD, Australia. [36]Amsterdam Institute for Life and Environment (A-LIFE), Systems Ecology Section, Vrije Universiteit, Amsterdam, Netherlands. [37]Programa de pós-graduação em Ecologia, Departamento de Biologia Vegetal, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, SP, Brazil. [38]Biology Department, McGill University, Montréal, QC, Canada. [39]Experimental Interaction Ecology, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany. [40]Institute of Biology, Leipzig University, Leipzig, Germany. [41]National Agency for National Parks, Libreville, Gabon. [42]Instituto de Investigación Interdisciplinaria (I$^3$), Vicerrectoría Académica, Universidad de Talca, Talca, Chile. [43]Biology Program, Centre College, Danville, KY, USA. [44]International Institute of Tropical Forestry, USDA Forest Service, Río Piedras, PR, USA. [45]German Centre for Integrative Biodiversity Research, Leipzig, Germany. [46]School of Biology and Environmental Science, Queensland University of Technology, Brisbane, QLD, Australia. [47]Departamento de Ciencias de la Vida, Universidad de Alcalá, Alcalá de Henares, Spain. [48]Department of Environmental System Science/Faculty of Science and Engineering, Doshisha University, Kyotanabe, Japan. [49]Department of Biology/Faculty of Science/Ecology and Biodiversity, Utrecht University, Utrecht, Netherlands. [50]Faculty of Science/School of Ecosystem and Forest Sciences, The University of Melbourne, Creswick, VIC, Australia. [51]Research Institute for the Environment and Livelihoods, Charles Darwin University, Darwin, NT, Australia. [52]Systematic Botany and Functional Biodiversity, Leipzig University, Leipzig, Germany. [53]Ecosystem Processes, TERN (Australian Terrestrial Ecosystem Research Network), Cairns, QLD, Australia. [54]Biont Research, Utrecht, Netherlands. [55]Ecology and Biodiversity, Institute of Environmental Biology, Department of Biology, Science Faculty, Utrecht University, Utrecht, Netherlands. [56]Land and Water, CSIRO, Wembley, WA, Australia. [57]School of Biological Sciences, Terrestrial Ecosystem Research Network, University of Adelaide, Adelaide, SA, Australia. [58]Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan. [59]Global Ecology Unit, CREAF-CSIC, Barcelona, Spain. [60]Department of Plant Biology, Institute of Biology, University of Campinas, Campinas, SP, Brazil. [61]European Commission, Joint Research Centre, Ispra, Italy. [62]Department of Biology, Université de Montréal, Montréal, Quebec, Canadá. [63]Departament de Biologia Animal, Biologia Vegetal i Ecologia, Universitat Autònoma de Barcelona, Barcelona, Spain. [64]Global Ecology Unit, CSIC, Bellaterra Barcelona, Spain. [65]Global Ecology Unit, CREAF, Cerdanyola del Valles Barcelona, Spain. [66]Biology Department/Functional and Ecosystem Ecology Lab, Universidad del Rosario, Bogota DC, Colombia. [67]Department of Forest Ecology, Silva Tarouca Research Institute for Landscape and Ornamental Gardening, Brno, Czechia. [68]School of Biological Sciences, The University of Western Australia, Crawley, WA, Australia. [69]UMR ECOFOG/Laboratoire des Sciences du Bois, CNRS, Kourou GF, France. [70]Soil Analytical Services, Soil Testing Laboratory, CSIR-Soil Research Institute, Kumasi Ashanti Region, Ghana. [71]Department of Crop and Forest Sciences, University of Lleida, Lérida, Spain. [72]School of Life Sciences and Engineering, Southwest University of Science and Technology, Mianyang, China. [73]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena- Leipzig, Leipzig, Germany. [74]Geoinformatics and Remote Sensing, Leipzig University, Leipzig, Germany. [75]Department of Biological Sciences, Wright State University, Dayton, OH, USA. [76]Department of Integrative Biology, University of South Florida, Tampa, FL, USA. [77]School of Science, Edith Cowan University, Joondalup, WA, Australia. [78]Agriculture and Environment, The University of Western Australia, Nedlands, WA, Australia. [79]Departamento de Ecologia e Conservação, Universidade Federal de Lavras, Lavras, MG, Brazil. [80]Land and Water, CSIRO, Canberra, ACT, Australia. [81]Southern Research Station, USDA Forest Service, Athens, GA, USA. [82]Odum School of Ecology, University of Georgia, Athens, GA, USA. [83]Instituto de Matemática Aplicada de San Luis (IMASL), CONICET, Universidad Nacional de San Luis, San Luis, Argentina. [84]Department of Biological Sciences, International Center of Tropical Biodiversity, Institute of Environment, Florida International University, Miami, FL, USA. [85]Department of Ecological Science, Faculty of Science, Vrije Universiteit, Amsterdam, Netherlands. [86]Department of Terrestrial Ecology, NIOO-KNAW, Wageningen, Netherlands. [87]School of Forest, Fisheries, and Geomatics Sciences, University of Florida, Gainesville, FL, USA. [88]School of Natural Sciences, University of Tasmania, Hobart, TAS, Australia.

***Corresponding author. Email: aezanne@gmail.com** †Present address: Department of Biology, San Diego State University, San Diego, CA, USA. ‡Present address: College of Life and Environmental Sciences, University of Exeter, Exeter, UK. §Present address: School of Biological Sciences, University of Bristol, Bristol, UK. ¶Present address: IMASL-CONICET, Universidad Nacional de San Luis, San Luis, Argentina. #Present address: Kellogg Biological Station, Michigan State University, Hickory Corners, MI, USA. **Present address: Smithsonian Tropical Research Institute, Balboa, Panama. ††Present address: Organismic and Evolutionary Biology, Harvard, Cambridge, MA, USA. ‡‡Present address: Observatoire Midi-Pyrénées/Géosciences Environnement Toulouse, IRD, Toulouse OCC, France. §§Present address: Department of Geography, University of Exeter, Exeter, UK. ¶¶Present address: Swift Current Research and Development Centre, Agriculture and Agri-Food Canada, Swift Current SK, Canada.

models, precipitation was not a significant predictor of microbial decay (Fig. 2B and table S4). When termites discovered wood, decay rates were higher at low elevations and where temperature was high (Fig. 1C, fig. S2, and table S2). Further, decay rates in termite-discovered wood were more sensitive to changes in temperature (decay effective $Q_{10}$ of 6.85; 95% CI, 4.73 to 9.92) compared with decay rates in undiscovered wood, where microbes dominated decay. In multivariate models, precipitation was not a significant predictor of decay for termite-discovered wood (Fig. 2C and table S5).

The termite-discovered wood decay effective $Q_{10}$ is much steeper than any previously recorded for microbes (*2*, *6*), which suggests that a different mechanism determines termite versus microbial wood decay. The observed high consumption rate by termites at warm sites may be related to termite assemblage composition, large population numbers, high activity, or some combination of these mechanisms. Consequently, subtropical, tropical, or global models using a single microbial-derived decay effective $Q_{10}$ are likely to (i) underpredict wood decay, (ii) overpredict terrestrial carbon storage (all else being equal—e.g., inputs into deadwood pools), and (iii) underpredict temperature sensitivity of decay. Use of termite-corrected decay effective $Q_{10}$ values, which may vary on the basis of termite assemblage composition, location, and/or wood substrate, should improve the accuracy of modeled wood decay under current and future climate predictions. Such model modifications can capitalize on empirical measures in the literature, such as ours for termites and (*20*) for insects more broadly. Our results suggest that precipitation variation influences the discovery, but not the decay, phase of termite wood decay. However, strong temperature and precipitation interaction influences on discovery mean that termites increased overall decay most in subtropical deserts and tropical seasonal forests and savannas (Fig. 1C). Further, even though microbial abundance is sensitive to precipitation (*4*, *5*), temperature was a stronger driver than precipitation for microbial-driven decay, perhaps mediated through effects on enzyme kinetics (*21*). Differences in decay sensitivity to precipitation were small, with only microbial-mediated wood decay weakly sensitive to precipitation; microbial decay largely occurs through the release of moisture-sensitive extracellular enzymes (*7*, *8*), whereas termites can conserve moisture, buffering aridity effects (*16*–*18*). Although low termite discovery in warm humid locations remains surprising, competitive interactions among decayers (*11*, *13*), biome-specific adaptations to moisture, variation in resource availability affecting foraging behavior, etc., may reduce discovery.

Given the high sensitivities of both termite wood discovery and decay to temperature, termites will likely expand their range in a warming world, with important consequences for carbon cycling. Using data-driven estimates of temperature and precipitation effects on termite discovery (table S3), we estimated discovery rates across the globe, restricting predictions to the range in mean annual precipitation covered by our sites ±10%. Termites today have the potential to discover large amounts of deadwood (>50%) at sites across 30.2% of the land surface (assuming our estimated discovery rates apply across wood and termite species; Fig. 3). To bracket potential climate change effects on discovery, we used our estimated climate relationships with all available midcentury Coupled Model Intercomparison Project Phase 6 (CMIP6) climate models for shared socioeconomic pathway (SSP) scenarios 1-2.6 and 5-8.5 (*22*). All scenarios predicted an expansion of termite discovery in tropical and subtropical regions, with the degree of expansion depending strongly

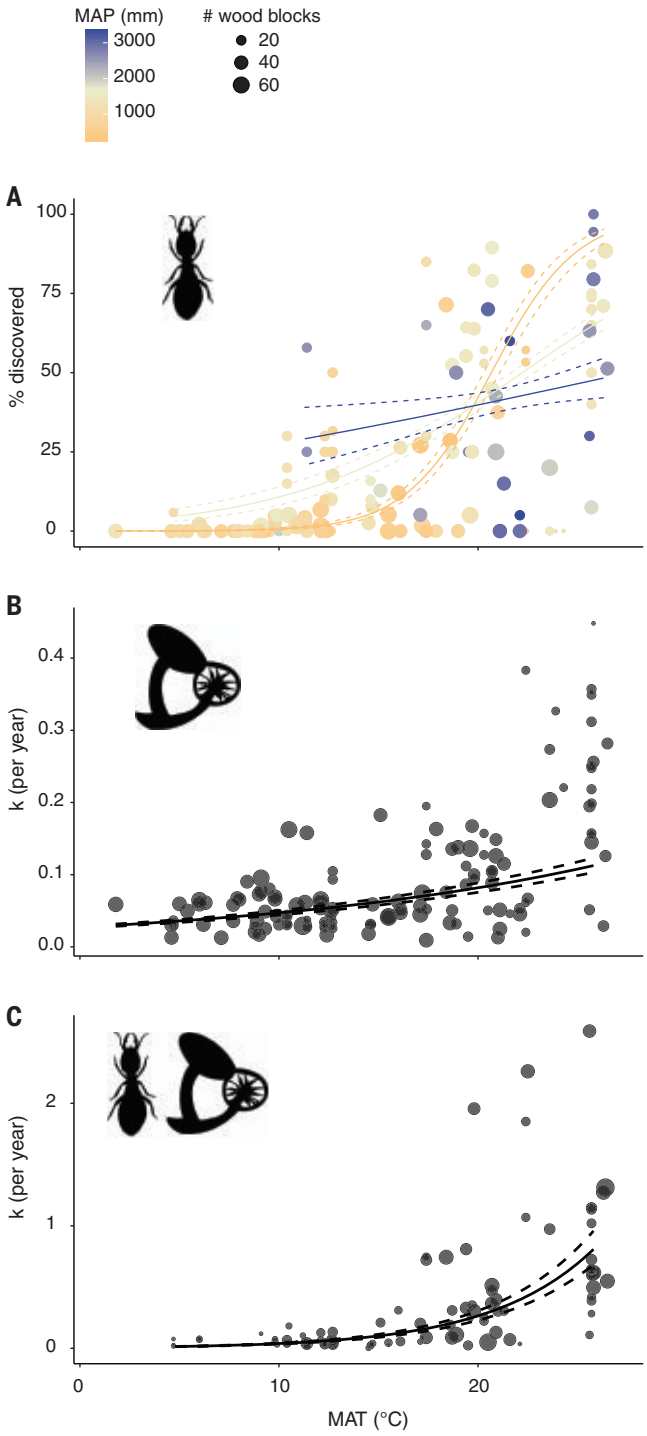



**Fig. 2. Discovery and decay of wood based on significant climatic predictors.** See tables S3 to S5 for full models. (**A**) Termite discovery rate; the estimated percentage of wood blocks in the microbes + termites treatment across all sites with evidence of termites per year, across MAT and MAP. (**B**) Decay rates of termite-undiscovered wood across MAT. (**C**) Decay rates of termite-discovered wood across MAT (note, MAP was not a significant predictor of termite-undiscovered or -discovered wood decay). Dot size represents number of wood blocks. Symbols in the upper left corner of each plot denote the role of wood-feeding termites and/or wood-dwelling microbes. Solid lines represent logistic (A) or linear [(B) and (C)] regression predictions and [for (A)] those at 250-mm MAP (orange; representative of mean desert and savanna biomes), 2000-mm MAP (cream; representative of mean temperate biomes), and 2700-mm MAP (blue; representative of mean tropical and temperate humid biomes). Dashed lines represent 95% CIs around predictions. The *y* axes for (B) and (C) are ln-transformed, but tick labels represent untransformed values for decay.

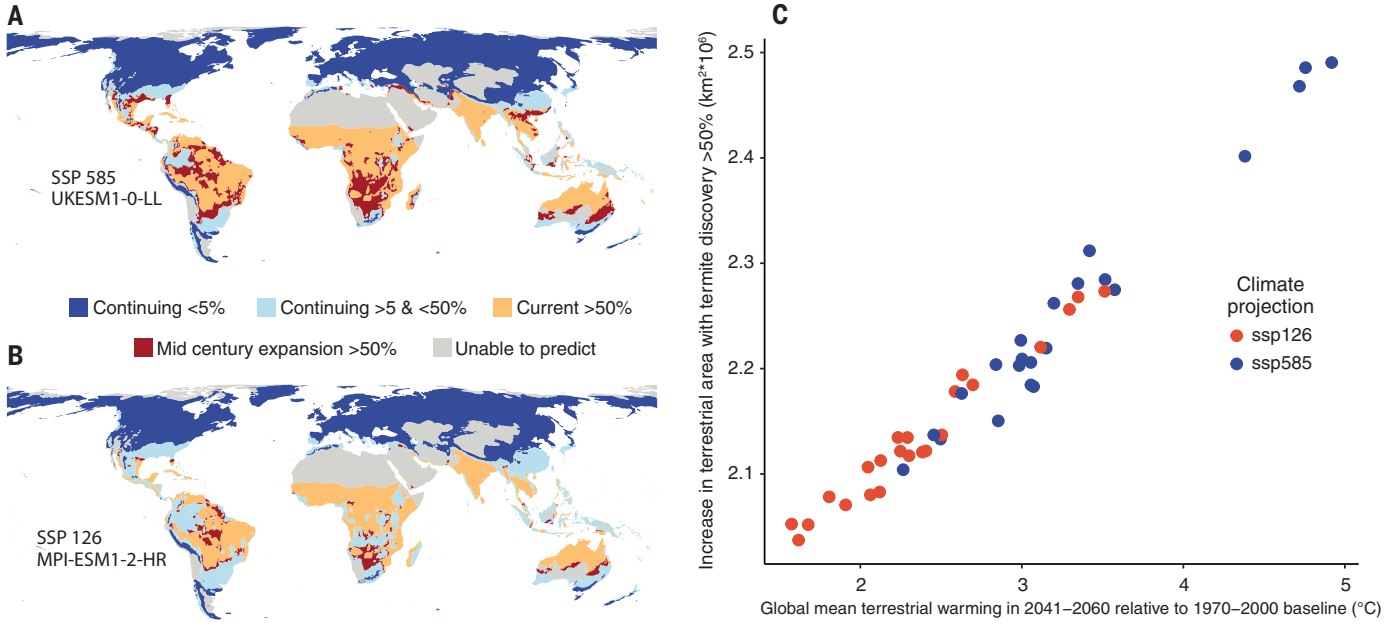



**Fig. 3. Predicted termite discovery by midcentury under different climate projections.** Global maps showing minimum and maximum termite expansion scenarios based on the model in table S3 and CMIP6 forecasts for 2041 to 2060. (**A** and **B**) Stronger climate change scenarios (SSP 5-8.5 UKESM1-0-LL) had the largest expansion in discovery rates (A), and weaker climate change scenarios (SSP 1-2.6 MPI-ESM1-2-HR) had the smallest (B). Termite discovery categories were rare (<5%, blue), continuing low (>5% and <50%, light blue), current high (>50%, orange), midcentury expansion to high (>50%, red), and unable to predict (gray), restricting predictions to the range in MAP covered by our sites (±10%). We did not model the transitions from rare (<5%, blue) to continuing low (>5% and <50%, light blue) discovery. (**C**) Forecast increases in terrestrial area (in square kilometers) with discovery >50% by midcentury versus forecast mean terrestrial warming relative to a historical baseline. Each point denotes a forecast based on one individual CMIP6 SSP 5-8.5 (blue) or SSP 1-2.6 (red) climate model. The *x* axis of (C) is the mean forecast of 2041 to 2060 warming above the 1970 to 2000 baseline for terrestrial areas only.

on the extent of global terrestrial warming (Fig. 3). Warming shifts to more tropical climates are occurring in many ecosystems (*23*), and temperature sensitivities demonstrated in this study suggest that termite contributions to wood decay will expand both within and beyond the tropics with such tropicalization. Our estimates may even underpredict termite effects in areas where fungus-growing termites occur (i.e., Africa and Asia) (*12*, *16*), which merits future research. The impact of termites on wood decay is both large and expected to increase (Fig. 3), and it also has a different functional form than microbial decay, with a clear two-step process—discovery and decay.

## REFERENCES AND NOTES


1. G. Kindermann, I. McCallum, S. Fritz, M. Obersteiner, *Silva Fenn.* **42**, 387–396 (2008).
2. M. E. Harmon *et al.*, *Carbon Balance Manag.* **15**, 1 (2020).
3. J. Mackensen, J. Bauhus, *Soil Biol. Biochem.* **35**, 177–186 (2003).
4. A. T. Austin, P. M. Vitousek, *J. Ecol.* **88**, 129–138 (2000).
5. B. N. Marais, C. Brischke, H. Militz, J. H. Peters, L. Reinhardt, *Forests* **11**, 1284 (2020).
6. N. Fierer, B. P. Colman, J. P. Schimel, R. B. Jackson, *Global Biogeochem. Cycles* **20**, GB3026 (2006).
7. W. K. Cornwell *et al.*, *Glob. Change Biol.* **15**, 2431–2449 (2009).
8. A. D. A'Bear, T. H. Jones, E. Kandeler, L. Boddy, *Soil Biol. Biochem.* **70**, 151–158 (2014).
9. M. D. Ulyshen, *Biol. Rev.* **91**, 70–85 (2016).
10. H. M. Griffiths, L. A. Ashton, C. L. Parr, P. Eggleton, *New Phytol.* **231**, 2142–2149 (2021).
11. R. A. Clement *et al.*, *Front. Ecol. Evol.* **9**, 657444 (2021).
12. H. M. Griffiths, L. A. Ashton, T. A. Evans, C. L. Parr, P. Eggleton, *Curr. Biol.* **29**, R118–R119 (2019).
13. M. A. Bradford *et al.*, *Ecology* **102**, e03484 (2021).
14. C. Guo, B. Tuo, H. Ci, E.-R. Yan, J. H. C. Cornelissen, *J. Ecol.* **109**, 1578–1590 (2021).
15. O. F. Palin *et al.*, *Biotropica* **43**, 100–107 (2011).
16. J. A. Bonachela *et al.*, *Science* **347**, 651–655 (2015).
17. L. A. Ashton *et al.*, *Science* **363**, 174–177 (2019).
18. A. W. Cheesman, L. A. Cernusak, A. E. Zanne, *Austral Ecol.* **43**, 257–267 (2018).
19. Materials and methods are available as supplementary materials.
20. S. Seibold *et al.*, *Nature* **597**, 77–81 (2021).
21. G. Wang, W. M. Post, M. A. Mayes, *Ecol. Appl.* **23**, 255–272 (2013).
22. V. Eyring *et al.*, *Geosci. Model Dev.* **9**, 1937–1958 (2016).
23. A. Vergés *et al.*, *Proc. Biol. Sci.* **281**, 20140846 (2014).
24. R. E. Ricklefs, *The Economy of Nature* (Macmillan, 2008).
25. A. Zanne, H. Flores-Moreno, J. Powell, W. Cornwell, J. W. Dalling, Global survey of termite and microbial contributions to wood decay, dataset, figshare (2022); https://doi.org/10.6084/m9.figshare.19920416.v1.
26. H. Flores-Moreno, J. Powell, A. Zanne, W. Cornwell, Zanne-Lab/global_termite_microbe_wd: Global survey of termite and microbial contributions to wood decay, version v1.0.2, Zenodo (2022); https://doi.org/10.5281/zenodo.7064570.


## ACKNOWLEDGMENTS


For materials, field, and laboratory assistance, we thank the Australian Landscape Trust Team, J. Martínez de Aragón, E. Roy, H. Wilson, K. Brice, B. Coleman, B. Delroy, C. Deveautour, S. Donn, G. Assis Neves, R. Marques Barreiros, A. Lochner, Pousada Vale do Céu, H. Robles, M. M. Rivera, and G. Kirker. We also thank the World Climate Research Programme for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies that support CMIP6 and ESGF. S. Allison, T. Crowther, and three anonymous reviewers provided helpful comments on the manuscript, and A. Yatsko helped with figure and manuscript formatting. **Funding:** This study received support from the following sources: US National Science Foundation (NSF) DEB-1655759 (A.E.Z.); US NSF DEB-2149151 (A.E.Z.); US NSF DEB-1713502 (M.A.); US NSF DEB-1713435 (M.A.); US NSF DEB-1647502 (N.A.B.); US NSF DEB-1546686 (G.G.); US NSF DEB-1831952 (G.G.); George Washington University (A.E.Z.); USDA Forest Service (G.G.); Centre College Faculty Development Funds (M.L.G.); Australia Terrestrial Ecosystem Research Network National Collaborative Research Infrastructure Strategy (P.R.G., M.K., M.L., M.M.B., R.P.S., J.S., L.B.H., M.N., S.M.P., T.J.W., and S.K.A.); Royal Society-FCDO Africa Capacity Building Initiative (C.L.P., G.W.Q., S.A.-B., K.B., F.E.O., and M.P.R.); New Phytologist Foundation (A.T.A.); Fondecyt grant 1160329 (C.D.); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES) (E.v.d.B., A.S.Mou., R.F.M., F.F.S., T.M.C.-S., R.S.O., and A.M.D.); Department of Ecology and Conservation of the Federal University of Lavras (T.M.C.-S.); CNPq (E.v.d.B. and R.S.O.); FAPEMIG (E.v.d.B.); Australian Academy of Science 2017 Thomas Davies Research Grant (J.R.P.); Australian Research Council DP160103765 (W.K.C., J.R.P., and A.E.Z.); UK National Environment Research Council NE/L000016/1 (L.A.A.); Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil NERC - FAPESP 19/07773-1 (R.S.O. and A.M.D.); Environment Research and Technology Development Fund ERTDF, JPMEERF15S11420 of the Environmental Restoration and Conservation Agency of Japan (A.S.Mor. and K.O.); COLCIENCIAS no. FP44842-046-2017 (J.M.P.); Spanish government PID2019-110521GB-I00 (J.Pe., G.P., and R.O.); Catalan government grant SGR 2017-1005 (J.Pe., G.P., and R.O.); Fundación Ramón Areces ELEMENTAL-CLIMATE (J.Pe., G.P., and R.O.); National Agency for the Promotion of Research, Technological Development and Innovation, Scientific and Technological Research Project 2018-01561 PICT 2018-01561 (F.P.T.); ANID PIA/BASAL FB210006 (A.Fa.); Millennium Science Initiative Program NCN2021-050 (A.Fa.); iDiv German Research Foundation DFG–FZT 118, 202548816 (N.E.); and European Research Council Horizon 2020 research and innovation program no. 677232 (N.E.). **Author contributions:** Conceptualization: A.E.Z. Methodology: A.E.Z., A.W.C., L.A.C., M.K., and J.W.D. Investigation: A.E.Z., J.R.P., J.W.D., A.T.A., A.T.C., P.E., K.O., C.L.P., E.C.A.,

S.A.-B., M.A., C.A.-G., D.A., R.A., M.A., S.K.A., L.A.A., N.A.B., J.Bea., M.P.B., J.Ber., M.M.B., J.A.B., K.B., T.J.B., D.C., D.C.-F., L.A.C., A.W.C., T.M.C.-S., J.R.C., J.H.C.C., T.J.C., A.M.D., C.D., N.E., F.E.O., A.Fa., R.D.F., A.Fe., M.A.L.F., M.L.G., G.G., F.G., P.R.G., E.G., H.M.G., M.G.L., M.H., M.M.H., N.H.-N., L.B.H., J.J., A.K., M.K., J.A.K., T.L., M.L., C.M., C.M.-N., R.F.M., M.S.M., W.S.M., A.S.Mor., A.S.Mou., M.N., R.O., R.S.O., A.O., J.Pa., G.P., J.Pe., L.I.P., J.M.P., C.M.P., T.P., S.M.P., J.Pr., G.W.Q., V.R.d.D., R.R., M.P.R., L.F.R., M.A.R., C.S., R.P.S., M.C.S., F.F.S., M.G.S., J.S., M.K.T., F.P.T., D.Y.P.T., D.T., M.T., M.D.U., O.J.V.-B., E.v.d.B., R.S.P.v.L., G.F.V., J.G.V., T.J.W., G.W., C.W., M.J.W., and P.-C.Z. Visualization: H.F.-M., J.R.P., and W.K.C. Funding acquisition: A.E.Z., J.R.P., J.W.D., A.T.A., A.T.C., P.E., K.O., C.L.P., E.C.A., S.A.-B., M.A., C.A.-G., D.A., R.A., M.A., S.K.A., L.A.A., N.A.B., J.Bea., M.P.B., J.Ber., M.M.B., J.A.B., K.B., T.J.B., D.C., D.C.-F., L.A.C., A.W.C., T.M.C.-S., J.R.C., J.H.C.C., T.J.C., A.M.D., C.D., N.E., F.E.O., A.Fa., R.D.F., A.Fe., M.A.L.F., M.L.G., G.G., F.G., P.R.G., E.G., H.M.G., M.G.L., M.H., M.M.H., N.H.-N., L.B.H., J.J., A.K., M.K., J.A.K., T.L., M.L., C.M., C.M.-N., R.F.M., M.S.M., W.S.M., A.S.Mor., A.S.Mou., M.N., R.O., R.S.O., A.O., J.Pa., G.P., J.Pe., L.I.P., J.M.P., C.M.P., T.P., S.M.P., J.Pr., G.W.Q., V.R.d.D., R.R., M.P.R., L.F.R., M.A.R., C.S., R.P.S., M.C.S., F.F.S., M.G.S., J.S., M.K.T., F.P.T., D.Y.P.T., D.T., M.T., M.D.U., O.J.V.-B., E.v.d.B., R.S.P.v.L., G.F.V., J.G.V., T.J.W., G.W., C.W., M.J.W., and P.-C.Z. **Competing interests:** L.F.R. is also affiliated with Eldorado Brasil Celulose, Três Lagoas, Brazil. P.R.G. is also affiliated with Michigan State University, Hickory Corners, Michigan, United States. The authors declare no other competing interests. **Data and materials availability:** Data are available at figshare (*25*), and code are available at Zenodo (*26*). 

### SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abo3856
Materials and Methods
Figs. S1 to S4
Tables S1 to S12
References (*27*–*59*)
MDAR Reproducibility Checklist

Submitted 13 February 2022; accepted 15 August 2022
10.1126/science.abo3856

ATMOSPHERE

# Water vapor injection into the stratosphere by Hunga Tonga-Hunga Ha'apai

**Holger Vömel[1]*, Stephanie Evan[2], Matt Tully[3]**

**Large volcanic eruptions, although rare events, can influence the chemistry and the dynamics of the stratosphere for several years after the eruption. Here we show that the eruption of the submarine volcano Hunga Tonga-Hunga Ha'apai on 15 January 2022 injected at least 50 teragrams of water vapor directly into the stratosphere. This event raised the amount of water vapor in the developing stratospheric plume by several orders of magnitude and possibly increased the amount of global stratospheric water vapor by more than 5%. This extraordinary eruption may have initiated an atmospheric response different from that of previous well-studied large volcanic eruptions.**

Water vapor is the most abundant radiatively active trace gas in our atmosphere, and despite its low concentration in the stratosphere, changes in stratospheric water vapor can influence our climate (*1*, *2*). Volcanic eruptions eject large amounts of gases, notably water vapor, nitrogen, and carbon dioxide. However, owing to the large concentrations of these gases in the atmosphere, their emission from volcanic eruptions is not believed to have a substantial climate effect (*3*). In addition to ash, sulfur-containing gases are generally believed to be the most important gases injected into the stratosphere by volcanic activity (*3*, *4*). A complex interaction of the freshly formed sulfuric acid aerosols, radiation, and chemistry can lead to a global decrease in the surface temperature and accelerated catalytic destruction of stratospheric ozone (*3*, *5*). Direct volcanic injection of water vapor into the stratosphere could possibly moderate the climate impacts of volcanic aerosols (*6*), but owing to the low frequency of such events, volcanic emissions are not considered a major source for stratospheric water vapor. Since the beginning of stratospheric water vapor observations in the middle of the past century, only a few volcanic eruptions were large enough to have added detectable amounts of water vapor into the stratosphere, and only minor injections have been reported (*7*–*9*). After the eruption of Pinatubo, one of the largest of the past century, an increase in stratospheric water vapor was not reported. Changes in stratospheric water vapor after the Pinatubo eruption are believed to be mostly related to changes in the tropopause temperature (*10*, *11*).

The eruption of the submarine volcano Hunga Tonga-Hunga Ha'apai on 15 January 2022 injected vast amounts of material, including substantial amounts of water, into the stratosphere. The eruption started on 13 January and peaked in a cataclysmic hydromagmatic explosion on 15 January 2022, which generated a volcanic eruption column reaching 58 km (*12*) and triggered a tsunami that devastated the nearby island nation of Tonga. Other large eruptions of the past few centuries, including Tambora in 1815, El Chichon in 1982, and Pinatubo in 1991, had substantial global atmospheric impacts but originated from land-locked volcanoes. Submarine eruptions can draw large parts of their explosive energy from the interaction of water and hot magma (*13*, *14*). This process allows for very large amounts of water and steam to be entrained in the eruption column. Most of these hydromagmatic eruptions have been small, leading to the formation of new islands, e.g., Surtsey in 1963. The Hunga Tonga-Hunga Ha'apai eruption may well have been the largest documented hydromagmatic eruption with an eruption column reaching well into the stratosphere.

Data from the operational upper air network using the Vaisala RS41 radiosonde provide in situ observations of large amounts of water vapor injected into the stratosphere by this eruption and allow a description of the early development of the volcanic plume as it dispersed. These instruments can detect stratospheric water vapor of this magnitude and provide high–vertical resolution observations from a large observing network (see data and methods in the supplementary materials).

The eruption of Hunga Tonga-Hunga Ha'apai (20.54°S, 175.38°W) was first detected by geostationary satellites on 15 January at 04:10 UTC. Ten hours after the eruption started, the volcanic plume reached the longitude of Fiji (17.77°S, 177.45°E) but largely stayed to the south.

A radiosounding at Nadi, Fiji (Fig. 1), was first to penetrate the outer regions of the stratospheric plume 19 hours after the eruption, after the visible cloud lost much of its definition in the satellite infrared images. The plume showed layers of strongly enhanced water vapor over an altitude range from 19 to 28 km. The largest mixing ratio of more than 1300 parts per million by volume (ppmv) was observed at the highest altitude and at the warmest stratospheric temperature reached by the

[1]National Center for Atmospheric Research, Boulder, CO 80301, USA. [2]LACy, Laboratoire de l'Atmosphère et des Cyclones, UMR8105, CNRS, Université de la Réunion, Météo-France, Saint-Denis, La Réunion, France. [3]Bureau of Meteorology, Melbourne, VIC 3001, Australia.
*Corresponding author. Email: voemel@ucar.edu

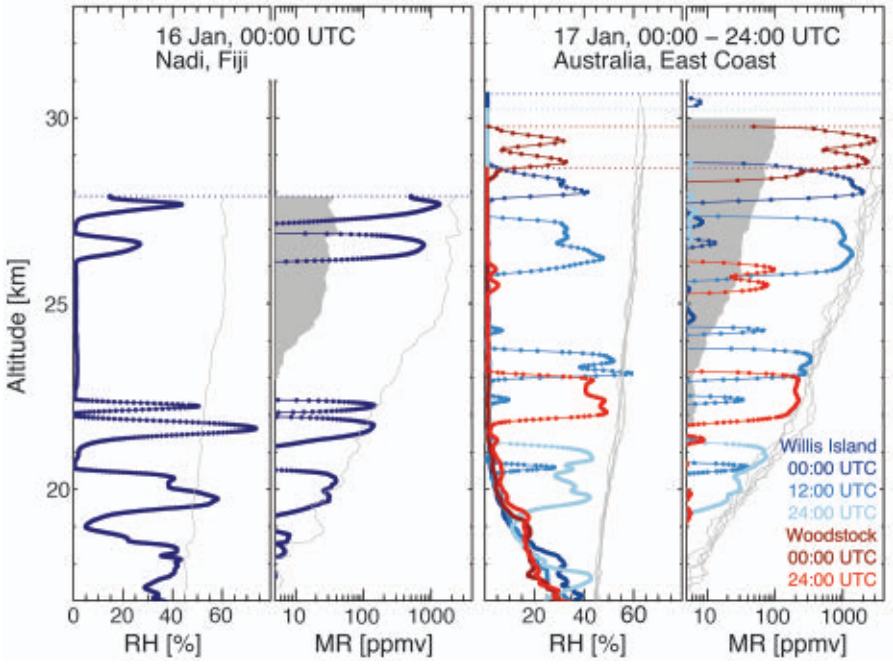



**Fig. 1. First soundings within 48 hours after eruption.** Left panels: Sounding at Nadi, Fiji, 16 January, 00:00 UTC, 19 hours after the eruption started. Right panels: Five soundings at two Australian east coast stations on 17 January. Ice saturation is shown as thin gray line. Balloon burst altitudes are indicated by dotted horizontal lines. The mixing ratio was calculated by first subtracting a 1% relative humidity baseline. The gray shaded area indicates the estimated detection limit.

sounding. The balloon burst within the layer of the highest water vapor mixing ratio, and we have no information about the altitude to which the increased water vapor extended.

The enhanced water vapor layers were detected next at the east coast of Australia at the two stations, Willis Island and Woodstock (Fig. 1). At both stations, high water vapor was observed on 17 January, 00:00 UTC, above 28 km, then descending in altitude in the soundings 12 and 24 hours later. The largest observed mixing ratio was 2900 ppmv, an increase by a factor of 580 compared with stratospheric background concentrations of 5 ppmv.

Because the tropical stratosphere becomes warmer with increasing altitude, the higher an initial volcanic eruption column reaches the more water vapor it can deposit, which is reflected in the observed increase in the mixing ratio peaks with altitude in these early soundings roughly in accordance with the saturation mixing ratio profiles. Owing to the shearing and layering of the eruption plume, hydrometeors injected to these altitudes do not fall out of the stratosphere but evaporate as they settle, moistening dry stratospheric layers underneath. The sounding at Fiji indicated two saturated layers and a possible presence of hydrometeors; the sounding at Willis Island on 17 January, 12:00 UTC, indicated one thin layer of saturation. Therefore, hydration by the evaporation of initially injected hydrometeors was largely complete by 17 January.

The apparent decrease in the altitude at which the layer was observed over the 24-hour period on 17 January is due to slanting of the plume by the wind shear in the Western Pacific region, with easterly winds linearly decreasing from 30 m/s at 30 km to 20 m/s at 20 km. The progressive slanting of the eruption plume is again observed at the subsequent stations under the volcanic plume.

On 18 January 2022 at 00:00 UTC, nine Australian radiosonde stations between the west coast and the east coast (Fig. 2) detected the volcanic water vapor. The mixing ratio profiles show the layers with the highest mixing ratios (>1000 ppmv) at the highest altitudes in the western part and layers with smaller mixing ratios at the lower altitudes in the eastern part of Australia. Several of these soundings terminate within the layer and do not map its total depth.

These soundings roughly map the spatial extent over which the stratospheric plume of enhanced water vapor layer had expanded, which we approximate as a rectangle of 2700 km by 1600 km, covering an area of $4.32 \times 10^6$ km$^2$.

The increase of stratospheric water vapor in the volcanic layers at that synoptic time ranges from 40 g m$^{-2}$ at Broome to 1.4 g m$^{-2}$ at Gove (table S1). Using the mean of all nine soundings, we estimate a total amount of water vapor injected into the stratosphere of about $50 \times 10^{12}$ g of water. Assuming a total amount of stratospheric water vapor of $10^{15}$ g (*15*), this eruption may have increased stratospheric water vapor by at least 5% on a global average. This estimate is a lower bound and less than what has been reported by satellite observations (*16*). Three soundings in the western part of the domain only touched the upper layer and did not map its full vertical extent. In addition, the sondes provide no information on whether water vapor was injected above the highest observed layers. For our calculation, we defined 20 km as the lowest altitude, because it becomes more challenging to separate the volcanic injection from ordinary tropopause processes below that. We also assume that the water vapor layers had not yet traveled much past the station at Broome (122.3°E), because Learmonth at 114.1°E had not yet detected these layers. Lastly, we low-biased the relative humidity profiles by 1% to remove the altitude-dependent baseline in mixing ratio (see data and methods in the supplementary materials).

We tracked the water vapor layer by operational Vaisala RS41 radiosondes for more than 6 weeks, during which it circled the globe more than two times and dispersed longitudinally and latitudinally. A sequence of soundings along 20°S is shown in Fig. 2C. The water vapor mixing ratio peak for these soundings ranged between 237 and 675 ppmv in the altitude range between 25.4 and 27.4 km. Instead of a Vaisala RS41 radiosonde, the sounding at Reunion Island on 22 January used a Cryogenic Frostpoint Hygrometer (CFH), which is used to monitor normal stratospheric water vapor concentration (*17*). This sounding showed a peak of 358 ppmv.

The time series of latitude and longitude of stations at which the layer was observed is shown in Fig. 3. The volcanic water vapor layer had crossed Australia by 20 January and was detected there again on 30 January after the first global circumnavigation. Stations in South America first detected the layer on 24 January and continuously from then on. Tarawa, Republic of Kiribati, (1.4°N, 172.9°E), detected the volcanic layer after it had circled the globe on 30 January.

During the first weeks of tracking, the layer spread more toward the north and stayed confined on its southern extent. Meekatharra, Australia (26.6°S), observed this layer only 3 days after the eruption, which is the farthest south the radiosonde network detected it despite the ability of other Australian and South American stations farther south, until a sounding at Adelaide, Australia (35°S), detected it on 28 February. By contrast, the water vapor layer was observed as far north as Bogotá (4.7°N) by 25 January 2022. The dispersion

pattern indicates transport within the tropics toward and across the equator and a transport barrier toward more southern latitudes above 20 km, consistent with our general understanding of the stratospheric Brewer Dobson circulation (*18*–*20*).

This dispersion also reduced the water vapor concentration within these layers (Fig. 3). Although there were nine soundings showing >1000 ppmv during the first days after the eruption, no sounding detected >100 ppmv after 9 February. After 12 February, the decrease in the water vapor concentration within the layer had substantially slowed and showed a peak concentration of 38 ± 17 ppmv at a center altitude of 24.7 ± 1.1 km and a layer thickness of 1.2 ± 0.8 km. In this layer, the water vapor mixing ratio is increased by up to a factor of 7.

With the decrease in peak mixing ratio, the ability to detect the volcanic water vapor layer using Vaisala RS41 radiosondes decreased as well. The dispersion after 8 weeks may span over a wider latitude range than indicated by this technique. Satellite observations and soundings using the CFH are more suited to observe the dispersion of the cloud after this initial phase. However, compared to satellite observations, in situ soundings are much better suited for showing the vertical distribution of the volcanic water vapor layer (fig. S3).

An air mass injected from a volcano into the stratosphere is not immediately in radiative and thermodynamic equilibrium. Its outflow layer will be at the same density as the environment, but the different chemical composition and the presence of aerosol particles are likely to cause different radiative heating and cooling than its environment, which will cause the volcanic plume to ascend or descend in altitude. Figure 4 shows the weekly averaged profiles of the volcanic layer observed with the entire latitude band after 29 January. The layer maximum decreased from >1000 ppmv to several tens of ppmv within the first 3 weeks after the eruption, and the altitude of the maximum descended from 29 km to about 25 km. This descent may indicate that the radiative cooling by water vapor may have overwhelmed the radiative heating that may have been expected by the presence of sulfate aerosol. Because the detection limit of our technique increases with altitude, the dispersion of the plume at 29 km is not as well quantified as that at 25 km.

Cooling of the water vapor layer is also indicated by the temperature anomaly within this layer. The soundings showing the water vapor layer are on average 2 K colder within and just above the layer compared with the average of all soundings during this study period. For individual soundings, the cooling within the layer maybe even stronger compared with the weekly average. Heating by additional aerosol that may be present within this layer is not observed. This may indicate that a radiative effect of additional aerosol injected by the eruption is small compared to the cooling by water vapor. Previous studies have indicated that an increase in stratospheric water vapor may contribute to stratospheric cooling (*2*, *21*). Therefore, the injected amounts of water vapor may contribute to stratospheric cooling and surface warming over the months to come. This would also imply that the climate effect of this eruption is likely to be different from, for example, that of the Pinatubo eruption, where large amounts of aerosol caused considerable surface cooling (*3*). Because the removal of this excess water vapor is controlled only by stratospheric dynamics and—unlike stratospheric aerosols—not by gravitational settling, the effect of this water vapor increase is likely to last longer than that caused by injected aerosol (*6*).

The large amounts of water in the plume of the Hunga Tonga-Hunga Ha'apai eruption may have limited the amount of $SO_2$ that reached the stratosphere. $SO_2$ is generally believed to be the most important gas emitted by volcanoes,

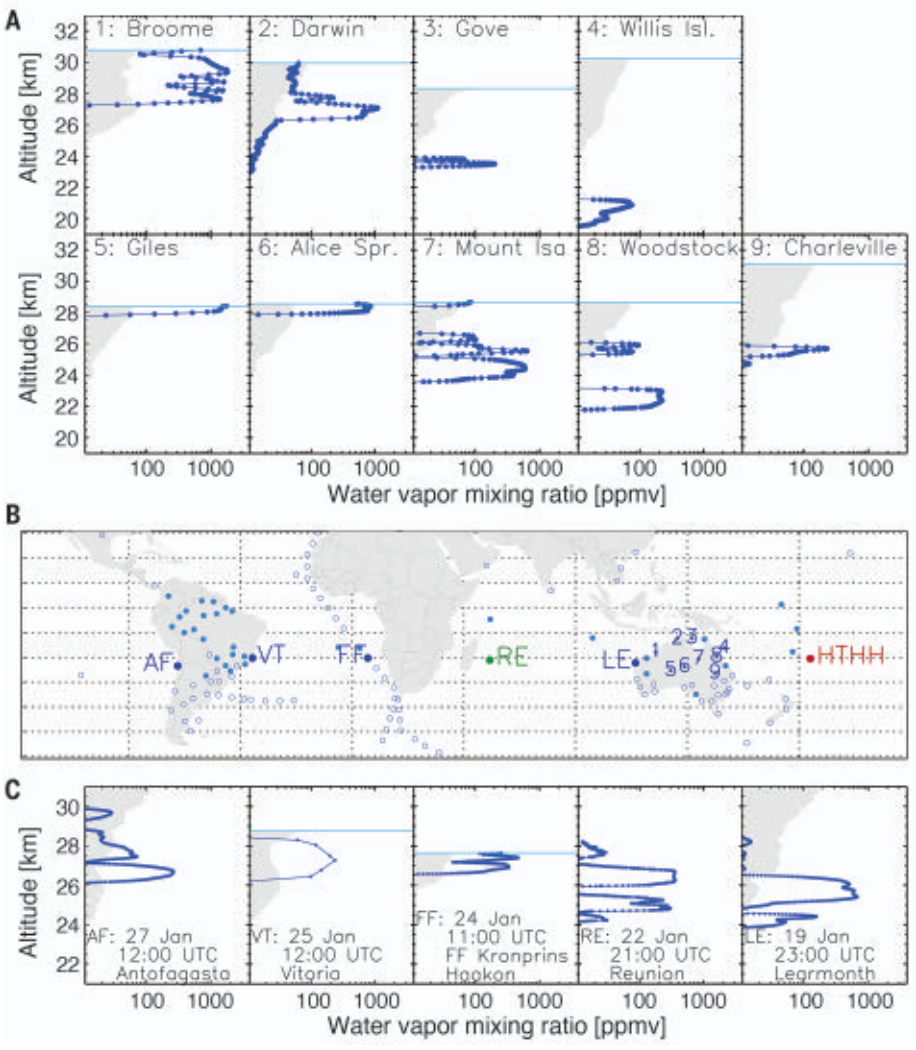



**Fig. 2. Soundings during the first global circumnavigation.** (**A**) All nine Australian soundings, which detected the volcanic water vapor layer on 18 January, 00:00 UTC. Detection limits are shown as grey areas. (**B**) Map of all Vaisala RS41 radiosonde stations used in this study. Stations indicated by solid dots detected the volcanic water vapor layer at least once during the 6 weeks after the eruption. Numbered stations correspond to the profiles shown in (A). The eruption site on 15 January is shown as HTHH. (**C**) Profiles along 20°S [lettered stations in (B)] showing the water vapor layer between 24 and 28 km moving from Learmonth (LE) on 19 January (right panel) to Antofagasta (AF) on 27 January (left panel). The station at Reunion (RE) launched a cryogenic frostpoint hygrometer, which has a detection limit well below 0.1 ppmv, instead of a Vaisala RS41. The research vessel *FF Kronprins Haakon* sailed from Antarctica to Tromsø, Norway, and detected the volcanic plume on 24 and 25 January.

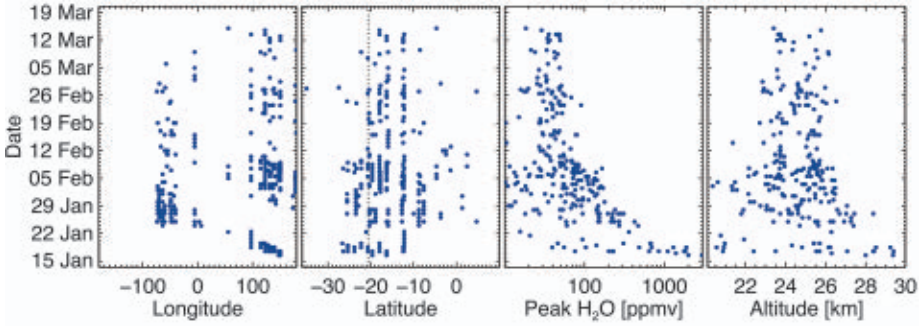
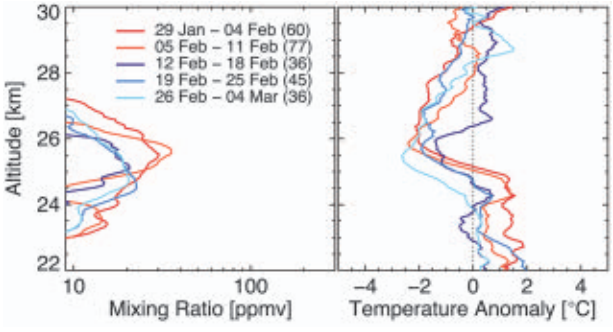



**Fig. 3. Time series of 374 Vaisala RS41 soundings in which the water vapor layer was observed.** The panels show the longitude and latitude of the sounding, the mixing ratio maximum, and its altitude within the observed stratospheric layer. The stations at which these soundings were launched correspond to those shown in Fig. 2. The latitude of the eruption is indicated as a dashed line.

**Fig. 4. Weekly averages of the volcanic water vapor layer.** Left: Mixing ratio average. Right: Temperature anomaly average within the profiles detecting enhanced water vapor derived from the simultaneous temperature observations.




which influences global climate through the formation of stratospheric sulfate aerosols. Washout and the liquid-phase conversion to sulfate in the presence of large amounts of water may have reduced the amount of $SO_2$ during the initial convective ascent into the stratosphere (*4*, *22*). In addition, the large water vapor concentrations that did reach the stratosphere may have accelerated the conversion of $SO_2$ to $H_2SO_4$ (*23*), especially during the early days after the eruption, when the water vapor concentrations were very large.

These large amounts of water vapor are likely to modify the chemical cycles that control the amount of ozone in the stratosphere. The increase in stratospheric water vapor is much larger and occurs at much warmer temperatures than what has been used in chemical transport models (*24*). Owing to the much larger initial amounts of water vapor within this volcanic layer, the increased concentration of OH radicals through the reaction of $H_2O$ with $O(^1D)$ oxygen atoms may decrease the amount of ozone within the water vapor layer. As dispersion decreases the concentration of water vapor, the importance of this process is likely to decrease as well. However, detailed studies will be required to quantify the effect on the amount of ozone because other chemical reactions may play a role as well. Water-soluble compounds such as iodine, bromine, and chlorine from ocean water may have been injected into the stratosphere in substantial amounts. Though it is not yet understood, the role of these trace compounds in this eruption event may further modify the ozone chemistry.

Our observations have been made possible by the availability of a high-quality radiosonde, which is operationally used for weather forecasting and which uses a sensor that is capable of detecting these extremely elevated amounts of stratospheric water vapor. The effort by the World Meteorological Organization to encourage National Hydro-Meteorological Services to share their radiosonde observations in high resolution provided the ability to study this event in detail. Owing to the unusually large amounts of water vapor (*16*) and a possible interference by volcanic aerosol with the remote-sensing observations from satellites, the plume during the first few days after the eruption could only be quantified by the radiosonde network. If volcanic eruptions inject large amounts of water vapor into the stratosphere after the end of life of the current satellite instruments monitoring stratospheric water vapor, balloon-borne instruments may be the only tool to study these events.

## REFERENCES AND NOTES


1. P. M. de F. Forster, K. P. Shine, *Geophys. Res. Lett.* **29**, 10-1–10–4 (2002).
2. S. Solomon *et al.*, *Science* **327**, 1219–1223 (2010).
3. A. Robock, *Rev. Geophys.* **38**, 191–219 (2000).
4. S. Kremser *et al.*, *Rev. Geophys.* **54**, 278–335 (2016).
5. S. Solomon, *Rev. Geophys.* **37**, 275–316 (1999).
6. M. M. Joshi, G. S. Jones, *Atmos. Chem. Phys.* **9**, 6109–6118 (2009).
7. D. G. Murcray, F. J. Murcray, D. B. Barker, H. J. Mastenbrook, *Science* **211**, 823–824 (1981).
8. M. J. Schwartz *et al.*, *Geophys. Res. Lett.* **40**, 2316–2321 (2013).
9. C. E. Sioris, A. Malo, C. A. McLinden, R. D'Amours, *Geophys. Res. Lett.* **43**, 7694–7700 (2016).
10. A. E. Dessler *et al.*, *J. Geophys. Res. Atmos.* **119**, 12588–12598 (2014).
11. S. Fueglistaler, *J. Geophys. Res.* **117** (D13), n/a (2012).
12. S. Bates, M. Carlowicz, Tonga Volcano Plume Reached the Mesosphere (2022); https://earthobservatory.nasa.gov/images/149474/tonga-volcano-plume-reached-the-mesosphere.
13. S. A. Colgate, T. Sigurgeirsson, *Nature* **244**, 552–555 (1973).
14. M. F. Sheridan, K. H. Wohletz, *J. Volcanol. Geotherm. Res.* **17**, 1–29 (1983).
15. G. Stenchikov *et al.*, *J. Geophys. Res. Atmos.* **126**, e2020JD033829 (2021).
16. L. Millan *et al.*, Hunga Tonga-Hunga Ha'apai Hydration of the Stratosphere, Earth and Space Science Open Archive (2022); https://doi.org/10.1002/essoar.10511266.1.
17. H. Vömel, D. E. David, K. Smith, *J. Geophys. Res. D Atmos.* **112**, D08305 (2007).
18. P. Chen, J. R. Holton, R. Swinbank, *Geophys. Res. Lett.* **22**, 259–262 (1995).
19. R. A. Plumb, *J. Meteorol. Soc. Jpn.* **80** (4B), 793–809 (2002).
20. T. Birner, H. Boenisch, *Atmos. Chem. Phys.* **11**, 817–827 (2011).
21. P. M. de F. Forster, K. P. Shine, *Geophys. Res. Lett.* **26**, 3309–3312 (1999).
22. M. Clyne *et al.*, *Atmos. Chem. Phys.* **21**, 3317–3343 (2021).
23. T. Reiner, F. Arnold, *J. Chem. Phys.* **101**, 7399–7407 (1994).
24. S. Robrecht *et al.*, *Atmos. Chem. Phys.* **19**, 5805–5833 (2019).
25. National Weather Service (NWS), GTS BUFR Radiosonde Data, Version 1.0 (2019); https://doi.org/10.26023/VTYW-80GB-2001.
26. H. Vömel, S. Evan, M. Tully, Hunga Tonga-Hunga Ha'apai stratospheric water vapor from Vaisala RS41 radiosondes, Version 1.0. UCAR/NCAR - GDEX (2022); https://doi.org/10.5065/P328-Z959.


## ACKNOWLEDGMENTS


We acknowledge W. Pene of the Fiji Meteorological Service at Nadi, Fiji, for providing the raw data of their soundings on 16 January and 17 January 0:00 UT and the Deutscher Wetterdienst and the Norwegian Polar Institute for providing the raw data for the soundings from the research vessel *FF Kronprins Haakon*. We thank J. Witte, K. Rosenlof, and H. Selkirk for their comments. **Funding:** National Science Foundation (NSF) Cooperative Agreement no. 1852977. **Author contributions:** Conceptualization: H.V. Methodology: H.V. Data curation: H.V., S.E., M.T. Writing – original draft: H.V .Writing – review and editing: S.E., M.T., H.V. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Radiosonde data in TAC were obtained at the University of Wyoming at http://weather.uwyo.edu/upperair/sounding.html. Radiosonde data in BUFR are archived at NCAR at https://doi.org/10.26023/VTYW-80GB-2001 (*25*). The profiles showing the water vapor layer were taken from the collections above and in a few cases high-resolution data were provided to the authors. All profiles are accessible at https://doi.org/10.5065/p328-z959 (*26*).



## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abq2299
Data and Methods
Figs. S1 to S3
Table S1
References (*27*–*31*)


Submitted 2 April 2022; accepted 19 August 2022
10.1126/science.abq2299

NEUROSCIENCE

# Dendritic axon origin enables information gating by perisomatic inhibition in pyramidal neurons

Alexander Hodapp[1]†, Martin E. Kaiser[1]†, Christian Thome[1,2,3], Lingjun Ding[4,5,6], Andrei Rozov[1,7,8], Matthias Klumpp[1], Nikolas Stevens[1], Moritz Stingl[1], Tina Sackmann[1], Nadja Lehmann[9], Andreas Draguhn[1], Andrea Burgalossi[4,5], Maren Engelhardt[2,9], Martin Both[1]*

Information processing in neuronal networks involves the recruitment of selected neurons into coordinated spatiotemporal activity patterns. This sparse activation results from widespread synaptic inhibition in conjunction with neuron-specific synaptic excitation. We report the selective recruitment of hippocampal pyramidal cells into patterned network activity. During ripple oscillations in awake mice, spiking is much more likely in cells in which the axon originates from a basal dendrite rather than from the soma. High-resolution recordings in vitro and computer modeling indicate that these spikes are elicited by synaptic input to the axon-carrying dendrite and thus escape perisomatic inhibition. Pyramidal cells with somatic axon origin can be activated during ripple oscillations by blocking their somatic inhibition. The recruitment of neurons into active ensembles is thus determined by axonal morphological features.

A hallmark of neuronal network activity is the selective recruitment of neurons into active ensembles, which form transiently stable patterns of activity (*1*–*3*). In the mammalian hippocampus, the repetitive activation of such neuronal ensembles during ripple (~200 Hz) oscillations supports the consolidation of spatial and declarative memories (*4*, *5*). A central question is how individual neurons are selected for participation in these patterns of coactivity. The activation of specific neurons has been suggested to result from the convergence and use-dependent plasticity of excitatory synapses (*6*, *7*). By contrast, the global, strong perisomatic inhibition would provide a common, unspecific gain control mechanism for all local neurons and an oscillating temporal scaffold for the embedded spatiotemporal activity patterns (*8*). Recent work has revealed a marked morphological and functional heterogeneity among principal cells in cortical networks (*9*–*12*), including the nonsomatic (dendritic) origin of axons in a subset of pyramidal cells (Fig. 1, A to C) (*13*, *14*). The differential recruitment of individual neurons could be determined by the morphological feature of axon onset. We studied this possibility in hippocampal ripple oscillations, which activate specific neuronal ensembles and recruit pronounced perisomatic inhibition (*15*, *16*).

We recorded juxtacellular activity from single neurons in the CA1 region of awake, head-fixed mice together with local field potentials representing the overall network state (Fig. 1, D and E). In CA1, ~50% of pyramidal cells have an axon originating from a basal dendrite (fig. S1) (*13*), opening the possibility of a functional distinction between axon-carrying dendrite cells (AcD cells) and canonical non-AcD cells. This distinction is further supported by a bimodal distribution of the distance between soma and axon initial segment (fig. S1C). Cells were filled with biocytin, reconstructed ex vivo, and classified into two groups with respect to the site of axon origin (Fig. 1, B and C, and figs. S1 to S3). During in vivo recordings, the firing probability of AcD cells during ripples was ~4.5-fold higher than for non-AcD cells, and the firing frequency during ripples was ~2.5-fold higher (Fig. 1, F and G, and table S1). By contrast, there was no difference in firing frequency outside of ripples (Fig. 1H and table S1). A more detailed analysis of field potentials revealed a difference in the power of spike-accompanied ripples, with larger power for AcD than non-AcD cells. AcD cells therefore fire spikes during cycles with particularly strong inhibition, in contrast to non-AcD cells (Fig. 1I and table S1) (*15, 16*).

We hypothesized that the preferred ripple-associated firing of AcD cells is caused by their morphology: Excitatory inputs to the axon-carrying basal dendrite escape perisomatic inhibition and allow action potential (AP) generation even during pronounced activation of γ-aminobutyric acid (GABA)–ergic interneurons. We tested this mechanism in acute hippocampal slices, which allowed us to study cell-network coupling under well-controllable conditions (*17*). The preferential activation of AcD cells during ripples was maintained in this preparation (Fig. 2, A and B, and table S2). First, we assessed whether AcD and non-AcD cells receive different synaptic input during ripples. Subthreshold inhibitory and excitatory postsynaptic potentials were recorded at different membrane potentials and revealed similar inhibition-excitation conductance ratios (I/E) (*18*) for both cell types. Likewise, the inhibitory and excitatory current ratio was not different (Fig. 2C and tables S2 and S3). There was no difference in the relative timing of inhibitory and excitatory events (fig. S4, A and B, and tables S2 and S3). A direct analysis of perisomatic inhibition by paired recordings from fast-spiking parvalbumin-positive inhibitory interneurons and pyramidal cells confirmed identical coupling probability and strength for AcD and non-AcD cells, respectively (Fig. 2D; fig. S4, C to E; and tables S4 and S5). Is the proposed privileged role of the axon-carrying basal dendrite reflected in peculiar morphological features? Both AcD and non-AcD cells had a similar number of basal dendrites (median = 3) (Fig. 2, E and F, and tables S3 to S5) and no differences in branching pattern, total dendritic length, and spine density (Fig. 2, G and H, and fig. S5). However, the AcD was longer than basal dendrites of non-AcD cells and made up for ~35% of basal dendritic length (~12% of total dendritic length in our reconstructed neurons) (Fig. 2I; supplementary materials, materials and methods; fig. S5C; and tables S3 to S5), which indicates a significant weight of synaptic input to this particular dendrite. Nevertheless, synaptic inhibition, I/E ratio, and dendritic arborization were largely similar between both cell types. Thus, factors other than synaptic input seem to determine the preferential firing of AcD cells. This hypothesis is further supported by the more negative threshold of ripple-associated APs in AcD cells, which is indicative of a noncanonical site of AP initiation (Fig. 2K and table S2) (*17*).

To assess the impact of axon origin on firing probability and threshold under different conditions of synaptic input, we used a detailed multicompartment cellular computer model (Fig. 3). Cells consisted of a soma and three dendrites with axon origin at the soma (non-AcD) or a basal dendrite (AcD), respectively (Fig. 3A). First, we modeled transient excitatory input to the AcD together with transient perisomatic inhibition, with typical postsynaptic kinetics observed during ripples. AcD cells fired APs more readily than did non-AcD cells and generated AP waveforms indicative of distal AP initiation (Fig. 3A). A systematic variation of inhibitory and excitatory input strength revealed a much broader variety of synaptic input combinations that triggered APs in AcD compared with non-AcD cells

[1]Institute of Physiology and Pathophysiology, Medical Faculty, Heidelberg University, Heidelberg, Germany. [2]Institute of Anatomy and Cell Biology, Medical Faculty, Johannes Kepler University, Linz, Austria. [3]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University, Stanford, CA, USA. [4]Institute of Neurobiology, University of Tübingen, Tübingen, Germany. [5]Werner-Reichardt Centre for Integrative Neuroscience, Tübingen, Germany. [6]Graduate Training Centre of Neuroscience, IMPRS, Tübingen, Germany. [7]Federal Center of Brain Research and Neurotechnologies, Moscow, Russian Federation. [8]OpenLab of Neurobiology, Kazan Federal University, Kazan, Russian Federation. [9]Institute of Neuroanatomy, Mannheim Center for Translational Neuroscience (MCTN), Medical Faculty Mannheim, Heidelberg University, Heidelberg, Germany.
*Corresponding author. Email: mboth@physiologie.uni-heidelberg.de
†These authors contributed equally to this work.

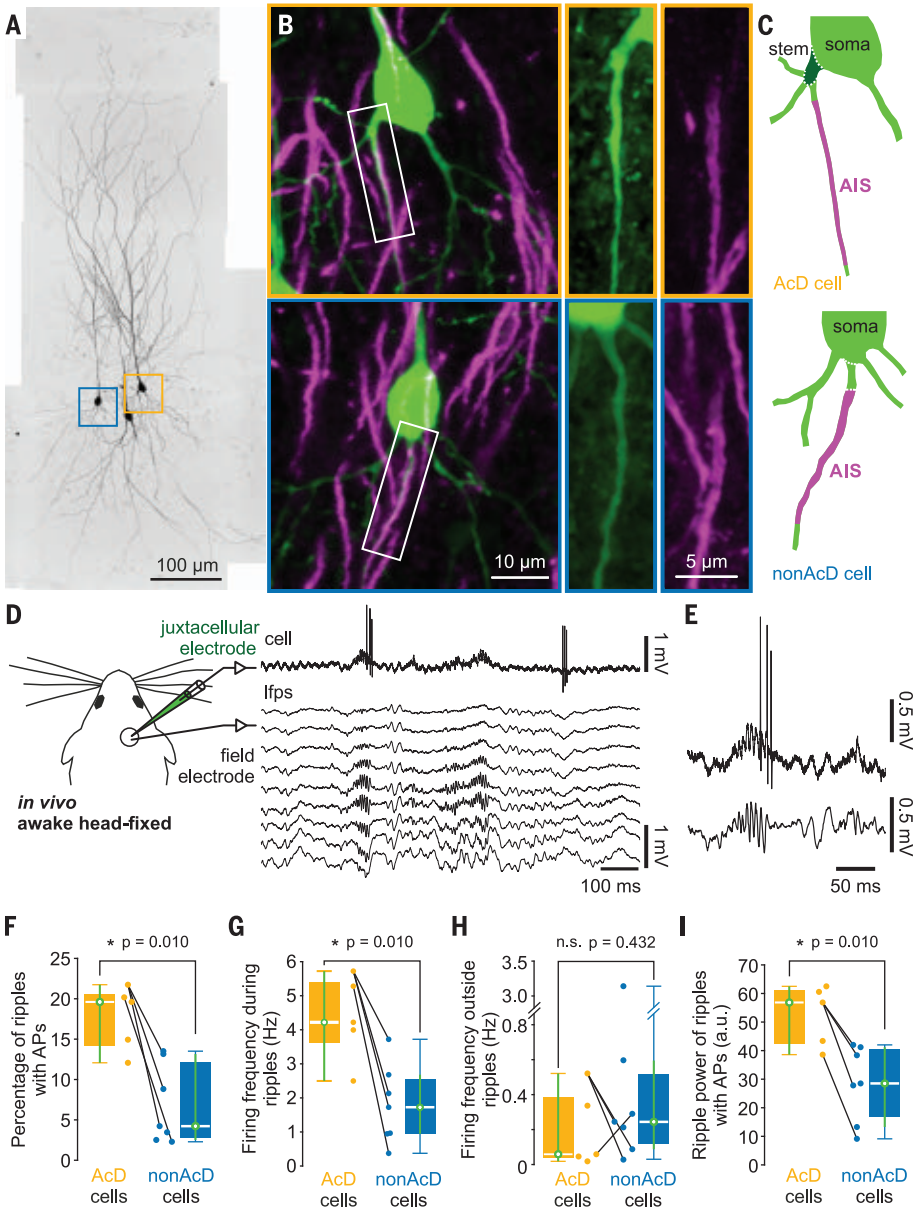



**Fig. 1. Preferential recruitment of AcD cells during ripples in vivo.** (**A** to **C**) Distinction between AcD and non-AcD cells. (A) Reconstruction of biocytin-filled pyramidal cells (left and right) together with a fast-spiking interneuron (middle) (Fig. 2 and fig. S4). (B) Soma-near region of the AcD (top) and non-AcD (bottom) cell. Cell is in green, axon initial segment (AIS) is in magenta, and the overlay is outlined in white. (C) Schematic representation showing the dendritic origin of the AcD cell. Details are available in figs. S1 and S2. (**D**) Single hippocampal CA1 pyramidal neurons were recorded juxtacellularly in awake, head-fixed mice. Simultaneously, local field potentials (lfps) were recorded with a 16-channel silicon probe. Ripples were identified by their characteristic frequency of 140 to 200 Hz. (**E**) Magnification of a representative ripple event from recording in (D). (**F**) Percentage of ripples with APs was larger for AcD cells than for non-AcD cells. (**G**) Similarly, AcD cells had a higher firing frequency during ripples than that of non-AcD cells. (**H**) Firing frequency outside of ripples was not different. n.s., not significant. (**I**) Mean power of ripples with APs was larger for AcD cells than for non-AcD cells. Cells recorded in the same animal are connected by black lines (five AcD cells from four animals and seven non-AcD cells from four animals). a.u., arbitrary unit.

(Fig. 3B). Firing thresholds were decidedly more negative for AcD cells, which is consistent with our experimental findings (Fig. 4C). Recording electrodes are typically located in the soma, yielding different apparent I/E conductance ratios compared with the site of origin of the synaptic conductances. Such apparent (somatically recorded) conductance ratios in a model cell are shown in Fig. 3B, bottom. Strong increases in local (dendritic) AMPA conductance can go along with small changes in apparent I/E ratio. Thus, APs in AcD cells may be caused by particularly strong excitation of the AcD, whereas somatically recorded I/E ratios appear similar for AcD and non-AcD cells. This mechanism implies that in the presence of perisomatic inhibition, excitatory input to the AcD becomes more efficient with increasing distance between axon and soma. This was confirmed in model calculations: The difference in excitability by input to the AcD versus non-AcD branch was increased by increasing axon-to-soma distance as well as by increasing perisomatic inhibition (Fig. 3C). The increased AP propensity of AcDs was markedly present even at short axon distances <5 µm, covering the empirical distribution of axon onsets (fig. S1, C and D). Thus, the model supports our hypothesis and emphasizes the privileged function of the AcD for participation in network activity.

The causal relationship between axon origin, perisomatic inhibition, and firing propensity predicts that functional differences between AcD and non-AcD cells should be diminished when perisomatic inhibition is reduced. We tested this by blocking GABA type A ($GABA_A$) receptors in individual pyramidal cells by means of picrotoxin loading through the intracellular pipette, which leaves the network-level I/E balance unaltered (Fig. 4, A and B). This procedure resulted in a strongly reduced intracellular I/E conduction ratio and increased firing probability (Fig. 4C and table S2). Under these conditions, non-AcD cells readily fired APs during ripples, in contrast to recordings with intact inhibition. Likewise, the apparent somatic AP threshold was shifted to more positive values, as predicted by our model (Fig. 4C, right, and table S2). Because of their higher firing propensity in this paradigm, the ripple-associated firing of non-AcD cells allowed for a correlation analysis between I/E ratio and firing probability. The result supports our proposed mechanism: Non-AcD cells showed a highly significant correlation, with lower I/E ratios favoring firing during ripples, whereas

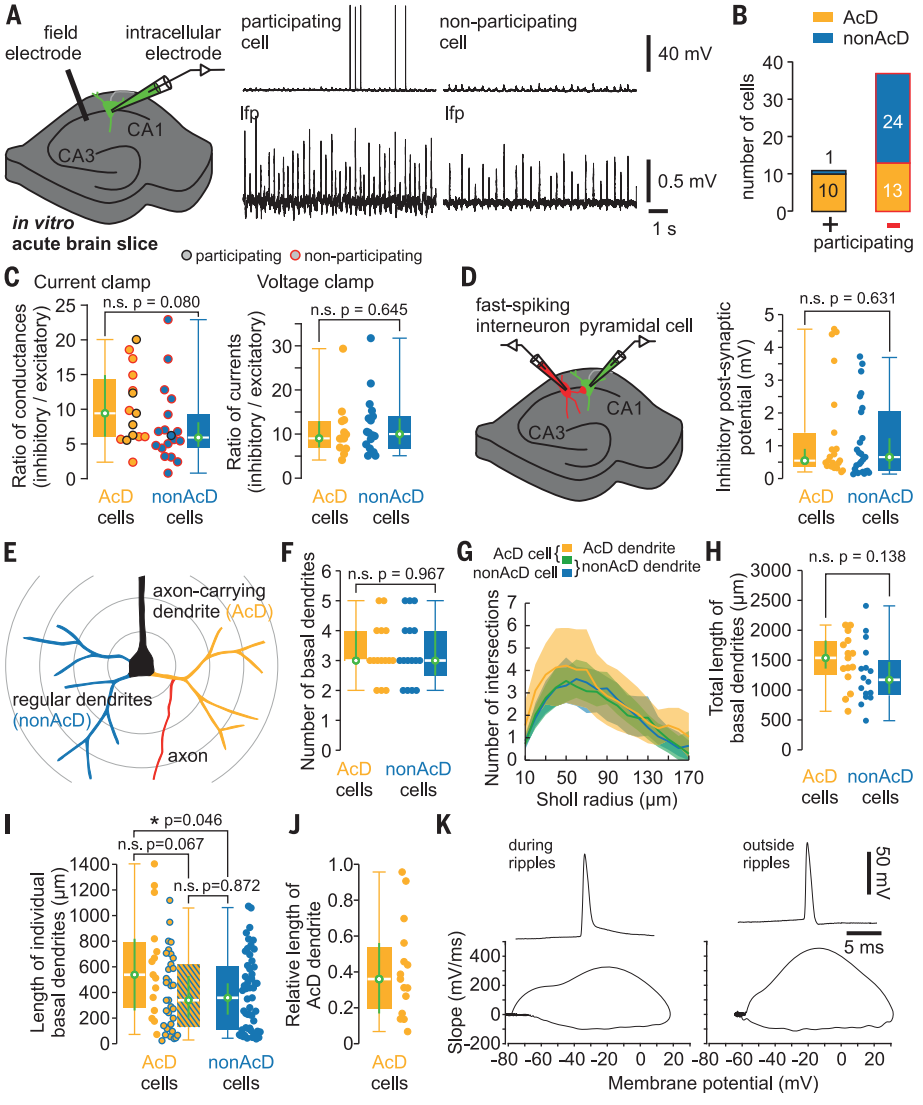



**Fig. 2. AcD and non-AcD cells receive similar synaptic input and have similar dendritic morphology.** (**A**) Single hippocampal CA1 pyramidal neurons were recorded with sharp electrodes in acute brain slices from mice. Simultaneously, the lfp was recorded with a single field electrode. Ripple oscillations occurred spontaneously and could be well detected by the prominent positive sharp wave (bottom right). Neurons were classified into participating cells (firing spikes during ripples) and nonparticipating cells. (**B**) Similar to in vivo ripples, AcD cells have a higher propensity to fire APs during ripples than that of non-AcD cells ($P$ = 0.002, Fisher's exact test). (**C**) Subthreshold excitatory and inhibitory synaptic events during ripples. Ratios of conductance changes (from current clamp experiments) and currents (from voltage clamp experiments) are similar in AcD and non-AcD cells. (**D**) Paired whole-cell recordings from presynaptic parvalbumin-positive fast-spiking inhibitory interneurons and postsynaptic CA1 pyramidal neurons. Connection probability (47.8% for AcD cells and 53.2% for non-AcD cells; $P$ = 0.680) and synaptic strength were not different between the two groups (22 connected of 46 recorded AcD cells versus 25 connected of 47 recorded non-AcD cells). In these experiments, high-chloride intracellular solution was used to artificially render inhibitory postsynaptic potentials depolarizing at resting membrane potential for more accurate quantification. (**E** to **J**) Morphology of basal dendrites was analyzed and compared between AcD and non-AcD cells. (F) The number of basal dendrites is not different between the two groups. (G) Sholl analysis reveals a tendency to higher complexity of AcDs compared with canonical dendrites in both AcD and non-AcD cells. (H) Total length of basal dendrites is not different between AcD and non-AcD cells. (I) AcDs are longer than canonical dendrites from non-AcD cells. (J) AcDs compose 36% of the total length of basal dendrites in AcD cells. Data were quantified from 15 AcD cells versus 16 non-AcD cells. (**K**) APs occurring during ripples have a more negative voltage threshold than APs occurring outside ripples. Phase plots show that APs recorded at the soma during ripples consist of two distinct phases during the upstroke, indicating an electrically distant origin of the AP.

AcD cell firing was much less dependent on the I/E ratio (Fig. 4D).

Together, these findings reveal a mechanism for differential recruitment of pyramidal neurons into network activity, depending on their axon origin. AcD cells retain the ability to fire APs even in situations in which firing of canonical (non-AcD) pyramidal cells is largely prohibited by GABAergic inhibition (Fig. 4E). In such network states, activation of AcD cells is largely confined to excitatory inputs at the AcD. The AcD contains ~1/3 of all spines at basal dendrites and makes up a relevant part of the entire dendritic tree (Fig. 2I and fig. S5), likely receiving notable excitatory input. The privileged function of this dendrite allows for state-dependent switches of the functional connectivity of the network: During phases of strong perisomatic inhibition, excitatory inputs are most efficient at the AcD, whereas during less pronounced perisomatic inhibition, inputs to all dendrites contribute more equally (Fig. 4E). This morpho-functional mechanism explains how specific cells are preferentially activated during ripples in hippocampal networks (*19*–*21*). Thus, the site of axon origin in combination with perisomatic inhibition defines the group of potentially active neurons (Fig. 4F), whereas the individual members of active ensembles are likely selected by additional mechanisms, including the strength and plasticity of excitatory synaptic inputs (*22*). According to our model, a substantial portion of this input must arrive at the axon-carrying basal dendrite, which expresses supralinear signal integration (*13*) and may, therefore, contribute to the temporal precision of firing during high-frequency ripple oscillations (*23*). Axon distance from the soma is a continuous parameter (fig. S1, C and D), and hence, the degree of functional coupling to ripples may vary between cells. However, even our simplified, categorial classification shows large differences of AcD versus non-AcD cell recruitment (Figs. 1 and 2).

The selective activation of neurons is fundamental for information processing and memory formation in cortical networks (*1*–*3*, *24*). Although the underlying mechanisms are largely unresolved, most models emphasize differences in excitatory synaptic activation (*6*, *7*, *9*–*11*, *19*–*21*). Recent evidence shows that neurons that underwent learning-related plasticity subsequently display increased synaptic excitation and participation in ripple oscillations (*22*). Our data provide an additional, complementary mechanism for preselection of activatable neurons: We propose that the location of the axon is a key determinant of asymmetric recruitment in oscillating network states. Excitatory inputs on AcDs evade perisomatic inhibition, so that AcD cells are clearly more prone to participate in ripple oscillations. This discovery helps to explain how the

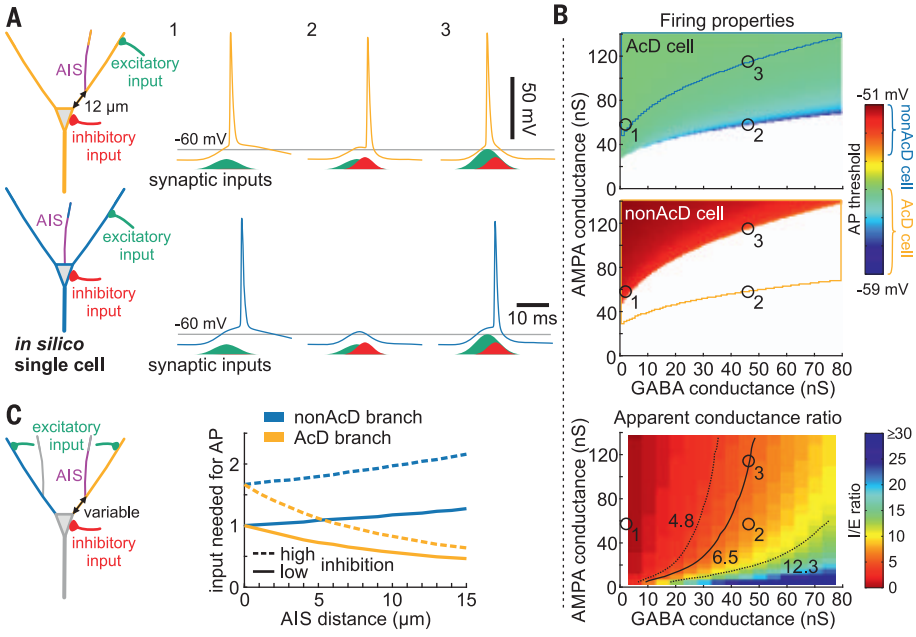



**Fig. 3. Single-cell multicompartment computer modeling predicts the observed differences in firing behavior between AcD and non-AcD cells.** (**A**) (Left) Schematics of an AcD cell with an axon origin at 12 µm from (top) the soma and (bottom) a non-AcD cell. (Right) Three different conditions of phasic synaptic inputs leading to different propensities of AP generation and AP thresholds. (**B**) Firing and AP threshold for a broad range of inhibitory (*x* axis) and excitatory (*y* axis) conductance changes. Colors indicate the firing threshold. White areas indicate that no AP was generated. The three different conditions shown in (A) are marked by circles. (Top) Firing of the model AcD cell. The blue line indicates limits of AP generation for the non-AcD cell. (Middle) Firing and AP threshold of the model non-AcD cell. The yellow line indicates the limits of AP generation for the AcD cell. (Bottom) Apparent I/E ratio as assessed from a simulated somatic recording. The 25th, 50th, and 75th percentile of conductance ratios recorded in single neurons in vitro (Figs. 2C and 4C) are indicated with black lines (numbers show the according I/E ratio). (**C**) In AcD cells, AcD branch has a higher propensity to elicit APs than do non-AcDs. This difference increases with axon distance (*x* axis) and with higher I/E ratio (dotted versus straight lines).

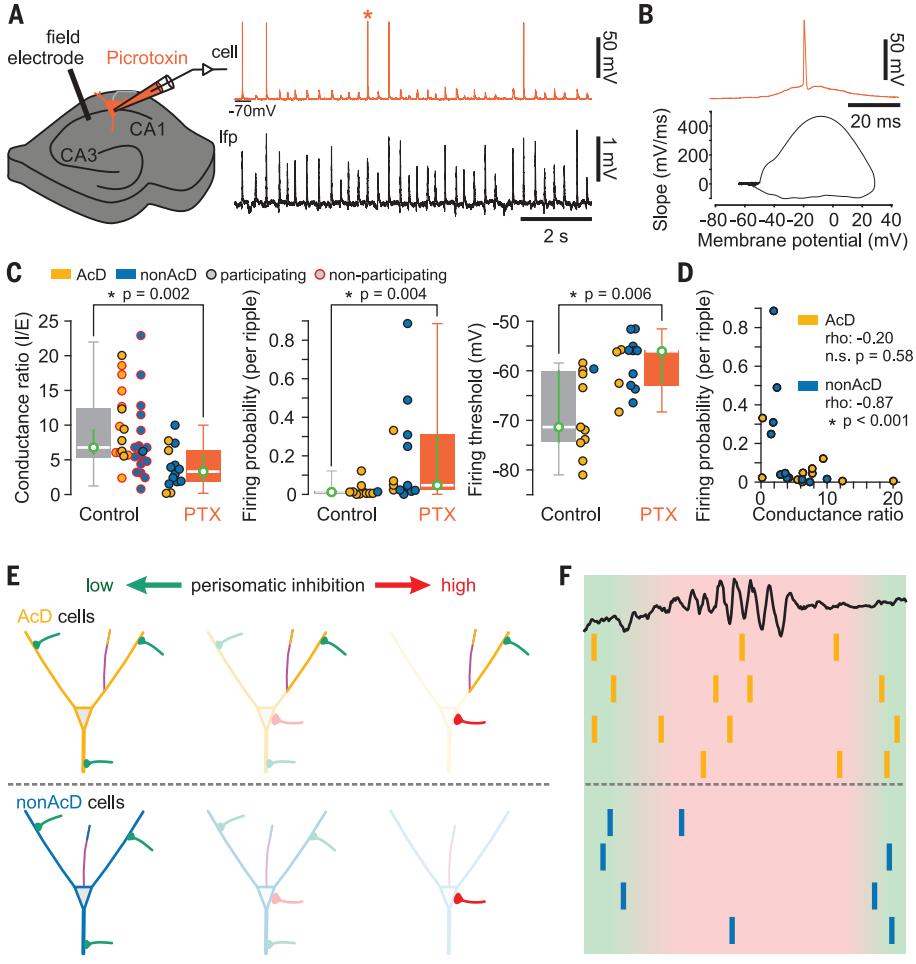



**Fig. 4. Differential gating of information processing by perisomatic inhibition in AcD and non-AcD cells.** (**A**) GABAergic transmission to the recorded cell was blocked by adding picrotoxin (1 mM) to the internal electrode solution, leaving global network oscillations unchanged. This enabled the recorded non-AcD cell to fire APs during ripples [classified as participating cells (Fig. 4C)]. (**B**) Phase plot of the AP marked by an asterisk in (A). There is similarity to Fig. 2K, right, which is indicative of a canonical location of AP generation. (**C**) Reducing perisomatic inhibition diminishes differences in firing characteristics of AcD and non-AcD cells during ripples. (Left) I/E ratio is strongly reduced by intracellular picrotoxin. (Middle) Picrotoxin-filled cells, including non-AcD cells, increase their firing probability during ripples. (Right) Additionally, firing thresholds shift to more positive values, which is typical for canonical AP generation. PTX, picrotoxin. (**D**) Firing probability is negatively correlated with I/E conductance ratio in non-AcD cells but not in AcD cells. (**E**) Schematic representation of the different excitability of AcD and non-AcD cells, respectively. Perisomatic inhibition increases from left to right. (Bottom) In non-AcD cells, AP generation is globally suppressed under perisomatic inhibition. In AcD cells, however, AP generation is still possible upon excitatory inputs at the AcD. (**F**) AcD cells maintain the possibility of firing APs even in situations of pronounced inhibition, such as ripple oscillations.

activation of selected pyramidal cells can be reconciled with the strong and global perisomatic inhibition during network oscillations. It will be important to study whether the site of axon origin undergoes activity-dependent structural plasticity, similar to the established homeostatic remodeling of the axon initial segment that regulates cellular excitability (*25*–*29*). Likewise, it remains to be shown whether AcDs receive excitatory input from specific upstream areas, especially during network states with strong inhibition. Given the abundance of similar axon morphologies in other cortical and subcortical areas of the vertebrate brain (*14*, *30*–*34*), it may well be that the selection of active neurons by their axon origin is a more widespread principle.

## REFERENCES AND NOTES


1. B. M. Sweis, W. Mau, S. Rabinowitz, D. J. Cai, *Curr. Opin. Neurobiol.* **67**, 199–206 (2021).
2. G. Dragoi, *Curr. Opin. Neurobiol.* **64**, 111–118 (2020).
3. L. Carrillo-Reid, R. Yuste, *Curr. Opin. Neurobiol.* **64**, 89–95 (2020).
4. G. Girardeau, K. Benchenane, S. I. Wiener, G. Buzsáki, M. B. Zugaro, *Nat. Neurosci.* **12**, 1222–1223 (2009).
5. A. Fernández-Ruiz *et al*., *Science* **364**, 1082–1086 (2019).
6. G. Buzsáki, *Neuron* **68**, 362–385 (2010).
7. D. O. Hebb, *The Organization of Behavior—A Neuropsychological Theory* (Psychology Press, 1949).
8. G. Buzsáki, D. Tingley, *Trends Cogn. Sci.* **22**, 853–869 (2018).
9. I. Soltesz, A. Losonczy, *Nat. Neurosci.* **21**, 484–493 (2018).
10. F. Sharif, B. Tayebi, G. Buzsáki, S. Royer, A. Fernandez-Ruiz, *Neuron* **109**, 363–376.e6 (2021).
11. M. Valero *et al*., *Nat. Neurosci.* **18**, 1281–1290 (2015).
12. L. Ding *et al*., *eLife* **11**, e71720 (2022).
13. C. Thome *et al*., *Neuron* **83**, 1418–1430 (2014).
14. P. Wahle *et al*., *eLife* **11**, e76101 (2022).
15. J. Gan, S. M. Weng, A. J. Pernía-Andrade, J. Csicsvari, P. Jonas, *Neuron* **93**, 308–314 (2017).
16. D. F. English *et al*., *J. Neurosci.* **34**, 16509–16517 (2014).
17. F. Bähner *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **108**, E607–E616 (2011).
18. L. J. Borg-Graham, C. Monier, Y. Frégnac, *Nature* **393**, 369–373 (1998).
19. J. Epsztein, M. Brecht, A. K. Lee, *Neuron* **70**, 109–120 (2011).
20. J. S. Lee, J. J. Briguglio, J. D. Cohen, S. Romani, A. K. Lee, *Cell* **183**, 620–635.e22 (2020).
21. K. Z. Tanaka *et al*., *Science* **361**, 392–397 (2018).
22. M. Mizunuma *et al*., *Nat. Neurosci.* **17**, 503–505 (2014).
23. S. Jahnke, M. Timme, R. M. Memmesheimer, *J. Neurosci.* **35**, 16236–16258 (2015).
24. J. Wolfe, A. R. Houweling, M. Brecht, *Curr. Opin. Neurobiol.* **20**, 306–312 (2010).
25. N. Jamann *et al*., *Nat. Commun.* **12**, 23 (2021).
26. H. Kuba, Y. Oichi, H. Ohmori, *Nature* **465**, 1075–1078 (2010).
27. M. S. Grubb, J. Burrone, *Nature* **465**, 1070–1074 (2010).
28. W. Wefelmeyer, D. Cattaert, J. Burrone, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 9757–9762 (2015).
29. K. Martinello *et al*., *Neuron* **85**, 346–363 (2015).
30. M. H. Kole, R. Brette, *Curr. Opin. Neurobiol.* **51**, 52–59 (2018).
31. M. S. Hamada, S. Goethals, S. I. de Vries, R. Brette, M. H. Kole, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14841–14846 (2016).
32. L. C. Triarhou, *Front. Neuroanat.* **8**, 133 (2014).
33. F. Höfflin *et al*., *Front. Cell. Neurosci.* **11**, 332 (2017).
34. M. Häusser, G. Stuart, C. Racca, B. Sakmann, *Neuron* **15**, 637–647 (1995).


## ACKNOWLEDGMENTS

We thank L. Menendez de la Prida, M. Valero, J. Epsztein, C. Bernard, and T. Künsting for insightful discussions and comments and N. Zuber, Y. Yanovsky, P. Geschwill, B. Throm, and J. M. Janssen for excellent technical assistance. We acknowledge the Nikon Imaging Center at Heidelberg University and Nikon for access to laser scanning confocal microscopy (A1 and AXR). **Funding:** This work was supported by the German Research Foundation (grant SFB1134, projects A01 and A03, DFG BO 3512/2-1, DFG EN 1240/2-1, DFG BU 3126/2-1, and DFG WBP Fellowship PN 458054460). **Author contributions:** M.B., M.E., A.B., and A.D. conceived of and designed the experiments; A.H., M.E.K., L.D., M.K., A.R., M.S., and M.B. performed experiments; A.H., M.E.K., C.T., M.E., A.R., N.S., N.L., T.S., and M.B. analyzed the data; and M.B. and A.D. wrote the original manuscript. All authors revised and edited the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data and analysis scripts that support the findings of this study are publicly available at https://doi.org/10.11588/data/JWLFFZ. 

## SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj1861
Materials and Methods
Figs. S1 to S5
Tables S1 to S6
References (*35*–*55*)
MDAR Reproducibility Checklist

Submitted 28 April 2021; resubmitted 3 June 2022
Accepted 19 August 2022
10.1126/science.abj1861

MAMMALIAN PHYSIOLOGY

# Retia mirabilia: Protecting the cetacean brain from locomotion-generated blood pressure pulses

**M. A. Lillie[1]*, A. W. Vogl[2], S. G. Gerard[1]†, S. Raverty[1,3], R. E. Shadwick[1]**

**Cetaceans have massive vascular plexuses (retia mirabilia) whose function is unknown. All cerebral blood flow passes through these retia, and we hypothesize that they protect cetacean brains from locomotion-generated pulsatile blood pressures. We propose that cetaceans have evolved a pulse-transfer mechanism that minimizes pulsatility in cerebral arterial-to-venous pressure differentials without dampening the pressure pulses themselves. We tested this hypothesis using a computational model based on morphology from 11 species and found that the large arterial capacitance in the retia, coupled with the small extravascular capacitance in the cranium and vertebral canal, could protect the cerebral vasculature from 97% of systemic pulsatility. Evolution of the retial complex in cetaceans—likely linked to the development of dorsoventral fluking—offers a distinctive solution to adverse locomotion-generated vascular pulsatility.**

Numerous cardiovascular adaptations allow cetaceans (whales, dolphins, and porpoises) and pinnipeds (seals, sea lions, and walruses) to make extraordinary breath-hold dives, but some of the adaptations are group specific (*1*). The blood supply to the cetacean brain and spinal cord differs radically from that in pinnipeds, passing through a series of massive retia mirabilia, or vascular networks located in the thorax, vertebral canal, and cranial cavity (*2*–*4*) (Fig. 1). Such differences indicate that diving cetaceans and pinnipeds face different vascular challenges.

In general, locomotion can generate internal pressures that affect cardiorespiratory performance (*5*, *6*). In galloping horses, dorsoventral flexion of the lumbosacral spine generates pulsatile pressures in the thorax and abdomen (*7*, *8*), and similar flexion occurs in fluking (swimming) cetaceans but not in pinnipeds (*9*). Spinal flexion pushes abdominal viscera against the diaphragm, displacing air from the lungs (*7*). Air outflow moderates pressures in terrestrial animals, but because cetaceans swim on a breath-hold, fluking can generate substantial pulsatile pressures in the thorax and abdomen (*10*–*12*) that can be transmitted through the arterial and venous systems into cerebral beds (*13*). Normally, capillary flow is nonpulsatile (*14*), but because the brain has high flow, its microvasculature is particularly exposed to pulsatile damage (*15*, *16*). We hypothesized that the retia protect the brain from fluking-generated arterial and venous pulses and sought to determine the mechanism involved.

We examined retial morphology in a mysticete, or baleen whale (*Balaenoptera physalus*, fin whale), combined with morphology from a previous study of an odontocete, or toothed whale (*Tursiops truncatus*, bottlenose dolphin) (*17*), to identify a mechanism by which the retia could affect cerebral hemodynamics. As previously described (*2*–*4*, *18*, *19*), arterial inflow to the retial system originates from branches off the thoracic aorta and enters a thoracic rete immediately dorsal to the lungs (Fig. 1). This thoracic rete is composed almost exclusively of small, intercommunicating arteries (Fig. 2A). Extensions of the thoracic rete pass through the intervertebral foramina into the vertebral canal and merge with an extensive spinal rete—a mix of arteries and veins in close association with the spinal cord and extradural veins (Fig. 2, B, C, and E). In *T. truncatus*, the spinal rete starts in the lumbar region and

[1]Department of Zoology, University of British Columbia, Vancouver, BC, Canada. [2]Life Sciences Institute and Department of Cellular & Physiological Sciences, University of British Columbia, Vancouver, BC, Canada. [3]Animal Health Centre, Ministry of Agriculture, Abbotsford, BC, Canada.
***Corresponding author. Email: lillie@zoology.ubc.ca**
†Present address: School of Energy, British Columbia Institute of Technology, Burnaby, BC, Canada.

connects through the foramen magnum to a small cranial rete (*17*). In *B. physalus*, both the thoracic and the spinal retia are short, starting at about the second thoracic vertebra, but they connect to a large cranial rete (Fig. 2D) before arteries coalesce into four efferent vessels that supply the brain. The cranial rete forms a thick, sponge-like mass ventral and lateral to the brain. A transverse section at the level of the pituitary shows arteries surrounded by a similar volume of veins (Fig. 2F) and frequently running through veins (Fig. 2, G and H), making the pressure outside an artery equal to venous blood pressure. Because the spinal and cranial retia (referred to here as the spinocranial rete) lie within the rigid vertebral canal and cranium, pressure pulses are transmitted virtually instantaneously throughout these compartments, thereby sharing arterial and venous pulses through the extravascular cerebrospinal fluid (CSF).

Although cerebral capillary flow is nearly smooth in mammals, almost all cerebral arterial, venous, and CSF flow is pulsatile. The pulses are synchronous with the heart, but the amplitudes of their pressure waveforms differ. Blood vessels are compliant and coupled to the CSF, which allows direct pressure transfer, but amplitude depends on total intracranial compliance (*15*, *20*–*23*). Cardiac pulses are dampened by a Windkessel mechanism, which accommodates a pulse with capacitive vessel expansion (Fig. 3C), but arteries cannot expand within a fixed-volume compartment like the cranium unless an equal volume of

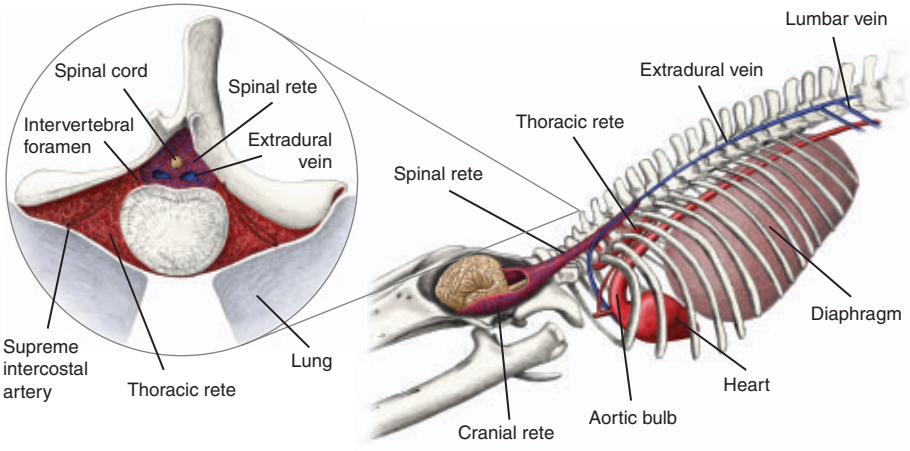



**Fig. 1. Principal vascular components supplying the spinal cord and brain in *B. physalus*.** Blood flow from the aorta to the brain passes through the thoracic, spinal, and cranial retia mirabilia. The thoracic and spinal retia connect between adjacent vertebrae, shown in the inset [from (*4*)]. Two lumbar veins illustrate connections between abdominal venous beds (not shown) and extradural veins in the vertebral canal. Components are simplified and not drawn to scale. Copyright 2019 by Alex Boersma (used with permission).

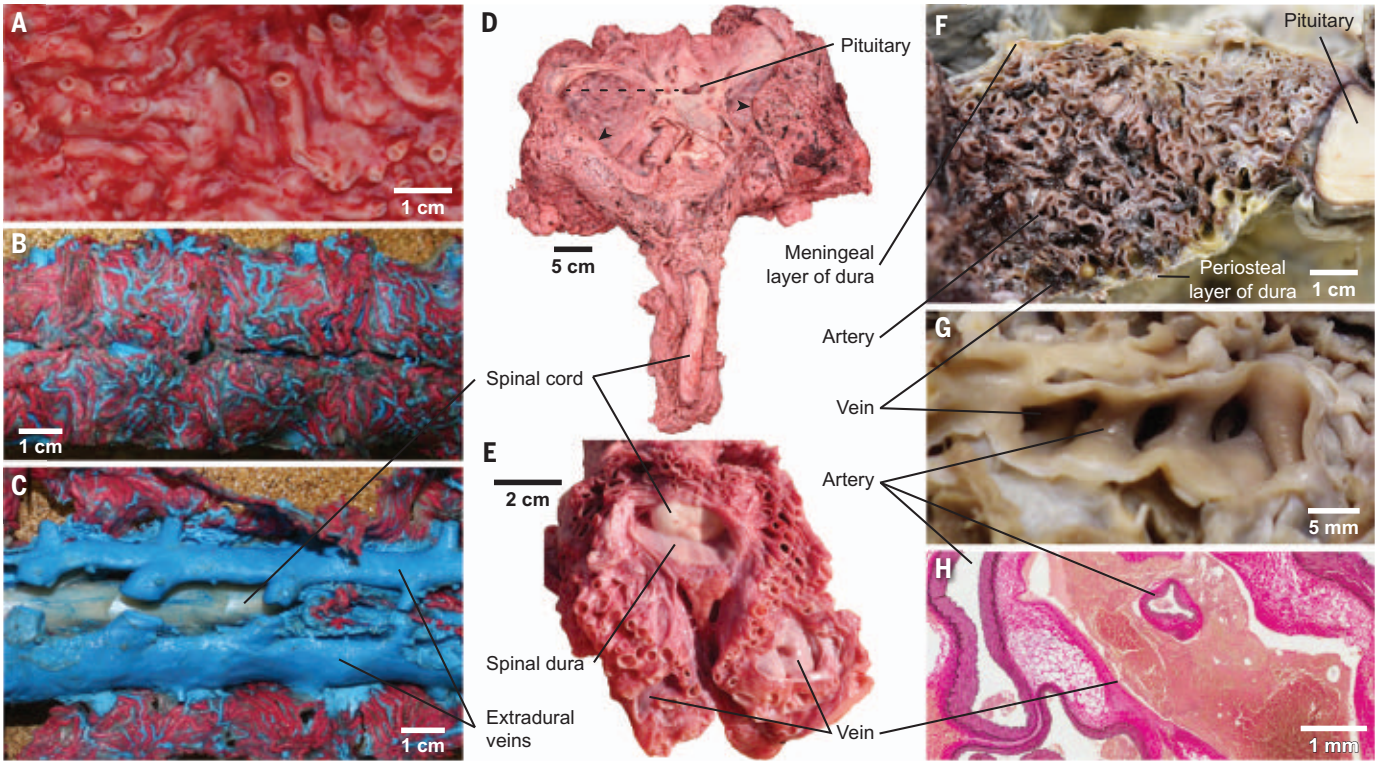



**Fig. 2. Retial morphology in *B. physalus* and *T. truncatus*.** Images are from *B. physalus* [(A) and (D) to (H)] and *T. truncatus* [(B) and (C)]. (**A**) Thoracic rete composed of arteries. (**B**) Ventral view of latex-injected spinal rete showing intimate juxtaposition of arteries (red) and veins (blue). (**C**) As in (B) but cut at the midline to expose extradural veins with midline anastomoses (cut). (**D**) Dorsal view of spinocranial rete. Arrowheads show edge of walls flopped down (see Fig. 1). Dashed line shows location of cut for (F). (**E**) Parallel intercommunicating arteries and veins in spinal rete near the entrance to the skull. (**F**) Transverse cut through cranial rete shows thick layer of arteries and blood-filled veins. (**G**) Close-up of three arteries running through venous structure cut from (F). (**H**) Histology from (F) showing artery surrounded by venous blood (Van Gieson's stain). (B) and (C) are from (*17*), used with permission.

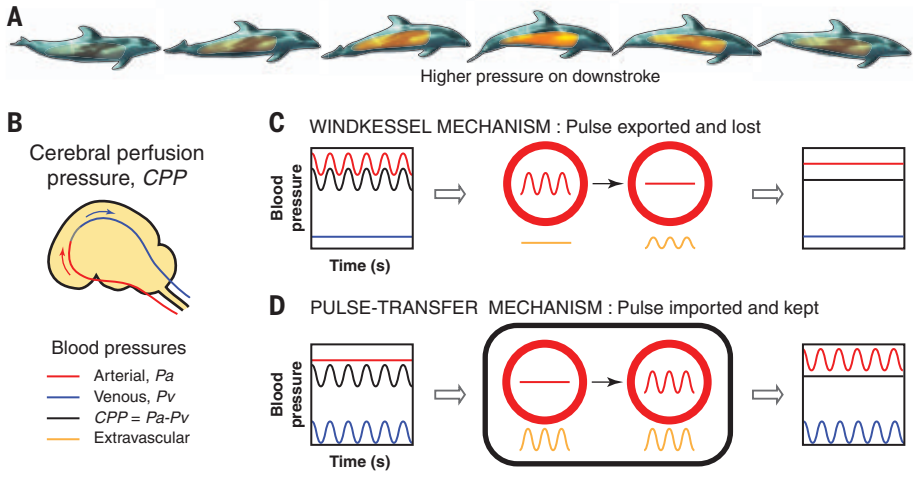



**Fig. 3. Mechanisms to reduce fluking-generated pulsatility in cerebral flow.** (**A**) Proposed widespread pressurization of extravascular and vascular tissue in the abdomen and thorax on each fluke downstroke. Pulses are rapidly transmitted through the retia into the cerebral vasculature (**B**). Pulse amplitude entering cerebral beds depends on amplitudes entering the retia and retial hemodynamics. CPP is pulsatile when arterial and venous pulses have different amplitudes. (**C** and **D**) Retia could reduce pulsatility in CPP using a Windkessel mechanism to dampen an arterial pulse (C) or a pulse-transfer mechanism to generate an arterial pulse from an extravascular pulse (D). Thick black line represents constraining bony wall.

CSF or venous blood is expelled to compensate. Such outflow increases the virtual intracranial compliance (*23*, *24*) and keeps pressure amplitudes low. In general, unobstructed outflow of CSF and venous blood keeps the arterial component of the CSF pulse small, and amplitudes increase when outflow is obstructed (*25*, *26*). However, in cetaceans, pulses occupy the vertebral canal as well as the cranium, and because CSF is constrained to the spinocranial space, CSF outflow cannot moderate pulse amplitude. Additionally, fluking generates venous and arterial pulses. Resistance to arterial outflow is higher than that to venous, which makes it harder to accommodate a venous pulse and keep amplitudes low.

Fluking-generated systemic arterial and venous pulses reach the spinocranial rete by different routes, so their amplitudes will differ. We hypothesize that rather than dampen these pulses, the retia maintain and spread them within the vertebral canal and cranium, creating arterial and venous pulses matched in amplitude and phase. In this work, we propose a pressure pulse-transfer mechanism, in which the instantaneous and quantitative transmission of pulsatile pressures among arteries ($P_a$), veins ($P_v$), and CSF prevents pulsatility in the cerebral perfusion pressure (CPP), defined here as $P_a - P_v$ (Fig. 3B and materials and methods). CPP drives longitudinal flow through the cerebral beds, and when venous pulses are quantitatively transmitted into the CSF, it also defines the arterial transmural pressure that determines contained blood volumes. When $P_a$ and $P_v$ rise equally and synchronously, CPP is not pulsatile (Fig. 3D). The retia's effectiveness in controlling pulsatility is expressed as ∿CPP, the fraction of systemic pulsatility that is transmitted through the retia into the CPP. The sine wave symbol (∿) signifies the pulsatile component only. The more effective the rete, the lower the fraction and the more constant the CPP. Where the pulse-transfer mechanism works perfectly, mean CPP is unaltered, and ∿CPP = 0.

In the pulse-transfer mechanism, to promote transmission of venous pulses into retial arteries, compensating arterial outflow is specifically hindered. For the spinocranial rete, outflow rate depends on the time constant $\tau = R_{sc}C_{sc}$, where $R_{sc}$ is the outflow resistance and $C_{sc}$ is the capacitance of the retial arteries. Increasing $\tau$ slows outflow. To transfer pulse amplitude, the rete benefits from a long $R_{sc}C_{sc}$ outflow time and a short pulse duration—i.e., a fast fluking frequency $f$. Therefore, minimizing ∿CPP requires maximizing $f \times R_{sc}C_{sc}$. $R_{sc}$ represents the resistance to flow between the heart and brain, which cannot be increased without reducing cerebral flow. $R_{sc}$ is lower in larger animals because of their greater cerebral flow. Similarly, fluking frequency cannot be increased because it is inversely proportional to body length (*27*). However, $C_{sc}$ can be modulated by increasing retial bed volume. Larger animals could have relatively larger retial beds to compensate for their lower $R_{sc}$ and lower fluking frequency. If cetaceans use a pulse-transfer mechanism in the spinocranial rete, we predict first that the product $f \times R_{sc}C_{sc}$ will be the same for all species, which requires that the outflow time in each species is inversely proportional to their fluking frequency, $R_{sc}C_{sc} \propto f^{-1}$. Second, we predict that the level of ∿CPP will be low for all species.

To test this hypothesis, we developed a computational model on the basis of observed morphological and mechanical parameters to predict the hemodynamic response to fluking-generated arterial, venous, and extravascular abdominal pressures (fig. S1 and materials and methods). We tested the pulse-transfer mechanism for the spinocranial rete by plotting the observed $R_{sc}C_{sc}$ values in 11 species against fluking frequency (Fig. 4A). The red line shows the best fit to the data from a log-log analysis of these points (fig. S2A). The 95% confidence interval for the line exponent (−1.38, −0.64) includes the value of −1 predicted for the pulse-transfer mechanism, which indicates that retial capacitance is modulated according to fluking frequency and cerebral flow. The data points overlie a set of contour lines showing the predicted ∿CPP. ∿CPP is lower at higher $R_{sc}C_{sc}$ values and faster fluking frequencies. From the clustering of the observed points around the 0.1 and 0.2 contour lines, the simulation predicts that the spinocranial rete transmits only ~15% of a systemic pulse into the CPP in all species. A reduction to 15% would sufficiently protect the cerebral beds when systemic pulses are small, but locomotion can generate large pressures. Internal pressures have not been measured in fluking cetaceans, but abdominal pressure swings of 10 to 16 kPa have been measured in galloping horses with an open respiratory system (*8*), and humans performing a Valsalva maneuver (forced expiration against a closed glottis) can raise internal and systemic arterial pressures by 30 kPa (*28*). We concluded that the spinocranial rete alone would be inadequate to deal with pulses of this magnitude, so we incorporated the more proximal thoracic rete into our simulations.

We hypothesized that the thoracic rete preconditions systemic arterial pulses to more closely match venous pulses arriving at the spinocranial rete. Consistent with this hypothesis, the time constant for the thoracic rete, $R_{thor}C_{thor}$, decreases with fluking frequency (Fig. 4B and fig. S2B), and the 95% confidence interval for the exponent (−1.95, −0.68) includes the predicted value of −1. We first considered whether the thoracic rete could use a Windkessel mechanism to reduce fluking-generated pulses in ∿CPP but found it counterproductive. Nagel and colleagues (*29*) have demonstrated that the thoracic and spinocranial retia can function as a Windkessel in anaesthetized dolphins and have speculated about whether dampening cardiac pulses represented the retia's primary role. However,

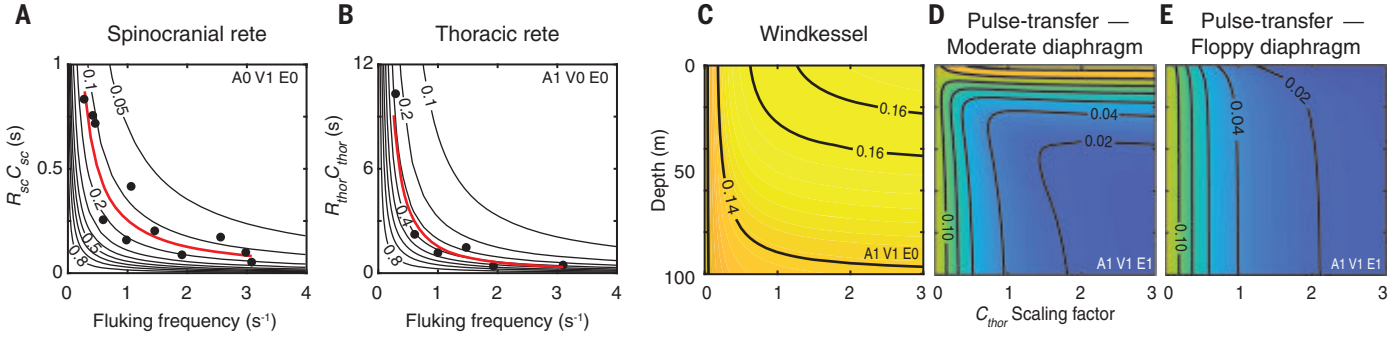



**Fig. 4. Simulated reduction of ∿CPP by the retia in a generic cetacean.** ∿CPP is the fraction of systemic pulsatility transmitted through the retia into the CPP. Input of a simulated arterial, venous, or extravascular pulse is indicated as A1, V1, or E1, respectively, and no input is indicated as A0, V0, or E0. (**A**) Reduction of ∿CPP (contour lines) with greater spinocranial rete time constants ($R_{sc}C_{sc}$) and fluking frequency. Thoracic rete excluded from simulation. Red line shows best fit using log-log analysis of mean observed values (circles): $R_{sc}C_{sc}$ = $0.26f^{-1.01}$ (fig. S2A). (**B**) As in (A) but for thoracic rete. $R_{thor}C_{thor}$ is the thoracic rete time constant. Spinocranial rete and aortic bulb excluded from simulation. Best-fit line: $R_{thor}C_{thor} = 1.54f^{-1.32}$ (fig. S2B). (**C** to **E**) Impact of thoracic rete capacitace (*x* axis) and depth on ∿CPP (contour lines and color). $C_{thor}$ scaling factor = 1 equals the observed $C_{thor}$ value. Spinocranial rete, thoracic rete, and aortic bulb included in simulations. (C) Aortic bulb dominates any Windkessel response of thoracic rete, preventing reduction of ∿CPP. (D and E) Thoracic rete reduces ∿CPP using pulse-transfer mechanism. Losses resulting from lung dampening occur at the surface with a moderately stiff diaphragm (D) but not with a floppy diaphragm (E).

where fluking-generated systemic arterial and venous pulses are already largely matched, damping the arterial pulse alone would create a mismatch. The aortic bulb in the cetacean thorax is a major Windkessel that dampens cardiac pulses (Fig. 1). Simulations show the bulb would also have this counterproductive impact on ∿CPP by dampening fluking-generated arterial pulses (Fig. 4C and fig. S3, A and B). In Fig. 4, C to E, the yellow contours represent ∿CPP = 0.15, the spinocranial rete contribution, and any beneficial, further reduction of ∿CPP from increasing thoracic rete capacitance (*x* axis) would shift the response toward bluer contours. This does not happen in Fig. 4C: The bulb dominates the response, and thoracic rete capacitance has no effect. This simulation reveals a basic problem that cetaceans face—the aortic bulb is thought necessary to maintain flow over the long diastoles of the diving response, but the bulb would dampen both cardiac and fluking-generated arterial pulses, causing matched fluking-generated systemic pulses to arrive at the spinocranial rete unmatched. The thoracic rete, therefore, may have developed to rematch the pulses.

We suggest that the thoracic rete provides a pulse-transfer mechanism that transfers fluking-generated extravascular pressure pulses, originating in the abdomen, into the thoracic retial arteries. Abdominal pressures push the diaphragm forward and upward into the lungs and, in turn, against the thoracic rete (Fig. 1, inset). The observed thoracic rete capacitance (i.e., scaling factor = 1) appears optimized to transfer an extravascular pulse into an arterial pulse (Fig. 4D and fig. S3, C and D). Some abdominal pulsatile energy will be lost before reaching the thoracic rete—either dampened by the lungs near the surface (materials and methods) or simply constrained by the diaphragm. However, such losses can be compensated for by relaxing the diaphragm to increase transmission of abdominal pressures into the thorax (Fig. 4E). With sufficiently reduced diaphragm stiffness, the thoracic rete can capture and transmit extravascular pulses into the retial arteries to largely match the venous pulses entering the spinocranial rete. Linking an arterial pulse-generating function of the thoracic rete to the arterial pulse-dampening action of the bulb explains the nearly exclusive arterial composition of the thoracic rete. The combined effect of the thoracic and spinocranial retia, when both use the pulse-transfer mechanism, can reduce ∿CPP to ~3%, with a benefit available at all depths.

Our study shows that thoracic and spinocranial retia can protect cetacean cerebral microvasculature from fluking-generated pressure pulses using a pulse-transfer mechanism. Such a mechanism ensures that equal arterial and venous pulses arrive simultaneously to the brain without dampening the pulses themselves. Previous pressure-based hypotheses had identified neither the source of damaging pressures nor the protective mechanism (*3*, *4*, *29*, *30*). Although retia have an arterial Windkessel ability (*29*), which our model corroborated (fig. S4), our analyses showed that the Windkessel mechanism was counterproductive to protecting cerebral beds from fluking-generated pulses. All animals deal with vascular pulsatility, and cardiorespiratory processes are generally coupled to locomotion to manage internal pressures. But in cetaceans, the combination of locomotion with dorsoventral fluke oscillation while breath-holding has created an extraordinary set of hemodynamic demands that appear to be met by the evolution of two morphologically distinct retia. Our results link the evolution of cetacean retia mirabilia to dorsoventral fluking during breath-hold diving and explain why retia are not present in other aquatic vertebrates that have different modes of locomotion.

## REFERENCES AND NOTES


1. P. J. Ponganis, *Diving Physiology of Marine Mammals and Seabirds* (Cambridge Univ. Press, 2015).
2. F. D. Ommanney, *Discov. Rep.* **5**, 327–362 (1932).
3. W. L. Mcfarland, M. S. Jacobs, P. J. Morgane, *Neurosci. Biobehav. Rev.* **3**, 1–93 (1979).
4. R. Walmsley, *Contrib. Embryol.* **27**, 107–178 (1938).
5. R. L. Kirby, S. T. Nugent, R. W. Marlow, D. A. MacLeod, A. E. Marble, *J. Appl. Physiol.* **66**, 323–329 (1989).
6. M. A. Daley, D. M. Bramble, D. R. Carrier, *PLOS ONE* **8**, e70752 (2013).
7. I. S. Young, R. Alexander, A. J. Woakes, P. J. Butler, L. Anderson, *J. Exp. Biol.* **166**, 19–31 (1992).
8. R. Slocombe, K. Brock, G. Covelli, W. Bayly, in *Equine Exercise Physiology 3*, S. G. Persson, L. B. Jeffcott, Eds. (ICEEP Publications, 1991), pp. 83–91.
9. F. E. Fish, *Am. Zool.* **36**, 628–641 (1996).
10. M. Bonato *et al.*, *J. Exp. Biol.* **222**, jeb198457 (2019).
11. P. B. Cotten *et al.*, *J. Morphol.* **269**, 1520–1538 (2008).
12. M. A. Lillie *et al.*, *J. Exp. Biol.* **220**, 3464–3477 (2017).
13. A. D. Guerci *et al.*, *Circ. Res.* **56**, 20–30 (1985).
14. J. E. Hall, M. E. Hall, *Guyton and Hall Textbook of Medical Physiology* (Elsevier, 2010).
15. G. A. Bateman, C. R. Levi, P. Schofield, Y. Wang, E. C. Lovett, *Neuroradiology* **50**, 491–497 (2008).
16. J. Stone, D. M. Johnstone, J. Mitrofanis, M. O'Rourke, *J. Alzheimers Dis.* **44**, 355–373 (2015).
17. A. Costidis, "The morphology of the venous system in the head and neck of the bottlenose dolphin (*Tursiops truncatus*) and the Florida manatee (*Trichechus manatus latirostris*)," thesis, University of Florida, Gainesville, FL (2012).
18. A. W. Vogl, H. D. Fisher, *J. Morphol.* **174**, 41–56 (1982).
19. C. E. Rowlands *et al.*, *J. Morphol.* **282**, 1415–1431 (2021).
20. C. J. Avezaat, J. H. van Eijndhoven, *Neurosurg. Rev.* **9**, 113–120 (1986).
21. M. Ursino, *Ann. Biomed. Eng.* **16**, 379–401 (1988).
22. M. Egnor, A. Rosiello, L. Zheng, *Pediatr. Neurosurg.* **35**, 284–298 (2001).
23. M. E. Wagshul, P. K. Eide, J. R. Madsen, *Fluids Barriers CNS* **8**, 5 (2011).
24. M. Zamir, M. E. Moir, S. A. Klassen, C. S. Balestrini, J. K. Shoemaker, *Front. Physiol.* **9**, 940 (2018).
25. J. Hamer, E. Alberti, S. Hoyer, K. Wiedemann, *J. Neurosurg.* **46**, 36–45 (1977).
26. R. A. Bhadelia, A. R. Bogdan, R. F. Kaplan, S. M. Wolpert, *Neuroradiology* **39**, 258–264 (1997).
27. K. Sato *et al.*, *Proc. R. Soc. B* **274**, 471–477 (2007).

28. D. Blazek *et al*., *Biol. Sport* **36**, 373–386 (2019).
29. E. L. Nagel, P. J. Morgane, W. L. McFarland, R. E. Galliano, *Science* **161**, 898–900 (1968).
30. E. J. Slijper, *Whales: The Biology of the Cetaceans* (Basic Books, 1962).

## ACKNOWLEDGMENTS

We are grateful to A. Costidis and C. Rowlands (Virginia Aquarium and Marine Science Center Foundation) and D. A. Pabst, W. A. McLellan, and S. Rommel (University of North Carolina at Wilmington) for ongoing discussions that contributed substantially to this study. We thank P. Cottrell, Marine Mammal Coordinator, Pacific Region Fisheries and Oceans Canada; K. Loftsson and the staff at Hvalur hf, Iceland; and D. Halldórsson at the Marine and Freshwater Research Institute, Reykjavík, Iceland. The post mortem room staff of the Animal Health Centre, Abbotsford, Canada, are gratefully acknowledged for assistance with transection and archiving of samples. **Funding:** This work was supported by Natural Sciences and Engineering Research Council of Canada discovery grant RGPIN-2019-04235 (to R.E.S.) and Natural Sciences and Engineering Research Council of Canada discovery grant RGPIN-2018-03727 (to A.W.V.). **Author contributions:** Conceptualization: M.A.L., A.W.V., and R.E.S. Investigation: M.A.L., A.W.V., S.G.G., and S.R. Resources: S.R. and R.E.S. Software: M.A.L. Funding acquisition: R.E.S. and A.W.V. Writing – original draft: M.A.L. Writing – review & editing: M.A.L., A.W.V., S.G.G., S.R., and R.E.S. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. 

## SUPPLEMENTARY MATERIALS



Submitted 26 November 2021; accepted 5 August 2022
10.1126/science.abn3315

PURDUE UNIVERSITY | Department of Industrial and Physical Pharmacy

## FACULTY POSITION IN PHARMACEUTICAL BIOTECHNOLOGY
### Assistant or Associate Professor

The Department of Industrial and Physical Pharmacy at Purdue University is seeking applications to fill a faculty position for a tenure-track/tenured Assistant Professor or Associate Professor in the general area of Pharmaceutical Biotechnology. The ideal candidate will establish an externally funded research program and will teach courses in the undergraduate, graduate, and professional (PharmD) programs. It is expected that the individual will collaborate on various research activities within the department and in multidisciplinary research throughout the University. The candidate must have expertise in fundamental and applied research related to the development, design, evaluation, and manufacturing of biopharmaceutical products. Representative areas of expertise include, but are not limited to the formulation-manufacturing interface (e.g., stability, analysis, and/or manufacturing) of peptide, protein, cell-based and/or nucleic acid-based therapeutics, vaccines, and/or immunotherapeutics.

Established in 1869, Purdue is Indiana's land-grant university, a comprehensive educational and research institution which is a member of the prestigious American Association of Universities (AAU). The Department of Industrial and Physical Pharmacy encompasses multidisciplinary research and teaching in the areas of drug development, biotechnology, pharmacokinetics, drug delivery systems and manufacturing technologies.

Purdue University offers a thriving interdisciplinary environment for education and research. The university has a student body of over 49,000 undergraduate, graduate, and professional students, including 11 percent of underrepresented minority students and 18 percent international students.

**Qualifications:** The candidate must hold a PhD degree or equivalent in a relevant discipline such as pharmaceutical sciences, engineering, biochemistry, analytical chemistry and biophysics or a closely related field. The candidate will have an outstanding record of scientific and technical accomplishment, an exceptional potential for world-class research and a commitment to both undergraduate and graduate education. Purdue University's Department of Industrial and Physical Pharmacy is committed to advancing diversity in all areas of faculty effort including scholarship, instruction, and engagement. Candidates should address at least one of these areas in a separate Diversity and Inclusion Statement, indicating their past experiences, current interests or activities, and/or future goals to promote a climate that values diversity and inclusion.

**Application:** For consideration, please upload a single PDF file that contains the following at https://careers.purdue.edu/job-invite/21594/

- Candidate's curriculum vitae
- Letter of interest providing a summary of qualifications for the position
- Statement of research interests and goals
- Statement of teaching philosophy
- Statement of diversity and inclusion
- Names and contact information for three references

Incomplete applications will not be considered. Review of applications will begin September 1, 2022 and will continue until the position is filled. A background check will be required for employment in this position. The desired starting date is August 14, 2023, but is flexible and negotiable.

For further information concerning the search please contact Mary Ellen Hurt, Faculty Search Committee, Department of Industrial and Physical Pharmacy, College of Pharmacy, Purdue University, West Lafayette, IN 47907-2091 **mhurt@purdue.edu.**

*Purdue University is an EOE/AA employer. Purdue University is committed to maintaining a community which recognizes and values the inherent worth and dignity of every person. In pursuit of its goal of academic excellence, the University seeks to develop and nurture diversity. All qualified applicants for employment will receive consideration without regard to race, religion, color, sex, national origin or ancestry, genetic information, marital status, parental status, sexual orientation, gender identity and expression, disability or status as a veteran.*

---

PURDUE UNIVERSITY | Department of Industrial and Physical Pharmacy

## FACULTY POSITION IN ADVANCED MANUFACTURING
### Assistant or Associate Professor

The Department of Industrial and Physical Pharmacy at Purdue University is seeking applications to fill a faculty position for a tenure-track/tenured Assistant Professor or Associate Professor in the general area of Advanced Manufacturing. The ideal candidate will establish an externally funded research program and will teach courses in the undergraduate, graduate, and professional (PharmD) programs. It is expected that the individual will collaborate on various research activities within the department and in multidisciplinary research throughout the University. The candidate must have expertise in fundamental and applied research related to advanced manufacturing. Representative areas of expertise include, but are not limited to advanced manufacturing approaches using novel technologies such as, but not limited to, continuous manufacturing, drying technology, aseptic processing, and 3D printing.

Established in 1869, Purdue is Indiana's land-grant university, a comprehensive educational and research institution which is a member of the prestigious American Association of Universities (AAU). The Department of Industrial and Physical Pharmacy encompasses multidisciplinary research and teaching in the areas of drug development, biotechnology, pharmacokinetics, drug delivery systems and manufacturing technologies.

Purdue University offers a thriving interdisciplinary environment for education and research. The university has a student body of over 49,000 undergraduate, graduate, and professional students, including 11 percent of underrepresented minority students and 18 percent international students.

**Qualifications:** The candidate must hold a PhD degree or equivalent in a relevant discipline such as pharmaceutical sciences, engineering, biochemistry, analytical chemistry and biophysics or a closely related field. The candidate will have an outstanding record of scientific and technical accomplishment, an exceptional potential for world-class research and a commitment to both undergraduate and graduate education. Purdue University's Department of Industrial and Physical Pharmacy is committed to advancing diversity in all areas of faculty effort including scholarship, instruction, and engagement. Candidates should address at least one of these areas in a separate Diversity and Inclusion Statement, indicating their past experiences, current interests or activities, and/or future goals to promote a climate that values diversity and inclusion.

**Application:** For consideration, please upload a single PDF file that contains the following at **https://careers.purdue.edu/job-invite/21768/**

- Candidate's curriculum vitae
- Letter of interest providing a summary of qualifications for the position
- Statement of research interests and goals
- Statement of teaching philosophy
- Statement of diversity and inclusion
- Names and contact information for three references

Incomplete applications will not be considered. Review of applications will begin September 1, 2022 and will continue until the position is filled. A background check will be required for employment in this position. The desired starting date is August 14, 2023, but is flexible and negotiable.

For further information concerning the search please contact Mary Ellen Hurt, Faculty Search Committee, Department of Industrial and Physical Pharmacy, College of Pharmacy, Purdue University, West Lafayette, IN 47907-2091 **mhurt@purdue.edu.**

*Purdue University is an EOE/AA employer. Purdue University is committed to maintaining a community which recognizes and values the inherent worth and dignity of every person. In pursuit of its goal of academic excellence, the University seeks to develop and nurture diversity. All qualified applicants for employment will receive consideration without regard to race, religion, color, sex, national origin or ancestry, genetic information, marital status, parental status, sexual orientation, gender identity and expression, disability or status as a veteran.*

*By* **Rasha Shraim**

# Out of isolation

This summer, I walked into the lecture hall at the first in-person conference I'd attended during my Ph.D. I expected to feel bored and distracted, as I had at many Zoom presentations during the previous 2 years. But as the session proceeded, I was surprised to feel curious and inspired. With the presenters standing right in front of me, I stayed engaged, taking in information about exciting new findings and methods. Suddenly, I felt something else new: an urge to dive back into my own research. I realized how much I'd been struggling during the pandemic—and what changes I needed to make.

When I started graduate school in 2020, in the midst of the COVID-19 pandemic, I set up my primary office at home, where I could easily work on my computational biology research. After COVID-19 restrictions eased, I began spending 1 day per week on campus. But I resisted going in more than that because my external monitor, books, and sticky notes were at home and it was easier to work from there. As time went on, though, the convenience wasn't enough. I began to feel increasingly uninspired about my work, although I wasn't sure why that was happening or what to do about it.

**"I could feel my enthusiasm for science returning."**

The conference, which brought together Ph.D. students from across Ireland, drew me out of isolation, allowing me to interact with other researchers for days at a time. For the first time since I started my Ph.D., I had the opportunity to talk at length with other students about our research, careers, and other shared experiences as we wandered around campus and went out for meals. I could feel my enthusiasm for science returning.

That feeling continued to grow later in the week when it was my turn to present. Ahead of my talk, I experienced the usual nerves, especially because I had been regularly underperforming at online presentations. I braced myself for the worst, but I ended up delivering a presentation I could be proud of, speaking with more rhythm and confidence than during any other presentation I'd given throughout my Ph.D. It helped to be able to move around and see other people's faces and encouraging reactions, rather than sitting and watching my own face on a computer screen.

Afterward, other students expressed interest in my research and inquired about the methods. One even suggested I look into a variable I hadn't considered. This gave me confidence that my research is interesting and important after all—something I had lost sight of after spending months largely alone, focused on the tedious steps necessary to meet the next deadline.

I left the conference feeling excited about my Ph.D., and I wanted that feeling to continue. I decided that working from home was part of my problem and that I needed more informal conversations during my workdays. If I was stuck on a problem or looking for a new direction, I wanted to be able to pop in to see one of my supervisors or bounce ideas off another Ph.D. student.

So, soon after the conference, I moved my entire work set up to my desk on campus. I now interact with other students regularly while working in our shared office, crossing paths in the hallways, or eating lunch. These little interactions are not only good for my mental health and well-being, but they are also helping me move my research forward more rapidly and think about what I want to do after I graduate. For instance, one conversation alerted me to a new method I could use in my research. Another conversation, which included senior Ph.D. students, helped me think through different career paths and how internships might help me choose one. I've spent my career so far in academia, but some of the other students had industry experience and I found it helpful to hear their perspective.

Although hybrid events and flexible work options are beneficial for many reasons, my in-person conference experience helped me figure out that working almost entirely from home wasn't right for me. I'm hopeful I'll be able to carry my newfound excitement forward in my research now that I'm surrounded by researchers again. ■

Rasha Shraim is a Ph.D. student at Trinity College Dublin. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

# *You're only an essay away!*

## Check out our Awards and Prizes that are currently seeking submissions.

**Immunology**
Submissions Opened: May 1, 2022
Submissions Close: Oct 1, 2022

---

**Scientific Innovation & Entrepreneurship**
Submissions Opened: Jun 1, 2022
Submissions Close: Nov 1, 2022

---

**NOSTER & Science Microbiome Prize**
Submissions to Open: Sep 15, 2022
Submissions Close: Jan 24, 2023

---

**Science & PINS Prize for Neuromodulation**
Submissions to Open: Nov 1, 2022
Submissions Close: Mar 15, 2023

*Science* Awards and Prizes seeks to find the next generation of early-career researchers on the cusp of changing the scientific landscape. Start your journey by finding the Award or Prize that fits you and find out how to apply today!

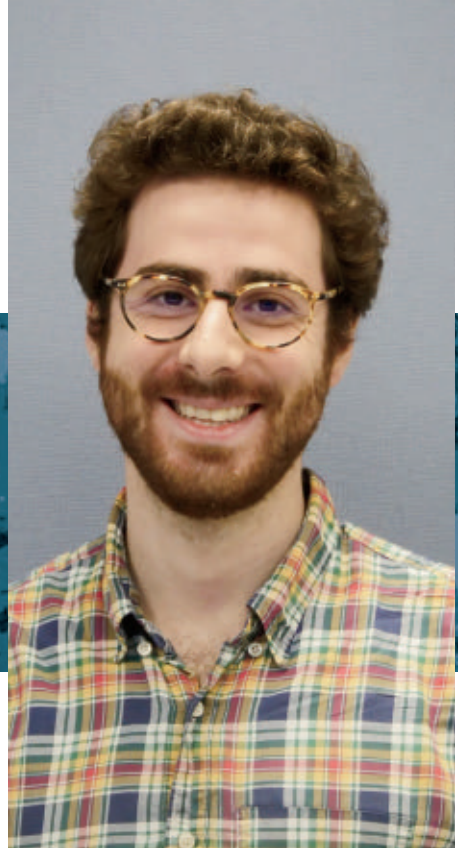
Science
AAAS

bit.ly/3qKxMN0

**Scan the code to start your application process!**

MICHELSON PHILANTHROPIES & Science PRIZE FOR IMMUNOLOGY

# REWARDING HIGH-RISK RESEARCH. SUPPORTING EARLY-CAREER SCIENTISTS. HELPING TO FIND CURES FASTER. **APPLY TODAY**



**Now accepting applications for the Michelson Philanthropies & *Science* Prize for Immunology.**

The Michelson Philanthropies and Science Prize for Immunology focuses on transformative research in human immunology, with trans-disease applications to accelerate vaccine and immunotherapeutic discovery. This international prize supports investigators 35 and younger, who apply their expertise to research that has a lasting impact on vaccine development and immunotherapy. It is open to researchers from a wide range of disciplines including computer science, artificial intelligence/machine learning, protein engineering, nanotechnology, genomics, parasitology and tropical medicine, neurodegenerative diseases, and gene editing.

**Application deadline: Oct. 1, 2022.**

For more information visit:
www.michelsonmedicalresearch.org

**#MichelsonPrizes**

*"The Michelson Philanthropies & Science Prize for Immunology will greatly impact my future work. As I am just starting my scientific career, it will illuminate my work, spark interest and support me to continue my research in this field."*

**Paul Bastard, MD, PhD,**
Laboratory of Human Genetics of Infectious Diseases, Imagine Institute (INSERM, University of Paris), Paris, France; and The Rockefeller University, New York.

Dr. Bastard received the inaugural Grand Prize for his essay: "Why do people die from COVID-19: Autoantibodies neutralizing type I interferons increase with age."

**GRAND PRIZE:**
**$30,000**

**FINALIST PRIZE:**
**$10,000**