Tomorrow's Earth
pp. 886, 889, & 908

The microbiome influences metabolic disease *p. 903*

A topological insulator goes superconducting *pp. 922 & 926*

Science

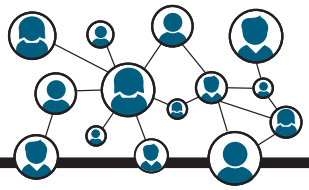
$15
23 NOVEMBER 2018
sciencemag.org

AAAS

EXPOSING THE CAMBRIAN

New light on an ancient explosion of animal forms *p. 880*

CONTENTS



886, 889, & 908
Social basis of conservation

23 NOVEMBER 2018 • VOLUME 362 • ISSUE 6417

880
Fossil treasures from the Canadian Rockies

NEWS

IN BRIEF

IN DEPTH

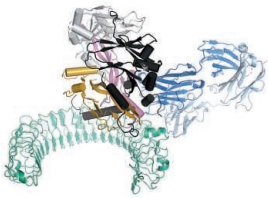
FEATURE

INSIGHTS

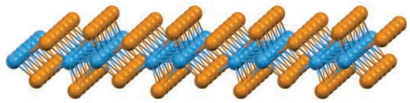
LETTERS

PERSPECTIVES

Published by AAAS



952
Molecular underpinnings of immunosuppression



922 & 926
The versatile tungsten ditelluride

ON THE COVER

A paleontologist in Canada's Kootenay National Park uses a rock saw to extract slabs bearing Cambrian fossils. In 1909, 40 kilometers up this valley, paleontologists discovered the Burgess Shale, a 500-million-year-old formation that records animals' soft features, such as eyes and guts. Today, paleontologists have shown that this type of preservation extends across many kilometers, opening up new vistas of the dawn of modern animals. See page 880.
Photo: John Lehmann

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. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1808; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery): $89. 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; 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 is granted by AAAS to libraries and others who use Copyright Clearance Center (CCC) Pay-Per-Use services provided that $35.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Science
AAAS
1200 New York Avenue, NW, Washington, DC 20005
Clarendon House, Clarendon Road, Cambridge, UK CB2 8FH

SUBSCRIPTION SERVICES For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

INSTITUTIONAL SITE LICENSES 202-326-6730 **REPRINTS:** Author Inquiries 800-635-7181 **COMMERCIAL INQUIRIES** 803-359-4578 **PERMISSIONS** 202-326-6765, permissions@aaas.org **AAAS Member Central Support** 866-434-2227 www.aaas.org/membercentral.

INFORMATION FOR AUTHORS See www.sciencemag.org/authors/science-information-authors

Published by AAAS

EDITORIAL

Fixing the internet

Data breaches at Facebook and Google—and along with Amazon, those firms' online dominance—crest a growing wave of anxiety around the internet's evolving structure and its impact on humanity. Three keys to the decades-long global expansion of the internet and the World Wide Web are breaking down.

The first key is the "procrastination principle," a propensity to "set it and forget it" without attempting to predict and avert every imaginable problem. The networks' framers established a set of simple and freely available protocols for communicating over the internet, then stepped back to let competitive markets and cooperative pursuits work their magic.

The second key is the networks' layered architecture. For the internet, this meant that people could concern themselves with, say, writing applications to read and send email without having to know anything about what happens "below," such as how bits find their way from sender to recipient. By the same token, those rolling out physical infrastructure didn't need to know or predict anything about how it would be used by the applications "above."

The third key flows from the first two: decentralization. The internet and the web were designed not to create new gatekeepers, in part because regulatory bodies had little awareness of these protocols, let alone a hand in structuring them. A website hosted in Romania would still be just a click away for a user in Canada, without authorization by some centralized party.

Today, the principles of layers and decentralization are badly fraying, which risks transforming the principle of procrastination into one of abdication.

First, the issue of centralization. Surfing the web can now mean simply jumping among Amazon Web Services' hosting servers. If such a major network of servers—or one of the top domain name resolution providers—were to stop working, whole swaths of the internet would go down with it.

Second, formerly separate layers of the internet's architecture are blurring. The runaway success of a few startups has created new, proprietized one-stop platforms. Many people are not really using the web at all, but rather flitting among a small handful of totalizing apps like Facebook and Google. And those application-layer providers have dabbled in providing physical-layer internet access. Facebook's Free Basics program has been one of several experiments that use broadband data cap exceptions to promote some sites and services over others.

What to do? Columbia University law professor Tim Wu has called upon regulators to break up giants like Facebook, but more subtle interventions should be tried first. Web inventor Tim Berners-Lee's Contract for the Web offers a set of principles for governments, companies, and individuals, focusing on internet accessibility, user privacy, and a form of "re-decentralization" to revitalize one key to the network's success. On the technical side, he has launched Solid, a "re-layerizing" separation of data from application: Users can maintain their own data (whether in a server in their living room or in the hands of a trusted proxy), and application providers would have to negotiate access rather than hoard the data themselves. And as Yale University law professor Jack Balkin and I have argued, those firms that do leverage users' data should be "information fiduciaries," obliged to use what they learn in ways that reflect a loyalty to users' interests. These interventions represent meaningful action, while procrastinating a bit longer on the stronger medicine of forced corporate breakup.

The internet was designed to be resilient and flexible, without need for drastic intervention. But its trends toward centralization, and exploitation of its users, call for action.

–Jonathan Zittrain

Jonathan Zittrain
(@zittrain) is the George Bemis Professor of International Law and Professor of Computer Science at Harvard University, Cambridge, MA, USA, and is cofounder of Harvard's Berkman Klein Center for Internet & Society. zittrain@law.harvard.edu

10.1126/science.aaw0798

NEWS

> “I will not allow a generation of children to become addicted to nicotine through e-cigarettes.”
>
> **U.S. Food and Drug Administration Commissioner Scott Gottlieb**, in a statement announcing plans to tighten restrictions on the sale of flavored e-cigarettes, among other measures.

IN BRIEF

Edited by **Kelly Servick**

INFECTIOUS DISEASE

Old drug, new hope for sleeping sickness

Health workers screen for human African trypanosomiasis in the Democratic Republic of the Congo.

An improved treatment for human African trypanosomiasis, also known as sleeping sickness, could help thousands of patients in West and Central Africa. The disease, caused by a parasite transmitted by the tsetse fly, disrupts sleep patterns and causes aggression, psychosis, and, ultimately, death. Ten years ago, the main treatment was an arsenic-based drug that killed 5% of those who took it. Current treatments must be administered in a hospital, making them inaccessible to most people in countries where the disease is most common: the Democratic Republic of the Congo, the Central African Republic, Guinea, and Chad. But fexinidazole, an antiparasitic drug first developed in the 1980s, is effective against the disease when taken as a pill daily for 10 days. Researchers at the Drugs for Neglected Diseases initiative (DNDi) in Geneva, Switzerland, rediscovered the drug and partnered with French drugmaker Sanofi to advance it. Last week, the European Medicines Agency, reviewing the drug in cooperation with the World Health Organization, gave a “positive opinion,” clearing the way for individual African countries to approve it. DNDi says it should reach patients by mid-2019.

NSF lifts one-proposal rule

RESEARCH FUNDING | The National Science Foundation’s (NSF’s) biology directorate announced last week it will no longer limit researchers to submitting only one proposal in which they are listed as a principal investigator (PI) or co-PI. The restriction, announced in August, applied only to the directorate’s three core tracks, which include environmental, molecular, and cellular biology. It was intended, in part, to reduce the number of rejected proposals resubmitted without major changes. But scientists pushed back, fearing the change would discourage collaborations and disadvantage those who rely on the PI designation for career advancement. NSF officials say opposition from the community—and a dip in grant submissions to the biology directorate—prompted them to reverse course.

India plans Venus mission

PLANETARY SCIENCE | The Indian Space Research Organisation (ISRO) says it will launch an orbiter to Venus in 2023 and will field international proposals to include research instruments. Officials from the agency announced the plan on 6 November at a meeting of the NASA Venus Exploration Analysis Group in Laurel, Maryland. ISRO has preselected 12 instruments proposed by Indian scientists, including several cameras and chemical probes, but is inviting additional proposals from the global research community until 20 December. Because Venus has a mass and gravity similar to Earth’s, researchers hope studying the runaway warming of its dense, carbon dioxide–rich atmosphere will offer insight about climate change scenarios on Earth. “NASA is looking forward to working with ISRO to collaborate and help India succeed in this endeavour,” says Lori Glaze, acting director of NASA’s Planetary Science division in Washington, D.C.

NFL funds brain research

NEUROSCIENCE | Two years after announcing it would contribute $40 million to traumatic brain injury research,

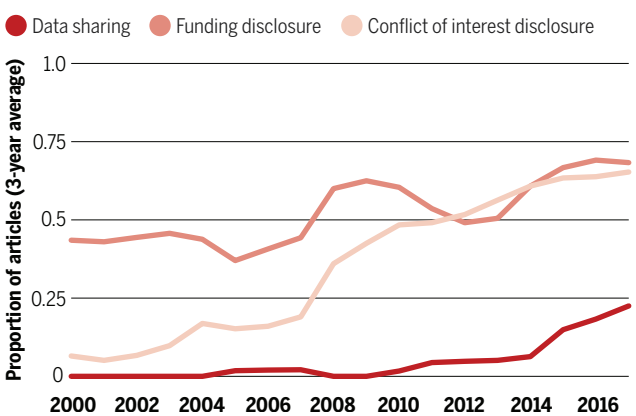
PHOTO: NEIL BRANDVOLD/DNDi

the National Football League (NFL) has revealed the winners of five awards worth a combined $35 million. The projects were selected from among 129 proposals by an NFL-constituted scientific advisory board; its chair will determine what research to fund with the remaining $5 million. The projects include a $9.4 million, 3-year study of 6000 Canadian high school athletes led by epidemiologist Carolyn Emery at the University of Calgary in Canada. It will evaluate tools for diagnosing concussions, as well as factors that predict outcome and ways to prevent and manage such injuries. A team led by sports medicine physician William Meehan at Boston Children's Hospital will receive $14.7 million to neurologically assess and track as many as 2500 former NFL players and study potential therapies for repeated brain trauma. In 2016, NFL came under scrutiny for trying to control an earlier, "unrestricted" gift of $30 million in brain research money to the Foundation for the National Institutes of Health. NFL at that time said it would fund its own research going forward (*Science*, 23 September 2016, p. 1345).

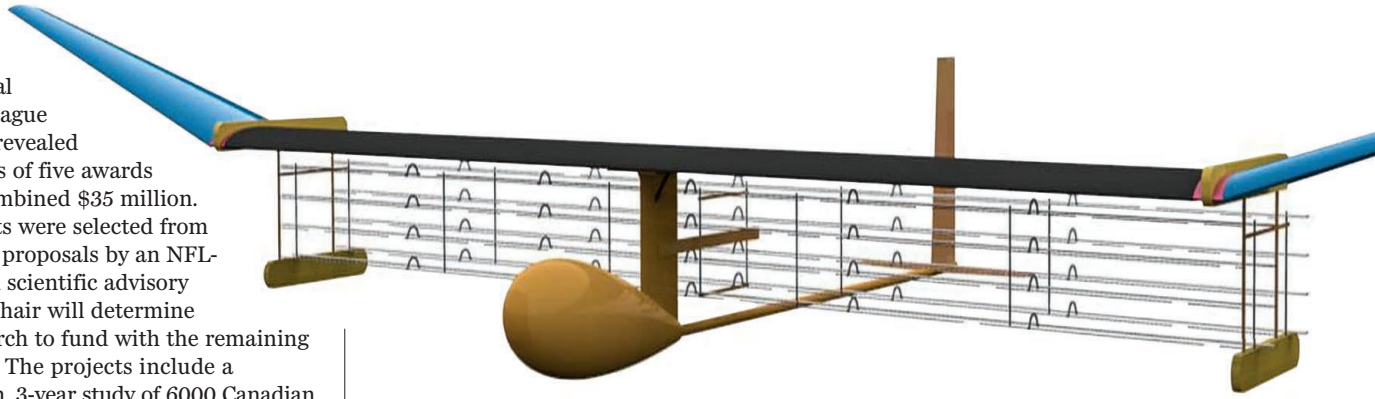
Authors open up—slowly

PUBLISHING | Biomedical researchers are sharing more of their data and conflicts of interest, but they still have a long way to go. That's the conclusion of a study in this week's issue of *PLOS Biology* that compared a random sample of 149 articles published from 2015 to 2017 with a similar set of papers from a longer stretch: 2000 to 2014. In 2017, about two-thirds of papers reported funding sources and conflicts of interest, up from less than half (see graph, below). Conflict disclosures have risen steadily since 2006, the team from Yale University and Stanford University in Palo Alto, California, reports. Only 21% of 2017 papers included a statement about how to obtain the raw data, but almost none did a decade ago. The changes reflect stricter requirements from funding agencies and journals aimed at improving transparency and reproducibility, the authors surmise.

Transparency on the rise in biomedicine



CREDITS: (IMAGE) H. XU *ET AL.*, *NATURE*, 10.1038/S41586-018-0707-9 (2018); (GRAPHIC) J. KAISER/*SCIENCE*; (DATA) J. D. WALLACH *ET AL.*, *PLOS BIOLOGY* 16, **11** (2018)



AVIATION

A Wright Flyer for the 21st century

The quieter, greener airplanes of the future might look something like this. Engineers have now successfully flown an unmanned prototype of this "electroaerodynamic" airplane, which doesn't rely on propellers, turbines, or fossil fuels. Instead, its solid-state propulsion system uses batteries to generate an electric field, creating a wind of charged particles that collide with air molecules to give the plane a push forward. In 10 test flights, the prototype, which has a 5-meter wingspan, flew up to 60 meters—a little farther than the Wright brothers' first flight—with an average altitude of 0.47 meters, researchers report this week in *Nature*. The propulsion system's efficiency must now be improved before the plane can fly farther or carry a payload.

Ocean warming estimate revised

CLIMATE CHANGE | Statistical flaws are tempering a recent study's warning that the world's oceans are warming faster than previously thought. The paper, published 1 November in *Nature*, found warming in recent decades was 60% higher than some previous estimates. The new estimate was based on measures of atmospheric oxygen and carbon dioxide, which increase as water warms, and not on actual temperature readings. On 9 November, co-author Ralph Keeling at the Scripps Institution of Oceanography in San Diego, California, announced the team was preparing a correction after being alerted to errors. They stemmed from how the scientists estimated carbon dioxide and oxygen absorption and release on land and how they dealt with errors in oxygen measurements, Keeling said. He expects the correction to have a small impact on the rate of temperature increase, but to widen the uncertainty. That means the findings could be less useful for estimating how fast the world will warm under different future emissions scenarios.

Taiwan reef under threat

ENVIRONMENTAL SCIENCE | Taiwanese scientists and environmentalists are protesting the planned construction of a liquefied natural gas port off the coast of Taoyuan in northwestern Taiwan that they say will damage a unique algal reef ecosystem nearby. The 27-kilometer-long Datan reef was built up over thousands of years by pink and purple algal species named crustose coralline algae. The state-owned constructor of the port, CPC, says the environmental impact will be minimal and argues the site is the only place where construction can finish on time to meet Taiwan's growing energy demands. But scientists fear the structure would cause changes in water currents that would erode the reef and threaten animal species living in its vicinity, including hammerhead sharks.

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

OCEAN SCIENCE

Luxe research ship to explore the deep ocean

REV Ocean ramps up with $500 million from Norwegian fisheries and oil tycoon

By **Erik Stokstad**

"A dream vessel" is what Joana Xavier, a sponge expert at the University of Porto in Portugal, calls a new research ship due to launch in 2021. Funded by a Norwegian billionaire, the 183-meter-long *Research Expedition Vessel* (*REV*) will be the largest such ship ever built, more than twice the length of most rivals. Engineered to endure polar ice, punishing weather, and around-the-world voyages, the *REV* will not only be big and tough, but packed with top-of-the-line research gear—and luxurious accommodations. Its full capabilities were detailed for the first time last week at a meeting on deep-sea exploration at The Royal Society in London.

The $350 million ship, under construction in a Black Sea shipyard in Romania, is owned by Kjell Inge Røkke, 60, who made his fortune in fishing, offshore oil, and other marine industries. In October, he promised an additional $150 million to REV Ocean in Fornebu, Norway, to operate the ship for at least 3 years, giving scientists free access. Røkke started the foundation last year to find solutions to climate change, ocean acidification, overfishing, and marine pollution. "The scale of the investment and commitment is astounding," says Victor Zykov, science director of the Schmidt Ocean Institute, a charity in Palo Alto, California, that has its own research vessel, the *Falkor*.

Many national research fleets are aging and shrinking. Since 2005, for example, the U.S. academic fleet has declined from 27 vessels to 18, and by 2025 it will it drop to 16 ships. As a result, marine scientists can face long waits for ship time. "If I want to know what's happening in a particular place, it might not work out within a decade," says Antje Boetius, an oceanographer and director of the Alfred Wegener Institute in Bremerhaven, Germany. Philanthropists have launched several vessels to help shorten the queue (see table, left), but few are dedicated to research, and all are dwarfed by the *REV*.

It offers room for 60 researchers and large areas for science and engineering. It will have trawls for capturing marine life and a remotely operated vehicle (ROV) for on-the-spot observations, a rare combination, and much else. "The idea that all the assets are on the ship, and you can pick and choose, that is tremendous," says Ajit

Big boat, big science

Private ships have helped compensate for declining access to government research vessels. Once complete, the *Research Expedition Vessel* (*REV*) will be the largest and best equipped.

SHIP	MAIN PURPOSE	LENGTH (M)	RESEARCH BERTHS	RANGE (KM)	BUILT	REFIT	PATRON
REV	Research	181	60	39,000	2021	N/A	Kjell Inge Røkke
Falkor	Technology development	83	17	15,000	1981	2012	Eric and Wendy Schmidt
Nautilus	Exploration/ education	64	31	24,000	1967	2008	Bob Ballard
Alucia2	Exploration/ media	87	50	41,000	2010	2019	Ray Dalio
Petrel	World War II shipwrecks	76	35	Not provided	2003	2017	Paul Allen

Twice as big as most research ships, the *REV* (seen in an artist's concept) can operate in polar regions.

Subramaniam, a microbial oceanographer at the Lamont-Doherty Earth Observatory of Columbia University. The ROV, capable of 6000-meter descents, can be launched through large side doors or a moon pool in the hull. A pair of ship-borne helicopters can release smaller autonomous underwater vehicles (AUVs), which don't need tethers to the main vessel. "Think of it as an aircraft carrier for robotics," says Chris German, a marine geochemist at Woods Hole Oceanographic Institution in Massachusetts. The *REV* will also have a crewed submersible, probably one capable of descending 2000 meters.

The main trawl, designed by Røkke's company Aker BioMarine for harvesting krill in the Southern Ocean, can remain 3000 meters deep while funneling fish to a tube that quickly pumps them up to the ship's wet labs. This offers the tantalizing possibility of collecting jellyfish and other soft organisms that normally don't survive the slow trip to the surface when the trawl is winched up, opening a porthole into marine food webs. "If the gear can sample with less damage, this would really help," says biological oceanographer Xabier Irigoien, science director of AZTI, a nonprofit institute for marine research in Pasaia, Spain.

> ***"Think of it as an aircraft carrier for robotics."***
>
> **Chris German**, Woods Hole Oceanographic Institution

The *REV* could also make a significant contribution to understanding fisheries on the high seas, Irigoien adds. The intergovernmental organizations that regulate fishing beyond national jurisdictions don't own ships and can rarely afford to pay for time. Free access to the *REV* could help scientists fill the gaps. They might be able to track tagged tuna or sharks with AUVs, for instance, while sizing up schools of fish with the ship's high-tech sonar. By combining data from the trawl and sonar, Irigoien says, researchers could chart potential fisheries in the deep sea before they're exploited. The same technologies would be useful for investigating far-flung marine protected areas.

Most research vessels are spartan, but on the *REV* scientists will have nearly full run of the ship, including its lounges, gym, dining room, and seven-story atrium. Magne Furuholmen, an artist and former keyboardist of 1980s pop group A-ha, is choosing the art collection. The *REV* is also eco-friendly: It's fuel efficient with low emissions and a broad, stable hull designed to reduce noise pollution. If it encounters a garbage patch, booms can collect up to 5 tons a day of plastic to incinerate onboard for energy.

Alex Rogers, an oceanographer at the University of Oxford in the United Kingdom, starts next month as the full-time science director for REV Ocean. He says scheduling an expedition on the *REV* could be quicker and more flexible than on government research vessels, which are sometimes limited by range, budget, or scientific focus. On the other hand, working with philanthropists is not like dealing with a research funding agency. "You have to explain what you're doing," Rogers says. "Be prepared to communicate with them."

Røkke's history could raise concerns about hidden agendas. "I think there will always be some level of suspicion from the public that a person like Røkke—who made a fortune in ocean industries—that somehow there are strings attached," Rogers admits. So he is working with the Research Council of Norway to design an independent review process that will select projects for ship time. Rogers says the only expectation is that researchers focus on solutions and share their data after they publish. "If Alex is involved, I have faith," says Kerry Howell, a deep-sea ecologist at the University of Plymouth in the United Kingdom. "He's not the kind of person who would work for the dark side."

As for Røkke, he has no plans to run the foundation and is "very meticulous about this being fully independent and objective," says Nina Jensen, CEO of REV Ocean. "He is serious about making a difference for the oceans." Jensen, who studied marine biology and previously led the environmental advocacy group WWF Norway, says she told Røkke she will resign if one of his companies, Aker BP, drills for oil in Norway's Lofoten islands, which boast rich fisheries and the largest known deep-water coral reefs.

To help cover the costs of operation, for 4 months a year the *REV* will open 60% of its berths on research expeditions to paying eco-tourists. For another 4 months, the entire ship will be available as a luxury yacht. Jensen hopes benefactors will charter it as a "floating think tank" to win more support for ocean protection. Any extra funds raised will go to support early-career scientists.

It's an unproven model, Jensen concedes, but the *REV* won't sink or float on its fundraising prowess. Røkke's pledge last month to support operations was only his first, Jensen says. "It will not be the last." ■

SPACE

NASA to pay private space companies for moon rides

Experiments on commercial landers could confirm a wetter, more active moon

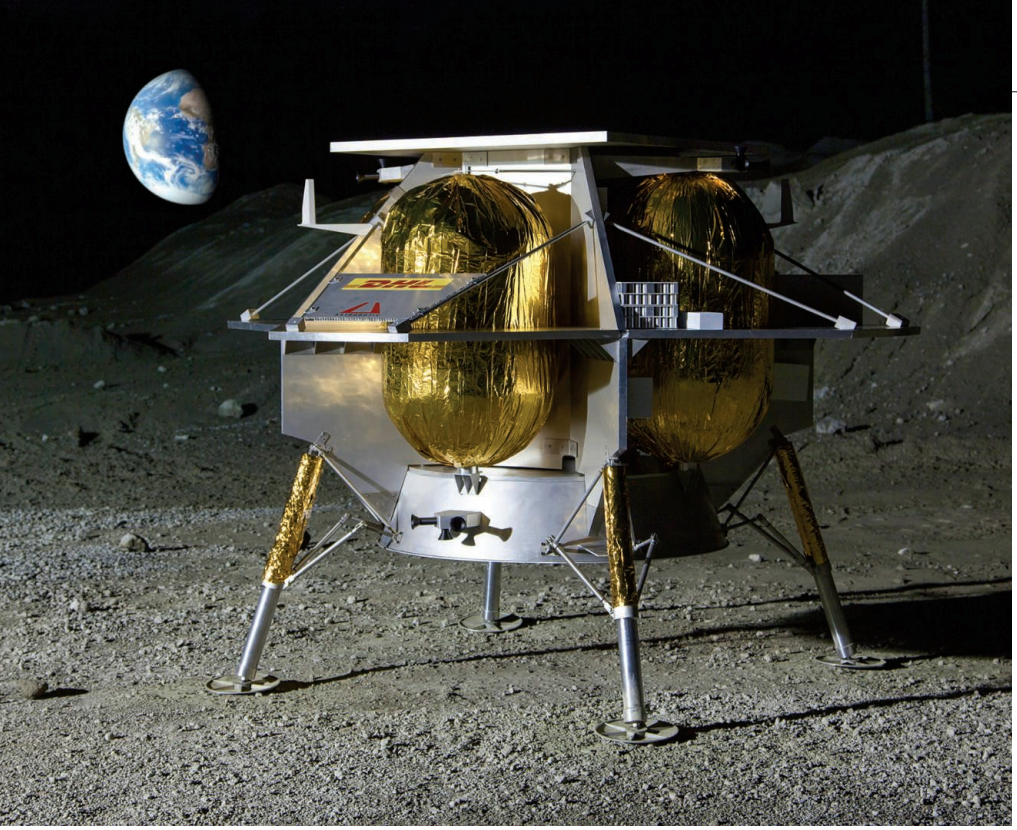
By **Paul Voosen**

Next month, almost a half-century since the United States last landed a spacecraft on the moon, NASA is expected to announce plans for a return. But the agency will just be along for the ride. Rather than unveiling plans for its own spacecraft, NASA will name the private companies it will pay to carry science experiments to the moon on small robotic landers.

Under a program called Commercial Lunar Payload Services (CLPS), NASA would buy space aboard a couple of launches a year, starting in 2021. The effort is similar to an agency program that paid private space companies such as Elon Musk's SpaceX to deliver cargo to the International Space Station (ISS). "This a new way of doing business," says Sarah Noble, a planetary scientist at NASA headquarters in Washington, D.C., who is leading the science side of NASA's lunar plans.

Scientists are lining up for a ride. "It really feels like the future of lunar exploration," says Erica Jawin, a planetary scientist at the Smithsonian Institution's National Museum of Natural History in Washington, D.C. She and other attendees at the annual meeting of the Lunar Exploration Analysis Group in Columbia, Maryland, last week were eager to show NASA why their small experiments would be worthy hitchhikers on the landers.

Several companies, including Astrobotic, Moon Express, and iSpace, are vying to establish a commercial moon market. Buying rides to the moon from launch providers like Rocket Lab, each firm hopes to become the go-to carrier for other companies seeking to prospect the moon for rocket fuel ingredients, or to gather rocks to sell for study. But a contract with NASA is the real prize. Moon Express, for example, has designed the MX-1, a lander roughly the



Smaller than a compact car, Astrobotic's Peregrine lander could put experiments on the moon for NASA.

size and shape of *Star Wars*'s R2-D2. But, "We won't pull the trigger until we know we have a CLPS award," says Moon Express CEO Robert Richards in Cape Canaveral, Florida.

The companies selected for CLPS must deliver at least 10 kilograms of payload by the end of 2021, NASA says. It is scrambling to find instruments that are ready to fly. "What do you have sitting on shelf now that you can throw onto the mission immediately?" Noble says. "We're looking for flight spares, engineering models, student-built projects. It's a little bit of a weird call for us." The agency is planning to pay up to $36 million to adapt eight to 12 existing scientific instruments to the initial small landers; by the middle of next decade it aims to build a pipeline of instruments for bigger landers that might also carry rovers.

The first small commercial landers will pale in capability next to traditional NASA missions. Some will likely fail, as NASA's science chief, Thomas Zurbuchen, has repeatedly warned. They will not survive the lunar night, 2 weeks when surface temperatures drop to −173°C. They may not be capable of landing in a specific spot. But scientists are still excited to get cameras and other instruments back to the surface of the moon, says Clive Neal, a lunar scientist at the University of Notre Dame in South Bend, Indiana. "It's a great start."

NASA is still working out destinations for the commercial landers. Earlier this year, lunar scientists compiled a list of 16 sites key to testing an emerging picture of the moon as more active volcanically and richer in water than was thought. For example, 4 years ago scientists studying the Ina caldera, a collection of smooth, small volcanic mounds on the moon's nearside, noticed it was relatively free of craters. The observation suggested that instead of ending a billion years ago, volcanism—a sign of interior heat—persisted until a few million years ago, smoothing the landscape. If true—and some dispute the finding—it would upend theories of how the moon, and potentially the rocky planets, cool over time.

> ***"It really feels like the future of lunar exploration."***
>
> **Erica Jawin**, Smithsonian Institution's National Museum of Natural History

At the 2-kilometer-tall Aristarchus plateau, scientists want to study abundant volcanic ash deposits, which were created in explosive, gas-driven eruptions, a rarity on the moon. Thanks to its fine granularity, the ash could also make an excellent building block for human habitats. Samples from Marius Hills, a shield volcano that likely erupted for a substantial time, could shed light on how the moon's endowment of water, carbon monoxide, and other volatiles evolved over time. And a look inside permanently shadowed craters at the moon's poles could confirm whether some of its water is frozen there, says Brett Denevi, a planetary geologist at the Johns Hopkins University Applied Physics Laboratory in Laurel, Maryland.

The first small landers allow only small steps toward these science goals. But the agency could eventually support commercial robotic sample return missions, which Astrobotic and Moon Express hope to offer. "They could say, 'I want 2 kilograms of lunar regolith from such and such location,'" says John Thornton, CEO of Astrobotic in Pittsburgh, Pennsylvania. Returned samples could help researchers with a perennial goal: dating the moon's old and young craters, which dictate age estimates for surfaces across the solar system.

NASA also wants to fly people back to the vicinity of the moon—but in its own spacecraft. It is building the Gateway, a small outpost, that, by 2024, would host astronauts for a couple of months at a time in a pressurized volume one-tenth the size of the ISS. The Gateway, which will cost NASA at least $3 billion for its first few sections, would not orbit the moon but, rather, would follow a weeklong loop around a distant gravitational balance point—a poor vantage for lunar observations. "We're not 100% sure of its value for lunar science," says Ryan Watkins, a lunar scientist at the Planetary Science Institute in St. Louis, Missouri. Noble acknowledges that the Gateway may be more valuable for studying the sun or the rest of the universe.

Ben Bussey, NASA's chief scientist for human exploration, says the agency is trying to accommodate scientists' concerns. For example, it will prioritize equipping the station with a robotic arm, needed for mounting experiments on its exterior. And it is investigating the possibility of a reusable "tug" spacecraft that could ferry landers, samples, and instruments between the Gateway and low lunar orbit, Bussey says.

Looming over these lunar plans is the fear that they will change. Republicans in Congress proposed a moon return under former President George W. Bush, only to have the administration of former President Barack Obama emphasize a visit to an asteroid in deep space, as a stepping stone to Mars. So far, the Republican-led Congress has fully funded the agency's moon plans: Draft 2019 spending bills contain $500 million for the Gateway and more than $200 million for NASA's initial lander and science plans. Now, lunar scientists need to convey their support to newly empowered Democrats, Neal says. If NASA funds a couple of small landers and the program changes again, Denevi adds, "that's just going to be another wasted decade." ■

IMAGE: ASTROBOTIC TECHNOLOGY

SYNTHETIC BIOLOGY

Artificial cells gain communication skills

Cell mimics make and pass on proteins that influence their neighbors

By **Mitch Leslie**

No biologist would mistake the microscopic "cells" that chemical biologist Neal Devaraj and colleagues are whipping up at the University of California, San Diego (UCSD), for the real thing. Instead of the lipid membrane that swaddles our cells, these cell mimics wear a coat of plastic—polymerized acrylate. And although they harbor a nucleuslike compartment containing DNA, it lacks a membrane like a real cell's nucleus, and its main ingredients are minerals found in clay.

Yet these mock cells are cutting-edge, "the closest anyone has come to building an actual functioning synthetic eukaryotic cell," says synthetic biologist Kate Adamala of the University of Minnesota in Minneapolis, who was not part of the work. Like real cells, the spheres can send protein signals to their neighbors, triggering communal behavior. And as Devaraj and his team revealed in a preprint recently posted on the bioRxiv site, the "nucleus" talks to the rest of the cell, releasing RNA that sparks the synthesis of proteins. The artificial nuclei can even respond to signals from other cell mimics. "This may be the most important paper in synthetic biology this year," Adamala says.

Synthetic biologists have big dreams for artificial cells. Compared with simpler synthetic structures, such as the liposomes that are already being used to transport certain drugs in the body, they could be more sensitive to their environment and perform a greater variety of jobs. In the future, artificial cells may deliver drugs more precisely to their targets, hunt down cancer cells, detect toxic chemicals, or improve the accuracy of diagnostic testing. Arrays of interacting synthetic cells could form artificial tissues and smart materials that sense and adapt to their surroundings. As scientists struggle to devise cell facsimiles, they may also learn more about how life originated and overcame some of the same engineering challenges.

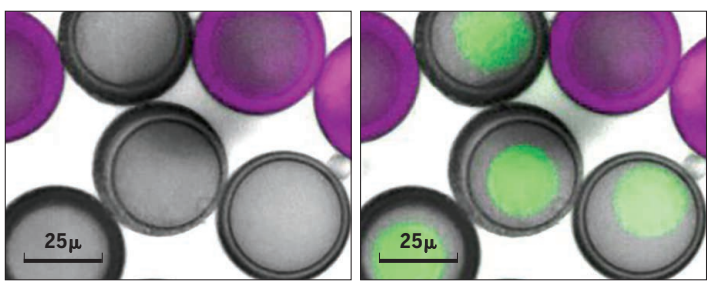
Performing some functions of a cell, such as manufacturing proteins and duplicating DNA, in isolation won't be enough. "If we are going to develop synthetic materials, we need to have the individual units cooperate," Devaraj says. Researchers had already devised synthetic cells that can communicate with each other by exchanging relatively small molecules such as sugars and hydrogen peroxide. However, Devaraj notes, many of the molecular signals in our bodies, including the hormone insulin and the cytokines that fire up our immune cells, are proteins and are typically much larger.

To make a more cell-like cell mimic, Devaraj and his colleagues stepped away from nature. Their latest pseudocells "look a little bit like natural cells, but they are made of completely artificial materials," says co-author Henrike Niederholtmeyer, a synthetic biologist at UCSD. The researchers used a silicon chip with microscopic fluid-filled channels to extrude tiny droplets that contain raw materials such as DNA, minerals from clay, and individual acrylate molecules. Ultraviolet light and chemical treatment spurred a porous membrane to form around each droplet. At the same time, the minerals and DNA inside the droplet condensed into a gel with the texture of a soft contact lens, creating a version of the nucleus, Devaraj says.

The result was a cell replica with new powers of communication. For some experiments, Devaraj's team equipped the nuclei of the cell mimics with DNA that encodes green fluorescent protein (GFP). They also outfitted some of their creations with a trap, a sticky stretch of DNA that captures GFP molecules. By adding a mixture of enzymes and other necessities for protein synthesis, such as ribosomes, to the fluid surrounding the ersatz cells, the investigators switched them on. This molecular machinery crossed the porous membrane, read the genetic information in the nucleus, and sparked synthesis of GFP.

The scientists then mixed cell mimics designed to produce GFP with receiver cells that couldn't make the marker themselves but did harbor the DNA trap for GFP. After 2 hours, receiver cells that were adjacent to senders were aglow, indicating that they had picked up the GFP message from neighbors. In a similar experiment, the team crafted mimics that released a different protein signal that switches on GFP synthesis in recipients. Like real cells, these cell mimics could communicate with nearby counterparts and stimulate them to produce proteins.

The imitation cells also displayed another lifelike attribute called quorum sensing, in which cells' behavior changes once they become abundant enough. This ability came to light when researchers tested solutions containing different densities of cell mimics, all of which released the activator of GFP synthesis and could make GFP as well when triggered. If a solution contained only a few of the synthetic cells, almost none turned green. After they reached a threshold density, however, nearly all of them lit up. Before they can begin to make GFP, the artificial cells apparently need to absorb a certain minimal amount of the activating protein from their surroundings.

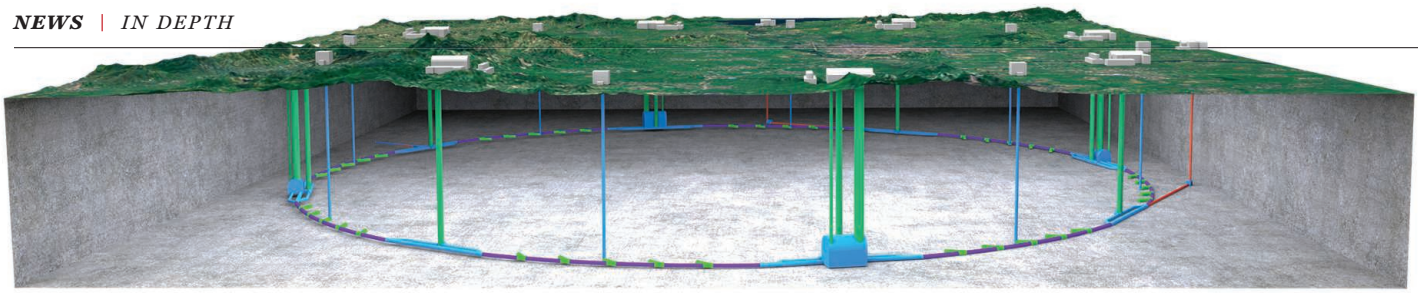


In a mix of artificial cells, one kind (purple) makes and releases a fluorescent protein, which is received and trapped within a second kind (gray), turning those mimics green.

The cell mimics are tough, remaining undamaged after 2 years in a freezer. Their durability may make them good environmental sensors—one of several applications for the structures that the UCSD team is now exploring. And Devaraj and colleagues hope to equip these or other synthetic cells with the ability to grow and divide.

Bioengineer Yuval Elani of Imperial College London is impressed with the design of the cell mimics. "The concept of using these nonbiological components is a powerful one." But the artificial components could also be a drawback for applications, he notes, if they prove incompatible with "natural" components making up artificial cells that other researchers are developing. ■

IMAGES: HENRIKE NIEDERHOLTMEYER



China's Circular Electron Positron Collider would be built underground in a 100-kilometer-circumference tunnel at an as-yet-undetermined site.

HIGH ENERGY PHYSICS

Asia set to take center stage in Higgs studies

China unveils full design for a $5 billion collider; Japan nears decision on its own plan

By **Dennis Normile**, *in Beijing*

The center of gravity in high energy physics could move to Asia if either of two grand plans is realized. At a workshop here last week, Chinese scientists unveiled the full conceptual design for the proposed Circular Electron Positron Collider (CEPC), a $5 billion machine to tackle the next big challenge in particle physics: studying the Higgs boson. (Part of the design was published in the summer.) Now, they're ready to develop detailed plans, start construction in 2022, and launch operations around 2030—if the Chinese government agrees to fund it.

Meanwhile, Japan's government is due to decide by the end of December whether to host an equally costly machine to study the Higgs, the International Linear Collider (ILC). How Japan's decision might affect China's, which is a few years away, is unclear. But it seems increasingly likely that most of the future action around the Higgs will be in Asia. Proposed "Higgs factories" in Europe are decades away and the United States has no serious plans.

The Higgs boson, key to explaining how other particles gain mass, was discovered at CERN, the European particle physics laboratory near Geneva, Switzerland, in 2012—more than 40 years after being theoretically predicted. Now, scientists want to confirm the particle's properties, how it interacts with other particles, and whether it contributes to dark matter. Having only mass but no spin and no charge, the Higgs is really a "new kind of elementary particle" that is both "a special part of the standard model" and a "harbinger of some profound new principles," says Nima Arkani-Hamed, a theorist at the Institute for Advanced Study in Princeton, New Jersey. Answering the most important questions in particle physics today "involves studying the Higgs to death," he says.

"Physicists want at least one machine," says Joao Guimaraes da Costa, a physicist at the Chinese Academy of Sciences's Institute of High Energy Physics (IHEP) here, which put together the Chinese proposal. "Ideally, both should be built," because each has its scientific merits, adds Hitoshi Murayama, a theoretical physicist at the University of California, Berkeley, and the University of Tokyo's Kavli Institute for the Physics and Mathematics of the Universe in Kashiwa, Japan.

The CERN discovery relied on the Large Hadron Collider, a 27-kilometer ring in which high-energy protons traveling in opposite directions are steered into head-on collisions. This produces showers of many types of particles, forcing physicists to sift through billions of events to spot the telltale signal of a Higgs. It's a bit like smashing together cherry pies, Murayama says: "A lot of goo flies out when what you are really looking for is the little clinks between pits."

Smashing electrons into their antimatter counterparts, positrons, results in cleaner collisions that typically produce one Z particle and one Higgs boson at a time, says Bill Murray of The University of Warwick in Coventry, U.K. How Z particles decay is well understood, so other signals can be attributed to the Higgs "and we can watch what it does," Murray says.

Japan's plan to build an electron-positron collider grew from international investigations in the 1990s. Physicists favored a linear arrangement, in which the particles are sent down two straight opposing raceways, colliding like bullets in rifles put muzzle to muzzle. That design promises higher energies, because it avoids the losses that result when charged particles are sent in a circle, causing them to shed energy in the form of x-rays. Its disadvantage is that particles that don't collide are lost; in a circular design they continue around the ring for another chance at colliding.

Along the way, Japan signaled it might host the machine and shoulder the lion's share of the cost, with other countries contributing detectors, other components, and expertise. A 2013 basic design envisioned a 500-giga-electronvolt (GeV) linear collider in a 31-kilometer tunnel costing almost $8 billion, not counting labor. But by then, the CERN team had already pegged the Higgs mass at 125 GeV, making the ILC design "overkill," Murayama says. The group has since revised the plan, aiming for a 250-GeV accelerator housed in a 20-kilometer-long tunnel and costing $5 billion, says Murayama, who is also deputy director of the Linear Collider Collaboration, which coordinates global R&D work on several future colliders.

IHEP scientists made their own proposal just 2 months after the Higgs was announced. They recognized the energy required for a Higgs factory "is still in a range where circular is better," Murray says. With its beamlines buried in a 100-kilometer-circumference tunnel at a site yet to be chosen, the CEPC would collide electrons and positrons at up to 240 GeV.

Both approaches have their advantages. The CEPC will produce Higgs at roughly five times the rate of ILC, allowing research to move faster. But Murayama notes that the ILC could easily be upgraded to higher energies by extending the tunnel by another couple of kilometers. Most physicists don't want to choose. The two colliders "are quite complementary," Murray says.

Whether politicians and funding agencies agree remains to be seen. Construction of the CEPC hinges on funding under China's next 5-year plan, which starts in 2021, says IHEP Director Wang Yifang. IHEP would then also seek international contributors. Murayama says Japan needs to say yes to the ILC in time to negotiate support from the European Union under a particle physics strategy to be hammered out in 2019. Missing that opportunity could mean delaying the collider by 20 years, he says—and perhaps ceding the field to China. ■

IMAGE: IHEP

PALEONTOLOGY

Giant mammal cousin rivaled early dinosaurs

Ancient plant eater grew as large as today's elephants

By **Gretchen Vogel**

Imagine if you crossed a rhino with a giant turtle and then supersized the result: You might get something like *Lisowicia bojani*, a newly discovered Triassic mammal cousin that had a body shaped like a rhinoceros, a beak like a turtle, and weighed as much as an African elephant, about 9 tons. Paleontologists say this startling creature offers a new view of the dawn of the age of the dinosaurs. "Who would have ever thought that there were giant, elephant-sized mammal cousins living alongside some of the very first dinosaurs?" marvels Stephen Brusatte, a vertebrate paleontologist at The University of Edinburgh.

Researchers had thought that during the Late Triassic, from about 240 million until 201 million years ago, early mammals and their relatives "retreated to the shadows while dinosaurs rose up and grew to huge sizes," Brusatte says. "That's the story I tell my students in my lectures. But this throws a wrench into that simple tale," suggesting the same evolutionary forces that favored giant dinosaurs were at work on other creatures as well.

The new fossil, a partial skeleton described online this week in *Science*, is an ancient plant eater called a dicynodont; the name means "two dog tooth," referring to the characteristic tusks on the upper jaw, which resemble oversize canines. Apart from the tusks, dicynodonts were mostly toothless, with a horny beak like modern-day turtles. They're part of the large evolutionary group called synapsids, which includes our mammal ancestors, and they were some of the most abundant and diverse land animals from the mid-Permian period into the Middle Triassic, from 270 million until about 240 million years ago.

Dicynodonts "are the first group of vertebrates that were successfully able to eat plants," says Tomasz Sulej, a paleontologist at the Polish Academy of Sciences's Institute of Paleobiology in Warsaw.

Dicynodonts evolved a striking range of forms: One burrowed like modern-day moles, another is the first known vertebrate to live in trees. Some grew as large as today's hippos, which weigh about 1.5 tons. However, the fossil record suggests the group was in decline by the time *L. bojani* lumbered into view. And even in dicynodonts' heyday, they did not come close to early dinosaurs in size.

Sulej, with the Institute of Paleobiology's Jerzy Dzik and Grzegorz Niedźwiedzki, a paleontologist at Uppsala University in Sweden, discovered the new fossil in a clay pit, once quarried for brickmaking, in the village of Lisowice, about 100 kilometers northwest of Krakow in southern Poland. In 2006, the team got a tip that someone had found bone fragments at the site. On their first visit, they found fossils within 15 minutes; during 11 years of fieldwork, they excavated more than 1000 bones.

They didn't immediately recognize the new dicynodont as such—in part because it is so big, Sulej says. "Our first idea was that it was a sauropod," which were the largest known herbivores in this period, reaching 11 meters long. But skull fragments and limb bones identified the animal as the biggest, most recent dicynodont ever found. The team named it after the village and 18th century comparative anatomist Ludwig Heinrich Bojanus; they estimate it was more than 4.5 meters long and 2.6 meters tall.

Most dicynodonts had a posture that seems awkward to the modern eye: Their hind limbs were straight, like those of today's mammals, but their forelimbs sprawled, lizard-style, with a bend at the elbow. The team suggests that because of the way *L. bojani*'s upper arm bone connected with its shoulder, its front legs must have been oriented vertically, giving it a more erect stance than in modern reptiles. This posture, like that of sauropod dinosaurs and modern mammals, might have helped support its massive weight. But others caution that reconstructing posture without soft tissue can be difficult.

L. bojani's bones also lacked the lines that, in most dicynodont fossils, mark periods when bone growth slowed. The animal may have grown unusually quickly, or wasn't yet full grown when it died. Given the "truly amazing" size of the creature, "it likely grew fast," says paleontologist Jennifer Botha-Brink of the Bloemfontein Palaeosystems Centre and the National Museum in South Africa. But she adds that lines signaling slower growth might have been erased when the bone was remodeled during adulthood, which happens in elephants today.

Researchers have hypothesized that sauropods grew big to avoid getting eaten. That may have been true for *L. bojani*, too, Sulej says. The Lisowice bone bed also contains the remains of a 5-meter-long predator—likely a dinosaur—and coprolites (fossilized feces) containing dicynodont bones.

The researchers will seek more specimens farther east in Russia and Ukraine. "There is definitely more to discover," Niedźwiedzki says. "How many surprises are still waiting for us in the rocks?" ■

This elephant-size mammalian relative, *Lisowicia bojani*, walked on Earth in the Late Triassic, just when dinosaurs were evolving to giant sizes.

ILLUSTRATION: KAROLINA SUCHAN-OKULSKA

FEATURES

CRACKING THE CAMBRIAN

New fossils and sites are helping make sense of the mysterious flowering of animal life half a billion years ago

By **Joshua Sokol**, *in Kootenay National Park in Canada*; *Photography by* **John Lehmann**

The drumming of the jackhammer deepens. Then, a block of shale butterflies open, exposing to crisp mountain air a surface that hasn't seen sunlight in half a billion years. "Woo!" says paleontologist Cédric Aria of the Nanjing Institute of Geology and Palaeontology in China, bracing the top slab of rock upright.

Its underside bears charcoal-colored smudges that look vaguely like horseshoe crabs or the Millennium Falcon from *Star Wars*. "It's a spaceship landing area here," says expedition leader Jean-Bernard Caron, curator of invertebrate paleontology at the Royal Ontario Museum (ROM) in Toronto, Canada.

Those "spaceships" are carapaces, molted onto a long-vanished ocean floor by a species new to science. This field season they've been spilling out of the rocks here, where Caron's team has spent the past few years unearthing groundbreaking animal fossils from the Cambrian period, the coming-out party for animal life on Earth. During the Cambrian, which began about 540 million years ago, nearly all modern animal groups—as diverse as mollusks and chordates—leapt into the fossil record. Those early marine animals exhibited a dazzling array of body plans, as though evolution needed to indulge a creative streak before buckling down. For more than a century, scientists have struggled to make heads or tails—sometimes literally—of those specimens, figure out how they relate to life today, and understand what fueled the evolutionary explosion.

FOSSIL: ROYAL ONTARIO MUSEUM

Jean-Bernard Caron shows off the "mothership," an enigmatic Cambrian life form his team found in the Canadian Rockies this summer.

Gingerly, Aria and Caron place the top piece of their slab aside. Space is hard to come by in the quarry, perched on a ledge the size of a small bedroom at an altitude of 2500 meters, far above Tokumm Creek. For years, an equally forbidding site about 40 kilometers northwest of this valley offered the clearest window on the Cambrian. There, in 1909, U.S. paleontologist Charles Doolittle Walcott discovered the Burgess Shale, a fossil formation that preserves not only hard shells, but also soft features such as the legs, eyes, and guts of Cambrian creepy-crawlies.

But in recent years, Caron has shown that the richest fossil-bearing rock extends many kilometers beyond Walcott's site. This summer's excavation marks his latest visit to this long Cambrian tapestry. Each new stop has offered striking views of unfamiliar animals, many already described in high-profile papers: the little fish relative *Metaspriggina*, a vertebrate ancestor that Caron now speculates clustered in schools; the pincered *Tokummia*; and the ice cream cone–shaped fossils called hyoliths, which Caron's Ph.D. student Joseph Moysiuk last year linked to shelled animals called brachiopods, some of which persist today.

Other sites around the world are also opening new vistas of the Cambrian. Scientists can now explore the animal explosion with a highlight reel of specimens, along with results from new imaging technologies and genetic and developmental studies of living organisms. "There have been a host of new discoveries," says paleontologist Doug Erwin of the Smithsonian Institution's National Museum of Natural History in Washington, D.C. Researchers may be closer than ever to fitting these strange creatures into their proper places in the tree of life—and understanding the "explosion" that birthed them.

Each new find brings the simple joy of unearthing and imagining a seemingly alien creature. On a break, Caron cautiously shows off this year's crown jewel, found about a week earlier. It's an intact, hand-size carapace with a center spine, like a Prussian spiked helmet frozen in ancient rock. Another undescribed species, it seems to be related to the spaceships. Caron's team calls it the mothership.

He's nervous just holding it. Burgess Shale fossils are so valuable that Parks Canada keeps the exact locations of Caron's sites secret, monitors them with cameras, and prosecutes fossil poachers. ROM once insured a Burgess Shale specimen for half a million Canadian dollars when it went on loan, he says—and that was an animal known through multiple fossils. This is one of a kind.

"It's going to be iconic," Caron says. "It's the most extraordinary fossil I've ever found."

FOR YEARS, CARON SUSPECTED Walcott's site might be rivaled elsewhere in the Rocky Mountains. The breakthrough came in 2012, near an area called Marble Canyon, where a 2003 wildfire had burned off the trees. While crossing an avalanche chute filled with broken tiles of rock, his reconnaissance party found itself surrounded by impressions of soft-bodied creatures, many with unfamiliar shapes. "It was clear that nobody had ever been walking over this pile of rocks before with this purpose in mind," says Bob Gaines, a geochemist from Pomona College in Claremont, California, who has joined Caron's expeditions since the beginning.

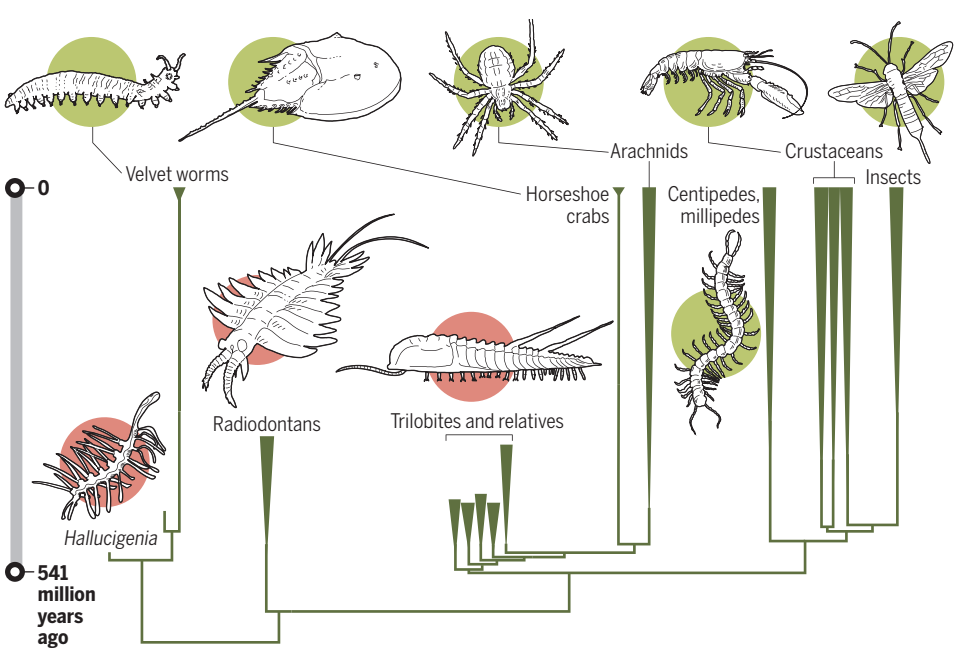
They returned to excavate in 2014. At least one in five of the animals they found at Marble Canyon belongs to species new to science, the team concluded. Now, they've moved on to other sites along the valley.

How Cambrian species are related to today's animals has been debated since the fossils first came to light. Walcott classified his oddities within known groups, noting that some Burgess Shale fossils, such as the brachiopods, persisted after the Cambrian or even into the present. So, for example, he concluded almost all the creatures resembling today's arthropods were crustaceans.

But later paleontologists had other ideas. Harvard University's Stephen Jay Gould perhaps best captured the charisma of Cambrian life in his 1989 book *Wonderful Life: The Burgess Shale and the Nature of History*, in which he lavished attention on the "weird wonders" excavated from Walcott's city block–size quarry. Gould argued that oddballs such as the aptly named *Hallucigenia*, a worm with legs and hard spines, seem unrelated to later animals. He slotted the unusual forms into their own phyla and argued that they were evolution's forgotten experiments, later cast aside by contingencies of fate.

Contemporary paleontologists have settled on yet another way to understand them. Consider the arthropods, arguably Earth's most successful animals. In a family tree, the spray of recent branches that end in living arthropods—spiders, insects, crustaceans—constitutes a "crown" group. But some animals in the Burgess Shale probably come from earlier "stems" that branched off before the crown arthropods. These branches of the tree don't have surviving descendants, like a childless great-uncle grinning out from a family photo. In that view, many of Gould's weird wonders are stem group organisms, related to the ancestors of current creatures although not ancestors themselves. Newer fossils

All in the family A partial schematic of a proposed family tree of arthropods shows the complex relations among living and extinct groups. Some extinct Cambrian creatures (red) may belong to "stem" groups that branched off the arthropod tree before the common ancestor of living groups like arachnids and insects.



CREDITS: (GRAPHIC) N. DESAI/*SCIENCE*; (DATA) JO WOLFE, MIT

from the Canadian Rockies help support that view. Caron argued in 2015, for example, that his specimens of *Hallucigenia* have features suggesting the animal belongs on a stem group of the velvet worms, creatures that still crawl around in tropical forests spitting slime.

Similar analysis awaits the spaceships. At first glance, Caron's team thinks they are a new species or group of radiodontans, stem arthropods that also include *Anomalocaris*, the Cambrian's charismatic apex predator—a clawed, fearsome-jawed swimmer half a meter long. Filling out the branches of that stem group gives a "step-by-step view of how an arthropod built its body" through evolutionary time, says paleontologist Allison Daley at the University of Lausanne in Switzerland.

Throughout much of Cambrian paleontology, that's the game—a high-stakes, sometimes contentious race to find diagnostic body parts on known or new fossils, make arguments about what taxonomic groups they belong to, and maybe revise evolutionary history in the process.

In the past few years, paleontologists have approached the problem with an array of new techniques. Those include scanning electron microscopes, which can discern a specimen's chemical makeup as well as image it, and computerized tomography (CT) scans, which can penetrate fossils without scraping away material. Those tools have also illuminated a startling series of internal features: fossilized Cambrian brains. Beginning in 2011, paleontologist Xiaoya Ma, now at the University of Exeter in the United Kingdom, published a string of papers tracing nervous tissue in exceptionally preserved Chinese fossils. Those nervous system architectures offer a parallel way to sort animals into evolutionary groups, beyond the usual anatomical structures, and other teams have presented their own compelling specimens.

In fossils of the shrimplike *Chengjiangocaris kunmingensis* from southwest China, for example, "we have this structure that looks almost like a pearl necklace," running almost head to tail, says Javier Ortega-Hernández, an incoming professor at Harvard. His team, led by Jie Yang at Yunnan University in Kunming, China, argued in 2016 that the necklace is a nerve cord studded with smaller clusters of neurons, themselves sprouting tiny nerve fibers. Living arthropods no longer have those fibers. But today's velvet worms and priapulid worms do, implying kinship between long-vanished stem arthropods and those groups.

Critics argue that paleontologists such as Ma and Ortega-Hernández overinterpret some fossils, spotting nervous tissues that aren't there. Many of those structures, the critics say, might just be "halos," biofilms formed when microbes broke down internal parts like muscles or guts after death. But other researchers are convinced. "If you look at the best-preserved nervous systems, there's no doubt" that the features are real, says Graham Budd, a paleontologist at Uppsala University in Sweden and an architect of the current stem-and-crown concept.

Bold claims that use anatomy to revise family trees engender similar controversy throughout the field. One argument that *Hallucigenia* fits with the velvet worms, for example, depends on the exact shape of its claws. But other teams counter that the claws aren't diagnostic of ancestry.

The uncertainties leave paleontologists ever hungry for newer, better specimens. "When there is a debate, you bring a new fossil and say, 'Look, this is the feature we see,'" Caron says, warming up in a tent perched high above Tokumm Creek. "Without fossils, it's speculation."

THE FOSSILS MAKE UP for the discomfort: 6 weeks in tents above the tree line warding off grizzlies with an electrified fence, contending with hot days and snow days and wildfire smoke, obeying the smelly requirement to carry everything—everything—out of the national park at the expedition's end.

It's a chilly August morning, 1 day before a helicopter comes to take all human traces

CREDITS: (PHOTO) JOHN LEHMANN; (MAP) N. DESAI/*SCIENCE*

The Burgess Shale, a storied rock formation that offered some of the earliest glimpses of half-a-billion-year-old creatures, crops out high above Tokumm Creek.

away. Today is the last chance to stumble on a fossil that could crack a mystery—say, to find the body that belongs in the mothership carapace.

The nine-member team hikes from camp to their quarry, up steep, rock-littered slopes. Ridged trilobite fossils poke out from exposed layers, but on this expedition, they don't even warrant a second glance. At the quarry, most people split rock while Caron's grad students help ROM curator Maryam Akrami pack away the most recent finds in swimming-pool noodles. "It's the last day," Caron says. "No injuries!"

Each successive excavation in this valley has targeted the same band of rock, which records a single slice of geologic time. But each dig has yielded a different array of new species. That's because conditions varied across the ancient sea floor, favoring different animals. Such variation is "not a shock to anybody that has ever strapped on a snorkel and swum around," Gaines says. But this vast, wide-open valley captures that kind of diversity at a single moment, allowing glimpses of how the earliest animal ecosystems were structured.

As Caron's quarries bring this moment into ever-sharper focus, other sites have opened portals on other stages of the Cambrian. Nearly everywhere, the fossils preserve levels of squishy detail that are absent in specimens from later in Earth's fossil record. In 2012, Gaines and colleagues proposed a reason: Perhaps unique chemical conditions suffused Cambrian seas. After dead animals settled into mud on the sea floor, low levels of sulfates could have slowed decay by sulfur-loving bacteria while alkaline chemistry encased the dead animals in coats of carbonate, sealing soft tissues inside.

In summer 1984, for example, paleontologist Hou Xian-guang of Yunnan University uncovered an arthropod glistening in Cambrian mudstone, its legs seemingly alive. He had discovered the Chengjiang biota, a trove of immaculate fossils that sprawls over a region in southwest China.

Slightly older than the Burgess Shale—about 518 million years old compared with the Burgess's roughly 507 million years—those deposits showcase related animals in a different style of preservation. Unlike Caron's sites, where geologic processes have squashed the fossils almost flat, the Chengjiang animals still retain some depth. Since 2015, Chinese researchers, including Hou, have capitalized on that by using CT scans to make 3D images of the specimens without destroying them. Today, three rival Chinese teams, each with international collaborators, compete to pull out new discoveries from the site. "There is an absolute landslide of material," Ortega-Hernández says.

Add to that sites such as Emu Bay in Australia, where paleontologists announced in 2011 that they had unearthed radiodontan fossils revealing their complex, multifaceted eyes; and Morocco's Fezouata Formation, which paleontologist Peter Van Roy at Ghent University in Belgium reported in 2010. Each site offers distinct insights. "Every fossil assemblage is horrifically biased," says paleontologist Nick Butterfield of the University of Cambridge in the United Kingdom, "but they're horrifically biased in different ways."

The Moroccan samples, for example, date to a little after the Cambrian, and they show a blend between the Cambrian's signature oddities and the more familiar fauna that dominated later periods. "We are still at the point of unpacking fossils," says Daley, a collaborator on that research. "This is a chance to study why some taxa go extinct and why others are able to replace them."

ALTHOUGH SHOW-STOPPING ANIMALS keep falling out of the strata, the full significance of the Cambrian explosion remains a mystery. Arthropods, the most diverse and common creatures known from the time, littered Cambrian ecosystems. Judging by the fossils, Daley argued in a paper in May, the Cambrian witnessed both the birth and step-by-step diversification of many modern groups. Another approach yields a different answer, however. Geneticists use a tool called molecular clocks to trace back down the tree of life. By starting with genetic differences between living animals, which have accrued as a result of random mutations over the eons, molecular clocks can rewind time to the point where branches diverged.

According to recent studies using that method, modern animals began to march off into their separate phyla some 100 million years before the Cambrian. The finding implies that those groups then hung out, inconspicuous or unnoticed in the fossil record, before suddenly stepping on stage.

Paleontologists have a cryptic set of clues about life before the explosion. Long before the odd beasts of the Cambrian

Perched in their quarry 2500 meters up, paleontologists hammer open slabs of shale to expose the rare fossils inside.

evolved, an even more alien set of ocean organisms left impressions on sedimentary rocks now seen in Namibia and Australia. The Ediacarans, as those fossils are called, taunt paleontologists with the same kind of interpretive challenge as the Cambrian's weird wonders. But they're even weirder. Their imprints suggest some grew in fractal patterns; others had three-part symmetry. Unhelpfully, they don't have obvious mouths, guts, or appendages. "That's where the freak flags are going now," says Jo Wolfe, a paleontologist at the Massachusetts Institute of Technology in Cambridge.

Most Ediacarans vanished before Cambrian deposits, perhaps perishing in the world's first mass extinction. But many researchers suspect some belong on the tree of animal life, perhaps as early stems. One Ediacaran, *Kimberella*, looks like an animal: a snail or slug that grazed along the sea floor. In August, *Stromatoveris*, a frond-like Cambrian creature already thought to be an animal, was pegged as an Ediacaran survivor on the basis of its fractal branches. That would make its Ediacaran relatives animals, too. And in September, researchers announced that an iconic Ediacaran fossil called *Dickinsonia*, which looks like a halved Christmas ham, contained lipid molecules that resemble those of living animals.

"We're seeing the beginning of the advent of animals in the Ediacaran," says paleobiologist Mary Droser of the University of California, Riverside. "It's more fun and exciting than just the Cambrian explosion."

And yet even as the Ediacarans shove Cambrian creatures off their perch as the first animals, Cambrian science itself continues to explode. Caron and others keep hunting for fossil features that could reveal the relationships among Ediacaran, Cambrian, and present-day groups. Other researchers struggle to explain what caused the explosion of animal forms. Atmospheric oxygen may have spiked, enabling animals to grow bigger, stronger, and more active. Or erosion could have dumped toxic calcium into the oceans, prompting organisms to shunt it into building hard skeletons.

Or biology itself could have led the way. Inventions such as predation, free swimming, and burrowing into the sea floor—all first seen in or shortly before the Cambrian—could have transformed a placid global ecology into a high-stakes contest, spurring waves of call-and-response innovation between groups. The explosion might also mark the moment when, after millions of years of quiet progress, animals had finally accrued the developmental recipes to build body parts and improvise on basic themes. That genetic toolkit, Butterfield argues, is "absolutely, astronomically, inconceivably complex. It just took a while to figure that out." Or, of course, multiple causes could have piled up together.

A lace crab—a long-extinct "stem" arthropod called *Marrella splendens*—turns up in a slab.

AFTER A LUNCH BREAK, the paleontologists chisel into a few more slabs. Gaines takes rock samples from each layer of their quarries, hoping to reconstruct each environment's chemistry. Then Caron delivers the announcement: "It's over, guys. No more digging."

The next day, its last, the ROM team breaks camp. Over several hours, a helicopter ferries nets sagging with fossils toward a staging area by the highway, making the roughly 10-minute trip again and again. Some specimens, like the spaceships, will be rushed to publication in coming months, now that visiting journalists have seen them. Other finds will sit in drawers, awaiting new techniques or the graduate student who asks the right question.

As the team huddles, waiting for its helicopter ride, tiny, rabbitlike mammals called pikas cry out from the hills. Each helicopter trip erases the signs of human presence one by one, until only carved-out quarries remain. More fossils still rest inside, pressed between folio sheets of rock, waiting for the next season. ■

Joshua Sokol is a journalist in Boston.

PHOTOS: JOHN LEHMANN, FOSSIL FROM ROYAL ONTARIO MUSEUM

INGENUITY: NEXTGEN'S VISION

Nurturing connections to the environment

It seems increasingly clear that our future depends on people caring more about the environment, yet as technology develops, people seem only more disconnected from nature. We asked young scientists: **How can scientists in your field help people see the value of our natural environment?** Scientists from across the world, with expertise in a variety of fields, suggested ways that scientists can make nature more relevant and use research results to facilitate sustainable behaviors. Read a selection of the best responses here. —**Jennifer Sills**

Relate to health and safety

Ecologists can help people recognize the value of the services provided by natural ecosystems, such as the clean water and flood safety provided by healthy river valleys and the crop pollination provided by diverse insect communities. Scientists can explain how natural, undisturbed systems provide us with multiple services vital to our physical, mental, and even financial well-being.

Barbara Pietrzak
Department of Hydrobiology, University of Warsaw, 02-089 Warsaw, Poland. Email: b.pietrzak@uw.edu.pl

Clinical and medical researchers can educate patients and laypeople about the health hazards of pollution, toxins, and pesticides in water, soil, and air. If people know the importance to their health and well-being of keeping their surroundings clean and conserving nature, they will appreciate and care for the natural environment.

Vandana Sharma
Department of Hematology, All India Institute of Medical Sciences, New Delhi, Delhi 110029, India. Email: dr.vandanasharma.phd@gmail.com

Pharmaceutical scientists should educate patients and the general public about the natural origins of their commercial drug products. This will allow laypeople to connect the pharmaceutics they consume to the natural environment, which in turn delivers the message that nature is vital to the sustainability of people.

Dhanuka Wasalathanthri
Sanofi, Fiskdale, MA 01518, USA. Email: dhanuka02@gmail.com

The influence of environmental disturbances on the emergence of disease outbreaks is increasingly evident. Geneticists and those in related fields should disseminate the idea that environmental equilibrium is a public health issue.

Joel Henrique Ellwanger
Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 91501-970, Brazil. Email: joel.ellwanger@gmail.com

Scientists can show the value of nature by explaining that healthy river valleys provide clean water and prevent flooding.

Inspire with stories and imagery

Stories can promote environmental stewardship by bridging the gap between scientific knowledge and public engagement. Researchers can conduct joint field trips with members of the public where they mutually share stories about human activities and the natural environment. The stories will help instill curiosity about the environment, while the field trip will promote scientific collaboration. The natural environment provides an excellent backdrop for storytelling through the diverse audio-visual and olfactory stimuli. Indigenous communities have many traditions involving storytelling while experiencing nature. The stories help the communities to value the natural environment and foster behaviors that promote conservation.

Edmond Sanganyado
Marine Biology Institute, Shantou University, Shantou, Guangdong 515063, China.
Email: esang001@ucr.edu

Humans harm nature unwittingly, so conservation science should encourage people to act sustainably even without thinking. This will not be achieved with data alone because automatic cognitive processes are not formed in the rational parts of our brains. People need to be inspired by conservation science, not put off by constant messages of doom and gloom. We can do this by framing our hard-earned data in metaphors portraying nature as resilient and formidable, rather than fragile and controllable. Perhaps it's time that conservation scientists add poetry and art to our studies of plants and animals.

Falko Buschke
Centre for Environmental Management, University of the Free State, Bloemfontein 9300, South Africa.
Email: falko.buschke@gmail.com

Technology helps archaeologists document the complexity of change over intervals of centuries using time-lapsed animations and three-dimensional visual mapping. It is now possible to align environmental perturbations, sea-level changes, droughts, and El Niños with man-made alterations to landforms, then on top of that overlay oral histories that narrate these changes in generational time. These composite images captivate our visual and auditory senses, and we should share this experience with the larger community. Connections created through social media provide the ideal forum for this—advancing a global outreach for science while inviting benefactors, citizen scientists, and young explorers to experience our changing world for themselves.

Felicia Rounds Beardsley
Department of Sociology and Anthropology, University of La Verne, La Verne, CA 91750, USA.
Email: fbeardsley@laverne.edu

By developing robust models to understand environmental trends and forecast how current unchecked practices will affect the natural environment, data scientists can present a well-reasoned, quantitative analysis and inspire people to be more engaged in sustainable and environment-friendly practices. Data scientists are uniquely poised to harness the power of patterns and tell compelling stories about nature. Statistical models might not be able to change people's reliance on technology, but they can certainly use that reliance to heighten our awareness about our natural environment.

Divyansh Agarwal
Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: divyansh@upenn.edu

We have unprecedented access to monumental volumes of data due to the increasing ubiquity of sensors for everything from air quality to ocean temperature. However, data alone are not enough to change public perception. More important than our ability to collect information is our ability to present it. The XY plots and bar graphs ubiquitous in the scientific literature are not visually compelling to most people. To more effectively communicate with the public, we must move past simple plots and work toward creating stunning visualizations that accurately convey data in context as part of compelling narratives.

Mark Martin Jensen
Department of Biomedical Engineering, University of Utah, Salt Lake City, UT 84112, USA.
Email: m.martin.jensen@utah.edu

Appeals and images that trigger emotional responses in people—a skinny polar bear or an oily tern—may have some impact, but to really get the public to consider the value of the natural world, it helps to engage their curiosity, too. Public engagement activities should not patronize but rather provide new information and the opportunity to think and problem-solve. Activities should leave the open questions in science open. When people are truly engaging with scientific problems, they can see both the value of

TOMORROW'S EARTH
Read more articles online at scim.ag/TomorrowsEarth

ILLUSTRATION: ADAM SIMPSON/HEART AGENCY

curiosity and the need for action to preserve and protect the natural environment.

Timothy L. Easun
School of Chemistry, Cardiff University, Cardiff CF10 3AT, UK. Email: easuntl@cardiff.ac.uk

Support accountability

Corruption, inefficiency, and inadequate spending undermine efforts to steer humanity toward sustainability. Blockchain, a secure method of storing decentralized data, could be used to audit the environmental impact of firms or consumers in real time. The reliability, security, and auditability of data stored through blockchain can ensure that actors are held accountable, which could transform how organizations and individuals fund and execute environmental programs.

Han Lin
School of Information Engineering, Jiangsu Key Laboratory of Auditing Information Engineering, Nanjing Auditing University, Nanjing, Jiangsu 211815, China. Email: linhan@nau.edu.cn

Real-time feedback and interaction are powerful in promoting environmentally friendly behaviors. Through advanced sensing and measurement technologies in the Internet of Things, we can collect the various fine-grained behavioral data of people's daily life and activities and then measure their environmental impact with big data analytics. Wearable devices or mobile apps can in turn provide real-time feedback and personalized recommendations.

Kaile Zhou
School of Management, Hefei University of Technology, Hefei, Anhui 230009, China, and City University of Hong Kong, Kowloon, Hong Kong SAR. Email: zhoukaile@hfut.edu.cn

Identify systemic barriers

If scientists, in any field, are worried about the preservation of the environment, they would do well to point out the structural causes of environmental degradation. Roughly half of the human inhabitants of Earth produce no net carbon emissions. For the rest of the world's human population, a substantial part of the pollution output comes from areas in which people have only a marginal say, such as where they can afford to live and how their electricity is produced. These problems require systemic solutions, such as increased production of dense public housing (not detached single-family homes) situated along public transit lines and the elimination of fossil fuel–powered cars and power plants. The primary opposition to such endeavors does not come from average people who do not care about the environment, but rather from powerful corporations and the handful of obscenely wealthy people who own them. This is what scientists should be making clear to the public.

E. Joseph Jordan
Department of Biochemistry and Biophysics, Stockholm University, 171 65 Solna, Sweden. Email: e.jjordan12@gmail.com

Behavior analysts have been developing research on pro-environmental behaviors since the 1970s. Simple arrangements such as manipulating the location and number of trash bins and providing more information about recycled items in schools, offices, and universities have been shown to promote considerable increase in pro-environmental behaviors. Scientists could conduct more research on the topic and also focus on the dissemination of knowledge of effective strategies.

Fernanda Suemi Oda
University of Houston–Clear Lake, Houston, TX 77058, USA. Email: fsuemioda@gmail.com

Behavior analysis indicates that small changes, such as making trash bins more accessible, can increase sustainable behavior.

The people of the world are presently facing a future where the prospects for their continued survival under any kind of decent conditions are extremely bleak. The reason we are in this position is not because individuals do not care enough about the value of the environment, but because our economic system is built from the ground up on rapidly growing consumption fueled by the extraction of finite resources, with absolutely no regard for the environmental consequences. Individuals have next to no power to escape such a system, much less create a new one that is compatible with our long-term survival. Environmental causes already have widespread support, but they require a level of global cooperation that has been very easy to undermine politically. The focus on appreciating nature diverts attention away from a much-needed critique of our unsustainable global economic system.

Harry MacKay
Department of Pediatrics, Baylor College of Medicine, Houston, TX 77054, USA. Email: harry.m@gmail.com

Engage educators and the media

As a former researcher and current high school teacher, I think the single most important thing that practicing scientists in any field can do is reach out to teachers and students at a local high school. Many teachers may be interested in having a scientist speak to their students, or in taking a field trip to an ecological, evolutionary, or plant biology laboratory, or in a nature walk guided by a local ecologist. But time is precious, and teachers may not know who is doing relevant work in their area. Scientists could even arrange a Skype call or offer to answer student questions about their scientific journeys. Scientists offer a unique perspective, and teachers would love to collaborate to share that perspective with their students.

Erin Coffey
Trinity School, New York, NY 10024, USA. Email: erin.coffey@trinityschoolnyc.org

Biologists and environmental educators have incredible tools to facilitate authentic engagement. However, the science community, including funding agencies, does not provide adequate support for science communication and evidence-based outreach and engagement plans. People need the opportunity to experience nature as it works for them and fits into their lives. Nature-based apps, virtual reality experiences, and citizen science projects show how technology can serve as a supplement to a flexible and adaptable experience rather than a competing demand.

Rachel Yoho
Department of Biology, Miami University, Oxford, OH 45056, USA. Email: yohora@miamioh.edu

When I describe to people how I feel about the inherent beauty of nature, they listen. Once they share my enthusiasm, I take that opportunity to talk about the importance of biodiversity. To spread the message further, scientists who feel comfortable doing so should use every opportunity to make others aware of nature, including speaking to newspaper and radio journalists, giving a TED talk, giving talks to laypeople, and even creating nature-inspired music, as I have.

Marten Winter
German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Synthesis Centre for Biodiversity Sciences (sDiv), 04103 Leipzig, Germany. Email: marten.winter@idiv.de

10.1126/science.aav9402

BEHAVIOR

How behavioral science can help conservation

Leveraging cognitive biases and social influence can make conservation efforts more effective

By **Joshua Cinner**

Most conservation initiatives require changes in human behavior. For example, the establishment of a protected area will typically require some people to change their land-use or fishing practices. Yet conventional attempts to encourage pro-environmental behavior through awareness campaigns, financial incentives, and regulation can prove ineffective (*1*, *2*). Insights into inducing behavior change from the social and behavioral sciences are therefore of critical importance for conservation scientists and practitioners (*2*–*4*). Conservation initiatives have begun to leverage a wide range of such behavioral insights (*5*) particularly regarding cognitive biases and social influence (see the figure). However, their application in the diverse socioeconomic and cultural contexts in which many conservation programs operate raises important ethical and implementation-related challenges.

LEVERAGING COGNITIVE BIASES

Numerous cognitive biases affect the conscious and unconscious decisions that humans make (*6*, *7*). These biases can make people behave in seemingly strange but predictable ways that have important implications for conservation (see the figure).

People have a strong tendency to avoid making difficult decisions, and as a result, they are prone to accepting whatever default option they are presented with—even when this option is not in their own, or society's, best interest. This status quo bias means that if people are asked to opt into a conservation program voluntarily (such as choosing electricity generated from renewables), they most likely won't, even if they think it is a good idea. Participation in programs intended to promote energy efficiency or green energy increases markedly when these options are the customer's default settings (*5*). Likewise, a university was able to reduce paper use by 15% simply by switching the default printer settings to double-sided (*5*). Of course, people should always be able to opt out (*6*).

People also have a cognitive bias that causes them to disproportionately weight initial information when making decisions (*7*). When a group of students were first asked to write down the last two digits of their social security number before estimating how much they would be willing to pay for a range of goods (such as chocolate and wine), those with the highest last two digits of their social security numbers were willing to pay three times as much as those with the lowest (*7*). The social security number, even though it was arbitrary and irrelevant, acted as an anchor—a starting point for decision-making under uncertainty.

This anchoring bias has conservation applications. Many fisheries do not allow fishers to retain fish below a minimum (usually reproductively viable) size. This minimum size often becomes the anchor for fishers, but it is easy to slip just below this anchor and keep fish that are too small. To reduce the likelihood of recreational fishers keeping fish that are too small, management authorities have used persuasive messaging that encourages people to aim for larger fish to reset people's conceptual anchor well above the minimum size (*8*).

Although these cognitive biases have already been leveraged in conservation activities spanning fisheries, energy use, waste production, and land use (*5*, *8*), further opportunities exist to integrate people's cognitive biases directly into operational tools used for conservation planning. For example, there is a cognitive bias that causes people to perceive that losses hurt about twice as much as gains feel good, often referred to as loss aversion or prospect theory (*6*). Yet, when determining where to place protected areas, conservation planners tend to give the same weight to potential losses from establishing a reserve in a specific location as to potential gains. To more accurately reflect people's loss aversion bias, systematic conservation planning could assign a heavier weight to losses (this could even be directly integrated into the software used to plan locations of reserves).

Likewise, loss aversion bias could be integrated into the framing of incentive-based conservation activities to improve motivation and participation. Many conservation issues are framed to highlight the benefits associated with sustainable resource use (*9*). Yet because of people's loss aversion bias, motivations to use resources sustainably can be higher when issues are framed to highlight the potential losses from not doing so, rather than the gains. A high-tech manufacturing company was able to increase output by experimentally manipulating the framing of bonuses. Total team productivity increased by 1% when bonuses were framed as a loss (a deduction from a maximum when benchmarks were not met) compared to when they were framed as a gain (a bonus for meeting a benchmark), even though the bonus system was actually the same (*10*). A similar type of loss-framing could be built into incentive-based conservation programs that provide rewards for achieving sustainability benchmarks, such as reforestation or curbing non–point source pollution, to help motivate participation (*11*).

TOMORROW'S EARTH

Read more articles online at scim.ag/TomorrowsEarth

Additionally, conservation interventions could also leverage what is known as the decoy effect. The decoy effect is the phenomenon that people tend to change their preference between two options when presented with a third option that is meant to be inferior in some regard (a decoy). A third option can thus be used to make one of the initial options seem more attractive. The decoy effect is often used in car sales and marketing to get people to purchase a more expensive product, but recent applications have shown that it can also be harnessed to promote cooperation (*12*). When faced with the options to cooperate or not, the presence of a third option—cooperate and reward someone else—vastly improved cooperation rates, even though this third option was rarely used (*12*). Some voluntary conservation initiatives could leverage the decoy effect to improve participation and cooperation.

ARC Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, QLD 4811, Australia. Email: joshua.cinner@jcu.edu.au

LEVERAGING SOCIAL INFLUENCE

Human behavior is also profoundly influenced by our innate desire for prestige, reputation, conformity, and reciprocity (see the figure). Social influence refers to the ways in which our decisions and actions are shaped by perceptions (whether accurate or not) of what other people do and what they approve of (*1*, *4*).

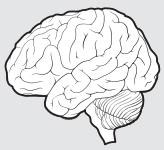
Some power and water companies have managed to dramatically reduce resource use simply by including in bills a comparison between the usage rates of the customer, their neighbors, and the most efficient users (*9*). This comparison conveys social norms about both what is right to do and what other people are doing (*2*) and enables feelings of agency and self-control (*1*). People tend to respond best to such comparisons when they feel that the comparison groups are like themselves. To date, social comparisons have generally been targeted at individual customers, but they offer the potential for scaled-up impact when they are targeted at leaders, policymakers, and even engineers. One experiment found that project designs were 28% more sustainable when civil engineers had social norms about sustainability reinforced just prior to starting the project (*13*).

People also behave differently when they think they are being observed (*4*). Some power companies have leveraged this knowledge by making customers' power usage public, resulting in substantial reductions in electricity consumption (*3*). Likewise, prosocial behavior such as compliance with conservation rules can be encouraged when stylized images of watching eyes are visible (i.e., on signs or posters), which make people feel as if they are being watched (*8*, *14*). However, it is unclear how long the effects of these types of behavior change intervention last, and it remains to be shown whether people eventually get desensitized.

Of course, we are also influenced by the source of our information. Spreading conservation ideas and practices can be facilitated through use of trusted messengers (i.e., popular actors, athletes, or public figures), key players in social networks, and block leaders (volunteers who help to inform people about an issue) (*1*, *15*). Other social influence levers include obtaining publicly visible commitments or pledges to change behavior—for instance, to start recycling (*1*, *3*). Inducing behavioral change through social influence levers works best when the behavior is visible—such as curbside recycling or hiking on as opposed to off the trail—and when there is strong social group cohesion, communication, and identity (*1*, *4*).

Leveraging behavioral insights for conservation

Cognitive biases



The status quo bias
Most people prefer to maintain the status quo. This can be addressed by setting the default option so that people need to "opt out" rather than "opt in" to sustainable options.

Anchoring
People tend to rely on initial information. This bias can be leveraged by setting cognitive anchors early and far from critical thresholds.

Issue framing
People have a strong aversion to losses. Highlighting what they stand to lose by unsustainable practices and policies helps to catalyze action.

Decoys
When people have trouble making decisions, the desirability of sustainable options can be emphasized with the use of less desirable "decoy" options.

Social influences

Social norms
People want to fit in with what "most people do" and what "should be done." Communicating social norms about conservation can help to encourage sustainable behaviors.

Observability
People behave more prosocially when they think others know what they are doing. Increasing observability can promote sustainable behaviors.

Block leaders
Whom we receive information from can be as powerful as what we receive. Trusted messengers and block leaders can amplify uptake.

Public commitments
People want to maintain prestige and reputation, which can be leveraged through public commitments or pledges to change behavior.

IMPLEMENTATION CHALLENGES

Behavioral insights have been used to inform a broad range of public policies, including health care and tax compliance. They also hold much promise for conservation, but applications to date have yielded mixed results (*5*), and critical questions remain about the context under which they can be most effective (*2*). Integration of behavioral insights into a wider range of conservation issues will require the development of an operational framework that helps practitioners to determine the types of behavioral interventions that might work best and be most appropriate in a particular context. Different types of implementing organizations, such as private companies, government agencies, NGOs, and community-based organizations, have different rights and degrees of social license to implement behavioral interventions. The best choice of intervention may also depend on whether the targets are consumers, resource users, leaders, policy-makers, project designers, or other types of stakeholders. Likewise, certain types of interventions may be more effective depending on whether the target behavior is a one-off or repeated decision; costly or cheap; under high or low uncertainty; visible or discrete; and whether the resource being conserved is public, private, or common-pool. Finally, sociocultural contexts such as attitudes, beliefs, norms, prices, and policies can all influence people's behavior and may help to inform the types of behavioral interventions that may work best (*2*, *3*).

Most behavioral interventions, and even the behavioral studies upon which they are founded, have been conducted in developed countries (*1*) and may not necessarily reflect people's behavior in the diverse social, economic, and cultural settings where many conservation programs operate. It is therefore crucial that researchers test many of the behavioral insights in field settings. These applications will, however, raise new ethical challenges, particularly when applied to traditionally marginalized peoples who may lack the agency to opt out of behavior change interventions. The development of a code of conduct may help to ensure that behavior change interventions are not viewed as coercive, a perception that could ultimately undermine support for conservation initiatives (*2*). Further research is also needed to explore the potential for behavior change interventions to displace certain norms and motivations, such as engaging in conservation because it is "the right thing to do" (*2*). ■

REFERENCES AND NOTES

1. W. Abrahamse, L. Steg, *Global Environ. Change* **23**, 1773 (2013).
2. S. M. W. Reddy *et al.*, *Conserv. Lett.* **10**, 248 (2017).
3. E. Yoeli *et al.*, *Behav. Sci. Pol.* **3**, 68 (2017).
4. K. Nyborg *et al.*, *Science* **354**, 42 (2016).
5. H. Byerly *et al.*, *Front. Ecol. Environ.* **16**, 159 (2018).
6. R. H. Thaler, *Am. Econ. Rev.* **108**, 1265 (2018).
7. D. Ariely, G. Loewenstein, D. Prelec, *Q. J. Econ.* **118**, 73 (2003).
8. M. Mackay, S. Jennings, E. I. van Putten, H. Sibly, S. Yamazaki, *Mar. Policy* **95**, 256 (2018).
9. S. Díaz *et al.*, *Science* **359**, 270 (2018).
10. T. Hossain, J. A. List, *Manage. Sci.* **58**, 2151 (2012).
11. P. Ferraro, K. D. Messer, S. Wu, *Choices* **32**, 1 (2017).
12. Z. Wang *et al.*, *Nat. Commun.* **9**, 2954 (2018).
13. T. Shealy *et al.*, *Sustain. Cities Soc.* **40**, 182 (2018).
14. S. Pfattheicher, J. Keller, *Eur. J. Soc. Psychol.* **45**, 560 (2015).
15. E. K. Mbaru, M. L. Barnes, *Biol. Conserv.* **210**, 222 (2017).

ACKNOWLEDGMENTS

I thank M. Barnes, J. Kittinger, G. Gurney, E. D. Cinner, C. Rojas, and S. Gelcich for helpful comments. Supported by the Australian Research Council (CE140100020, FT160100047), the Pew Charitable Trust, and the Paul M. Angell Family Foundation.

10.1126/science.aau6028

GRAPHIC: N. CARY/*SCIENCE*

CELL BIOLOGY

Cellular networks in wound healing

Macrophage and fibroblast heterogeneity and plasticity control tissue fibrosis

By **Sebastian Willenborg**[1] *and* **Sabine A. Eming**[1,2,3]

Restoration of tissue integrity and homeostasis after injury is fundamental to most organisms. Throughout evolution, there is diversity in whether the healing response replaces the damaged tissue through regeneration or by depositing collagenous connective tissue, defined as fibrosis. The molecular basis that underlies the repair response is complex. The postnatal decline of regenerative capacity in mammals as well as the regenerative heterogeneity among diverse organs is an unresolved obstacle in medicine. Limited fibrosis to temporarily stabilize newly forming tissue is important for healing but ultimately impairs tissue function. Disorders characterized by excessive fibrosis—including interstitial lung disease; fibrosis of the liver or kidney; sclerosing skin diseases; and localized fibrotic manifestations associated with, for example, late stage venous insufficiency, skin fragility disorders, trauma, or cancer—contribute considerably to patient discomfort and morbidity as well as high numbers of deaths worldwide. Nonetheless, progression in the development of antifibrotic therapeutics is slow (*1*). On page 909 of this issue, Shook *et al.* (*2*) identify a critical role of fibroblast and immune cell heterogeneity and communication in promoting efficient skin wound healing. These findings add to the understanding of fibrosis and could guide us toward better treatments for fibrosing diseases.

Skin wound healing is a stringently organized process that proceeds through a series of dynamic overlapping immune responses associated with migration and proliferation of different cell types, extracellular matrix (ECM) deposition, and tissue remodeling (*3*) (see the figure). Fibroblasts derived from the lower, reticular dermis represent the first wave of myofibroblasts in skin wound healing; they produce α–smooth muscle actin and high amounts of collagen, which are well-established hallmarks of repair (*4*, *5*). In addition, various phenotypes of macrophages have been recently identified as pivotal in the quality of repair during the healing response; potentially, they are important therapeutic targets (*6*).

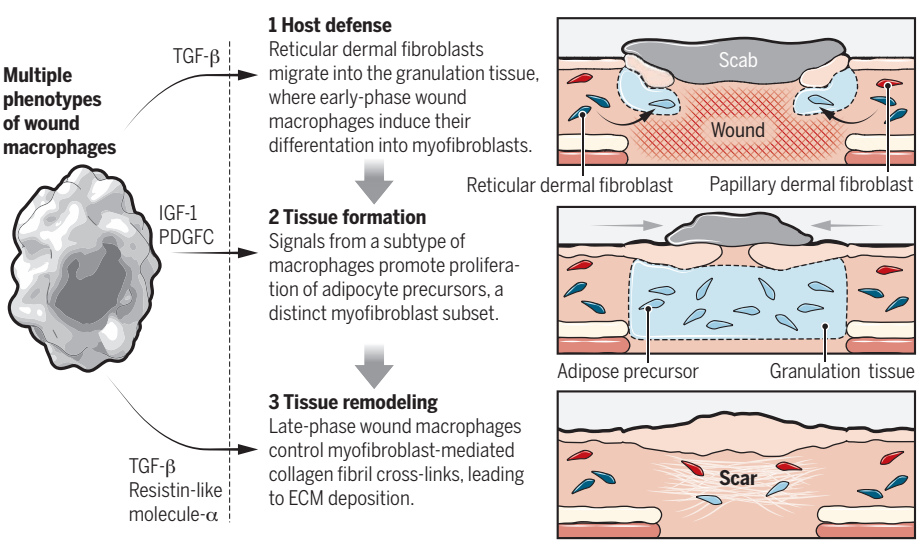
Shook *et al.* identified a distinct subpopulation of myofibroblasts [which are precursors to adipocytes (fat cells)] that proliferate at the wound site. This proliferation is regulated by a distinct population of activated CD301b$^+$ macrophages. Yet the ultimate consequences of this specific cellular cross-talk for tissue morphology and function merit further investigation. Supposing the interaction between specific macrophage and myofibroblast subpopulations is critical for the outcome of the repair response; then is there a window during which the harmful profibrotic cascade can be interrupted or even reversed to improve healing? In soft connective tissue healing in mammals, during the macrophage-fibroblast cross-talk, profibrotic programs, such as collagen–cross-linking pathways that usually only occur in hard connective tissue such as bone and cartilage, become activated and lead to scarring (*7*). Regulation of this transition from early to late tissue ECM consolidation is little understood and the subject of investigation (*8*). Probably, the combination of different experimental model systems will pave the way to address these challenging questions; interesting approaches are being developed (*9*, *10*).

The current view suggests that the healing response in most tissues follows the sequence of events described in skin repair. Skin repair is perhaps the best studied because of its accessibility. Shook *et al.* used skin wounding as a model to investigate the role of fibroblast–immune cell interactions in tissue repair. Whether their findings represent a general principle of fibroblast–immune cell interactions in other systems is an important question, especially because tissue- and organ-specific mechanisms of fibrosis have recently been described (*11*).

A complex network of multiple cytokines and mediators facilitates communication between macrophages and myofibroblasts. We need to understand whether there is a hierarchical framework that induces and controls fibrosing pathways. Transforming growth factor–β (TGF-β) and type 2 immune responses have been identified as two of the core drivers of tissue repair and fibrosis. Interestingly, *Mgl2* gene expression (which encodes CD301b) in macrophages as identified by Shook *et al.* is induced by type 2 cytokines (*6*). As previously shown, wound-associated macrophages activated by type 2 cytokines initiate a profibrotic

Dynamics of macrophage-fibroblast cross-talk in tissue repair

In the skin, wound macrophages within granulation tissue acquire a spectrum of phase-specific phenotypes that affects fibroblast function, including myofibroblast differentiation, myofibroblast numbers, and ECM remodeling. IGF-1, insulin-like growth factor 1; PDGFC, platelet-derived growth factor C.



GRAPHIC: C. BICKEL/*SCIENCE*

[1]*Department of Dermatology, University of Cologne, Kerpenerstraße 62, 50937 Köln, Germany.* [2]*Center for Molecular Medicine Cologne, University of Cologne, 50931 Cologne, Germany.* [3]*Cologne Cluster of Excellence on Cellular Stress Responses in Aging-Associated Diseases, University of Cologne, 50931 Cologne, Germany. Email: sabine.eming@uni-koeln.de*

pathway that induces in fibroblasts the expression of *Plod2* [which encodes lysyl hydroxylase 2 (LH2)], a critical enzyme in the formation of profibrotic collagen cross-links (*7*). Intriguingly, Shook *et al.* found that adipocyte precursors highly express *Plod2*. How type 2 immune responses and profibrotic signaling pathways connect, specifically with regard to macrophage and myofibroblast heterogeneity, is far from understood. This challenging biological complexity is not specific for fibrosing diseases, and moving toward a cytokine-based disease taxonomy, as previously suggested for other immune-mediated diseases (*12*), may benefit the wound healing field.

The study of Shook *et al.* also raises interesting questions about the role of adipocyte-related cells in tissue repair. Recent studies in fruitflies show that adipocytes might be important for preventing wound infection and clearing cell debris, thus expanding the potential consequences of immune cell–adipocyte communication (*13*).

How can we translate the knowledge uncovered by Shook *et al.* to the clinic? The authors found differences in the heterogeneity of fibroblast–immune cell subpopulations in wound repair in young mice, in ageing-impaired repair, and in skin fibrosis, indicating a correlation between cellular heterogeneity and fibrotic outcome in mice. A critical step will be to investigate cellular heterogeneity in human fibrosing disorders. Shook *et al.* identified a similar composition of fibroblast subsets in uninjured human skin compared with mouse skin, with human fibroblasts resembling a more profibrotic phenotype. Interestingly, they also found a high abundance of CD301^{+} macrophages in keloid scars, a pathological scarring entity in human skin, supporting a possible clinically-relevant profibrotic function of this cell type. Ongoing antifibrosis clinical trials, which target some of the outlined pathways such as type 2 immune responses, highlight the exciting possibility that we can target tissue fibrosis to improve organ function after injury (*14*, *15*). ■

REFERENCES

1. T. A. Wynn, *Nat. Med.* **18**, 1028 (2012).
2. B. A. Shook *et al.*, *Science* **362**, eaar2971 (2018).
3. T. J. Shaw, P. Martin, *Curr. Opin. Cell Biol.* **42**, 29 (2016).
4. P. Pakshir, B. Hinz, *Matrix Biol.* **68–69**, 81 (2018).
5. R. R. Driskell *et al.*, *Nature* **504**, 277 (2013).
6. P. J. Murray, *Annu. Rev. Physiol.* **79**, 541 (2016).
7. J. A. Knipper *et al.*, *Immunity* **43**, 803 (2015).
8. M. Pensalfini *et al.*, *Acta Biomater.* **65**, 226 (2018).
9. R. Richardson *et al.*, *J. Invest. Dermatol.* **133**, 1655 (2013).
10. J. L. Morris *et al.*, *Dev. Biol.* **441**, 4 (2018).
11. D. R. Lemos, J. S. Duffield, *Sci. Transl. Med.* **10**, eaan5174 (2018).
12. G. Schett *et al.*, *Nat. Med.* **19**, 822 (2013).
13. A. Franz *et al.*, *Dev. Cell* **44**, 460 (2018).
14. A. Nyström *et al.*, *EMBO Mol. Med.* **7**, 1211 (2015).
15. O. Kowal-Bielecka *et al.*, *Ann. Rheum. Dis.* **76**, 1327 (2017).

10.1126/science.aav5542

PHOTO: WALTER STEIN/GETTY IMAGES

PALEOECOLOGY

The decline of Africa's largest mammals

Did hominins play a role in the loss of megaherbivores?

By **René Bobe**[1,2,3] *and* **Susana Carvalho**[1,2,3,4]

The human species is causing profound climatic, environmental, and biotic disruptions on a global scale. In the present time (now called the Anthropocene), most species of large terrestrial herbivores are threatened with extinction as their populations decline and their geographic ranges collapse under the pressure of human hunting, poaching, and encroachment (*1*). Although the scale of ongoing anthropogenic ecological disruptions is unprecedented, human-driven extinctions are not new: There is strong evidence that humans played a major role in the wave of megafaunal losses at the end of the Pleistocene, between about 10,000 and 50,000 years ago (*2*). But when did humans, or our ancestors, begin to have such a profound effect on large herbivores to the point of causing extinctions? On page 938 of this issue, Faith *et al.* (*3*) provide evidence to help answer this question. They track the number of megaherbivore species (mammals weighing more than 1000 kg) in eastern Africa from the late Miocene, about 7 million years ago, to the present. Their analysis indicates that megaherbivore diversity began to decline in the early Pliocene, about 4.6 million years ago.

Hominins—species on our side of the evolutionary divergence that separated us from the chimpanzees—first appeared in Africa in the late Miocene, about 7 million years ago. The late Miocene was a time of global climatic and environmental change, with an expansion of grasslands (with C_4 photosynthesis) in tropical latitudes and an increasing frequency of fires (*4*). At that time, large mammals were abundant in eastern Africa. At Lothagam, for example, in the Turkana Basin of Kenya, there were 10 species of megaherbivores (*5*), including the earliest records of the elephant family, Elephantidae. Near Lothagam is the early Pliocene site of Kanapoi, with a rich fossil record dated to about 4.2 million years ago. Kanapoi has the earliest record of the hominin genus *Australopithecus* (*6*), which coexisted with at least 11

[1]*Primate Models for Behavioural Evolution Lab, Institute of Cognitive and Evolutionary Anthropology, School of Anthropology, University of Oxford, Oxford, UK.* [2]*Gorongosa National Park, Sofala, Mozambique.* [3]*Interdisciplinary Center for Archaeology and Evolution of Human Behaviour (ICArEHB), Universidade do Algarve, Portugal.* [4]*Centre for Functional Ecology–Science for People & the Planet, Universidade de Coimbra, Portugal. Email: rene.bobequinteros@anthro.ox.ac.uk*

African elephants (*Loxodonta africana*), such as this matriarch from Gorongosa National Park in Mozambique, are survivors from the deep past.

species of megaherbivores: five proboscideans, two rhinocerotids, two giraffids, and two hippopotamids, along with a diverse fauna of large carnivores that included giant otters, hyenas, two species of saber-tooth felids, and three species of crocodiles (*7*). Kanapoi, at about 4.2 million years ago, with an area of 32 km^2, had twice the number of megaherbivore species as the entire continent of Africa has today. Among the five species of proboscideans, there were the ancestors of the modern African and Asian elephants, *Loxodonta* and *Elephas*, respectively. From their first appearance in eastern Africa, both elephants fed predominantly on increasingly abundant grasses (*8*). It is probable that these elephants and other megaherbivores played a beneficial role for early hominins by opening up wooded environments, thereby resulting in the mix of woodlands and grasslands where hominins seemed to thrive (*5*, *9*). Thus, at Kanapoi and other sites in Africa during the Pliocene (5.3 to 2.6 million years ago) early hominins lived in proboscidean-dominated landscapes. These hominins were relatively small (~45 kg) and left no known traces of tool-assisted behaviors. They were primarily terrestrial omnivores that used trees for refuge and food, and it is likely that they competed for resources with other mammals like suids and monkeys, but had little impact on megaherbivores. Moving forward in time from 4.2 million years ago to 1.5 million years ago, three hominin species, including *Homo erectus*, possibly *Homo habilis*, and *Paranthropus boisei*, inhabited another set of sites in the Turkana Basin. By then, hominins were procuring animal resources by using stone tools, as the abundant archaeological evidence makes clear, and they were probably beginning to use fire (*10*). There is no evidence of systematic hunting of very large mammals, so direct effects of hominins on megaherbivores were limited. But if hominins were beginning to use fire 1.5 million years ago, what was the impact on the landscape and on the foraging patterns of the largest terrestrial mammals?

Early hominins in eastern Africa lived in a rapidly changing landscape, in which grasslands were becoming more prevalent, and a whole suite of grazing mammals were rapidly evolving and diversifying (*11*). *Elephas* was successfully adapting to these changing African landscapes during the Pleistocene when it became a dominant megaherbivore. But even this widely distributed and abundant grazer became extinct in Africa toward the late Pleistocene, surviving in Asia until today. *Loxodonta* became rare after the early Pliocene, only to reemerge in the middle and late Pleistocene as a browser and mixed-feeder rather than a grazer (*8*).

Faith *et al.* propose that expanding grasslands, declining concentrations of atmospheric CO_2, and competition with smaller browsing mammals all contributed to the decline of megaherbivore species richness. However, it is not clear what ecological roles hominins played throughout the long evolutionary history of megaherbivores in Africa, and how these roles changed over time and varied across geographic space. Another question is when hominins became systematic predators of animals larger than themselves. Could periods of marked seasonality, with peaks of drought, result in a "fallback" predation strategy, whereby carnivores preyed heavily on proboscideans? As documented in modern African landscapes (*12*), predation on young, vulnerable individuals in species with very slow life histories could have irreparable effects on the long-term viability of megaherbivore populations. As fires became more frequent beginning in the late Miocene (*4*), they may have altered landscapes and affected the foraging, ranging patterns, and dispersal of large herbivores. Also, what new pathogens emerged with the new landscapes of eastern Africa and what was their effect in the oldest and youngest individuals within these large mammal populations? The causes of megaherbivore decline are probably complex, multidimensional, and varied across time and space. The precise timing of key hominin behavioral innovations remains poorly constrained by the current archaeological and paleontological records. Indeed, the role that hominins played is still open to question. ■

REFERENCES AND NOTES

1. W. J. Ripple *et al.*, *Sci. Adv.* **1**, e1400103 (2015).
2. C. Sandom, S. Faurby, B. Sandel, J.-C. Svenning, *Proc. R. Soc. B* **281**, 20133254 (2014).
3. J. T. Faith, J. Rowan, A. Du, P. L. Koch, *Science* **362**, 938 (2018).
4. W. J. Bond, *Front. Plant Sci.* **5**, 749 (2015).
5. L. Werdelin, W. J. Sanders, Eds., *Cenozoic Mammals of Africa* (Univ. California Press, Berkeley, 2010).
6. C. V. Ward, J. M. Plavcan, F. K. Manthi, *J. Hum. Evol.* 10.1016/j.jhevol.2017.07.008 (2018).
7. J. M. Harris, M. G. Leakey, Eds., *Geology and Vertebrate Paleontology of the Early Pliocene Site of Kanapoi, Northern Kenya* (Natural History Museum of Los Angeles County, Los Angeles, 2003), vol. 498.
8. T. E. Cerling, J. M. Harris, M. G. Leakey, *Oecologia* **120**, 364 (1999).
9. J. Kingdon, in *Mammals of Africa*, J. Kingdon *et al.*, Eds. (Bloomsbury, London, 2013), vol. 1, pp. 75–100.
10. G. L. Isaac, B. Isaac, Eds., *Plio-Pleistocene Archaeology: Koobi Fora Research Project* (Clarendon, Oxford, 1997), vol. 5.
11. T. E. Cerling *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11467 (2015).
12. A. J. Loveridge, J. E. Hunt, F. Murindagomo, D. W. Macdonald, *J. Zool. (Lond.)* **270**, 523 (2006).

ACKNOWLEDGMENTS

We thank J. Kingdon and D. R. Braun for discussions on topics presented here.

10.1126/science.aav6883

CANCER

The paradox of mutations and cancer

Esophageal tissue often contains driver mutations in people without cancer

By **Stephen J. Chanock**

The past decade has witnessed the cataloging of genetic mutations in cancer genomes, providing new insights into how and in what ways cancer can develop and spread (*1*, *2*). The focus has been on defining specific "driver" mutations, genetic errors in cancer cells that reveal basic biological processes gone awry that are required for cancer initiation and progression. These drivers are the target of new therapies—this concept is central to precision oncology efforts to treat patients according to the genetic changes that are present in their tumors (*3*). Along the way, it has also become apparent that cancer genomes harbor many additional "passenger" mutations (*4*). Patterns of driver and passenger DNA mutations derived from cancer genomes have provided clues about the different ways that cancer can manifest as a disease of genetic mutations (*5*, *6*). In some circumstances, they can be linked to strong environmental carcinogens (for example, mutation patterns caused by tobacco smoke, ultraviolet radiation, or the fungal toxin aflatoxin) (*7*). Moreover, these forensic mutational patterns can be used to estimate how long it has taken for a tumor to develop (*5*). On page 911 of this issue, Martincorena *et al.* (*8*) turned their atten-

"These findings suggest that transition from healthy tissue to cancer may be more complicated than we have modeled."

Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD 20890, USA. Email: chanocks@mail.nih.gov

tion toward the detection of mutations in normal tissue, addressing a long-standing paradox that mutations arise in normal tissues but do not necessarily lead to cancer.

Martincorena *et al.* found that a larger-than-anticipated fraction of cells of the esophagus harbor mutations in cancer driver genes, which increases with age. Although only nine individuals without evidence of cancer were studied, the results raise two important questions. What is the role of mutations in tumorigenesis, and what factors are required to progress to cancer, including additional mutations or local events in the microenvironment or immune system? The findings also reveal that genome integrity deteriorates with age and perhaps in tissue-specific ways, which could reflect distinct external exposures and tissue-specific properties.

The authors compared the genome sequences of nondiseased esophagus and skin (*9*) to mutations in 74 known cancer driver genes. The total number of mutations was higher in the skin, which is not unexpected because the outer layers of skin are continually exposed to a strong mutagen, ultraviolet light. However, it came as a surprise that esophageal epithelial cells have a higher number of driver mutations, particularly mutations in the *NOTCH1* and *TP53* genes, which often occur in esophageal squamous carcinoma (*10*).

It is intriguing to speculate how such driver cancer mutations might be induced in the esophagus. The authors propose that the accumulation of mutations increases with age, mainly owing to errors in intrinsic processes such as aberrant DNA replication or repair. It is also plausible that the complex mix of what we swallow daily could contribute to the high mutation burden. The cumulative lifetime probability of developing esophageal squamous carcinoma in the United States is about 1 in 125, which is substantially smaller than that which would be predicted by the occurrence of the observed *NOTCH1* and *TP53* mutations. This raises the question of what other events are needed to induce cancer. Alternatively, perhaps there are mechanisms that prevent cancer from developing.

These findings suggest that the transition from healthy tissue to cancer may be more complicated than we have modeled (see the figure). A next step will be to survey the entire genome (rather than only selected driver genes), where we may find mutations in new genes as well as the elusive "dark matter" between genes. It is important to focus on this intergenic space to understand how regulation of the genome is disrupted through mutations or chemical modification, a hallmark of epigenetics. Larger comparative studies are also needed to more accurately estimate the actual distribution of mutational events in other tissues, particularly those with high cancer rates, such as the lung, and tissues with very low cancer rates, such as the heart. A long-term goal should be to generate a genetic atlas of mutations in tissues of healthy individuals, sampled across the age spectrum and from distinct ancestries. Armed with a more comprehensive catalog of aging-associated mutations, we should be in a better position to understand the trigger events for cancer.

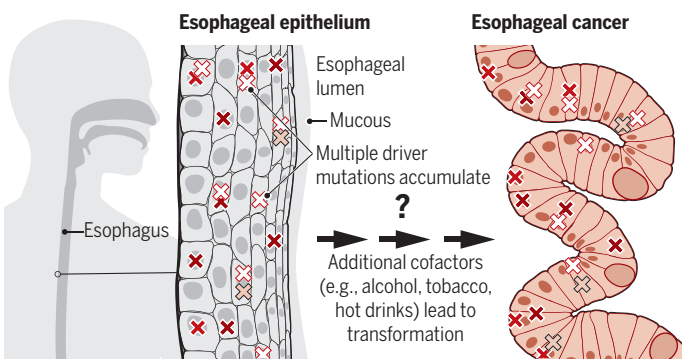
Although more comprehensive cataloging of mutations will inform our understanding of the genetic events that underlie cancer development, it is also necessary to investigate the suppression of tumor development through epigenetic changes as well as the influences of the host immune system and the local tissue microenvironment. Investment in prospective studies, which follow individuals sampled serially over time, will enable researchers to understand the timing and dynamics of transformation to cancer. Additionally, it could be interesting to prospectively study individuals in various locations because there are substantial differences in the rates and regional risk factors for esophageal squamous carcinoma and other cancer types. The potential implications of such future work are enormous because this study suggests that it might be more daunting than anticipated to develop genetic tests as a diagnostic tool for cancer.

For some time, there have been hints that the genome deteriorates with age, including telomere shortening at the ends of chromosomes (*11*), and the detection of subpopulations of blood cells that harbor cancer-associated mutations, known as clonal mosaicism. These observations have arisen from the analyses of circulating blood cells drawn from tens of thousands of apparently healthy individuals (*12*). The most common large event involving an entire chromosome, mosaic loss of the Y chromosome in blood cells, has been shown to increase with age and tobacco use, prompting some to speculate its use as a biomarker for chronic diseases associated with age (*13*). Similarly, point mutations in driver genes that are particularly important in hematologic cancers occur more frequently with age (*14*). This is known as clonal hematopoiesis and can be associated with increased risk for cardiovascular disease or a hematological cancer (*14*, *15*).

We should not restrict the application of the study of Martincorena *et al.* to cancer. Imagine that the generation of mutations in ostensibly healthy cells could be important in the pathogenesis of other chronic diseases of aging, such as diabetes, heart disease, and neurodegenerative disorders. If we can comprehend the basic mechanisms that underlie differences in mutational rates by tissue sites, together with how mutated cells are held in check, we could be a step closer to slowing down the destabilization of the genome with age. ■

Mutational events and aging in the esophagus

Genetic mutations accumulate in normal esophageal epithelium over time—often this does not initiate cancer. The initial mutations can be generated by intrinsic processes as well as induction by environmental factors (e.g., alcohol, tobacco). The progression to esophageal squamous carcinoma requires additional factors, which could include environmental triggers, additional mutations, ineffective immune surveillance and local changes to the microenvironment.



GRAPHIC: VERONICA FALCONIERI/*SCIENCE*

REFERENCES

1. C. Hutter, J. C. Zenklusen, *Cell* **173**, 283 (2018).
2. B. Vogelstein *et al.*, *Science* **339**, 1546 (2013).
3. F. S. Collins, H. Varmus. *N. Engl. J. Med.* **372**, 793 (2015).
4. P. A. Futreal *et al.*, *Nat. Rev. Cancer* **4**, 177 (2004).
5. I. Martincorena *et al.*, *Cell* **171**, 1029 (2017).
6. L. B. Alexandrov *et al.*, *Nature* **500**, 415 (2013).
7. L. B. Alexandrov *et al.*, *Science* **354**, 618 (2016).
8. I. Martincorena *et al.*, *Science* **362**, 911 (2018).
9. I. Martincorena *et al.*, *Science* **348**, 880 (2015).
10. Cancer Genome Atlas Research Network *et al.*, *Nature* **541**, 169 (2017).
11. S. A. Stewart, R. A. Weinberg, *Annu. Rev. Cell Dev. Biol.* **22**, 531 (2006).
12. M. Machiela *et al.*, *Am. J. Hum. Genet.* **96**, 487 (2015).
13. E. Loftfield *et al.*, *Sci. Rep.* **8**, 12316 (2018).
14. P. Jaiswal *et al.*, *N. Engl. J. Med.* **377**, 111 (2017).
15. S. Abelson *et al.*, *Nature* **559**, 400 (2018).

10.1126/science.aav5697

STEM CELLS

Changing concepts in hematopoietic stem cells

Analyses of platelet and red blood cell lineages are redefining how we consider stemness

By **Ryo Yamamoto[1,2], Adam C. Wilkinson[1,2], Hiromitsu Nakauchi[1,2,3]**

Hematopoietic stem cells (HSCs) can produce all cell lineages within the adult blood system, and they have provided a flagship model in which to study stem cell biology. Concepts developed from studying HSCs have influenced how we consider other stem cell systems. HSCs are also one of the few stem cell types with a long history of clinical application, in the form of bone marrow transplantation. Recent technical advances have brought about a major revision in our understanding of the HSC compartment. These necessitate new models and nomenclature to describe HSC heterogeneity, self-renewal, and differentiation potential, as well as having broader implications for how we consider stemness.

HSC transplantation assays determine cellular potential by transferring a donor cell (or cells) into a host (recipient) that lacks a functional hematopoietic system. Functional HSCs transplanted into a lethally irradiated mouse engraft and display multipotent differentiation into all mature blood cell types to maintain long-term hematopoiesis. This assay has long been the gold-standard for functionally defining HSCs. Self-renewal capacity can be further investigated by serial transplantation, which resolves long-term (LT) and short-term (ST) reconstituting HSCs. By contrast, hematopoietic progenitor cells (HPCs) only transiently reconstitute the hematopoietic system (*1*). On the basis of in vivo and in vitro reconstitution capacity, various types of HPCs have been phenotypically and functionally defined. These efforts propagated the idea that HSCs initially differentiate in a linear hierarchy [LT-HSCs initially differentiate into ST-HSCs, which then progress into multipotent progenitors (MPPs)], followed by lineage branching at the MPP stage into oligo-potent and unilineage progenitors (see the figure).

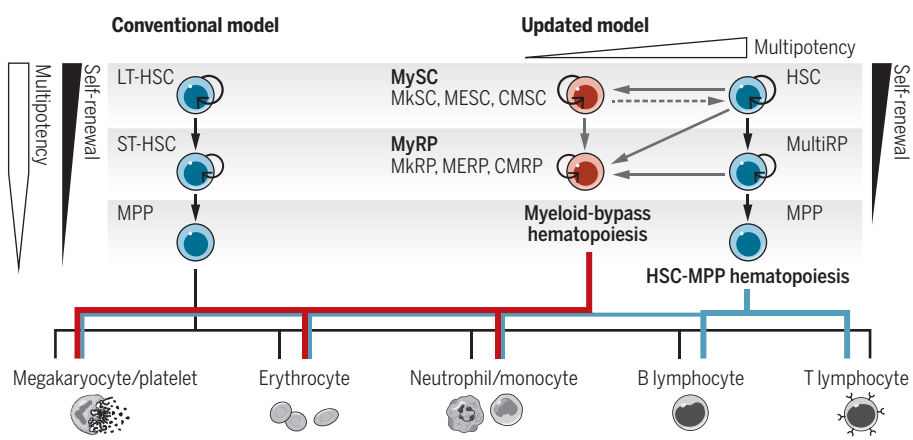
To track HSC lineage commitment, the transplantation assay has been combined with the CD45-congenic system, which provided an elegant approach to discriminate donor and recipient blood cells. This system is based on two different alleles of *CD45* (which encodes a surface protein), *CD45.1* and *CD45.2*, the expression of which can be distinguished by monoclonal antibodies. CD45 is expressed in neutrophil-monocyte, T cell, and B cell lineages, affording three-lineage chimerism analysis. Sensitive flow cytometric analysis of CD45 antibody staining even affords tracking of the progeny of single HSCs and led to the demonstration that single mouse bone marrow HSCs (immunophenotypically defined as $CD34^{-/lo}KIT^{+}SCA1^{+}Lineage^{-}$) display both multipotency and self-renewal (*2*). Functional HSCs have since been further resolved with additional cell surface markers, allowing close to 100% phenotypic purity (*1*). Using these systems along with cellular barcoding, researchers also identified skewing in HSC differentiation toward myeloid (neutrophil-monocyte) or lymphoid (T and B cell) production, leading to the concept of lineage-biased HSCs (*3, 4*).

Although the CD45-congenic system has provided a simple way to track HSC progeny, it is not the perfect stem cell assay. One limitation is that CD45 is not expressed on erythrocytes (red blood cells) or platelets (derived from megakaryocytes). This precludes quantification of HSC differentiation into these two abundant and essential blood components; erythrocytes and platelets together constitute >99% of peripheral blood. Only with the development of fluorescently labeled transgenic donor mice has it been possible to simultaneously analyze five blood lineages (*1, 5, 6*). Importantly, these new approaches have changed concepts in HSC biology. For example, a class of myeloid-restricted repopulating progenitor (MyRP) cells were identified, which only generated cells within the platelet, platelet-erythrocyte, or platelet-erythrocyte-neutrophil-monocyte lineages [defined as megakaryocyte repopulating progenitors (MkRPs), megakaryocyte-erythrocyte repopulating progenitors (MERPs), and common myeloid repopulating progenitors (CMRPs), respectively] (*1*). These repopulating progenitors (RPs) have the capacity to reconstitute parts of the hematopoietic system and can still be observed in recipients after 6 months, but they cannot be serially transplanted (*5, 7*). More recently, it has been demonstrated that myeloid lineage restriction can also occur within self-renewing stem cells (*5*), thereby functionally identifying a myeloid-restricted stem cell (MySC) population.

Five blood-lineage analysis has also identified a new HSC type in aged mouse bone marrow. In a clonal analysis of HSC aging, cells with myeloid-restricted differentiation output in primary recipients but multipotent differentiation output in secondary recipients were identified (*7*). Notably, this new stem

A new model of self-renewal and lineage potential in hematopoiesis

In the conventional model, LT-HSCs form ST-HSCs, which form MPPs prior to lineage commitment. In the updated model, stem cells can be multipotent (HSCs) or myeloid-restricted (MySCs). HSCs can generate MySCs via the myeloid-bypass pathway, but it is unclear if MySCs can generate HSCs. Loss of self-renewal generates repopulating progenitors (RPs), which can be multipotent (multiRPs) or myeloid-restricted (MyRPs).



GRAPHIC: A. KITTERMAN/*SCIENCE*

[1]*Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Lorry I. Lokey Stem Cell Research Building, 265 Campus Drive, Stanford, CA 94305, USA.* [2]*Department of Genetics, Stanford University, Stanford, CA, USA.* [3]*Division of Stem Cell Therapy, Center for Stem Cell Biology and Regenerative Medicine, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan. Email: nakauchi@stanford.edu*

cell type, called latent-HSCs, was only identified with the aid of five blood-lineage analysis; latent-HSCs often only produce platelets in primary recipients (*7*). Latent-HSCs were frequently found when aged mouse bone marrow cells were transplanted into young recipient mice, but this cell type has not been observed when young bone marrow cells were transplanted. These results suggest that there may be a certain level of plasticity (ability to change cellular potential) in the aged stem cell compartment.

The study of HSC activity in native (nontransplantation) hematopoiesis has also benefited from improved technology, including the development of cellular barcoding and lineage-tracing tools (*8*–*11*). These systems have allowed lineage output to be observed in unperturbed hematopoiesis. Consistent with evidence that transplantable HSCs are largely quiescent (reversibly nondividing) in the bone marrow, HSCs were found to minimally contribute to steady-state native hematopoiesis. Instead, phenotypic progenitor cell populations (particularly MPPs) appear to be the major population responsible for functionally maintaining blood system homeostasis (*8, 9, 11*). However, application of five blood-lineage analysis to molecular barcoding systems has identified the HSC compartment as an important source of megakaryocytes in native hematopoiesis (*11*). It is estimated that half of these platelet precursors are generated via this myeloid-bypass pathway during steady-state hematopoiesis (*11*), although this may be higher in stress conditions (*6*).

These five blood-lineage studies have revised our understanding of hematopoiesis. It is now clear that HSC-MPP differentiation is not the only route to produce lineage-restricted progenitors; HSCs may also differentiate directly into MyRPs, in a route called the myeloid-bypass pathway, which does not proceed through the MPP stage (see the figure). There is evidence that this pathway is active in both posttransplantation and native hematopoiesis in mice (*1, 6, 11*), as well as humans (*12*). Given the reproducible output in transplantation assays, it is assumed that these differentiation pathways are at least partially regulated by intrinsic mechanisms. However, recent evidence indicates that these different stem cell types may be spatially separated within the bone marrow (*13*), suggesting a potential role for particular bone marrow microenvironments (niches) in regulating cell potential. The apparent plasticity of latent-HSCs in young recipients (*7*) also suggests a role for young bone marrow niches in regulating stem cell potential.

"...five blood-lineage studies have revised our understanding of hematopoiesis."

These new insights necessitate an update in our concepts of stem cell self-renewal and multipotency. It is now clear that functional HSCs can lose self-renewal capacity via at least two differentiation routes, through differentiation into an MPP or differentiation into a MyRP (via the myeloid-bypass pathway) (*1*). These findings shift our understanding of HSC biology: self-renewal and multipotency do not have to be shared cellular properties. To emphasize this, we propose limiting the use of "stem cell" to those HSCs and MySCs with bona fide self-renewal activity (based on capacity to serially transplant into secondary recipients), but not ST-HSCs or MyRPs that can only reconstitute primary recipients. By contrast, "repopulating progenitor" describes these short-term reconstituting cell types. Because of their failure to provide long-term chimerism following transplantation, we believe that ST-HSCs are more accurately considered as multipotent-repopulating progenitors (multiRPs). In this definition, multipotency is not necessary for stem cell activity; unipotent self-renewing primitive (immature) cell types are considered stem cells.

Another important revision is that there are at least two functional self-renewing stem cell populations within the hematopoietic system. Both are found within the immunophenotypically defined bone marrow HSC compartment, although MySCs are reportedly enriched within the VWF^+ subfraction of this compartment in mice (*5*). To make clear the functional distinction between these multipotent and myeloid-restricted stem cell types, we term them HSCs and MySCs. It is apparent that MySCs come in several "flavors," and these oligo- and unilineage stem cell populations can be further functionally defined as megakaryocyte stem cells (MkSCs), megakaryocyte-erythrocyte stem cells (MESCs), and common myeloid stem cells (CMSCs). Notably, the megakaryocyte (platelet) lineage is the only one to always be associated with self-renewal activity (assayed by secondary transplantation) (*5, 7*). This terminology also standardizes stem cell nomenclature with current naming of downstream progenitor populations (for example, CMPs, MEPs, and MkPs) and helps to distinguish lineage bias from lineage restriction; myeloid-HSCs denotes myeloid-biased (but still multipotent) HSCs, whereas MySC denotes myeloid-restricted stem cells. Of course, the strong myeloid output of myeloid-biased HSCs could be due to the preferential production of myeloid-restricted stem or progenitor cells by these myeloid-HSCs.

These changes to the model of hematopoiesis have implications for how we approach HSC biology as well as clinical use of HSCs and self-renewing lineage-restricted cells. There is increasing interest in describing the molecular and biochemical aspects of self-renewal and differentiation within the hematopoietic system. However, understanding the cellular properties of stem cells and progenitors is necessary to focus such research. Similarly, efforts to genetically correct HSCs in gene therapies for various hematological diseases need to ensure that the appropriate stem cell population is targeted to achieve long-term and high-level reconstitution in the affected blood lineage.

These new insights may also have relevance to disease, particularly the origin of malignancies such as leukemia. Self-renewing HSCs have long been proposed as the origin of leukemias as well as the more recently described clonal hematopoiesis, a preleukemic state in which a somatic mutation (or mutations) clonally spreads through the human hematopoietic system (*14*). The potent self-renewal of MySCs suggests potential alternative cellular origins of these somatic mutations.

We believe that this new model and nomenclature for describing the activity of functional hematopoietic stem and progenitor cells will help to unify these recent findings. Additionally, these concepts may have application in other adult stem cell systems, for example, in the neural system where the stem cell hierarchy has recently been questioned (*15*). Moreover, it is important to determine the relevance of these HSC and MySC populations in humans. We await new technological advances to drive our biological understanding of these rare but important cell populations. ■

REFERENCES AND NOTES

1. R. Yamamoto *et al.*, *Cell* **154**, 1112 (2013).
2. M. Osawa, K. Hanada, H. Hamada, H. Nakauchi, *Science* **273**, 242 (1996).
3. C. J. Eaves, *Blood* **125**, 2605 (2015).
4. S. H. Naik *et al.*, *Nature* **496**, 229 (2013).
5. J. Carrelha *et al.*, *Nature* **554**, 106 (2018).
6. S. Haas *et al.*, *Cell Stem Cell* **17**, 422 (2015).
7. R. Yamamoto *et al.*, *Cell Stem Cell* **22**, 600 (2018).
8. J. Sun *et al.*, *Nature* **514**, 322 (2014).
9. K. Busch *et al.*, *Nature* **518**, 542 (2015).
10. W. Pei *et al.*, *Nature* **548**, 456 (2017).
11. A. E. Rodriguez-Fraticelli *et al.*, *Nature* **553**, 212 (2018).
12. F. Notta *et al.*, *Science* **351**, aab2116 (2016).
13. S. Pinho *et al.*, *Dev. Cell* **44**, 634 (2018).
14. M. Jan, B. L. Ebert, S. Jaiswal, *Semin. Hematol.* **54**, 43 (2017).
15. G.-A. Pilz *et al.*, *Science* **359**, 658 (2018).

ACKNOWLEDGMENTS

The authors receive funding from the California Institute of Regenerative Medicine (LA1_C12-06917, DISC1-10555), the National Institutes of Health (R01DK116944, R01HL147124), Japan Agency for Medical Research and Development, the Ludwig Foundation, the Siebel Foundation, and the Leukemia and Lymphoma Society (3385-19). R.Y. and A.C.W. contributed equally to this work.

10.1126/science.aat7873

RETROSPECTIVE

Thomas A. Steitz (1940–2018)

Biochemistry giant who illuminated ribosome structure

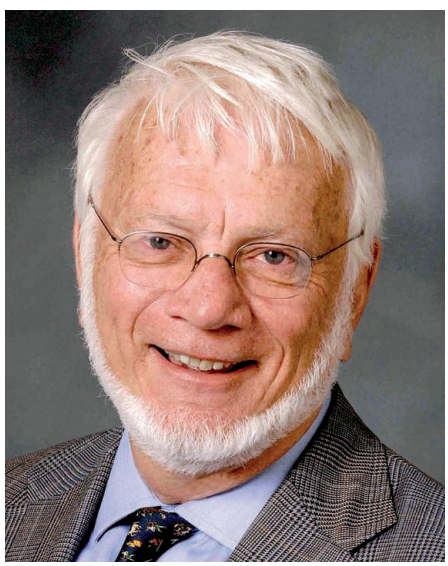
By **Venki Ramakrishnan** *and* **Richard Henderson**

Thomas A. Steitz, distinguished molecular and structural biologist, died on 9 October at the age of 78. Tom was known for his unerring judgment in picking fundamentally important problems, and persisting, sometimes for over a decade, until he solved them. His work on the information flow from DNA to RNA to protein culminated in work on the structure of the 50*S* subunit of the ribosome, for which he shared the 2009 Nobel Prize in Chemistry.

Born on 23 August 1940 in Milwaukee, Wisconsin, Tom attended Lawrence College in Wisconsin on a scholarship and earned his degree in chemistry in 1962. He then began his graduate work at Harvard University. Although Tom initially planned to work on biophysical studies of nucleic acids, he was thrilled by lectures given by Max Perutz, the pioneer of protein crystallography, in which Max presented the first atomic-resolution structure of the protein myoglobin. Inspired to become a protein crystallographer, Tom joined the group of chemist and crystallographer W. N. (Bill) Lipscomb to work on the structure and mechanism of carboxypeptidase A. During this time, he met and married Joan Argetsinger Steitz, who became an equally distinguished molecular biologist.

In 1966, Tom earned his Ph.D. in biochemistry and molecular biology at Harvard. The next year, he and Joan moved to the Medical Research Council (MRC) Laboratory of Molecular Biology (LMB) in Cambridge, United Kingdom, as postdocs, where Tom worked with crystallographer and biophysicist David Blow on the mechanism of protein hydrolysis by the serine protease chymotrypsin. The work led to an understanding of how all serine proteases cleave peptide bonds, as well as how the homologous trypsin recognizes positively charged side chains using a negatively charged aspartic acid at the bottom of its specificity pocket. Tom enjoyed discussions about research strategies with Brian Hartley and other pioneers of molecular and structural biology at LMB. He described his time at LMB as "the most important time in the development of my future research direction and my perspective on how creative science should best be done."

Next, Tom and Joan moved briefly to the University of California, Berkeley, where Tom had been offered a faculty job. As Tom described it, "I asked if there was any possibility of a job for Joan. The department chairman… said, 'She's a woman. Women do not run their own lab; they work in the lab of their husband.'" So they departed for Yale University, where Fred Richards offered both Tom and Joan faculty positions as part of his visionary



transformation of the Molecular Biophysics and Biochemistry Department into a powerhouse of structural and molecular biology. Joan said recently, "I was very, very lucky to have married Tom. He really believed that I should have as equal an opportunity to succeed as he." They were a star couple at Yale, both becoming hugely influential leaders in their respective and sometimes overlapping fields. They even published an important paper together on the role of metal ions in catalysis by DNA and RNA polymerases.

At Yale, Tom began by determining the structure of hexokinase, an enzyme that phosphorylates glucose. The enzyme was one of the early examples of induced fit, whereby the substrate induces a change in enzyme conformation to bring catalytic residues into their active positions. His lab then focused on the structural mechanisms underlying the central dogma of molecular biology: how information flows from DNA to RNA to protein. His group worked out the structures of a series of important enzymes and cofactors that form the central core of structural molecular biology. This work culminated in his collaboration with Peter Moore, whose lifetime of work on the ribosome complemented Tom's crystallographic expertise. Together, they and their colleagues worked out the structure of the 50*S* ribosomal subunit, which definitively established that the ribosome is a ribozyme, and obtained the structures of complexes of the ribosome with many different antibiotics.

In his steadfast determination to get to the heart of a problem, Tom was not afraid to make bold guesses about how a molecule works. This occasionally led to dramatically incorrect proposals. When he solved one of the first structures of a DNA-binding protein, the CAP repressor, he suggested that it bound to left-handed DNA. He joked that this was such an egregious mistake that he would have to solve five more structures before he would be invited to speak at a conference again. Later, he suggested a mechanism of peptidyl transfer in the 50*S* ribosomal subunit. Both of these ideas proved incorrect. In each case, Tom exemplified how science should progress from mistakes by admitting the error and correcting it. His lab went on to show that CAP bound to a sharply bent form of DNA, a tour de force at the time. His lab also obtained a series of structures of substrate and transition-state analogs bound to the 50*S* subunit that helped to establish the correct mechanism and was described by the Nobel Committee as the "jewel in the crown" of studies on the mechanism of peptidyl transfer. Rather than be overly wedded to his own ideas, Tom received corrections from others with interest as well. His ultimate goal was always to help establish the truth.

Tom was known for his direct manner, which, along with his enviable track record, could intimidate anyone who was insecure. However, this was simply another aspect of his eagerness to get to the crux of the matter without wasting time. It also meant that he could be bluntly critical, often with an amusing pun, when encountering sloppy science, but generous with praise when he came across beautiful work. His students and colleagues never doubted where they stood with him. Even so, he was a supportive and egalitarian mentor who believed that talent and drive should determine success. These traits, along with his commitment to working only on problems he thought were truly important, helped him to train many generations of outstanding protégés, who along with achievements from his own lab form part of the huge legacy he has left behind. ■

MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. Email: ramak@mrc-lmb.cam.ac.uk; rh15@mrc-lmb.cam.ac.uk

10.1126/science.aav8253

PHOTO: YALE UNIVERSITY/UPI/NEWSCOM

POLICY FORUM

The alleged Golden State Killer, Joseph DeAngelo, appears at his arraignment in Sacramento, California, in late April.

SCIENCE AND LAW

Is it time for a universal genetic forensic database?

Bias and privacy concerns cloud police use of genetics

By **J. W. Hazel[1,2], E. W. Clayton[1,2,3], B. A. Malin[2,4,5,6], C. Slobogin[2,3]**

DNA is an increasingly useful crime-solving tool. But still quite unclear is the extent to which law enforcement should be able to obtain genetic data housed in public and private databases. How one answers that question might vary substantially, depending on the source of the data. Several countries—the United Kingdom, Kuwait, and Saudi Arabia among them—have even toyed with creating a "universal" DNA database, populated with data from every individual in society, obviating the need for any other DNA source (*1*). Although this move would be controversial, it may not be as dramatic as one might think. In the United States, for example, the combination of state and federal databases (containing genetic profiles of more than 16.5 million arrestees and convicts) and public and private databases (containing genetic data of tens of millions of patients, consumers, and research participants) already provides the government with potential access to genetic information that can be linked to a large segment of the country, either directly or through a relative (*2*, *3*). We discuss here how, if correctly implemented, a universal database would likely be more productive and less discriminatory than our current system, without compromising as much privacy.

Current law enforcement methods of genetic investigation are both haphazard and underregulated. In early 2018, U.S. law enforcement officers investigating the Golden State Killer case were able to home in on a suspect after querying GEDmatch, a publicly accessible database that encourages consumers to upload genetic data coupled with personal identifiers in order to gain insights into their genealogy. Without authorization from a court, law enforcement simply pretended to be the donor of what was, in fact, crime scene DNA. Through that ruse, officers found a match to a person in the database who was distantly related to Joseph DeAngelo, the man ultimately arrested for the crimes. Since these revelations came to light last spring, multiple law enforcement agencies have used similar long-range familial searches of publicly accessible databases to close 13 cold cases, including several murders (*2*, *4*).

In the Golden State case, the government could justify its action by pointing to the fact that GEDmatch is advertised as a publicly accessible database—one that does not specifically ban the type of deception police used in that case. But publicly accessible databases are not the only source of genetic information that law enforcement might query. For instance, if accessing such a database fails to yield a useful result, which will often be the case, law enforcement could resort to private databases, such as those maintained by direct-to-consumer (DTC) companies, e.g., 23andMe and Ancestry.com. Although these databases are not as easily exploited as databases meant to be accessed by the public, in most jurisdictions in the United States and throughout the world a subpoena is all that law enforcement needs to force those companies to determine whether they have a match with crime scene data. A subpoena only

[1]*Center for Biomedical Ethics and Society, Vanderbilt University, Nashville, TN 37203, USA.* [2]*Center for Genetic Privacy and Identity in Community Settings, Vanderbilt University Medical Center, Nashville, TN 37203, USA.* [3]*Vanderbilt University Law School, Nashville, TN 37203, USA.* [4]*Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, USA.* [5]*Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37203, USA.* [6]*Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN 37203, USA. Email: james.w.hazel.1@vumc.org*

requires showing that the data sought are relevant to an investigation and is therefore much simpler to get than a warrant based on probable cause.

Until now law enforcement has largely focused its efforts on targeting publicly accessible resources such as GEDmatch. But requests for privately maintained data are likely to become much more frequent in the future, given the increasing value of genetic data to law enforcement, the low level of justification required for a subpoena, and the tremendous amount of effort that can be associated with long-range familial searching by using a resource such as GEDmatch, which might generate dozens or hundreds of possible leads in a given case (*2*).

If publicly accessible databases and DTC companies are of no help, law enforcement might try to access genetic data in the possession of healthcare providers and researchers. Again, only a subpoena is needed to obtain genetic information contained in patients' electronic medical records under the U.S. Health Insurance Portability and Accountability Act of 1996 (*5*). And although biomedical research efforts are often protected by government-issued Certificates of Confidentiality (*6*), which purport to assure participants that research data are immune from court orders, the enhanced protections recently conveyed by the U.S. 21st Century Cures Act of 2016 remain largely untested in the courts. Further, because Certificates of Confidentiality typically apply only to research funded by the National Institutes of Health and other agencies within the U.S. Department of Health and Human Services, genetic research funded by other sources remains largely unprotected unless a request for a Certificate of Confidentiality is made and granted.

Last, in addition to these public and private resources, a government interested in using DNA to help solve crimes can maintain its own database. In the United States, many states and the federal government maintain DNA profiles not only of convicted felons but also of those simply arrested for a felony or, in some cases, even a misdemeanor (*1*). The U.S. Supreme Court has given its imprimatur to such databases (*7*). As we explain below, this development is one of the most potent reasons for considering establishment of a more comprehensive genetic database.

UNIVERSAL DATABASE

The first obvious benefit of a universal database is its potential for solving or deterring serious crimes such as murder, rape, robbery, and burglary. As both research and anecdotal reports indicate, DNA matches have often been crucial in catching the perpetrators of such crimes and useful in identifying bodies and remains as well (*8*, *9*). Unfortunately, from law enforcement's perspective, forensic databases that contain only genetic data of arrestees and those convicted of crimes have serious limitations, a fact demonstrated by law enforcement's increasing reliance on publicly accessible and private databases, composed almost entirely of "innocent" individuals. And when law enforcement chooses the latter route, a match is by no means guaranteed; additionally, considerable inefficiency is likely if the effort to find a match requires consulting numerous companies, all of which may need to re-analyze their sample to generate the relevant profile.

Just as important, a universal database would eliminate or reduce problems associated with the current haphazard genetic investigative regime. First, such a database would virtually erase the government's incentive to conduct long-range familial DNA searches of the type used in the Golden State Killer case. It would thus markedly alleviate the impact on innocent people who happen to be related to criminals and whom police are likely to treat as suspects unless and until countervailing evidence surfaces.

Second, a universal database would eliminate the temptation on the part of law enforcement to use public, DTC, or research databases for investigative purposes. Indeed, for reasons we give below, universal database legislation should prohibit law enforcement officials from trawling nongovernmental DNA sources such as GEDMatch, Ancestry.com and research-oriented databases. That in turn might enhance research into diseases, treatments, and other socially beneficial avenues because studies indicate that many people, especially those of color, are reluctant to provide genetic information to researchers out of fear it will be misused by the government (*10*, *11*).

Last, a universal database would be less discriminatory than the government's current method of compelling genetic samples. If the government collects DNA only from convicted individuals or only from individuals arrested for serious crimes—as is true as a matter of law in the United Kingdom and as a practical matter with the U.S. Combined DNA Index System (CODIS)—there is real concern that the resulting databases will be skewed against the disadvantaged because they are the ones most likely to be the focus of such convictions and arrests.

The situation in the United States has been exacerbated by federal, state, and local governments now creating "shadow" databases (*9*)—not only of people arrested for any crime but also of people who are merely stopped on suspicion of having committed a crime without being arrested (the so-called "stop-and-spit" and "swab-and-go" practices). As a result, arrest-based DNA databases contain a huge proportion of the young nonwhite male population and a much smaller representation of other groups (*9*, *12*). Indeed, that is why police had to rely on a publicly accessible database to catch the Golden State Killer, a white former police officer; in sharp contrast to government DNA caches such as CODIS, public and DTC databases tend to contain the genetic data of predominately white individuals, generally from higher income brackets.

Despite these advantages of a universal database, many concerns have been raised about its privacy implications and the associated potential for misuse of genetic information. As a result, in some countries a universal database is clearly prohibited. In *S. and Marper* v. *United Kingdom* (*13*), the European Court of Human Rights concluded that the indefinite retention of biological samples and profiles (including not only genetic data but also fingerprints and other biological information) is a violation of the right to privacy protected under the European Convention of Human Rights. That not only spells doom for universal genetic databases, it also prohibits long-term databases composed of profiles of people who are arrested but not convicted. In response to Marper, the United Kingdom, which had been retaining the DNA samples of virtually all arrestees, now destroys such samples immediately if collected from individuals charged with minor crimes and after 3 years for those arrested for serious crimes. Although Marper applies only in the Council of Europe's 47 member countries, many other countries follow its dictates (*1*).

ALLAYING CONCERNS

To some extent, the decision in Marper is based on fear that those in the database will be associated with criminality. But that drawback is specific to databases that focus on arrestees; the criminal stigma of being in a database is eliminated if everyone's DNA is acquired. More relevant is Marper's objection that broad collection of genetic material might increase "the risk of abuse and arbitrariness" (*13*). These concerns would clearly be raised by the establishment of a universal database, but they can be allayed in a number of ways.

Most important to recognize is that a forensic database would only require a subset of genetic markers with little to no medical relevance. Profiles would consist of a few dozen short-tandem repeats, with perhaps a modest expansion of the 20 CODIS loci currently used to improve the identification of degraded samples or the addition of a limited subset of "forensic" single-nucleotide polymorphisms to enhance the identification of more distant relatives in the rare instances in which familial searches were still needed (*3*). As a result, these law enforce-

ment profiles would reveal substantially less sensitive information than the thousands (or hundreds of thousands) of genetic variants, often coupled with individual and family medical histories, that are found in the healthcare, research, or DTC ecosystems that law enforcement might otherwise be tempted to commandeer.

Many other protections against misuse of DNA databases can and should be created by the relevant legislative body (which, in the United States, would be Congress, given the nationwide impact of the law). For instance, legislation could require that genetic data not only be uncoupled from any personal identifiers within the system, as it is in CODIS, but also establish a more robust "unmasking" process that limits law enforcement access to any personal information until an association has been made and confirmed (a procedure better monitored through one central system than state-by-state or company-by-company). In further contrast to the current system, legislation might limit access to the database to specific circumstances, such as investigations into felonies and identification of missing persons' remains.

> ***"...legislators, their kin, and their key constituents will be affected just like everyone else."***

Universal database legislation should also require that the DNA database be housed in an independent agency and that access to it be authorized by a warrant (not just a subpoena) based on probable cause to believe a match will produce a perpetrator (a showing that is usually impossible with a database that is not universal). Most important, the law should require that the physical samples analyzed to create the database be destroyed after obtaining the relevant genetic information, to mitigate the risk that the sample will be subjected to further analysis or used for purposes other than populating the database.

Additional privacy protection could be realized through emerging cryptographic protocols that control access to genomic data through multiple keys. Where more than one organization is required to "turn the key" to decrypt any record, the risk of a rogue individual or agency misusing the resource is substantially mitigated (*14*). Simultaneously, because law enforcement needs would be fully met, Congress should (and probably would) severely restrict the ability of law enforcement to search other health-care, research, or DTC databases, increasing trust in these activities and avoiding government access to the more complete genetic information housed there.

Whatever its precise structure, the most important point is that the population-wide nature of the database would all but guarantee the adoption of strong security measures such as those just described, as well as the enactment of harsh penalties for abuse of the type currently associated with misuse of data in CODIS (a fine of up to $250,000 or imprisonment for up to 1 year). That is because members of Congress would know that government DNA harvesting would no longer focus solely on out-groups but would also sweep in their own DNA. As the ubiquity of federal and state legislation strictly regulating the privacy of communications records and tax information suggests, slippery-slope concerns about government collection and exploitation of every citizen's full genetic makeup dissipate in a regime in which legislators, their kin, and their key constituents will be affected just like everyone else.

These concerns are further minimized in jurisdictions such as the United States and Europe that, unlike many countries that have considered a universal database, have codified basic privacy protections that would mitigate the potential for abuse or misuse of the data (for example, the Privacy Act of 1974 and the Genetic Information Nondiscrimination Act in the U.S., and the General Data Protection Regulation in Europe).

IMPLEMENTATION ISSUES

There remain implementation issues that would need to be debated by the public and ultimately resolved by Congress, including whether a universal database should be populated by obtaining samples from all newborns or instead through a census-style effort (or a combination of both), how to collect the DNA of visitors from other countries, and the appropriate incentives to promote compliance with the program.

The ethical objections to mandating forensic profiling of newborns and/or compelling every citizen or visitor to submit to a buccal swab or to spit in a cup when they have done nothing wrong are not trivial. But newborns are already subject to compulsory medical screening, and people coming from foreign countries to the United States already submit to fingerprinting. It is also worth noting that concerns about coercion or invasions of privacy did not give pause to legislatures (or, for that matter, even the European Court) when authorizing compelled DNA sampling from arrestees, who should not forfeit genetic privacy interests simply by virtue of being arrested.

A universal database would not be cheap; extrapolating from a $20- to $40-per-profile estimate for the existing CODIS system calculated in 2010 (*15*), compiling a database of ~350 million people could cost between $7.5 billion and $15 billion dollars. Although this figure does not include implementation costs (which are difficult to estimate), the economies of scale associated with a universal system, coupled with the declining cost of forensic profiling, would likely drive this figure lower.

In addition, the societal and economic benefits that could be derived from the system could easily offset these costs. Criminal activity is extremely expensive for private citizens (both monetarily and in terms of intangible harms to victims), third parties (such as businesses and insurers), and the government (for police investigations and incarceration). There is evidence that existing forensic databases have more than made up for their initial costs by increasing the efficiency, accuracy, and success rate of ongoing criminal investigations and by deterring would-be criminals (*15*).

At the very least, putting the idea of a universal forensic database on the table would spur a long overdue debate about the deficiencies of the current system and, more broadly, our societal commitment to privacy, fairness, and equal protection under the law. ■

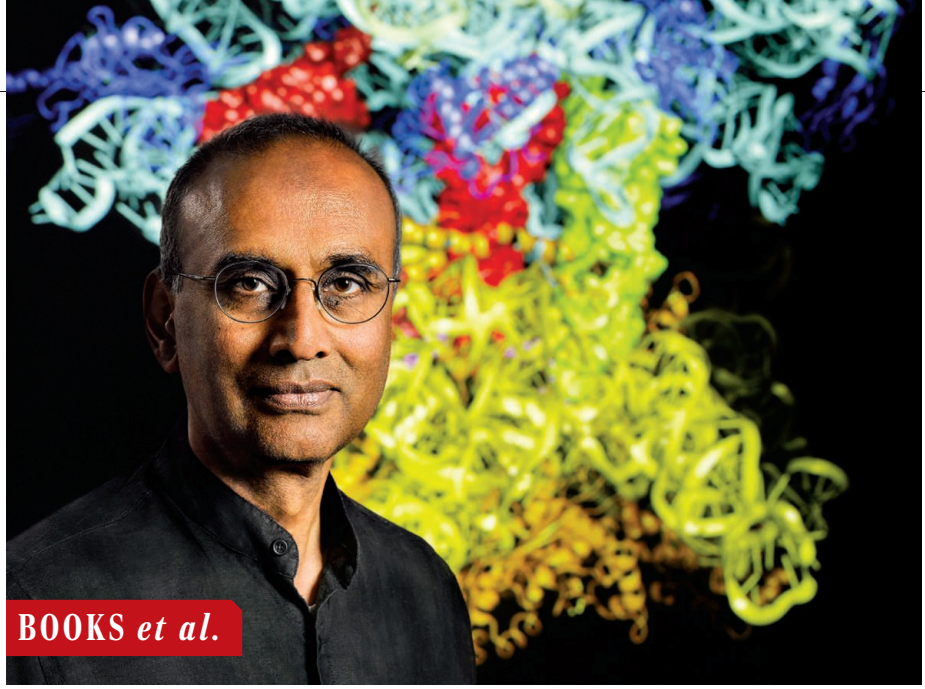
REFERENCES AND NOTES

1. H. M. Wallace, A. R. Jackson, J. Gruber, A. D. Thibedeau, *Egyptian J. Forens. Sci.* **4.3**, 57 (2014).
2. Y. Erlich, T. Shor, I. Pe'er, S. Carmi, *Science* **362**, 690 (2018).
3. J. Kim, M. D. Edge, B. F. B. Algee-Hewitt, J. Z. Li, N. A. Rosenberg, *Cell* **175**, 848 (2018).
4. A. Regalado, *MIT Technol. Rev.* (13 September 2018); www.technologyreview.com/s/612001/hundreds-of-crimes-will-soon-be-solved-using-dna-databases-genealogist-predicts.
5. 45 Code of Federal Regulations 164.512(f).
6. 42 U.S. Code 241(d)(1)(E).
7. *Maryland* v. *King*, 569 U.S. 435 (2013).
8. H. Safir, *Forens. Mag.* (1 September 2007); www.forensicmag.com/article/2007/10/dna-technology-effective-tool-reducing-crime..
9. T. Jones, *Guardian* (9 April 2018); www.theguardian.com/politics/2010/apr/10/dna-analysis-crime-solving.
10. E. W. Clayton, C. M. Halverson, N. A. Sathe, B. A. Malin, *PLOS ONE* 10.1371/journal.pone.0204417 (2018).
11. L. G. Landry, N. Ali, D. R. Williams, H. L. Rehm, V. L. Bonham, *Health Aff.* **37**, 780 (2018).
12. R. Brame, S. D. Bushway, R. Paternoster, M. G. Turner, *Crime Delinq.* **60**, 471 (2014).
13. *S. and Marper* v. *the United Kingdom*, Grand Chamber of the European Court of Human Rights, Strasbourg (4 December 2008); https://rm.coe.int/168067d216.
14. H. Cho, D. Wu, B. Berger, *Nat. Biotech.* **36.6**, 547 (2018).
15. J. L. Doleac, *Am. Econ. J. Appl. Econ.* **9.1**, 165 (2017).

ACKNOWLEDGMENTS

The authors thank members of the Center for Genetic Privacy and Identity in Community Settings (GetPreCiSe) for their thoughtful comments and feedback. **Funding:** This work was funded by the Center for Genetic Privacy and Identity in Community Settings (NIH/NHGRI RM1HG009034). **Author contributions:** All authors contributed equally to this work. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data is available in the main text.

10.1126/science.aav5475



Venki Ramakrishnan shared the 2009 Nobel Prize in chemistry for his work on the ribosome structure.

PHOTO: MAX ALEXANDER/SCIENCE SOURCE

BOOKS *et al.*

STRUCTURAL BIOLOGY

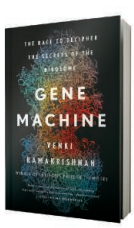
The ribosome, revealed

A Nobel laureate recounts the story of his quest to discover the structure of a key molecular machine

By **Daniel N. Wilson**

By the late 1970s, the structure of DNA had already been "solved," and the genetic code already "cracked," leading many to conclude that there was little left to learn in this arena. However, the quest to understand the process by which the ribosome converts genetic information into proteins was still in its infancy. In *Gene Machine*, Venki Ramakrishnan provides a frank behind-the-scenes account of the 30-year race to solve the structure of the ribosome that earned him a share of the Nobel Prize in chemistry in 2009.

The first chapters of *Gene Machine* cover Ramakrishnan's early career as a researcher, revealing how he serendipitously stumbled onto ribosomes. He had wanted to study membrane proteins in Don Engelman's laboratory at Yale, but there was no vacancy when he inquired. A position was available, however, in the neighboring laboratory of Peter Moore. Moore, incidentally, worked closely with the late Thomas Steitz, who would become Ramakrishnan's biggest competitor, illustrating the interconnectedness—or, as Ramakrishnan describes it, the "inbred" nature—of science.



Gene Machine
The Race to Decipher the Secrets of the Ribosome
Venki Ramakrishnan
Basic Books, 2018.
288 pp.

Ramakrishnan would go on to attain high-resolution structures of the small ribosomal subunit and subsequently the entire 70*S* ribosome, providing fundamental insights into the process of translation and shedding light on the mechanism of action exploited by antibiotics.

Despite the ribosome's complicated architecture, it was evident by the early 1980s that obtaining three-dimensional crystals of ribosomal particles was feasible. Here in his retelling, Ramakrishnan reveals a little-known story about Hasko Paradies, a German pediatrician who claimed to have crystallized ribosomes in 1974. Leading crystallographers questioned Paradies's findings, and he soon "disappeared from the world of biological structures." But, as Ramakrishnan notes, "Often, just being told something is possible … spurs people to try things." A number of crystallographers reported that Paradies's spurious findings encouraged them to continue their own research.

The colorful characters that were involved in the race to crystallize the ribosome provide a richness to the narrative. Ramakrishnan was the "dark horse"—someone who "came out of nowhere and surprised everyone," according to biophysicist Joachim Frank—and his postdoc, Brian Wimberly, was "Ferrari Boy" (a nickname bestowed on him after Ramakrishnan likened Wimberly's excitement over a set of electron density maps to "giving a teenager the keys to a Ferrari").

The text is also full of comical anecdotes. In telling the story of how Nigel Unwin and Carlo Taddei obtained two-dimensional protein crystals from lizard oocytes in 1977—the first demonstration that ribosomes were indeed crystallizable—Ramakrishnan reveals how Taddei set off the fire alarms with his cigars and how the pair's lizards escaped and were seen scampering around the grounds of the Medical Research Council Laboratory of Molecular Biology for years afterward.

As mentioned earlier, among Ramakrishnan's main competitors was Thomas Steitz, a talented Yale crystallographer, who was renowned for his directness and—as Ramakrishnan observes—a slightly Amish appearance due to his chin-strap beard (on page 897 of this issue, Ramakrishnan and Richard Henderson reflect on Steitz's legacy). Others included the Israeli crystallographer Ada Yonath, who led a large team split between the Weizmann Institute in Israel and the Max Planck Institutes in Berlin and Hamburg, and the "Sage of Santa Cruz," biochemist Harry Noller, an RNA expert with "the demeanor of a mellow, pot-smoking California hippie."

Ramakrishnan, Steitz, and Yonath would share the Nobel Prize in chemistry in 2009 for their work on the ribosome. Although Noller would go on to receive the $3 million Breakthrough Prize in 2016, the Nobel committee's decision was not without controversy. (Another presumed contender for the 2009 award, Joachim Frank, received the Nobel Prize in chemistry in 2017.)

Ramakrishnan includes some insightful thoughts on the topic of prizes and the politics of recognition. In addition to arguing that limiting the Nobel Prize to only three people is inappropriate, he also identifies some common ailments associated with it, including "pre-Nobelitis," a condition in which frustration sets in each year when one does not receive the Nobel Prize, and "post-Nobelitis," an affliction to which many Nobel laureates succumb that compels them to pontificate on topics in which they are not specialists.

Gene Machine is not an unbiased historical account but rather a memoir that documents a momentous endeavor in scientific history. In the end, however, Ramakrishnan's personal reflections add color and humanity to the story. ■

10.1126/science.aav4624

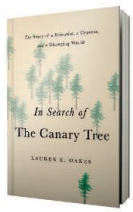
The reviewer is at the Institute for Biochemistry and Molecular Biology, University of Hamburg, 20146 Hamburg, Germany. Email: daniel.wilson@chemie.uni-hamburg.de

FOREST ECOLOGY

Eco-optimism in dark times

Even as extinction looms for her subjects, an ecologist maintains hope for a better future

In Search of the Canary Tree
The Story of a Scientist, a Cypress, and a Changing World
Lauren E. Oakes
Basic Books, 2018. 288 pp.

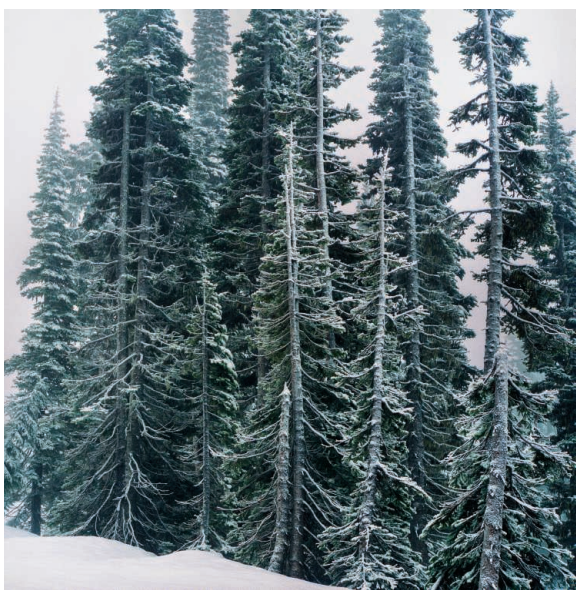


By **Sarah Boon**

In the introduction to *In Search of the Canary Tree*, ecologist Lauren Oakes establishes a hopeful tone for what is, for many who study environmental change, an emotionally taxing topic: "It is a book about finding faith…as a force that summons local solutions to a global problem, that helps me live joyfully and choose what matters most in seemingly dark times."

Oakes focuses on two key questions that environmental researchers are often asked: "How do you live with what you know?" and "Do you have hope about the future?" As she writes, "If you believe climate change is occurring… these questions are not alarmist; they are…the [logical] by-products of doing the research itself; the results of the fact that science is not removed from lived experience."

In Search of the Canary Tree begins in 2010, with Oakes searching for a topic for her Ph.D. research in Alaska. She interviews local scientists, forest service employees, and park rangers about their research and how she might be able to contribute. She finds herself fascinated by the yellow-cedar tree (*Cupressus nootkatensis*), which is dying across huge swaths of the northwest coast of North America, from British Columbia to Alaska. The cause: a lack of snow, which normally insulates and protects the tree's roots during the region's harsh winter.



In the absence of adequate snow, yellow-cedar trees are dying off.

For the field component of her research, Oakes set up a sampling strategy called a "chronosequence"—a collection of forested sites of different ages, from which a researcher can infer the time-dependent development of an ecosystem. This is one of many scientific concepts that Oakes describes well in plain language.

Oakes selected sites according to the time since yellow-cedar death and then measured a range of ecological variables to quantify what grew back. She found that stands with dead yellow-cedar trees were transitioning to forests dominated by western hemlock, with fewer mosses and ferns and more shrubs. Stands with living yellow-cedar, however, had the conditions necessary to nurture new yellow-cedar seedlings.

Oakes also interviewed 45 locals in Juneau and Sitka—members of the First Nations, loggers, scientists, and local business owners—asking them about the yellow-cedar decline. Did it affect them personally? Did they know what was causing it? Had it made them change anything in their everyday lives? Yellow-cedar, she found, is important to many Alaskans for a range of reasons: cultural heritage and art, logging, and even as symbols of larger moral issues related to climate change and human culpability.

Oakes is as meticulous with her writing as she is with her science. She draws on field notes, transcripts of interviews, research papers, emails, letters, and journals to ensure the book's accuracy. This attention to detail helps her craft a compelling book with a narrative driven by people.

We get to know her mentors—Paul Hennon, Ashley Steel, and Greg Streveler—and her field team: Odin Miller, a Juneau native, "giant with size 14 shoes," and student at the University of Alaska; Paul Fischer, an experienced ecologist and forestry student at the University of Washington whom Oakes affectionately nicknamed "P-Fisch;" and Kate Cahill, vice president of the University of California, Berkeley, forestry club who came with her own nickname ("Maddog"). A talented artist, Oakes enlisted Cahill to illustrate the various stages of yellow-cedar decline.

In Search of the Canary Tree is reminiscent of Hope Jahren's 2016 memoir, *Lab Girl*, but where that book was a wild, fairly disorganized romp through Jahren's career, Oakes's storytelling is more methodical. In addition to the benefits readers might derive from this story of a scientist's early career, the book is also a good read for new researchers considering how to set up their own ecological studies.

This is also one of the rare books that captures the reality of the fieldwork experience. It may sound romantic to live in the wilderness for weeks on end, but as a former Arctic researcher myself, I could relate to Oakes's descriptions of the austere living conditions and to the fact that reentry into society can be difficult.

In Search of the Canary Tree includes many anecdotes from Oakes's personal life. In one passage, for example, she breaks up with her boyfriend over a satellite phone but hardly has time to process it before she has to hang up and chase away a brown bear. After her father dies unexpectedly, she compares the loss of a parent with the loss of a culturally important tree such as the yellow-cedar. Rather than detract from the story, these personal moments are an important part of it, reminding readers that scientists are people, not objective automatons.

Oakes has few preconceptions about where her research will lead and is willing to be surprised. She's surprised by what she finds in her ecological studies of yellow-cedar stands and by what she learns from the people she interviews. True, there is an attendant grief, given the grim nature of her research, but there is also hope for the future. "This isn't a situation where any one person or group has the blueprint," she writes. "[W]e craft it together through actions big and small, through care for one another, through openness to a form of society that isn't what was or is today, but, rather, is still yet to come." ■

The reviewer is a freelance science writer and editor and cofounder of Science Borealis, *Canada's science blog aggregator. Email: snowhydro1@gmail.com*

10.1126/science.aav2345

PHOTO: BOB STEFKO/GETTY IMAGES

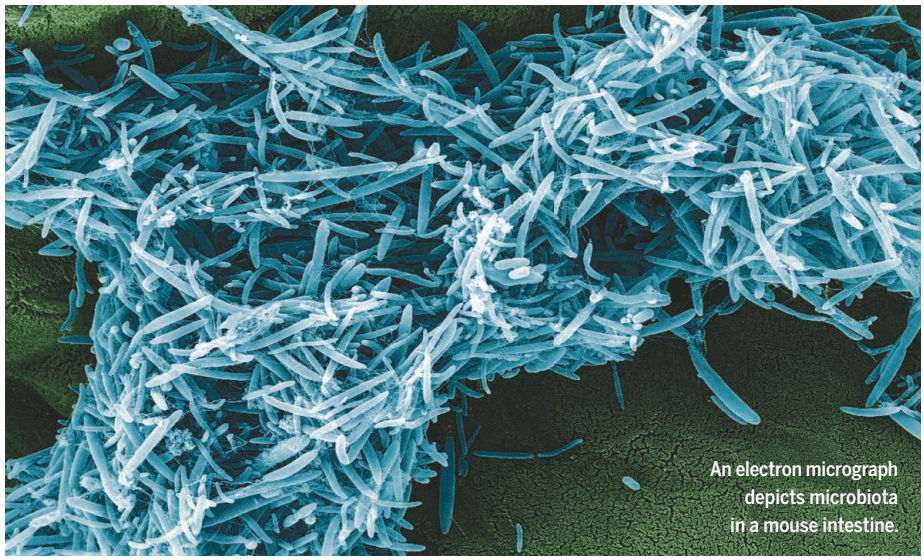
PRIZE ESSAY

TRANSLATIONAL MEDICINE

Microbiome dynamics in obesity

The microbiome influences central features of metabolic disease

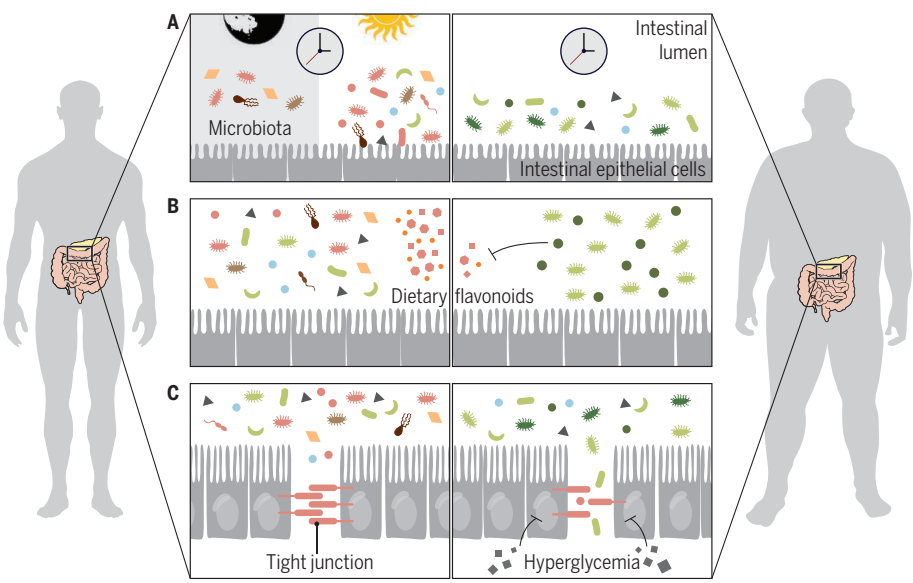
By **Christoph A. Thaiss**



An electron micrograph depicts microbiota in a mouse intestine.

Temporal and spatial microbiome dynamics in obesity

(**A**) The microbiome undergoes oscillations in composition, functional activity, and localization over the course of a day. These diurnal patterns are aberrant in obesity. (**B**) The obesity-associated microbiota degrade diet-derived flavonoids, thereby diminishing energy expenditure of the host and predisposing it to accelerated weight regain after successful dieting. (**C**) Hyperglycemia disrupts epithelial barrier function and facilitates the influx of microbiota-derived products into the systemic circulation, thereby driving chronic inflammation.



The prevalence of obesity has increased at an astounding rate over the past decades. More than 44% of the global population is estimated to be overweight, and more than 300 million individuals are affected by morbid adiposity (*1*). Obesity is a major risk factor for a number of co-occurring diseases, including type II diabetes mellitus, nonalcoholic fatty liver disease, and ischemic cardiovascular disease (*2*). Thus, the obesity pandemic has far-reaching consequences on life expectancy, quality of life, and health-care costs.

What has caused this rapid increase in obesity within less than a generation? The past century has seen dramatic changes in human lifestyle, ranging from new dietary patterns to improved hygiene and altered sleep-wake cycles. However, the molecular and cellular mechanisms by which these environmental factors predispose humans to obesity remain largely unknown. As a graduate student in Eran Elinav's laboratory at the Weizmann Institute of Science, I investigated three phenomena of human obesity that are tightly linked to the modern human lifestyle. For all three, we discovered an unexpected role for temporal and spatial dynamics of the intestinal microbiome—the community of trillions of microorganisms that inhabit the gastrointestinal tract (see the figure).

CIRCADIAN REGULATION OF MICROBIOME OSCILLATIONS

The body's circadian clock evolved in most species to adjust the organism's physiology to the daily environmental fluctuations that are caused by Earth's rotation on its axis (*3*). Disruption of the circadian clock—by shift work, frequent jet lag, or exposure to artificial light during the night—is a widespread consequence of the modern human lifestyle. Epidemiologically and experimentally, clock disturbances have been linked to metabolic diseases, including obesity and hyperglycemia (*4*, *5*).

In our initial experiments, we discovered that the gut microbiome in both mice and humans undergoes oscillations in a 24-hour rhythm (*6*). These diurnal fluctuations encompass multiple characteristics of the microbiota, such as taxonomic composition, metagenomic function, metabolite secretion, and biogeographical localization within the intestine.

We then showed that daily oscillations in the intestinal microbial community influence the circadian biology of the host, by determining the oscillatory pattern of serum polyamines. These, in turn, orchestrate the circadian transcriptional program in metabolic tissues (*7*). The host-microbiome interface can thus be viewed as a dynamic interplay that undergoes hour-scale fluctuations over the course of a day. Aberrations in this coordinated daily interplay may contribute to the development of obesity and its complications.

Science & SciLifeLab PRIZE For Young Scientists

CREDITS: (PHOTO) C. A. THAISS, M. LEVY, E. ELINAV; (GRAPHIC) N. DESAI/*SCIENCE* BASED ON CHRISTOPH A. THAISS

Published by AAAS

GRAND PRIZE WINNER: TRANSLATIONAL MEDICINE

Christoph A. Thaiss

Christoph A. Thaiss performed his undergraduate studies at the University of Bonn, Yale University, and ETH Zurich, and received his Ph.D. in immunology from the Weizmann Institute of Science in Rehovot, Israel. After the completion of his graduate studies, he became an assistant professor at the Perelman School of Medicine at the University of Pennsylvania. His lab studies environmental impacts on metabolic and inflammatory diseases.

FINALIST: ECOLOGY AND ENVIRONMENT

Matthew Savoca

Matthew Savoca received his undergraduate degree from Cornell University and his Ph.D. in Ecology from the University of California, Davis. He is currently a postdoctoral researcher at the Hopkins Marine Station of Stanford University studying the foraging behavior of baleen whales and their possible interactions with plastic debris.

www.sciencemag.org/content/362/6417/904.1

FINALIST: CELL AND MOLECULAR BIOLOGY

Ruixue Wan

Ruixue Wan received her undergraduate degree from Sun Yat-sen University in Guangzhou, China and her Ph.D. from Tsinghua University in Beijing. She is currently a postdoctoral fellow at Tsinghua, where she is conducting research on structural and biochemical investigations of the spliceosome to elucidate the mechanism of pre-mRNA splicing and the regulation of the spliceosome.

www.sciencemag.org/content/362/6417/904.2

FINALIST: GENOMICS AND PROTEOMICS

Tim Wang

Tim Wang received his undergraduate degree from the University of California, Berkeley, and his Ph.D. from the Massachusetts Institute of Technology. He is currently investigating the structural and functional organization of neocortical circuits as a postdoctoral fellow at the Janelia Research Campus in the laboratory of K. Svoboda.

www.sciencemag.org/content/362/6417/904.3

We found that disruption of host circadian rhythmicity, either genetically or by jet lag, provokes the development of an altered microbial community whose properties predispose the host to obesity and glucose intolerance (*6*). Circadian rhythms, it seems, are a new principle in host-microbiome interactions with a profound impact on metabolism.

MICROBIOME "MEMORY" IN RECURRENT OBESITY

We also investigated microbiome dynamics on larger time scales. One of the biggest conundrums in the clinical management of obesity is the rapid weight regain of formerly obese individuals after successful weight loss, a phenomenon commonly known as the "yo-yo effect" of recurrent obesity.

We found, in mice, that a period of obesity induced long-lasting alterations in the composition of the microbiome that persisted even after the host organism returned to normal weight (*8*). This "memory-like" behavior of the microbiome mediated the susceptibility of the formerly obese host to accelerated weight regain.

The weight-modulating effect of the microbiome was achieved by the degradation of dietary flavonoids, which resulted in reduced metabolic activity of adipose tissue and diminished energy expenditure. When we restored flavonoid levels, energy expenditure was normalized, and weight regain was ameliorated. Manipulation of the microbiota and intestinal metabolites as a means of achieving and maintaining weight loss is now being tested in humans.

MICROBIOME INVASION FACILITATED BY HYPERGLYCEMIA

The third phenomenon we addressed is the enhanced susceptibility of obese and diabetic individuals to enteric infection and systemic inflammation. Intestinal epithelial cells normally provide a tight barrier that separates intestinal bacteria from the systemic circulation.

In analyses of obese and diabetic mice, we found that chronic hyperglycemia led to a breakdown of the epithelial barrier and facilitated translocation of intestinal bacterial components into the bloodstream, where they promoted systemic inflammation (*9*). The detrimental effect of chronically high levels of glucose was mediated by functional reprogramming of epithelial metabolism and transcription. These cellular alterations compromised the assembly of tight junction complexes, which are critical for the maintenance of an intact barrier.

We validated these findings in a human cohort, in which chronic hyperglycemia strongly correlated with the amount of microbial products detected in the circulation. These findings suggest a mechanism linking obesity with an enhanced risk for mucosal infection and systemic inflammation. They also suggest that glycemic control may be critically important to achieve local containment of the intestinal microbiota in diabetic individuals.

TOWARD A POSTBIOTIC FUTURE

My Ph.D. work highlights temporal and spatial microbiome dynamics as an important element of host-microbiome interactions in obesity. We hypothesize that alterations in this dynamic interplay are centrally involved in the impact of the modern human lifestyle on the development of obesity and associated comorbidities.

A common theme emerging from the three features discussed above is that the impact of the microbiome on human health and disease is largely mediated by the exchange of metabolites (for example, polyamines act on the circadian clock, flavonoids on adipose tissue, and glucose on intestinal epithelial cells). As such, this work provides evidence for a powerful new therapeutic concept, which—rather than interfering with the intestinal microbial community itself (via prebiotics and probiotics)—may be based on the modulation of intestinal metabolites with a direct effect on host metabolism (analogously termed "postbiotics") (*10*, *11*). Exploring the postbiotic universe of microbiota-derived metabolites that modulate host physiology may yield interesting insights, enabling us to understand and, indeed, counteract the rapid rise of obesity worldwide. ■

Microbiology Department, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA. Email: thaiss@pennmedicine.upenn.edu

REFERENCES AND NOTES

1. M. Ng *et al.*, *Lancet* **384**, 766 (2014).
2. J. P. Despres, I. Lemieux, *Nature* **444**, 881 (2006).
3. J. Bass, J. S. Takahashi, *Science* **330**, 1349 (2010).
4. O. M. Buxton *et al.*, *Sci. Trans. Med.* **4**, 129ra43 (2012).
5. F. A. Scheer, M. F. Hilton, C. S. Mantzoros, S. A. Shea, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 4453 (2009).
6. C. A. Thaiss *et al.*, *Cell* **159**, 514 (2014).
7. C. A. Thaiss *et al.*, *Cell* **167**, 1495 (2016).
8. C. A. Thaiss *et al.*, *Nature* **540**, 544 (2016).
9. C. A. Thaiss *et al.*, *Science* **359**, 1376 (2018).
10. C. A. Thaiss, N. Zmora, M. Levy, E. Elinav, *Nature* **535**, 65 (2016).
11. K. Tsilingiri, M. Rescigno, *Benef. Microbes* **4**, 101 (2013).

ACKNOWLEDGMENTS

I am grateful to my Ph.D. adviser, E. Elinav, whose enthusiasm, vision, and mentorship have greatly inspired me. I also thank all the members of the Elinav laboratory and my collaborators for their advice, help, and friendship.

10.1126/science.aav6870

PRIZE ESSAY

FINALIST: ECOLOGY AND ENVIRONMENT

Matthew Savoca

Matthew Savoca received his undergraduate degree from Cornell University and his Ph.D. in Ecology from the University of California, Davis. He is currently a postdoctoral researcher at the Hopkins Marine Station of Stanford University studying the foraging behavior of baleen whales and their possible interactions with plastic debris.

www.sciencemag.org/content/362/6417/904.1

ECOLOGY AND ENVIRONMENT

The ecology of an olfactory trap

Chemical cues lead marine organisms to consume plastic debris

By **Matthew Savoca**

Clean up after yourself. We are taught this adage from a young age, and yet these words often seem to be discarded before adulthood, especially in regard to the global commons (*1*). Our largest global commons is the ocean. It is also the sink for much of our waste.

Plastics are a ubiquitous form of marine litter; these synthetic materials have been found in oceans worldwide as well as in sea ice, seafloor sediment, and distressingly, in seafood (*2*, *3*). To date, nearly 1000 marine species, ranging from invertebrates to whales, have been found to ingest plastic debris.

Since the late 1960s, researchers have focused on documenting not only which species are eating plastic but in what quantities, in what regions, and if/how ingested plastic affects the consumers and associated food webs. Despite burgeoning scientific interest on the interactions between plastic waste and marine wildlife, far less research has focused on the behavioral mechanisms underlying the maladaptive decision to ingest plastic. Most of what has been published is largely speculative and based solely on the visual cues of plastic debris.

Think of the common explanation: "Sea turtles eat plastic bags because they look like jellyfish." My doctoral research found that in addition to having the appearance of food, plastic also releases a biochemical signature that may trick organisms into confusing our trash for their treasure (see the figure).

The story begins with the odoriferous algal-derived molecule dimethyl sulfide (DMS). When phytoplankton are consumed by zooplankton, DMS is released en masse—a result of dimethylsulfoniopropionate (DMSP) catabolism—from the depredated phytoplankton (*4*). Marine predators—including certain species of fish, seabirds, and sea turtles—track relative concentrations of DMS over the open ocean in order to find productive regions to forage [reviewed in (*5*)]. I first used a 50-year seabird diet database, coupled with decades of data on seabirds' olfactory preferences, to discover that DMS-tracking is an ecologically predictable trait.

I determined that procellariiform seabirds—a group that includes burrow-nesting petrels and shearwaters—were likely to follow an odor trail of DMS, not to find the phytoplankton producing it but rather to locate nutrient-rich zooplankton (such as krill) (*6*). Zooplankton consumption of phytoplankton is often responsible for local elevations of DMS (*4*). In addition, these recruited seabirds provide limiting nutrients such as iron via their waste to the phytoplankton (*5*, *6*). This work provided evidence of a chemically mediated tritrophic mutualism in the marine environment, but the connection to marine debris was yet to be explored.

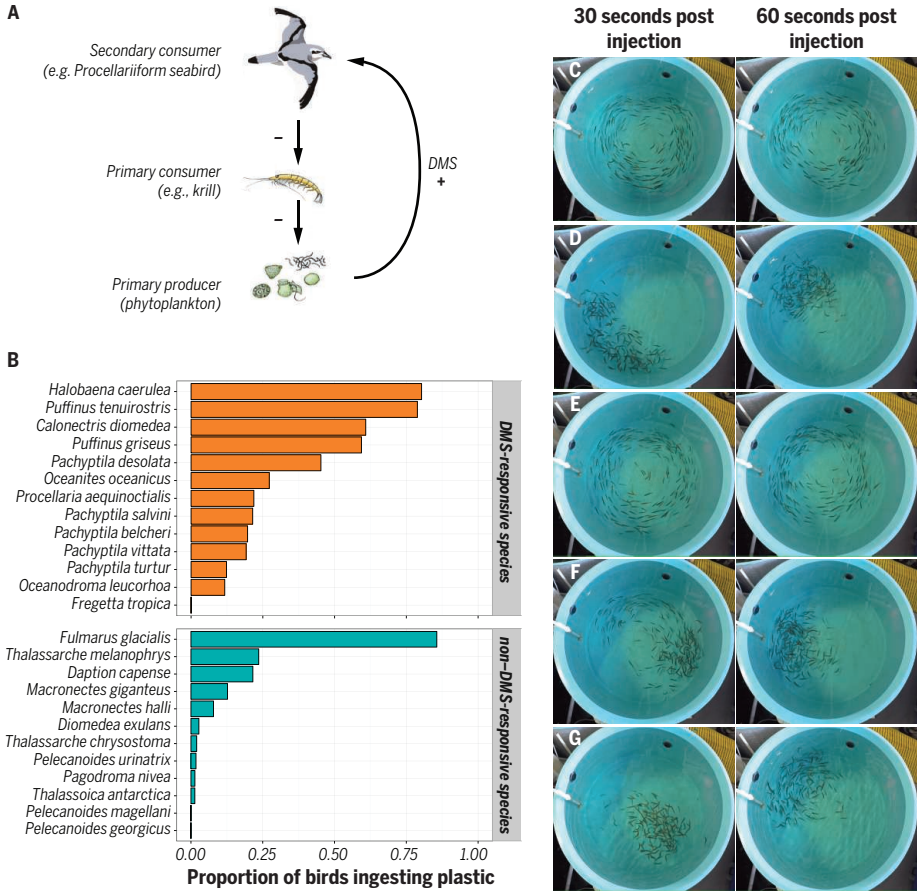
Procellariiform seabirds possess an acute olfactory sense and are also severely affected by the ingestion of plastic (*7*). The notion that these birds consume plastic simply because of its appearance is incongruous with decades of research demonstrating that procellariiformes are highly reliant on odor cues for foraging (*8*–*10*). In addition, some species never ingest plastic, whereas others seem unable to avoid it. This unexplained interspecific heterogeneity is puzzling.

To confront this problem scientifically, I created a database by amassing every published record of plastic ingestion in all procellariiform species. This database ultimately included 73 studies, with plastic ingestion data from 20,922 individuals of 65 species. For approximately half of these species, their behavioral attraction, or lack thereof, to DMS was known. When I analyzed how often each species ingested plastic debris in relation to their olfactory attraction to DMS, the result was startling: Those species that use DMS to forage were almost six times more likely to consume plastic than those that are not attracted to DMS (*11*). Finding a result this pronounced on a dataset this robust is unusual in ecology, so I knew I needed to dive deeper.

The next question to consider was, what does ocean-seasoned plastic smell like to a foraging seabird? To answer this, I tethered three types of preproduction plastic—

Hopkins Marine Station, Department of Biology, Stanford University, Pacific Grove, CA, USA. Email: msavoca13@gmail.com

PHOTO: MATT KOLLER



Evidence for an olfactory trap. (**A**) Simplified diagram of the tritrophic interaction between phytoplankton and top predators (such as Procellariiform seabirds), mediated by DMS. [from (*5*).] (**B**) Plastic ingestion and DMS responsiveness among procellariform species. Horizontal bars represent plastic ingestion prevalence by species ordered from high to low in each group. The red bars depict DMS-responsive species (n = 13), and the blue bars depict non–DMS-responsive species (n = 12) [Modified from (*11*)]. (C to G) School-wide behavioral responses of northern anchovy (*E. mordax*) to treatments of (**C**) control (dyed seawater), (**D**) food, (**E**) clean plastic odor, (**F**) biofouled plastic odor, and (**G**) food odor, taken from videos 30 and 60 s after treatment introduction. [Modified from (*12*)]

PHOTO: N. DESAI/*SCIENCE* BASED ON (12) AND MATTHEW SAVOCA

polypropylene and high- and low-density polyethylene, which together make up two-thirds of consumer plastics—to buoys at two different locations off the coast of California during the summers of 2013 and 2014.

After the plastic was left at sea for a month, I brought these samples to the Robert Mondavi Institute for Wine and Food and Sciences at the University of California, Davis. I used a gas-chromatography technique that is typically used on wine and food products and found that every sample of ocean-weathered (i.e. biofouled) plastic emitted DMS at concentrations six orders of magnitude above the detection threshold for procellariiforms and three orders of magnitude above background DMS in the environment (*11*). Furthermore, there was no DMS signature on clean plastics (*11*).

The distinct DMS signature associated with marine plastic debris originates from the algal biofilm that rapidly coats plastic at sea. The implication is that marine plastics may falsely amplify an olfactory signal that certain species associate with foraging opportunities.

This compelling evidence required further support, in the form of a behavioral assay that would test how the chemical signature of marine plastic affected the foraging behavior of a marine organism. In the summer of 2015, I tested this question on Northern Anchovy (*Engraulis mordax*), an ecologically critical forage fish species and known plastic consumer (*2*). Anchovy responded to odors of food and plastic debris similarly, increasing clustering behavior and reducing school-wide rheotaxis in a manner that is characteristic of olfactory search (*12*). These behaviors are consistent with foraging in this species.

To understand a problem as pervasive and perplexing as plastic debris ingestion requires careful consideration of how the organisms we study experience their world. Taken together, these results suggest that plastic debris may be more confusing and appetizing to marine organisms than previously thought possible. Moreover, the results of my work underscore the urgency of mitigating plastic pollution at a global scale. ■

REFERENCES

1. G. Hardin, *Science* **162**, 1243 (1968).
2. C. M. Rochman *et al.*, *Sci. Rep.* **5**, 14340 (2015).
3. C. Ioakeimidis, F. Galgani, G. Papatheodorou, in *The Handbook of Environmental Chemistry*, H. Takada, H. K. Karapanagioti, Eds. (Springer, 2017), pp. 1–28.
4. J. W. Dacey, S. G. Wakeham, *Science*. **233**, 1314 (1986).
5. M. S. Savoca, *Biogeochemistry*. **138**, 1 (2018).
6. M. S. Savoca, G. A. Nevitt, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4157 (2014).
7. C. Wilcox, E. Van Sebille, B. D. Hardesty, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11899 (2015).
8. T. C. Grubb, *Nature* **237**, 404 (1972).
9. C. Verheyden, P. Jouventin, *Auk*. **111**, 285 (1994).
10. G. A. Nevitt, K. Reid, P. Trathan, *J. Exp. Biol.* 207, 3537 (2004).
11. M. S. Savoca, M. E. Wohlfeil, S. E. Ebeler, G. A. Nevitt, *Sci. Adv.* 2, e1600395 (2016).
12. M. S. Savoca, C. W. Tyson, M. McGill, C. J. Slager, *Proc. R. Soc. B Biol. Sci.* 10.1098/rspb.2017.1000 (2017).

10.1126/science.aav6873

PRIZE ESSAY

FINALIST: CELL AND MOLECULAR BIOLOGY

Ruixue Wan

Ruixue Wan received her undergraduate degree from Sun Yat-sen University in Guangzhou, China, and her Ph.D. from Tsinghua University in Beijing. She is currently a postdoctoral fellow at Tsinghua, where she is conducting research on structural and biochemical investigations of the spliceosome to elucidate the mechanism of pre-mRNA splicing and the regulation of the spliceosome.

www.sciencemag.org/content/362/6417/904.2

CELL AND MOLECULAR BIOLOGY

A key component of gene expression, revealed

High-resolution microscopy sheds light on the molecular mechanisms of the spliceosome

By **Ruixue Wan**

In eukaryotes, genetic information stored in DNA is first transcribed into precursor messenger RNAs (pre-mRNAs), in which noncoding (introns) and coding regions (exons) appear alternatively. Introns must be removed, and exons are ligated to generate a mature mRNA that contains continuous protein-coding sequences through a sophisticated process called pre-mRNA splicing, which was discovered 40 years ago (*1, 2*). Splicing comprises two sequential transesterification reactions, branching and exon ligation, executed by a multi-megadalton, ribonucleoprotein (RNP) complex known as the spliceosome (*3*).

The spliceosome is composed of five uridine-rich, small nuclear RNPs (U1, U2, U4, U5, and U6 snRNPs—each of which contains one U snRNA and associated proteins), the NineTeen complex (NTC), the NTC-related complex (NTR), and various protein factors. To catalyze splicing, the spliceosome must recognize the three conserved splice sites (5′ and 3′ splice sites at exon-intron junctions and the branch point sequence) and undergo precise stepwise assembly and remodeling (Fig. 1A).

In the three decades between the discovery of the spliceosome (*4*–*6*) and the beginning of this century, scientists sought to identify its constitutional components and biochemically analyze the various spliceosomal complexes by immunoprecipitation, cross-linking mass spectrometry, in vitro splicing essay, and so on. However, our understanding of the molecular mechanism was impeded owing to the lack of high-resolution structures of the fully assembled spliceosome.

The spliceosome exhibits exceptional conformational flexibility and compositional dynamics. Compared to the relatively stable structures of RNA polymerase and the ribosome, an atomic structure of the spliceosome was long thought to be insurmountable. Electron microscopy (EM) studies achieved only moderate resolutions, with the highest at 20 to 29 Å, which only unveiled the overall shape and global features (*7*). For my Ph.D. thesis research, I chose to confront this challenging target, investigating the molecular mechanism of the pre-mRNA splicing process by capturing the atomic structures of the spliceosome using the rapidly developing cryo-EM technology.

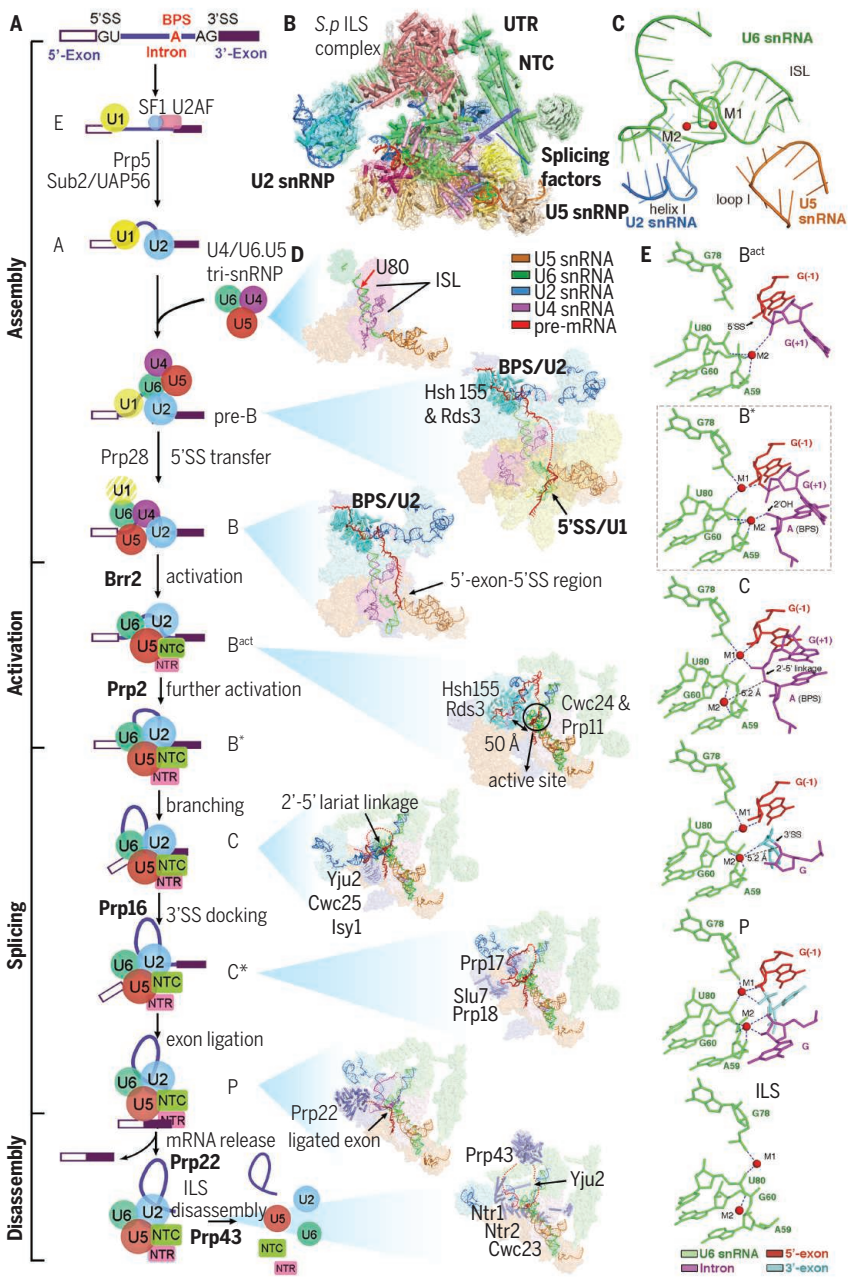
By tagging Cdc5 with TAP-tag in *Schizosaccharomyces pombe*, my co-workers and I were able to purify the intact spliceosome, which eventually yielded a structure at the overall resolution of 3.6 Å (*8, 9*) (Fig. 1B). This structure, containing 37 proteins and four RNA molecules, reveals the organizational principles of a functional spliceosome and the RNA-based splicing active site arrangement for the first time (Fig. 1B, C).

The active site is located in the heart of the spliceosome, consisting of the intramolecular stem-loop (ISL) of U6 snRNA, the associated Mg^{2+} ions, helix I of the U2/U6 duplex, and loop I of U5 snRNA. Two catalytic metal ions, M1 and M2, are coordinated by zigzagged U6 snRNA in the active site. Thus, the spliceosome is in essence a protein-directed metalloribozyme. Analysis suggests that it represents a late-stage spliceosome known as the intron-lariat spliceosome or ILS complex because of the absence of exon sequences and the 20 Å distance separated from the lariat junction away to the catalytic ions.

To further investigate the mechanism of spliceosome assembly and catalysis, we sought to capture the structures of the spliceosome at other working states by tagging different protein components and overexpressing adenosine triphosphatase (ATPase)–defective mutants. Prior to my thesis defense, we solved the atomic or near-atomic structures of nearly all key intermediates of the spliceosome, including pre-B and B (*10*), B^{act} (*11*), C (*12*), C^{*} (*13*), P (*14*), and ILS complexes (*15*) from *S. cerevisiae* (Fig. 1D). We also determined the structure of the preassembled U4/U6.U5 tri-snRNP from *S. cerevisiae* (*16*). Local resolution of all the structures in the core regions reaches at least 3.0 Å, allowing us to accurately identify the catalytic metal coordination during the splicing reaction (Fig. 1E).

Beijing Advanced Innovation Center for Structural Biology, School of Medicine, Tsinghua University, Beijing 100084, China
Email: wrxruisnow@163.com

PHOTO: COURTESY OF RUIXUE WAN/BEIJING ADVANCED INNOVATION CENTER FOR STRUCTURAL BIOLOGY



Molecular mechanism of pre-mRNA splicing revealed by structures of the spliceosome at different states. (**A**) A scheme of the assembly and catalysis cycle of the spliceosome. This process includes four phases: assembly based on the pre-mRNA substrate, activation to prepare for the reaction, two-step splicing reaction, and disassembly of the spliceosome. State transition of the spliceosome is mainly driven by eight DExD/H ATPases/helicases (colored red). (**B**) The cryo-EM structure of the ILS complex from *S. pombe* (PDB code 3JB9). (**C**) A close-up view of the active site of *S.pombe* ILS complex. (**D**) Cryo-EM structures of the U4/U6.U5 tri-snRNP, pre-B, B, B^{act}, C, C*, P, and ILS complex from *S. cerevisiae*. Structures are color-coded by subcomplexes related to (A). RNA components are highlighted in cartoon style. The orientation and scale are presented the same for comparison. (**E**) Coordination of catalytic metal ions during splicing reaction (from B^{act} to ILS complex, B* is predicted). The RNA elements are shown in the same orientations.

These structures provide unprecedented clarity toward a mechanistic understanding of pre-mRNA splicing. It appears that the spliceosome catalyzes the two steps of splicing reaction using a single active site, which remains almost unchanged once formed in the B^{act} complex. The U6 snRNA coordinates the catalytic metal ions and plays a notable role in catalyzing the two steps of splicing. U1, U2, U5, and U6 snRNAs serve to recognize and transfer the RNA elements on the pre-mRNA, including 5′ and 3′ splice sites, branch point sequence (BPS) and 5′ exon (Fig. 1D). U4 snRNA plays an important role in sequestering U6 snRNA before the reaction.

Throughout the splicing reaction, at least 20 components (mainly from U6 snRNA, U5 snRNP, NTR, and a small portion of U2 snRNA and NTC) act as a rigid scaffold, maintaining the conformation of the active site. The 3′ end of U2 snRNA and ~13 proteins (part of U2 snRNP and NTC) are compositionally unchanged, but mobile, facilitating the delivery of critical reaction groups on the RNA molecules into the active site for the two-step reactions. The compositional change and remodeling of the spliceosome are mainly driven by ATPase/helicases and its cofactors. Thus, a near-complete cycle of pre-mRNA splicing can be recapitulated in atomic details (Fig. 1).

As an essential step in eukaryotic gene expression, splicing is relevant to many diseases (*17*). For example, Hsh155 is frequently mutated in hematological malignancies (*18*). Based on the structure of the B^{act} complex, we mapped *36* cancer-derived mutations to Hsh155. Most of these residues bind upstream of BPS, thus compromising the binding ability and adversely affect splicing.

Forty years after the discovery of pre-mRNA splicing, we have finally begun to understand its underlying structural mechanism. This work will enable the understanding of the intricacies of an essential biological process. It will also aid drug discovery for numerous diseases that are linked to splicing malfunction, including many cancers, blindness, and other genetic diseases. The next challenge is to investigate the more complex regulation mechanism of the spliceosome, such as alternative splicing. ■

REFERENCES AND NOTES

1. S. M. Berget, C. Moore, P. A. Sharp, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3171 (1977).
2. L. T. Chow, R. E. Gelinas, T. R. Broker, R. J. Roberts, *Cell* **12**, 1 (1977).
3. M. J. Moore, C. C. Query, P. A. Sharp, *Splicing of Precursors to mRNA by the Spliceosome*. The RNA World (Cold Spring Harbor Monograph Archive, 1993).
4. E. Brody, J. Abelson, *Science* **228**, 963 (1985).
5. D. Frendewey, W. Keller, *Cell* **42**, 355 (1985).
6. P. J. Grabowski, S. R. Seiler, P. A. Sharp, *Cell* **42**, 345 (1985).
7. R. Lührmann, H. Stark, *Curr. Opin. Struct. Biol.* **19**, 96 (2009).
8. C. Yan *et al.*, *Science* **349**, 1182 (2015).
9. J. Hang, R. Wan, C. Yan, Y. Shi, *Science* **349**, 1191 (2015).
10. R. Bai, R. Wan, C. Yan, J. Lei, Y. Shi, *Science* **360**, 1423 (2018).
11. C. Yan, R. Wan, R. Bai, G. Huang, Y. Shi, *Science* **353**, 904 (2016).
12. R. Wan, C. Yan, R. Bai, G. Huang, Y. Shi, *Science* **353**, 895 (2016).
13. C. Yan, R. Wan, R. Bai, G. Huang, Y. Shi, *Science* **355**, 149 (2016).
14. R. Bai, C. Yan, R. Wan, J. Lei, Y. Shi, *Cell* **171**, 1589 (2017).
15. R. Wan, C. Yan, R. Bai, J. Lei, Y. Shi, *Cell* **171**, 120 (2017).
16. R. Wan *et al.*, *Science* **351**, 466 (2016).
17. M. M. Scotti, M. S. Swanson, *Nat. Rev. Genet.* **17**, 19 (2016).
18. C. N. Hahn, H. S. Scott, *Nat. Genet.* **44**, 9 (2011).

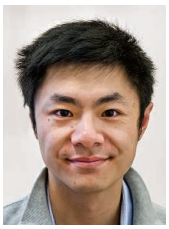
ACKNOWLEDGMENTS

I am extremely grateful to my Ph.D. adviser, Y. Shi, for his mentorship. I also thank all of the members of the Shi lab, both past and present, especially L. Zhou and R. Bai.

10.1126/science.aav6875

GRAPHIC: RUIXUE WAN

PRIZE ESSAY



FINALIST: GENOMICS AND PROTEOMICS

Tim Wang

Tim Wang received his undergraduate degree from the University of California, Berkeley, and his Ph.D. from the Massachusetts Institute of Technology. He is currently investigating the structural and functional organization of neocortical circuits as a postdoctoral fellow at the Janelia Research Campus in the laboratory of K. Svoboda.

www.sciencemag.org/content/362/6417/904.3

GENOMICS AND PROTEOMICS

Paring down to the essentials

A genome-wide library of single gene knockouts enables high-throughput functional analysis of mammalian genes

By **Tim Wang**

In the early 1950s, television sets were complicated and expensive, putting them beyond the reach of most American households. To lower prices, businessman and engineer Earl "Madman" Muntz devised an ingenious solution (*1*). Armed with a pair of snippers, he would take a competitor's product and proceed to pluck off component after component until, inevitably, the device would break. At this point, he would put back the last piece removed and thereby restore function. In the end, he was left with a minimal design that contained only the essential circuit elements, which could then be manufactured at a reduced cost.

This practice of "Muntzing" is not only a clever way to dissect electronic circuitry but also a powerful method by which to probe the genetic circuitry of biological organisms. By disrupting individual genes (either at random or in a systematic fashion) and observing the resultant phenotypes, one can efficiently identify the set of genes underlying any cellular pathway.

In several model organisms, large-scale genetic screens have provided insights into fundamental biological processes, including the cell cycle (*2*), programmed cell death (*3*), and embryonic development (*4*). In mammalian cells, however, the development of similar techniques has lagged behind, leaving our understanding of human genes incomplete.

As a graduate student working jointly in David M. Sabatini's laboratory at the Whitehead Institute and Eric Lander's laboratory at the Broad Institute, I set out to develop a general strategy for conducting genome-wide loss-of-function screens in human cells (see the figure). To accomplish this task, I used the clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9 system, which functions as an adaptive immune system in bacteria [reviewed in (*5*, *6*)]. During my second year in the laboratory, two groups had reported a method for mammalian genome editing via heterologous expression of two components of the CRISPR pathway: a Cas9 nuclease that is directed to cut specific genomic sequences by means of a single-guide RNA (sgRNA) (*7*, *8*).

To develop a platform that would be amenable for large-scale genetic screening, I designed a library of sgRNAs that targets all protein-coding genes in the human genome (*9*). This library could be introduced into any cultured cell line to generate a collection of "knockout" mutants in which a single gene is inactivated in each cell. Using high-throughput sequencing, the abundance of each mutant in the population can be tracked during the course of an experiment.

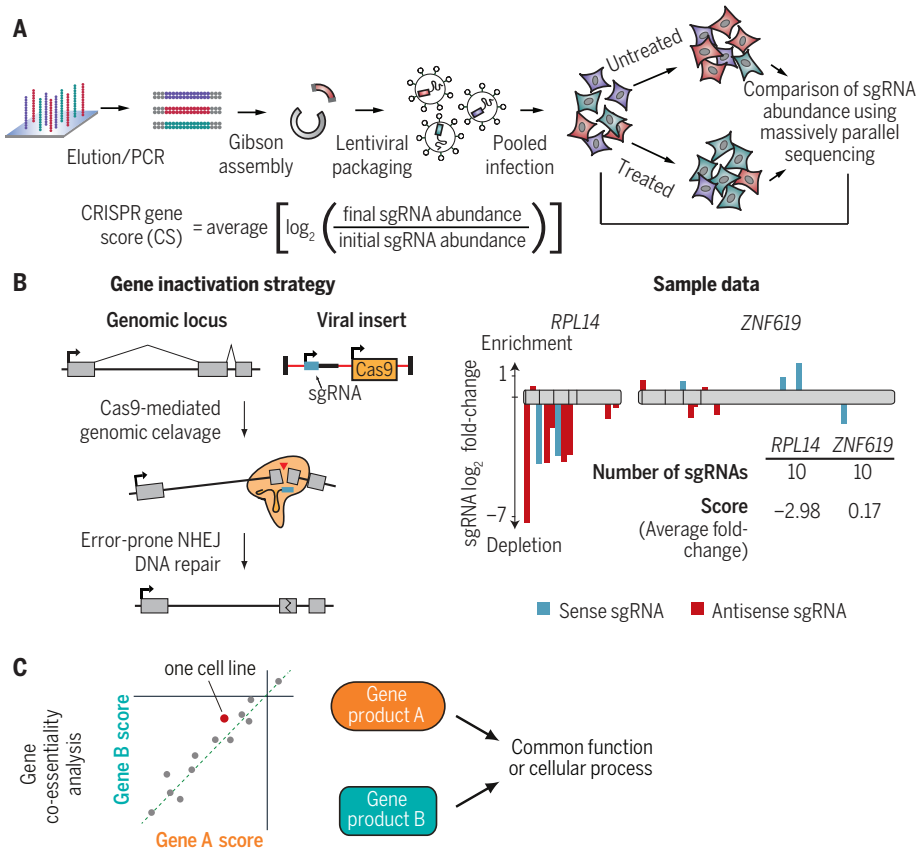
In an initial study, I applied this technique to search for genes that are necessary for human cells to grow and divide in culture (*10*). This survey produced a provisional catalog of all cell-essential genes. As a whole, these genes show broad conservation across species, are expressed at high levels, and contain few inactivating polymorphisms in the human population. The majority of these genes encode well-defined components of fundamental cellular pathways (such as ribosomal subunits).

Surprisingly, this study also uncovered a large group of genes (more than 300) for which no molecular function had been previously ascribed. For this set of genes, bioinformatics analysis and comparison to proteomic datasets revealed substantial enrichment in proteins found in the nucleolus and those containing domains related to RNA processing. Indeed, directed studies confirmed key roles for three of these genes in various stages of posttranscriptional RNA processing. More broadly, these findings indicate that the molecular components of many critical cellular processes in mammalian cells have yet to be fully defined.

Next, I profiled gene essentiality across a much larger collection of human cancer cell lines derived from acute myeloid leukemias, which are the most common form of adult leukemia (*11*). Whereas most genes were either essential or dispensable in all lines, a small fraction showed variable essentiality across the cell lines. This latter class of genes provided two major avenues for further investigation.

Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. Email: wangt@janelia.hhmi.org

PHOTO: CEAL CAPISTRANO



CRISPR/Cas9–based genetic screening. (**A**) Outline of sgRNA library construction and genetic screening strategy. Briefly, sgRNAs are synthesized on a microarray, cloned, and packaged into lentiviral particles that can be used to transduce a cell line of interest to generate a mutant library. Cells can be cultured under various conditions, and the fold-change in abundance between the initial seeding and final collection of each mutant can be quantified by means of high-throughput sequencing. (**B**) (Left) Gene inactivation through Cas9-mediated genomic cleavage is directed by a 20–base pair sequence at the 5′ end of the sgRNA. Lesions repaired by the error-prone nonhomologous end joining (NHEJ) pathway frequently result in nonfunctional, frame-shifting mutations. (Right) sgRNAs that target essential genes are depleted from the population, whereas those that target dispensable genes are maintained, as illustrated from the data of two neighboring genes: *RPL14*, an essential ribosomal protein gene, and *ZNF619*, a dispensable gene that encodes a zinc finger protein. (**C**) Strategy for identifying functionally related sets of genes by using the gene essentiality dataset in (*11*). Because genes acting in the same cellular pathway should show similar patterns of essentiality across cell lines, functional gene networks can be mapped through correlation-based analysis. [Adapted from (*9*–*11*)]

First, I used the "gene essentiality" dataset to pinpoint genetic dependencies in particular cancer subtypes. In contrast to core essential genes, variably essential genes are attractive targets for anticancer therapies because their inhibition is less likely to be broadly cytotoxic.

Because the genomes of these cell lines had already been characterized, it was possible to relate patterns of gene essentiality with the presence of specific mutations. This approach could not only identify the mutant oncogene driving the proliferation of each line but, importantly, could also detect vulnerabilities in genes that were not mutated (and thus not detectable with genome sequencing). For example, I uncovered several liabilities associated with oncogenic *Ras*, a frequently mutated oncogene that has proven difficult to inhibit directly (*12*). In the future, gene essentiality profiling of cancer cell lines, and even patient tumor cells, could guide personalized cancer treatment and allow for the discovery of previously unidentified gene targets for therapeutic intervention.

Second, the dataset was used to map functional gene networks. I noticed that genes acting in the same cellular pathway tended to show similar patterns of essentiality across cell lines. This observation suggested that functionally related sets of genes can be recognized by analyzing correlations in gene essentiality. By applying a "guilt-by-association" approach, genes were systematically clustered by function, which in turn revealed new gene relationships, the essential substrates of enzymes, and the molecular functions of uncharacterized proteins.

In a number of collaborative studies, CRISPR/Cas9–based screens have been performed to uncover the molecular underpinnings of diverse biological processes. Screens in a $CD4^+$ cell line, for example, have identified host factors that are required for HIV infection (*13*). Genetic analysis of cells cultured under limiting nutrient conditions or in the presence of chemical inhibitors have yielded key insights into mitochondrial function (*14*, *15*).

Others have greatly extended this CRISPR-based approach to examine noncoding regions of the genome (*16*), to dissect pairwise gene interactions (*17*), to perform gain-of-function experiments (*18*), and to analyze richer phenotypes by use of single-cell RNA sequencing (*19*).

Taken together, these efforts have enabled genetic analyses of mammalian cells to be conducted with a degree of rigor and completeness previously possible only in microorganisms. With further advances in the design of sgRNA libraries, Cas9 effector proteins, and readout strategies, it is tremendously exciting to imagine what discoveries will be made by "Muntzing" around in the cell. ■

REFERENCES AND NOTES

1. R. C. Post, *Technol. Culture* **46**, 773 (2005).
2. L. H. Hartwell, J. Culotti, B. Reid, *Proc. Natl. Acad. Sci. U.S.A.* **66**, 352 (1970).
3. H. M. Ellis, H. R. Horvitz, *Cell* **44**, 817 (1986).
4. C. Nüsslein-Volhard, E. Wieschaus, *Nature* **287**, 795 (1980).
5. L. A. Marraffini, E. J. Sontheimer, *Nat. Rev. Gen.* **11**, 181 (2010).
6. P. Horvath, R. Barrangou, *Science* **327**, 167 (2010).
7. L. Cong *et al.*, *Science* **339**, 819 (2013).
8. P. Mali *et al.*, *Science* **339**, 823 (2013).
9. T. Wang *et al.*, *Science* **343**, 80 (2014).
10. T. Wang *et al.*, *Science* **350**, 1096 (2015).
11. T. Wang *et al.*, *Cell* **168**, 890 (2017).
12. A. D. Cox *et al.*, *Nat. Rev. Drug Discov.* **13**, 828 (2014).
13. R. J. Park *et al.*, *Nat. Gen.* **49**, 193 (2017).
14. K. Birsoy *et al.*, *Cell* **162**, 540 (2015).
15. N. Kory *et al.*, *Science* **362**, 791 (2018).
16. N. E. Sanjana *et al.*, *Science* **353**, 1545 (2016).
17. F. J. Najm *et al.*, *Nat. Biotechnol.* **36**, 179 (2018).
18. S. Konermann *et al.*, *Nature* **517**, 583 (2014).
19. B. Adamson *et al.*, *Cell* **167**, 1867 (2016).

ACKNOWLEDGMENTS

The author thanks R. A. Saxton, E. S. Lander, and D. M. Sabatini for reviewing this essay and the Howard Hughes Medical Institute for funding. **Competing interests:** The author is an inventor on a patent for functional genomics using CRISPR/Cas (U.S. patent application 15/141,348) and a scientific cofounder of KSQ Therapeutics, which is using CRISPR-based genetic screens to identify to drug targets.

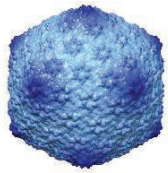
10.1126/science.aav6872

PHOTO: N. DESAI/*SCIENCE*, BASED ON (*9*–*11*) AND TIM WANG

RESEARCH

A new approach can weigh individual virus particles

Dominguez-Medina et al., *p. 918*



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

Artist's rendering of a prehistoric African landscape

IMAGE: DOMINGUEZ-MEDINA *ET AL.*; MAURICIO ANTON/SCIENCE SOURCE

PALEOECOLOGY

Megaherbivore extinctions in Africa

Human ancestors have been proposed as drivers of extinctions of Africa's diverse large mammal communities. Faith *et al*. challenge this view with an analysis of eastern African herbivore communities spanning the past ~7 million years (see the Perspective by Bobe and Carvalho). Megaherbivores (for example, elephants, rhinos, and hippos) began to decline about 4.6 million years ago, preceding evidence for hominin consumption of animal tissues by more than 1 million years. Instead, megaherbivore decline may have been triggered by declining atmospheric carbon dioxide and expansion of grasslands. —AMS

Science, this issue p. 938; see also p. 892

GENETICS

The mutational burden of aging

As people age, they accumulate somatic mutations in healthy cells. About 25% of cells in normal, sun-exposed skin harbor cancer driver mutations. What about tissues not exposed to powerful mutagens like ultraviolet light? Martincorena *et al*. performed targeted gene sequencing of normal esophageal epithelium from nine human donors of varying age (see the Perspective by Chanock). The mutation rate was lower in esophagus than in skin, but there was a strong positive selection of clones carrying mutations in 14 cancer-associated genes. By middle age, more than half of the esophageal epithelium was colonized by mutant clones. Interestingly, mutations in the cancer driver gene *NOTCH1* were more common in normal esophageal epithelium than in esophageal cancer. —PAK

Science, this issue p. 911; see also p. 893

SUPERCONDUCTIVITY

A monolayer of many talents

Superconductors with a topologically nontrivial band structure have been predicted to exhibit exotic properties. However, such materials are few and far between. Now, two groups show that the monolayer of the material tungsten ditelluride (WTe_2)—already known to be a two-dimensional topological insulator—can also go superconducting. Fatemi *et al*. and Sajadi *et al*. varied the carrier density in the monolayer by applying a gate voltage and observed a transition from a topological to a superconducting phase. The findings may lead to the fabrication of devices in which local gating enables topological and superconducting phases to exist in the same material. —JS

Science, this issue p. 926, p. 922

METALLURGY

Nanoparticle superalloy

Improving the strength of a metal alloy is hard to do without sacrificing the ductility. Yang *et al*. designed an iron-nickel-cobalt (Fe-Ni-Co) alloy laced with aluminum-titanium (Al-Ti) nanoparticles with both high strength and ductility. The key was getting the composition tuned correctly, because the Fe-Ni-Co matrix reacts with the Al-Ti nanoparticles. This was vital for avoiding environmental embrittlement, enhancing work hardening, and improving ductility. —BG

Science, this issue p. 933

IMMUNOLOGY

Fine-tuning pyroptosis with ESCRT-III

Pyroptosis is an inflammatory form of cell death induced by select caspases downstream of inflammasome complexes.

These caspases cleave gasdermin D (GSDMD), whose N-terminal fragments quickly form large permeability pores that induce cell death. However, a large percentage of cells with active inflammasomes are resistant to pyroptosis. Rühl *et al*. found that the membrane-remodeling ESCRT-III machinery was recruited to the plasma membrane upon GSDMD activation. ESCRT-III–dependent membrane repair limited proinflammatory cytokine secretion and pyroptosis after activation of inflammasomes. —STS

Science, this issue p. 956

CANCER

Taking aim at a childhood cancer

Rhabdomyosarcoma is a difficult-to-treat soft tissue pediatric tumor. In a particular subtype of this cancer, a fusion protein generated by a chromosomal abnormality drives chemoresistance and aggressive progression. Bharathy *et al*. investigated why the histone deacetylase inhibitor entinostat shows promise in treating this rhabdomyosarcoma subtype. Entinostat altered epigenetic regulation to inhibit translation of the fusion protein. Without the fusion protein, rhabdomyosarcoma growth in mouse models slowed, and tumors were sensitized to the chemotherapeutic drug vincristine. —LKF

Sci. Signal. **11**, eaau7632 (2018).

Normally cooperative ants change their behavior toward sick individuals.

EPIDEMIOLOGY

Protecting the colony

When we get a cold and then stay home from work, we are not only taking care of ourselves but also protecting others. Such changes in behavior after infection are predicted in social animals but are difficult to quantify. Stroeymeyt *et al*. looked for such changes in the black garden ant and found that infected workers did alter their behavior—and healthy workers altered their behavior toward the sick. The changed behavior was especially valuable for protecting the most important and vulnerable members of the colony. —SNV

Science, this issue p. 941

ROBOTIC SENSING

A firm, but gentle, touch

Many seemingly simple manual tasks require the ability to detect force direction as well as magnitude. Boutry *et al*. developed an electronic skin inspired by human skin and nature. Tiny pyramids, similar to hill-like structures in human skin, were arranged in spirals, like the center of a sunflower. These microstructures enabled the sensor array to differentiate pressure applied perpendicularly from pressure applied at an angle—a key feature for dexterity. The pressure information was used to interrupt automatic movement of a robotic arm. The robot was able to touch a fresh raspberry and reverse motion quickly enough to avoid damaging the fruit. —RLK

Sci. Robot. **3**, eaau6914 (2018).

EMERGING INFECTIONS

Antibodies to abrogate Andes hantavirus

Andes hantavirus circulates in rodent reservoirs and can cause hantavirus cardiopulmonary syndrome in humans. This results in a potentially lethal disease for which no vaccines or targeted treatments exist. Garrido *et al*. screened memory B cells from people that had been infected with Andes hantavirus. Antibodies were isolated from one individual with a high viral neutralization capacity. Two of these antibodies were fully protective against disease in a hamster model, even when given several days after infection. These antibodies target distinct epitopes on the viral glycoprotein and could be developed for use alone or as a combination therapy. —LP

Sci. Transl. Med. **10**, eaat6420 (2018).

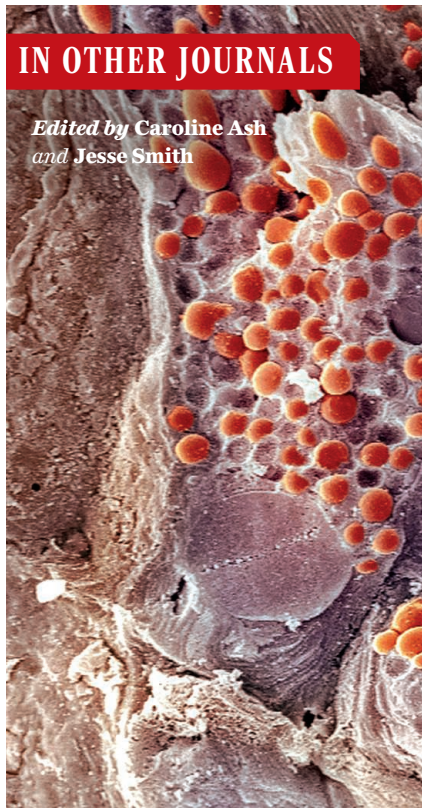


PHOTO: DANI DANIAR/ALAMY STOCK PHOTO

PHYSIOLOGY

Could microbes be diabetogenic?

Microbes that live in the gut secrete metabolites that enter the bloodstream and may influence the health of the host organism. Koh *et al*. found that human gut microbes can produce the amino acid metabolite imidazole propionate, which is abundant in blood from human patients with type 2 diabetes and might contribute to their disease. Germ-free mice injected with imidazole propionate developed glucose intolerance and disrupted insulin signaling, like the diabetic patients. Imidazole propionate appeared to act, at least in part, through the p38γ mitogen-activated protein kinase to activate the mechanistic target of rapamycin complex 1 (mTORC1) protein kinase complex. —LBR

Cell **175**, 947 (2018).

MICROFLUIDICS

Flexibility via fiber drawing

Traditional microfluidic devices are fabricated either by building up layers or by etching solid

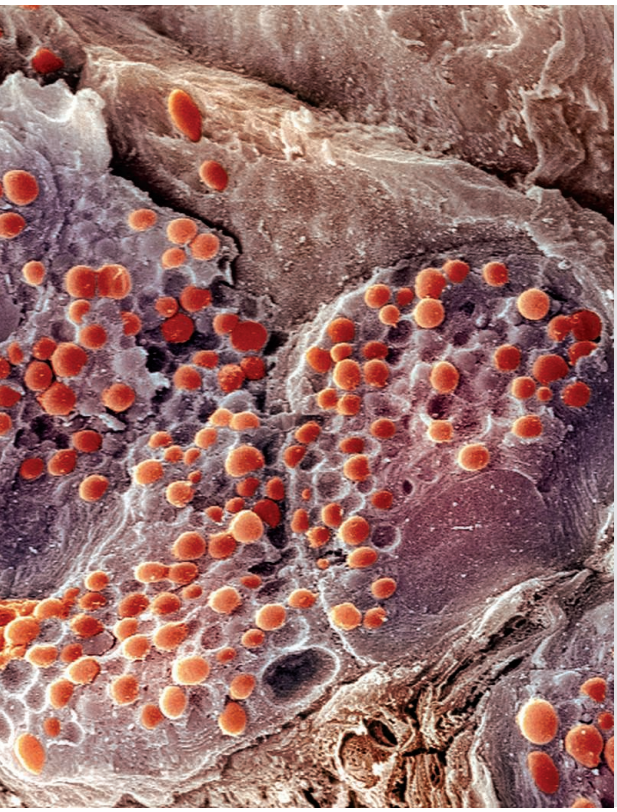


IMAGE: STEVE GSCHMEISSNER/SCIENCE SOURCE

REGENERATION

The making of a salivary gland

Several glands in the body generate and release material via a duct. The salivary gland represents one such exocrine gland, which is composed of several cell types and has a specific role in digestion and swallowing. Regenerative therapies are needed for salivary glands because various diseases and medical procedures can compromise their function. Tanaka *et al*. used organoid technology to regenerate salivary gland tissue in mice. Sox9 and Foxc1 were identified as organ-inductive signals that mediated the differentiation of mouse embryonic stem cells into induced salivary gland primordium (iSG). When iSGs were transplanted into mice lacking salivary glands, the transplanted iSGs looked like normal embryonic salivary gland rudiments and secreted saliva. —BAP

Nat. Comm. **9**, 4216 (2018).

Organoids can be used to regenerate salivary gland tissue.

materials, but both ways primarily result in devices with rectangular, triangular, or circular cross sections and limited flow geometries. Yuan *et al*. show that complex shapes like crosses and stars or even arbitrary shapes can be designed into millimeter-scale objects and then reduced in dimension, but not in cross-sectional profile, through a fiber drawing process. Channels with concave cross sections were used to study inertial flow effects, whereas coextrusion of conductive wires allowed for inertial dielectrophoretic particle manipulation. Sorted particles could then be extracted through the addition of a fiber-to-world connector that can split flow streams without disturbing the laminar flow. —MSL

Proc. Natl. Acad. Sci. U.S.A. **115**, E10830 (2018).

DIVERSITY

What's good for me is good for you

U.S. funding agencies are committed to bringing gender equality to STEM. As these programs begin to mature, is there evidence that they benefit more than just the women involved? Smith *et al*. designed and tested a gender-diversity program that supported the autonomy, relatedness, and competence of female faculty. Over the course of 3 years, researchers collected data from male and female tenure track faculty involved with the program and found positive changes in job satisfaction over time for everyone. Specifically, the more a faculty member felt involved with the project, the more positive changes were reported, suggesting that although diversity programs may target only one group, the results can have positive impacts on everyone who feels involved. —MMc

J. Divers. High. Educ. 10.1037/dhe0000066 (2017).

NANOMATERIALS

Subnanometer WS_2 pores

Nanopores in monolayers of transition metal dichalcogenides such as molybdenum disulfide (MoS_2) and tungsten disulfide (WS_2) have potential applications in molecular separations, but how small can the pores be made? For example, defects created by removal of single Mo atoms from MoS_2 are refilled by mobile Mo atoms. Ryu *et al*. used electron-beam irradiation in a transmission electron microscope to remove single W atoms from a monolayer WS_2 sheet. At 500°C, which is sufficient to allow some sulfur vacancy migration, bond rearrangements created either smaller triangular or larger circular subnanometer pores that were stabilized in part through W–W bonding and by bond rotations about the pore periphery. The triangular holes were less stable and, under irradiation, evolved into the circular structures. —PDS

ACS Nano 10.1021/acsnano.8b07051 (2018).

NEURODEVELOPMENT

Gender in the brain

The hippocampus of the brain is important for higher-order cognitive function, such as learning and memory. Derangements in this region are associated with autism spectrum, attention-deficit hyperactivity disorder, and epilepsy. While looking at the development of neurons during formation of the hippocampus in mice, Chung and Bailey found differences between males and females. In the early postnatal period, when hippocampal circuitry is being built, these neurons depend on signaling through nicotinic acetylcholine receptors. Whole-cell electrophysiology showed that these nascent hippocampal neurons in female mice showed greater excitability, higher resting membrane potential, and greater input resistance than neurons of male mice. These sex-dependent differences occurred during development of circuits that, at maturity, facilitate cognitive functions. —PJH

Dev. Neurobiol. 10.1002/dneu.22646 (2018).

MICROBIOTA

Model for maintaining integrity

The relationships between the gut microbiota, host cells, and food are complicated. Dissecting these interactions is impossible in mice, let alone in humans. Shin and Kim developed a human "gut inflammation–on-a-chip" microfluidic model, complete with villi and peristalsis to explore gut biology. They investigated events following application of the irritant dextran sodium sulfate (DSS), which is often used in mice to stimulate gut inflammation. By disrupting the chip epithelium and removing mucus, DSS facilitated contact between the dissociated epithelial cells and peripheral immune cells and triggered oxidative stress. In the presence of *Escherichia coli*, the epithelial cells produced inflammatory cytokines. After removal of DSS, epithelial regrowth and mucus production restored gut barrier function within days, but probiotics were only helpful before DSS was applied. —CA

Proc. Natl. Acad. Sci. U.S.A. **115**, E10539 (2018).

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

AGRICULTURE

The future of farming

In the mid-20th century, food production from agriculture sharply increased worldwide; however, this was achieved through heavy use of agrochemicals. Extensive collateral damage from excessive use of pesticides, herbicides, and fertilizers has occurred to the wider environment. This has led to biodiversity loss, pesticide resistance and the emergence of new pests, pollution and decline of freshwater supplies, and soil degradation and erosion, as well as direct harm to health. In a Review, Pretty examines the alternative approaches that can achieve sustainable intensification of farming systems by integrating pest management with agroecological systems to minimize costs, maximize yields, restore ecosystem services, and ensure environmental enhancement. —CA

Science, this issue p. 908

SKIN REPAIR

Myofibroblast diversity with injury and aging

Fibroblasts deposit extracellular matrix (ECM) molecules to regulate tissue strength and function. However, if too much ECM is deposited, fibrosis and scarring results. Shook *et al*. examined cells during mouse skin wound healing, fibrosis, and aging (see the Perspective by Willenborg and Eming). They identified distinct subpopulations of myofibroblasts, including cells identified as adipocyte precursors (APs). In cellular ablation mouse models, CD301b-expressing macrophages selectively activated proliferation of APs, but not other myofibroblasts. Myofibroblast composition and gene expression changed during aging. Thus, macrophage-fibroblast interactions are important during tissue repair and aging, which may have therapeutic implications for chronic wounds and fibrotic disease. —BAP

Science, this issue p. 909; see also p. 891

DEVELOPMENTAL BIOLOGY

Maternal factor sets axis

The vertebrate body form changes from the round shape of the fertilized egg to a cylindrical shape when the body plan is established. However, it is unknown whether a maternal factor controls this body axis formation. Yan *et al*. identified such a maternal factor and named it Huluwa. Loss of maternal Huluwa, a transmembrane protein, in zebrafish or frog eggs resulted in embryos that lacked the body axis and were missing the head and dorsoanterior tissues. Huluwa promoted Axin degradation, likely independent of Wnt ligand–receptor signaling, to protect β-catenin from degradation and induce body axis development during embryogenesis. —BAP

Science, this issue p. 910

TOPOLOGICAL MATTER

A messy topological wire

Adding irregularity to a system can lead to a transition from a more orderly to a less orderly phase. Meier *et al*. demonstrated a counterintuitive transition in the opposite direction: Controlled fluctuations in the system's parameters caused it to become topologically nontrivial. The starting point was a one-dimensional lattice of ultracold rubidium atoms in momentum space whose band structure was topologically trivial. The researchers then introduced fluctuations in the tunneling between the lattice sites and monitored the atomic "wires" as the amplitude of the fluctuations increased. The wires first became topologically nontrivial and then went back to trivial for sufficient disorder strengths. —JS

Science, this issue p. 929

NEUROSCIENCE

Egocentric representation of objects

The lateral entorhinal cortex (LEC) and medial entorhinal cortex (MEC) are the two major cortical projections to the hippocampus. The discovery of a variety of functional cell types in MEC has greatly advanced our understanding of the functional anatomy of entorhinal-hippocampal circuits. However, the function of LEC and the behavioral correlates of LEC cells are still not fully understood. Wang *et al*. analyzed the firing properties of LEC and MEC neurons. They found that LEC and MEC used different reference frames, with LEC encoding objects egocentrically. —PRS

Science, this issue p. 945

DEVELOPMENTAL BIOLOGY

A single myosin sets chirality at all scales

When viewed externally, most organisms appear symmetric between the left and right sides. However, many organs are left-right asymmetric. Whether macroscopic asymmetries are directly related to molecular-level chirality remains an open question. Working in *Drosophila*, Lebreton *et al*. found that the conserved molecular motor myosin 1D induced stereotyped chirality at all biological scales—from F-actin turning in vitro to the organ level and even organismal behavior. Thus, a single conserved myosin can generate de novo nano-to-macroscopic changes in form and direction through chiral interaction with the actin cytoskeleton. —BAP

Science, this issue p. 949

IMMUNOLOGY

Visualizing TGF-β1 regulation by GARP

Regulatory T cells (T_{regs}) can suppress immune responses through a variety of mechanisms. One such mechanism involves the activation of a surface-bound latent form of the cytokine transforming growth factor–β1 (TGF-β1). Within the cell, newly synthesized pro-TGF-β1 homodimers form disulfide bonds with the transmembrane protein GARP, which acts to chaperone and orient the cytokine for activation at the cell surface. Liénart *et al*. reveal how GARP interacts with TGF-β1, using a crystal structure in which the complex was stabilized using a Fab fragment from a monoclonal antibody (MHG-8) that binds to the complex. In so doing, they also demonstrate how MHG-8 prevents membrane-associated TGF-β1 release. These structural and mechanistic insights may inform treatments of diseases with altered TGF-β1 functionality and dysfunctional T_{reg} activity, including cancer immunotherapy. —STS

Science, this issue p. 952

BEHAVIOR

From behavior change to conservation success

Between 1970 and 2014, wild vertebrate populations have declined by 60%, and natural systems worldwide are under increasing pressure from human actions. Changes in human behavior will be key to addressing these conservation challenges. In a Perspective, Cinner discusses a range of cognitive biases and social perceptions that can be leveraged in conservation programs. For example, losses tend to hurt more than gains feel good. Conservation efforts can take account of this loss aversion bias by stressing the potential losses from not pursuing them. However, integration of this and other behavioral insights into conservation is not straightforward. It is crucial that behavior change interventions are not seen as coercive and that they

are carefully tailored to the situation to which they are applied. —JFU and SNV

Science, this issue p. 889

STEM CELLS

Rethinking stemness

Stem cells are defined by their ability to produce multiple cell lineages and to self-renew. New technology has allowed the investigation of lineage relationships in hematopoiesis, the process that maintains the cellular constituents of blood. In a Perspective, Yamamoto *et al*. discuss the recent changes in our definitions of what a stem cell is, as well as in our understanding of the lineage hierarchy in hematopoiesis. They also discuss how this might have consequences for patients receiving stem cell transplants. Such changes in concepts may apply to other systems, such as neurogenesis in the brain, for which stem cell lineages are also being questioned. —GKA

Science, this issue p. 895

INNATE LYMPHOID CELLS

Fishing for ILCs in zebrafish

Most studies on innate lymphoid cells (ILCs) have focused on their functions in mammals. Both lymphoid cells and adaptive immunity are not unique to mammals but are shared by all vertebrates. Hernández *et al*. studied lymphoid cells in *rag1*-deficient zebrafish that lack both B and T cells to elucidate the functions of ILC-like cells. They used single-cell RNA sequencing to profile gene expression in gut-resident lymphocytes of *rag1*-deficient zebrafish after immune challenge. They identified lymphocytes in zebrafish that correspond to mammalian ILC1, ILC2, and ILC3 lineages. —AB

Sci. Immunol. **3**, eaau5265 (2018).

MASS SPECTROMETRY

Bridging the mass gap

Viruses and many large biomolecule complexes are in a mass range that is challenging to measure with conventional mass spectrometry methods. Nanomechanical resonators can determine masses of impacting molecules, but separation methods often lose too much of the sample to be efficient. Dominguez-Medina *et al*. used an aerodynamic lens that improved separation and focusing of nebulized molecules with increasing mass. The mass of both filled and empty viral capsids was determined with an array of 20 nanoresonators. —PDS

Science, this issue p. 918

REVIEW SUMMARY

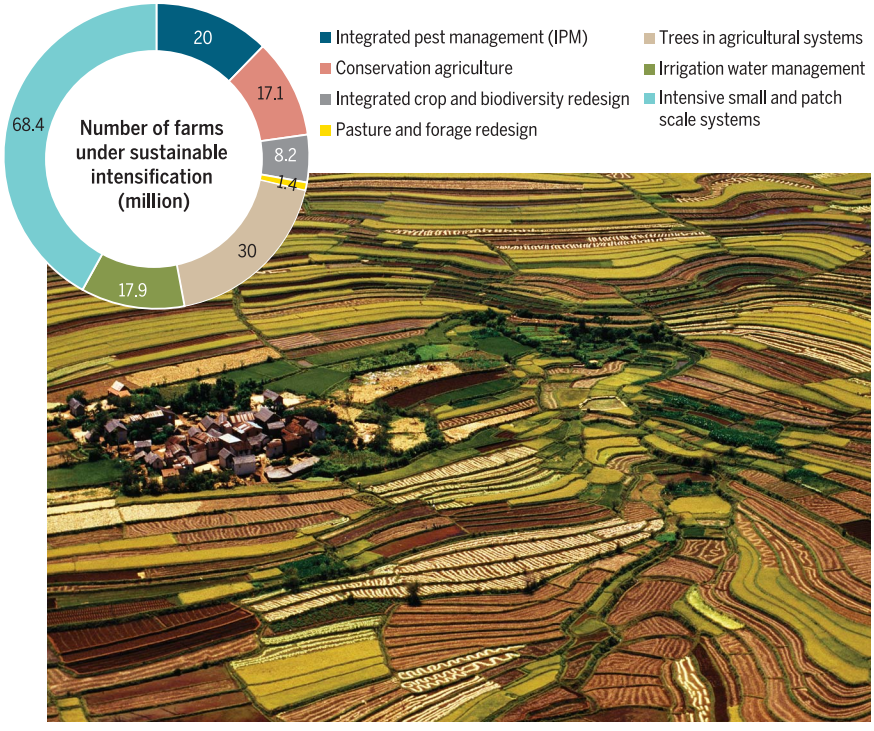
AGRICULTURE

Intensification for redesigned and sustainable agricultural systems

Jules Pretty*

BACKGROUND: The mid-20th century brought agricultural transformation and the "Green Revolution." New crop varieties and livestock breeds—combined with increased use of inorganic fertilizers, manufactured pesticides, and machinery—led to sharp increases in food production from agriculture worldwide. Yet this period of agricultural intensification was accompanied by considerable harm to the environment. This imposed costs on economies and made agricultural systems less efficient by degrading ecosystem goods and services. The desire for agriculture to produce more food without environmental harm, and even to make positive contributions to natural and social capital, has been reflected in many calls for more sustainable agriculture. Sustainable intensification (SI) comprises agricultural processes or systems in which production is maintained or increased while progressing toward substantial enhancement of environmental outcomes. It incorporates these principles without the cultivation of more land and loss of unfarmed habitats and with increases in system performance that incur no net environmental cost.

SI seeks to develop synergies between agricultural and landscape-wide system components and is now a priority for the Sustainable Development Goals of the United Nations. The concept is open; emphasizes outcomes rather than means; can be applied to any size of enterprise; and does not predetermine technologies, production type, or design components. SI can thus be distinguished from earlier manifestations of intensification because of the explicit emphasis on a wider set of environmental as well as socially progressive outcomes. Central to SI is an acceptance that there will be no perfect end point. No designed system is expected to succeed forever, and no single package of practices is able to fit the dynamics of every ecosystem.



SI at the landscape scale. In most landscapes worldwide, SI requires engagements by large numbers of farmers to deliver both productivity improvements and benefits to ecosystem services. Redesign will be a continuing effort of transformation and improvement.

ADVANCES: Three nonlinear stages in transition toward sustainability have been proposed to occur: efficiency, substitution, and redesign. Although both efficiency and substitution are important, they are not sufficient for maximizing coproduction of favorable agricultural and beneficial environmental outcomes without redesign. Whereas efficiency and substitution tend to be additive and incremental within current production systems, redesign should be the most transformative. Redesign presents social and institutional as well as agricultural challenges.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aav0294

It is now clear that SI is spreading to increasing numbers of farmers and is being practiced on a growing area of farmland. By 2018, it was estimated from these initiatives that across some 100 countries, 163 million farms had crossed an important substitution-redesign threshold by using SI methods in at least one farm enterprise, and over an area approaching 453 million ha of agricultural land. This is equivalent to 29% of all farms worldwide and 9% of agricultural land.

OUTLOOK: Pest management exemplifies the need for continuing active intervention for SI; the job is never done. Ecological and economic conditions will change, and agroecosystems will have to be adaptable in order to deliver a range of ecosystem services, including food production but also water and soil conservation, soil carbon storage, nutrient recycling, and pest control. Cooperation—or at least individual actions that collectively result in additive or synergistic benefits—is needed for SI to have a transformative impact across landscapes. Farmers will have to be given the confidence to innovate in a flexible way as conditions change. Every example of successful redesign for SI at scale has involved the prior building of social capital. Widespread adoption of IPM needs new knowledge economies for agriculture. Technologies and practices are growing, but new knowledge needs to be collectively created and deployed and needs to give equal emphasis to ecological and technological innovations. The concept and practice embodied in the SI model of agriculture will be a process of adaptation, driven by a wide range of actors cooperating in new agricultural knowledge economies. ■

School of Biological Sciences, University of Essex, Essex, UK.
*Corresponding author. Email: jpretty@essex.ac.uk
Cite this article as J. Pretty, *Science* **362**, eaav0294 (2018).
DOI: 10.1126/science.aav0294

TOMORROW'S EARTH
Read more articles online at scim.ag/TomorrowsEarth

REVIEW

AGRICULTURE

Intensification for redesigned and sustainable agricultural systems

Jules Pretty*

Redesign of agricultural systems is essential to deliver optimum outcomes as ecological and economic conditions change. The combination of agricultural processes in which production is maintained or increased, while environmental outcomes are enhanced, is currently known as sustainable intensification (SI). SI aims to avoid the cultivation of more land, and thus avoid the loss of unfarmed habitats, but also aims to increase overall system performance without net environmental cost. For example, large changes are now beginning to occur to maximize biodiversity by means of integrated pest management, pasture and forage management, the incorporation of trees into agriculture, and irrigation management, and with small and patch systems. SI is central to the Sustainable Development Goals of the United Nations and to wider efforts to improve global food and nutritional security.

The mid-20th century brought agricultural transformation and the "Green Revolution." New crop varieties and livestock breeds—combined with increased use of inorganic fertilizers, manufactured pesticides (*1*), and machinery as well as better water control and increased field size—led to sharp increases in food production from agriculture worldwide. As a result, aggregate world food production more than tripled during the past 50 years (*2*). The intensity of production on agricultural lands has also risen (*3*). The area under irrigation has doubled, and consumption of nitrogen (N) fertilizers increased sevenfold. At the same time, food production per person has grown, despite considerable population growth (Fig. 1). For each person today, there is 50% more food compared with each person in 1961 (*2*).

Yet this period of agricultural intensification was accompanied by considerable harm to the environment (*4*–*6*). This imposed costs on economies and made agricultural systems less efficient by degrading ecosystem goods and services, including through pollution of groundwater and losses of beneficial insects. Concern about these negative effects shifted ideas about how agricultural systems could be more effective at both food production and reductions in harm to the environment. The desire for agriculture to produce more food without environmental harm, and even to make positive contributions to natural and social capital, has been reflected in many calls for more sustainable agriculture. These have variously been evoked as a doubly green revolution (*7*), alternative agriculture (*8*, *9*), evergreen agriculture (*10*), agroecological intensification (*11*), save and grow (*12*, *13*), diversified agroecosystems (*14*), and sustainable intensification (SI) (*15*–*17*).

SI comprises agricultural processes or systems in which production is maintained or increased while progressing toward substantial enhancement of environmental outcomes. It incorporates these principles without the cultivation of more land and loss of unfarmed habitats and with increases in system performance that incur no net environmental cost (*18*–*20*). However, some controversy surrounds the SI term (*21*). Does the term imply no more than business as usual? Is it a vehicle to smuggle into agriculture potentially harmful technologies? Will it lead to losses of productivity as environmental goods are prioritized? At the same time, concepts of land-sparing and land-sharing have brought into sharp focus the need to improve the intensification of agricultural resources without expanding into nonagricultural and usually highly biodiverse habitats (*22*). SI seeks to make better use of natural and human resources (such as land, water, biodiversity, and knowledge) and technologies.

In many farmed landscapes, the need for effective SI is urgent. Environmental degradation is reducing the asset base of existing agricultural lands (*6*, *23*), expansion of urban and road infrastructure has removed agricultural land [in the current countries of the European Union, agricultural area fell by 31 Mha over 50 years; in the United States and Canada, 0.5 Mha are lost annually (*24*, *25*)], and climate change and extreme weather events create new stresses that test the resilience of agricultural systems. SI seeks to develop synergies between agricultural and landscape-wide system components and is now a priority for the Sustainable Development Goals of the United Nations (*26*). The concept is open; emphasizes outcomes rather than means; can be applied to any size of enterprise; and does not predetermine technologies, production type, or design components. It can thus be distinguished from earlier manifestations of intensification because of the explicit emphasis on a wider set of environmental as well as socially progressive outcomes. Central to SI is an acceptance that there will be no perfect end point. No designed system is expected to succeed forever, and no single package of practices is able to fit the dynamics of every ecosystem (*27*).

Redesign framework for SI

Three nonlinear stages in transitions toward sustainability have been proposed to occur: efficiency, substitution, and redesign. Although both efficiency and substitution are important, they are not sufficient for maximizing coproduction of favorable agricultural and beneficial environmental outcomes without redesign (*28*, *29*).

Efficiency aims to make better use of on-farm and imported resources within existing farm configurations. Many agricultural systems are wasteful, permitting natural capital degradation within the farm or the escape of agrochemical inputs across system boundaries, which causes external costs on-farm and beyond. Post-harvest losses reduce food availability, and tackling them contributes directly to efficiency gains and amplifies the benefits of yield increases generated by other means. On-farm efficiency gains can arise from targeting and rationalizing inputs of fertilizer, pesticide, and water to focus impact, reduce use, and cause less damage to natural capital and human health. Precision farming requires sensors, detailed soil mapping, drone mapping, scouting for pests, weather and satellite data, information technology, robotics, improved diagnostics, and delivery systems to ensure that targeted inputs (such as pesticide, fertilizer, and water) are applied at an appropriate rate and time to the right place only when needed (*11*, *25*, *30*). Automatic control and satellite navigation of agricultural vehicles and machinery can enhance energy efficiency and limit soil compaction.

Substitution focuses on the replacement of technologies and practices. The development of new crop varieties and livestock breeds deploys substitution to replace less efficient system components with alternatives, such as plant varieties that are better at converting nutrients to biomass, that tolerate drought and/or increases in salinity, and with resistance to specific pests and diseases. Other forms of substitution include the release of biological control agents to substitute for agrochemical inputs, the use of RNA-based gene-silencing pesticides, replacement of the use of soil in hydroponics, and no-tillage systems that use new forms of direct seeding and weed management to replace inversion ploughing.

The third stage is fundamental for SI to achieve sustainability at scale. The redesign of agroecosystems is essential to harness ecological processes such as predation, parasitism, allelopathy, herbivory, N fixation, pollination, trophic dependencies, and others (*31*, *32*). A prime aim is to modulate greenhouse gas emissions; provide clean water; maximize carbon sequestration;

School of Biological Sciences, University of Essex, Essex, UK.
*Corresponding author. Email: jpretty@essex.ac.uk

promote biodiversity; and disperse and ameliorate the effects of pests, pathogens, and weeds. Whereas efficiency and substitution tend to be additive and incremental within current production systems, redesign should be the most transformative.

Redesign presents social and institutional as well as agricultural challenges (*31*–*34*). Unintended consequences must also be identified and mitigated as part of the redesign process.

SI impacts on productivity

The two key questions to ask of an SI system is first whether it actually generates more food, fiber, and other valued products while simultaneously improving natural capital, and second, can this be done without harming key renewable capital assets? Farmers who adopt various SI approaches can increase food outputs by multiplicative or by additive means (*35*). Multiplicative approaches improve yields per hectare by combining use of new and improved varieties with changes to agronomic-agroecological management. Additive methods require diversification of farms into a range of new crops, livestock, or fish that add to the existing staples or vegetables already being cultivated. Additive components range from use of fish ponds or concrete tanks; raised beds and vegetable cultivation; rehabilitation of degraded land; fodder grasses and shrubs for livestock (which can increase milk productivity); new crops or trees brought into rotations with staple crops such as clovers, soybean, and indigenous trees; to the adoption of short-maturing varieties (such as sweet potato and cassava) that permit the cultivation of two crops per year instead of one.

An early large-scale assessment of SI was commissioned by the U.S. National Research Council (NRC) (*8*). Partly driven by increased costs of fertilizer and pesticide inputs, plus growing scarcity of natural resources (such as groundwater for irrigation), and continued soil erosion, farmers had been adopting new approaches in a wide variety of farm systems. The NRC noted that "alternative agriculture" was not a single system of farming practices but rather used a mix of crop rotations, integrated pest management (IPM), soil- and water-conserving tillage, animal production systems that emphasized disease prevention without antibiotics, and genetic improvement of crops to resist pests and disease and to use nutrients more efficiently. Well-monitored alternative farming systems nearly always used less synthetic pesticide, fertilizer, and antibiotics per unit of production than comparable conventional farms. They also required more information and management skills of farmers per unit of production. The NRC (*9*) conducted follow-up studies on 10 of the original farms. These included integrated crop-livestock enterprises, fruit and vegetable farms, a beef cattle ranch, and one rice farm. After 22 years, there were four common features of these farms: (i) accumulation and maintenance of a natural resource base and maximization of internal resources; (ii) environmental sustainability and closed nutrient cycles; (iii) careful soil management, the use of crop rotations and cover crops, and for livestock, management practices that did not use hormones or antibiotics; and (iv) taking advantage where possible of direct sales markets (via farmers markets and/or internet sales), with some sold at a premium with labeled traits and products (such as organic, naturally raised livestock).

Substantial progress toward SI has also been made in developing countries over the past two decades. One study analyzed 286 projects in 57 countries, and a later one assessed 40 projects in 20 African countries (*36*, *37*). In both, several million farmers on tens of megahectares had adopted practices that had led to yield increases of 79% (study 1) and 113% (study 2). The time scale for these improvements varied from 3 to 10 years. A further analysis of 85 IPM projects from 24 countries in Asia and Africa implemented over a 25-year period (1990 to 2014) further illustrated the potential for productivity improvement and substantial reductions in pesticide costs (*38*). Overall, mean yields increased by 41%, and pesticide use declined to 31% of prior use (Fig. 2). Compared with the benchmark preproject point, 30% of the crop combinations resulted in a transition to zero pesticide use.

Although pesticide reductions with IPM should be expected, explanations for yield increases induced by IPM are more complex. IPM may, for example, reduce the incidence of severe-loss years, although yield increases in a normal year may not be evident, but mean production does increase across years. Many IPM projects involve interventions focused on more than just pest management. For example, they may involve a substantial component of farmer training [for example, through farmer field schools (FFS)], in which case, farmers' capabilities at innovating in several areas of their agroecosystems may also have increased, such as in soil and water management (*39*). Farmer training through FFS has resulted in greater and continuing innovation, with positive outcomes for both productivity and environmental services (*34*, *39*).

Global extent of SI redesign

It is now clear that SI is spreading to increasing numbers of farmers and is being practiced on a growing area of farmland. A recent global assessment screened 400 SI projects, programs, and initiatives worldwide (*20*). The intention was to assess where agricultural innovation had scaled to have potentially positive landscape-scale outcomes on ecosystem services (Table 1).

There are some 570 million farms worldwide, 84% of which are landholdings of less than 2 ha (*40*). These small farms make up only 12% of total agricultural area. Of all farms, 74% are in Asia (of which 35% are in China and 24% are in India), 9% in sub-Saharan Africa, 7% in Central Europe and Central Asia, 3% in Latin America and the Caribbean, and 3% in Middle East and North Africa. Only 4% of farms are in industrialized countries. To be effective, SI will have to encompass larger numbers of farms in less developed countries and larger farms of smaller numbers in industrialized countries.

In the analysis summarized in Table 1, 47 of the SI initiatives exceeded the 10^4 scale for either hectares or farm numbers, of which 17 exceeded the 10^5 threshold and 14 exceeded 10^6 (*21*). Many SI initiatives worldwide show promise but remain limited in scale. By 2018, it was estimated from these initiatives that in some 100 countries, 163 million farms had crossed an important

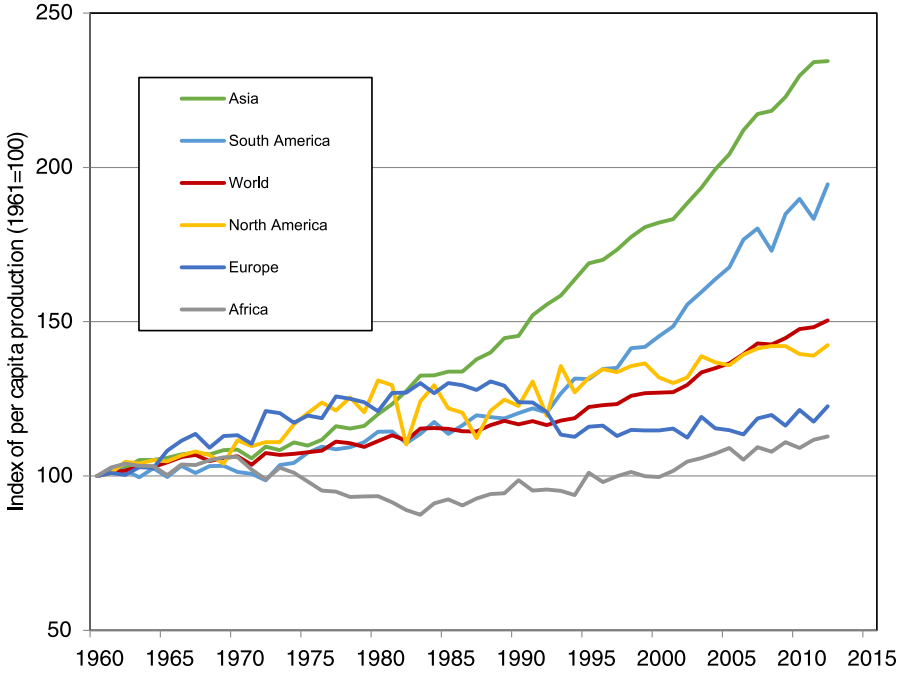


Fig. 1. Global per capita agricultural production. 1961 = 100. [Source data, (*2*)]

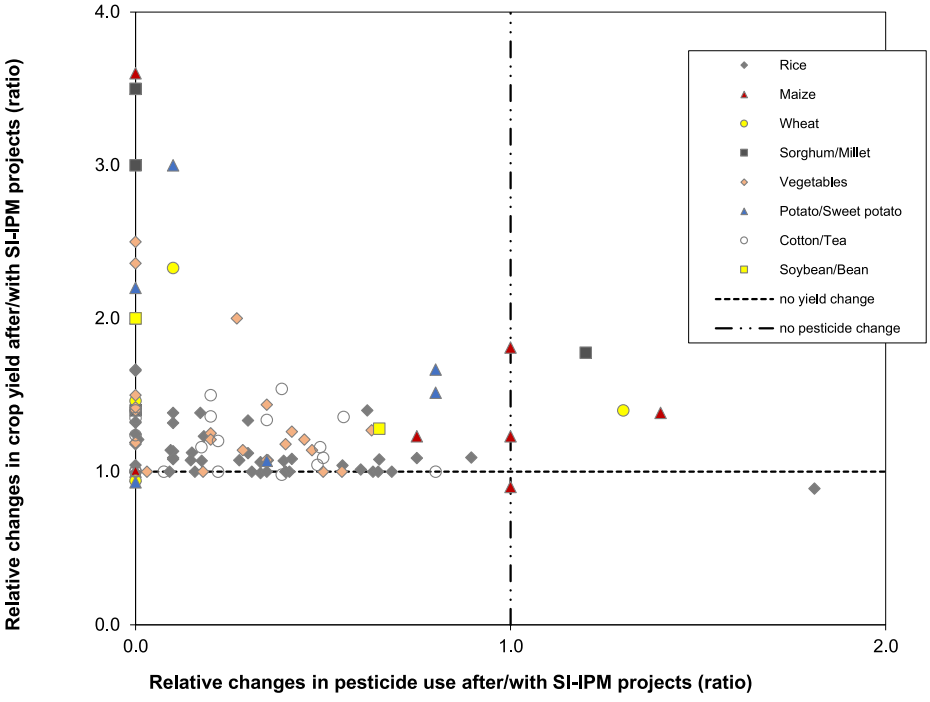


Fig. 2. Impacts of SI-IPM projects and programs in Asia and Africa on pesticide use and crop yields. Shown are 85 projects in 24 countries. [Source data (*38*)]

substitution-redesign threshold using SI methods in at least one farm enterprise, and on an area approaching 453 million ha of agricultural land (not counting the SI initiatives in home and urban gardens and on field boundaries). This is equivalent to 29% of all farms worldwide and 9% of agricultural land (total worldwide crop and pasture land is 4.9×10^9 hectares).

Such a global assessment might imply that numbers of farms and hectares are fixed. Flux may arise from farmer choice and agency but equally from the actions of vested interests, agricultural input companies, consolidation of small farms into larger operations, changes in agricultural policy or shifts in market demand, and discrepancies between on-paper claims and what farmers have implemented. Efficiency-substitution adoption was not included in this assessment. For example, European Union regulations require all farms to use IPM, but this has not yet led to major redesign of agricultural practices that substantially benefit ecosystem services (*25*, *30*).

Cost of pest management by pesticides

Pathogens, weeds, and invertebrates cause substantial crop losses worldwide. Although the reporting of pesticide use and market data are patchy, the use of synthetic pesticides in agriculture has grown steadily to 3.5 billion kg of active ingredient (AI) per year (*38*). The value of the global market is now US$45 billion per year, with herbicides accounting for 42%, insecticides 27%, fungicides 22%, and disinfectants and other agrochemicals 9% of sales. China, the United States, and Argentina account for 70% of world pesticide use in agriculture (2.44 billion kg of AI annually) (*38*), and six countries each consume between 50 million and 100 million kg (Brazil, Canada, France, Italy, Japan, Thailand). In the past 20 years, pesticide consumption has grown fourfold in China, eightfold in Argentina, threefold in Brazil , fivefold in Bangladesh, and fourfold in Thailand.

Pesticides are intended to be hazardous to life, and there will be risks associated with their use; their full costs illustrate the often hidden harm of nonsustainable deployment. The value of pesticides lies in their ability to kill unwanted organisms, but their toxicity can also cause unintended harm on and beyond the farm (external costs). The collateral effects of pesticide use show features commonly found across the agricultural sector. The costs of unintended harm are often neglected, in part because they may occur after a time lag and may damage groups whose interests are not well-represented. Furthermore, it is not always clear where harmful compounds in the environment may have come from. In studies of pesticide externalities in China, Germany, Thailand, the United Kingdom, and the United States, costs have been calculated to range from $4 to 19 (€3 to 15) per kg of AI (*41*–*45*). These costs put annual pesticide externalities worldwide in the range of $10 billion to $60 billion (for use of 3.5 billion kg and for a market size of $45 billion).

Additional private costs are borne by farmers themselves and tend not to be included in calculations of damage, such as the costs of personal ill-health resulting from exposure to pesticides (*46*) or from increased pest, weed, or fungal resistance. Worldwide, weed species have evolved resistance to every herbicide class, and more than 550 arthropod species have gained resistance to at least one insecticide (*47*). New research has also shown that residues of some classes of pesticide (such as neonicotinoid insecticides) are more ubiquitous than previously assumed, suggesting that external costs may be underestimated: 97% of neonicotinoids brought back in pollen by bees in arable landscapes originates from nearby wildflowers rather than from crops themselves (*48*). At the same time, it has been found that the total flying insect biomass in central Europe has declined by 75% over a 27-year period (*22*). The ecosystem services provided by wild insects have been estimated at $57 billion annually in the United States (*49*). Such private and external costs reveal that some forms of agriculture are less effective and efficient than might appear from productivity data alone, indicating the need for new metrics and system design (*50*).

Redesign for SI-integrated pest management and ecosystem services

Redesign is critical as ecological, economic, social, and political conditions change across whole landscapes. The rapidly changing nature of pest, disease, and weed threats illustrates the continuing challenge to respond with agility. New pests and diseases can suddenly emerge because of resistance to pesticides, which can then lead to secondary pest outbreaks owing to pesticide overuse. Climate change has facilitated invasions of pests and pathogens, the accidental long-distance transfer of organisms, as well as long-distance trade (for example, of bees, pets, and plants). For example, wheat blast fungus (*Magnaporthe oryzae*) has recently emerged as a crop pathogen in Bangladesh (2016), and the Fall Army Worm (*Spodoptera frugiperda*) is spreading across sub-Saharan Africa (2017). The papaya mealybug (*Paracoccus marginatus*) is native to Mexico but spread to the Caribbean in 1994; then to the Pacific islands by 2002; and then to Indonesia, India, and Sri Lanka by 2008; and is currently found in West Africa. Although the mealybug's preferred host is papaya, it has now adapted to mulberry, cassava, tomato, and eggplant (*51*). Each geographic spread, each shift of host, requires redesign of local agricultural systems and rapid responses from research and extension services. Such new pests and diseases may also affect crop pollinators, as illustrated by host shifts and the anthropogenic spread of bee parasites (such as *Varroa* mites) and pathogens (such as *Nosema ceranae*) (*5*).

A further example is the cassava pink mealybug, which was first reported in the greater Mekong region of Thailand in 2008 and caused an immediate 27% drop in cassava production (*52*). An IPM program was developed with multiple tactics, involving ploughing and drying soil, soaking stalk cuttings in insecticide, burning of infested plants, banning transport of infested plant materials, and the release of *Anagyrus lopezi* parasitoids. In 2010–2011, 6 million pairs of *Anagyrus* were released in Thailand, which brought the pest completely under control and enabled a lasting recovery of fresh root yields. This further underlies how important ecologically based tactics are to the SI of agriculture.

Old pests can return. The brown planthopper (BPH) has been called the "ghost of green revolutions past" (*53*). It was the primary threat to rice in the 1960s yet has resurfaced as a major pest threat in the 2000s owing to resistance to insecticides coupled with the heavy use of N fertilizers. BPH outbreaks are often triggered by overuse of insecticides, which reinforces farmers' fears of insect pests, provoking in them the wish to apply more. In China, between 6 and 9 Mha were infested with BPH in 2005–2007, up from 2 Mha in the 1990s (*51*). Farmers in China apply on average 180 kg N/ha to rice as fertilizer, and N-enriched plants are known to enhance size, performance, and abundance of herbivorous pests.

IPM consists of a toolbox of interventions that combine the use of targeted compounds with agronomic and biological techniques to control different classes of crop pests. Complementary and alternative modes of pest control that exploit specificities in pest ecologies have been gaining increasing attention. The use of on- and off-farm biodiversity is key in IPM because biodiverse agroecosystems experience less pest damage and have more natural pest enemies than those of nonbiodiverse ones (*27*, *54*). At the same time, both social and human capital are important for successful outcomes (*33*). IPM is knowledge-intensive. For successful IPM, farmers need to monitor pests and natural enemies, understand thresholds for decisions, and be competent in the deployment of a range of different methods.

IPM approaches span the efficiency-substitution-redesign (ESR) framework (Table 2). These range from targeted use of pesticide compounds to habitat and agroecological design. In only rare cases—such as the aerial release of the parasitic wasp *Epidinocarsis lopezi* to control cassava mealybug in West and Central Africa (*55*)—can IPM be implemented without farmer engagement. Recent years have seen a substantial increase in understanding how to increase farmers' knowledge so that they are able to husband crops and livestock while reducing or eliminating pesticides.

Social capital matters greatly. IPM strategies have now transitioned from individual field-based practice to coordinated, community-scale decision-making covering wider landscapes. Although this improves the effectiveness of pest control, it presents a considerable obstacle to wider adoption by presenting a collective-action dilemma: How can farmers as individual businesses be persuaded to work together for personal as well as wider landscape benefits (*33*)? FFS, which were started in the 1980s (*56*, *57*), are among the most important mechanisms for the development and spread of IPM. FFS are not an extension method; they increase knowledge of agroecology, problem-solving skills, group building, and political strength. They can be particularly effective where there are simple messages (for example, do not use insecticides in the first 40 days of rice cultivation because herbivore-damaged plants recover with no yield loss) (*58*). FFS have been used in 90 countries (*34*, *59*, *60*), with 19 million farmer graduates, and now some 20,000 FFS graduates are now running FFS for other farmers.

One of the most effective IPM-redesign systems is the "push-pull" system (in which pests and beneficial insects are pushed and pulled into and away from valued crops), which is yielding notable successes in monocropped cereal systems (*16*). This method has been deployed with great effect against *Striga* weed and stemborer infestations in maize, millet, and sorghum (*61*, *62*) and involves the use of interplanted "decoy" crops. Across Kenya, Uganda, Tanzania, and Ethiopia, push-pull systems are used on about 130,000 small farms. Interplanting of the leguminous forage crop *Desmodium* suppresses *Striga* and repels stemborer moths while attracting their natural enemies; planting *Napier* grass as a border crop attracts stemborer moths out of the crop. The interplanted fodder crop not only fixes N but has also provided an additional resource that has enabled farmers to diversify into dairy and poultry production, which in turn provides animal manure for application on fields. As a result, yields of maize and sorghum have increased substantially, with an up to threefold increase over control plots. Better quality of fodder for dairy animals has increased milk yields by at least 2 liters daily and ultimately gives considerably higher economic returns to the farmer than does monocropping.

This kind of redesign has been deployed in many agroecosystems, resulting in increased rotational diversity, use of wildflowers for pollinators and other beneficial insects, and deployment of conservation headlands and trap crops (*63*, *64*), often with large reductions in input use without yield compromise (*65*). In tropical systems, fish, crab, and duck reintroduced into rice systems reduce pest and weed incidence, often eliminate the need for pesticides, and increase gross productivity through the provision of animal protein outputs (*66*).

Toward collective action and landscape-scale change

Pest management exemplifies the need for continuing active intervention for SI; the job is never done. Ecological and economic conditions will

Table 1. Redesign for SI. Subtypes of SI, farm numbers, and hectares (at 2018). Some subtypes span several types (for example, "organic agriculture" also appears in elements of 4 and 7). [Source (*20*)]

Redesign SI type	Illustrative redesign subtypes of SI intervention	Farm numbers (million)	Hectares under SI (million)
IPM	IPM through farmer field schools; integrated plant and pest management; push-pull systems	20.03	17.41
Conservation agriculture (CA)	Conservation agriculture practices; zero- and low-tillage; soil conservation and soil erosion prevention; enhancement of soil health	17.10	181.03
Integrated crop and biodiversity redesign	Organic agriculture; rice-fish systems; systems of crop and rice intensification; zero-budget natural farming; science and technology backyard platforms; farmer wisdom networks; landcare and watershed management groups	8.18	63.31
Pasture and forage redesign	Mixed forage-crop systems; management intensive rotational grazing systems; agropastoral field schools	1.43	81.85
Trees in agricultural systems	Agroforestry; joint and collective forest management; leguminous fertilizer trees and shrubs	30.00	61.21
Irrigation water management	Water user associations; participatory irrigation management; watershed management; micro-irrigation technologies	17.90	33.00
Intensive small and patch scale systems	Community farms, allotments, backyard gardens, and raised beds; vertical farms; group purchasing associations and artisanal small producers (in Community Supported Agriculture, tekei groups, and guilds); micro-credit groups for small-scale intensification; integrated aquaculture	68.41	15.52

Table 2. ESR options for integrated pest management and SI. [Source: adapted from (38)]

IPM SI type	Examples of application
	Efficiency
Management of application of pesticides	Targeted spraying
	Threshold spraying prompted by decision-making derived from observation and data on pest, disease or weed incidence
	Substitution
Substitution of pesticidal products with other compounds	Synthetic pesticide with high toxicity substituted by another product with low toxicity
	Use of agrobiologicals or biopesticides (e.g., derived from neem)
Releases of antagonists, predators or parasites to disrupt or reduce pest populations	Sterile breeding of male pest insects to disrupt mating success at population level
	Identification and deliberate release of parasitoids or predators to control pest populations
Deployment of pheromone compounds to move or trap pests	Sticky and pheromone traps for pest capture
Crop and livestock breeding	Deliberate introduction of resistance or other traits into new varieties or breeds (for example, recent use of genetic modification for insect resistance and/or herbicide tolerance)
	Redesign
Agroecological system and habitat redesign	Seed and seed bed preparation
	Deliberate use of domesticated or wild crops/plants to push-pull pests, predators, and parasites
	Use of crop rotations and multiple-cropping to limit pest, disease, and weed carryover across seasons or viability within seasons
	Adding host-free periods into rotations
	Adding stakes to fields for bird perches

change, and appropriate responses will be needed. Agroecosystems will have to be adaptable consistently to deliver a range of ecosystem services, including food production, but also water and soil conservation, soil carbon storage, nutrient recycling, and pest control.

Cooperation, or at least individual actions that collectively result in additive or synergistic benefits, is needed for SI to have a transformative impact across landscapes. Farmers will have to be given the confidence to innovate in a flexible way as conditions change. Every example of successful redesign for SI at scale has involved the prior building of social capital (20). Such initiatives require relations of trust, reciprocity and exchange, common rules, norms and sanctions, and connectedness in groups. As social capital reduces the costs of working together, it facilitates cooperation, which gives people the confidence to invest in collective activities, knowing that others will do so too. Individuals are then less likely to get away with free-rider actions that cause resource degradation.

Widespread adoption of IPM needs new knowledge economies for agriculture (67). Technologies and practices are growing, but new knowledge needs to be collectively created and deployed and needs to give equal emphasis to ecological and technological innovations. Extension systems and FFS must give equal consideration to environmental as well as agronomic skills (34). For example, the Landcare movement in Australia consists of 6,000 groups of farmer-led watershed councils. The agroecosystem research network in the United States, the French network of agroecology farms, and the Farmer Cluster Initiatives in the United Kingdom (68, 69) are all important examples from industrialized countries that are delivering practices to address locally specific problems of erosion, nutrient loss, pathogen escape, and waterlogging. In Cuba, the *Campesino-a-Campesino* movement has built agroecological methods with knowledge and technologies spread through exchange and cooperatives. As a result, the productivity of 100,000 farmers has increased by 150% over 10 years, and pesticide use has fallen to 15% of former levels (70). In Bangladesh, innovation platforms have driven adoption of direct seeding and use of early-maturing rice (71). In China, Science and Technology Backyard (STB) platforms operate in 21 provinces, covering many cereal, root, and fruit crops (72). STB platforms bring agricultural scientists to live in villages and use field demonstrations and farm schools for the development of innovations. They are successful because they center on in-person communication, sociocultural bonding, and trust developed among farmer groups of 30 to 40 individuals.

Concluding comments

In general, policymakers and regulators find it easier to seek to prevent practices or problems, such as the regulation of certain pesticide compounds or the establishment of safe drinking water limits for certain compounds. It has been harder to encourage positive practices. In most contexts, state policies for SI remain poorly developed or counterproductive. In the European Union, farm subsidies have increasingly been shifting toward targeted environmental outcomes rather than payments for production (73), but this seldom guarantees synergistic benefits across whole landscapes. Ethical and sustainable sourcing by food manufacturers, processors, and retailers would help drive up demand, particularly if producers connect directly with consumers (74). There are some regional-scale exemplars of positive policy practice. One example is India's state of Andhra Pradesh, where the state government has made explicit its support to zero-budget natural farming (a local form of uncertified organic farming that does not require the expenditure of farmer income on inputs), aiming to reach 6 million farmers by 2027 (75). The greening of the Sahel through agroforestry began when national tree ownership regulations were changed to favor local people (18), and in China, where the 2016 No. 1 Central Document emphasizes innovation, coordination, greening, and sharing as key parts of a new strategy for SI (76).

There are arguments from some quarters that we would not need to increase agricultural production if less food was wasted and less energetically inefficient meat was consumed by the affluent. These would help, but there is no magic wand of redistribution. Most if not all farmers need to raise yields while improving environmental services. As the evidence shows, redesign

of agroecosystems around SI can achieve both yield increases and resilience. The evidence from farms of redesign and transformations toward SI offers scope for optimism. A full transition from increased efficiency through substitution to redesign will be essential. The concept and practice embodied in the SI model of agriculture will be a process of adaptation, driven by a wide range of actors cooperating in new agricultural knowledge economies.

REFERENCES AND NOTES

1. In this paper, the term "pesticide" incorporates all synthesized pest, disease, weed, fungal, and other control compounds. There is also no single accepted terminology for grouping of types of countries. Terms relate to past stages of development (developed, developing, or less developed), state of economy or wealth (such as industrialized or affluent), geographic location (such as global south or north), or membership [Organisation for Economic Cooperation and Development (OECD) or non-OECD]. None are perfect. Here, I have simply used "industrialized" and "less-developed" and acknowledge the shortcomings.
2. Food and Agricultural Organization of the United Nations (FAO), FAOSTAT Online database (FAO, 2018); www.fao.org/faostat/en.
3. D. Tilman, C. Balzer, J. Hill, B. L. Befort, Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20260–20264 (2011). doi: 10.1073/pnas.1116437108; pmid: 22106295.
4. Government Office for Science, Foresight, *The Future of Global Food and Farming* (Government Office for Science, 2011).
5. D. Goulson, E. Nicholls, C. Botías, E. L. Rotheray, Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* **347**, 1255957 (2015). doi: 10.1126/science.1255957; pmid: 25721506
6. J. Rockström *et al*., Sustainable intensification of agriculture for human prosperity and global sustainability. *Ambio* **46**, 4–17 (2017). doi: 10.1007/s13280-016-0793-6; pmid: 27405653
7. G. R. Conway, *The Doubly Green Revolution* (Penguin, 1997).
8. NRC, *Alternative Agriculture* (National Academies Press, 1989)
9. NRC, *Towards Sustainable Agricultural Systems in the 21st Century* (National Academies Press, 2010).
10. D. P. Garrity *et al*., Evergreen Agriculture: A robust approach to sustainable food security in Africa. *Food Secur.* **2**, 197–214 (2010). doi: 10.1007/s12571-010-0070-7
11. K. Garbach *et al*., Examining multi-functionality for crop yield and ecosystem services in five systems of agroecological intensification. *Int. J. Agric. Sustain.* **15**, 11–28 (2017). doi: 10.1080/14735903.2016.1174810
12. FAO, *Save and Grow* (FAO, 2011).
13. FAO, *Save and Grow: Maize, Rice and Wheat: A Guide to Sustainable Crop Production* (FAO, 2016).
14. IPES-Food, From uniformity to diversity: A paradigm shift from industrial agriculture to diversified agroecological systems (2016); www.ipes-food.org.
15. J. Pretty, The sustainable intensification of agriculture. *Nat. Resour. Forum* **21**, 247–256 (1997). doi: 10.1111/j.1477-8947.1997.tb00699.x
16. Royal Society, *Reaping the Benefits: Science and the Sustainable Intensification of Global Agriculture* (Royal Society, 2009).
17. R. M. Gunton, L. G. Firbank, A. Inman, D. M. Winter, How scalable is sustainable intensification? *Nat. Plants* **2**, 16065 (2016). doi: 10.1038/nplants.2016.65; pmid: 27243658
18. H. C. J. Godfray *et al*., Food security: The challenge of feeding 9 billion people. *Science* **327**, 812–818 (2010). doi: 10.1126/science.1185383; pmid: 20110467
19. P. Smith, Delivering food security without increasing pressure on land. *Glob. Food Secur.* **2**, 18–23 (2013). doi: 10.1016/j.gfs.2012.11.008
20. J. Pretty *et al*., Global assessment of agricultural system redesign for sustainable intensification. *Nat. Sustain.* **1**, 441–446 (2018). doi: 10.1038/s41893-018-0114-0
21. T. Garnett, H. J. C. Godfray, 2012. *Sustainable Intensification in Agriculture* (Oxford Martin Programme on the Future of Food, University of Oxford, 2012).
22. B. Phalan, M. Onial, A. Balmford, R. E. Green, Reconciling food production and biodiversity conservation: Land sharing and land sparing compared. *Science* **333**, 1289–1291 (2011). doi: 10.1126/science.1208742; pmid: 21885781
23. C. A. Hallmann *et al*., More than 75 percent decline over 27 years in total flying insect biomass in protected areas. *PLOS ONE* **12**, e0185809 (2017). doi: 10.1371/journal.pone.0185809; pmid: 29045418
24. C. A. Francis *et al*., Farmland conversion to non-agricultural uses in the US and Canada: Current impacts and concerns for the future. *Int. J. Agric. Sustain.* **10**, 8–24 (2012). doi: 10.1080/14735903.2012.649588
25. A. Buckwell *et al*., *The Sustainable Intensification of European Agriculture* (RISE Foundation, 2014).
26. UN Sustainable Development Platform. Sustainable Development Goals. (2017); www.un.org/sustainabledevelopment/sustainable-development-goals.
27. H. Sandhu *et al*., Significance and value of non-traded ecosystem services on farmland. *PeerJ* **3**, e762 (2015). doi: 10.7717/peerj.762; pmid: 25737811
28. S. Hill, *Alternatives* **12**, 32–36 (1985).
29. S. Hill, in *Organic Farming, Prototype for Sustainable Agriculture*, S. Bellone, S. Penvern, Eds. (Springer, 2014).
30. N. H. Lampkin *et al*., 2015. *The Role of Agroecology in Sustainable Intensification* (Organic Research Centre, 2015).
31. S. R. Gliessman, M. Rosemeyer, Eds. *The Conversion to Sustainable Agriculture* (CRC Press, 2009).
32. G. M. Gurr *et al*., Multi-country evidence that crop diversification promotes ecological intensification of agriculture. *Nat. Plants* **2**, 16014 (2016). doi: 10.1038/nplants.2016.14; pmid: 27249349
33. J. Pretty, Social capital and the collective management of resources. *Science* **302**, 1912–1914 (2003). doi: 10.1126/science.1090847; pmid: 14671287
34. FAO, *Farmer Field School Guidance* (FAO, 2016).
35. J. Pretty, Z. P. Bharucha, *Ann. Bot. (Lond.)* **205**, 1–26 (2014).
36. J. N. Pretty *et al*., Resource-conserving agriculture increases yields in developing countries. *Environ. Sci. Technol.* **40**, 1114–1119 (2006). doi: 10.1021/es051670d; pmid: 16572763
37. J. Pretty, C. Toulmin, S. Williams, Sustainable intensification in African agriculture. *Int. J. Agric. Sustain.* **9**, 5–24 (2011). doi: 10.3763/ijas.2010.0583
38. J. Pretty, Z. P. Bharucha, Integrated pest management for sustainable intensification of agriculture in Asia and Africa. *Insects* **6**, 152–182 (2015). doi: 10.3390/insects6010152; pmid: 26463073
39. W. Settle, M. Soumaré, M. Sarr, M. H. Garba, A.-S. Poisot, Reducing pesticide risks to farming communities: Cotton farmer field schools in Mali. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20120277 (2014). doi: 10.1098/rstb.2012.0277; pmid: 24535387
40. S. K. Lowder, J. Skoet, T. Raney, The number, size, and distribution of farms, smallholder farms, and family farms worldwide. *World Dev.* **87**, 16–29 (2016). doi: 10.1016/j.worlddev.2015.10.041
41. E. M. Tegtmeier, M. D. Duffy, External costs of agricultural production in the United States. *Int. J. Agric. Sustain.* **2**, 1–20 (2004). doi: 10.1080/14735903.2004.9684563
42. A. W. Leach, J. D. Mumford, Pesticide environmental accounting: A method for assessing the external costs of individual pesticide applications. *Environ. Pollut.* **151**, 139–147 (2008). doi: 10.1016/j.envpol.2007.02.019; pmid: 17604888
43. A. W. Leach, J. D. Mumford, Pesticide environmental accounting: A decision-making tool estimating external costs of pesticides. *J. Verbr. Lebensm.* **6**, 21–26 (2011). doi: 10.1007/s00003-011-0674-7
44. S. Praneetvatakul, P. Schreinemachers, P. Pananurak, P. Tipraqsa, Pesticides, external costs and policy options for Thai agriculture. *Environ. Sci. Policy* **27**, 103–113 (2013). doi: 10.1016/j.envsci.2012.10.019
45. D. Norse, X. Ju, Environmental costs of China's food security. *Agric. Ecosyst. Environ.* **209**, 5–14 (2015). doi: 10.1016/j.agee.2015.02.014
46. H. Andersson, D. Tago, N. Treich, *Pesticides and Health: A Review of Evidence on Health Effects, Valuation of Risks, and Benefit-Cost Analysis* (Toulouse School of Economics, WP TSE-477, 2014).
47. F. Gould, Z. S. Brown, J. Kuzma, Wicked evolution: Can we address the sociobiological dilemma of pesticide resistance? *Science* **360**, 728–732 (2018). doi: 10.1126/science.aar3780; pmid: 29773742
48. C. Botías *et al*., Neonicotinoid residues in wildflowers, a potential route of chronic exposure for bees. *Environ. Sci. Technol.* **49**, 12731–12740 (2015). doi: 10.1021/acs.est.5b03459; pmid: 26439915
49. J. E. Losey, M. Vaughan, *A.I.B.S. Bull.* **56**, 311–323 (2006).
50. P. Sukhdev, Smarter metrics will help fix our food system. *Nature* **558**, 7 (2018). doi: 10.1038/d41586-018-05328-1; pmid: 29872209
51. S. Myrick, G. W. Norton, K. N. Selvaraj, K. Natarajan, R. Muniappan, Economic impact of classical biological control of papaya mealybug in India. *Crop Prot.* **56**, 82–86 (2014). doi: 10.1016/j.cropro.2013.10.023
52. K. A. Wyckhuys *et al*., Biological control of an invasive pest eases pressures on global commodity markets. *Environ. Res. Lett.* **13**, 094005 (2018). doi: 10.1088/1748-9326/aad8f0
53. D. G. Bottrell, K. G. Schoenly, Resurrecting the ghost of green revolutions past: The brown planthopper as a recurring threat to high-yielding rice production in tropical Asia. *J. Asia Pac. Entomol.* **15**, 122–140 (2012). doi: 10.1016/j.aspen.2011.09.004
54. D. K. Letourneau *et al*., Does plant diversity benefit agroecosystems? A synthetic review. *Ecol. Appl.* **21**, 9–21 (2011). doi: 10.1890/09-2026.1; pmid: 21516884
55. H. R. Herren, P. Neuenschwander, R. D. Hennessey, W. N. O. Hammond, Introduction and dispersal of *Epidinocarsis lopezi* (Hym., Encyrtidae), an exotic parasitoid of the cassava mealybug, *Phenacoccus manihoti* (Hom., Pseudococcidae), in Africa. *Agric. Ecosyst. Environ.* **19**, 131–144 (1987). doi: 10.1016/0167-8809(87)90013-2
56. P. Kenmore *et al*., *J. Plant Protect. Trop.* **1**, 19–38 (1984).
57. H. van den Berg, J. Jiggins, Investing in farmers—The impacts of farmer field schools in relation to integrated pest management. *World Dev.* **35**, 663–686 (2007). doi: 10.1016/j.worlddev.2006.05.004
58. K. L. Heong, M. M. Escalada, N. H. Huan, V. Mai, Use of communication media in changing rice farmers' pest management in the Mekong Delta, Vietnam. *Crop Prot.* **17**, 413–425 (1998). doi: 10.1016/S0261-2194(98)00036-2
59. A. Braun, D. Duveskog, *The Farmer Field School Approach* (IFAD, 2009).
60. J. W. Ketelaar *et al*., in *Agricultural Development and Sustainable Intensification* (Routledge, 2018).
61. Z. R. Khan *et al*., Achieving food security for one million sub-Saharan African poor through push-pull innovation by 2020. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20120284 (2014). doi: 10.1098/rstb.2012.0284; pmid: 24535391
62. C. A. Midega, J. O. Pittchar, J. A. Pickett, G. W. Hailu, Z. R. Khan, A climate-adapted push-pull system effectively controls fall armyworm, *Spodoptera frugiperda* (J E Smith), in maize in East Africa. *Crop Prot.* **105**, 10–15 (2018). doi: 10.1016/j.cropro.2017.11.003
63. R. F. Pywell *et al*., Wildlife-friendly farming increases crop yield: Evidence for ecological intensification. *Proc. Biol. Sci.* **282**, 20151740 (2015). doi: 10.1098/rspb.2015.1740; pmid: 26423846
64. A. C. Franke, G. J. van den Brand, B. Vanlauwe, K. E. Giller, Sustainable intensification through rotations with grain legumes in sub-Saharan Africa: A review. *Agric. Ecosyst. Environ.* **261**, 172–185 (2018). doi: 10.1016/j.agee.2017.09.029; pmid: 29970946
65. M. Lechenet, F. Dessaint, G. Py, D. Makowski, N. Munier-Jolain, Reducing pesticide use while preserving crop productivity and profitability on arable farms. *Nat. Plants* **3**, 17008 (2017). doi: 10.1038/nplants.2017.8; pmid: 28248316
66. FAO, *Aquatic Biodiversity in Rice-Based Ecosystems* (FAO, 2014)
67. T. MacMillan, T. G. Benton, Agriculture: Engage farmers in research. *Nature* **509**, 25–27 (2014). doi: 10.1038/509025a; pmid: 24784199
68. A. Campbell, J. Alexandra, D. Curtis, Reflections on four decades of land restoration in Australia. *Rangeland J.* **39**, 405–416 (2018). doi: 10.1071/RJ17056
69. S. Spiegal *et al*., Evaluating strategies for sustainable intensification of US agriculture through the Long-Term Agroecosystem Research network. *Environ. Res. Lett.* **13**, 034031 (2018). doi: 10.1088/1748-9326/aaa779
70. P. M. Rosset, B. M. Sosa, A. M. Jaime, D. R. Lozano, The *Campesino*-to-*Campesino* agroecology movement of ANAP in Cuba: Social process methodology in the construction of sustainable peasant agriculture and food sovereignty. *J. Peasant Stud.* **38**, 161–191 (2011). doi: 10.1080/03066150.2010.538584; pmid: 21284238
71. A. J. B. Malabayabas, K. Kajisa, M. A. Mazid, F. G. Palis, D. E. Johnson, Impacts of direct-seeded and early-maturing

varieties of rice on mitigating seasonal hunger for farming communities in northwest Bangladesh. *Int. J. Agric. Sustain.* **12**, 459–470 (2014). doi: 10.1080/14735903.2014.927980
72. W. Zhang *et al.*, Closing yield gaps in China by empowering smallholder farmers. *Nature* **537**, 671–674 (2016). doi: 10.1038/nature19368; pmid: 27602513
73. C. Morris *et al.*, *Asp. Appl. Biol.* **136**, 19–26 (2017).
74. J. Poore, T. Nemecek, Reducing food's environmental impacts through producers and consumers. *Science* **360**, 987–992 (2018). doi: 10.1126/science.aaq0216; pmid: 29853680
75. V. T. Kumar, *Zero-Budget Nature Farming* (Government of Andhra Pradesh, 2017).
76. Xinhua, CPC and State Council Guide Opinion on Using New Development Concepts to Accelerate Agricultural Modernisation and Realise Moderate Prosperity Society, (2016); http://news.xinhuanet.com/fortune/2016-01/27/c_1117916568.htm.

ACKNOWLEDGMENTS

Competing interests: The author declares that there are no conflicts of interest regarding this research, nor funding or remuneration from consultancies or industrial links. There are no external funding sources for this paper nor paid consultancies for industry. J.P. is chair of the Kansas State University Sustainable Intensification Innovation Lab, funded by the United States Agency for International Development (USAID).

10.1126/science.aav0294

RESEARCH ARTICLE SUMMARY

SKIN REPAIR

Myofibroblast proliferation and heterogeneity are supported by macrophages during skin repair

Brett A. Shook*, Renee R. Wasko, Guillermo C. Rivera-Gonzalez, Emilio Salazar-Gatzimas, Francesc López-Giráldez, Biraja C. Dash, Andrés R. Muñoz-Rojas, Krystal D. Aultman, Rachel K. Zwick, Vivian Lei, Jack L. Arbiser, Kathryn Miller-Jensen, Damon A. Clark, Henry C. Hsia, Valerie Horsley*

INTRODUCTION: Fibroblasts produce extracellular matrix (ECM) molecules that regulate tissue strength and resilience. Imbalanced ECM maintenance leads to tissue dysfunction. Although multiple populations of fibroblasts support uninjured skin function, the extent of fibroblast diversity and the presence of functionally distinct subsets in adult fibrotic skin are poorly defined. The inability to find a single molecular marker that identifies all activated "myofibroblasts" during wound healing suggests the existence of multiple subsets of ECM-producing myofibroblasts. Furthermore, little is known about the cellular and molecular mechanisms that promote the expansion of individual myofibroblast subsets and support cellular heterogeneity. Although macrophages influence myofibroblast numbers and ECM deposition after injury, subpopulations of wound bed macrophages have only recently been defined, and specific interactions with fibroblasts have not been explored.

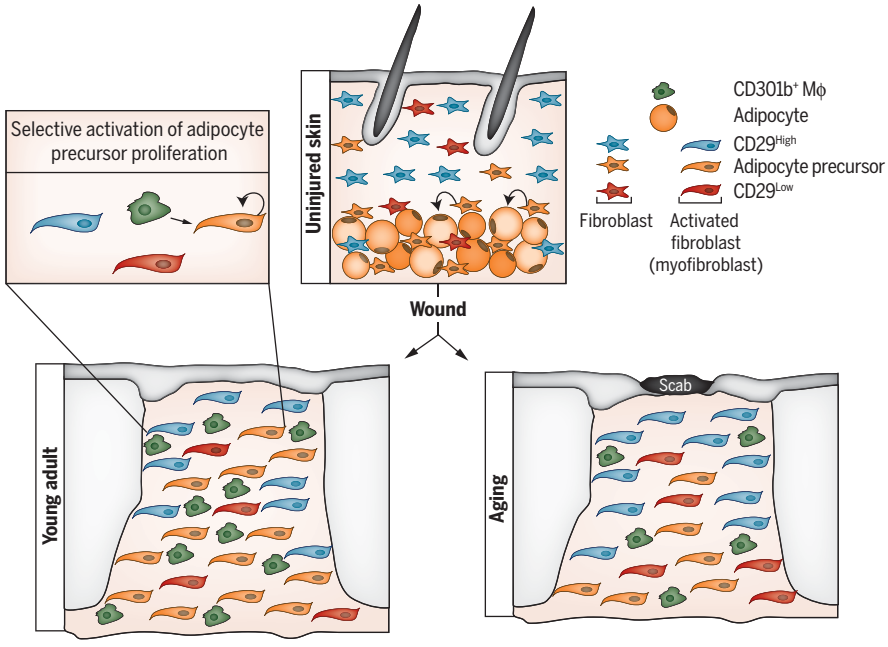
RATIONALE: Variation in healing and scarring rates in multiple tissues suggests that fibroblast diversity exists. To develop therapies targeting fibrotic responses, functionally distinct subsets of myofibroblasts and the mechanisms that support individual populations must be uncovered. We used a comprehensive, hierarchical fluorescence-activated cell sorting strategy to define myofibroblast subsets in skin wound beds from adult mice. This strategy revealed distinct subsets of wound bed myofibroblasts, including an abundant population that contains the cell surface marker profile of adipocyte precursor cells (APs). We examined myofibroblast subsets in different cutaneous fibrotic contexts and explored mechanisms that selectively promote the expansion of APs.

RESULTS: Genetic lineage tracing and flow cytometry revealed distinct subsets of wound bed myofibroblasts that express smooth muscle actin and collagen. The most abundant populations were CD26-expressing APs and a subset with high cell surface levels of CD29 ($CD29^{High}$). Although transcriptomic analysis revealed that each myofibroblast subset has a distinct gene expression profile, functional analyses suggest that myofibroblast subsets make both overlapping and distinct contributions to repair. APs were significantly reduced and $CD29^{High}$ cells were more abundant in wound beds from aged mice and skin from mice that underwent bleomycin-induced fibrosis, suggesting that the fibrotic environment influences myofibroblast composition. Injury and repair-related changes in AP transcription implicated macrophage signaling in the modulation of AP gene expression. Genetic ablation and cell transplantations of different myeloid cells revealed that macrophages expressing macrophage galactose *N*-acetylgalactosamine–specific lectin 2 (*Mgl2*/CD301b) directly stimulate proliferation in a subset of APs and not in other myofibroblast subsets. By combining in vitro cytokine stimulation with in vivo signaling pathway inhibition, we identified multiple $CD301b^{+}$ macrophage–secreted factors (platelet-derived growth factor C and insulin-like growth factor 1) that selectively stimulate AP proliferation, thus supporting the heterogeneity of wound bed myofibroblasts.

CONCLUSION: We identified multiple populations of skin myofibroblasts and observed that the composition of myofibroblasts is dependent upon the fibrotic environment. Distinct interactions allow $CD301b^{+}$ macrophage–derived signaling to selectively activate the proliferation of APs and not other myofibroblasts. These results have potential for the development of therapies that target multiple cellular populations or signaling pathways under conditions associated with excessive or deficient ECM deposition, such as wound healing and fibrosis. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aar2971



Regulation of myofibroblast diversity in skin wounds. After injury, multiple subsets of fibroblasts become activated myofibroblasts that contribute to tissue repair and scar formation. Wound bed macrophages expressing CD301b selectively activate proliferation in APs and not other myofibroblasts. With age, impaired healing is associated with a reduction in $CD301b^{+}$ macrophages (Mϕ) and APs. These findings identify distinct cellular and molecular interactions that support myofibroblast heterogeneity.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: brett.shook@yale.edu (B.A.S.); valerie.horsley@yale.edu (V.H.)
Cite this article as B. A. Shook *et al.*, *Science* **362**, eaar2971 (2018). DOI: 10.1126/science.aar2971

RESEARCH ARTICLE

SKIN REPAIR

Myofibroblast proliferation and heterogeneity are supported by macrophages during skin repair

Brett A. Shook[1]*, Renee R. Wasko[1], Guillermo C. Rivera-Gonzalez[1], Emilio Salazar-Gatzimas[2], Francesc López-Giráldez[3], Biraja C. Dash[4], Andrés R. Muñoz-Rojas[5], Krystal D. Aultman[1], Rachel K. Zwick[1], Vivian Lei[1], Jack L. Arbiser[6], Kathryn Miller-Jensen[1,5], Damon A. Clark[2], Henry C. Hsia[4], Valerie Horsley[1,7]*

During tissue repair, myofibroblasts produce extracellular matrix (ECM) molecules for tissue resilience and strength. Altered ECM deposition can lead to tissue dysfunction and disease. Identification of distinct myofibroblast subsets is necessary to develop treatments for these disorders. We analyzed profibrotic cells during mouse skin wound healing, fibrosis, and aging and identified distinct subpopulations of myofibroblasts, including adipocyte precursors (APs). Multiple mouse models and transplantation assays demonstrate that proliferation of APs but not other myofibroblasts is activated by CD301b-expressing macrophages through insulin-like growth factor 1 and platelet-derived growth factor C. With age, wound bed APs and differential gene expression between myofibroblast subsets are reduced. Our findings identify multiple fibrotic cell populations and suggest that the environment dictates functional myofibroblast heterogeneity, which is driven by fibroblast-immune interactions after wounding.

Tissues sustain resilience and strength through the maintenance of extracellular matrix (ECM) molecules by mesenchymal cells. Under disease states, profibrotic conditions lead to excessive and disordered ECM deposition that impairs tissue function (*1*, *2*). Additionally, dysregulated ECM is associated with aged skin and age-related defective wound healing (*3*–*8*). Variability in rates of wound healing, scarring, and fibrosis may result from functionally distinct mesenchymal cells (*9*, *10*). Thus, identifying distinct mesenchymal cell populations that contribute to fibrosis and the mechanisms that drive cellular diversity has substantial implications for disease treatment (*2*, *11*–*13*).

Prior experiments in mice have demonstrated that embryonic mesenchymal precursors expressing *Engrailed* (*En1*) or *Delta-like homolog 1* (*Dlk1*/*Pref1*) generate skin fibroblast and adipocyte lineages (*14*–*16*). During skin repair after injury, myofibroblasts expressing alpha smooth muscle actin (SMA), *Pdgfra*, *Sca1*, *Itga8*, *CD34*, and *Dpp4* (CD26) migrate, proliferate, and deposit ECM (*17*–*19*). Myofibroblasts do not form lipid-filled adipocytes within regenerated tissue after a standard small skin injury (*14*, *15*, *20*, *21*) but can form adipocytes that regenerate hair follicles in large wounds (*20*). How environmental conditions alter functional cellular diversity and the contribution of mesenchymal subsets to tissue fibrosis are not well understood.

In this study, we uncover unappreciated heterogeneity within wound bed myofibroblasts that is dependent on the tissue environment. In particular, we show that the predominant population of myofibroblasts is adipocyte precursor cells

[1]Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06511, USA. [2]Interdepartmental Neuroscience Program, Yale University, New Haven, CT 06511, USA. [3]Yale Center for Genome Analysis, Yale School of Medicine, New Haven, CT 06510, USA. [4]Department of Surgery (Plastic), Yale School of Medicine, New Haven, CT 06510, USA. [5]Department of Biomedical Engineering, Yale University, New Haven, CT 06511, USA. [6]Department of Dermatology, Atlanta Veterans Administration Health Center, Emory University, Atlanta, GA 30322, USA. [7]Department of Dermatology, Yale University, New Haven, CT 06511, USA.
***Corresponding author. Email: brett.shook@yale.edu (B.A.S.); valerie.horsley@yale.edu (V.H.)**

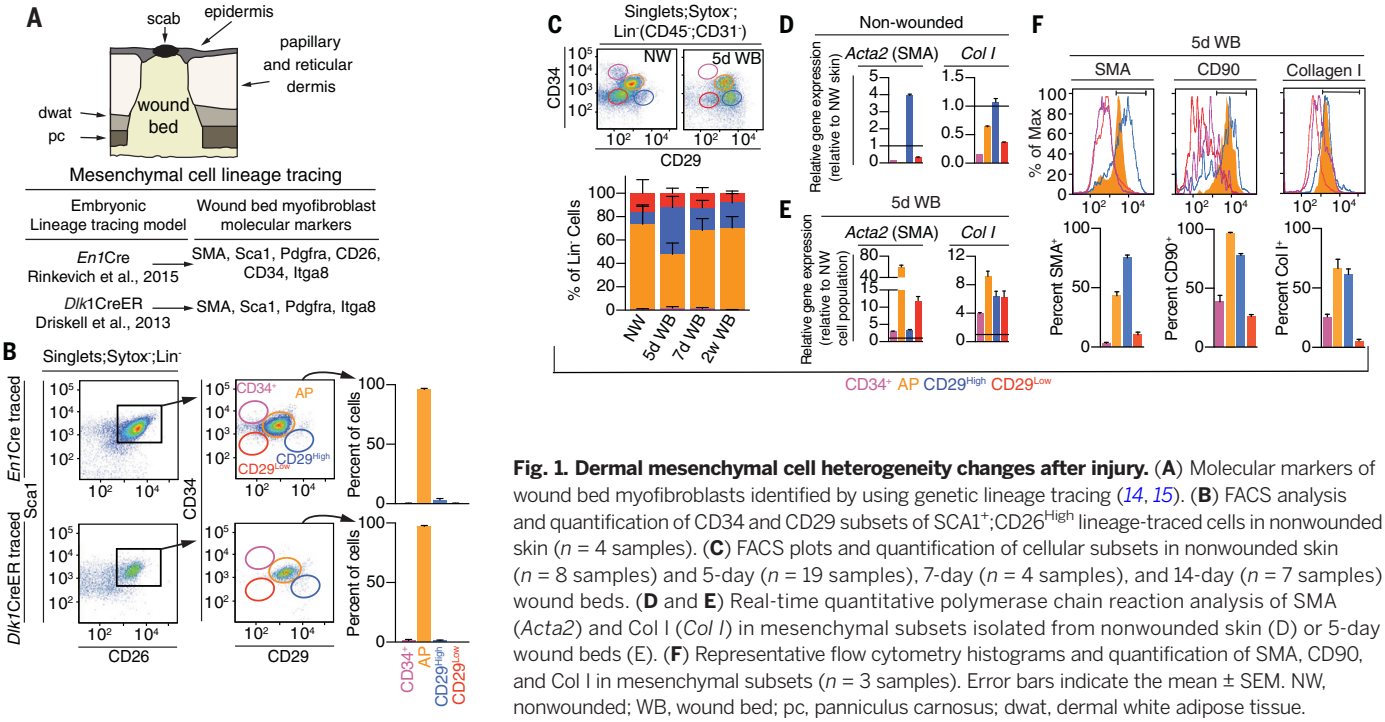


Fig. 1. Dermal mesenchymal cell heterogeneity changes after injury. (**A**) Molecular markers of wound bed myofibroblasts identified by using genetic lineage tracing (*14*, *15*). (**B**) FACS analysis and quantification of CD34 and CD29 subsets of SCA1^{+};CD26^{High} lineage-traced cells in nonwounded skin (n = 4 samples). (**C**) FACS plots and quantification of cellular subsets in nonwounded skin (n = 8 samples) and 5-day (n = 19 samples), 7-day (n = 4 samples), and 14-day (n = 7 samples) wound beds. (**D** and **E**) Real-time quantitative polymerase chain reaction analysis of SMA (*Acta2*) and Col I (*Col I*) in mesenchymal subsets isolated from nonwounded skin (D) or 5-day wound beds (E). (**F**) Representative flow cytometry histograms and quantification of SMA, CD90, and Col I in mesenchymal subsets (n = 3 samples). Error bars indicate the mean ± SEM. NW, nonwounded; WB, wound bed; pc, panniculus carnosus; dwat, dermal white adipose tissue.

(APs) derived from *En1* lineage–traced fibroblasts (*14*, *15*, *21*, *22*), which contribute to tissue repair and ECM production and modulation. We show that, in wound beds of aged mice, APs are markedly reduced and wound bed myofibroblast subpopulations become more homogeneous in their gene expression profiles and localization. Our data indicate that CD301b$^+$ macrophage–derived platelet-derived growth factor C (PDGFC) and insulin-like growth factor (IGF) signaling contributes to myofibroblast heterogeneity by selectively promoting the proliferation of wound bed APs and not other myofibroblast subsets. These findings define major subsets of wound bed myofibroblasts and identify immune and molecular interactions that promote functional cellular heterogeneity under distinct fibrotic conditions.

Mesenchymal cell heterogeneity under fibrotic conditions

Myofibroblasts within wound beds express PDGF receptor α (PDGFRα), CD34, and SCA1 and derive from embryonic precursors that express *En1* or *Dlk1/Pref1* (*14*–*16*) (*14*, *15*, *23*, *24*) (Fig. 1A). Because PDGFRα, CD34, and SCA1 define APs (SCA1$^+$;CD34$^+$;CD29$^+$) (*25*, *26*), we sought to determine whether APs were derived from *En1*- and *Dlk1*-expressing precursors. We confirmed that *En1*Cre;*Rosa26*-LSL-tdTomato– and *Dlk1*CreER; mT/mG–traced cells expressed CD26 and SCA1 (*14*, *15*) (Fig. 1B). Ninety-six percent of tdTomato$^+$; SCA1$^+$;CD26$^+$ cells in *En1*Cre;tdTomato and 97% of GFP$^+$;SCA1$^+$;CD26$^+$ cells in *Dlk1*CreER;mTmG nonwounded skin express AP markers CD29 and CD34 (Fig. 1B and fig. S1).

Flow cytometry analysis of immune and endothelial lineage–negative cells (Lin$^-$) isolated from uninjured dermis and 5-day wound beds revealed four populations of cells: CD29$^+$;CD34$^+$, CD34$^+$, CD29^{High} (cells with high surface levels of CD29), and CD29^{Low} cells (Fig. 1C). We define CD29$^+$;CD34$^+$ cells as APs because 90% retain SCA1$^+$ expression after injury and have adipogenic potential in vitro and in vivo (*26*–*28*) (fig. S2A). Although AP and CD29^{Low} cells were the most abundant populations in nonwounded skin, wound beds contained increased proportions of CD29^{High} cells (Fig. 1C).

To determine which populations were myofibroblasts, we analyzed the expression of SMA and collagen I (Col I). Within wound beds, each cell population up-regulated SMA and Col I mRNA expression compared with that in cell populations from uninjured skin (Fig. 1, D and E); however, flow cytometry revealed that only APs and CD29^{High} cells were enriched for SMA, Col I, and the fibroblast marker CD90 in wound beds (Fig. 1F).

To further analyze the fibrotic nature of these cell populations, we examined the expression of profibrotic proteins SCA1, CD9, CD26, and PDGFRα (*14*–*16*, *22*, *29*) (figs. S1B and S2, A to D). A greater percentage of APs and CD29^{High} and CD29^{Low} cells expressed CD9 after injury (fig. S2D), and CD9$^+$ APs have decreased in vitro adipogenic potential compared with CD9$^-$ APs (fig. S2, E and F). These data suggest that fibrotic cells are

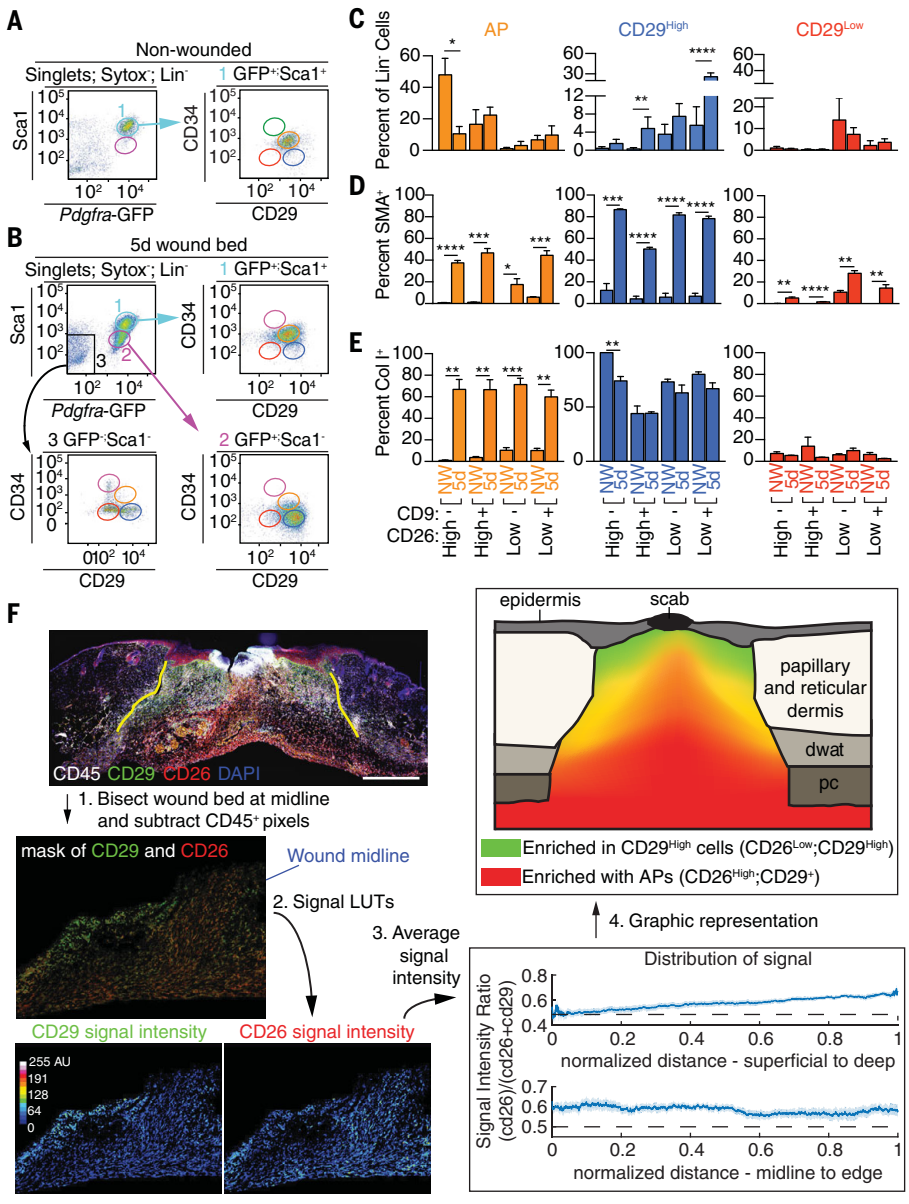


Fig. 2. Skin wounds contain multiple myofibroblast subsets. (**A** and **B**) FACS plots detailing the gating strategy to define mesenchymal subpopulations. (**C** to **E**) Quantification of the relative abundances of prevalent profibrotic subsets (n = 6 samples) (C) and colocalization with SMA (D) and Col I (E) in nonwounded skin and 5-day wound bed mesenchymal subsets (n = 3 samples). (**F**) Pipeline for processing immunostained tissue sections to infer the locations of APs (CD29$^+$;CD26^{High}) and CD29^{High} cells in day 5 wound beds. Yellow lines delineate wound edges. Scale bar, 250 μm. Error bars indicate the mean ± SEM. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. AU, arbitrary units; LUT, look-up table.

heterogeneous and distinct between uninjured and injured skin. Further, at least two major populations of fibrotic mesenchymal cells exist in skin wounds: APs and CD29^{High} cells.

We next examined CD29 and CD34 populations in bleomycin-induced fibrosis. CD29^{High} cells increased more robustly and fewer APs were observed after bleomycin treatment than in the context of wound healing (fig. S2, G and H). Colocalization with other profibrotic markers was not markedly changed (fig. S2, B to D). Thus, the profibrotic cellular composition in bleomycin-treated skin is distinct from that in wound healing, suggesting that distinct strategies are required to treat tissue fibrosis under different pathological conditions.

Because SMA and CD9 expression increased in multiple populations of mesenchymal cells within skin wounds, we sought to design a comprehensive hierarchical marker panel to delineate mesenchymal heterogeneity in skin wounds by using six fibrotic markers (Fig. 2, A and B, and fig. S1). Lin$^-$ cells were subdivided on the basis of PDGFRα and SCA1 (*14*, *15*). Whereas PDGF

signaling is central to fibrosis (*2*, *21*, *30*–*32*), we included PDGFRα$^-$ cells in our analysis because PDGFRα$^-$ profibrotic cells may also contribute to repair (*14*, *33*). We further subdivided populations on the basis of CD29 and CD34 expression and then by the presence of CD26 (high or low) and CD9 (CD9$^+$ or CD9$^-$). This analysis revealed that 54% of Lin$^-$ cells in nonwounded skin contained surface markers that prospectively identify APs: PDGFRα$^+$;SCA1$^+$;CD29$^+$;CD34$^+$ (Fig. 2A and figs. S2A and S3A). Within the AP pool, 66% were CD26^{High};CD9$^-$ cells (Fig. 2, A to C, and fig. S4, A and B). The only other nonimmune or nonvascular populations that contribute more than 3% of total cells in nonwounded skin were PDGFRα$^-$;SCA1$^-$;CD29^{High} (CD29^{High}) cells (~7%) and PDGFRα$^-$; SCA1$^-$;CD29^{Low} (CD29^{Low}) cells (~15%) (Fig. 2, A to C, and figs. S2A and S3A).

In 5-day wound beds, the relative abundance of APs decreased as PDGFRα$^+$;CD29^{High} cells and CD29^{Low} cells became more abundant (Fig. 2, A to C, and figs. S3 and S4, A and B). Because CD34$^+$ cells were mostly negative for SMA (Fig. 1F) and not abundant during repair, we excluded them from further analysis. CD29^{High} and CD29^{Low} cells remained predominantly CD26^{Low};CD9$^+$ and CD26^{Low};CD9$^-$, respectively (Fig. 2, A to C, and fig. S4, A and B). Forty-six percent of APs were CD26^{High};CD9$^+$ cells in 5-day wound beds, compared with 24% in nonwounded skin (Fig. 2, A to C, and fig. S4, A and B). The shift from CD9$^-$ to CD9$^+$ APs persisted 14 days after injury (fig. S3E). Immunofluorescent staining of fluorescence-activated cell sorting (FACS)–isolated cells and wound beds showed greater intensity of CD26 staining in APs than in other cells, confirming our flow cytometry results (fig. S4, C and D). SMA and Col I were expressed by APs and CD29^{High} cells in wound beds regardless of the cells' CD26 or CD9 expression; however, few CD29^{Low} cells expressed SMA, Col I, CD90, or ER-TR7 (Figs. 1, D to F, and 2, D and E, and fig. S4, C to H).

To explore the spatial organization of myofibroblast subsets, we regionally dissected wound beds for flow cytometry and examined the signal intensity of CD26 and CD29 in nonimmune (CD45$^-$) cells in tissue sections. Although APs are found throughout the wound bed, CD29^{High} cells were biased toward the most superficial region of the outer wound bed edge (Fig. 2F and fig. S5). CD29^{High} cells were more abundant in the upper dermis of nonwounded skin, which contributes to the superficial regenerating dermis (*15*). Human skin has a composition of mesenchymal cells similar to that of mouse skin. However, greater percentages of these populations were CD26^{High} and CD9$^+$ cells (fig. S6), indicating that human skin may be more biased toward fibrotic responses.

To determine whether repair-related myofibroblast heterogeneity arises from conversion between cellular subsets, we performed genetic lineage tracing by using inducible Cre-lox mouse lines that label profibrotic cells: *Dlk1*CreER;mT/mG (postnatal labeling) and *Pdgfra*CreER;mT/mG (adult labeling) (fig. S7A) (*15*). In uninjured skin, several mesenchymal populations were labeled in *Dlk1*CreER mice, whereas *Pdgfra*CreER mice predominantly labeled APs (~94%) (fig. S7, B and C). *Pdgfra*CreER lineage–traced cells in 5-day wound beds contributed to APs, SCA1$^-$;CD29^{Low} cells, and a rare population of SCA1$^+$;CD29^{High} cells (fig. S7, C to F). SCA1$^+$;CD29^{High} cells are CD9$^+$ during repair and similar in size to APs (fig. S7, E and F), suggesting that they are profibrotic. Two weeks after injury, *Pdgfra*CreER lineage–traced cells comprised ~80% APs, ~10% SCA1$^+$;CD29^{High} cells, and 7% SCA1$^-$;CD29^{Low} cells (fig. S7D). These data suggest that APs contribute to multiple myofibroblast subpopulations; however, they do not contribute to the expansion of SCA1$^-$;CD29^{High} cells. Yet, because non-AP, CD29$^+$ cells are labeled in *Pdgfra*CreER mice, we cannot rule out the possibility that the proliferation of CD29$^+$ cells (fig. S12) also contributes to myofibroblast heterogeneity after injury.

Myofibroblast subsets have distinct gene expression profiles

The comparison of transcriptional profiles of CD29^{Low} and CD29^{High} cells and APs that were either CD9 positive or negative by RNA sequencing (RNA-seq) (n = 2 profiles for each population) confirmed significant diversity among wound bed myofibroblasts (Fig. 3, A and B). Although transcriptomic analysis revealed that CD9$^+$ and CD9$^-$ APs were similar, each mesenchymal subset expressed distinct mRNAs (figs. S8 and S9 and table S1). For instance, CD29^{High} cells had elevated expression of *Pdgfrβ*, *CD146*, *NG2*, and other perivascular cell or pericyte markers (*34*) compared with APs and CD29^{Low} cells and had elevated *Acta2* expression in nonwounded skin (Fig. 1D).

Although transcriptomes were distinct between cellular subsets, Ingenuity Pathway Analysis (IPA) predicted common active biofunctions and similar upstream activators of gene expression

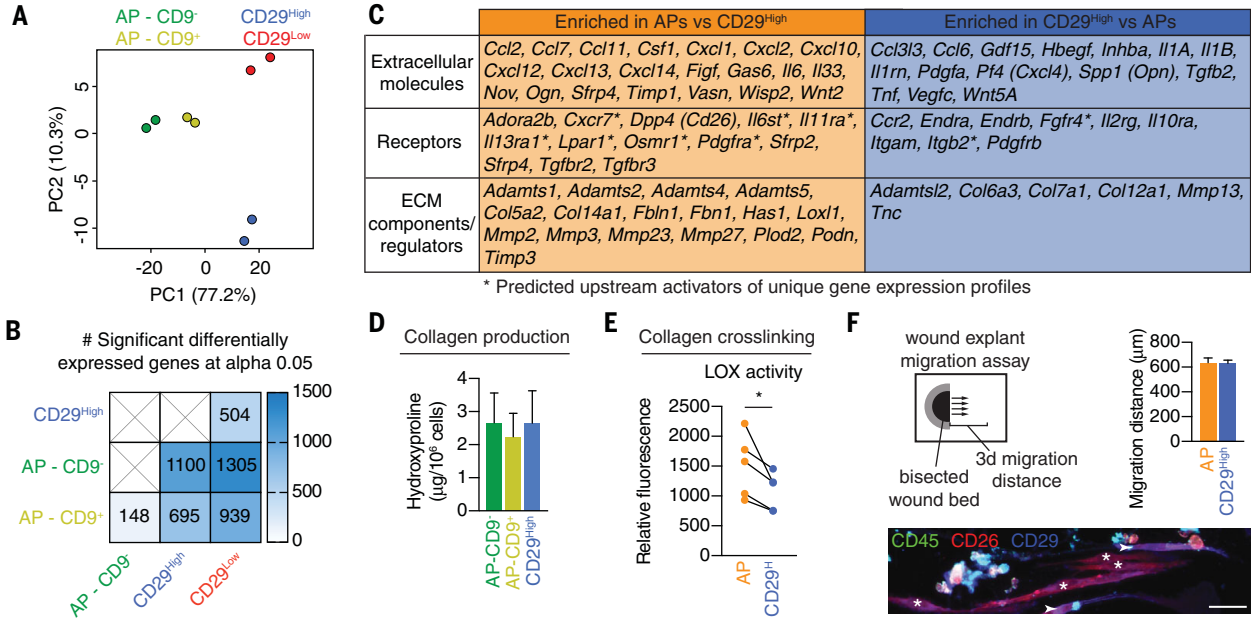


Fig. 3. Myofibroblast subsets can distinctively regulate repair. (**A**) Transcriptomic principal components analysis of myofibroblast subsets. (**B**) Table of genes with statistically significant differences in expression between cellular subsets. (**C**) Wound healing–related genes enriched in APs (CD9$^-$ and CD9$^+$ AP populations) or CD29^{High} cells. (**D** and **E**) Quantification of hydroxyproline content (n = 7 samples) (D) and lysyl oxidase (LOX) activity (n = 4 samples; *P = 0.0416) (E) in cells from day 5 wounds. (**F**) Migration distance of APs (CD26^{High}) (asterisks) and CD29^{High} cells (arrowheads) from cultured wound beds (n ≥ 250 cells from 3 wound beds). Scale bar, 10 μm. Error bars indicate the mean ± SEM.

profiles (figs. S10 and S11 and table S2), suggesting some functional redundancy among myofibroblasts. However, many genes differentially expressed between APs and CD29^{High} cells have been implicated in wound healing (Fig. 3C and figs. S8 and S9). Additionally, each myofibroblast population was enriched for different ECM components and modifiers (Fig. 3C and figs. S8 and S9). Both CD9^{+} and CD9^{-} APs were enriched for many cytokine genes (*Ccl2*, *Cxcl1*, *Cxcl10*, and *Cxcl12*) and ECM components (*Col5a2*, *Fbln1*, *Fbn1*, *Has1*, and *Loxl1*) that promote rapid ECM deposition (*35*–*37*). The enrichment of genes involved in repair and fibrosis changed among myofibroblast populations between day 5 and day 14 of repair (fig. S8B), indicating that myofibroblast subsets can distinctively influence both the proliferative and maturation phases of tissue repair.

To determine whether myofibroblast subpopulations were functionally distinct, we examined collagen production, collagen cross-linking, and the migration of APs and CD29^{High} cells. We did not observe differences in collagen production or cellular migration; however, we detected an increased ability of APs to cross-link collagen compared with CD29^{High} cells, consistent with elevated *Lox* expression in APs (Fig. 3, C to F, and fig. S8A).

Because myofibroblast numbers increase after injury (Fig. 1C and fig. S3), we examined in vivo proliferation within the different mesenchymal subsets (APs and CD29^{High} and CD29^{Low} cells) during tissue repair. Proliferation increased in APs and CD29^{High} and CD29^{Low} cells after injury, and CD26^{Low} cells were more proliferative than CD26^{High} cells within each cellular subset (fig. S12, A to D). Taken together, these data demonstrate that the dermis contains tremendous heterogeneity within profibrotic cells, which have distinct functions during tissue repair.

Myofibroblast composition and gene expression are altered during aging

Age-related defects in repair are associated with reduced myofibroblasts and dysfunctional ECM deposition (*3*–*6*) (fig. S13, A and B). To determine whether mesenchymal populations were altered with age, we analyzed 5-day wound beds in young and aged mice. The relative abundance of APs decreased and CD29^{High} cells increased in wound beds from aged mice (Fig. 4, A to D), with reduced percentages of CD9^{+} cells in all mesenchymal populations (Fig. 4C), suggesting that fibrotic cells are lost or unstimulated with age.

Analysis of transcriptional changes in myofibroblasts during aging by RNA-seq (n = 2 profiles) revealed fewer differentially expressed genes between myofibroblast subsets (fig. S14, A to C) because of age-related down-regulation of many genes within individual populations (fig. S14D). Comparing the transcriptomes of myofibroblast populations in young versus aged mice revealed age-related changes in gene expression of extracellular molecules (Fig. 4E) and increased expression of multiple metalloproteases in myofibroblasts, consistent with the ability of aged fibroblasts to break down ECM faster than young fibroblasts and impair healing (*4*, *6*).

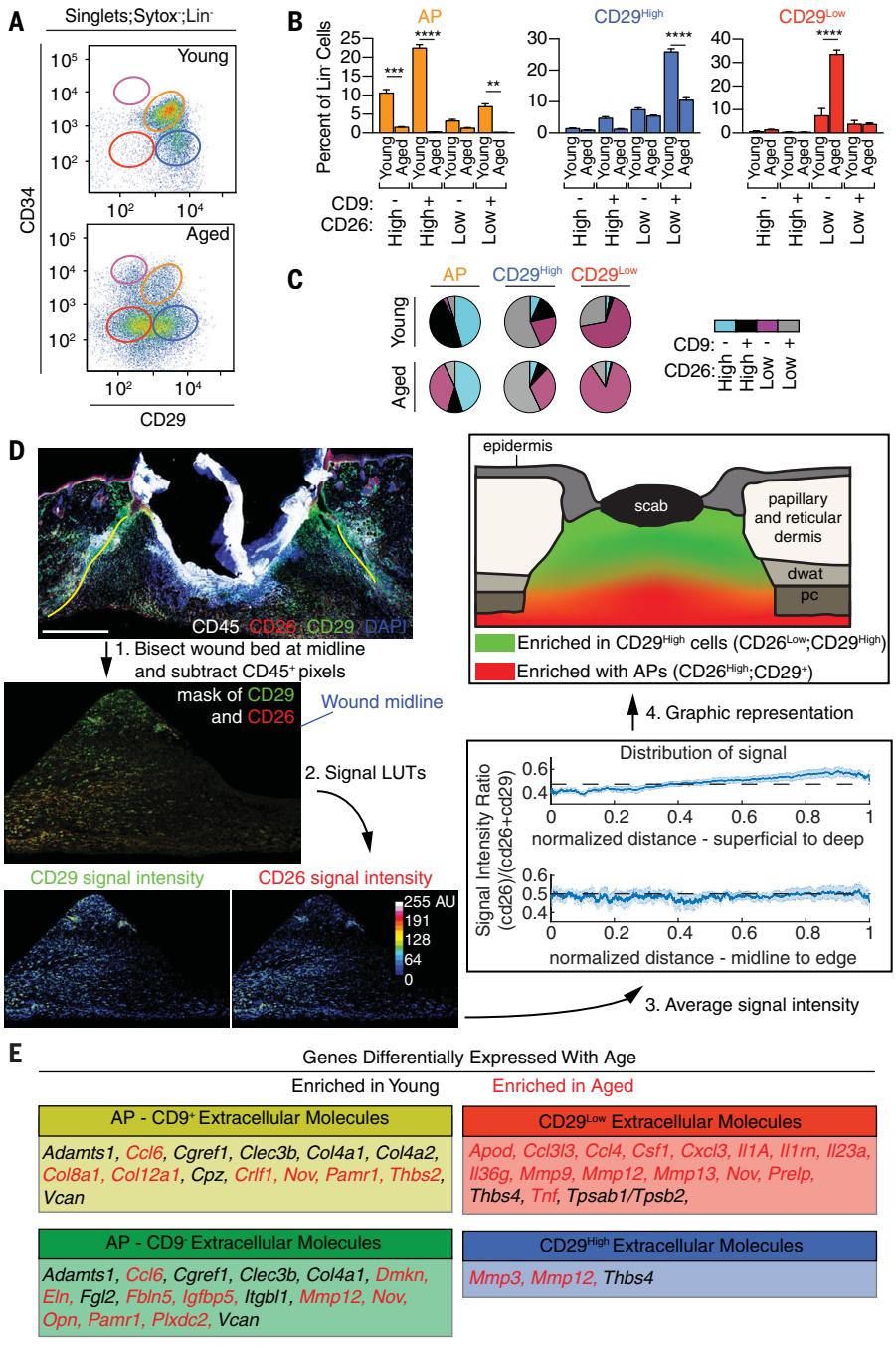


Fig. 4. Myofibroblast composition and gene expression is altered during aging. (**A**) FACS plots for 5-day wounds from young and aged mice. (**B**) Quantification of the relative abundance of prevalent profibrotic subsets in 5-day wounds (n = 4 samples). (**C**) Pie charts depicting CD9 and CD26 colocalization. (**D**) Pipeline for processing immunostained wound bed sections to infer AP and CD29^{High} cell locations in day 5 wound beds from aged mice. Yellow lines delineate wound edges. Scale bar, 250 μm. (**E**) Genes with altered differential expression with age. Black and red text indicates enrichment in young and aged mice, respectively. Error bars indicate the mean ± SEM. $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

APs become fibrotic after injury

To identify molecular mechanisms regulating AP myofibroblasts during repair, we isolated APs from uninjured skin and 5-day wound beds and performed RNA-seq (n = 2 profiles). Injury and repair up-regulated *Acta2* (SMA) and several secreted factors implicated in tissue repair (fig. S15, A to C). Several adipogenic genes and in vitro adipogenic potential were reduced in wound bed

APs (fig. S15, D and E). Thus, APs displayed dramatic alterations within the wound environment that limit their adipogenic potential, promote myofibroblast gene expression, and may explain the myofibroblast origin of adipocytes in large wounds (*20*).

Macrophage signaling selectively activates proliferation of wound bed APs

Because delayed healing in aged mice is associated with decreased APs and APs rapidly increase from days 3 to 7 after injury, when new dermal tissue is generated, we investigated potential signaling pathways that could affect AP numbers during repair. IPA predicted that injury-related changes in AP gene expression could result from monocyte-macrophage–derived ligands (fig. S15F). Macrophage ablation reduces wound bed myofibroblast numbers, impairs myofibroblast function, and impairs wound healing (*38*–*41*); however, the underlying mechanisms are ill defined. To examine the contribution of monocytes and macrophages to myofibroblast heterogeneity, we ablated macrophages by using *LysM*Cre;iDTR mice (*38*, *41*–*43*) (Fig. 5A). Ablating monocytes and macrophages reduced all AP subsets (fig. S16) and diminished AP proliferation in wounds without significantly changing $CD29^{High}$ or $CD29^{Low}$ populations (Fig. 5B and fig. S16A). Additionally, pharmacological reduction of macrophages decreased the percentage of dividing APs from 25 to 9% in controls (fig. S17), indicating that the myeloid lineage in 5-day wound beds selectively activates the proliferation of APs and not other myofibroblasts.

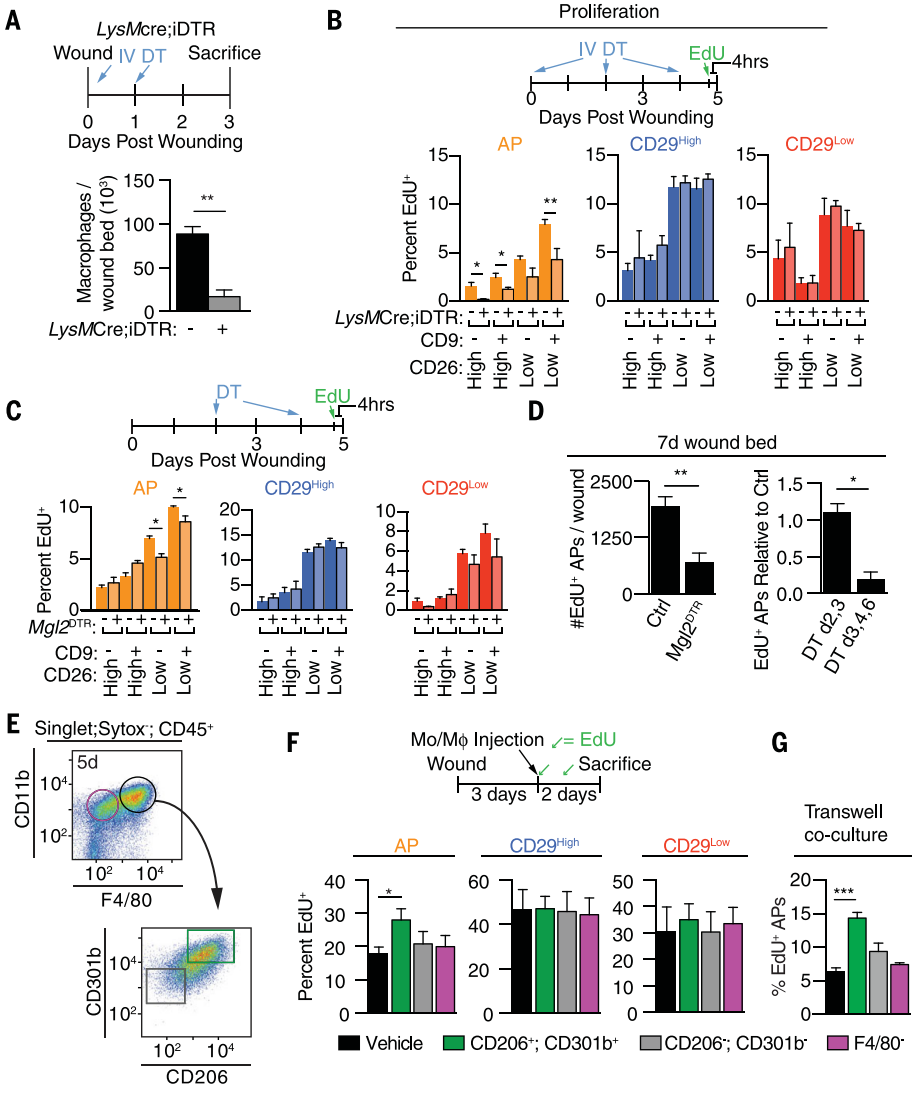


Fig. 5. $CD301b^{+}$ macrophages selectively stimulate AP proliferation during wound healing. (**A**) Quantification of wound bed macrophage depletion ($n \geq 3$ samples; $P = 0.004$). (**B**) Quantification of myofibroblast proliferation in wound beds ($n = 4$ samples). (**C**) Quantification of cell proliferation in wound beds of $CD301b^{+}$ macrophage–depleted mice ($Mgl2^{DTR}$) ($n = 3$ samples). (**D**) Quantification of EdU-incorporating APs in mice receiving DT on days 2, 4, and 6 after injury (left) ($n = 3$ samples, $P = 0.0469$) and DT on days 2 and 3 or 3, 4, and 6 after injury (right) ($n = 3$ samples, $P = 0.0116$). Mice were given two injections of EdU per day from days 3 through 7 after injury. (**E**) FACS plots of immune cell populations isolated for transplants. (**F** and **G**) Quantification of EdU-incorporating cells after injection of select immune cell subsets in vivo ($n = 5$ samples, $P = 0.0146$) (F) or Transwell coculture ($n = 6$ samples, $P = 0.001$) (G). Error bars indicate the mean ± SEM. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. IV DT, intravenous DT; Mo/Mϕ, monocyte-macrophage.

$CD301b^{+}$ macrophages activate AP proliferation during wound healing

During the mid-phase of wound healing (days 3 to 7), the myeloid lineage is composed predominantly of monocyte and macrophage subsets (*41*, *44*, *45*) (fig. S18A). We have previously shown that wound bed macrophages expressing macrophage galactose-type C-type lectin 2 (*Mgl2*/CD301b) contribute to repair by promoting proliferation and fibroblast repopulation and that $CD301b^{+}$ macrophages are ablated in $Mgl2^{DTR}$ mice (*41*, *46*, *47*). Whereas ablating all macrophages in *LysM*Cre;iDTR mice decreased the proliferation of all subsets of APs, ablating CD301b-expressing macrophages reduced the proliferation of $CD26^{Low}$ APs in 5-day wounds, with no change in $CD29^{High}$ or $CD29^{Low}$ cell proliferation (Fig. 5C). This suggests that diversity in wound bed macrophages (*48*) allows the proliferation of different AP subsets to be differentially regulated, thus promoting myofibroblast heterogeneity. Reduced AP proliferation in diphtheria toxin (DT)–treated $Mgl2^{DTR}$ mice resulted in a ~50% reduction in 5-ethynyl-2′-deoxyuridine–positive (EdU^{+}) APs in 7-day wound beds that was observed only when newly generated $CD301b^{+}$ macrophages were ablated during the proliferative phase of repair (Fig. 5D). Consistent with this model, cell transplantation of $CD301b^{+}$ macrophages, and not that of other immune cells, increased AP proliferation (from 18 to 28%), whereas $CD29^{High}$ and $CD29^{Low}$ cells were unaltered (Fig. 5, E and F). Further, cultured $CD301b^{+}$ macrophages doubled AP proliferation in vitro (Fig. 5G), demonstrating direct signaling between $CD301b^{+}$ macrophages and APs.

To identify signaling molecules that activate AP proliferation during repair, we compared the transcriptomes of $CD301b^{+}$ macrophages with those of $F4/80^{-}$ immune cells isolated from day 5 wounds (Fig. 6A and fig. S18, B and C) ($n = 2$ profiles per group). We identified ligands enriched in $CD301b^{+}$ macrophages that bind to receptors on APs (Fig. 6B and fig. S18D) and validated these results by quantifying protein secretion (fig. S18E). Cultured APs were treated with candidate molecules, and only PDGFC and IGF1 induced proliferation (Fig. 6C). To determine whether PDGFC and IGF1 signaling pathways contribute to AP proliferation in vivo, we administered ligand-neutralizing antibodies or receptor antagonists after injury (Fig. 6D). Local injection of PDGFC- or IGF1-neutralizing antibodies in vivo reduced AP proliferation; however, no change in the proliferation of other cells was detectable.

Additionally, the inhibition of PDGFRα and IGF1 receptor (IGF1R) or downstream phosphatidylinositol 3-kinase (PI 3-kinase) signaling selectively reduced AP proliferation (Fig. 6D). We did not observe spatial biasing of CD301b$^+$ macrophages in wound beds (Fig. 6E), and the expression of wound healing–associated genes changed minimally in myofibroblasts from 5-day wound beds of *Mgl2*^{DTR} mice relative to controls (fig. S19). These data suggest that the distinct gene expression profile of each myofibroblast subset results from interactions with other tissue resident cells, such as keratinocytes (*49*). As a result, the delayed re-epithelialization and revascularization observed in *Mgl2*^{DTR} mice (*41*) may result from CD301b$^+$ macrophages interacting with keratinocytes and endothelial cells. IGF1 can stimulate repair, potentially by promoting the migration and proliferation of keratinocytes and fibroblasts (*35*, *50*–*53*), yet the contribution of PDGFC to healing has not previously been explored. To examine the contribution of PDGFC signaling to wound healing, we locally injected a PDGFC-neutralizing antibody at the periphery of wound beds and examined skin repair. We did not observe gross changes in re-revascularization and myofibroblast repopulation in 5-day wound beds compared with controls; however, we observed a slight decrease in re-epithelialization (fig. S20). These data demonstrate that multiple ligands produced by CD301b$^+$ macrophages activate the proliferation of APs, and not other myofibroblast subsets, during wound healing.

Numbers of CD301b$^+$ macrophages increase in wounds as AP abundance increases (*41*). However, wound beds from aged mice contain fewer CD301b$^+$ cells than those from young controls, and human keloid scars, which have been shown to contain many CD26$^+$ fibroblasts (*54*), are enriched with CD301$^+$ cells (fig. S21). Thus, the interaction between CD301b$^+$ macrophages and mesenchymal cells may provide a therapeutic target for fibrosis-related diseases.

Discussion

Dermal cells, including fibroblasts and adipocytes, support epidermal functions and integrity (*11*, *12*). Although a common embryonic precursor for SMA$^+$ wound bed myofibroblasts exists (*14*, *15*, *20*, *33*, *55*, *56*), the diversity of mesenchymal cells during adult tissue repair is ill defined. In this study, we discovered that after injury, skin wound beds contain tremendous mesenchymal heterogeneity, similar to what is observed during lung fibrosis (*57*). We identified two major classes of SMA$^+$ and Col I$^+$ myofibroblasts that arise from different cellular origins: cells with a cell surface marker profile of APs, and CD29^{High} cells. During tissue repair, a greater percentage of APs express profibrotic cell surface proteins CD26^{High} and CD9, with reduced adipogenic potential. Spatially, these two myofibroblast populations are distinct, with APs evenly distributed within wounds and CD29^{High} cells biased toward superficial, outer regions of wound beds. RNA-seq and functional analysis of these myofibroblast subsets revealed that each subset has distinct transcriptomes with some functional overlap.

With age, the abundance of APs decreases and CD29^{High} cells are more prevalent, as differential gene expression between myofibroblast subsets is reduced. Although myofibroblasts are dynamic after injury in mouse skin, human dermal fibroblasts express profibrotic cell surface proteins in uninjured skin, possibly resulting in stronger fibrotic biasing in humans. These studies illuminate distinct functional subsets of fibrotic cells, providing a stepping stone to develop therapeutic strategies that promote efficient wound healing and treat fibrosis.

Regulation of functional myofibroblast diversity

In this study, we have shown that CD26^{High} myofibroblasts are largely CD34$^+$;CD29$^+$ APs that function as myofibroblasts in regenerating mouse skin. Although previous reports did not observe the same degree of CD34$^+$ and CD29$^+$ colocalization on myofibroblasts (*14*, *16*), these differences likely result from changes in fibroblast surface

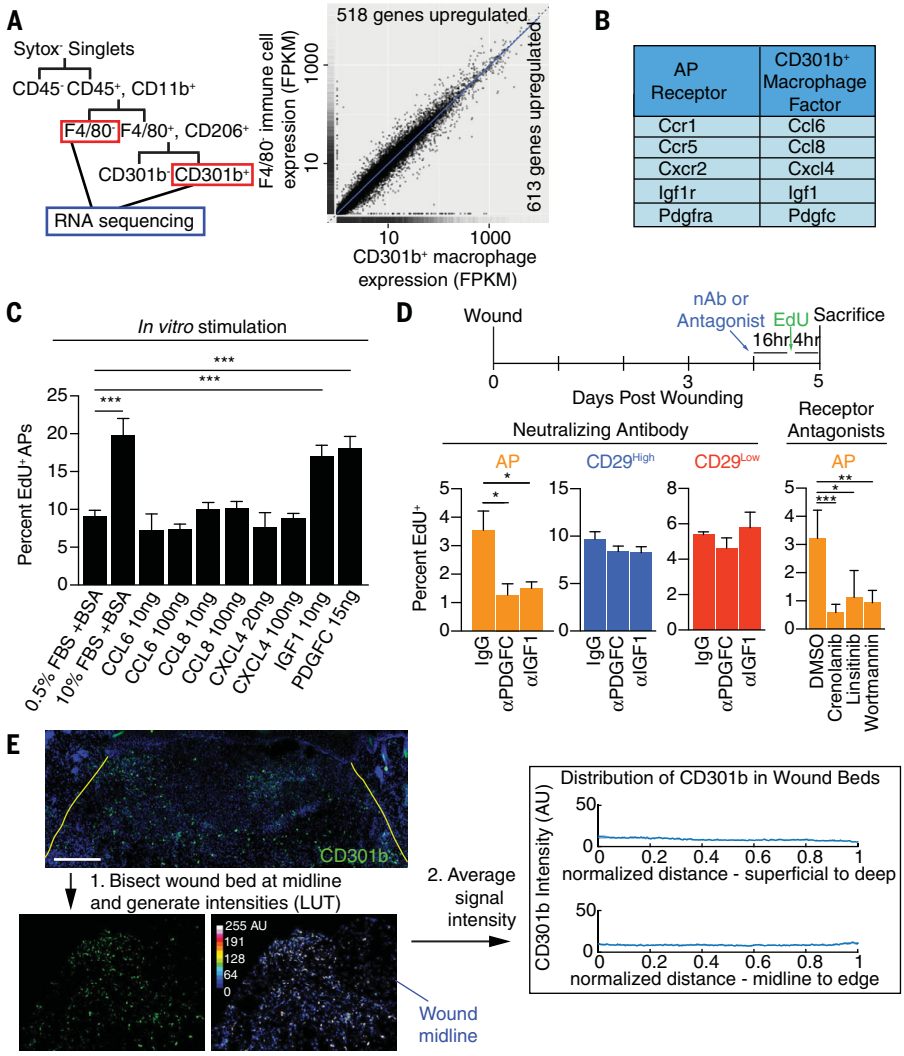


Fig. 6. CD301b$^+$ macrophage–derived ligands activate AP proliferation. (**A**) Cell populations isolated from 5-day wound beds for RNA-seq (left) and FPKM (fragments per kilobase per million) scatter-plotting (right). (**B**) Table of ligands enriched in CD301b$^+$ macrophages that bind to receptors on APs. (**C**) Quantification of AP proliferation after the administration of ligands. Fetal bovine serum (FBS) at 10% is a positive control (n = 5 samples, ***P < 0.001). BSA, bovine serum albumin. (**D**) Quantification of in vivo cellular proliferation after the administration of PDGFC (n = 6 samples, *P = 0.0337)– and IGF1 (n = 6 samples, *P = 0.0436)–neutralizing antibodies (nAbs) or antagonists against PDGFRα (crenolanib) (n = 6 samples, ***P = 0.0001), IGFR1 (linsitinib) (n = 4 samples, *P = 0.01017), or PI 3-kinase (wortmannin) (n = 4 samples, **P = 0.0028). DMSO, dimethyl sulfoxide. (**E**) Pipeline for processing immunostained wound bed sections to infer the distribution of CD301b$^+$ macrophages in day 5 wounds (n = 6 samples). Yellow lines delineate wound edges. Scale bar, 250 μm. Error bars indicate the mean ± SEM.

marker expression associated with different ages and hair follicle stages (*29*). Our data reveal that biased proliferation and plasticity of fibroblast subsets promote myofibroblast heterogeneity in skin wounds. Our lineage tracing data suggest that a combination of proliferation and plasticity supports fibroblast heterogeneity within regenerating skin.

Although multiple signals likely influence myofibroblast heterogeneity, our study highlights the importance of myofibroblast-macrophage interactions and, particularly, of PDGFC and IGF1 in promoting myofibroblast heterogeneity and repair. These data correspond with the function of macrophages in tissue fibrosis (*1*, *58*), the ability of exogenous PDGFC to rescue delayed skin wound healing in diabetic mice (*59*), and the promotion of fibroblast proliferation and repair by IGF1 (*35*, *50*–*53*). In various tissues, macrophages express *Pdgfc* and *Igf1* after injury or under pathological conditions (*60*–*64*). PDGF signaling and IGF signaling cooperate synergistically to promote fibroblast proliferation and enhance wound healing without increasing scarring (*65*–*67*). Treatments aimed at fine-tuning the number of CD301b$^+$ macrophages could be of tremendous clinical value, as reduction in CD301b$^+$ macrophages and CD26-expressing fibroblasts is associated with aging and defective wound healing and keloid scars contain excessive ECM, CD26-expressing fibroblasts, and CD301$^+$ cells. Further understanding of how myofibroblast subsets function and are influenced by the microenvironment during fibrosis and pathologies with irregular ECM homeostasis will allow optimization of treatments for these encumbering diseases.

REFERENCES AND NOTES

1. T. A. Wynn, K. M. Vannella, Macrophages in tissue repair, regeneration, and fibrosis. *Immunity* **44**, 450–462 (2016). doi: 10.1016/j.immuni.2016.02.015; pmid: 26982353
2. S. Ebmeier, V. Horsley, Origin of fibrosing cells in systemic sclerosis. *Curr. Opin. Rheumatol.* **27**, 555–562 (2015). doi: 10.1097/BOR.0000000000000217; pmid: 26352735
3. A. Gosain, L. A. DiPietro, Aging and wound healing. *World J. Surg.* **28**, 321–326 (2004). doi: 10.1007/s00268-003-7397-6; pmid: 14961191
4. C. B. Ballas, J. M. Davidson, Delayed wound healing in aged rats is associated with increased collagen gel remodeling and contraction by skin fibroblasts, not with differences in apoptotic or myofibroblast cell populations. *Wound Repair Regen.* **9**, 223–237 (2001). doi: 10.1046/j.1524-475x.2001.00223.x; pmid: 11472619
5. T. Fujiwara *et al*., Age-associated intracellular superoxide dismutase deficiency potentiates dermal fibroblast dysfunction during wound healing. *Exp. Dermatol.* (2017). doi: 10.1111/exd.13404; pmid: 28677217
6. N. Vedrenne, B. Coulomb, A. Danigo, F. Bonté, A. Desmoulière, The complex dialogue between (myo)fibroblasts and the extracellular matrix during skin repair processes and ageing. *Pathol. Biol. (Paris)* **60**, 20–27 (2012). doi: 10.1016/j.patbio.2011.10.002; pmid: 22099331
7. E. C. Naylor, R. E. B. Watson, M. J. Sherratt, Molecular aspects of skin ageing. *Maturitas* **69**, 249–256 (2011). doi: 10.1016/j.maturitas.2011.04.011; pmid: 21612880
8. E. Kohl, J. Steinbauer, M. Landthaler, R. M. Szeimies, Skin ageing. *J. Eur. Acad. Dermatol. Venereol.* **25**, 873–884 (2011). doi: 10.1111/j.1468-3083.2010.03963.x; pmid: 21261751
9. T. Tabib, C. Morse, T. Wang, W. Chen, R. Lafyatis, SFRP2/DPP4 and FMO1/LSP1 define major fibroblast populations in human skin. *J. Invest. Dermatol.* **138**, 802–810 (2018). doi: 10.1016/j.jid.2017.09.045; pmid: 29080679
10. C. Philippeos *et al*., Spatial and single-cell transcriptional profiling identifies functionally distinct human dermal fibroblast subpopulations. *J. Invest. Dermatol.* **138**, 811–825 (2018). doi: 10.1016/j.jid.2018.01.016; pmid: 29391249
11. R. R. Driskell, F. M. Watt, Understanding fibroblast heterogeneity in the skin. *Trends Cell Biol.* **25**, 92–99 (2015). doi: 10.1016/j.tcb.2014.10.001; pmid: 25455110
12. J. M. Sorrell, A. I. Caplan, Fibroblast heterogeneity: More than skin deep. *J. Cell Sci.* **117**, 667–675 (2004). pmid: 14754903
13. R. G. Marangoni *et al*., Myofibroblasts in murine cutaneous fibrosis originate from adiponectin-positive intradermal progenitors. *Arthritis Rheumatol.* **67**, 1062–1073 (2015). doi: 10.1002/art.38990; pmid: 25504959
14. Y. Rinkevich *et al*., Identification and isolation of a dermal lineage with intrinsic fibrogenic potential. *Science* **348**, aaa2151 (2015). doi: 10.1126/science.aaa2151; pmid: 25883361
15. R. R. Driskell *et al*., Distinct fibroblast lineages determine dermal architecture in skin development and repair. *Nature* **504**, 277–281 (2013). doi: 10.1038/nature12783; pmid: 24336287
16. D. Jiang *et al*., Two succeeding fibroblastic lineages drive dermal development and the transition from regeneration to scarring. *Nat. Cell Biol.* **20**, 422–431 (2018). doi: 10.1038/s41556-018-0073-8; pmid: 29593327
17. I. A. Darby, B. Laverdet, F. Bonté, A. Desmoulière, Fibroblasts and myofibroblasts in wound healing. *Clin. Cosmet. Investig. Dermatol.* **7**, 301–311 (2014). pmid: 25395868
18. B. Hinz *et al*., Recent developments in myofibroblast biology: Paradigms for connective tissue remodeling. *Am. J. Pathol.* **180**, 1340–1355 (2012). doi: 10.1016/j.ajpath.2012.02.004; pmid: 22387320
19. T. A. Wynn, T. R. Ramalingam, Mechanisms of fibrosis: Therapeutic translation for fibrotic disease. *Nat. Med.* **18**, 1028–1040 (2012). doi: 10.1038/nm.2807; pmid: 22772564
20. M. V. Plikus *et al*., Regeneration of fat cells from myofibroblasts during wound healing. *Science* **355**, 748–752 (2017). doi: 10.1126/science.aai8792; pmid: 28059714
21. T. Iwayama *et al*., PDGFRα signaling drives adipose tissue fibrosis by targeting progenitor cell plasticity. *Genes Dev.* **29**, 1106–1119 (2015). doi: 10.1101/gad.260554.115; pmid: 26019175
22. G. Marcelin *et al*., A PDGFRα-mediated switch toward CD9^{high} adipocyte progenitors controls obesity-induced adipose tissue fibrosis. *Cell Metab.* **25**, 673–685 (2017). doi: 10.1016/j.cmet.2017.01.010; pmid: 28215843
23. R. Higashiyama *et al*., Differential contribution of dermal resident and bone marrow-derived cells to collagen production during wound healing and fibrogenesis in mice. *J. Invest. Dermatol.* **131**, 529–536 (2011). doi: 10.1038/jid.2010.314; pmid: 20962852
24. O. Kanisicak *et al*., Genetic lineage tracing defines myofibroblast origin and function in the injured heart. *Nat. Commun.* **7**, 12260 (2016). doi: 10.1038/ncomms12260; pmid: 27447449
25. R. Berry, M. S. Rodeheffer, Characterization of the adipocyte cellular lineage in vivo. *Nat. Cell Biol.* **15**, 302–308 (2013). doi: 10.1038/ncb2696; pmid: 23434825
26. G. C. Rivera-Gonzalez *et al*., Skin adipocyte stem cell self-renewal is regulated by a PDGFA/AKT-signaling axis. *Cell Stem Cell* **19**, 738–751 (2016). pmid: 27746098
27. C. D. Church, R. Berry, M. S. Rodeheffer, Isolation and study of adipocyte precursors. *Methods Enzymol.* **537**, 31–46 (2014). doi: 10.1016/B978-0-12-411619-1.00003-3; pmid: 24480340
28. M. S. Rodeheffer, K. Birsoy, J. M. Friedman, Identification of white adipocyte progenitor cells in vivo. *Cell* **135**, 240–249 (2008). doi: 10.1016/j.cell.2008.09.036; pmid: 18835024
29. B. M. Lichtenberger, M. Mastrogiannaki, F. M. Watt, Epidermal β-catenin activation remodels the dermis via paracrine signalling to distinct fibroblast lineages. *Nat. Commun.* **7**, 10537 (2016). doi: 10.1038/ncomms10537; pmid: 26837596
30. T. Ostendorf, F. Eitner, J. Floege, The PDGF family in renal fibrosis. *Pediatr. Nephrol.* **27**, 1041–1050 (2012). doi: 10.1007/s00467-011-1892-z; pmid: 21597969
31. L. J. Reigstad, J. E. Varhaug, J. R. Lillehaug, Structural and functional specificities of PDGF-C and PDGF-D, the novel members of the platelet-derived growth factors family. *FEBS J.* **272**, 5723–5741 (2005). doi: 10.1111/j.1742-4658.2005.04989.x; pmid: 16279938
32. J. C. Bonner, Regulation of PDGF and its receptors in fibrotic diseases. *Cytokine Growth Factor Rev.* **15**, 255–273 (2004). doi: 10.1016/j.cytogfr.2004.03.006; pmid: 15207816
33. S. Dulauroy, S. E. Di Carlo, F. Langa, G. Eberl, L. Peduto, Lineage tracing and genetic ablation of ADAM12$^+$ perivascular cells identify a major source of profibrotic cells during acute tissue injury. *Nat. Med.* **18**, 1262–1270 (2012). doi: 10.1038/nm.2848; pmid: 22842476
34. S. Paquet-Fifield *et al*., A role for pericytes as microenvironmental regulators of human skin tissue regeneration. *J. Clin. Invest.* **119**, 2795–2806 (2009). pmid: 19652362
35. S. Werner, R. Grose, Regulation of wound healing by growth factors and cytokines. *Physiol. Rev.* **83**, 835–870 (2003). doi: 10.1152/physrev.2003.83.3.835; pmid: 12843410
36. J. Ding, E. E. Tredget, The role of chemokines in fibrotic wound healing. *Adv. Wound Care* **4**, 673–686 (2015). doi: 10.1089/wound.2014.0550; pmid: 26543681
37. M. Xue, C. J. Jackson, Extracellular matrix reorganization during wound healing and its impact on abnormal scarring. *Adv. Wound Care* **4**, 119–136 (2015). doi: 10.1089/wound.2013.0485; pmid: 25785236
38. I. Goren *et al*., A transgenic mouse model of inducible macrophage depletion: Effects of diphtheria toxin-driven lysozyme M-specific cell lineage ablation on wound inflammatory, angiogenic, and contractive processes. *Am. J. Pathol.* **175**, 132–147 (2009). doi: 10.2353/ajpath.2009.081002; pmid: 19528348
39. T. Lucas *et al*., Differential roles of macrophages in diverse phases of skin repair. *J. Immunol.* **184**, 3964–3977 (2010). doi: 10.4049/jimmunol.0903356; pmid: 20176743
40. R. Mirza, L. A. DiPietro, T. J. Koh, Selective and specific macrophage ablation is detrimental to wound healing in mice. *Am. J. Pathol.* **175**, 2454–2462 (2009). doi: 10.2353/ajpath.2009.090248; pmid: 19850888
41. B. Shook, E. Xiao, Y. Kumamoto, A. Iwasaki, V. Horsley, CD301b+ macrophages are essential for effective skin wound healing. *J. Invest. Dermatol.* **136**, 1885–1891 (2016). doi: 10.1016/j.jid.2016.05.107; pmid: 27287183
42. B. E. Clausen, C. Burkhardt, W. Reith, R. Renkawitz, I. Förster, Conditional gene targeting in macrophages and granulocytes using LysMcre mice. *Transgenic Res.* **8**, 265–277 (1999). doi: 10.1023/A:1008942828960; pmid: 10621974
43. T. Buch *et al*., A Cre-inducible diphtheria toxin receptor mediates cell lineage ablation after toxin administration. *Nat. Methods* **2**, 419–426 (2005). doi: 10.1038/nmeth762; pmid: 15908920
44. J. M. Daley, S. K. Brancato, A. A. Thomay, J. S. Reichner, J. E. Albina, The phenotype of murine wound macrophages. *J. Leukocyte Biol.* **87**, 59–67 (2010). doi: 10.1189/jlb.0409236; pmid: 20052800
45. R. Mirza, T. J. Koh, Dysregulation of monocyte/macrophage phenotype in wounds of diabetic mice. *Cytokine* **56**, 256–264 (2011). doi: 10.1016/j.cyto.2011.06.016; pmid: 21803601
46. Y. Kumamoto *et al*., CD301b$^+$ dermal dendritic cells drive T helper 2 cell-mediated immunity. *Immunity* **39**, 733–743 (2013). doi: 10.1016/j.immuni.2013.08.029; pmid: 24076051
47. B. Yang *et al*., IL-27 facilitates skin wound healing through induction of epidermal proliferation and host defense. *J. Invest. Dermatol.* **137**, 1166–1175 (2017). doi: 10.1016/j.jid.2017.01.010; pmid: 28132857
48. M. S. Hu *et al*., Delivery of monocyte lineage cells in a biomimetic scaffold enhances tissue repair. *JCI Insight* **2**, e96260 (2017). doi: 10.1172/jci.insight.96260; pmid: 28978794
49. S. Werner, T. Krieg, H. Smola, Keratinocyte-fibroblast interactions in wound healing. *J. Invest. Dermatol.* **127**, 998–1008 (2007). doi: 10.1038/sj.jid.5700786; pmid: 17435785
50. Y. Ando, P. J. Jensen, Epidermal growth factor and insulin-like growth factor I enhance keratinocyte migration. *J. Invest. Dermatol.* **100**, 633–639 (1993). doi: 10.1111/1523-1747.ep12472297; pmid: 8491986
51. I. Haase, R. Evans, R. Pofahl, F. M. Watt, Regulation of keratinocyte shape, migration and wound epithelialization by IGF-1- and EGF-dependent signalling pathways. *J. Cell Sci.* **116**, 3227–3238 (2003). pmid: 12829742
52. J. G. Simmons, J. B. Pucilowska, T. O. Keku, P. K. Lund, IGF-I and TGF-beta1 have distinct effects on phenotype and proliferation of intestinal fibroblasts. *Am. J. Physiol. Gastrointest. Liver Physiol.* **283**, G809–G818 (2002). doi: 10.1152/ajpgi.00057.2002; pmid: 12181198
53. S. M. Rudman, M. P. Philpott, G. A. Thomas, T. Kealey, The role of IGF-I in human skin and its appendages: Morphogen as well as mitogen? *J. Invest. Dermatol.* **109**, 770–777 (1997). doi: 10.1111/1523-1747.ep12340934; pmid: 9406819
54. Y. Xin *et al*., Expansion of CD26 positive fibroblast population promotes keloid progression. *Exp. Cell Res.* **356**, 104–113 (2017). pmid: 28454879
55. H. Suga *et al*., Tracking the elusive fibrocyte: Identification and characterization of collagen-producing hematopoietic lineage cells during murine wound healing. *Stem Cells* **32**, 1347–1360 (2014). doi: 10.1002/stem.1648; pmid: 24446236
56. B. Hinz, Formation and function of the myofibroblast during tissue repair. *J. Invest. Dermatol.* **127**, 526–537 (2007). doi: 10.1038/sj.jid.5700613; pmid: 17299435
57. J. R. Rock *et al*., Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to

mesenchymal transition. *Proc. Natl. Acad. Sci. U.S.A.* **108**, E1475–E1483 (2011). doi: 10.1073/pnas.1117988108; pmid: 22123957
58. K. M. Vannella, T. A. Wynn, Mechanisms of organ injury and repair by macrophages. *Annu. Rev. Physiol.* **79**, 593–617 (2017). doi: 10.1146/annurev-physiol-022516-034356; pmid: 27959618
59. D. G. Gilbertson *et al*., Platelet-derived growth factor C (PDGF-C), a novel growth factor that binds to PDGF α and β receptor. *J. Biol. Chem.* **276**, 27406–27414 (2001). doi: 10.1074/jbc.M101056200; pmid: 11297552
60. A. Triantafyllopoulou *et al*., Proliferative lesions and metalloproteinase activity in murine lupus nephritis mediated by type I interferons and macrophages. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 3012–3017 (2010). doi: 10.1073/pnas.0914902107; pmid: 20133703
61. H. R. Chang, H. J. Kim, X. Xu, A. W. Ferrante Jr., Macrophage and adipocyte IGF1 maintain adipose tissue homeostasis during metabolic stresses. *Obesity (Silver Spring)* **24**, 172–183 (2016). doi: 10.1002/oby.21354; pmid: 26663512
62. P. Ramachandran *et al*., Differential Ly-6C expression identifies the recruited macrophage phenotype, which orchestrates the regression of murine liver fibrosis. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E3186–E3195 (2012). doi: 10.1073/pnas.1119964109; pmid: 23100531
63. J. Tonkin *et al*., Monocyte/macrophage-derived IGF-1 orchestrates murine skeletal muscle regeneration and modulates autocrine polarization. *Mol. Ther.* **23**, 1189–1200 (2015). doi: 10.1038/mt.2015.66; pmid: 25896247
64. D. Pohlers, R. Huber, B. Ukena, R. W. Kinne, Expression of platelet-derived growth factors C and D in the synovial membrane of patients with rheumatoid arthritis and osteoarthritis. *Arthritis Rheum.* **54**, 788–794 (2006). doi: 10.1002/art.21670; pmid: 16508943
65. S. E. Lynch, J. C. Nixon, R. B. Colvin, H. N. Antoniades, Role of platelet-derived growth factor in wound healing: Synergistic effects with other growth factors. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7696–7700 (1987). doi: 10.1073/pnas.84.21.7696; pmid: 3499612
66. S. E. Lynch, R. B. Colvin, H. N. Antoniades, Growth factors in wound healing. Single and synergistic effects on partial thickness porcine skin wounds. *J. Clin. Invest.* **84**, 640–646 (1989). doi: 10.1172/JCI114210; pmid: 2788174
67. S. E. Lynch *et al*., The effects of short-term application of a combination of platelet-derived and insulin-like growth factors on periodontal wound healing. *J. Periodontol.* **62**, 458–467 (1991). doi: 10.1902/jop.1991.62.7.458; pmid: 1920013

ACKNOWLEDGMENTS

We thank members of the Horsley lab, R. Atit, and A. MacLeod for critical reading of the manuscript. We thank B. Hogan and C. Barkauskas for sharing *Pdgfra*CreER mice and C. Wolfrum for sharing *Dlk1*CreER mice before the initial mouse line publications and R. Atit for *En1*Cre;*Rosa26*-LSL-tdTomato mice. **Funding:** This work was supported in part by NIH grants to V.H. from NIAMS (AR060295 and AR069550) and NIA through pilot project grants from the Claude D. Pepper Older Americans Independence Center at Yale (NIA P30AG21342) awarded to V.H. This research was supported in part by a New Investigator award to H.C.H. from the Yale Department of Surgery Ohse Fund. B.A.S. is a New York Stem Cell Foundation–Druckenmiller Fellow. This research was supported by the New York Stem Cell Foundation. **Author contributions:** B.A.S. conceived, designed, and oversaw the experiments with suggestions from V.H., J.L.A., K.M.-J., D.A.C., and H.C.H., R.R.W, G.C.R.-G., and A.R.M.-R. performed experiments and analyzed the data. B.C.D., E.S.-G., K.D.A., R.K.Z., and V.L. assisted with collection and analysis of data. F.L.-G. analyzed the RNA-seq data. B.A.S. and V.H. wrote the manuscript. **Competing interests:** The authors have no competing interests. **Data and materials availability:** RNA-seq data are available at the Gene Expression Omnibus (GEO) under accession numbers GSE105788, GSE105789, and GSE105790.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/eaar2971/suppl/DC1
Materials and Methods
Figs. S1 to S21
Tables S1 to S3
References (*68*–*78*)

23 October 2017; resubmitted 20 July 2018
Accepted 4 October 2018
10.1126/science.aar2971

RESEARCH ARTICLE SUMMARY

DEVELOPMENTAL BIOLOGY

Maternal Huluwa dictates the embryonic body axis through β-catenin in vertebrates

Lu Yan*, Jing Chen*, Xuechen Zhu*, Jiawei Sun, Xiaotong Wu, Weimin Shen, Weiying Zhang, Qinghua Tao†, Anming Meng†

INTRODUCTION: The formation of the body axis in vertebrate embryos relies on the formation of the dorsal organizer at the onset of gastrulation, which is called the Spemann-Mangold organizer in frogs and the embryonic shield in zebrafish. Previous studies have indicated that nuclear β-catenin signaling is essential for induction of the dorsal organizer.

RATIONALE: Maternal Wnt ligand/receptor signaling has been proposed for stabilization and nuclear transportation of cytoplasmic β-catenin during early blastulation for the dorsal organizer formation. It is unknown whether β-catenin signaling is activated for organizer formation by Wnt-independent mechanisms.

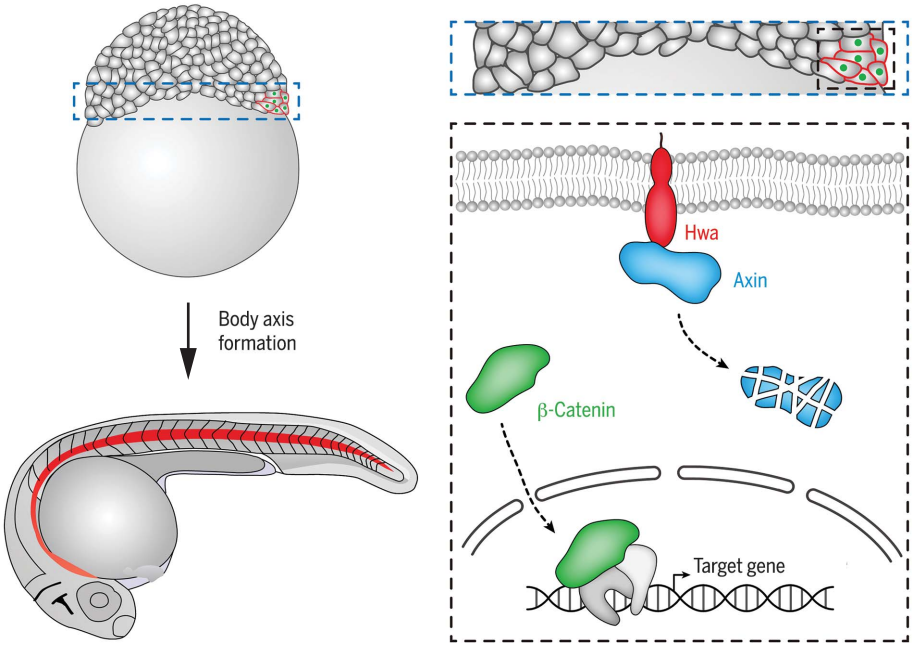
RESULTS: In a spontaneous maternal-effect mutant line of zebrafish, none of the maternal mutant embryos formed the dorsal organizer at the shield stage, nor did they have a head or other dorsoanterior tissues at later stages of development. Through positional cloning and candidate gene testing, we found the mutant gene to be a previously uncharacterized locus that we designate *huluwa* (*hwa*). The defects of M*hwa* mutants could be fully rescued by overexpression of wild-type *hwa* mRNA. Using antisense oligodeoxynucleotides or morpholino, we found that maternal depletion of *Xenopus hwa* transcript in oocytes also causes loss of the body axis and dorsal tissues in the derived embryos. Zebrafish Hwa protein consists of 294 amino acids comprising a 23–amino acid extracellular domain, a 23–amino acid transmembrane domain, and a 248–amino acid intracellular domain. Immunostaining revealed that Hwa protein is located on the plasma membrane of blastomeres only in a region during zebrafish early blastulation in which β-catenin is translocated into nuclei. By performing rescue experiments in *hwa* mutant embryos using different *hwa* mutant mRNAs, we found that the transmembrane domain and the intracellular ^{164}VPPNSP^{169} and ^{184}SLRRSST^{190} motifs are essential for Hwa activity. Overexpression of β-*catenin* efficiently rescued the defects in zebrafish M*hwa* mutants. Ectopic expression of *hwa* mRNA efficiently induced a secondary axis in zebrafish and *Xenopus* embryos, which absolutely required β-catenin. Although mammalian homologs of zebrafish Hwa have not yet been identified, we found that transfection of zebrafish Hwa in human HEK293T cells enhances β-catenin signaling. Zebrafish Hwa could directly bind to human Axin protein, and transfection of zebrafish *hwa* into mammalian cells promoted Axin degradation with participation of tankyrases. Consistent with in vitro data, levels of Axin1 and Axin2 proteins in M*hwa* mutant embryos were up-regulated, and overexpression of the dominant negative form *Axin1ΔRGS* or *Axin1ΔDIX* of Axin1 in M*hwa* mutant embryos could rescue the dorsal organizer and the body axis. Therefore, we conclude that maternal Hwa is absolutely required for the formation of the dorsal organizer and the body axis by protecting β-catenin from Axin-mediated degradation in vertebrate embryos.

Overexpression of the Wnt antagonist gene *DKK1*, the dominant negative Wnt8a form *dnwnt8*, or the dominant negative LRP5 form *LRP5ΔC* in zebrafish embryos neither disrupted the dorsal organizer nor blocked the organizer- and body axis–inducing activity of ectopic *hwa*; treatment with the Wnt inhibitor Wnt-C59 also had no such effects. *Xenopus* embryos derived from oocytes depleted of *lrp6* or *wnt11b* transcript could form the full body axis after being injected with *hwa* mRNA at the four-cell stage.

CONCLUSION: Maternal Hwa protein in vertebrate embryos is essential for the dorsal organizer and body axis formation, which activates β-catenin signaling during early blastulation in a Wnt ligand/receptor-independent fashion. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat1045



Huluwa (Hwa) is essential for the organizer and body axis formation. Hwa protein is located on the plasma membrane of the prospective dorsal blastomeres in the zebrafish blastula. Hwa binds to and promotes the degradation of Axin in a way independent of Wnt ligand/receptor signaling, resulting in stabilization and nuclear translocation of β-catenin for activating organizer-specific target gene expression. The notochord (red) is an organizer-derived tissue.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: mengam@mail.tsinghua.edu.cn (A.M.); qhtaolab@mail.tsinghua.edu.cn (Q.T.)
Cite this article as L. Yan *et al.*, *Science* **362**, eaat1045 (2018).
DOI: 10.1126/science.aat1045

RESEARCH ARTICLE

DEVELOPMENTAL BIOLOGY

Maternal Huluwa dictates the embryonic body axis through β-catenin in vertebrates

Lu Yan[1]*, Jing Chen[1]*, Xuechen Zhu[2]*, Jiawei Sun[1], Xiaotong Wu[1], Weimin Shen[1], Weiying Zhang[1], Qinghua Tao[2]†, Anming Meng[1]†

The vertebrate body is formed by cell movements and shape change during embryogenesis. It remains undetermined which maternal signals govern the formation of the dorsal organizer and the body axis. We found that maternal depletion of *huluwa*, a previously unnamed gene, causes loss of the dorsal organizer, the head, and the body axis in zebrafish and *Xenopus* embryos. Huluwa protein is found on the plasma membrane of blastomeres in the future dorsal region in early zebrafish blastulas. Huluwa has strong dorsalizing and secondary axis–inducing activities, which require β-catenin but can function independent of Wnt ligand/receptor signaling. Mechanistically, Huluwa binds to and promotes the tankyrase-mediated degradation of Axin. Therefore, maternal Huluwa is an essential determinant of the dorsal organizer and body axis in vertebrate embryos.

The form of the vertebrate embryo changes markedly through extensive morphogenetic movements during development. Before fertilization, an animal-vegetal axis forms within the oocyte, with the maternal nucleus closer to the animal end and (in some species) most of the yolk in the vegetal hemisphere. After fertilization in *Xenopus* and zebrafish, maternal dorsal determinants in the vegetal pole are transported to the future dorsal side through microtubule bundles, to subsequently stabilize β-catenin and allow its nuclear translocation. β-Catenin is the key effector of canonical Wnt signals in the future dorsal blastomeres (*1*) and it induces the formation of the organizer, which acts as a signaling center for correct embryonic patterning and gastrulation cell movements (*2*–*7*). The dorsal organizer is fully formed at the onset of gastrulation, which is called the Spemann-Mangold organizer in frogs and the embryonic shield in zebrafish. Subsequently, two major body axes form: the rostral-caudal (or anteroposterior) axis and the dorsoventral axis.

Loss-of-function studies in frogs, fish, and mice demonstrate that β-catenin plays an essential role in the formation of the dorsal organizer and the anteroposterior axis (*8*–*13*). In the absence of Wnt signaling, cytoplasmic β-catenin forms a complex with the β-catenin destruction complex components Axin, Gsk3β, APC, and CK1 and is subjected to degradation (*14*). Upon binding of Wnt ligand to the Wnt receptors Frizzled and Lrp5/6, Dishevelled proteins participate in a signal transduction pathway to release cytoplasmic β-catenin, allowing its nuclear translocation and activation of target gene transcription. However, accumulating evidence suggests that β-catenin can also be released in a Wnt ligand/receptor-independent fashion. For instance, epidermal growth factor, hepatocyte growth factor, insulin-like growth factor, and platelet-derived growth factor have been found to activate β-catenin via the kinases Akt, Erk, or C-Abl, resulting in dissociation of β-catenin from the E-cadherin adherens complex, inactivation of Gsk3β, or displacement of Axin and Gsk3β (*15*–*20*).

It is widely believed that the role of β-catenin during embryonic development is dependent on Wnt ligand/receptor signaling (*15*–*20*). One supporting observation is that a maternal Dishevelled protein appears to be actively transported to the future dorsal side in early *Xenopus* embryos (*21*); however, Dishevelled proteins can have other functions in addition to their implication in Wnt/β-catenin signaling (*22*). In the zebrafish *wnt8* zygotic mutant *Df(LG14)wnt8*^{w8}, the embryonic shield is expanded at the shield stage and

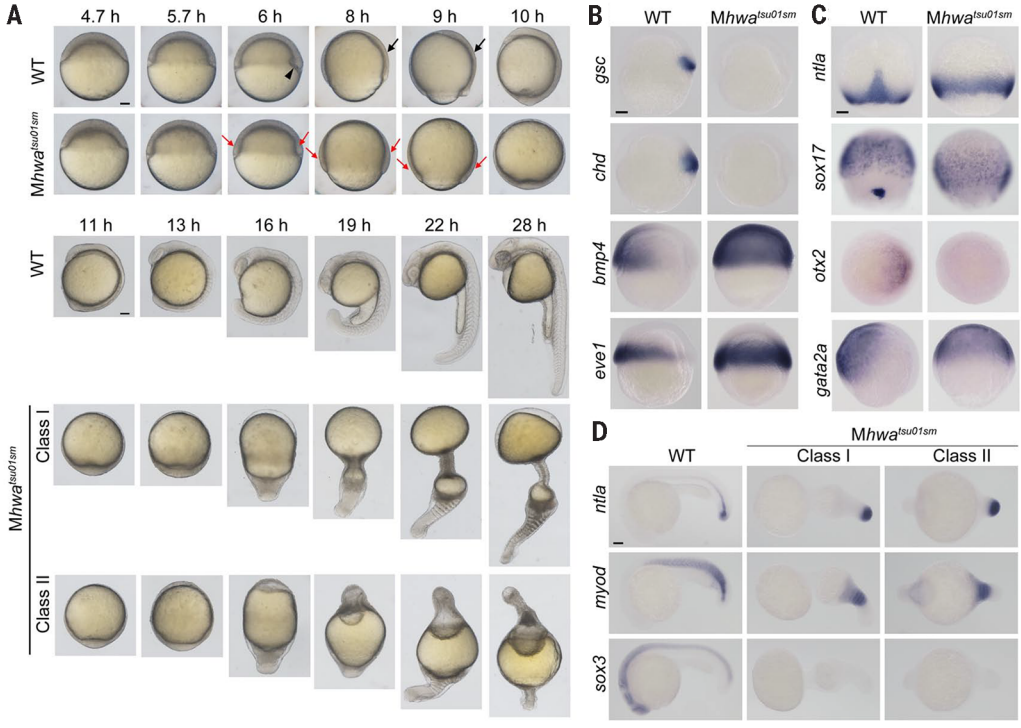


Fig. 1. Mhwa^{tsu01sm} mutant phenotype. (**A**) Morphogenesis of live wild-type (WT) and mutant embryos. The black arrowhead and arrows indicate the dorsal organizer and the thickening body axis, respectively; the red arrows in mutants indicate the thicker blastodermal marginal zone. See also movie S1. (**B** and **C**) Germ layer marker expression detected by WISH during gastrulation. Orientation: *ntla* and *sox17*, dorsal views with animal pole to the top at 75% epiboly stage; others, lateral views (*gsc*, *chd*, *bmp4*, *eve1*, and *gata2a*) or animal-pole view (*otx2*) with dorsal to the right. (**D**) Marker expression at 24 hpf. Lateral views with anterior to the left. Scale bars, 100 μm.

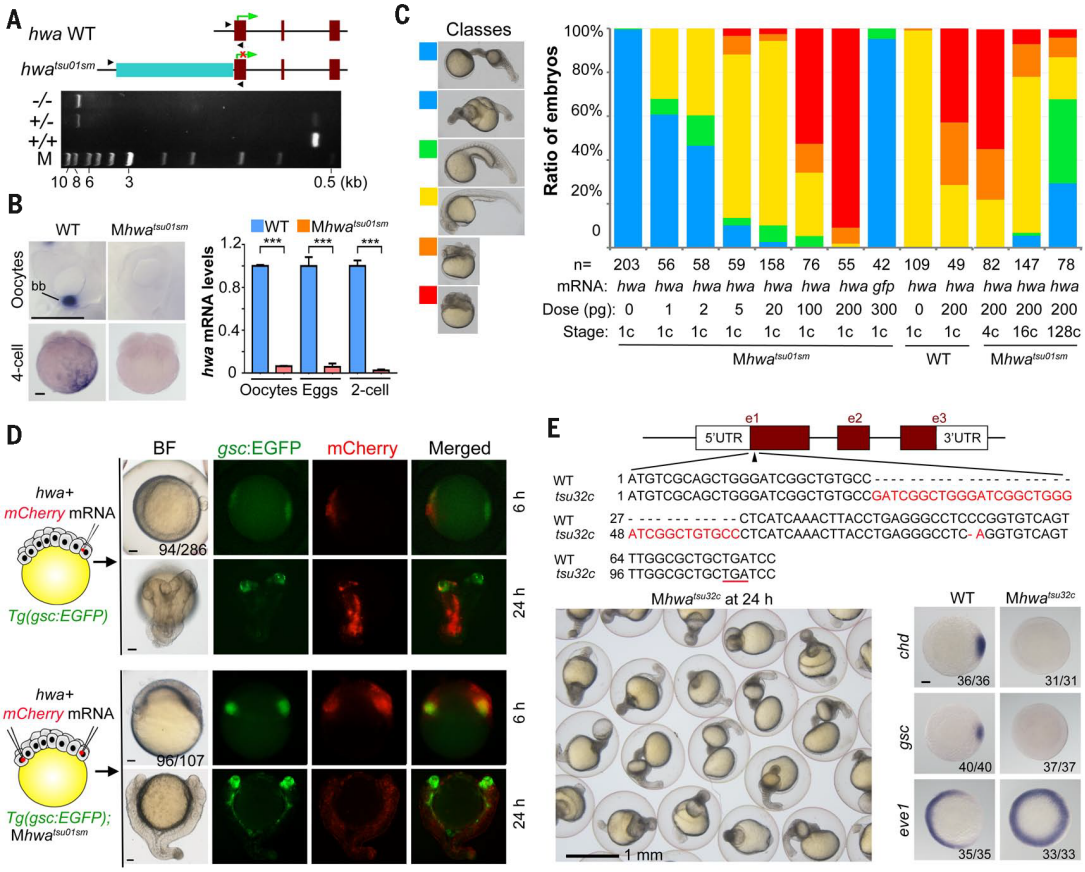


Fig. 2. Identification and verification of *tsu01sm* mutant gene. (**A**) Top: Illustration of the zebrafish *hwa* wild-type and *tsu01sm* mutant alleles. Arrowheads indicate the positions of PCR primers. The green arrow indicates the translation start site. Bottom: Electrophoretic result of PCR products in wild-type, heterozygous, and homozygous mutant embryos. M, molecular weight markers. (**B**) *hwa* mRNA detection by WISH (left) and by qRT-PCR (right). bb, Balbiani body in stage I oocyte. For qRT-PCR, oocytes are a pool of stage I–III oocytes and eggs are squeezed from females. The relative *hwa* mRNA levels are averages (±SEM) from three independent experiments, normalized to *eif4g2a* levels. ***$P < 0.001$. Scale bars, 100 μm. (**C**) M*hwa*^{tsu01sm} mutant embryos form the body axis or are dorsalized by *hwa* overexpression. Left: Morphology of classified embryos at 24 hpf. Right: Ratios of embryos in different classes. (**D**) Induction of secondary body axis by ectopic *hwa* mRNA overexpression. One blastomere of 32-cell stage *Tg(gsc:EGFP)* transgenic embryos or two opposite blastomeres of 32-cell stage *Tg(gsc:EGFP);*M*hwa*^{tsu01sm} mutant embryos were injected with 50 pg of *hwa* and 50 pg of *mCherry* mRNAs, as illustrated at left. Embryos were observed dorsally (top) or laterally (bottom) at 6 hpf under a fluorescence dissection microscope, and those with two dorsal organizers (EGFP-positive) were observed again at 24 hpf. The ratio of embryos with double organizers is indicated. Scale bars, 100 μm. (**E**) Generation of a *hwa* mutant allele by Cas9 knockout. Top: Illustration of the mutant allele with an underlined premature stop codon. Bottom left: A group of M*hwa*^{tsu32c} mutant embryos at 24 hpf. Bottom right: Alteration of the dorsal markers *gsc* and *chd* and the ventral marker *eve1* at shield stage; scale bar, 100 μm.

posterior/ventral tissues are lost at later stages, suggesting a role for zygotic Wnt/β-catenin signaling in restricting the organizer size and dorsal development (*23*, *24*). The role of *Xwnt8* in constraining dorsal cell fate during late blastulation has also been reported in *Xenopus* (*25*). Depletion of zygotic *Wnt3A* or *Lrp5* in mouse embryos results in truncation of the trunk, supporting an important role of Wnt signaling in the development of posterior tissues (*26*–*28*). In *Xenopus*, the use of antisense oligodeoxynucleotides to deplete *wnt11b* or *lrp6* mRNA in oocytes leads to delayed gastrulation and the loss of axial structures (*29*, *30*), which could be ascribed to decreased cytoplasmic β-catenin levels during oocyte maturation as opposed to after fertilization. It is unclear whether β-catenin is stabilized for the organizer and body axis formation by maternal Wnt ligand/receptor signaling from fertilization to early blastulation stages.

Here, we identified a maternal-effect mutant line, *tsu01sm*, that we have named *huluwa* (*hwa*). Unlike previously known β-catenin–related, maternal-effect mutants showing variable degrees of ventralization in zebrafish (*10*, *31*–*37*), all *huluwa* maternal mutant embryos lack the dorsal organizer and fail to form the head and other dorsoanterior tissues. We demonstrate that *huluwa* encodes a transmembrane protein and acts as an essential determinant of the vertebrate body axis formation by activating β-catenin in a Wnt ligand/receptor-independent fashion.

M*hwa* mutant embryos lack a body axis

The maternal-effect *tsu01sm* mutants were found during normal breeding of Tübingen fish and were then used to establish mapping families and preservation families. As described below, some mutant embryos look like calabashes (bottle gourds). We then named the mutant line *huluwa*, the Chinese equivalent of calabash boy. (In the Chinese animation TV series *Calabash Brothers*, seven calabash brothers each have a special power to defeat monsters.) *hwa*^{tsu01sm} zygotic mutants (*Zhwa*$^{-/-}$) from *hwa*^{tsu01sm} heterozygous fish crosses undergo normal development and grow to adulthood. When *Zhwa* mutant adult females mate with wild-type or heterozygous male fish, all resultant embryos (M*hwa* or M*Zhwa*) develop normally during cleavage and blastula stages but lack the embryonic shield at the shield stage (Fig. 1A). The mutant phenotype is stable among batches and generations. Through morphological assessment at 24 hours post-fertilization (hpf), these embryos can be categorized into two classes: class I, calabash-like shape with the yolk splitting into two parts and a tail at the vegetal pole; class II, onion-like shape with one yolk protrusion at the original animal pole and a tail-like protrusion at the vegetal pole (Fig. 1A and movie S1). Both classes of embryos lack a thick body axis on the yolk ball.

As revealed by whole-mount in situ hybridization (WISH), the dorsal organizer markers *gsc*

[1]Laboratory of Molecular Developmental Biology, State Key Laboratory of Membrane Biology, Tsinghua-Peking Center for Life Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China. [2]MOE Key Laboratory of Protein Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China.
*These authors contributed equally to this work.
†Corresponding author. Email: mengam@mail.tsinghua.edu.cn (A.M.); qhtaolab@mail.tsinghua.edu.cn (Q.T.)

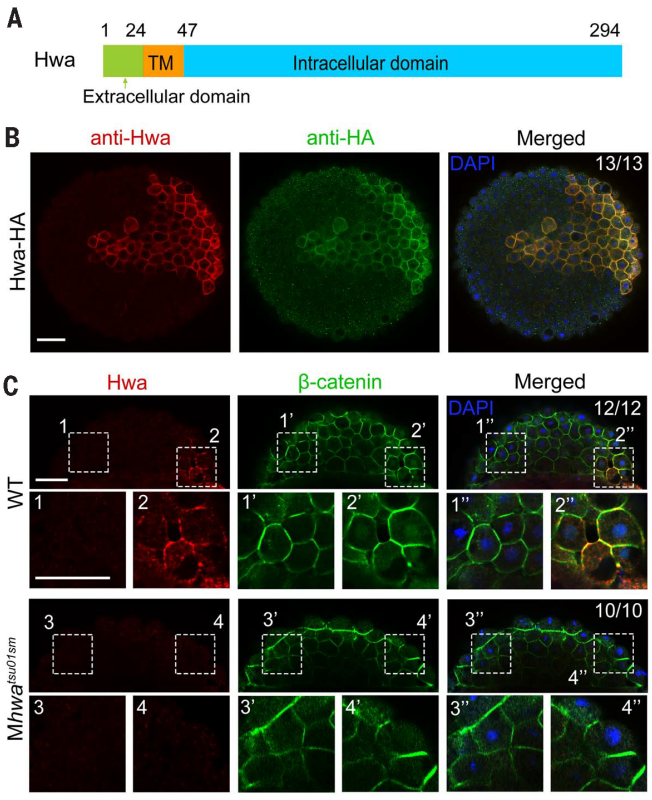


Fig. 3. Hwa is localized on the plasma membrane of dorsal blastomeres. (**A**) Illustration of zebrafish Hwa protein structure. Positions of residues at the start of the domain are indicated except the last residue position. (**B**) Detection of exogenous Hwa-HA by immunostaining using anti-Hwa antibody: 50 pg of *hwa-HA* mRNA was injected into one blastomere of eight-cell stage M*hwa*^{tsu01sm} mutant embryos, and the embryos were collected at 4 hpf for immunostaining with anti-HA and anti-Hwa antibodies. Nuclei were also stained with 4′,6-diamidino-2-phenylindole (DAPI). Embryos were imaged by confocal microscopy in animal pole view. Note that exogenous Hwa-HA is enriched on the plasma membrane. Scale bar, 50 μm. (**C**) Location of endogenous Hwa protein. Zebrafish wild-type and M*hwa*^{tsu01sm} mutant embryos were collected at the 512-cell stage for coimmunostaining of endogenous Hwa and β-catenin. Embryos were imaged by confocal microscopy and are shown in lateral view with the animal pole to the top. In all 12 wild-type embryos with β-catenin in nuclei of the dorsal blastomeres, endogenous Hwa protein is enriched on the plasma membrane in the same region (top panel). In M*hwa* mutants (bottom panel), Hwa signal is absent on the plasma membrane and no blastomeres have nuclear β-catenin. For each embryo, two numbered areas are enlarged for better view. The ratio of embryos with the representative pattern is indicated. Scale bars, 50 μm.

and *chd* are eliminated in M*hwa*^{tsu01sm} mutants at the shield stage, whereas the ventral markers *bmp4* and *eve1* are expanded (Fig. 1B), indicative of embryonic ventralization. The mesodermal marker *ntla* and the endodermal marker *sox17* are expressed in mutant embryos at the shield stage, but expression of *ntla* and *myod* (a myogenic marker) is detected only in the tail-like region at 24 hpf (Fig. 1, C and D). Mutant embryos do not express the anterior neuroectoderm marker *otx2* at the shield stage, nor do they express the central nervous system marker *sox3* at 24 hpf (Fig. 1, C and D). Instead, the epidermis marker *gata2a* is expanded throughout the ectoderm in mutants at the shield stage. These data suggest that in the absence of maternal *hwa*, the mesendodermal fate is specified and the ectoderm completely transforms into the epidermis fate.

During gastrulation in wild-type embryos, marginal cells committed to the mesendodermal fate internalize, and the internalized cells in the lateral and ventral margin migrate dorsoanteriorly to form a thick body axis along the dorsal midline. A blastodermal margin zone close to the margin in M*hwa* mutants (Fig. 1A, red arrows) is thicker than the other blastodermal region during gastrulation, which suggests that blastodermal marginal cells have internalized. As shown in movie S1, blastodermal cells in M*hwa* mutants internalize and migrate vegetally without dorsally converging during gastrulation, and some internalized cells move toward the animal pole until halting and aggregating at variable positions on the yolk to form a furrow during the segmentation period. If the furrow is formed at an equatorial position, the calabash shape (class I) is created; if the furrow is closer to the animal pole, the onion shape (class II) appears. By observing dynamic migration of enhanced green fluorescent protein (EGFP)–labeled mesodermal cells in *Tg(efnb2b:EGFP)*;M*hwa*$^{-/-}$ embryos and endodermal cells in *Tg(sox17:EGFP)*:M*hwa*$^{-/-}$ embryos (figs. S1 and S2 and movies S2 and S3), we confirmed that mesendodermal cells in mutants undergo internalization but fail to converge dorsally. Furthermore, transplantation experiments revealed that transplanted M*hwa* mutant cells and wild-type cells converge normally in wild-type host embryos during gastrulation but are unable to converge dorsally in M*hwa* mutant embryos (fig. S3 and movie S4). Taken together, these data suggest that the failure of gastrulation cell convergence in M*hwa* mutants accounts for the failure to form the dorsal organizer, head, and trunk. It has been proposed that during zebrafish gastrulation, high Bmp levels inhibit the convergence and extension of internalized mesendodermal precursors, but their vegetal migration is promoted (*38*). In M*hwa* mutants, up-regulated expression of *bmp4* across the dorsoventral axis with loss of the gradient (Fig. 1B) may account for the failure of cell convergence.

The *hwa* locus is responsible for the mutant phenotype

By combining positional and candidate cloning, we found that in M*hwa*^{tsu01sm}, the uncharacterized locus *si:dkey-121h17.7* on chromosome 21 carries a 7313–base pair (bp) insertion element at position +118 upstream of the first exon (Fig. 2A). *hwa* transcripts were enriched in the Balbiani body in wild-type stage I oocytes and in the vegetal pole region upon fertilization (Fig. 2B and fig. S4A). During the early cleavage period, *hwa* transcripts in the vegetal pole region were transported toward the animal pole from one side of the yolk (indicated by arrows in fig. S4A), but transcripts appeared to be distributed evenly in blastomeres. *hwa* mRNA levels decreased at 2 hpf (fig. S4, A and B), indicating maternal expression. However, *hwa* transcripts were undetected in M*hwa* mutant oocytes or early mutant embryos by WISH, and quantitative reverse transcription polymerase chain reaction (qRT-PCR) assays revealed low levels of *hwa* mRNA in mutants, which suggests that *hwa*^{tsu01sm} is a null allele.

Next, we performed rescue experiments. Injection of 5 to 20 pg of *hwa* mRNA into M*hwa* mutant embryos at the one-cell stage efficiently restored the body axis at 24 hpf, and higher doses led to embryonic dorsalization of both mutant and wild-type embryos (Fig. 2C). In addition, *hwa* overexpression also rescued the expression of the dorsal marker *gsc* and the ventral marker *eve1* (fig. S5). Some of the rescued mutant embryos could survive to adulthood. Interestingly, injection of *hwa* mRNA into a single blastomere of mutant embryos at four-cell, 16-cell, or 128-cell stages also produced a rescuing effect (Fig. 2C and fig. S5B). Injection of *hwa* mRNA into one blastomere of the 16- to 32-cell stage wild-type embryos induced a secondary organizer/body axis, and injection of *hwa* mRNA into two distant blastomeres of M*hwa*^{tsu01sm} mutant embryos at 16- to 32-cell stages efficiently induced two organizers/body axes (Fig. 2D). These data indicate that Hwa has potent organizer/body axis–inducing activity and that the dorsalizing activity of maternal *hwa* may not depend on asymmetrical transportation of dorsal determinants after fertilization.

To further confirm the function of *hwa*, we used Cas9 technology to generate the *hwa* mutant line *hwa*^{tsu32c}, in which exon 1 of the *hwa* locus harbors a 33-bp insertion, a 1-bp deletion, and a 1-bp substitution allowing a premature stop codon to appear at the 36th codon. M*hwa*^{tsu32c} mutant embryos at 24 hpf exhibited a calabash- or onion-like shape without the major body axis in addition

to alterations of the dorsal and ventral marker expression pattern (Fig. 2E). The phenotype of M*hwa*^{tsu32c} is identical to that of M*hwa*^{tsu01sm} mutants.

The predicted Hwa protein contains a 23–amino acid extracellular domain, a 23–amino acid transmembrane domain, and a 248–amino acid intracellular domain (Fig. 3A). However, no regions of Hwa protein share high homology with any known protein motifs in the prediction databases. An antibody recognizing an extracellular region of zebrafish Hwa could detect overexpressed hemagglutinin (HA)–tagged Hwa by immunofluorescence in zebrafish embryos (Fig. 3B), showing an enrichment on the plasma membrane. Endogenous Hwa was detected around the 512-cell stage using this antibody. We found that endogenous Hwa is enriched on the plasma membrane of blastomeres only in a small region in which β-catenin is translocated into nuclei and in which the dorsal organizer will form (Fig. 3C). In contrast, no M*hwa*^{tsu01sm} embryos (n = 10) displayed Hwa protein on the plasma membrane in any region (Fig. 3C). These results suggest that, like nuclear β-catenin, Hwa protein is present in future dorsal blastomeres.

By performing rescue experiments in *hwa* mutant embryos using different *hwa* mutant mRNAs, we found that the transmembrane domain and an 84–amino acid intracellular domain (positions 137 to 220) of H*wa* protein are essential for function (fig. S6, A and B). Fine mapping identified two essential motifs, ^{164}VPPNSP^{169} (Val-Pro-Pro-Asn-Ser-Pro) and ^{184}SLRRSST^{190} (Ser-Leu-Arg-Arg-Ser-Ser-Thr) (fig. S6C).

Potential homologs of zebrafish Hwa protein have been identified in sea cucumber (urochordate), *Xenopus laevis* (amphibian), and lizard (reptile), which share a pairwise identity of 17.3 to 34% in amino acid sequence (fig. S7). The VPPNSP motif of zebrafish Hwa is conserved in frog and lizard Hwa, whereas its SLRRSST motif is only conserved in bony fish Hwa (fig. S6C). We have not found Hwa homologs in mammals, most likely because of the extremely low homology with Hwa proteins in the lower organisms.

Next, we examined the developmental role of the *Xenopus hwa* gene. We found that *Xenopus hwa* transcripts are present in granules at the vegetal cortex of oocytes and move dorsally along with cortical rotation after fertilization (Fig. 4A) so that its transcripts are highly enriched in two dorsal blastomeres at the four-cell stage (Fig. 4B). The *hwa* mRNA level in one-cell stage embryos (stage 1) was much lower than that of oocytes, suggesting a quick turnover after fertilization (Fig. 4C). Injection of *hwa* mRNA into one ventral blastomere of the four-cell stage *Xenopus* embryos induced a secondary axis (Fig. 4D) and caused an extra expression domain of the organizer markers *nodal3.1*, *chordin*, and *goosecoid* with inhibition of the ventral markers *wnt8a* and *ventx1.2* (Fig. 4E). Use of antisense oligodeoxynucleotides or morpholino to deplete *hwa* in *Xenopus* oocytes resulted in embryos lacking the body axis and dorsal tissues with reduced dorsal marker expression but enhanced ventral marker expression (Fig. 4F and figs. S8 and S9). These defects could be efficiently rescued by co-injection with a modified *hwa* mRNA that may not be targeted by either antisense oligodeoxynucleotides or morpholino (figs. S8 and S9). These results indicate an evolutionarily conserved function of maternal Hwa. We note that injection of *hwa* morpholino into fertilized eggs had no effect on embryonic development, implying that maternal Hwa protein may play a key role.

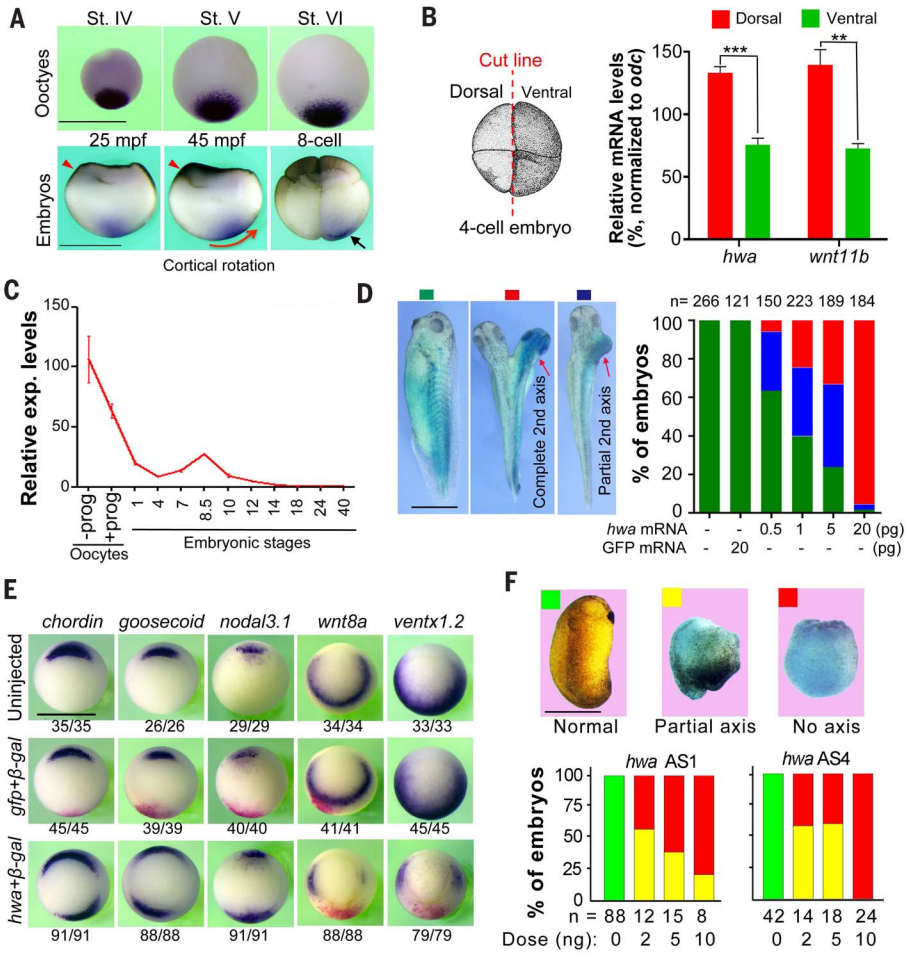


Fig. 4. Roles of *Xenopus hwa* during embryonic body axis formation. (**A**) Expression of *hwa* in oocytes and embryos by WISH detection. Red arrowheads indicate the sperm entry site. St, stage; mpf, minutes post-fertilization. The red arrow shows the cortical rotation direction; the black arrow indicates asymmetrically located *hwa* transcripts. (**B**) Quantification of *hwa* transcripts in two dorsal and two ventral blastomeres at four-cell stage by qRT-PCR. Left: Illustration of cutting of embryos. Right: qRT-PCR result showing average levels from three technical repeats (±SEM) normalized to *odc* level. The dorsal marker *wnt11b* serves as an indicator. $^{**}P < 0.01$, $^{***}P < 0.001$. (**C**) Quantification of *hwa* mRNA levels by qRT-PCR at the indicated stages; prog, progesterone. *hwa* expression level is averaged from technical repeats (±SEM) and normalized to *odc* level at the indicated stages. (**D**) Induction of secondary axis by *hwa* in *Xenopus* embryos. One ventral blastomere of four-cell stage embryos was injected with different doses of *hwa-myc* or 20 pg of *GFP* mRNA plus 200 pg of β-*gal* mRNA and observed at stage 32. Left: Classes of embryos. Right: Ratios of classified embryos. (**E**) Organizer and ventral marker expression patterns detected by WISH at stage 10 (vegetal view). One ventral blastomere of four-cell stage embryos was injected with 20 pg of *hwa* or 100 pg of *GFP* mRNA plus 200 pg of β-*gal* mRNA. The ratio of embryos with the representative pattern is indicated. (**F**) Maternal knockdown effect of *hwa*. Oocytes were injected with different doses of antisense oligodeoxynucleotides (AS1 or AS4) and the resulting embryos were observed and categorized at stage 24. All scale bars, 1 mm.

Hwa function is dependent on β-catenin but independent of Wnt ligand/receptor signaling

Given that β-catenin is essential for body axis formation and its coexistence with Hwa in the future dorsal blastomeres in zebrafish early blastulas (Fig. 3), we investigated the functional relationship between Hwa and β-catenin. Immunofluorescence results showed that β-catenin is present on the plasma membrane but not in nuclei of any blastomeres in M*hwa*^{tsu01sm} mutants at

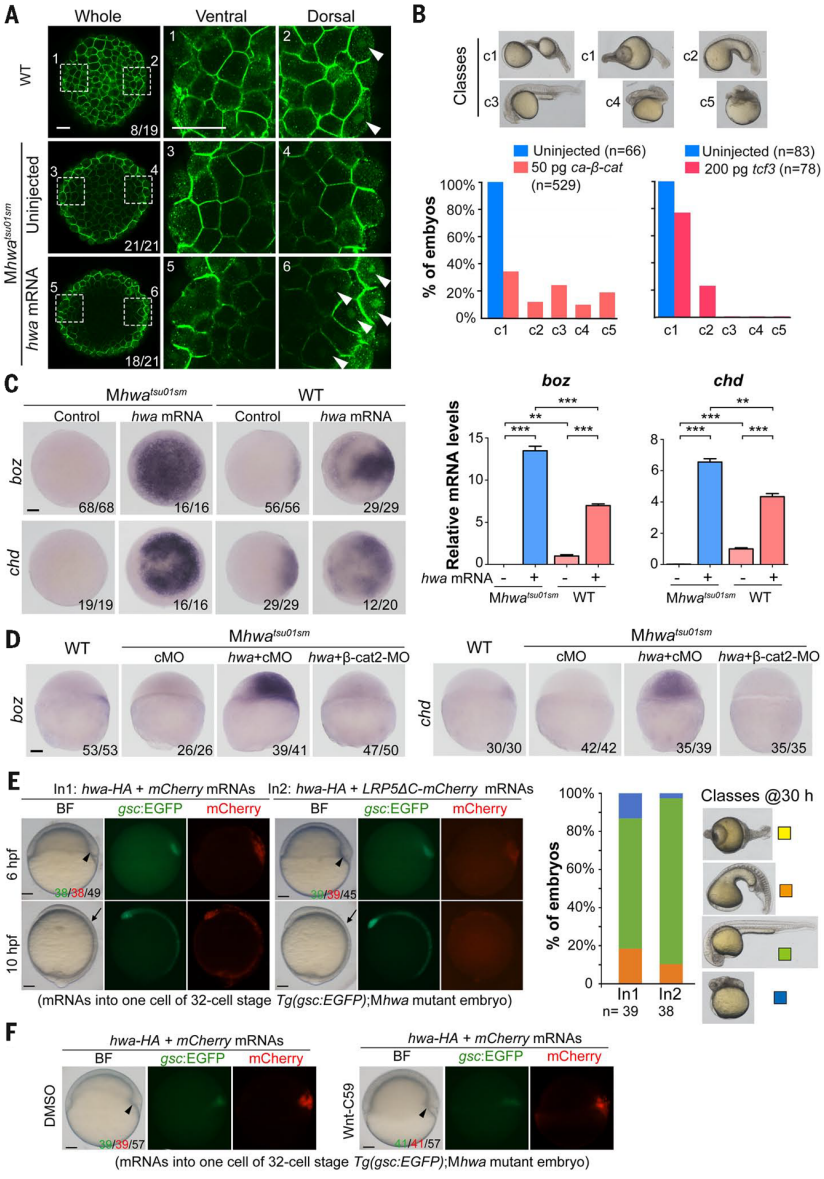


Fig. 5. *hwa* can function through β-catenin in a Wnt ligand/receptor-independent manner in zebrafish embryos. (**A**) Immunostaining of β-catenin. After injection of 50 pg of *hwa* mRNA at one-cell stage, embryos were immunostained for β-catenin at 512-cell stage and observed by confocal microscopy. Whole embryos are viewed from the animal pole; two opposite marginal areas are viewed at a higher magnification. The ratio of embryos with the representative pattern is indicated. (**B**) Recovery of the body axis in M*hwa*^{tsu01sm} embryos by *ca-β-catenin* or *tcf3* overexpression. Embryos were injected at one-cell stage and observed at 24 hpf. The embryos are categorized into five classes (top panel); ratios are shown below. (**C**) *hwa* overexpression induces the expression of the β-catenin target genes *boz* and *chd*. Wild-type or M*hwa*^{tsu01sm} embryos were injected with 50 pg of *hwa*-HA mRNA at the one-cell stage and collected at 4 hpf for WISH or qRT-PCR analysis. Left: WISH results (animal-pole views). Right: qRT-PCR results showing averages from three independent experiments (±SEM). $^{**}P < 0.01$, $^{***}P < 0.001$. (**D**) *hwa* overexpression fails to induce *boz* and *chd* in β-*cat2* morphants. Wild-type or M*hwa*^{tsu01sm} embryos were injected with 50 pg of *hwa-HA* mRNA together with 20 ng of control MO (cMO) or β-cat2-MO at the one-cell stage and subjected to WISH at 4 hpf. Embryos are laterally viewed. (**E** and **F**) Inhibition of Wnt ligand/receptor signaling has no effect on *hwa*'s organizer/axis-inducing activity. (E) One blastomere of *Tg(gsc:EGFP)*;M*hwa*^{tsu01sm} embryos at about 32-cell stage was injected with 50 pg of *hwa* plus 200 pg of *mCherry* (In1) or *LRP5ΔC-mCherry* mRNA (In2). Left: Embryos are shown at 6 and 10 hpf. Embryo ratios are expressed as GFP$^+$/mCherry$^+$/total. Right: Bar graph showing ratios of classes of 30-hpf embryos growing from GFP$^+$ 6-hpf embryos. Note that the class I mutant phenotype did not show in any injected embryos and served only for comparison. (F) *Tg(gsc:EGFP)*;M*hwa*^{tsu01sm} embryos injected with *hwa* and *mCherry* mRNAs as in (E) were incubated in DMSO or 20 μM Wnt-C59. The organizer and body axis are indicated by arrowheads and arrows, respectively. Scale bars, 10 μm (A), 100 μm (all others).

the 512-cell stage, and that nuclear β-catenin is recovered by injection of *hwa* mRNA (Fig. 5A). Overexpression of mRNA encoding the constitutively active (ca) β-catenin in M*hwa*^{tsu01sm} mutants could restore the embryonic body axis and even dorsalize the embryos, whereas *tcf3* overexpression had little rescue effect (Fig. 5B). The expression of the β-catenin target genes *boz* and *chd* was eliminated in M*hwa*^{tsu01sm} mutants, which could be recovered by injection of *hwa* mRNA (Fig. 5C). Note that *hwa* overexpression in the mutants induced higher levels of *boz* and *chd* than in wild-type embryos, which suggests that M*hwa* mutants have a larger number of cells that are responsive to activity of ectopic Hwa relative to wild-type embryos. When β-*catenin2* was silenced, however, the expression of *boz* and *chd* could not be induced by *hwa* mRNA overexpression in M*hwa*^{tsu01sm} mutants (Fig. 5D). These data suggest that Hwa functions through β-catenin.

It has been proposed that during early vertebrate embryogenesis, β-catenin stabilization in dorsal blastomeres depends on Wnt ligand/receptor-mediated signaling (*39*–*41*), but compelling genetic evidence is lacking. In zebrafish, *wnt8a* is believed to be an essential maternal Wnt ligand for the formation of the organizer (*42*). However, a recent study showed that M*wnt8a* mutants have no defects in axis formation and that MZ*wnt8a* mutants display an anterodorsalized phenotype similar to zygotic *wnt8a* mutants (*43*). Given the existence of many maternally expressed Wnt ligands including *wnt8a* and *wnt6a* (*42*, *43*), it is possible that several Wnt ligands work cooperatively to activate β-catenin in early embryos. Nonetheless, we decided to investigate the requirement of maternal Wnt ligands/receptors for dorsal organizer and body axis formation as well as *hwa* function in these aspects. We found that in M*hwa*^{tsu01sm} mutant embryos at the one-cell stage and 4 hpf, the mRNA levels of *wnt8a*, β-*catenin2*, *axin1*, *axin2*, *dvl2*, *lrp5*, *lrp6*, *fzd3b*, *fzd5*, *fzd7b*, and *fzd8b*, which are highly expressed in oocytes (*44*), are comparable to those of wild-type embryos (fig. S10); this suggests that Wnt ligand/receptor signaling, even if activated, is not sufficient for organizer and body axis formation in the absence of Hwa. Furthermore, zebrafish embryos injected at the one-cell stage with *dnwnt8a-mCherry* (encoding mCherry fused to a dominant negative form of Wnt8a), *DKK1-EGFP* (encoding human DKK1-EGFP fusion protein), or *LRP5ΔC-mCherry* mRNA (encoding mCherry fused to human LRP5 with a truncation of the intracellular domain) or treated with the Wnt inhibitor Wnt-C59 all exhibit a typical dorsalized phenotype instead of a ventralized phenotype at 24 hpf and show no obvious reduction of the dorsal marker genes *boz* and *chd* (figs. S11 to S14). The *dnwnt8a*-induced dorsalized phenotype was also reported by Lu *et al.* (*42*), but a complete loss of *chd* expression in a *dnwnt8a*-overexpressing embryo was shown at the sphere stage (4 hpf) [figure 2C of (*42*)], which we did not observe in our assays (fig. S11C).

This inconsistency may be due to the different experimental conditions between the studies. In our study, DKK1-EGFP and DnWnt8a-mCherry proteins in the injected embryos become fluorescently visible as early as the two- to eight-cell stages (figs. S11A and S12A) and thus should be capable of antagonizing existing maternally derived Wnt ligands/receptors. Thus, maternal Wnt/receptor signaling may not be required for organizer formation from the period of fertilization to the onset of gastrulation during zebrafish embryogenesis. Conversely, the dorsalizing and axis-inducing activity of ectopic *hwa* mRNA in zebrafish embryos is not prevented by the coexpression of *dnwnt8a-mCherry*, *DKK1-EGFP*, or *LRP5ΔC-mCherry* or by Wnt-C59 treatment (Fig. 5, E and F, and figs. S11 to S14). Therefore, Hwa can function in a Wnt ligand/receptor-independent fashion.

We next assessed the overexpression of *hwa* or *wnt8* during *Xenopus* oocyte maturation. We found that overexpression slows down the post-fertilization degradation of cytosolic β-catenin accumulated in oocytes (Fig. 6A), which in turn suggests that Hwa helps to stabilize cytosolic β-catenin. When β-catenin MO was injected into *Xenopus* one-cell stage embryos followed by *hwa* mRNA injection into one blastomere at the four-cell stage, no body axis formed (Fig. 6B), confirming that β-catenin is required for Hwa function. Ectopic *hwa* or *wnt8* in *Xenopus* embryos efficiently induced a secondary axis, and the induction of *wnt8* but not *hwa* activity was inhibited by coexpression of *dkk1* (Fig. 6C). As demonstrated earlier (*29*, *30*), depletion of *lrp6* or *wnt11b* mRNA in *Xenopus* oocytes by antisense oligodeoxynucleotides leads to a loss of the body axis and reduction of dorsal organizer marker expression with increased ventral marker expression in the resulting embryos (fig. S15). However, when one blastomere at the four-cell stage was injected with *hwa* mRNA, these embryos derived from oocytes depleted of *lrp6* or *wnt11b* mRNA could form the full body axis and restore dorsal organizer gene expression (Fig. 6, D and E, and fig. S15). These data suggest that *Xenopus* Hwa can also function in a Wnt ligand/receptor-independent fashion.

Hwa promotes Axin degradation

Given that Axin is the rate-limiting component of the β-catenin destruction complex because of its low concentration (*45*, *46*), we wanted to know whether Hwa acts on Axin. Transfection of zebrafish *hwa* into human embryonic kidney (HEK) 293T cells stimulated the expression of the Wnt/β-catenin responsive Topflash reporter and increased the amount of β-catenin in the cytosol (fig. S16). Next, we tested the potential interaction between zebrafish Hwa and mouse Axin in HEK293T cells and detected Myc-Axin2 in the Hwa-Flag immunoprecipitate (Fig. 7A), whereas Hwa and Axin1 did not coimmunoprecipitate. However, an association of Hwa-Flag and Axin1-HA was detected in the presence of XAV939 (Fig. 7B), an inhibitor of tankyrases (Tnks) that mediate poly-ADP ribosylation (PARsylation) and subsequent degradation of Axin (*47*). Western blotting results of whole-cell lysates showed a reduction of Axin1-HA and Myc-Axin2 levels in HEK293T cells when zebrafish Hwa was coexpressed (Fig. 7, A and B). To assess whether Hwa directly binds to Axin, we used the His-AXIN1 (N351) protein, which contains the N-terminal 351 amino acids of human AXIN1, and purified the intracellular domain [GST-Hwa(ΔN46)] of zebrafish Hwa after *Escherichia coli* expression. We found that GST-Hwa(ΔN46) can be pulled down by His-AXIN1 (N351) (Fig. 7C),

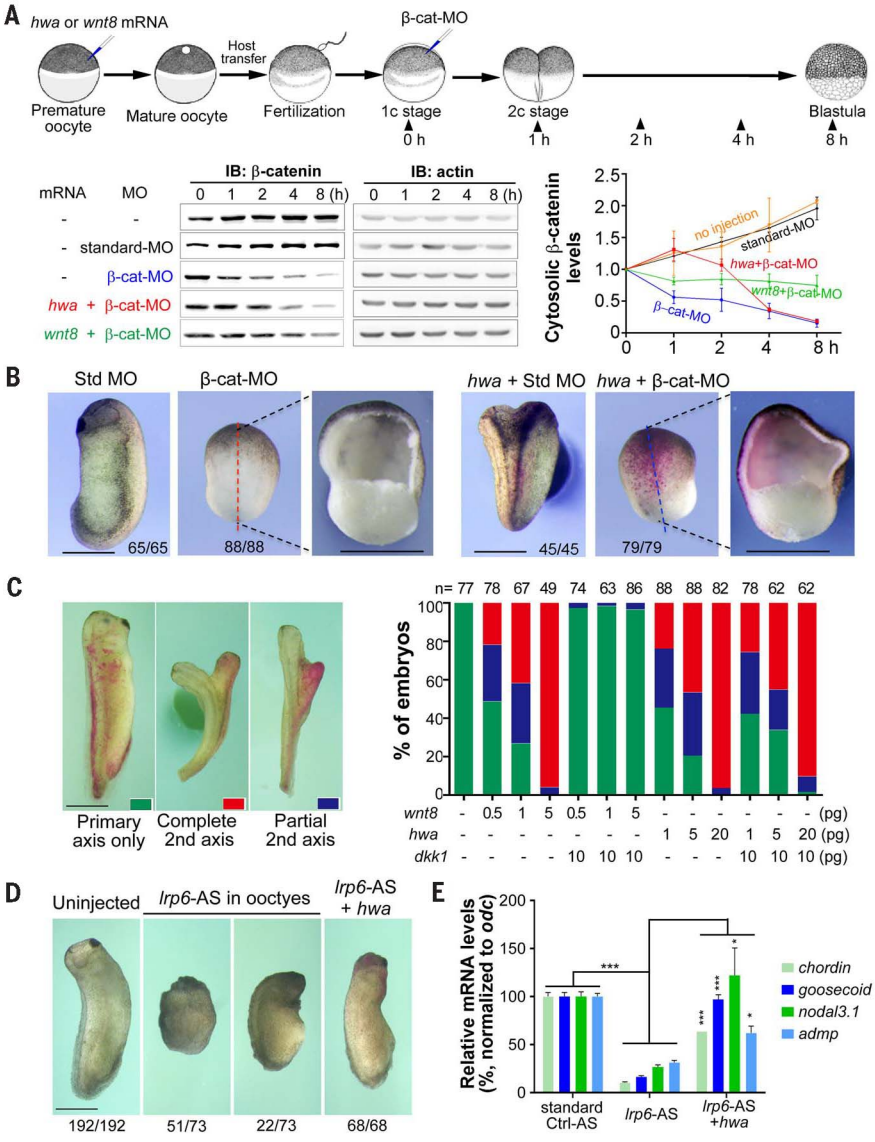


Fig. 6. *hwa* can function through β-catenin in a Wnt ligand/receptor-independent manner in *Xenopus* embryos. (**A**) *hwa* or *wnt8* overexpression delays cytosolic β-catenin degradation. Top: Illustration of experiment design (dosage: *hwa* or *wnt8* mRNA, 200 pg; β-catenin MO or standard MO, 20 ng). Bottom: Immunoblotting results and quantification of cytosolic β-catenin levels. For immunoblotting, membrane fraction in embryo lysate was removed using digitonin buffer. Quantitative data are derived from three experiments. (**B**) *hwa* fails to induce the body axis in embryos depleted of β-catenin. Embryos were first injected with 20 ng of β-cat-MO or standard MO at one-cell stage and then injected with 200 pg of *hwa* and 200 pg of β-*gal* mRNAs into one ventral blastomere at four-cell stage, and were observed at stage 26. (**C**) *dkk1* overexpression inhibits secondary axis induction of *wnt8* but not *hwa*. mRNAs were injected into one ventral blastomere of four-cell stage embryos. Left: Categories of embryos at stage 32. Right: Ratios of embryos in each category. (**D**) *hwa* overexpression induces the body axis of embryos derived from oocytes depleted of *lrp6*. Oocytes were injected with 6 ng of *lrp6* antisense oligodeoxynucleotide (*lrp6*-AS) and one blastomere of the resulting embryos at two-cell stage was injected with 20 pg of *hwa* mRNA. Embryos were observed at stage 28 (left panel). (**E**) Embryos from the experiment shown in (D) were also collected at stage 10 for RT-PCR analysis of dorsal markers. qRT-PCR results are based on four technical repeats. **P* < 0.05, ****P* < 0.001. Scale bars, 1 mm.

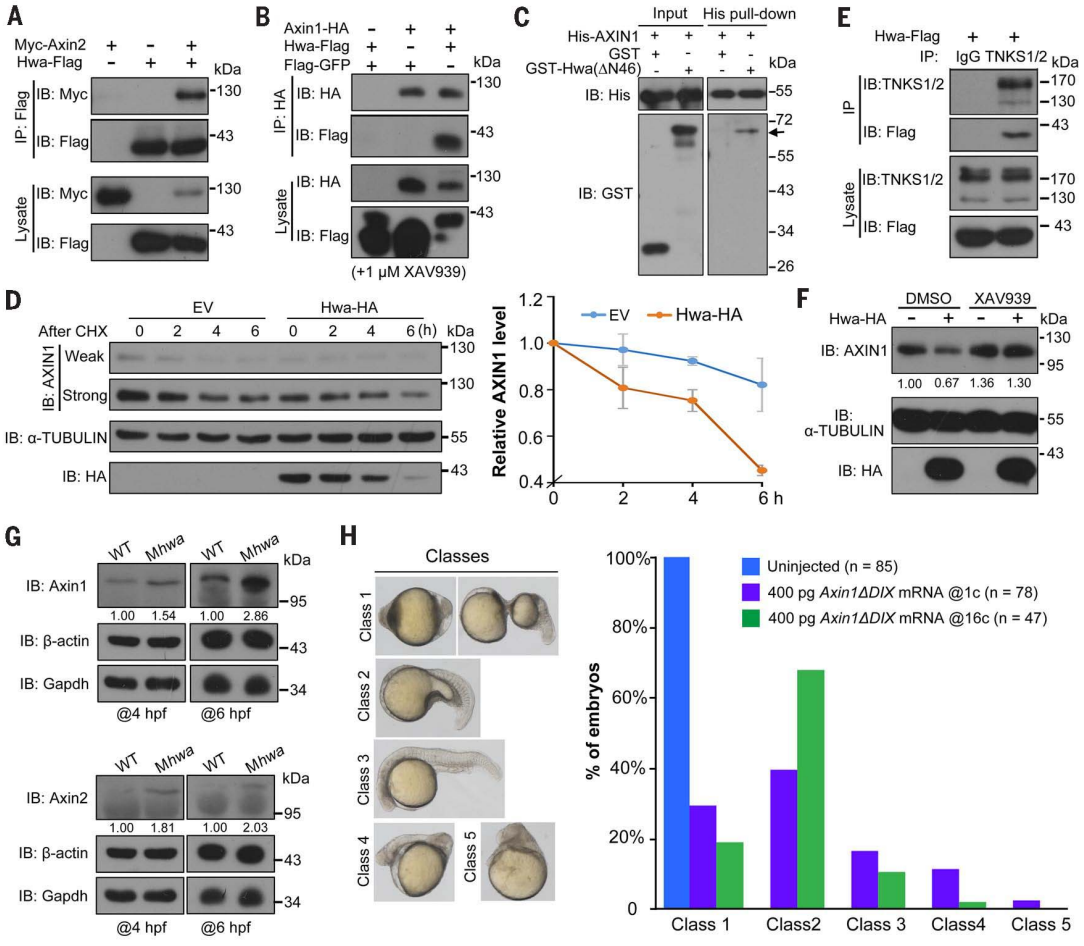


Fig. 7. Hwa binds to and promotes the degradation of Axin. (**A** and **B**) Hwa-Flag physically interacts with Myc-Axin2 (A) and Axin1-HA (B) in HEK293T cells. Zebrafish Hwa and mouse Axin1 and Axin2 were used. XAV939 is a tankyrase inhibitor. Note that Axin1 and Axin2 levels are reduced when zebrafish Hwa is expressed. (**C**) Direct binding of Hwa with AXIN1. His-AXIN1 (N351) protein of human AXIN1 origin was incubated with GST-Hwa(ΔN46) protein of zebrafish origin expressed in *E. coli*. GST-Hwa (indicated by an arrow) is detected by Western blotting in the His-AXIN1 precipitates. (**D**) Hwa-HA promotes degradation of endogenous AXIN1 in HEK293T cells. Left: Cells transfected with empty vector (EV) or Hwa-HA were treated with cycloheximide (CHX; 100 μg/ml) and harvested for immunoblotting at indicated time points. Right: Dynamic change of AXIN1 protein level (normalized to α-TUBULIN), shown as mean ± SD. (**E**) Hwa-Flag binds to endogenous TNKS1/2 in HEK293T cells. (**F**) Hwa-promoted degradation of endogenous AXIN1 in HEK293T cells is prevented by inhibition of tankyrases. Six hours after transfection, cells were treated with DMSO or 1 μM XAV939 for 18 hours. Relative AXIN1 levels (normalized to α-TUBULIN) are indicated. (**G**) Immunoblotting reveals an increase of endogenous Axin1 and Axin2 proteins in M*hwa*^{tsu01sm} mutants at 4 and 6 hpf. Relative levels are indicated. (**H**) *Axin1ΔDIX* overexpression rescues the body axes of M*hwa*^{tsu01sm} mutants. *Axin1ΔDIX* mRNA was injected into one-cell stage mutants or into one blastomere of 16-cell stage mutants. Embryos at 24 hpf were categorized into five classes (left) and the ratio of embryos in each class is shown (right).

which indicates a direct physical interaction of Hwa with Axin.

To confirm the effect of Hwa on Axin protein stability, we examined the degradation of existing AXIN1 protein in HEK293T cells treated with cycloheximide (CHX), a translation inhibitor. Degradation of endogenous AXIN1 protein was accelerated by Hwa-HA expression (Fig. 7D). Hwa-HA in HEK29T cells associated with endogenous TNKS1/2 (Fig. 7E), and XAV939 treatment efficiently blocked Hwa-induced degradation of endogenous AXIN1 (Fig. 7F). These data suggest that Hwa recruits Axin and Tnks to facilitate Tnks-mediated Axin PARsylation and degradation, thereby preventing the formation of the β-catenin destruction complex.

As expected, immunoblotting results showed an increase of endogenous Axin1 and Axin2 levels in zebrafish M*hwa*^{tsu01sm} mutant embryos at 4 and 6 hpf (Fig. 7G). By contrast, overexpression of *hwa* in zebrafish M*hwa*^{tsu01sm} embryos or in wild-type *Xenopus* embryos led to a marked reduction of Axin protein levels (fig. S17). It has been reported that the DIX domain of the Axin protein is required for its activity when forming the β-catenin destruction complex (*48*) and that overexpression of *axinΔDIX* lacking the DIX domain in the zebrafish dorsal axis–defective mutant *ichabod* can recover the dorsal tissues (*49*). Injection of *Axin1ΔDIX* mRNA into one-cell stage M*hwa*^{tsu01sm} mutants or into one blastomere of 16-cell stage M*hwa*^{tsu01sm} mutants rescued or dorsalized the embryonic body in more than 70% of the mutants at 24 hpf (Fig. 7H). Furthermore, overexpression of *Axin1ΔRGS*, which lacks the RGS domain and can act as a dominant negative form (*50*), rescued the embryonic body in approximately 10% of M*hwa*^{tsu01sm} mutants (fig. S18), showing a less efficient rescue effect relative to *Axin1ΔDIX* overexpression. These in vivo data support the notion that Hwa exerts the body axis–inducing effect by promoting Axin degradation.

Discussion

We have identified the transmembrane protein Hwa as a maternal determinant of the dorsal organizer and body axis formation in vertebrates (Fig. 8). Hwa protein is located on the plasma membrane of the future dorsal blastomeres in zebrafish early blastulas (Fig. 3C), in which β-catenin is translocated into the nucleus. Hwa appears to promote Axin degradation and stabilize cytosolic β-catenin during early embryonic development in a Wnt ligand/receptor-independent fashion.

Previous studies in *Xenopus* embryos, which resulted from oocytes depleted of *wnt11b* or *lrp6* mRNA using antisense oligodeoxynucleotides, suggested that Wnt ligand/receptor-mediated signaling is essential for β-catenin activation and dorsal axis formation during early embryogenesis (*29*, *30*). Our results show that the inhibition of Wnt ligand/receptor signaling during early embryogenesis does not impair the organizer and body axis formation (Fig. 5, E and F, and figs. S11 to S14), and instead results in a dorsalized phenotype that is typically ascribed to loss of zygotic Wnt signaling. Our observations are consistent with the finding that the overexpression of dominant

negative *lrp6* in *Xenopus* embryos does not hinder axis formation (*51*).

In zebrafish and *Xenopus* embryos, nuclear β-catenin is not detected until early blastula stages (*5*). We found that injection of *hwa* mRNA into one blastomere of 128-cell stage M*hwa* mutant embryos is able to rescue the body axis, although the rescue efficiency is lower than for earlier injections (Fig. 2C). In mice, although β-catenin signaling is activated through oocyte maturation and embryonic development, inhibition or overactivation of β-catenin does not influence embryonic development prior to implantation (*52*, *53*). These data raise the possibility that β-catenin signaling prior to early blastulation may not be essential for the formation of the dorsal organizer and body axis.

Hwa is a transmembrane protein, and zebrafish Hwa contains an extracellular domain of only 23 residues. Overexpression of a *hwa* mutant construct lacking the extracellular domain rescued M$hwa^{tsu01sm}$ mutants (fig. S6, A and B); therefore, Hwa is unlikely to be a ligand-binding receptor. We have also shown that Hwa may not functionally depend on Wnt receptors during early embryonic development. However, we cannot exclude the possibility that Hwa exerts its dorsalizing activity in concert with other signaling pathways. The mechanisms underlying Hwa function need to be explored further.

Materials and methods

Zebrafish

Tübingen (TU) and India strains were used. Embryos were raised at 28.5°C and staged as described (*54*). Fish maintenance followed the institutional animal care and use committee (IACUC) protocol, with the approval by the Tsinghua University Animal Care and Use Committee. Zygotic homozygous ($Zhwa^{tsu01sm/tsu01sm}$) mutant offspring derived from crosses between heterozygous fish develop and grow normally. M$hwa^{tsu01sm}$ mutant embryos derived from crosses between $Zhwa^{tsu01sm/tsu01sm}$ females and wild-type males are phenotypically identical to MZ$hwa^{tsu01sm}$ mutants generated by $Zhwa^{tsu01sm/tsu01sm}$ female/male crosses. Using a mapping strategy described by Zhou *et al.* (*55*), the mutant gene was initially mapped in a F_2 population established by crossing TU mutation carriers with India fish to the 1.2-Mb region between the markers CH73-29C22 and DKEY-93M18 on chromosome 21. Examination of the expression of 35 candidate genes residing in that region identified the uncharacterized gene *si:dkey-121h17.7* as the mutant gene, which was further confirmed by mRNA rescue experiments. To genotype $hwa^{tsu01sm}$ mutants by PCR, the upper primer (5′-AAAGTGGCTTCTGGCGACAT-3′) and the lower primer (5′-AAGTACTGGTGAGTTCGGCG-3′) were used. hwa^{tsu32c} mutants were generated by CRISPR/Cas9 approach. The guide RNA was designed to target the first exon of *si:dkey-121h17.7* (Ensembl Gene ID: ENSDARG00000094542). The target sequence including the PAM (underlined) is GCCCTCATCAAACTTACCTGAGG. Primers used for genotyping hwa^{tsu32c} were 5′-CACACCGAAGACACCCTGAT-3′ (upper primer) and 5′-CCTCTGGCGGGTGGAATATG-3′ (lower primer). The transgenic lines $Tg(efnb2b:EGFP)^{tsu2014}$ (*56*, *57*), $Tg(sox17:EGFP)^{s870}$ (*58*), and *Tg(gsc:EGFP)* (*59*) were described previously. For inhibition of Wnt signaling, zebrafish embryos were treated with Wnt-C59 (10 or 20 μM) from the one-cell stage to the desired stage.

The *Xenopus laevis* animals (J strain) were obtained from Nasco Inc. (Wisconsin, USA) and maintained according to the IACUC protocol. Approval was provided by the Tsinghua University Animal Care and Use Committee.

Whole-mount in situ hybridization (WISH) and immunofluorescence

WISH for zebrafish embryos was performed following the standard protocol. WISH for *Xenopus* embryos followed the previous procedures (*60*). The linearized plasmid was used as the template for in vitro synthesizing digoxigenin-UTP–labeled antisense RNA probe. Embryos after WISH were immersed in glycerol and photographed with a Nikon SMZ1500 stereomicroscope.

Whole-mount immunofluorescence was carried out based on the standard protocol using the following primary antibodies: rabbit anti–β-catenin antibody (C2206, Sigma, 1:1000 dilution) or mouse anti–β-catenin antibody (2677S, CST, 1:200 dilution), anti-HA antibody (sc-7392, Santa Cruz, 1:200 dilution). The rabbit anti-Hwa antibody was customized by Fantibody using a synthesized peptide (MSQLGSAVPSSNLPE) derived from the extracellular domain of zebrafish Hwa protein. For nuclear β-catenin staining, embryos were treated with 2 M hydrochloric acid for 30 min after rehydration and equilibrated with 100 mM Tris-HCl (pH 8.5) for 15 min. For Hwa immunostaining, the tyramide signal amplification (TSA)–based immunodetection protocol was used (*61*). After rehydration, embryos were treated with 0.3% H_2O_2 for 30 min. An HRP-conjugated goat anti-rabbit antibody (D110058, BBI, 1:200) was used as the secondary antibody. After six times of washing, using PBST with 0.1% Triton X-100, embryos were incubated in TSA reagent (TSA-Cy5, PerkinElmer, 1:50) for 1 hour at room temperature. Then, embryos were washed and mounted for observation. Immunostained embryos were photographed using a Zeiss 710 META confocal laser scanning microscope.

Morpholino, antisense oligodeoxynucleotides, and mRNA injection

Morpholino methods to knock down zebrafish β-*catenin2* and mismatched control morpholino were as described (*62*). Sequences of morpholinos and oligodeoxynucleotides for *Xenopus* genes are listed in table S1. Antisense morpholino for knocking down *Xenopus* β-*catenin* and oligodeoxynucleotides for *Xenopus wnt11b* and *lrp6* genes were as described (*9*, *29*, *30*). All morpholinos were purchased from Gene Tools LLC. Capped mRNAs were synthesized with the mMessage mMachine kit (Ambion), purified with RNeasy

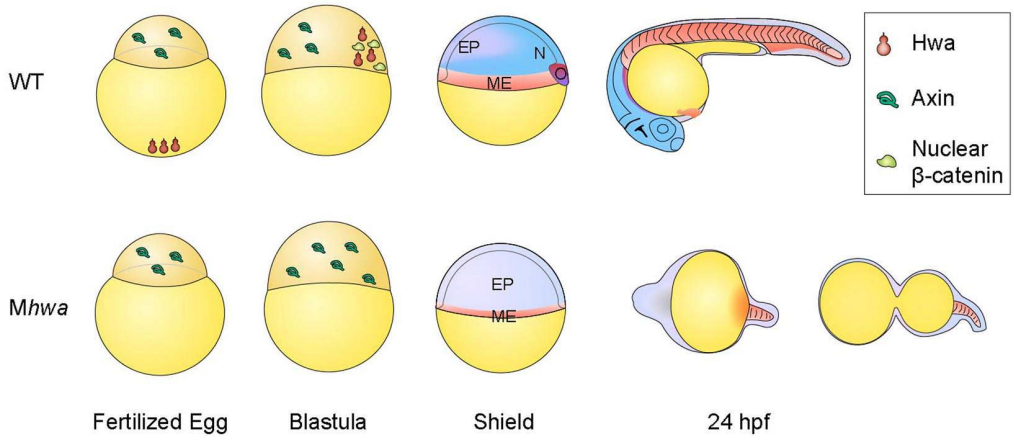


Fig. 8. Working model of Hwa in the zebrafish embryo. In the wild-type fertilized egg, *hwa* mRNA is localized in the vegetal pole region, whereas Axin and β-catenin proteins are located in the cytoplasm. During early blastulation, Hwa protein is located in the dorsal blastomeres, where it locally promotes Axin degradation and results in the stabilization and nuclear transportation of β-catenin. At the onset of gastrulation, the dorsal organizer, also known as the embryonic shield, is formed by the action of β-catenin. The organizer regulates the cell fates along the dorsal-ventral and anterior-posterior axes as well as cell movements during gastrulation, which leads to the formation of the main body. In the absence of *hwa* (as in M*hwa* mutants), cytosolic β-catenin is destroyed by the Axin-containing machinery. As a result, the organizer is not formed and the head, neural tissues, and body axis do not develop, but there are residual mesendodermal tissues in the tail-like region. O, organizer; N, neuroectoderm; EP, epidermis; ME, mesendoderm.

mini kit (Qiagen) according to the manufacturers' instructions, and injected into the yolk or the cytoplasm at the desired stage as elucidated in the main text.

Knockdown of Xenopus maternal genes through host transfer technique

Xenopus embryos depleted of maternal *hwa*, *wnt11b*, or *lrp6* were obtained through host transfer technique as described (*63*, *64*). Briefly, meiosis I (G_2/M)–arrested oocytes were isolated from small lobes of ovary through manual defolliculation using watchmaker's forceps and cultured in oocyte culture medium (OCM, glutamine plus L15 supplemented with 4% BSA and antibiotics). Isolated oocytes injected with or without antisense oligos or mRNA were cultured for 48 hours and then matured with 2 μM progesterone for 12 hours at 18°C. Mature oocytes were labeled with vital dyes (neutral red, Bismarck brown, Nile blue) and then transferred into the body cavity of an egg-laying female anesthetized with tricaine methanesulfonate (MS222, A5040, Sigma). The female recovered from the surgery was reared in 1.2× MMR (Marc's Modified Ringer) solution for 4 to 6 hours for egg collection. Eggs were then fertilized with sperm suspension. Fertilized eggs were dejellied using 2% cysteine in 0.1× MMR (pH 8.0) followed by thorough wash with 0.3× MMR. For microinjection after fertilization, eggs were cultured in 0.3× MMR + 4% Ficoll400. Microinjected embryos were transferred into fresh 0.3× MMR at stage 8 and cultured to the desired stages.

Embryonic cell transplantation

To investigate the cell migration behavior of M*hwa*^{tsu01sm}, lateral margin cells from wild-type or M*hwa*^{tsu01sm} embryos at 6 hpf were transplanted to the corresponding position in the host embryos. Donor cells were labeled by EGFP via injection of 100 pg *EGFP* mRNA at the one-cell stage. Embryos were embedded in 0.5% low–melting point agarose; time-lapse imaging started at about 8 hpf.

Live imaging

Zebrafish embryos were placed in a temperature-controlled sample holder and live-imaged under a Zeiss 710 META or Olympus FV1200 confocal microscope. GFP channel images are 3D views of z-stacks. For axis induction visualization, embryos were mounted in 5% methyl cellulose (Sigma) and photographed under an Olympus MVX10 fluorescence macro zoom microscope at indicated stages. For time-lapse movies showing the early development, embryos were photographed using a Nikon Eclipse Ti microscope and a Motic SMZ-161 stereoscope. Movies were processed with Imaris software 7.2.3 (Bitplane) and Adobe Photoshop CC.

Immunoblotting and immunoprecipitation

HEK293T cells were cultured in DMEM supplemented with 10% FBS and penicillin/streptomycin in a 37°C humidified incubator with 5% CO_2. Plasmids were transfected with polyethylenimine (23966, Polysciences) or VigoFect (T001, Vigorous) according to the manufacturers' instructions. For Topflash Luciferase reporter assay, the internal control pRL-TK plasmid was co-transfected. Coimmunoprecipitation and immunoblotting were performed essentially as described (*56*, *65*). Treatment of cultured cells with the tankyrase inhibitor XAV939 was performed as described (*47*). Briefly, cells were cultured in DMEM containing 1 μM XAV939 from 6 hours after transfection and harvested at 24 hours after transfection. To investigate the degradation of endogenous Axin1, CHX (100 μg/ml) was added to the culture medium at 30 hours after transfection and cells were then harvested for immunoblotting at desired time points. The band intensity was assessed using ImageJ software.

Immunoblotting using *Xenopus* embryos was performed as described (*66*). Briefly, embryos at desired stages were lysed with Digitonin buffer (0.15 mg/ml Digitonin in PBS pH7.6) or RIPA buffer (25 mM Tris-HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS). Protein amount equivalent to one embryo for each sample was loaded onto PAGE/SDS gel. β-Actin was detected as the loading control.

In vitro pull-down assay

GST and GST-Hwa(ΔN46) protein were produced in *E. coli* BL21 (DE3). Briefly, when OD_{600} reached 0.8 to 1.0, 1 mM IPTG was added to induce protein synthesis. After 3 hours of induction at 37°C, cells were harvested and sonicated. The cell lysates were centrifuged at 12,000*g* at 4°C for 40 min. GST protein present in the supernatant was purified using glutathione Sepharose 4B resin (GE Healthcare). To purify GST-Hwa(ΔN46) present in the pellet (in the form of inclusion bodies), the pellet was suspended and washed sequentially with wash buffer I [50 mM Tris-HCl (pH 8.0), 2 mM EDTA, 150 mM NaCl, 0.5% Triton X-100, 2 M urea], wash buffer II [50 mM Tris-HCl (pH 8.0), 2 mM EDTA, 150 mM NaCl, 3% Triton X-100], and wash buffer III [50 mM Tris-HCl (pH 8.0), 2 mM EDTA, 150 mM NaCl, 0.5% Triton X-100, 2 M guanidine hydrochloride], each followed by centrifugation at 1500*g* at 4°C for 30 min. After final wash, the precipitate was dissolved in lysis buffer [50 mM Tris-HCl (pH 8.0), 2 mM EDTA, 2 mM DTT, 8 M urea] at room temperature for 30 min and dialyzed at 4°C in dialysis buffer [50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 1 mM EDTA, 5% glycerol, 1% Triton X-100]. The purified protein had a purity of >90% as examined by Coomassie Brilliant Blue G-250 staining. *E. coli*–derived N-terminal 6× His-tagged human AXIN1 (amino acids 1 to 351) was purchased from Proteintech (Ag10079).

For detection of Hwa/AXIN1 interaction, the His-AXIN1 protein was incubated with GST or GST-Hwa(ΔN46) in 1 ml of dialysis buffer for 2 hours at 4°C. Then, ~10 μl of Ni-NTA-Agarose resin (30310, Qiagen) was added to the mixture and incubated at 4°C overnight. Finally, the resin was washed four times, 5 to 10 min each, dissolved in 2× SDS loading buffer, and examined by immunoblotting with anti-GST and anti-His antibodies.

Quantitative RT-PCR

Zebrafish embryos, oocytes, or eggs were collected at desired stages. Total RNA isolation, cDNA synthesis, and qRT-PCR were performed as described (*67*). Expression levels were normalized to the reference gene *eif4g2a* for zebrafish gene expression assays. The sequences of the qRT-PCR primers for zebrafish genes are listed in table S2.

For qRT-PCR analysis of *Xenopus* genes, oocytes or embryos at desired stages were lysed with extraction buffer [50 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM EDTA, 0.5% SDS, and Proteinase K (0.25 mg/ml)], extracted with acidic phenol/chloroform, and precipitated with 70% isopropanol. RNA pellets were then washed with 70% alcohol and air-dried. Purified RNA was resuspended in pure water; 2 μg of total RNA for each sample was used to synthesize cDNA. All quantitative PCR data were acquired using the BIOER qPCR machine (Hangzhou, China) following the manufacturer's instruction. The expression level of *odc* was used as the control reference for all samples. The sequences of qRT-PCR primers for *Xenopus* genes are listed in table S2.

Plasmid constructs

Full-length coding sequence and 3′UTR of *hwa* (XM_698288.8) was amplified and cloned into the EZ-T vector (GenStar) for making antisense RNA probe. The ORF of *Xenopus hwa* with an HA-tag at its 3′ end was cloned into pCS107 vector using restriction enzymes *EcoR*I/*Xho*I. The expression constructs Myc-tagged Axin2 (mouse) and β-catenin (mouse), HA-tagged Axin1 (mouse), and LRP5ΔN (human) were all based on pCS2(+) vector and were obtained from W. Wu, Tsinghua University. For zebrafish *hwa* expression constructs, full-length or partial *hwa* coding sequence was amplified and cloned into pCS2(+) vector, with addition of HA, Flag, or Myc tag in frame at the C terminus. Hwa tagged with HA or Flag was functional, as indicated by an efficient rescue effect on *hwa* mutant embryos. The construct pCS2(+)-LRP5ΔC-mCherry was made by fusing human LRP5 lacking the intracellular domain to mCherry, which was used as a dominant negative form of LRP5; pCS2(+)-Axin1ΔDIX and pCS2(+)-Axin1ΔRGS were constructed by deleting the DIX domain (amino acids 750 to 832) and the RGS domain (amino acids 88 to 211) of the mouse Axin1 from HA-tagged Axin1, respectively. For the production of zebrafish Hwa(ΔN46) protein, the coding sequence encoding amino acids 47 to 294 plus the His tag was cloned into the pGEX-6p-1 expression vector with GST fusion at the N terminus.

Antibodies and drugs

The following antibodies and reagents were used: anti-Flag (F1804, Sigma; M185-3L, MBL), anti–c-Myc (sc-40, Santa Cruz), anti-HA antibody (sc-7392, Santa Cruz), anti-GFP (sc-9996, Santa Cruz), anti-Axin1 (2087S, CST), anti–β-catenin (8480S, CST), anti–non-p-β-catenin (8814S, CST), anti-Axin2 (LS-C345085, LifeSpan BioSciences), anti-Tankyrase1/2 (sc-365897, Santa Cruz), anti-His

(AE003, Abclonal), anti-GST (AF0174, Beyotime), anti–α-tubulin (T6199, Sigma), anti–β-actin (AC026, Abclonal), anti-GAPDH (TA-08, ZSGB-BIO), XAV939 (13596, Cayman), CHX (C8030, Solarbio), and Wnt-C59 (S7037, Selleck). Antibodies used for immunoblotting of *Xenopus* embryo lysates included anti–β-catenin (2206, Sigma), anti-HA (11867423001, Roche), and anti–β-actin (BE0021, Easy Bio).

Statistical analysis

Significance of differences between means was analyzed using Student's *t* test. Data for statistical analysis are expressed as mean ± SEM unless otherwise stated.

REFERENCES AND NOTES

1. C. Niehrs, The complex world of WNT receptor signalling. *Nat. Rev. Mol. Cell Biol.* **13**, 767–779 (2012). doi: 10.1038/nrm3470; pmid: 23151663
2. M. Kikkawa, K. Takano, A. Shinagawa, Location and behavior of dorsal determinants during first cell cycle in Xenopus eggs. *Development* **122**, 3687–3696 (1996). pmid: 9012490
3. M. Sakai, The vegetal determinants required for the Spemann organizer move equatorially during the first cell cycle. *Development* **122**, 2207–2214 (1996). pmid: 8681801
4. S. Jesuthasan, U. Strähle, Dynamic microtubules and specification of the zebrafish embryonic axis. *Curr. Biol.* **7**, 31–42 (1997). doi: 10.1016/S0960-9822(06)00025-X; pmid: 9024620
5. S. Schneider, H. Steinbeisser, R. M. Warga, P. Hausen, β-catenin translocation into nuclei demarcates the dorsalizing centers in frog and fish embryos. *Mech. Dev.* **57**, 191–198 (1996). doi: 10.1016/0925-4773(96)00546-1; pmid: 8843396
6. B. A. Rowning *et al*., Microtubule-mediated transport of organelles and localization of β-catenin to the future dorsal side of Xenopus eggs. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1224–1229 (1997). doi: 10.1073/pnas.94.4.1224; pmid: 9037034
7. C. A. Larabell *et al*., Establishment of the dorso-ventral axis in *Xenopus* embryos is presaged by early asymmetries in β-catenin that are modulated by the Wnt signaling pathway. *J. Cell Biol.* **136**, 1123–1136 (1997). doi: 10.1083/jcb.136.5.1123; pmid: 9060476
8. J. Heasman *et al*., Overexpression of cadherins and underexpression of β-catenin inhibit dorsal mesoderm induction in early Xenopus embryos. *Cell* **79**, 791–803 (1994). doi: 10.1016/0092-8674(94)90069-8; pmid: 7528101
9. J. Heasman, M. Kofron, C. Wylie, β-catenin signaling activity dissected in the early Xenopus embryo: A novel antisense approach. *Dev. Biol.* **222**, 124–134 (2000). doi: 10.1006/dbio.2000.9720; pmid: 10885751
10. C. Kelly, A. J. Chin, J. L. Leatherman, D. J. Kozlowski, E. S. Weinberg, Maternally controlled β-catenin-mediated signaling is required for organizer formation in the zebrafish. *Development* **127**, 3899–3911 (2000). pmid: 10952888
11. H. Haegel *et al*., Lack of β-catenin affects mouse development at gastrulation. *Development* **121**, 3529–3537 (1995). pmid: 8582267
12. J. Huelsken *et al*., Requirement for β-catenin in anterior-posterior axis formation in mice. *J. Cell Biol.* **148**, 567–578 (2000). doi: 10.1083/jcb.148.3.567; pmid: 10662781
13. T. Valenta *et al*., Probing transcription-specific outputs of β-catenin in vivo. *Genes Dev.* **25**, 2631–2643 (2011). doi: 10.1101/gad.181289.111; pmid: 22190459
14. J. L. Stamos, W. I. Weis, The β-catenin destruction complex. *Cold Spring Harb. Perspect. Biol.* **5**, a007898 (2013). doi: 10.1101/cshperspect.a007898; pmid: 23169527
15. D. A. Cross, D. R. Alessi, P. Cohen, M. Andjelkovich, B. A. Hemmings, Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* **378**, 785–789 (1995). doi: 10.1038/378785a0; pmid: 8524413
16. C. Desbois-Mouthon *et al*., Insulin and IGF-1 stimulate the β-catenin pathway through two signalling cascades involving GSK-3β inhibition and Ras activation. *Oncogene* **20**, 252–259 (2001). doi: 10.1038/sj.onc.1204064; pmid: 11313952
17. O. G. Morali *et al*., IGF-II induces rapid β-catenin relocation to the nucleus during epithelium to mesenchyme transition. *Oncogene* **20**, 4942–4950 (2001). doi: 10.1038/sj.onc.1204660; pmid: 11526479
18. D. Fang *et al*., Phosphorylation of β-catenin by AKT promotes β-catenin transcriptional activity. *J. Biol. Chem.* **282**, 11221–11229 (2007). doi: 10.1074/jbc.M611871200; pmid: 17287208
19. H. Ji *et al*., EGF-induced ERK activation promotes CK2-mediated disassociation of α-Catenin from β-Catenin and transactivation of β-Catenin. *Mol. Cell* **36**, 547–559 (2009). doi: 10.1016/j.molcel.2009.09.034; pmid: 19941816
20. L. Yang, C. Lin, Z. R. Liu, P68 RNA helicase mediates PDGF-induced epithelial mesenchymal transition by displacing Axin from β-catenin. *Cell* **127**, 139–155 (2006). doi: 10.1016/j.cell.2006.08.036; pmid: 17018282
21. J. R. Miller *et al*., Establishment of the dorsal-ventral axis in *Xenopus* embryos coincides with the dorsal enrichment of dishevelled that is dependent on cortical rotation. *J. Cell Biol.* **146**, 427–437 (1999). doi: 10.1083/jcb.146.2.427; pmid: 10427095
22. C. C. Malbon, H. Y. Wang, Dishevelled: A mobile scaffold catalyzing development. *Curr. Top. Dev. Biol.* **72**, 153–166 (2006). doi: 10.1016/S0070-2153(05)72002-0; pmid: 16564334
23. A. C. Lekven, C. J. Thorpe, J. S. Waxman, R. T. Moon, Zebrafish wnt8 encodes two wnt8 proteins on a bicistronic transcript and is required for mesoderm and neurectoderm patterning. *Dev. Cell* **1**, 103–114 (2001). doi: 10.1016/S1534-5807(01)00007-7; pmid: 11703928
24. M. C. Ramel, A. C. Lekven, Repression of the vertebrate organizer by Wnt8 is mediated by Vent and Vox. *Development* **131**, 3991–4000 (2004). doi: 10.1242/dev.01277; pmid: 15269175
25. J. L. Christian, R. T. Moon, Interactions between Xwnt-8 and Spemann organizer signaling pathways generate dorsoventral pattern in the embryonic mesoderm of Xenopus. *Genes Dev.* **7**, 13–28 (1993). doi: 10.1101/gad.7.1.13; pmid: 8422982
26. S. Takada *et al*., Wnt-3a regulates somite and tailbud formation in the mouse embryo. *Genes Dev.* **8**, 174–189 (1994). doi: 10.1101/gad.8.2.174; pmid: 8299937
27. T. L. Greco *et al*., Analysis of the vestigial tail mutation demonstrates that Wnt-3a gene dosage regulates mouse axial development. *Genes Dev.* **10**, 313–324 (1996). doi: 10.1101/gad.10.3.313; pmid: 8595882
28. K. I. Pinson, J. Brennan, S. Monkley, B. J. Avery, W. C. Skarnes, An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**, 535–538 (2000). doi: 10.1038/35035124; pmid: 11029008
29. Q. Tao *et al*., Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos. *Cell* **120**, 857–871 (2005). doi: 10.1016/j.cell.2005.01.013; pmid: 15797385
30. M. Kofron *et al*., Wnt11/β-catenin signaling in both oocytes and early embryos acts through LRP6-mediated regulation of axin. *Development* **134**, 503–513 (2007). doi: 10.1242/dev.02739; pmid: 17202189
31. G. Bellipanni *et al*., Essential and opposing roles of zebrafish β-catenins in the formation of dorsal axial structures and neurectoderm. *Development* **133**, 1299–1309 (2006). doi: 10.1242/dev.02295; pmid: 16510506
32. H. Nojima *et al*., Genetic evidence for involvement of maternally derived Wnt canonical signaling in dorsal determination in zebrafish. *Mech. Dev.* **121**, 371–386 (2004). doi: 10.1016/j.mod.2004.02.003; pmid: 15110047
33. H. Nojima *et al*., Syntabulin, a motor protein linker, controls dorsal determination. *Development* **137**, 923–933 (2010). doi: 10.1242/dev.046425; pmid: 20150281
34. J. Lyman Gingerich, T. A. Westfall, D. C. Slusarski, F. Pelegri, hecate, a zebrafish maternal effect gene, affects dorsal organizer induction and intracellular calcium transient frequency. *Dev. Biol.* **286**, 427–439 (2005). doi: 10.1016/j.ydbio.2005.07.031; pmid: 16154557
35. X. Ge *et al*., Hecate/Grip2a acts to reorganize the cytoskeleton in the symmetry-breaking event of embryonic axis induction. *PLOS Genet.* **10**, e1004422 (2014). doi: 10.1371/journal.pgen.1004422; pmid: 24967891
36. W. Mei, K. W. Lee, F. L. Marlow, A. L. Miller, M. C. Mullins, hnRNP I is required to generate the Ca2+ signal that causes egg activation in zebrafish. *Development* **136**, 3007–3017 (2009). doi: 10.1242/dev.037879; pmid: 19666827
37. P. D. Campbell, A. E. Heim, M. Z. Smith, F. L. Marlow, Kinesin-1 interacts with Bucky ball to form germ cells and is required to pattern the zebrafish body axis. *Development* **142**, 2996–3008 (2015). doi: 10.1242/dev.124586; pmid: 26253407
38. D. C. Myers, D. S. Sepich, L. Solnica-Krezel, Bmp activity gradient regulates convergent extension during zebrafish gastrulation. *Dev. Biol.* **243**, 81–98 (2002). doi: 10.1006/dbio.2001.0523; pmid: 11846479
39. E. M. De Robertis, Spemann's organizer and self-regulation in amphibian embryos. *Nat. Rev. Mol. Cell Biol.* **7**, 296–302 (2006). doi: 10.1038/nrm1855; pmid: 16482093
40. A. F. Schier, W. S. Talbot, Molecular genetics of axis formation in zebrafish. *Annu. Rev. Genet.* **39**, 561–613 (2005). doi: 10.1146/annurev.genet.37.110801.143752; pmid: 16285872
41. Y. G. Langdon, M. C. Mullins, Maternal and zygotic control of zebrafish dorsoventral axial patterning. *Annu. Rev. Genet.* **45**, 357–377 (2011). doi: 10.1146/annurev-genet-110410-132517; pmid: 21942367
42. F. I. Lu, C. Thisse, B. Thisse, Identification and mechanism of regulation of the zebrafish dorsal determinant. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15876–15880 (2011). doi: 10.1073/pnas.1106801108; pmid: 21911385
43. H. Hino *et al*., Roles of maternal wnt8a transcripts in axis formation in zebrafish. *Dev. Biol.* **434**, 96–107 (2018). doi: 10.1016/j.ydbio.2017.11.016; pmid: 29208373
44. B. S. Zhao *et al*., m^6A-dependent maternal mRNA clearance facilitates zebrafish maternal-to-zygotic transition. *Nature* **542**, 475–478 (2017). doi: 10.1038/nature21355; pmid: 28192787
45. A. Salic, E. Lee, L. Mayer, M. W. Kirschner, Control of β-catenin stability: Reconstitution of the cytoplasmic steps of the wnt pathway in Xenopus egg extracts. *Mol. Cell* **5**, 523–532 (2000). doi: 10.1016/S1097-2765(00)80446-3; pmid: 10882137
46. E. Lee, A. Salic, R. Krüger, R. Heinrich, M. W. Kirschner, The roles of APC and Axin derived from experimental and theoretical analysis of the Wnt pathway. *PLOS Biol.* **1**, e10 (2003). doi: 10.1371/journal.pbio.0000010; pmid: 14551908
47. S. M. Huang *et al*., Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* **461**, 614–620 (2009). doi: 10.1038/nature08356; pmid: 19759537
48. S. Kishida *et al*., DIX domains of Dvl and axin are necessary for protein interactions and their ability to regulate β-catenin stability. *Mol. Cell. Biol.* **19**, 4414–4422 (1999). doi: 10.1128/MCB.19.6.4414; pmid: 10330181
49. F. Valenti *et al*., The increase in maternal expression of *axin1* and *axin2* contribute to the zebrafish mutant *ichabod* ventralized phenotype. *J. Cell. Biochem.* **116**, 418–430 (2015). doi: 10.1002/jcb.24993; pmid: 25335865
50. L. Zeng *et al*., The mouse Fused locus encodes Axin, an inhibitor of the Wnt signaling pathway that regulates embryonic axis formation. *Cell* **90**, 181–192 (1997). doi: 10.1016/S0092-8674(00)80324-4; pmid: 9230313
51. K. Tamai *et al*., LDL-receptor-related proteins in Wnt signal transduction. *Nature* **407**, 530–535 (2000). doi: 10.1038/35035117; pmid: 11029007
52. H. Xie *et al*., Inactivation of nuclear Wnt-β-catenin signaling limits blastocyst competency for implantation. *Development* **135**, 717–727 (2008). doi: 10.1242/dev.015339; pmid: 18199579
53. M. Kumar *et al*., Germ cell specific overactivation of WNT/βcatenin signalling has no effect on folliculogenesis but causes fertility defects due to abnormal foetal development. *Sci. Rep.* **6**, 27273 (2016). doi: 10.1038/srep27273; pmid: 27265527
54. C. B. Kimmel, W. W. Ballard, S. R. Kimmel, B. Ullmann, T. F. Schilling, Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310 (1995). doi: 10.1002/aja.1002030302; pmid: 8589427
55. Y. Zhou, L. I. Zon, The zon laboratory guide to positional cloning in zebrafish. *Methods Cell Biol.* **104**, 287–309 (2011). doi: 10.1016/B978-0-12-374814-0.00016-1; pmid: 21924169
56. Z. Liu *et al*., Global identification of SMAD2 target genes reveals a role for multiple co-regulatory factors in zebrafish early gastrulas. *J. Biol. Chem.* **286**, 28520–28532 (2011). doi: 10.1074/jbc.M111.236307; pmid: 21669877
57. J. Zhang, Z. Jiang, X. Liu, A. Meng, Eph/ephrin signaling maintains the boundary of dorsal forerunner cell cluster during morphogenesis of the zebrafish embryonic left-right organizer. *Development* **143**, 2603–2615 (2016). doi: 10.1242/dev.132969; pmid: 27287807
58. W. S. Chung, D. Y. Stainier, Intra-endodermal interactions are required for pancreatic β cell induction. *Dev. Cell* **14**, 582–593 (2008). doi: 10.1016/j.devcel.2008.02.012; pmid: 18410733
59. M. Doitsidou *et al*., Guidance of primordial germ cell migration by the chemokine SDF-1. *Cell* **111**, 647–659 (2002). doi: 10.1016/S0092-8674(02)01135-2; pmid: 12464177
60. X. Zhu, Z. Min, R. Tan, Q. Tao, NF2/Merlin is required for the axial pattern formation in the Xenopus laevis embryo. *Mech. Dev.* **138**, 305–312 (2015). doi: 10.1016/j.mod.2015.08.008; pmid: 26344136

61. L. Faget, T. S. Hnasko, Tyramide Signal Amplification for Immunofluorescent Enhancement. *Methods Mol. Biol.* **1318**, 161–172 (2015). doi: 10.1007/978-1-4939-2742-5_16; pmid: 26160574
62. M. Zhang, J. Zhang, S. C. Lin, A. Meng, β-Catenin 1 and β-catenin 2 play similar and distinct roles in left-right asymmetric development of zebrafish embryos. *Development* **139**, 2009–2019 (2012). doi: 10.1242/dev.074435; pmid: 22535411
63. A. M. Hulstrand, P. N. Schneider, D. W. Houston, The use of antisense oligonucleotides in Xenopus oocytes. *Methods* **51**, 75–81 (2010). doi: 10.1016/j.ymeth.2009.12.015; pmid: 20045732
64. D. J. Olson, A. M. Hulstrand, D. W. Houston, Maternal mRNA knock-down studies: Antisense experiments using the host-transfer technique in Xenopus laevis and Xenopus tropicalis. *Methods Mol. Biol.* **917**, 167–182 (2012). doi: 10.1007/978-1-61779-992-1_10; pmid: 22956088
65. X. Liu *et al.*, Araf kinase antagonizes Nodal-Smad2 activity in mesendoderm development by directly phosphorylating the Smad2 linker region. *Nat. Commun.* **4**, 1728 (2013). doi: 10.1038/ncomms2762; pmid: 23591895
66. Y. Ding, Y. Zhang, C. Xu, Q. H. Tao, Y. G. Chen, HECT domain-containing E3 ubiquitin ligase NEDD4L negatively regulates Wnt signaling by targeting dishevelled for proteasomal degradation. *J. Biol. Chem.* **288**, 8289–8298 (2013). doi: 10.1074/jbc.M112.433185; pmid: 23396981
67. S. Jia, D. Wu, C. Xing, A. Meng, Smad2/3 activities are required for induction and patterning of the neuroectoderm in zebrafish. *Dev. Biol.* **333**, 273–284 (2009). doi: 10.1016/j.ydbio.2009.06.037; pmid: 19580801

ACKNOWLEDGMENTS

We thank S. Su and W. Wu (Tsinghua University), Y.-G. Chen (Tsinghua University), and J. Yuan (Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences) for providing experimental materials and methods. We are also grateful to other members of the Meng laboratory for help and discussion and the staff at the Cell Facility in Tsinghua Center of Biomedical Analysis for technical assistance. **Funding:** National Natural Science Foundation (31330052, 31590832, and 31571531), the Tsinghua University Initiative Scientific Research Program (20141081299), and China Postdoctoral Science Foundation (2017M620748). **Author contributions:** L.Y. found and characterized *huluwa*^{tsu01sm} mutants, generated *huluwa*^{tsu32c} mutants, performed embryonic analyses, analyzed the data, and contributed to manuscript preparation; J.C. mapped the *hwa* gene, performed embryonic and biochemical analyses, analyzed the data, and contributed to manuscript preparation; X.Z. performed all *Xenopus*-related experiments with Q.H.T.'s supervision; Q.H.T. analyzed the *Xenopus*-related data; J.S., X.W., W.S., and W.Z. helped to carry out experiments; and A.M. conceived the project, analyzed the data, and wrote the manuscript. **Competing interests:** None. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/eaat1045/suppl/DC1
Figs. S1 to S18
Tables S1 and S2
Movies S1 to S4

24 January 2018; accepted 27 September 2018
10.1126/science.aat1045

RESEARCH ARTICLE

GENETICS

Somatic mutant clones colonize the human esophagus with age

Iñigo Martincorena[1]*†, Joanna C. Fowler[1]*, Agnieszka Wabik[1], Andrew R. J. Lawson[1], Federico Abascal[1], Michael W. J. Hall[1,2], Alex Cagan[1], Kasumi Murai[1], Krishnaa Mahbubani[3], Michael R. Stratton[1], Rebecca C. Fitzgerald[2], Penny A. Handford[4], Peter J. Campbell[1,5], Kourosh Saeb-Parsy[3], Philip H. Jones[1]†

The extent to which cells in normal tissues accumulate mutations throughout life is poorly understood. Some mutant cells expand into clones that can be detected by genome sequencing. We mapped mutant clones in normal esophageal epithelium from nine donors (age range, 20 to 75 years). Somatic mutations accumulated with age and were caused mainly by intrinsic mutational processes. We found strong positive selection of clones carrying mutations in 14 cancer genes, with tens to hundreds of clones per square centimeter. In middle-aged and elderly donors, clones with cancer-associated mutations covered much of the epithelium, with *NOTCH1* and *TP53* mutations affecting 12 to 80% and 2 to 37% of cells, respectively. Unexpectedly, the prevalence of *NOTCH1* mutations in normal esophagus was several times higher than in esophageal cancers. These findings have implications for our understanding of cancer and aging.

Somatic mutations occur in healthy cells throughout life (*1*–*3*). Most of these mutations do not alter cell behavior and accumulate passively (*4*). Occasionally, however, a key gene is altered in a way that provides mutant cells with a competitive advantage, leading to the formation of persistent mutant clones. Such clones are thought to be the origin of cancer and have also been linked to other diseases (*5*, *6*). Despite the importance of somatic mutation, understanding its extent in normal tissues has been challenging because of the difficulties of identifying mutations present in small numbers of cells.

The most highly mutated normal tissue reported to date is Sun-exposed human skin. Deep targeted sequencing of Sun-exposed skin from four middle-aged individuals revealed large numbers of mutant clones under positive selection, with around a quarter of skin cells carrying cancer-driving mutations (*7*). As most mutations were caused by ultraviolet (UV) light, it is unclear whether aged Sun-exposed skin represents a special case due to a lifetime of exposure to a powerful mutagen. This question motivated us to investigate the mutational landscape of esophageal epithelium, a tissue with a similar structure but very different exposure to mutagens. Like the skin, esophageal epithelium consists of layers of keratinocytes. Cells are shed from the surface throughout life and are replaced by proliferation. In addition, both the skin and the upper and mid-esophagus develop squamous cell cancers.

We performed ultradeep targeted sequencing of 844 small samples of normal esophageal epithelium from nine deceased organ transplant donors, ranging in age from 20 to 75 years (table S1). None of the donors had a known history of esophageal or other chronic disease, and none were taking prescription medication for gastroesophageal reflux. Four of the nine donors had a history of cigarette smoking. Upper and mid-esophageal epithelium was separated from the underlying stroma and cut into a contiguous grid of 2-mm^2 samples, allowing us to map clones that spanned multiple samples (methods S1). Each sample was examined under a dissecting microscope, and no lesions were seen. Histology and whole-mount confocal imaging of adjacent tissue were also normal (figs. S1 and S2). Deep targeted sequencing of 74 cancer genes was performed on each sample to a median on-target coverage after duplicate removal of 870× (methods S2). Twenty-one samples that were found to be dominated by large clones from the targeted sequencing data were also whole-genome sequenced to a median coverage of 37×. This captures the state of the genome of the cell whose progeny subsequently colonized the sample.

Detection of mutations in normal esophagus

To detect mutations present in only a small fraction of each sample from deep targeted sequencing data, we used the ShearwaterML algorithm (*7*, *8*). This algorithm uses the observed error rates per site from a large collection of normal samples to build a site-specific error model for every type of change in every targeted site (methods S3 and fig. S3). In this dataset, we identified 8919 somatic coding mutations across 844 samples from all donors (total area, ~17 cm^2). Of these, 6935 were considered to be independent events after mutations shared by nearby samples were merged into single clones (methods S3.3 and table S2). Most sites in the genome display error rates below 1×10^{-4} errors per base, which enables accurate identification of mutations at low allele frequencies (methods S3.2 and figs. S3 and S4). The median allele frequency of the mutations detected by ShearwaterML was 1.6%, with a third of all mutations below 1% (fig. S3). As the fraction of sequencing reads that carry a mutation is a function of the fraction of mutant cells within a sample and of the local copy number, we can integrate allele frequencies and sample areas to estimate the sizes of detectable mutant clones in normal esophagus, which ranged from 0.01 mm^2 to more than 8 mm^2 (methods S5).

The number of mutations identified per sample and their allele frequencies varied markedly across individuals, with both the number of detectable mutations and the sizes of mutant clones roughly increasing with donor age (Fig. 1, A and B). To better understand the passive rate of accumulation of mutations in healthy esophagus, we can estimate the mean number of mutations per cell in each individual by integrating allele frequencies (methods S5) (*7*). These are conservative lower-bound estimates, as they are limited to mutations present in detectable clones. On average, healthy cells in the esophageal epithelium carry at least several hundred mutations per cell in people in their 20s, rising to over 2000 mutations per cell late in life (Fig. 1C). Similar estimates were obtained from the whole-genome sequencing data (methods S5.1). These estimates of the mutation rate in normal esophagus are broadly comparable to the mutation rates reported for human stem cells of the colon, small intestine, and liver by sequencing of clonal organoids (*9*).

Widespread positive selection driving clonal growth

In middle-aged individuals, the number of mutations per cell in normal esophagus is about $^1/_{10}$ that in Sun-exposed skin (*7*), a difference due partially to the high degree of UV damage sustained by the skin. Given this, we anticipated that the frequency of cancer-driver mutations in esophagus would be much lower than that in skin. Unexpectedly, however, analysis of the frequency and size of mutant clones revealed a higher density of cancer-associated mutations in normal esophagus than in Sun-exposed skin, suggesting stronger positive selection of clones with mutations in cancer-associated genes.

To formally quantify the extent of selection driving clonal expansions in normal esophagus, we estimated the ratio of nonsynonymous to

[1]Wellcome Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK. [2]MRC Cancer Unit, Hutchison-MRC Research Centre, University of Cambridge, Cambridge CB2 0XZ, UK. [3]Department of Surgery and Cambridge NIHR Biomedical Research Centre, Biomedical Campus, University of Cambridge, Cambridge CB2 2QQ, UK. [4]Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. [5]Department of Haematology, University of Cambridge, Cambridge CB2 2XY, UK.
*These authors contributed equally to this work.
†Corresponding author. Email: im3@sanger.ac.uk (I.M.); pj3@sanger.ac.uk (P.H.J.)

synonymous mutation rates (dN/dS) across genes, which is a widely used measure of selection. We used the dNdScv model, an implementation of dN/dS for somatic data that controls for trinucleotide mutational signatures, sequence composition, and variable mutation rates across genes (*4*) (methods S6). This method has been shown to reliably identify genes under positive selection in cancer and normal tissues (*4*, *7*). In the context of this experiment, dN/dS ratios reveal how much more (or less) likely it is for a nonsynonymous mutation to reach a detectable clone size than a synonymous mutation (methods S6.2).

This analysis revealed strong evidence of selection driving clonal expansions in normal esophagus. At the gene level, we detected significant positive selection in 14 of the genes that we sequenced (Fig. 2, A to D, and table S3). This means that mutation of these genes confers a competitive advantage on mutant cells relative to neighboring cells. Sorted by mutation frequency, the list comprises *NOTCH1*, *TP53*, *NOTCH2*, *FAT1*, *NOTCH3*, *ARID1A*, *KMT2D*, *CUL3*, *AJUBA*, *PIK3CA*, *ARID2*, *TP63*, *NFE2L2*, and *CCND1*. Notably, the five most frequently mutated genes in normal esophagus also dominated the mutational landscape in Sun-exposed skin. Many of the positively selected genes play a role in keratinocyte differentiation through NOTCH signaling [*NOTCH1*, *NOTCH3*, and *TP53* (*10*–*13*)], through redox cellular stress [*NFE2L2*, *TP63*, and *CUL3* (*14*–*17*)], or through epigenetic regulation [*KMT2D* (*18*)]. Tilting cell fate balance away from differentiation toward proliferation may confer a competitive advantage on mutant cells in normal esophageal epithelium (*19*).

At least 11 of the 14 genes found under positive selection in normal esophagus are canonical drivers of esophageal squamous cell carcinomas (ESCCs) (methods S6.5) (*20*–*22*). Their presence in normal epithelium suggests that they act as early ESCC drivers, leading to the expansion of persisting clones that may undergo further mutation and malignant transformation. The landscape of selection in normal squamous epithelium of the esophagus more closely resembles that of ESCCs than that of esophageal adenocarcinomas (EACs) (*21*), consistent with the typical development of ESCCs from the squamous epithelium of the upper and mid-esophagus. By contrast, EACs evolve from epithelium close to the stomach junction and are associated with Barrett's metaplasia.

Colonization of the epithelium by *NOTCH1* mutant clones

One unexpected observation was the very high prevalence of *NOTCH1* mutations in normal esophagus (Fig. 2, A to C). Across the nine donors, we detected 2055 coding mutations in *NOTCH1*, of which more than 98% were nonsynonymous, with an average of ~120 different *NOTCH1* mutations per square centimeter of normal esophagus (Fig. 2A). *NOTCH1* acts as an oncogene in different leukemias but has a mutation pattern consistent with a tumor suppressor gene in squamous cell carcinomas (SCCs) of the skin, head and neck, esophagus, and lung (*23*). As in SCCs, mutations in *NOTCH1* in normal esophagus were enriched for truncating mutations (dN/dS > 50), including stop-gains, essential splice site mutations, and indels (Fig. 2B). Missense mutations were also frequent in *NOTCH1*, and they were concentrated in 5 of the 36 extracellular epidermal growth factor (EGF) repeat domains, EGF8 to EGF12 (Fig. 2E). These EGF repeats contain the binding domains for the Notch1 ligands Jagged and Delta. The most recurrent codon alterations occurred at sites predicted to affect structural residues (calcium-binding motifs, cysteine residues, and interdomain packing residues) or the contact surface with Notch1 ligands (Fig. 2F and supplementary materials) (*23*, *24*). The large number of positively selected *NOTCH1* mutations provides structural and functional insights into this key regulatory protein.

By integrating the allele fractions of the mutations and allowing for the possibility that

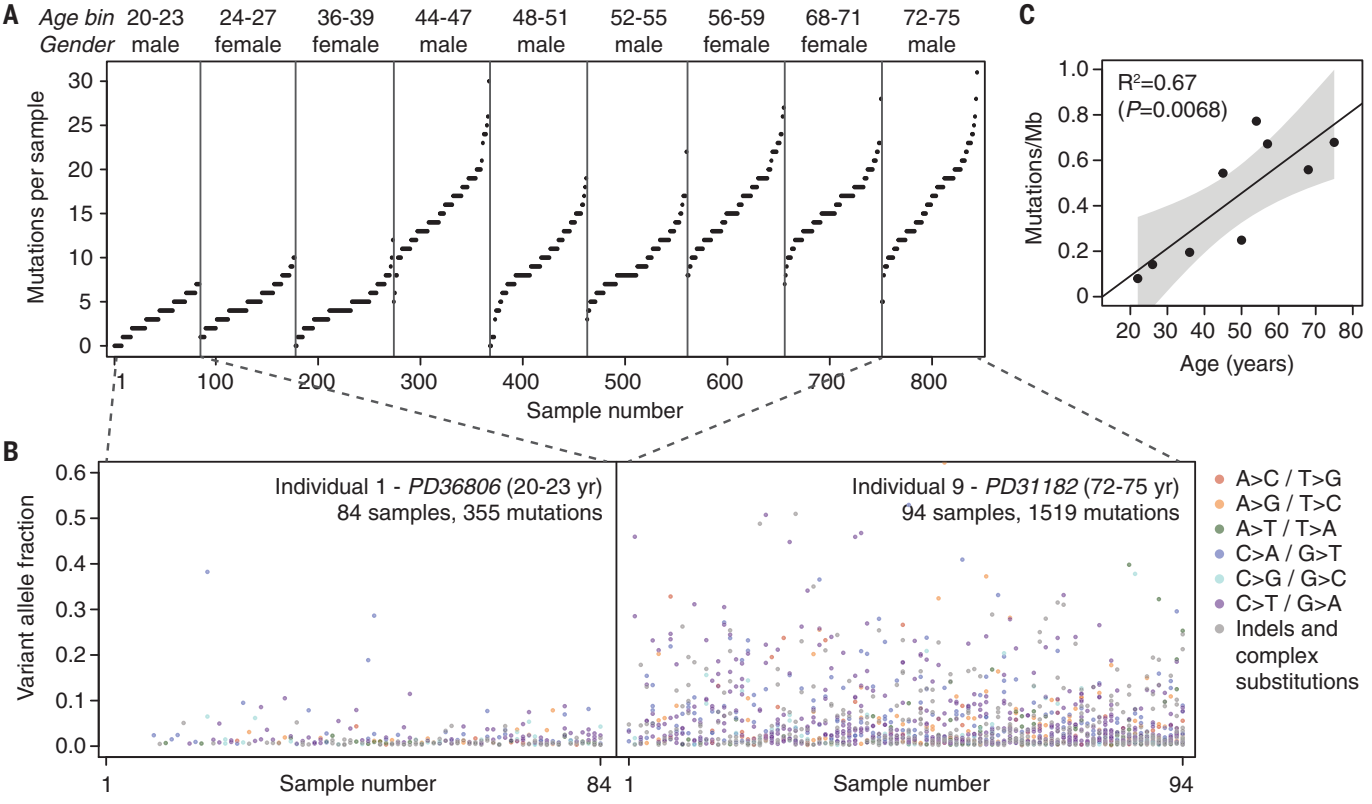


Fig. 1. Detection of somatic mutations in normal esophagus. (**A**) Number of mutations detected per sample across the 844 samples from the nine transplant donors (sorted by age). Donor age is shown in 4-year bins to increase sample anonymity. (**B**) Variant allele fractions (VAFs) for the mutations detected in the youngest and oldest donors, colored by mutation type. The VAF is the fraction of sequencing reads reporting a mutation within a sample. (**C**) Scatter plot of donor age and the estimated mean mutation burden per cell for each donor. The fitted line, R^2 value, and P value were obtained by linear regression.

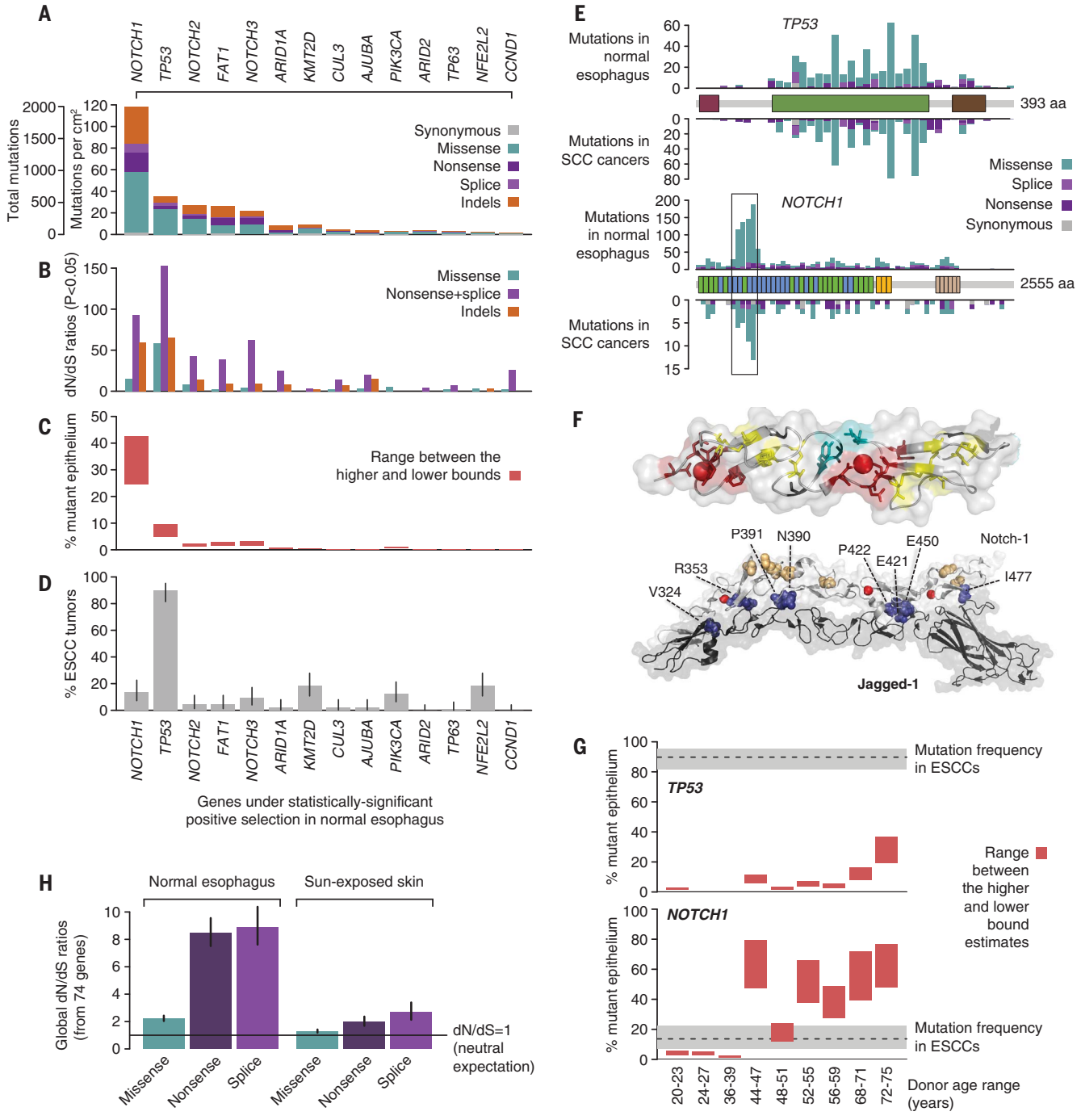


Fig. 2. Widespread positive selection of cancer-associated mutations in normal esophagus. (**A**) Number of mutations detected in each of the 14 genes found under positive selection. (**B**) Observed-to-expected ratios for missense substitutions, truncating (nonsense and essential splice site) substitutions, and indels. Observed-to-expected ratios for substitutions are dN/dS ratios. Only ratios with $P < 0.05$ are shown. (**C**) Estimated percentage of cells carrying a mutation in each gene (methods S5.3). (**D**) Percentage of ESCCs with a nonsynonymous substitution or an indel in each gene. Error bars depict 95% Poisson CIs. (**E**) Distribution of mutations within *TP53* and *NOTCH1* in normal esophagus (above the gene domain diagram) and in SCC cancers from The Cancer Genome Atlas (below). The region of EGF8 to EGF12 is boxed. aa, amino acids. (**F**) Consequences of *NOTCH1* missense mutations. (Top) Most *NOTCH1* missense mutations affect structural residues in EGF domains (shown in stick form) [Protein Data Bank (PDB) code 2VJ3]: calcium-binding consensus residues (red); hydrophobic interdomain packing residues (teal); cysteine residues, which form disulfide bonds (yellow); and conserved glycines (black). Calcium ions are shown as red spheres. (Bottom) Other residues affected by missense mutations (≥4 per residue) in the EGF8-to-EGF12 region are shown in space-filling representation. Many are predicted to disrupt the Notch receptor-ligand binding interface (shown in deep blue and labeled by residue number), whereas others are distal (colored wheat) (PDB code 5UK5). Single-letter abbreviations for the amino acid residues are as follows: E, Glu; I, Ile; N, Asn; P, Pro; R, Arg; and V, Val. (**G**) Estimated percentage of mutant epithelium per donor compared with ESCC mutation frequency. (**H**) dN/dS values estimated from all 74 target genes together in normal esophagus and Sun-exposed skin (*7*). Error bars depict 95% CIs.

mutations may affect one or two alleles per cell, we can estimate the fraction of mutant cells in a tissue for any given gene (methods S5.3). On average across the nine donors, 25 to 42% of the cells in normal esophagus harbored *NOTCH1* mutations (Fig. 2C). The frequency of *NOTCH1* mutant clones showed a large increase with age. About 30 to 80% of the cells in normal esophagus were *NOTCH1* mutated in five of the six middle-aged or elderly individuals, compared with 1 to 6% in the three individuals under 40 years of age (Fig. 2G). This observation is consistent with data from experimental mouse models showing that transgenic inhibition of Notch signaling in a small fraction of cells confers clonal advantage and enables these clones to colonize the normal esophageal epithelium (*19*, *25*).

This observation has potentially important implications. The *NOTCH1* gene has been widely assumed to be a driver in ESCCs because it is mutated in ~10% of tumors (*21*, *26*) (Fig. 2, D and G). The observation that, in middle-aged individuals, *NOTCH1* is typically mutated in 30 to 80% of the normal esophageal epithelium suggests that *NOTCH1* mutations are less frequent in cancers than in the background of normal tissue from which the cancers develop. This raises questions about the role of *NOTCH1* in the development of ESCCs.

The case of *NOTCH1* contrasts with that of *TP53* (Fig. 2G), which is mutated in more than 90% of ESCCs but in a minority of cells in the normal esophageal epithelium. *TP53* is the second most frequently mutated gene in normal esophagus, with ~35 mutations per square centimeter and strong positive selection for both truncating and missense mutations (dN/dS ratios ~150 and ~50, respectively) (Fig. 2, A and B). As in cancer genomes, the missense mutations affect mostly the central DNA binding domain (Fig. 2E). Across the nine donors, 5 to 10% of the epithelium carried a *TP53* mutation, a fraction that appeared to increase with age, with the oldest donor having *TP53* mutations in 20 to 35% of cells (Fig. 2G).

In summary, we found an unexpectedly high density of driver mutations in normal esophagus and positive selection acting on most of the main drivers of ESCC. By combining the 74 genes studied, global dN/dS ratios for missense and protein-truncating (nonsense and essential splice site) mutations were ~2.2 and ~8.6, respectively, with the enrichment of nonsynonymous mutations increasing rapidly with clone size (Fig. 2H and fig. S5B). This suggests that approximately 55% of all missense mutations and 88% of all truncating mutations identified in this dataset were actively driven to detectable clone sizes by positive clonal selection. Overall, by using dN/dS ratios and considering substitutions and indels

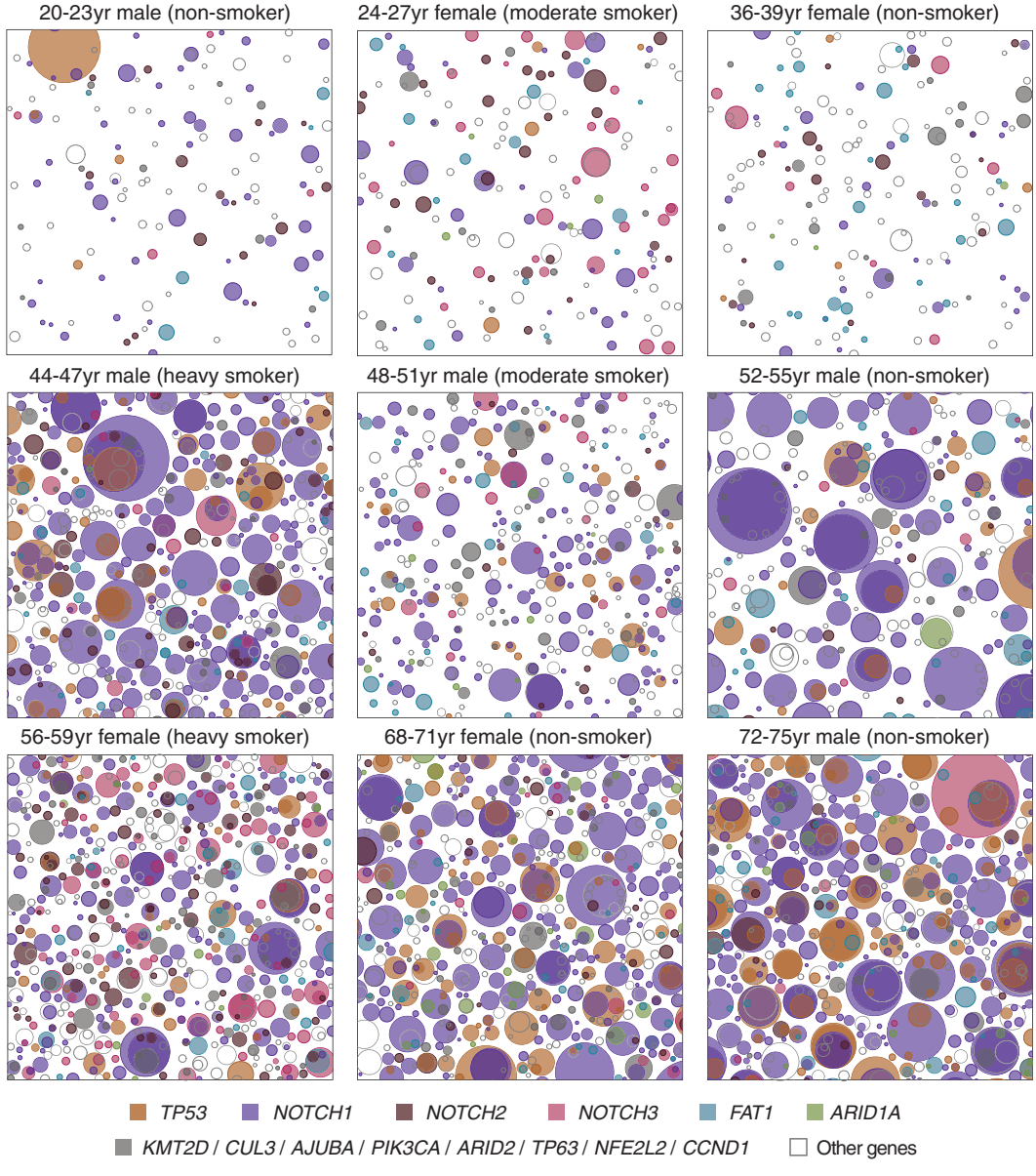


Fig. 3. Variation of the mutational landscape across the nine donors. Representative patchwork plots from each donor. Each panel is a schematic representation of the mutant clones in an average 1-cm^2 area of normal esophageal epithelium from each donor. To generate each figure, a number of samples from the donor were randomly selected to amount to 1 cm^2 of tissue, and all clones detected are represented as circles randomly distributed in space. The density and size of the clones are inferred from the sequencing data, and the nesting of clones and subclones is inferred from the data when possible and randomly allocated otherwise (methods S5.4).

in the 14 genes under significant selection, we estimate that there are 3915 [95% confidence interval (CI), 3829 to 3988) positively selected driver mutations in the ~17 cm^2 of normal esophageal epithelium sequenced in this study, of which 52% are in genes other than *NOTCH1* (methods S6.3). This number is comparable to the yield of driver mutations obtained from sequencing more than 1000 cancer genomes (*4*).

Variation of the mutational and selective landscape across donors

The patterns of somatic evolution varied greatly across the nine individuals in this study, with large differences in mutation density, clone sizes,

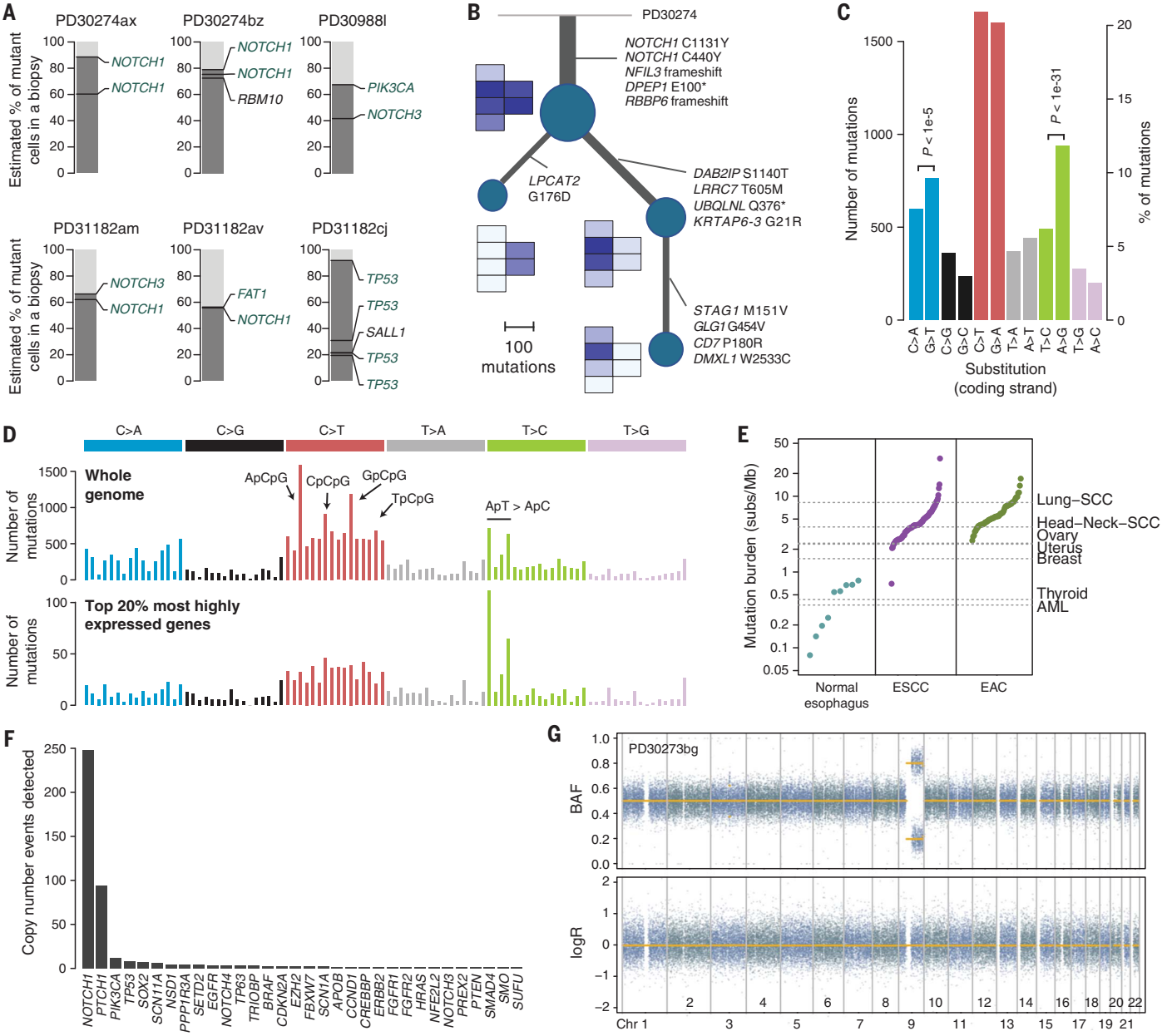


Fig. 4. Phylogenetic and mutational patterns in normal esophagus. (**A**) Representation of mutations co-occurring in the same clones by using the pigeonhole principle (see supplementary materials 5.5). (**B**) Phylogenetic reconstruction of the evolution of a large clone overlapping six samples by using whole-genome sequencing data and spatial information. A small heatmap of the six affected samples is shown next to each node in the tree, depicting the mean VAF for the mutations in each node. Single-letter abbreviations for the amino acid residues are as follows: C, Cys; D, Asp; G, Gly; M, Met; Q, Gln; S, Ser; T, Thr; W, Trp; and Y, Tyr. (**C**) Number of substitutions per mutation type as mapped to the coding (untranscribed) strand from all donors. *P* values reflect transcription strand asymmetry (exact Poisson test). (**D**) Ninety-six–mutation–class bar plot depicting the number of mutations in each of the possible 96 trinucleotides (strand independent). (Top) Whole-genome plot aggregating all 21 whole genomes. (Bottom) Spectrum for mutations occurring in the transcribed region of the top 20% most highly expressed genes. (**E**) Mutation burden in normal esophagus and in ESCC and EAC tumors (every point corresponds to a donor, sorted by mutation burden). AML, acute myeloid leukemia. (**F**) Number of copy number events detected in each gene across the 844 samples by using the targeted data. (**G**) Representative log *R* ratio and B-allele frequency (BAF) scatter plots for heterozygous SNPs from whole-genome data showing a copy-neutral LOH event affecting *NOTCH1* (sample *PD30273bg* is shown). Chr 1, chromosome 1.

and overall driver frequency (Fig. 3). Age is by far the strongest risk factor in ESCC, with cancer incidence rising near-geometrically with age (*27*, *28*). We used mixed-effect regression models to evaluate the association between the mutation landscape and age while controlling for other risk factors, such as gender and smoking status (methods S7). Despite the modest cohort size, this analysis revealed a significant increase in the number of mutations per sample (P = 0.009) and clone sizes (P = 0.027) with age. This is consistent with the significant increase in the mutation burden with age depicted in Fig. 1C and determined by standard linear regression (P = 0.0068; coefficient of determination R^2 = 0.67). We also noted that the two heavy smokers in the cohort could have a higher number of mutations than expected for their age (Figs. 1A and 3 and methods S7). However, larger cohorts will be needed to reliably study the effects of behavioral risk factors on the mutational landscape in the esophagus.

Despite the dominant effect of age, we found unexplained differences across individuals, including differences in the strength of selection on different genes across individuals, as suggested by the differently colored clones in Fig. 3. To formally quantify differences in selection pressure per gene across donors while removing the effect of variable mutation rates and signatures across individuals, we used an extension of dNdScv that compares two dN/dS ratios (methods S6.4). This confirmed significant differences in the driver landscape across donors (fig. S5, C to E). For example, across individuals, *NOTCH1* is mutated five times as frequently as *NOTCH3*. Yet, in one donor, we detected nearly the same number of mutations in *NOTCH1* and *NOTCH3* (fig. S5D) ($q = 5 \times 10^{-13}$, likelihood-ratio test). Similarly, the oldest donor showed a twofold relative enrichment in *TP53* mutations compared with other individuals ($q < 1 \times 10^{-15}$, likelihood-ratio test) (fig. S5E), consistent with the observation that 20 to 37% of normal esophageal epithelium was *TP53* mutated in this donor (Fig. 2G). Whether the variation in the driver landscape across donors reflects differences in exposure to environmental factors, the genetic background of each individual, or both is unclear. Nevertheless, differences in mutation rates, clone sizes, and driver preferences may have implications for understanding interindividual variation in cancer risk.

Given the large increase in driver mutant clones with age, many clones are expected to acquire more than one driver mutation over the course of a lifetime. Although the small clone sizes limit our ability to determine which mutations within a sample occur in the same cells, 25 samples had sufficiently large clones for us to confidently group mutations (methods S5.5) (Fig. 4A and fig. S6). Most cases (14 of 25) were examples of *NOTCH1* biallelic inactivation by two mutations. We also observed examples of clones carrying mutations in *NOTCH1* and *FAT1*, *NOTCH1* and *NOTCH3*, and *PIK3CA* and *NOTCH3*. In the oldest donor (in the age range from 72 to 75 years), whose samples showed an enrichment of *TP53* mutations, we found a large clone, measuring >4 mm^2, with a founder heterozygous *TP53* mutation and three separate subclones each carrying a different second *TP53* mutation (Fig. 4A). For a large clone extending over six samples and measuring >8.5 mm^2, we were able to integrate whole-genome data and spatial information to reconstruct the clone's phylogenetic history (methods S5.6). The tree shows that the ancestor cell underwent a large clonal expansion after losing both copies of *NOTCH1*, followed by branching evolution with two subclones dominating spatially distinct areas (Fig. 4B).

The whole-genome mutational landscape in normal esophagus

To better understand the contribution of different mutational processes and the extent of structural variation in normal esophagus, we performed whole-genome sequencing of 21 samples dominated by a major clone. Across all donors, C→T or G→A (C>T/G>A) mutations dominate the spectra, with a clear excess of mutations at CpG dinucleotides (Fig. 4, C and D, and fig. S7). These changes result from the deamination of 5-methylcytosine into thymine and are believed to occur spontaneously throughout life (*29*, *30*). Signature analysis revealed that the pattern of mutations largely resembles a combination of COSMIC (Catalogue of Somatic Mutations in Cancer) mutational signatures 1 and 5 (*30*) (methods S4 and figs. S7 and S8). Both signatures have been shown to dominate the accumulation of mutations in normal tissues such as colon, small intestine, and liver during life (*9*).

In addition, we observed two other mutational processes. There was a considerable rate of C>A/G>T changes with a modest but significant transcription-strand bias. Multiple mechanisms can lead to these types of changes, including smoking. Although four of the nine individuals were smokers, we did not observe a clear signature of tobacco-induced mutations (COSMIC signature 4) (methods S4). We also observed considerable variation in T>C changes across the 21 whole genomes, with a strong transcription-strand asymmetry (Fig. 4D and figs. S7 and S8). Stratification of the mutation spectra by gene expression level revealed that highly transcribed genes are targets of a process of transcription-coupled mutagenesis that induces T>C changes preferentially at ApT sites in the transcribed strand, a phenomenon previously described in liver cancers (*31*) (related to COSMIC signature 16) (Fig. 4D and fig. S8).

Overall, most mutations seem to be generated by intrinsic mutational processes associated with age or transcription (*30*, *31*), without clear evidence of external mutagenic processes. We found no evidence of COSMIC signatures 2 and 13 in the targeted or the whole-genome data. These signatures are believed to be caused by APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide–like) cytidine deaminases and contribute large numbers of mutations in esophageal cancers (*20*, *21*, *32*). This partially explains the observation that the mutation burden in normal esophagus is about an order of magnitude lower than the median mutation burden of ESCC and EAC cancers (Fig. 4E). The rarity of APOBEC mutagenesis in normal esophagus may suggest that this is acquired later in the evolution of ESCC or that ESCCs are more likely to evolve from rare clones displaying APOBEC mutagenesis.

Esophageal cancers are characterized by large numbers of copy number changes and structural rearrangements (*21*, *33*). To explore the extent of copy number changes in normal esophagus, we first analyzed the deep targeted sequencing data. We used a copy number detection algorithm designed to identify low-frequency subclonal loss of heterozygosity (LOH) on targeted data, exploiting the statistical phasing of heterozygous single-nucleotide polymorphisms (SNPs) to detect small allelic imbalances (*7*) (methods S3.4). *NOTCH1* loss was the most frequent copy number change identified, although caution must be exercised because statistical power varies across genes and donors. *NOTCH1* LOH was detected in nearly 30% of all samples (Fig. 4F) and in virtually all of the samples with a single high-frequency *NOTCH1* mutation, confirming that the loss of *NOTCH1* is typically biallelic. *PTCH1* sits on the same arm of chromosome 9 as *NOTCH1* and is often lost together with *NOTCH1*. We also detected less frequent but recurrent whole–chromosome 3 gains, which lead to the duplication of *PIK3CA/SOX2/TP63*, an event observed in approximately half of ESCCs (*21*) (Fig. 4F). Several instances of *TP53* LOH were also detected, largely concentrated in the oldest donor.

Copy number analysis of the 21 whole genomes confirmed that segmental loss of *NOTCH1* is typically mediated by copy-neutral LOH without detectable rearrangements (Fig. 4G, fig. S9, methods S3.5.2, and table S4). Such events may be generated by mitotic homologous recombination. Events varied in size, from whole-arm losses to focal events (Fig. 4G and figs. S9 and S10). With the exception of copy-neutral LOH events in *NOTCH1* and an instance of chromosome 3 gain, the 21 genomes appeared largely diploid, without evidence of other copy number changes that may be expected to accumulate by chance over time (Fig. 4G, fig. S9, methods S3.5.2, and table S4). The rarity of copy number changes in large clones, none of which had *TP53* mutations, suggests that the background rate of copy number changes is low in normal cells of the esophagus or that such changes are negatively selected. Either way, this represents a major difference between normal esophageal cells and ESCCs, suggesting that structural changes may occur late in the evolution of esophageal cancers (*33*).

Discussion

These data have unveiled a hidden world of somatic mutation and clonal competition in

normal esophagus. We have detected thousands of mutations per cell, hundreds of positively selected clones per square centimeter, and clones with cancer-associated mutations colonizing most of the esophageal epithelium with age, all without grossly detectable changes in histology.

The higher frequency of cancer-associated mutations in normal esophagus than in Sun-exposed skin is unexpected, particularly given the lower mutation rate in the esophagus. Although we found most of the common drivers of ESCC already under selection in normal esophageal epithelium, key differences remain between the genomes of cells in mutant clones in aging normal epithelium and those of cancer cells. These include a mutation burden in normal epithelium about $^{1}/_{10}$ that in many ESCCs, no evidence of APOBEC mutagenesis, and an apparent lack of chromosomal instability. Further, although clones carrying cancer-driver mutations are widespread, the average number of driver mutations per cell in normal esophagus is much lower than that in cancer cells (Fig. 2C), a result consistent with the multistage theory of carcinogenesis (*27*, *28*, *34*). Larger-scale genomic studies of normal tissues in healthy individuals and of premalignant lesions of different grades will help refine our understanding of the transition from normal cells to cancer (*3*, *28*, *33*, *35*).

An unexpected observation is the high frequency of *NOTCH1* mutation in aged normal esophagus compared with ESCCs. This may suggest that ESCCs are more likely to evolve from cells in the epithelium without *NOTCH1* mutations. By contrast, *TP53* mutations, which are less frequent than *NOTCH1* mutations, are almost ubiquitous in ESCCs, suggesting that cancers arise from the small fraction of *TP53* mutant cells. Cancer risk may therefore vary across the aging epithelium, depending on the colonizing mutations present. Interventions that decrease the proportion of mutant cells at a higher risk of transformation in normal epithelium may thus be beneficial.

We note that, even if they do not contribute to carcinogenesis, drivers of benign clonal expansions may still appear as recurrently mutated genes in cancer genomes, owing to their high mutation frequency in the normal cells from which tumors evolve. Better understanding of the mutational landscape in normal tissues may thus help refine current catalogs of cancer-driver genes, with important implications for early diagnosis and targeted therapy.

Positive selection of mutant clones has now been observed during normal aging in blood, Sun-exposed skin, and esophageal epithelium (*7*, *36*). This opens the theoretical possibility of clonal selection across tissues as a contributing factor in tissue and organismal aging (*4*, *37*, *38*). Somatic mutation has long been recognized as a possible factor contributing to aging, with mutations and other forms of damage deleterious to the carrying cells passively accumulating during life and progressively reducing cellular fitness (*39*). Widespread positive selection of mutant clones may be an additional contributory factor in aging, as it can greatly accelerate the accumulation of functional mutations and altered phenotypes. Throughout life, somatic mutations increasing cellular fitness can spread and even dominate tissues, independently of their cost to the organism. If the selected mutations negatively affect tissue function, the physiological integrity of the organism will decline, a hallmark of the aging process.

This study emphasizes how little we know about somatic evolution within normal tissues, a fundamental process that is likely to take place to varying degrees in every tissue of every species. Better understanding of the extent of somatic mutation and selection across tissues in health and disease promises to provide insights into the origins of cancer and aging.

REFERENCES AND NOTES

1. M. R. Stratton, P. J. Campbell, P. A. Futreal, *Nature* **458**, 719–724 (2009).
2. B. Vogelstein *et al.*, *Science* **339**, 1546–1558 (2013).
3. L. M. Merlo, J. W. Pepper, B. J. Reid, C. C. Maley, *Nat. Rev. Cancer* **6**, 924–935 (2006).
4. I. Martincorena *et al.*, *Cell* **171**, 1029–1041.e21 (2017).
5. M. S. Anglesio *et al.*, *N. Engl. J. Med.* **376**, 1835–1848 (2017).
6. S. Jaiswal *et al.*, *N. Engl. J. Med.* **377**, 111–121 (2017).
7. I. Martincorena *et al.*, *Science* **348**, 880–886 (2015).
8. M. Gerstung, E. Papaemmanuil, P. J. Campbell, *Bioinformatics* **30**, 1198–1204 (2014).
9. F. Blokzijl *et al.*, *Nature* **538**, 260–264 (2016).
10. S. Ohashi *et al.*, *Gastroenterology* **139**, 2113–2123 (2010).
11. K. Sakamoto *et al.*, *Lab. Invest.* **92**, 688–702 (2012).
12. K. Lefort *et al.*, *Genes Dev.* **21**, 562–577 (2007).
13. T. Yugawa *et al.*, *Mol. Cell. Biol.* **27**, 3732–3742 (2007).
14. A. Bhaduri *et al.*, *Dev. Cell* **35**, 444–457 (2015).
15. R. B. Hamanaka *et al.*, *Sci. Signal.* **6**, ra8 (2013).
16. N. Wakabayashi *et al.*, *Nat. Genet.* **35**, 238–245 (2003).
17. A. Kobayashi *et al.*, *Mol. Cell. Biol.* **24**, 7130–7139 (2004).
18. A. S. Hopkin *et al.*, *PLOS Genet.* **8**, e1002829 (2012).
19. M. P. Alcolea *et al.*, *Nat. Cell Biol.* **16**, 612–619 (2014).
20. L. Zhang *et al.*, *Am. J. Hum. Genet.* **96**, 597–611 (2015).
21. Cancer Genome Atlas Research Network; Analysis Working Group: Asan University; BC Cancer Agency; Brigham and Women's Hospital; Broad Institute; Brown University; Case Western Reserve University; Dana-Farber Cancer Institute; Duke University; Greater Poland Cancer Centre; Harvard Medical School; Institute for Systems Biology; KU Leuven; Mayo Clinic; Memorial Sloan Kettering Cancer Center; National Cancer Institute; Nationwide Children's Hospital; Stanford University; University of Alabama; University of Michigan; University of North Carolina; University of Pittsburgh; University of Rochester; University of Southern California; University of Texas MD Anderson Cancer Center; University of Washington; Van Andel Research Institute; Vanderbilt University; Washington University; Genome Sequencing Center: Broad Institute; Washington University in St. Louis; Genome Characterization Centers: BC Cancer Agency; Broad Institute; Harvard Medical School; Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins University; University of North Carolina; University of Southern California Epigenome Center; University of Texas MD Anderson Cancer Center; Van Andel Research Institute; Genome Data Analysis Centers: Broad Institute; Brown University; Harvard Medical School; Institute for Systems Biology; Memorial Sloan Kettering Cancer Center; University of California Santa Cruz; University of Texas MD Anderson Cancer Center; Biospecimen Core Resource: International Genomics Consortium; Research Institute at Nationwide Children's Hospital; Tissue Source Sites: Analytic Biologic Services; Asan Medical Center; Asterand Bioscience; Barretos Cancer Hospital; BioreclamationIVT; Botkin Municipal Clinic; Chonnam National University Medical School; Christiana Care Health System; Cureline; Duke University; Emory University; Erasmus University; Indiana University School of Medicine; Institute of Oncology of Moldova; International Genomics Consortium; Invidumed; Israelitisches Krankenhaus Hamburg; Keimyung University School of Medicine; Memorial Sloan Kettering Cancer Center; National Cancer Center Goyang; Ontario Tumour Bank; Peter MacCallum Cancer Centre; Pusan National University Medical School; Ribeirão Preto Medical School; St. Joseph's Hospital & Medical Center; St. Petersburg Academic University; Tayside Tissue Bank; University of Dundee; University of Kansas Medical Center; University of Michigan; University of North Carolina at Chapel Hill; University of Pittsburgh School of Medicine; University of Texas MD Anderson Cancer Center; Disease Working Group: Duke University; Memorial Sloan Kettering Cancer Center; National Cancer Institute; University of Texas MD Anderson Cancer Center; Yonsei University College of Medicine; Data Coordination Center: CSRA Inc.; Project Team: National Institutes of Health, *Nature* **541**, 169–175 (2017).
22. G. Sawada *et al.*, *Gastroenterology* **150**, 1171–1182 (2016).
23. C. S. Nowell, F. Radtke, *Nat. Rev. Cancer* **17**, 145–159 (2017).
24. V. C. Luca *et al.*, *Science* **355**, 1320–1324 (2017).
25. M. P. Alcolea, P. H. Jones, *Cell Cycle* **14**, 9–17 (2015).
26. Y. Song *et al.*, *Nature* **509**, 91–95 (2014).
27. P. Armitage, R. Doll, *Br. J. Cancer* **8**, 1–12 (1954).
28. I. Martincorena, P. J. Campbell, *Science* **349**, 1483–1489 (2015).
29. L. B. Alexandrov *et al.*, *Nature* **500**, 415–421 (2013).
30. L. B. Alexandrov *et al.*, *Nat. Genet.* **47**, 1402–1407 (2015).
31. N. J. Haradhvala *et al.*, *Cell* **164**, 538–549 (2016).
32. J. Chang *et al.*, *Nat. Commun.* **8**, 15290 (2017).
33. C. S. Ross-Innes *et al.*, *Nat. Genet.* **47**, 1038–1046 (2015).
34. D. E. Brash, *Science* **348**, 867–868 (2015).
35. P. Martinez *et al.*, *Nat. Commun.* **9**, 794 (2018).
36. G. Genovese *et al.*, *N. Engl. J. Med.* **371**, 2477–2487 (2014).
37. J. M. Smith, *Proc. R. Soc. London Ser. B* **157**, 115–127 (1962).
38. R. A. Risques, S. R. Kennedy, *PLOS Genet.* **14**, e1007108 (2018).
39. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, *Cell* **153**, 1194–1217 (2013).

ACKNOWLEDGMENTS

We are very grateful to the families of deceased donors for their consent and to the Cambridge Biorepository for Translational Medicine for access to human tissue. **Funding:** I.M. is funded by Cancer Research UK (C57387/A21777). P.J.C. is a Wellcome Trust Senior Clinical Fellow. This work was funded by a Cancer Research UK program grant to P.H.J. (C609/A17257), an MRC Centenary grant, Wellcome Trust core funding to the Wellcome Sanger Institute, and an MRC grant-in-aid to the MRC Cancer Unit. **Author contributions:** P.H.J. initiated the project. P.H.J. and J.C.F. designed the experiments. I.M. led data analysis with help from A.R.J.L., F.A., A.C., and M.W.J.H. and advice from M.R.S. and P.J.C. P.A.H. analyzed the structural implications of *NOTCH1* mutations. K.S.-P. and K.Ma. collected the samples. J.C.F., A.W., and K.Mu. performed experiments. R.C.F. contributed to a pilot study. I.M., J.C.F., and P.H.J. wrote the paper. **Competing interests:** M.R.S. is on the scientific advisory board of GRAIL. The other authors declare no competing interests. **Data and materials availability:** Sequencing data are deposited in the European Genome-phenome Archive (EGA) under accession numbers EGAD00001004158 and EGAD00001004159.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/911/suppl/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S4
References (*40*–*49*)

8 June 2018; accepted 3 October 2018
Published online 18 October 2018
10.1126/science.aau3879

REPORT

MASS SPECTROMETRY

Neutral mass spectrometry of virus capsids above 100 megadaltons with nanomechanical resonators

Sergio Dominguez-Medina[1,2,3], Shawn Fostner[4]*, Martial Defoort[4], Marc Sansa[4], Ann-Kathrin Stark[1,2,3], Mohammad Abdul Halim[1,2,3], Emeline Vernhes[5]†, Marc Gely[4], Guillaume Jourdan[4], Thomas Alava[4], Pascale Boulanger[5], Christophe Masselon[1,2,3]‡, Sébastien Hentz[4]‡

Measurement of the mass of particles in the mega- to gigadalton range is challenging with conventional mass spectrometry. Although this mass range appears optimal for nanomechanical resonators, nanomechanical mass spectrometers often suffer from prohibitive sample loss, extended analysis time, or inadequate resolution. We report on a system architecture combining nebulization of the analytes from solution, their efficient transfer and focusing without relying on electromagnetic fields, and the mass measurements of individual particles using nanomechanical resonator arrays. This system determined the mass distribution of ~30-megadalton polystyrene nanoparticles with high detection efficiency and effectively performed molecular mass measurements of empty or DNA-filled bacteriophage T5 capsids with masses up to 105 megadaltons using less than 1 picomole of sample and with an instrument resolution above 100.

The current lower limit to weighing individual objects is in the picogram range and is obtained with piezoelectric resonators (*1*). Twelve orders of magnitude lower on the mass scale, mass spectrometry (MS) uses ionization, electromagnetic fields to manipulate ions, and ensemble averaging of mass-to-charge ratios to identify species on the basis of their specific molecular mass. Since the advent of soft ionization techniques (*2*, *3*), commercial spectrometers have performed routine proteomics analysis of peptides and proteins in the range of 1 to 100 kDa (1 kDa = 1.66×10^{-24} kg = 1.66×10^{-9} pg). Recent MS research has expanded the mass range accessible with conventional techniques, up to the 10-MDa range (*4*). Alternative MS techniques, such as charge-detection MS (CDMS) (*5*, *6*), can analyze single ions with masses up to 1 GDa. Nevertheless, unmodified commercially available MS instruments measure masses only up to 1 MDa.

Nanomechanical resonators measure the mass of individual particles accreting on their surface. Because the frequency-to-mass relation scales with the resonator's characteristic dimension d^{-4}, researchers in the field of nanomechanics have engaged in a race to measure the smallest detectable mass with ever smaller resonators (*7*–*10*). These efforts were driven by the hope of competing with commercial MS down to the Dalton mass range, whereas the MS community has worked in the opposite direction to measure larger masses. Nanomechanical resonators fabricated by scalable silicon processes are actually ideally suited for analysis of masses in the mega- to gigadalton range, which includes most viruses, many disease biomarkers, and species with defined molecular mass.

The analysis of viruses by nanomechanical MS has been pursued for several years (*11*). Unfortunately, early attempts to perform nanomechanical MS were hampered by a combination of losses associated with ionization yield and ion transfer, and the small cross-section (typically a few square micrometers) of the capture area on nanoresonators. These issues led to prohibitively long analysis times and excessive sample consumption (*12*, *13*). However, nanomechanical MS of metallic nanoparticles of any charge state (ionized or neutral) (*14*) was demonstrated, and recently, the possibility of measuring particle stiffness simultaneously with mass and position was demonstrated with a system that operates without ion guides in the tens to hundreds of gigadaltons range (*15*).

To achieve the full potential of nanomechanical MS in the MDa to GDa range, we report here on a system architecture specifically designed for this mass range (Fig. 1) that features high efficiency and excellent resolution by circumventing the requirement for ionization. It relies on nebulization of analytes from solution at atmospheric pressure, exploits the particles' inertia to efficiently transfer and focus them without any need for electromagnetic fields, and determines the mass of individual particles by using arrays of nanomechanical resonators in high vacuum.

Nebulization of particles was performed by surface acoustic wave nebulization (SAWN) (*16*) or by electrospray ionization (ESI) (*4*). With SAWN, a liquid sample deposited on a piezoelectric surface is nebulized as a thin mist upon propagation of an acoustic wave across the surface. Compared to ESI, which produces a rapidly expanding jet through Coulomb repulsion, SAWN produces droplets with lower kinetic energy, resulting in very efficient uptake into the system's inlet capillary. Nevertheless, ESI was also used as spraying conditions have been extensively optimized for numerous species of interest, including viruses (*4*).

The small capture area of nanoresonators necessitates a low particle-beam–to–detector-size ratio for efficient particle detection. We used an aerodynamic lens composed of a pressure-limiting orifice followed by a series of apertures between relaxation volumes (Fig. 1). We designed and optimized this lens for the mass range of interest (MDa to GDa) following previously reported guidelines (*17*). Compared to conventional ion guides, which counteract inertia to maintain heavy particles on a stable trajectory, an aerodynamic lens exploits the particles' inertia for focusing, so its performance improves with increasing mass (*18*).

The detector consisted of 20 nanomechanical resonators (arranged in a four-by-five array covering a 50 μm by 230 μm area). We fabricated doubly clamped beams with electrostatic actuation and differential piezoresistive readout using large-scale silicon processes (*19*). Particles landing on the vibrating part of a resonator add to its total mass M and cause its resonance frequency f to down-shift ($f \propto \sqrt[3]{M}$). As these frequency shifts also depend on the landing position on the resonator's surface, the frequencies of two resonance modes were monitored simultaneously to resolve the two unknowns (i.e., mass and position) (*13*, *14*, *20*). With an array of 20 resonators multiplexed in time, an increase of more than an order of magnitude in capture cross-section was obtained without degrading mass resolution (*21*–*23*).

We initially analyzed ~45-nm-diameter polystyrene nanoparticles [PS NPs, National Institute of Standards and Technology (NIST) size standards]. The transmission and focusing performances of our system were characterized by exposing silicon targets to the NP beam produced by SAWN (Fig. 2A). The transmission efficiency, defined as the ratio of the number of NPs on the target to the number of NPs in the nebulized solution, ranged from 3.5 to 7.6%. This transmission was four to five orders of magnitude greater than that achieved with previously reported architectures combining ion guides with nanomechanical

[1]Université Grenoble Alpes, F-38000 Grenoble, France. [2]CEA, BIG, Biologie à Grande Echelle, F-38054 Grenoble, France. [3]Inserm, Unité 1038, F-38054 Grenoble, France. [4]Université Grenoble Alpes, CEA, LETI, 38000 Grenoble, France. [5]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Université Paris-Saclay, 91198, Gif sur Yvette cedex, France.
*Present address: Techinsights, 1891 Robertson Road, Suite 500, Ottawa, ON K2H 5B7, Canada. †Present address: Laboratoire de Microbiologie et Génétique Moléculaires (LMGM), Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, Université Paul Sabatier, 31062 Toulouse, France.
‡Corresponding author. Email: christophe.masselon@cea.fr (C.M.); sebastien.hentz@cea.fr (S.H.)

resonators (*13*), and a factor 30 greater than that reported by (*15*) where, unlike in the present study, efficiency was measured from sprayed NPs, in a two- to three-orders-of-magnitude higher mass range. The measured full width at half maximum (FWHM) of the beam profile at the detector position (1.5 mm) was smaller than the aerodynamic lens outlet and the skimmer's orifice (Fig. 2B). Moreover, the beam's solid angle was less than 1°, indicating near-perfect aerodynamic collimation of the NP beam.

A 1.7×10^7 NP·µl^{-1} (28.2 pM) solution was nebulized, and NPs were focused onto the resonator array. Frequency time traces obtained for one resonance mode of 15 resonators within the array over 10 min of acquisition (Fig. 2C) showed downward frequency jumps corresponding to individual NP landing "events." An event rate ranging from 0.3 to 1.8 NP·min^{-1}·resonator^{-1} was recorded over 128 min of nebulization at an average flow rate of 2 µl·min^{-1}. The NP average size and standard size deviation σ determined beforehand by scanning electron microscopy (SEM) (45 ± 3 nm and σ = 13 nm) were consistent with the NIST specifications (46 ± 2 nm and σ = 7 nm). Converting these size measurements to mass, we could expect a broad mass distribution, with a central mass ranging from 28 to 36 MDa (σ = 15 to 25 MDa) (*23*). The mass histogram constructed from the nanomechanical measurements of 173 individual NPs (Fig. 2D) showed that the central mass and standard deviation of the normal fit to this histogram (29.5 MDa and σ = 17 MDa) were in good agreement with the expected mass distribution. The absolute mass error in our measurement was mainly caused by errors in the determination of the resonators' effective mass. We estimated this absolute error to be between 1 and 2% (0.3 to 0.6 MDa) (*23*), which was well below the uncertainty in central mass expected from size measurements (4 to 6 MDa). Moreover, the standard deviation of

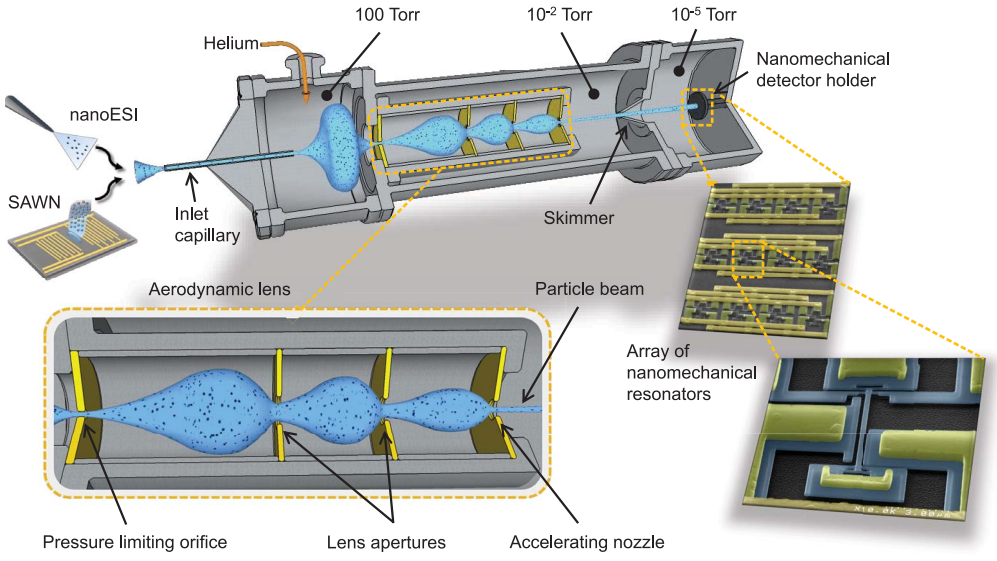


Fig. 1. High-transmission system architecture for nanomechanical resonator–based charge-independent single-particle mass sensing. The setup consists of three chambers with decreasing pressures. Analytes in solution are nebulized by SAWN or nano-ESI and aspirated through a heated metal capillary inlet at atmospheric pressure. An aerodynamic lens focuses the particle stream (shaded blue area), which is then transferred onto an array of frequency-addressed nanomechanical resonators. Left inset: Simplified schematic of the aerodynamic lens (actual design shown in fig. S8). Right inset: SEM image of a portion of an array showing 12 out of 20 resonators and magnified SEM image of a single resonator with its metallic layer and silicon material, falsely colored in yellow and gray, respectively. Resonator dimensions: 160 nm (thickness), 300 nm (width), and 7 to 10 µm (length).

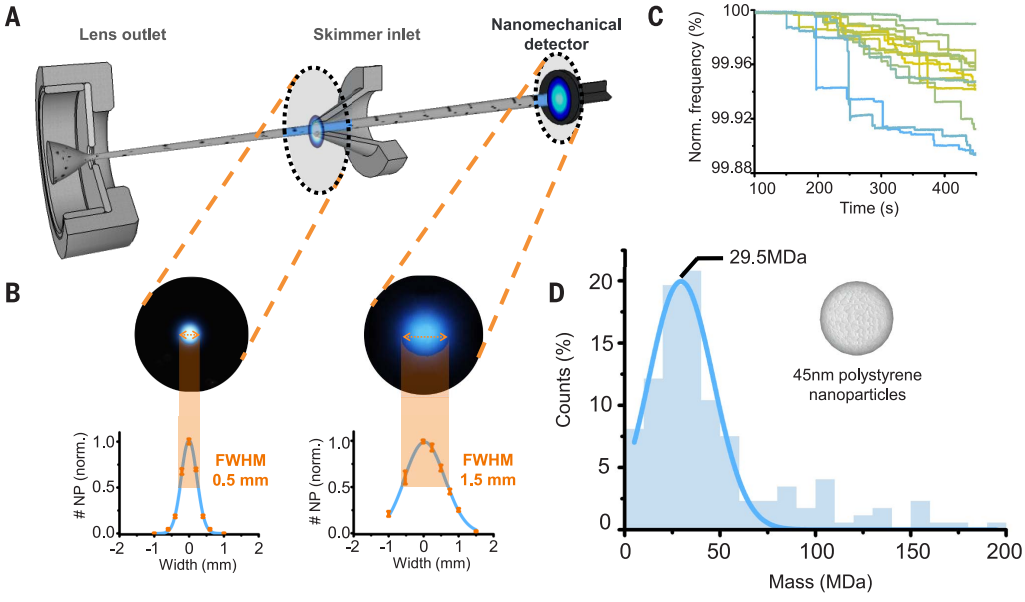
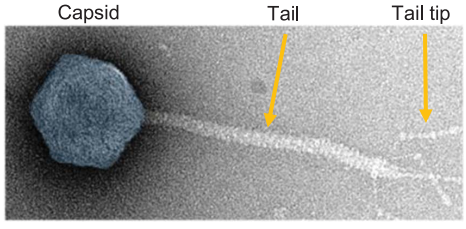


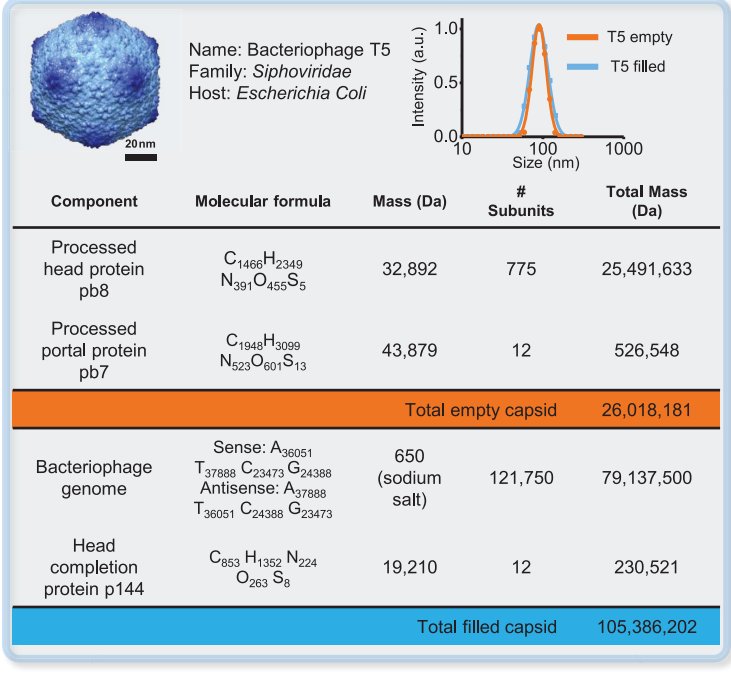
Fig. 2. Characterization of SAWN-based transmission and focusing efficiency, and high-throughput mass spectrum of PS NPs. (**A**) Diagram showing how the particle beam was characterized by placing silicon targets at the skimmer (~4 cm from aerolens outlet) and nanomechanical detector (~8 cm) positions. (**B**) Optical images of the targets observed after deposition of NIST PS NPs, with respective Gaussian fit to horizontal sections of the beam profile in particles per µm^2 (normalized scale) as measured on a series of SEM images. (**C**) Normalized raw frequency traces as a function of time for the fundamental mode of 15 nanoresonators exposed to the particle beam produced by SAWN. Each color represents a different nanomechanical device, and each step corresponds to a particle landing event. Some resonators were discarded because of poor signal (fig. S24). (**D**) Accumulated histogram of mass measurements of the same PS NPs for a nanoresonator array exposed to the particle beam, fit to a normal distribution.

the mass distribution expected from size measurement ranged from 15 (NIST) to 25 MDa (SEM), whereas we measured a 17-MDa deviation. This difference was not caused by our measurement noise level: The mass resolution for each measured NP could be inferred from the frequency noise and landing position on the resonator (*23*), and this estimation yielded an average mass resolution of 0.6 MDa for the 173 events.

The total amount of NPs consumed during this measurement was only 7.3 fmol (4.4×10^9 particles), and the detection efficiency was 1 NP per 2.2×10^7 in solution. These numbers are six orders of magnitude better than previous nanomechanical resonator–based systems using ion guides (*13*), but still two orders of magnitude lower than the system in (*15*) that used resonators with two orders of magnitude greater capture area. Indeed, this system operates in a mass range three orders of magnitude higher where efficient detection can be obtained without a focusing device, by accelerating particles through a nozzle with a viscous flow (low vacuum) and placing the resonator at close proximity, at the cost of a mass resolution four orders of magnitude higher than in our work.

We then performed measurements of biological particles: capsids of bacteriophage T5. Bacteriophage T5 is a *Siphoviridae* family member that infects *Escherichia coli* bacteria. Its ~90-nm icosahedral capsid is connected to a 250-nm tail, which plays a role in host cell recognition and genome delivery. The capsid itself is composed of 775 copies of the major capsid protein pb8 and 12 copies of the portal protein pb7. Well-controlled capsid assembly and expansion without ["empty" capsids (*24*)] and with ["filled" capsids (*25*)] genome packaging inside the capsid have been demonstrated. Once loaded with its 121.75-kbp double-stranded DNA viral genome, the capsid is completed by the head completion protein p144, which constitutes the docking site for tail attachment. Phage T5 is one of the only viruses with such large genome content amenable to in vitro studies. Unlike PS NPs, empty and filled capsids have well-defined molecular masses, calculated to be 26.0 and 105.4 MDa, respectively (see Fig. 3).





Component	Molecular formula	Mass (Da)	# Subunits	Total Mass (Da)
Processed head protein pb8	$C_{1466}H_{2349}N_{391}O_{455}S_5$	32,892	775	25,491,633
Processed portal protein pb7	$C_{1948}H_{3099}N_{523}O_{601}S_{13}$	43,879	12	526,548
			Total empty capsid	26,018,181
Bacteriophage genome	Sense: $A_{36051}T_{37888}C_{23473}G_{24388}$ Antisense: $A_{37888}T_{36051}C_{24388}G_{23473}$	650 (sodium salt)	121,750	79,137,500
Head completion protein p144	$C_{853}H_{1352}N_{224}O_{263}S_8$	19,210	12	230,521
			Total filled capsid	105,386,202

Fig. 3. Molecular mass of the bacteriophage T5 capsid. Top: Negatively stained electron microscopy image of the native bacteriophage T5. The capsid is falsely colored in blue. Bottom: Three-dimensional reconstruction of the assembled capsid measured here [EMDB-8423 (*26*)], also falsely colored in blue. Dynamic light-scattering size measurements of empty and filled capsids: Both capsids display very similar sizes, whereas their mass differs widely. The table shows the components of the capsid, with theoretical molecular mass calculations for both types of capsid.

Empty capsids were produced from infection of *E. coli* F cells by a T5st*AmN5* mutant defective in genome packaging and purified by anion exchange chromatography (*23*). Their expansion was triggered by decreasing the salt concentration to 0 mM NaCl through dialysis against 25 mM Hepes buffer at pH 7.2. Expanded empty capsids were stored at 4°C at a concentration of 0.46 mg/ml (~10^{13} capsids/ml). T5 DNA–filled capsids were produced as described in (*26*) by infection of *E. coli* F strain with the mutant *T5D18am-Δdec*, defective in tail assembly. The viral particles were purified by precipitation in NaCl–polyethylene glycol followed by centrifugation on a CsCl gradient. The capsids were then dialyzed in phage buffer for storage at a concentration of $2{\times}10^{13}$ capsids·ml^{-1}.

Ionization and transfer of large objects such as virus capsids with SAWN has not been reported in conventional MS experiments. Although capsids could be nebulized with SAWN and transferred to the detector chamber, relevant spectra could not be acquired, which was attributed to previously reported aggregation issues (*27*). We thus chose to use nano-ESI to nebulize the capsids, given the success of this approach for native MS of high-mass biological analytes (*4*). The capsid solutions were dialyzed against 25 mM ammonium acetate solution for at least 24 hours at 4°C, then slowly diluted to a final buffer concentration of 12.5 mM in 10% (v/v) methanol. The dialysis against ammonium acetate, addition of methanol, and an adequate capsid concentration were necessary for stable spray conditions. Empty and filled capsids were thus further diluted and sprayed at a concentration of ~5.9×10^{11} and ~8.8×10^{10} capsids·ml^{-1}, respectively. The dialysis procedure did not affect the stability of the capsids in solution, as almost identical dynamic light-scattering measurements were obtained for the filled capsids in their original buffer and after dialysis in ammonium acetate (fig. S21).

Empty and filled capsids were electrosprayed into the system and detected by the array of nanoresonators. Figure 4, A and B, show representative frequency traces of eight resonators for empty and filled T5 capsids, respectively. Spraying the buffer solution alone produced almost no detectable mass events nor measurable drift, suggesting efficient buffer desolvation given its volatility, the heated inlet capillary (150°C), and the high-vacuum environment. Figure 4C shows the mass distribution obtained for empty capsids. The spectrum displayed a bimodal distribution, which was fitted by two Gaussian peaks. The main peak fit had a central mass of 27.2 MDa (σ = 2.3 MDa) with the most abundant bin at 26 MDa, near the calculated 26.02 MDa; the second peak had a mass of 33.4 MDa (σ = 2.8 MDa). From the mass resolutions obtained for each single particle in the main peak, it was possible to deduce a 1.0-MDa FWHM resulting from the instrument alone (i.e., considering negligible sample heterogeneity) (fig. S22). This value corresponds to a resolution of 27 ($M/\Delta M_{\mathrm{FWHM}}$), which is comparable to charge-detection MS

measurements in this mass range (*6*). Moreover, it suggests that the two peaks could be ascribed to two distinct capsid populations present in the sample. This result was further substantiated by a differential scanning calorimetry experiment (fig. S20) showing two denaturation profiles, and could be attributed to the presence of a remaining short DNA fragment associated to some T5 capsids during purification. Finally, 363 events within the mass range of Fig. 4C were obtained with 645 fmol of capsids, over the course of 302 min (one particle detected per 1×10^9 in solution), with an event rate ranging from 1.15 to 1.35 capsid·min^{-1}. This result confirms the advantages of using our architecture over previous NEMS-MS systems using ion guides.

We then nebulized filled capsids and measured their mass distribution (Fig. 4D). In the filled capsid spectrum, a clear peak emerged with a fit at a mass of 108.4 MDa (σ = 6.0 MDa), and the most abundant bin at 107.5 MDa. The central mass of this distribution was within 2.5% of its calculated molecular mass (105.4 MDa). The slight remaining discrepancy and the peak asymmetry could be the result of salt adduction during the electrospray process, which reportedly yields masses slightly greater than theoretical estimates (*28*). This discrepancy is higher in absolute value for the filled than empty capsids (3.1 versus 1.2 MDa), presumably because the filled capsids' storage buffer contained a much higher salt concentration than that of the empty capsids. In addition, it is likely that some of the salt adducts bind to the viral DNA inside the filled capsids, making them more difficult to remove by dialysis. Individual polyhedral particles of the expected size (~93 nm) were clearly discernible on the nanomechanical resonators observed under SEM (fig. S26).

Notably, most capsids not only maintained their integrity during nebulization and transfer across the vacuum system, but they also survived landing on the nanoresonator surface at nearly supersonic speeds. The SEM pictures also show a finite contact area (i.e., not point contact) between capsids and silicon. Capsid stiffness could then induce a downward shift in the measured central mass and a broadening of the peak (*15*). Finite-element simulations have shown that not only is this effect negligible compared to both these quantities, but it is also below our average mass resolution (fig. S23). We attribute this to the use of an in-plane mode, to the moderate aspect ratio of our resonators and to the small displacement and strain induced by our high-frequency devices. Also, the measured capsid mass was very close to (even slightly greater than) the known molecular one. We detected 648 capsids within the mass range of Fig. 4D with a total sample of 286 fmol (one capsid detected per 2.6×10^8 in solution) over the course of 655 min with an event rate ranging from 0.8 to 1.15 capsid·min^{-1}. This fourfold improvement in detection efficiency compared to that of empty capsids was likely the result of improved focusing of these higher-mass particles. Notably, Fig. 4, E and F, show that the average mass resolution

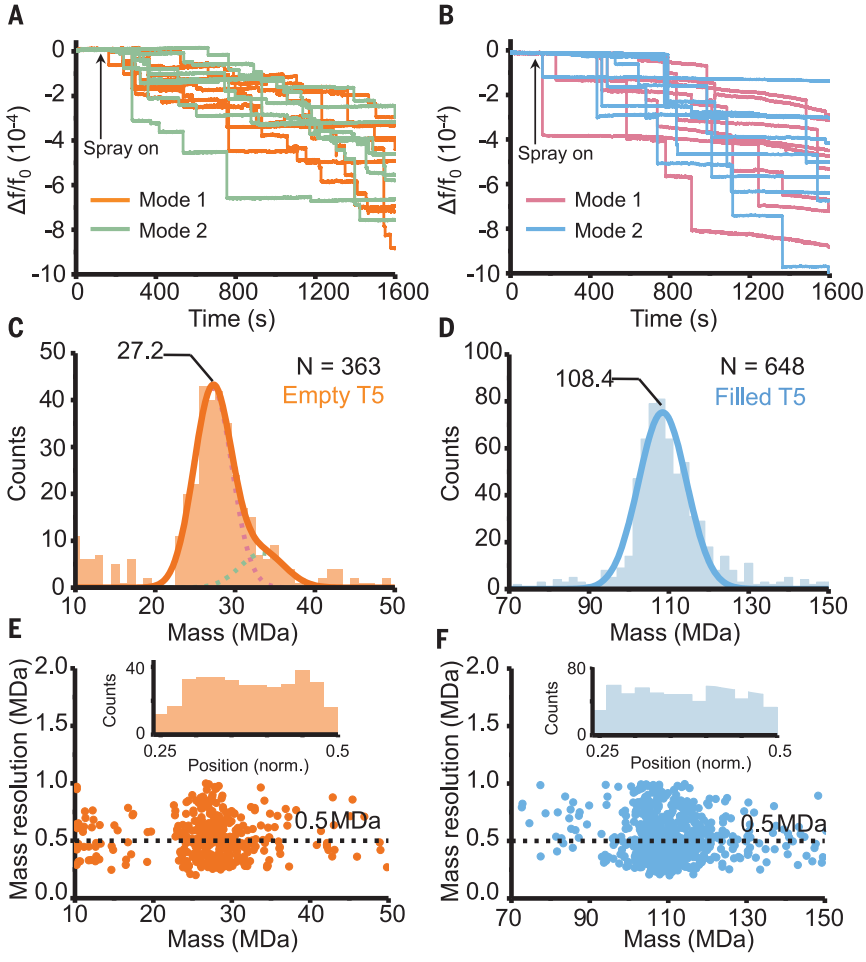


Fig. 4. Molecular mass measurement of empty and filled bacteriophage T5 capsids. (**A** and **B**) Representative raw frequency traces while spraying empty (A) and filled (B) capsids of bacteriophage T5, shown in fractional frequency change ($\Delta f/f_0$) with respect to their initial frequency f_0. For the sake of readability, the two modes of only eight nanoresonators are shown for 1600 s out of the whole experiment. (Other devices and time samples are shown in figs. S24 and S25.) (**C** and **D**) Accumulated mass histograms of 363 empty (C) and 648 filled (D) capsids nebulized using nano-ESI (with 1- and 2-MDa bin size, respectively), fitted to two and single normal distributions, respectively. Mass resolution per particle for empty (**E**) and filled capsids (**F**), with their respective particle landing position histograms (insets) in normalized units (beam extremities are at position 0 and center is at position 0.5). Capsids landing below 0.25 were discarded because of their poor mass resolution.

per detected filled capsid was the same as that obtained with empty capsids (~0.5 MDa), with capsids evenly spread over the beam area (Fig. 4, E and F, insets), and confirmed that nanomechanical MS had a higher resolution ($M/\Delta M_{FWHM}$) at higher masses. From the mass resolutions obtained for each single capsid in the main peak, we determined a 0.95-MDa FWHM from the instrument alone (i.e., in the absence of sample heterogeneity), which corresponds to an instrument resolution of 114 at 106 MDa.

The MS architecture can analyze ionized or neutral species and overcomes many of the limitations associated with earlier nanomechanical resonator–based systems, including inefficient detection. Simple nebulization techniques like SAWN can be used, leading to efficient uptake as ionization yield or Taylor cone expansion do not come into play. Moreover, these techniques may be less prone to dissociating noncovalently bound supramolecular analytes. Rather than being counteracted by electromagnetic fields, the inertia of massive particles is exploited for efficient guiding and focusing by using an aerodynamic lens. Nanomechanical resonators directly measure the inertial mass of individual analytes, avoiding separate measurements of *m*/*z* (mass-to-charge ratio) and charge. Experiments were performed with concentrations and sample quantities that were close to those typically used in MS experiments, and the total duration of experiments was only a few hours.

REFERENCES AND NOTES

1. A. Mujahid, F. L. Dickert, *Sensors (Basel)* **17**, 2716 (2017).
2. J. B. Fenn, *Angew. Chem. Int. Ed.* **42**, 3871–3894 (2003).

3. K. Tanaka, *Angew. Chem. Int. Ed.* **42**, 3860–3870 (2003).
4. J. Snijder, R. J. Rose, D. Veesler, J. E. Johnson, A. J. R. Heck, *Angew. Chem. Int. Ed.* **52**, 4020–4023 (2013).
5. T. Doussineau *et al*., *Angew. Chem. Int. Ed.* **55**, 2340–2344 (2016).
6. D. Z. Keifer, T. Motwani, C. M. Teschke, M. F. Jarrold, *Rapid Commun. Mass Spectrom.* **30**, 1957–1962 (2016).
7. B. Ilic, Y. Yang, H. G. Craighead, *Appl. Phys. Lett.* **85**, 2604–2606 (2004).
8. H.-Y. Chiu, P. Hung, H. W. C. Postma, M. Bockrath, *Nano Lett.* **8**, 4342–4346 (2008).
9. K. Jensen, K. Kim, A. Zettl, *Nat. Nanotechnol.* **3**, 533–537 (2008).
10. J. Chaste *et al*., *Nat. Nanotechnol.* **7**, 301–304 (2012).
11. J. J. Ruz, J. Tamayo, V. Pini, P. M. Kosaka, M. Calleja, *Sci. Rep.* **4**, 6051 (2014).
12. A. K. Naik, M. S. Hanay, W. K. Hiebert, X. L. Feng, M. L. Roukes, *Nat. Nanotechnol.* **4**, 445–450 (2009).
13. M. S. Hanay *et al*., *Nat. Nanotechnol.* **7**, 602–608 (2012).
14. E. Sage *et al*., *Nat. Commun.* **6**, 6482 (2015).
15. O. Malvar *et al*., *Nat. Commun.* **7**, 13452 (2016).
16. Y. Huang *et al*., *J. Am. Soc. Mass Spectrom.* **23**, 1062–1070 (2012).
17. X. Wang, F. E. Kruis, P. H. McMurry, *Aerosol Sci. Technol.* **39**, 611–623 (2005).
18. X. Wang, P. H. McMurry, *Int. J. Mass Spectrom.* **258**, 30–36 (2006).
19. E. Mile *et al*., *Nanotechnology* **21**, 165504 (2010).
20. S. Schmid, S. Dohn, A. Boisen, *Sensors (Basel)* **10**, 8092–8100 (2010).
21. M. Sansa *et al*., *Nat. Nanotechnol.* **11**, 552–558 (2016).
22. E. Sage *et al*., *Nat. Commun.* **9**, 3283 (2018).
23. See supplementary materials and methods.
24. O. Preux *et al*., *J. Mol. Biol.* **425**, 1999–2014 (2013).
25. Y. Zivanovic *et al*., *J. Virol.* **88**, 1162–1174 (2014).
26. E. Vernhes *et al*., *Sci. Rep.* **7**, 41662 (2017).
27. K. Yu, X. Wei, Z. Jiang, Y. Lei, F. Zhang, in *2016 IEEE 10th International Conference on Nano/Molecular medicine and engineering (NANOMED)* (2016), pp. 138–143.
28. P. Lössl, J. Snijder, A. J. R. Heck, *J. Am. Soc. Mass Spectrom.* **25**, 906–917 (2014).

ACKNOWLEDGMENTS

We thank V. Brun, E. Boeri Erba, and C. Breyton for support and fruitful discussions. We also thank M. Aumont-Nicaise for assistance with differential scanning calorimetry experiments and V. Zhong for help with electrospray ionization. This work has benefited from the facilities and expertise of the Macromolecular Interaction Platform of I2BC (UMR 9198). We also thank Malvern for the use of their facilities in Vénissieux, France. **Funding:** We acknowledge support from the European Union through the ERC Enlightened project (616251), from the LETI Carnot Institute NEMS-MS project, from the DGA Astrid NEMS-MS project, and from incoming CEA fellowships (M.S., S.D.-M.) from the CEA-Enhanced Eurotalents program, cofunded by FP7 Marie-Skłodowska-Curie COFUND program (Grant Agreement 600382). **Author contributions:** S.D.-M., S.F., M.D., A.-K.S., M.A.H., and T.A. performed the MS experiments. E.V. and P.B. prepared the T5 capsid samples and performed their characterization experiments. M.S., M.D., and G.J. worked on the operation set point and control of the nanomechanical arrays. M.G. fabricated the nanomechanical arrays. C.M. and S.H. supervised the research. All authors participated in the writing of the manuscript. **Competing interests:** C.M. and S.H. are co-inventors of the patents US9506852B2, EP2779209A1, and JP6352004B2. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/918/suppl/DC1
Materials and Methods
Figs. S1 to S26
Tables S1 to S8
References (*29*–*35*)

6 April 2018; resubmitted 20 July 2018
Accepted 11 October 2018
10.1126/science.aat6457

SUPERCONDUCTIVITY

Gate-induced superconductivity in a monolayer topological insulator

Ebrahim Sajadi[1], Tauno Palomaki[2], Zaiyao Fei[2], Wenjin Zhao[2], Philip Bement[1], Christian Olsen[1], Silvia Luescher[1], Xiaodong Xu[2,3], Joshua A. Folk[1]*, David H. Cobden[2]*

The layered semimetal tungsten ditelluride (WTe_2) has recently been found to be a two-dimensional topological insulator (2D TI) when thinned down to a single monolayer, with conducting helical edge channels. We found that intrinsic superconductivity can be induced in this monolayer 2D TI by mild electrostatic doping at temperatures below 1 kelvin. The 2D TI–superconductor transition can be driven by applying a small gate voltage. This discovery offers possibilities for gate-controlled devices combining superconductivity and nontrivial topological properties, and could provide a basis for quantum information schemes based on topological protection.

Many of the most important phenomena in condensed matter emerge from the quantum mechanics of electrons in a lattice. The periodic potential of the lattice gives rise to Bloch energy bands of independent fermions; on the more exotic side, electrons in a lattice can pair up into bosons and condense into a superconducting macroscopic quantum state, which conducts electricity with zero resistance. Relatively recently, it was realized that Bloch wave functions can have a nontrivial topology, leading to the discovery of topological insulators—materials that are electrically insulating in their interior but have conducting boundary modes (*1*). The first of these to be studied was the so-called 2D topological insulator (2D TI), in which 1D helical edge modes (spin locked to momentum) give rise to the quantum spin Hall effect (*2*–*4*).

Materials that combine nontrivial topology with superconductivity have been the subject of active investigation in recent years (*5*–*7*). Here, we report that monolayer WTe_2, recently shown (*8*–*13*) to be an intrinsic 2D TI, turns superconducting under moderate electrostatic gating. Several other nontopological layered materials superconduct in the monolayer limit, either intrinsically or under heavy doping using ionic liquid gates (*14*–*22*). In monolayer WTe_2, however, the phase transition to a superconducting state is from a 2D topological insulator, and it occurs at such a low carrier density that it can be readily induced by a simple electrostatic gate. The discovery may lead to gateable superconducting circuitry and may enable the development of topological superconducting devices in a single material, as opposed to the hybrid constructions currently required (*5*).

We present data from two monolayer WTe_2 devices, M1 and M2, with consistent superconducting characteristics. Each contains a monolayer flake of WTe_2 encapsulated along with thin platinum electrical contacts between hexagonal boron nitride (hBN) dielectric layers. Figure 1A shows an image of M1, which has seven contacts along one edge, together with a side view and a schematic showing the configuration used to measure the linear four-probe resistance, $R_{xx} = dV/dI$. Top and bottom gates—at voltages V_t and V_b and with areal capacitances c_t and c_b, respectively—can be used to induce negative or positive charge in the monolayer WTe_2, producing an areal doping density given by $n_e = (c_tV_t + c_bV_b)/e$, where e is the electron charge. Note that we do not interpret this as a carrier density because the insulating state may be of correlated nature (as in, for example, an excitonic insulator); in addition, Hall density measurements are challenging because of the 2D TI edge conduction. See (*23*) for details about gating, contact resistances, and capacitances.

Figure 1B illustrates the electrostatic tuning of M1 from p-doped conducting behavior at negative gate voltage, through an insulating state, to an n-doped highly conducting state at positive gate voltage. M1 is the same device whose insulating state was investigated in (*9*) and was demonstrated to be a 2D TI (*9*, *13*); at $n_e = 0$, R_{xx} is more than 10^7 ohms, owing to a meV-scale gap that blocks edge conduction below 1 K [see below and (*23*)]. For n_e above $n_{crit} \approx +5 \times 10^{12}$ cm^{-2}, however, the resistance drops drastically when the sample is cooled, reaching the noise floor of the experiment (~0.3 ohms) for $n_e > +7 \times 10^{12}$ cm^{-2} at 20 mK, indicating the appearance of superconductivity. Figure 1C is a phase diagram constructed from these and similar measurements discussed below. The emergence of a superconducting phase in direct proximity to a 2D TI phase, and at a doping level achievable with a single electrostatic gate, is the primary result of our work.

The transition from an insulating to a metallic/superconducting T dependence—the crossing of R_{xx} lines in Fig. 1B—occurs at 2.4 kilohms. This corresponds to a square resistivity $\rho \approx 20$ kilohms, with a substantial uncertainty because the precise distribution of current in the device is not known (*23*). The evolution of the T dependence with n_e is illustrated in Fig. 2A. For all densities shown, the collapse of R_{xx} with temperature is gradual, as expected for materials where the normal-state 2D conductivity is not much greater than e^2/h (where h is the Planck constant). We define a characteristic temperature, $T_{1/2}$, at which R_{xx} falls to half of its 1 K value. Although this specific definition is somewhat arbitrary, it is typical in the literature (*15*, *21*, *22*) and does not affect any of our conclusions (*23*). Measured values of $T_{1/2}$ are shown as red dots on the phase diagram in Fig. 1C to indicate the boundary of superconducting behavior.

The superconductivity is suppressed by a perpendicular ($B_\perp$) or in-plane ($B_{||}$) magnetic field (Fig. 2, B and C). For a perpendicular field, orbital effects are expected to dominate (*24*–*26*). The dependence of $T_{1/2}$ on $B_\perp$ (Fig. 2B, inset) in the low-field limit is consistent with the linear $B^\perp_{c2}(T)$ expected from Ginzburg-Landau theory. The characteristic perpendicular field in the low-temperature limit, based on the measurements in Fig. 2B (inset), is $B^\perp_{1/2}(T \to 0) \approx 25$mT, where $B^\perp_{1/2}$ is the magnetic field where R_{xx} falls to half its normal-state resistance. Estimates for the superconducting coherence length can be obtained either from the slope of $B^\perp_{1/2}(T)$ near $T_{1/2}$ or from $B^\perp_{1/2}(T \to 0)$, yielding $\xi_{meas} = 100 \pm 30$ nm in both cases (*23*).

The fact that ξ_{meas} is much larger than the estimated mean free path $\lambda = h/(e^2\rho\sqrt{g_s g_v \pi n_e}) \approx 8$ nm suggests that the system is in the dirty limit ($\lambda << \xi$). To calculate λ, we use spin and valley degeneracies $g_s = g_v = 2$, as well as density and normal-state resistivity reflecting the conditions for Fig. 2B: $n_e = 20 \times 10^{12}$ cm^{-2} and $\rho \approx 2$ kilohms, respectively. The coherence length expected in the dirty limit is $\xi \approx \sqrt{\hbar D/\Delta_0}$, for zero-temperature gap $\Delta_0 = 1.76 k_B T_c$ and diffusion constant D, where $\hbar$ is the reduced Planck constant and k_B is the Boltzmann constant. Indeed, if we use $T_{1/2} = 700$ mK for T_c, and if $D = 2\pi\hbar^2/g_s g_v m^* e^2 \rho \approx 12$ cm^2 s^{-1} (from the Einstein relation) with effective mass $m^* = 0.3m_e$ [where m_e is the electron mass (*27*)], the result is $\xi \approx 90$ nm, consistent with ξ_{meas}.

For the in-plane magnetic field, the atomic thinness of the monolayer makes orbital effects small. In the absence of spin scattering, superconductivity is then suppressed when the energy associated with Pauli paramagnetism in the normal state overcomes the superconducting condensation energy. This is referred to as the Pauli (Chandrasekar-Clogston) limit (*28*) and gives a critical field $B_P = 1.76 k_B T_c/g^{1/2}\mu_B$, where μ_B is the Bohr magneton. Assuming an electron g-factor of $g = 2$ and taking $T_c = 700$ mK gives $B_P \approx 1.3$ T. However, the data in Fig. 2, C and F, indicate superconductivity persisting to $B^{||}_{1/2} = 3$T.

[1]Stewart Blusson Quantum Matter Institute, University of British Columbia, Vancouver, BC V6T 1Z4, Canada, and Department of Physics and Astronomy, University of British Columbia, Vancouver, BC V6T 1Z1, Canada. [2]Department of Physics, University of Washington, Seattle, WA 98195, USA. [3]Department of Materials Science and Engineering, University of Washington, Seattle, WA 98195, USA.
***Corresponding author. Email: cobden@uw.edu (D.H.C.); jfolk@physics.ubc.ca (J.A.F.)**

Similar examples of $B_{1/2}^{\parallel}$ exceeding B_P have recently been reported in other monolayer dichalcogenides, MoS_2 and $NbSe_2$, but the Ising superconductivity mechanism (*15*, *21*) invoked in those works cannot explain an enhancement of $B_{1/2}^{\parallel}$ here because WTe_2 lacks the required in-plane mirror symmetry. One possible explanation in this case is a high spin-orbit scattering rate τ_{so}^{-1}. Fitting the predicted form for T_c in a parallel field (*29*) to the data in the inset of Fig. 2C gives $\tau_{so}^{-1} \approx 2 ps^{-1}$ (*23*). Another possibility is that the Pauli limit is not actually exceeded but that the effective g-factor in WTe_2 is smaller than 2 owing to the strong spin-orbit coupling.

The data in Fig. 2 display several other features worthy of mention. First, at intermediate magnetic fields, the resistance approaches a T-independent level as $T \rightarrow 0$ that is orders of magnitude below the normal-state resistance. The data from Fig. 2B are replotted versus $1/T$ in Fig. 2D to highlight the behavior below 100 mK. Similar behavior is seen at $B = 0$ (Fig. 2A) for intermediate n_e, adding to the growing body of evidence that this is a robust phenomenon occurring in thin films close to superconductivity (*30*). Second, even at the lowest temperature, R_{xx} rises smoothly from zero as a function of $B_\perp$ (Fig. 2E), whereas the onset of measurable resistance as a function of $B_{||}$ is relatively sudden, occurring above 2 T (Fig. 2F). Third, an intermediate plateau is visible in the $R_{xx} - T$ data at $B = 0$ over a wide range of n_e (Fig. 2A). It is extremely sensitive to $B_\perp$, almost disappearing at only 2 mT (Fig. 2B), whereas it survives in $B_{||}$ to above 2 T (Fig. 2C and inset of Fig. 2F). A similar feature has been reported in some other quasi-2D superconductors (*31*–*33*), but its nature, and the role of disorder, remain unresolved.

The high tunability of this 2D superconducting system invites comparison with theoretical predictions for critical behavior close to a quantum phase transition. Figure 3 shows how R_{xx} depends on doping at a series of temperatures, along the dashed lines in the phase diagram (upper inset). The T dependence changes sign at $n_{crit} \approx 5 \times 10^{12}$ cm^{-2}. In the lower inset, we show an attempt to collapse the data onto a single function of $|1 - n_e/n_{crit}|T^{-\alpha}$. The procedure is somewhat hindered by the fluctuations, which can be seen to be largely reproducible. The best-fit critical exponent $\alpha = 0.8$ is similar to that reported for some insulator-superconductor transitions in thin films (*34*), although we note that the anomalous behavior near n_{crit} mentioned above is not consistent with such a scaling.

Superconductivity induced by simple electrostatic gating in a monolayer of material that is not normally superconducting is intriguing, but perhaps even more interesting is that the ungated state is a 2D TI. This prompts the question of whether the helical edge channels remain when the superconductivity appears, and if so, how strongly they couple to it. In principle, R_{xx} includes contributions from edges as well as bulk. However, because in device M1 the edge conduction freezes out below 1 K, in order to investigate the combination of edge channels and superconductivity we turn to another device, M2, in which edge conduction persists to lower temperatures (*23*).

Figure 4 shows measurements of the conductance G between adjacent contacts in M2 as a function of gate doping. The figure includes schematics indicating the inferred state of the edge (red for conducting), as well as the bulk state (colored to match the phase diagram). Consider first the black trace, taken at 200 mK and $B_\perp = 0$. At low n_e, the bulk is insulating and edge conduction dominates, albeit with large mesoscopic fluctuations. For $n_e > 2 \times 10^{12}$ cm^{-2}, G increases as bulk conduction begins; then, once n_e exceeds n_{crit}, it increases faster as superconductivity appears, before leveling out at ~200 μS as a result of contact resistance. This interpretation is supported by warming to 1 K (red dotted trace), which destroys the superconductivity and so reduces G for $n_e > n_{crit}$, but enhances the edge conduction at low n_e toward the ideal value of $e^2/h = 39$ μS. (We note that this T dependence of the edge is associated with a gap of ~100 μeV, visible in the inset map of differential conductance versus bias and doping.) A perpendicular field $B_\perp$ of 50 mT (green trace) also destroys the superconductivity, causing the conductance to fall for $n_e > n_{crit}$ but barely affecting it at lower n_e. High magnetic fields have been shown (*9*) to suppress edge conduction in the 2D TI state by breaking time-reversal symmetry. This effect can be clearly seen in the $B_\perp = 1$ T data (orange trace in Fig. 4) as G falls to zero at low n_e. Comparison of the green ($B_\perp = 0.05$ T) and orange ($B_\perp = 1$T) traces shows that G falls by a similar amount at higher n_e, consistent with a scenario in which the edge conduction supplies a parallel contribution; this implies that helical edge states persist when $n_e > n_{crit}$ and at temperatures below T_c.

This discovery raises compelling questions for future investigation. It is likely that the helical edge modes persist when the superconductivity is restored by reducing the magnetic field to zero. Other techniques, such as scanning probe microscopy, may be needed to probe the edges separately from the bulk. The measurements presented here cannot determine the degree or nature of the coupling between superconductivity and edge conduction. One key question is whether the edge states also develop a superconducting gap, in which case they could host Majorana zero modes (*5*).

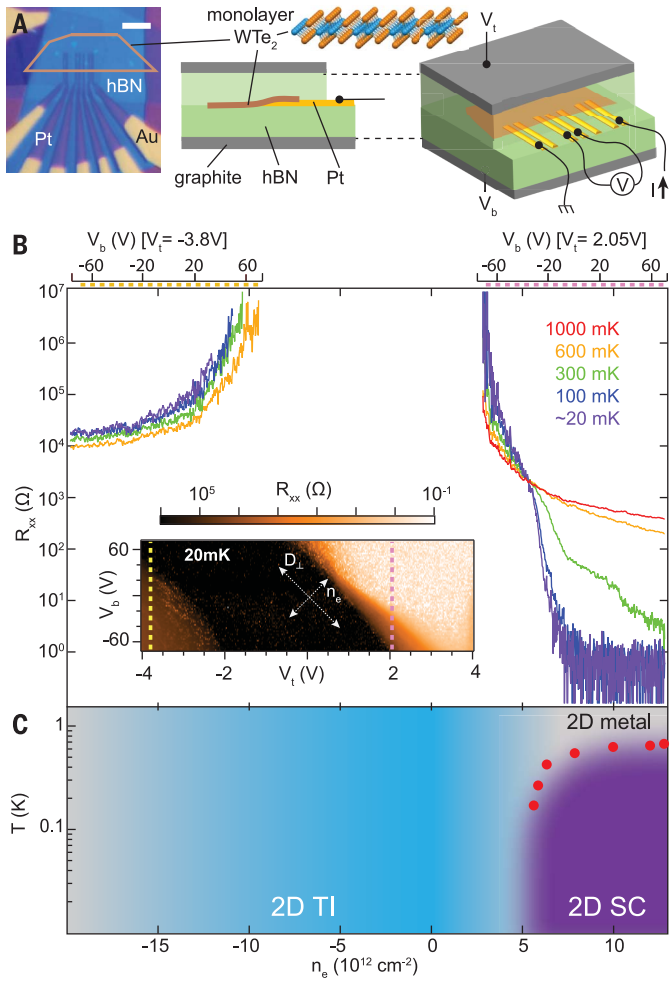


Fig. 1. Characteristics of monolayer WTe_2 device M1 at temperatures below 1 K. (**A**) Optical image (scale bar, 5 μm) of M1 and schematic device structure of a sample with two graphite gates, showing current, voltage contacts, and ground configuration for measuring the four-probe resistance R_{xx}. Inset: Schematic of the atomic structure of monolayer WTe_2. (**B**) R_{xx} as a function of electrostatic doping (n_e) at a series of temperatures. Inset: Variation of R_{xx} at 20 mK with top and bottom gate voltages, V_t and V_b, indicating the axes corresponding to doping n_e and transverse displacement field $D_\perp$. R_{xx} depends primarily on n_e and only weakly on $D_\perp$. The measurements in the main panel for $n_e > 0$ and $n_e < 0$ were made separately, sweeping V_b along the two colored dashed lines in the inset to avoid contact effects. (**C**) Phase diagram constructed from measurements in this paper.

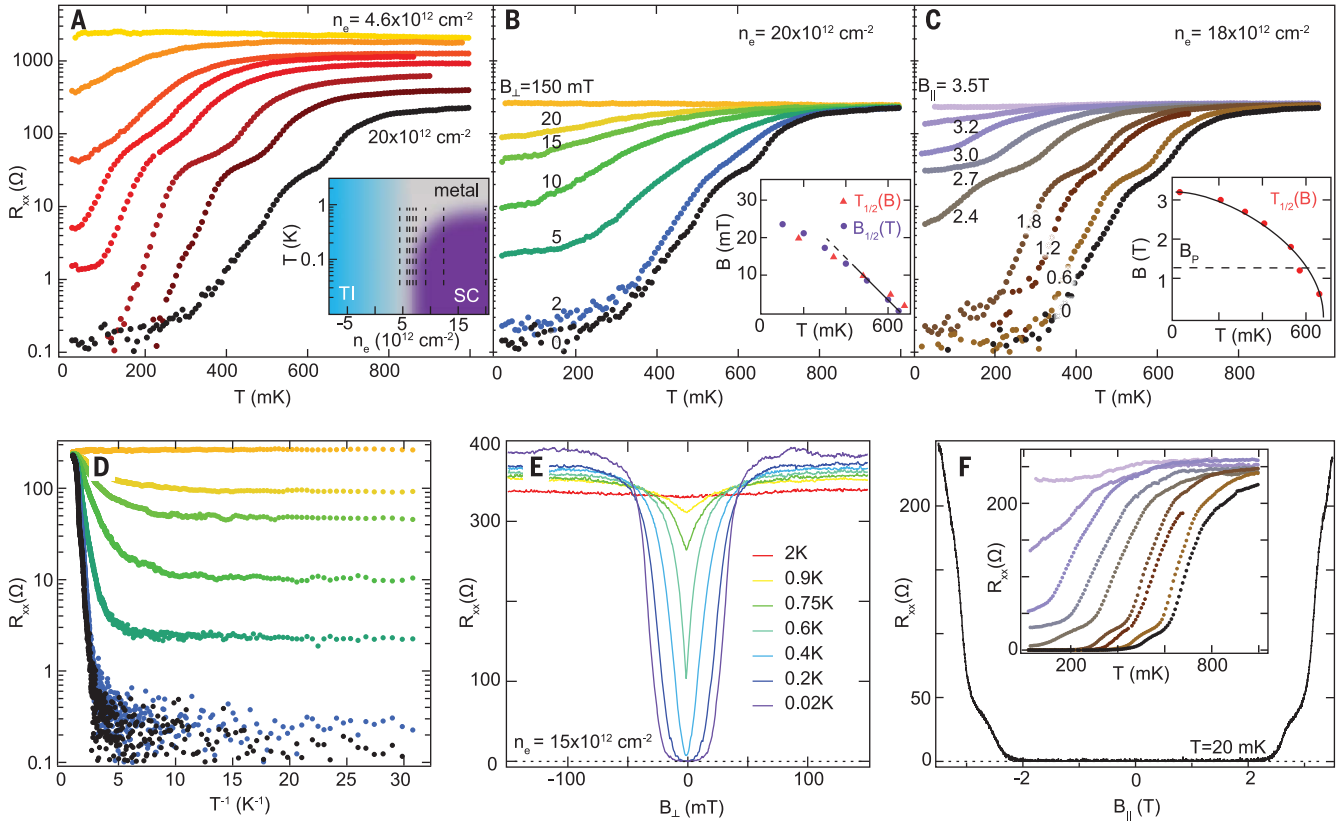


Fig. 2. Resistance characterization of device M1 in the superconducting regime. (**A**) R_{xx} on log scale versus temperature T at a series of positive-gate doping levels n_e [20, 12, 8.5, 6.7, 6.1, 5.6, 5, and 4.6 × 10^{12} cm^{-2}] showing a drop of several orders of magnitude at low T for larger n_e. Inset: Location of sweeps on the phase diagram. (**B**) Effect of perpendicular magnetic field $B_\perp$ on resistance at the highest n_e value in (A). (Demagnetization effects are neglected in light of the finite resistivity of the sample.) Inset: Characteristic temperatures $T_{1/2}$ obtained from these temperature sweeps, as well as characteristic fields $B_{1/2}$ measured from field sweeps under similar conditions. (**C**) Same as (B) but for the in-plane magnetic field $B_{||}$ (the $B_{||}$ = 0 data are for n_e = 19 × 10^{12} cm^{-2}; the remaining data are for n_e = 18 × 10^{12} cm^{-2}). Inset: Reduction of $T_{1/2}$ with $B_{||}$, fit to the expected form for materials with strong spin-orbit scattering (solid line). The Pauli limit B_P, assuming g = 2, is indicated by the dashed line. (**D**) Data from (B) replotted to highlight the saturation of R_{xx} at low T. (**E**) Sweeps of $B_\perp$ showing rise of resistance beginning at very low field. (**F**) Sweep of $B_{||}$ showing sharper onset of resistance relative to (E). Inset: Data from (C) on a linear scale.

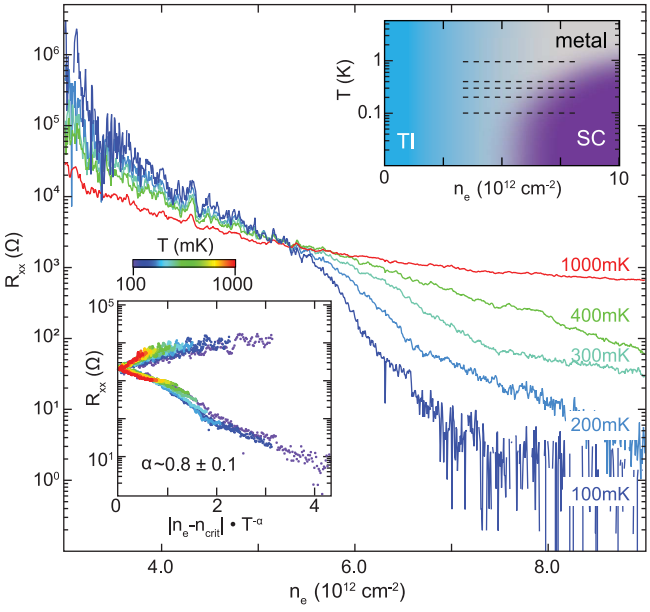


Fig. 3. Scaling analysis of the transition. Main panel: Multiple R_{xx} versus doping traces, taken at different temperatures, cross at a critical doping level n_{crit} ≈ 5 × 10^{12} cm^{-2}. Upper inset: Dashed lines locate these sweeps on the phase diagram. Lower inset: The same data presented on a scaling plot, taking critical exponent α = 0.8.

Another question concerns the nature of the superconducting order. It is striking that n_{crit} corresponds to only ~0.5% of an electron per W atom, which is about an order of magnitude lower than the doping level needed to observe superconductivity in other transition metal dichalcogenide monolayers (*18*). Many-layer WTe_2 is semimetallic (*35*–*38*) under ambient conditions, with near-perfect compensation of electrons and holes, but becomes superconducting as the ratio of electrons to holes increases at high pressure (*39*). Some related materials, such as $TiSe_2$, are known to switch from charge-density-wave to superconducting states at quite low doping (*40*) or under pressure (*41*). We therefore speculate that doping tips the balance in monolayer WTe_2 in favor of superconductivity, away from a competing insulating electronic ordering. Finally, given the topological band structure and likely strong correlations in this material, it is possible that the pairing is unconventional and perhaps topologically nontrivial.

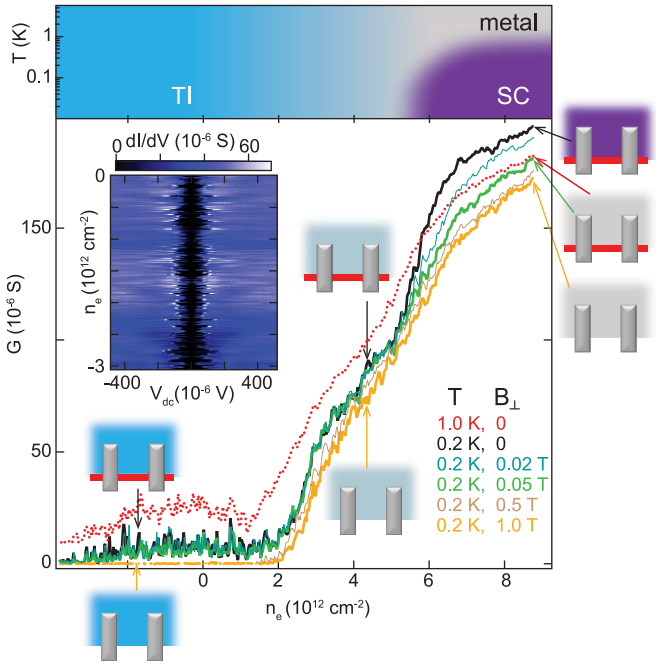


Fig. 4. Evidence for the presence of both edge conduction and superconductivity in device M2. The main panel shows the linear conductance between two adjacent contacts versus gate doping at the temperatures and perpendicular magnetic fields noted. Schematics indicate the state of edge and bulk conduction at different points; the bulk is colored to match the phase diagram reproduced above, and red indicates a conducting edge state. Superconductivity occurs for $n_e > 5 \times 10^{12}$ cm^{-2} at $B = 0$. The zero-resistance state, disguised by contact resistance in this figure, was confirmed in a separate four-wire measurement of R versus T (fig. S10); edge conduction dominates for $n_e < 2 \times 10^{12}$ cm^{-2} but appears to be present at all values of n_e. Inset: Color-scale plot of differential conductance versus dc voltage bias and doping level, revealing a gap of ~100 μeV that fluctuates rapidly as a function of doping level.

REFERENCES AND NOTES

1. M. Z. Hasan, C. L. Kane, *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
2. B. A. Bernevig, T. L. Hughes, S.-C. Zhang, *Science* **314**, 1757–1761 (2006).
3. C. L. Kane, E. J. Mele, *Phys. Rev. Lett.* **95**, 146802 (2005).
4. M. König *et al.*, *Science* **318**, 766–770 (2007).
5. L. Fu, C. L. Kane, *Phys. Rev. Lett.* **100**, 096407 (2008).
6. S. Sarma, M. Freedman, C. Nayak, *NPJ Quantum Inform.* **1**, 15001 (2015).
7. M. Sato, Y. Ando, *Rep. Prog. Phys.* **80**, 076501 (2017).
8. X. Qian, J. Liu, L. Fu, J. Li, *Science* **346**, 1344–1347 (2014).
9. Z. Fei *et al.*, *Nat. Phys.* **13**, 677–682 (2017).
10. S. Tang *et al.*, *Nat. Phys.* **13**, 683–687 (2017).
11. L. Peng *et al.*, *Nat. Commun.* **8**, 659 (2017).
12. Z.-Y. Jia *et al.*, *Phys. Rev. B* **96**, 041108 (2017).
13. S. Wu *et al.*, *Science* **359**, 76–79 (2018).
14. A. W. Tsen *et al.*, *Nat. Phys.* **12**, 208–212 (2015).
15. X. Xi *et al.*, *Nat. Phys.* **12**, 139–143 (2016).
16. Q. Y. Wang *et al.*, *Chin. Phys. Lett.* **29**, 4 (2012).
17. L. J. Li *et al.*, *Nature* **529**, 185–189 (2016).
18. Y. Fu *et al.*, *NPJ Quantum Mater.* **2**, 52 (2017).
19. Y. Saito, T. Nojima, Y. Iwasa, *Supercond. Sci. Technol.* **29**, 093001 (2016).
20. K. Ueno *et al.*, *Nat. Mater.* **7**, 855–858 (2008).
21. J. M. Lu *et al.*, *Science* **350**, 1353–1357 (2015).
22. D. Costanzo, S. Jo, H. Berger, A. F. Morpurgo, *Nat. Nanotechnol.* **11**, 339–344 (2016).
23. See supplementary materials.
24. R. A. Klemm, *Physica C* **514**, 86–94 (2015).
25. M. Tinkham, *Introduction to Superconductivity* (McGraw-Hill, ed. 2, 1996).
26. Y. Saito, T. Nojima, Y. Iwasa, *Nat. Rev. Mater.* **2**, 16094 (2016).
27. H. Y. Lv *et al.*, *Europhys. Lett.* **110**, 37004 (2015).
28. A. M. Clogston, *Phys. Rev. Lett.* **9**, 266–267 (1962).
29. R. A. Klemm, A. Luther, M. R. Beasley, *Phys. Rev. B* **12**, 877–891 (1975).
30. A. Kapitulnik, S. A. Kivelson, B. Spivak, arXiv:1712.07215 (19 December 2017) [cond-mat].
31. Y. Cheng, M. B. Stearns, *J. Appl. Phys.* **67**, 5038–5040 (1990).
32. L. Z. Deng *et al.*, *Phys. Rev. B* **93**, 054513 (2016).
33. S.-G. Jung *et al.*, *Supercond. Sci. Technol.* **30**, 085009 (2017).
34. A. M. Goldman, *Int. J. Mod. Phys. B* **24**, 4081–4101 (2010).
35. A. A. Soluyanov *et al.*, *Nature* **527**, 495–498 (2015).
36. M. N. Ali *et al.*, *Nature* **514**, 205–208 (2014).
37. D. MacNeill *et al.*, *Nat. Phys.* **13**, 300–305 (2017).
38. V. Fatemi *et al.*, *Phys. Rev. B* **95**, 041410 (2017).
39. D. Kang *et al.*, *Nat. Commun.* **6**, 7804 (2015).
40. E. Morosan *et al.*, *Nat. Phys.* **2**, 544–550 (2006).
41. A. F. Kusmartseva, B. Sipos, H. Berger, L. Forró, E. Tutiš, *Phys. Rev. Lett.* **103**, 236401 (2009).

ACKNOWLEDGMENTS

We thank O. Agam, A. Andreev, and B. Spivak for discussions, and J. Yan for the WTe_2 crystals. **Funding:** T.P., Y.F., W.Z., X.X., and D.H.C. were supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, awards DE-SC0002197 (D.H.C.) and DE-SC0018171 (X.X.); AFOSR FA9550-14-1-0277; NSF EFRI 2DARE 1433496; and NSF MRSEC 1719797. E.S., P.B., C.O., S.L., and J.A.F. were supported by the Canada Foundation for Innovation, the National Science and Engineering Research Council, CIFAR, and SBQMI. **Author contributions:** T.P., X.X., and D.H.C. conceived of the original experiment; T.P., Z.F., and W.Z. fabricated the devices; E.S., Z.F., P.B., and C.O. carried out the measurements under the primary supervision of J.A.F. and D.H.C. in a cryostat developed by E.S., S.L., and J.A.F.; all authors contributed to the analysis of the data; E.S., J.A.F., and D.H.C. wrote the manuscript; all authors contributed to the final editing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data shown in the paper are available at https://github.com/EbrahimSajadi/2D_SC_TI.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/922/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Table S1
References (*42*, *43*)

12 November 2017; accepted 5 October 2018
Published online 25 October 2018
10.1126/science.aar4426

SUPERCONDUCTIVITY

Electrically tunable low-density superconductivity in a monolayer topological insulator

Valla Fatemi[1]*, Sanfeng Wu[1]*†, Yuan Cao[1], Landry Bretheau[2], Quinn D. Gibson[3], Kenji Watanabe[4], Takashi Taniguchi[4], Robert J. Cava[5], Pablo Jarillo-Herrero[1]†

Turning on superconductivity in a topologically nontrivial insulator may provide a route to search for non-Abelian topological states. However, existing demonstrations of superconductor-insulator switches have involved only topologically trivial systems. Here we report reversible, in situ electrostatic on-off switching of superconductivity in the recently established quantum spin Hall insulator monolayer tungsten ditelluride (WTe_2). Fabricated into a van der Waals field-effect transistor, the monolayer's ground state can be continuously gate-tuned from the topological insulating to the superconducting state, with critical temperatures T_c up to ~1 kelvin. Our results establish monolayer WTe_2 as a material platform for engineering nanodevices that combine superconducting and topological phases of matter.

The intersection of superconductivity and topological insulators hosts a fertile landscape of interesting quantum phenomena, including non-Abelian topological excitations (*1*–*3*). Some topological insulators have been turned into superconductors via chemical doping (*4*, *5*), application of pressure (*6*–*8*), or via the proximity effect (*9*, *10*), methods that are either irreversible or ex situ. Proximity effect–based devices have, in particular, been engineered in an effort to realize Majorana physics; this has required fine-tuned interface engineering between distinct materials (*11*–*16*).

Here we report the observation of intrinsic superconductivity in the monolayer topological insulator WTe_2 induced by the electric field effect. This monolayer transition metal dichalcogenide has recently been established as a quantum spin Hall insulator with robust edge transport up to 100 K (*17*–*22*). Our field-effect device geometry (*23*) is illustrated in Fig. 1, A and B. The WTe_2 monolayer flake is van der Waals–encapsulated between two sheets of hexagonal boron nitride (BN) to protect it from chemical degradation and to improve transport characteristics. Top and bottom electrostatic gates, for which the same BN sheets serve as gate dielectrics, are fabricated to modulate the carrier density. An optical microscopy image of a typical device (device 1, with BN thicknesses of 15 nm, top, and 8 nm, bottom) is shown in Fig. 1B.

[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Laboratoire des Solides Irradiés, École Polytechnique, CNRS, CEA, 91128 Palaiseau Cedex, France. [3]Department of Chemistry, University of Liverpool, Liverpool L69 7ZX, UK. [4]Advanced Materials Laboratory, National Institute for Materials Science, 1-1 Namiki, Tsukuba 305-0044, Japan. [5]Department of Chemistry, Princeton University, Princeton, NJ 08544, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: swu02@mit.edu (S.W.); pjarillo@mit.edu (P.J.-H.)

Figure 1C displays the temperature dependence of the four-probe resistance $R(T)$ in device 1 when the monolayer is n-doped [bottom-gate voltage (V_{bg}) = 4 V; top-gate voltage (V_{tg}) = 5 V]. R drops to zero at low temperatures from 1.2 kilohm in the normal state, and this zero-resistance state is present for a wide range of gate-voltage parameters (inset of Fig. 1C). The dc voltage-current (V-I) characteristics at various temperatures (Fig. 1D) exhibit the transition from ohmic behavior (red curve) at ~1 K to highly nonlinear behavior (black curve) at low temperatures, including the characteristic zero-voltage plateau for finite current, typical of superconductivity. This nonlinearity is further captured by the measurement shown in Fig. 1E, where the four-probe differential resistance, dV/dI, is plotted as a function of dc current bias I_{dc} at base temperature. Under application of an out-of-plane magnetic field B, the nonlinear behavior is clearly suppressed (Fig. 1E), as expected for superconductivity. These observations confirm a superconducting state forming in the monolayer atomic sheet, in contrast to its three-dimensional (3D) parent crystal, in which no superconductivity is found unless high pressure is applied (*24*, *25*). A small shoulder appears around 400 mK in the $R(T)$ curve (Fig. 1C). Although this may be an intrinsic feature of WTe_2, it could also result from the appearance of inhomogeneous superconducting islands as the temperature is reduced. We then characterize the superconducting transition by quantifying several important temperature scales. The onset of superconductivity, defined as the temperature at which R drops to 90% from its normal state, occurs at ~820 mK. Zero resistance is achieved at ~350 mK. Furthermore, we have also confirmed the 2D nature of the superconductivity by checking that the data fit the picture of a 2D Berezinskii-Kosterlitz-Thouless transition, with estimated critical temperature T_{BKT} ~ 470 mK (fig. S1). For simplicity in our discussion, we define the critical transition temperature T_c as the temperature at which R equals 50% of the normal state value. For the curve in Fig. 1C, T_c ~ 580 mK, as indicated. Similar behavior is observed in device 2, in which T_c is found to be as high as ~950 mK, although the resistance never drops to zero, owing to a known imperfection in this device [figs. S2 and S3, (*23*)]. In contrast to device 1, the $R(T)$ characteristic of device 2 is a good fit to the standard Aslamazov-Larkin and Maki-Thompson fluctuation conductivity terms (*26*, *27*) (fig. S1).

The observed superconductivity, and hence the monolayer's ground state, is easily gate-tunable. We reveal this aspect by plotting $R(T)$ for different V_{bg} (Fig. 2A), while V_{tg} is kept fixed at 5 V (all data in Fig. 2 is for device 1). One can clearly see a critical gate voltage V_c^{MIT} ~ −0.75 V at which a metal-to-insulator (MIT) transition occurs. At large gate voltages $V_{bg} > V_c^{MIT}$ (i.e., toward higher electron density), R decreases with decreasing T, and a zero-resistance state is observed at low enough temperature, indicating a superconducting phase. Slightly above V_c^{MIT}, there may exist an intrinsic metallic (nonsuperconducting) phase at zero temperature, but a more detailed study is necessary to confirm this, as the observed behavior near the transition may also be consistent with the formation of both superconducting and nonsuperconducting regions as a result of density inhomogeneities. By contrast, when $V_{bg} < V_c^{MIT}$ (i.e., lower electron density), R increases with decreasing T, pointing to an insulating ground state. This insulating behavior is consistent with the expectation for monolayer WTe_2 at low density: A topologically nontrivial insulating state exists at low temperatures (*19*–*21*), and edge-state conduction is known to be marginal for the measurement configuration of device 1 (*19*). Figure S2 shows a similar transition behavior for device 2. Given that the field effect is applied through a moderate dielectric constant material (κ_{BN} ~ 3.5), these observations demonstrate that the ground state of the monolayer can be easily tuned between the two extremes of electron transport in materials, that is, from the insulating to the superconducting state. For comparison, similar field-effect behavior in existing superconducting systems is achieved only by extreme charge-density doping using ionic liquids (*28*), using ultrahigh-κ dielectric materials such as $SrTiO_3$ (*29*, *30*), or ferroelectric polarization (*31*).

The gate tunability of the superconducting state can be further characterized by the extraction of the critical temperature and measurements of the critical current. In the upper panel of Fig. 2B, we plot the extracted $T_c(V_{bg})$, where one finds increasing T_c with increasing V_{bg}. Extrapolating to T_c = 0 K (red curve, see also fig. S4), we find a critical gate voltage V_{c1} ~ −0.65 V. The maximum observed T_c ~ 0.61 K appears at the highest applicable gate voltages for this device. Figure 2C shows dV/dI versus I_{dc} at $V_{bg} = V_{tg}$ = 5 V, where the sharp

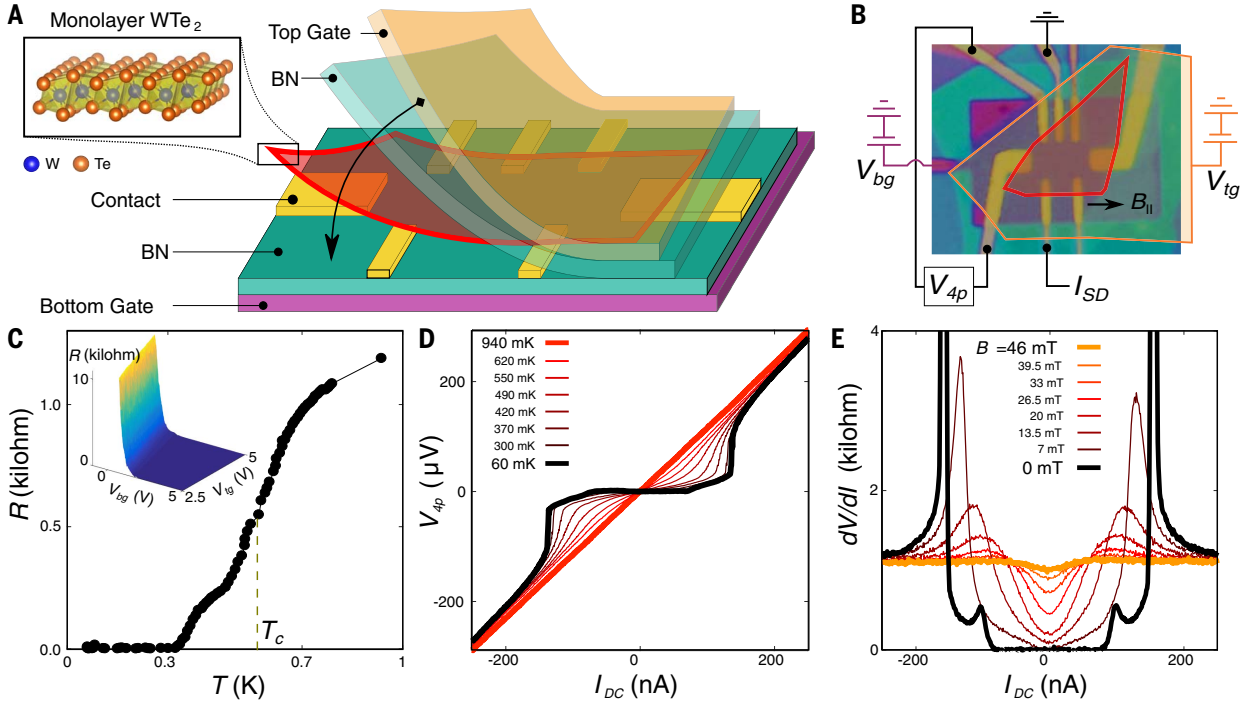


Fig. 1. Device schematic and superconductivity characteristics. (**A**) Cartoon illustration of the device structure and the crystal structure of monolayer WTe_2. (**B**) Optical microscopy image of device 1, with the monolayer WTe_2 (red) and graphite top gate (orange) highlighted. Circuit elements show the measurement configuration. V_{4p}, four-probe voltage; I_{SD}, source-to-drain current. (**C**) Temperature dependence of the resistance for V_{bg} = 4 V and V_{tg} = 5 V. The inset shows the resistance as a function of both gate voltages, at a base temperature of 60 mK. (**D**) *V-I* characteristics from base temperature (black) up to 940 mK (red). (**E**) Nonlinear *V-I* behavior, captured by differential resistance curves, at base temperature for different perpendicular magnetic fields.

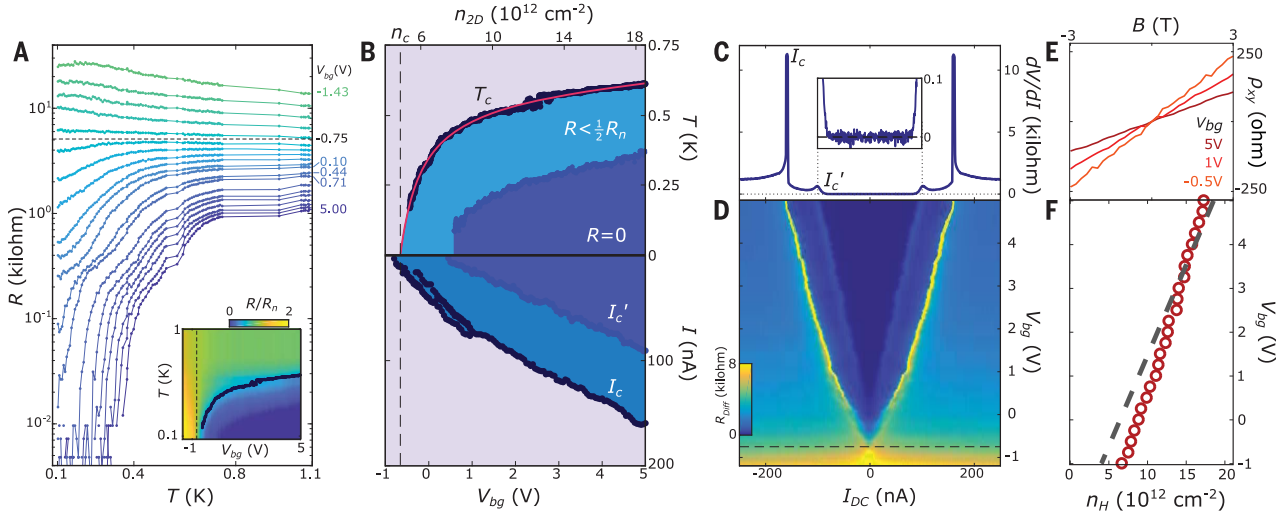


Fig. 2. Switching superconductivity on and off with an electrostatic gate. Shown are data from device 1. (**A**) *R*(*T*) characteristic for different gate voltages, showing the transition from a superconducting state to an insulating state, in evenly spaced increments between the labeled curves. The dashed line is a guide to the gate voltage, V_c^{MIT}, that separates the two regimes. The inset shows a color plot of the same data, normalized to the normal state resistance R_n, with T_c marked in black. (**B**) The upper panel shows the gate-dependent critical temperature T_c, summarized from (A). The zero-resistance region is shaded dark blue. The lower panel shows the gate-dependent critical current, I_c and I_c', summarized from (D). For V_{bg} < 1 V, I_c is found to bifurcate into two peaks, the values of which are determined by extrema of the second derivative of d*V*/d*I* (I_{dc}) data. The electron density (n_{2D}), corresponding to the bottom-gate voltage shown on the bottom axis, is estimated from the capacitance model and is shown on the top axis. (**C**) Differential resistance d*V*/d*I* versus current bias I_{dc} for V_{tg} = V_{bg} = 5V. The inset shows a zoom-in to the zero-resistance region. (**D**) Differential resistance (R_{diff}) d*V*/d*I* versus current bias I_{dc} and gate voltage V_{bg}. Close to V_{bg} ~ 1 V, the observed I_c trace bifurcates, as more clearly indicated in (B). The dashed line indicates the gate voltage at which the peaks in differential resistance merge at zero bias, indicating destruction of superconductivity. (**E**) Hall resistivity $\rho_{xy}(B)$ for three selected bottom-gate voltages taken at 2 K (V_{tg} = 5 V). See details in (*23*). (**F**) Extracted Hall density (n_H) as a function of the gate voltage (red circles) and the estimated density from the capacitance model (gray dashed line).

peak at the critical current $|I_{dc}| = I_c$ and the zero-resistance plateau for low currents demonstrate the expected behavior for superconductivity. The additional peak at $|I_{dc}| = I_c'$ in this dataset is consistent with the previously mentioned small shoulder in the $R(T)$ characteristic (Fig. 1C). The differential resistance is strongly modulated by gate voltage (Fig. 2D). The peaks at $\pm I_c$ monotonically shift toward zero bias with decreasing V_{bg} and eventually merge into a single peak at zero bias in the insulating region. This modulation of I_c is summarized in the lower panel of Fig. 2B, where we find that the critical gate voltage V_{c2}, at which I_c vanishes, falls in the range between −0.6 and −0.8 V.

Therefore, one finds that V_{c1}, V_{c2}, and V_c^{MIT} are very close to each other, which identifies the critical gate voltage as $V_{bg} = V_c \sim -0.7$ V (when $V_{tg} = 5$ V), above which superconductivity is exhibited. Using the capacitance model, we estimate the corresponding critical doping density as $n_c \sim 5 \times 10^{12}$ cm^{-2} (*23*). Hall effect measurements confirm a low absolute carrier density in the superconducting region that is consistent with the density estimates from the electrostatic capacitance model (Fig. 2, E and F, and details in fig. S5). The visible deviation appearing at lower doping in the plot may be related to inhomogeneous transport in the monolayer (e.g., the presence of conducting edge channels), which can introduce a minor factor to the measured Hall density. A similar critical density (~3×10^{12} cm^{-2}) is found in device 2 (*23*). These are among the lowest critical density values reported for 2D superconductors (*32*). Such low-density superconductivity in monolayer WTe_2 is the key to its extreme gate tunability. The maximum T_c obtained here (~0.6 K for device 1 and ~1 K for device 2, at carrier density $n_{2D} \sim 1.8 \times 10^{13}$ cm^{-2}) is relatively high, compared with other 2D superconductors with similar density (*30*, *32*).

The preliminary low-temperature electronic phase diagram of monolayer WTe_2 can now be summarized (Fig. 3A). The quantum spin Hall insulator (QSHI) phase resides near zero charge density, robust up to 100 K (*22*). With n-type doping above the critical density, n_c, the superconducting ground state develops at temperatures below T_c (superconductivity was not observed with p-type doping in the gate-accessible region). Above T_c, the system is metallic. The coexistence of QSHI and superconducting states in the same phase diagram establishes monolayer WTe_2 as a material for observing physics at the intersection of topological insulating states and superconductivity. In Fig. 3B, we show data from device 2, in which superconductivity and the quantized transport of the QSHI are observed in the same monolayer device. The helical edge transport in the QSHI phase in this device was characterized in detail in our previous report (*22*). Its superconducting behavior is characterized in fig. S3. These results therefore demonstrate the electrostatic field–effect switching of superconductivity on and off in a QSHI system. Moreover, in the same device, we also observed preliminary evidence that superconductivity can be introduced into the helical edge states by proximity effect via adjacent gate-induced superconducting regions (see details in fig. S6). In principle, such proximity-induced superconducting helical edge states can be used to construct devices hosting Majorana zero modes to study non-Abelian physics.

We further characterize the gate-tunable superconducting properties by examining the magnetic field dependence. As shown Fig. 4A, one finds an out-of-plane critical field at base temperature $B_{c2,\perp} \sim 30$ mT (for $V_{tg} = V_{bg} = 5$ V), at which half of the normal state resistance is recovered. The temperature and gate dependence of this critical field

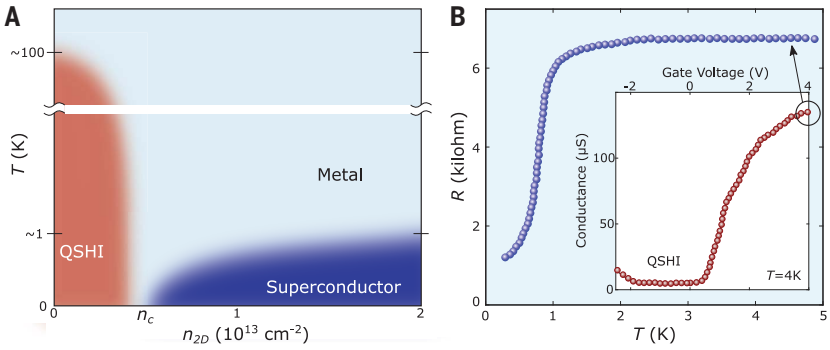


Fig. 3. Electronic phase diagram of monolayer WTe_2. (**A**) Schematic temperature-density phase diagram for the electronic ground state of monolayer WTe_2, reminiscent of the inset of Fig. 2A. (**B**) Resistance versus temperature for device 2, which shows a clear superconducting transition. The residual resistance at low temperature is caused by a known imperfection in this device, as described in (*23*). The inset shows the gate-dependent conductance of device 2, where the observed plateau corresponds to the QSHI phase. The arrow indicates the gate voltage at which the temperature dependence was recorded.

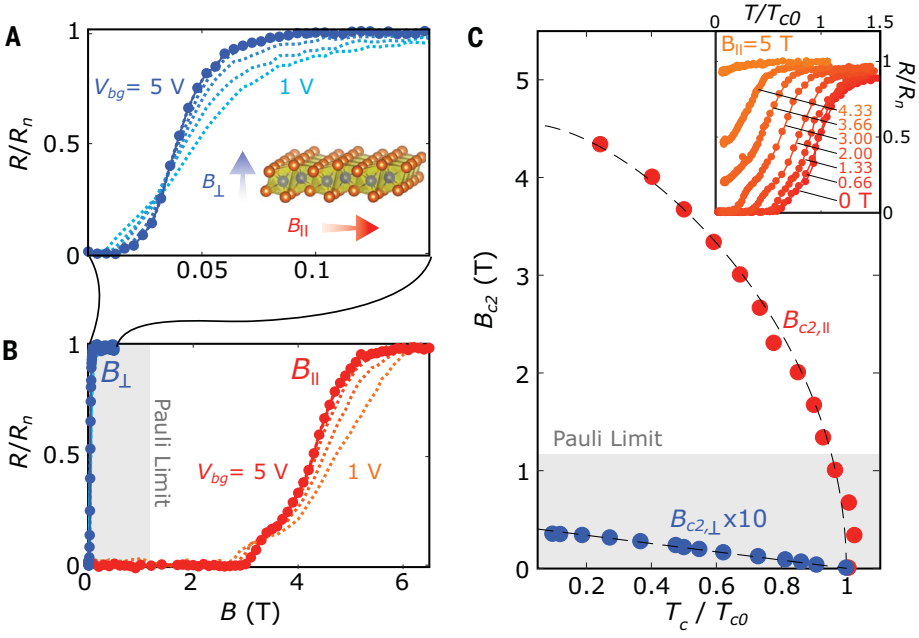


Fig. 4. Effect of magnetic field on the monolayer superconductivity. Shown are data from device 1. (**A**) Perpendicular magnetic field dependence of the resistance, $R(B_\perp)$, normalized to the normal state resistance R_n, at base temperature for different gate voltages: $V_{tg} = 5$ V and $V_{bg} = 5$, 4, 3, 2, or 1 V. The inset shows a schematic of the perpendicular and parallel field orientations. See Fig. 1B for the orientation of the parallel magnetic field in the plane of the device. (**B**) Comparison between effects of perpendicular and parallel fields. Orange represents the parallel magnetic field dependence of resistance, $R(B_\parallel)$, normalized to the normal state resistance R_n, for the same series of gate voltages as in (A). Blue represents the perpendicular magnetic field data of (A) plotted on the same scale for comparison. (**C**) B_{c2}-T_c phase diagram for both the parallel (orange) and perpendicular (blue) orientations. $V_{tg} = V_{bg} = 5$ V. Black dashed lines are fits to theoretical models (*23*). The inset shows the temperature dependence of resistance under different parallel magnetic fields.

is summarized in fig. S7, which indicates that the Ginzburg-Landau coherence length is close to 100 nm. This length is about an order of magnitude larger than the estimated transport mean free path (fig. S4), suggesting that the superconductivity is in the dirty limit. In contrast to the out-of-plane critical field, the in-plane critical field $B_{c2,\parallel}$ is substantially larger: About 4.3 T is required to recover half the normal state resistance (Fig. 4B). This field is about four times the conventional Pauli paramagnetic limit, which is given by $1.84T_c \sim 1.1$ T. We summarize the B_{c2}-T_c phase diagram for both the in-plane and out-of-plane cases in Fig. 4C. We note that the crystal symmetry and band structure of monolayer WTe_2 are very different from the transition metal dichalcogenides with hexagonal lattice structures, in which Ising-type superconductivity is responsible for exceeding the Pauli limit (*32*). Other possible mechanisms for exceeding the Pauli limit include reduced electron g-factor, spin-triplet pairing, or strong spin-orbit scattering (*33*, *34*). The superconductivity in the dirty limit is consistent with the spin-orbit scattering scenario (*33*, *34*), in which we extract a spin-orbit scattering time of ~244 fs [see fit in Fig. 4 and details in (*23*)]. However, we stress that our observations merit further study to understand the nature of the superconductivity and the exact mechanism for the large in-plane critical field in the WTe_2 monolayer.

We comment now on several interesting directions to explore. One is to look for an optimal doping, for example, to determine whether a superconducting dome exists in the phase diagram (*24*, *30*, *35*). Another is to investigate the effects of the material's strong anisotropy (*36*) and tunable noncentrosymmetry (*17*, *37*) on the superconductivity. A particularly interesting question is whether the observed superconductivity is topologically nontrivial. Although we currently do not have an answer, our results point to an exciting possibility of creating a 2D crystal–based topological superconductor using the proximity effect. Indeed, using local gates to achieve lateral modulation of superconducting and QSHI regions would enable the fabrication of tunable superconductor-QSHI-superconductor topological Josephson junctions (fig. S6) and the investigation of Majorana modes in a single material. In addition, van der Waals heterostructures that interface monolayer WTe_2 with materials such as the recently discovered 2D layered ferromagnets (*38*, *39*) can be developed for studying the interplay between superconductivity, magnetism, and topology.

REFERENCE AND NOTES

1. L. Fu, C. L. Kane, *Phys. Rev. Lett.* **100**, 096407 (2008).
2. M. Z. Hasan, C. L. Kane, *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
3. X.-L. Qi, S.-C. Zhang, *Rev. Mod. Phys.* **83**, 1057–1110 (2011).
4. Y. S. Hor *et al.*, *Phys. Rev. Lett.* **104**, 057001 (2010).
5. L. A. Wray *et al.*, *Nat. Phys.* **6**, 855–859 (2010).
6. M. Einaga *et al.*, *J. Phys. Conf. Ser.* **215**, 012036 (2010).
7. J. L. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 24–28 (2011).
8. P. P. Kong *et al.*, *J. Phys. Condens. Matter* **25**, 362204 (2013).
9. B. Sacépé *et al.*, *Nat. Commun.* **2**, 575 (2011).
10. S. Hart *et al.*, *Nat. Phys.* **10**, 638–643 (2014).
11. V. Mourik *et al.*, *Science* **336**, 1003–1007 (2012).
12. S. Nadj-Perge *et al.*, *Science* **346**, 602–607 (2014).
13. W. Chang *et al.*, *Nat. Nanotechnol.* **10**, 232–236 (2015).
14. H.-H. Sun *et al.*, *Phys. Rev. Lett.* **116**, 257003 (2016).
15. E. Bocquillon *et al.*, *Nat. Nanotechnol.* **12**, 137–143 (2017).
16. Q. L. He *et al.*, *Science* **357**, 294–299 (2017).
17. X. Qian, J. Liu, L. Fu, J. Li, *Science* **346**, 1344–1347 (2014).
18. F. Zheng *et al.*, *Adv. Mater.* **28**, 4845–4851 (2016).
19. Z. Fei *et al.*, *Nat. Phys.* **13**, 677–682 (2017).
20. S. Tang *et al.*, *Nat. Phys.* **13**, 683–687 (2017).
21. Z.-Y. Jia *et al.*, *Phys. Rev. B* **96**, 041108 (2017).
22. S. Wu *et al.*, *Science* **359**, 76–79 (2018).
23. See supplementary materials.
24. X.-C. Pan *et al.*, *Nat. Commun.* **6**, 7805 (2015).
25. D. Kang *et al.*, *Nat. Commun.* **6**, 7804 (2015).
26. V. V. Dorin, R. A. Klemm, A. A. Varlamov, A. I. Buzdin, D. V. Livanov, *Phys. Rev. B Condens. Matter* **48**, 12951–12965 (1993).
27. A. I. Larkin, A. A. Varlamov, *Theory of Fluctuations in Superconductors* (International Series of Monographs on Physics, Oxford Univ. Press, 2005).
28. Y. Saito, T. Nojima, Y. Iwasa, *Supercond. Sci. Technol.* **29**, 093001 (2016).
29. K. A. Parendo *et al.*, *Phys. Rev. Lett.* **94**, 197004 (2005).
30. A. D. Caviglia *et al.*, *Nature* **456**, 624–627 (2008).
31. C. H. Ahn *et al.*, *Science* **284**, 1152–1155 (1999).
32. Y. Saito, T. Nojima, Y. Iwasa, *Nat. Rev. Mater.* **2**, 16094 (2016).
33. R. A. Klemm, A. Luther, M. R. Beasley, *Phys. Rev. B* **12**, 877–891 (1975).
34. M. Tinkham, *Introduction to Superconductivity* (Dover, 2004).
35. J. T. Ye *et al.*, *Science* **338**, 1193–1196 (2012).
36. L. R. Thoutam *et al.*, *Phys. Rev. Lett.* **115**, 046602 (2015).
37. E. Bauer, M. Sigrist, Eds., *Non-Centrosymmetric Superconductors* (Lecture Notes in Physics Series, Springer, 2012), vol. 847.
38. B. Huang *et al.*, *Nature* **546**, 270–273 (2017).
39. C. Gong *et al.*, *Nature* **546**, 265–269 (2017).
40. S. Wu, Replication data for: Electrically tunable low-density superconductivity in a monolayer topological insulator. Harvard Dataverse (2018); https://doi.org/10.7910/DVN/UFGFY2.

ACKNOWLEDGMENTS

We thank L. Fu for helpful discussions and J. I.-J. Wang for assistance in operating the dilution refrigerator. **Funding:** This work was partly supported through AFOSR grant no. FA9550-16-1-0382 as well as the Gordon and Betty Moore Foundation's EPiQS Initiative through grant no. GBMF4541 to P.J.-H. Device nanofabrication was partly supported by the Center for Excitonics, an Energy Frontier Research Center funded by the DOE, Basic Energy Sciences Office, under award no. DE-SC0001088. This work made use of the Materials Research Science and Engineering Center's Shared Experimental Facilities supported by NSF under award no. DMR-0819762. Sample fabrication was performed partly at the Harvard Center for Nanoscale Science supported by the NSF under grant no. ECS-0335765. S.W. acknowledges the support of the MIT Pappalardo Fellowship in Physics. The WTe_2 crystal growth performed at Princeton University was supported by an NSF MRSEC grant, DMR-1420541. Growth of BN crystals was supported by the Elemental Strategy Initiative conducted by the MEXT, Japan, and JSPS KAKENHI grant nos. JP15K21722 and JP25106006. **Author contributions:** V.F. and S.W. fabricated the devices, performed electronic transport measurements, and analyzed the data. Y.C. and L.B. helped with measurements. Q.D.G. and R.J.C. grew the WTe_2 crystals. K.W. and T.T. grew the BN crystals. P.J.-H. supervised the project. V.F., S.W., and P.J.-H. co-wrote the manuscript with input from all co-authors. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Hexagonal boron nitride crystals were provided by K.W. and T.T. under a material transfer agreement with NIMS. All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials and have been deposited at (*40*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/926/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
References (*41*–*47*)

12 November 2017; accepted 9 October 2018
Published online 25 October 2018
10.1126/science.aar4642

TOPOLOGICAL MATTER

Observation of the topological Anderson insulator in disordered atomic wires

Eric J. Meier[1], Fangzhao Alex An[1], Alexandre Dauphin[2], Maria Maffei[2,3], Pietro Massignan[2,4]*, Taylor L. Hughes[1]*, Bryce Gadway[1]*

Topology and disorder have a rich combined influence on quantum transport. To probe their interplay, we synthesized one-dimensional chiral symmetric wires with controllable disorder via spectroscopic Hamiltonian engineering, based on the laser-driven coupling of discrete momentum states of ultracold atoms. Measuring the bulk evolution of a topological indicator after a sudden quench, we observed the topological Anderson insulator phase, in which added disorder drives the band structure of a wire from topologically trivial to nontrivial. In addition, we observed the robustness of topologically nontrivial wires to weak disorder and measured the transition to a trivial phase in the presence of strong disorder. Atomic interactions in this quantum simulation platform may enable realizations of strongly interacting topological fluids.

Topology and disorder share many surprising connections, from the formal similarity of one-dimensional (1D) pseudodisordered lattices and 2D integer quantum Hall Hofstadter lattices (*1*, *2*) to the deep connection between the symmetry classes of random matrices (*3*) and the classification of symmetry-protected topological phases (*4*). Recently, there has been great interest in exploring both disorder (*5*) and topology (*6*) through quantum simulation, stemming from the dramatic influences that these ingredients can have, separately, on the localization properties of quantum particles (*7*, *8*). When combined, disorder and topology can have a rich and varied influence on quantum transport (*9*). Indeed, one of the hallmark features of topological insulators (TIs) is the topologically protected boundary states that are immune to certain types of disorder up to some characteristic strength (*10*). The robust conductance of such boundary states, such as the 1D edge states of integer quantum Hall systems (*8*) or the 2D surface states of 3D TIs (*11*), serves as an important counterexample to the inevitability of localization in low-dimensional disordered systems (*7*, *12*). However, topological features can eventually disappear when the disorder strength becomes too large, and unusual critical phenomena related to the unwinding of the topology can accompany such transitions (*13*, *14*).

Conversely, static disorder can induce nontrivial topology when added to a trivial band structure. This disorder-driven topological phase, known as the topological Anderson insulator (TAI), was first predicted to occur in metallic 2D HgTe/CdTe quantum wells (*15*). There has been much interest in the TAI phase over the past decade (*15*–*17*), and many theoretical studies have shown the TAI phenomenon to be quite general, emerging across a range of disordered systems (*18*–*21*). However, owing to the lack of precise control over disorder in real materials and the difficulty in engineering both topology and disorder in most quantum simulators, the TAI has so far evaded experimental realization. In this work, we reach this goal by engineering synthetic 1D chiral symmetric wires with precisely controllable disorder using simultaneous, coherent control over many transitions between discrete quantum states of ultracold atoms.

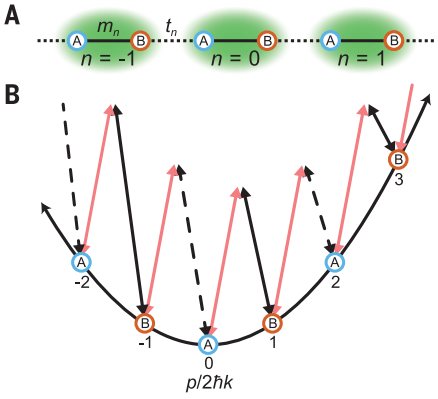


Fig. 1. Synthetic chiral symmetric wires engineered with atomic momentum states. (**A**) Schematic lattice of the nearest-neighbor-coupled chiral symmetric wire. Site-to-site links within the unit cell (solid) and those connecting different unit cells (dashed) have independent tunneling energies m_n and t_n, respectively. (**B**) Schematic of the experimental implementation of the tight-binding model depicted in (A), with tunnelings based on two-photon Bragg transitions between discrete atomic motional states with momentum p.

The topological band structures we consider are 1D TIs based on the Su-Schrieffer-Heeger model with a chiral, or sublattice, symmetry (*4*, *19*, *22*, *23*). We describe this system in terms of a tight-binding model with a two-site unit cell, consisting of sublattice sites A and B (depicted in Fig. 1A). We consider the Hamiltonian

$$H = \sum_n \left[m_n c_n^\dagger S c_n + t_n \left(c_{n+1}^\dagger \frac{(\sigma_1 - i\sigma_2)}{2} c_n + \text{h.c.} \right) \right] \quad (1)$$

where $c_n^\dagger = (c_{n,\text{A}}^\dagger, c_{n,\text{B}}^\dagger)$ creates a particle at unit cell n in sublattice site A or B, c_n is the corresponding annihilation operator, h.c. denotes the Hermitian conjugate, and σ_i are the Pauli matrices related to the sublattice degree of freedom (*23*). The m_n and t_n characterize the intra- and intercell tunneling energies, respectively. This model can describe chiral wires of the AIII or BDI symmetry classes, by choosing the intracell hopping term to be $S = \sigma_1$ (BDI) or $S = \sigma_2$ (AIII). Both the AIII (chiral unitary) and the BDI (chiral orthogonal) class models respect chiral symmetry—that is, they obey $\Gamma H \Gamma = -H$ with the chiral operator $\Gamma = \sigma_3 \otimes \mathbb{I}$—whereas the BDI class also obeys particle-hole and time-reversal symmetries (*4*). In this work we chose to study both BDI and AIII class systems because they represent all possible distinguishable chiral classes for which the $\mathbb{Z}$ topological invariant in one dimension, the winding number, is defined.

We experimentally implement effective tight-binding models of the form of Eq. 1 using the controlled, parametric coupling of many discrete momentum states of ultracold atoms (*24*). We start with a weakly trapped Bose-Einstein condensate (BEC) of ^{87}Rb atoms and apply a pair of counter-propagating laser fields with nominal wavelength λ and wave vector $k = 2\pi/\lambda$. These lasers are far-detuned from any atomic transition; however, their interference pattern couples to the atoms through the ac Stark effect. The spatial periodicity of the laser interference pattern, π/k, defines the set of momentum states whose momenta are separated by integer values of $2\hbar k$ (where $\hbar = 2\pi\hbar$ is the Planck constant). Atoms initially in the BEC, which is a source of atoms with essentially zero momentum, may undergo transitions between these many discrete momentum states, which represent the sites of our synthetic lattice. The effective tunneling of atoms between these sites is precisely controlled by simultaneously driving many two-photon Bragg transitions with the applied laser fields. The individual, spectroscopically resolved control over many such transitions is allowed for by the Doppler shifts experienced by the atoms, which are specific to the various Bragg transitions

[1]Department of Physics, University of Illinois at Urbana–Champaign, Urbana, IL 61801-3080, USA. [2]ICFO–Institut de Ciencies Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels, Barcelona, Spain. [3]Dipartimento di Fisica, Università di Napoli Federico II, Complesso Universitario di Monte Sant'Angelo, Via Cintia, 80126 Napoli, Italy. [4]Departament de Física, Universitat Politècnica de Catalunya, Campus Nord B4-B5, 08034 Barcelona, Spain.
***Corresponding author. Email: bgadway@illinois.edu (B.G.); hughest@illinois.edu (T.L.H.); pietro.massignan@upc.edu (P.M.)**

(Fig. 1B). This provides local (in momentum space) control of the intra- and intercell tunneling amplitudes and phases, directly through the amplitudes and phases of the corresponding Bragg laser fields (*24*). This control gives us access to both BDI and AIII class wires, in contrast to previous studies based on real-space superlattices that were restricted to exploring BDI wires, because quantum tunneling between stationary lattice sites is real-valued (*25*–*27*).

The ability to create precisely defined disorder in the off-diagonal tunneling terms is crucial for this study. Unlike the site-potential disorder that is more naturally realized in real-space cold atom experiments—for example, through optical speckle (*28*) or quasiperiodic lattice potentials (*29*)—pure tunneling disorder is important for preserving the chiral symmetry of our wires (*19*, *23*). In particular, we let

$$t_n = t(1 + W_1\omega_n) \tag{2}$$

$$m_n = t(m + W_2\omega'_n) \tag{3}$$

define the controlled fluctuations of our hopping terms, where t is the characteristic intercell tunneling energy, m is the ratio of intra- to intercell tunneling in the clean limit, ω_n and ω'_n are independent random real numbers chosen uniformly from the range [−0.5, 0.5], and W_1 and W_2 are the dimensionless disorder strengths applied to inter- and intracell tunneling.

We begin by considering the influence of disorder added to a BDI-class wire. The wire is strongly dimerized, as characterized by a small intracell-to-intercell tunneling ratio of $m = 0.100(5)$ (with $t/\hbar \approx 2\pi \times 1.2$ kHz), and hence is in the topological regime in the clean limit. We fix the disorder amplitudes to be $W \equiv W_2 = 2W_1$ and show in Fig. 2A the disorder-averaged topological phase diagram of this model as a function of W and m, as determined numerically by a real-space calculation of the winding number ν for a system with 200 unit cells, together with the critical phase boundary predicted for an infinite system based on the divergence of the localization length Λ (*23*, *30*).

The strong dimerization produces a large (in units of the bandwidth) energy gap in the band structure. Such large bandgaps are typically favorable for experimentally observing the topological nature of disorder-free nontrivial wires via adiabatic charge pumping (*25*, *26*) or the adiabatic preparation of boundary states (*31*). However, it is expected that in disordered chiral symmetric wires, the bulk energy gap will essentially vanish at moderate disorder strengths, well below those required to induce a change in topology (*23*). The energy gap is replaced by a mobility gap, and the band insulator of the clean system is replaced by an Anderson insulator that remains topological, with topology carried by localized states in the spectrum (*23*). Thus, without the spectral gap, experimental probes relying on adiabaticity are expected to fail in evidencing the topology of disordered wires.

We instead characterize the topology of our wires by monitoring the bulk dynamical response of atoms to a sudden quench. Specifically, we measure the mean chiral displacement of our atoms. This observable was recently introduced in the context of discrete-time photonic quantum walks (*32*); here we measure it for continuous-time dynamics. We define the expectation value of the chiral displacement operator as

$$\mathcal{C} = 2\langle \Gamma X \rangle \tag{4}$$

given in terms of the chiral operator Γ and the unit cell operator X (*32*). The dynamics of $\mathcal{C}$, in general, display a transient, oscillatory behavior, and their time- and disorder-average $\langle \bar{\mathcal{C}} \rangle$ converges to the winding number ν, or equivalently to the Zak phase φ_{Zak} divided by π, in both the clean and the disordered cases. Moreover, at topological critical points, $\langle \bar{\mathcal{C}} \rangle$ converges to the average of the invariants computed in the two neighboring phases (*30*, *32*, *33*).

For our experiment, we begin with all tunnel couplings turned off, and the entire atomic population localized at a single central bulk lattice site (site A of unit cell $n = 0$, for a system with 20 unit cells). We then quench on the tunnel couplings in a stepwise fashion. The projection of the localized initial state onto the quenched system's eigenstates leads to rich dynamics, as depicted in Fig. 2B for both weak ($W = 0.5$) and strong ($W = 5$) disorder. Such site-resolved dynamics of the atomic population distribution are directly measured by a series of absorption images taken after dynamical evolution under the Hamiltonian of Eq. 1 for a variable time τ (given in units

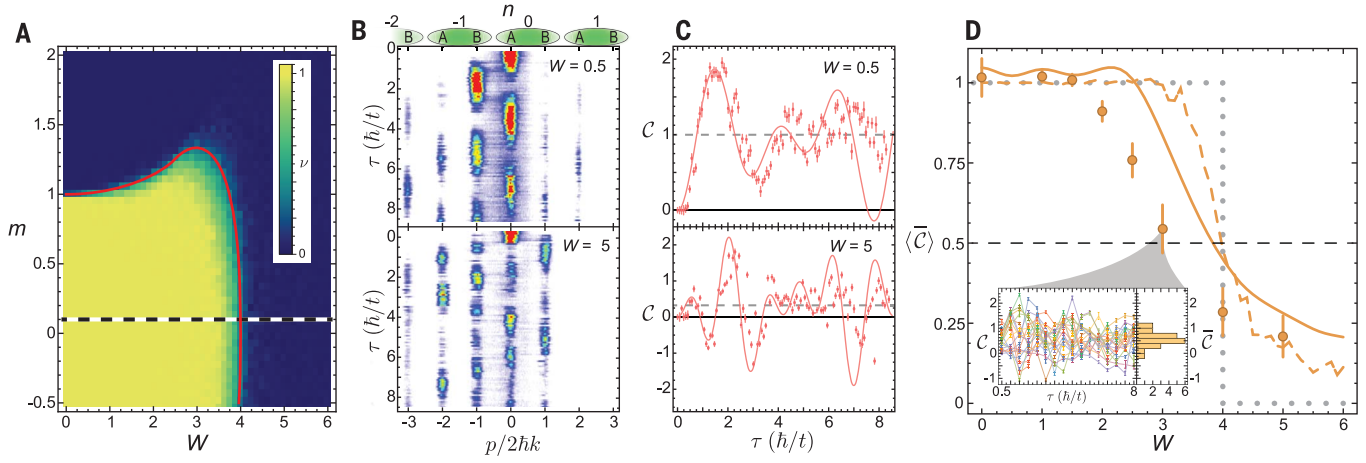


Fig. 2. Disorder-driven transition from topological to trivial wires. (**A**) Calculated topological phase diagram of the BDI wire model described in Eq. 1, showing the winding number ν (inset color scale) as a function of disorder strength W and tunneling ratio m with tunneling disorder strengths $W \equiv W_2 = 2W_1$. The striped black and white line at $m = 0.1$ indicates the region explored experimentally in (B) to (D). The solid red curve indicates the critical phase boundary (i.e., the set of points where the localization length Λ diverges for an infinite chain) (*23*, *30*). (**B**) Integrated absorption images of the bulk dynamics after a sudden quench of the tunnel couplings, for both weak disorder ($W = 0.5$) and strong disorder ($W = 5$), each for a single disorder configuration. (**C**) Dynamics of $\mathcal{C}$ calculated from the data shown in (B). The solid red curves are numerical simulations according to Eq. 1 with no free parameters (*30*). The dashed gray horizontal lines denote $\bar{\mathcal{C}}$ for each dataset. (**D**) $\langle \bar{\mathcal{C}} \rangle$ as a function of W for $m = 0.100(5)$. The data are averaged over 20 independent disorder configurations and times in the range 0.5 $\hbar/t$ to 8 $\hbar/t$ in steps of 0.5 $\hbar/t$. The solid gold line represents a numerical simulation according to Eq. 1, with no free parameters for 200 disorder configurations but with the same finite-time sampling as the data (*30*). The dashed gold line is based on the same simulation as the solid gold line but sampled to much longer times (τ = 1000 $\hbar/t$) in a wire with 250 unit cells (*30*). The dotted gray curve shows the topological index in the thermodynamic limit (*23*), which changes value at the same point as the red line position in (A). This topological index takes a value of 0.5 at the critical point, as indicated by the horizontal dashed line. The inset shows $\mathcal{C}$ for $W = 3$ as a function of time for all 20 disorder configurations (distinguished by color) with the number of disorder configurations corresponding to various values of $\bar{\mathcal{C}}$ shown in the histogram. All error bars in (C) and (D) denote one standard error of the mean.

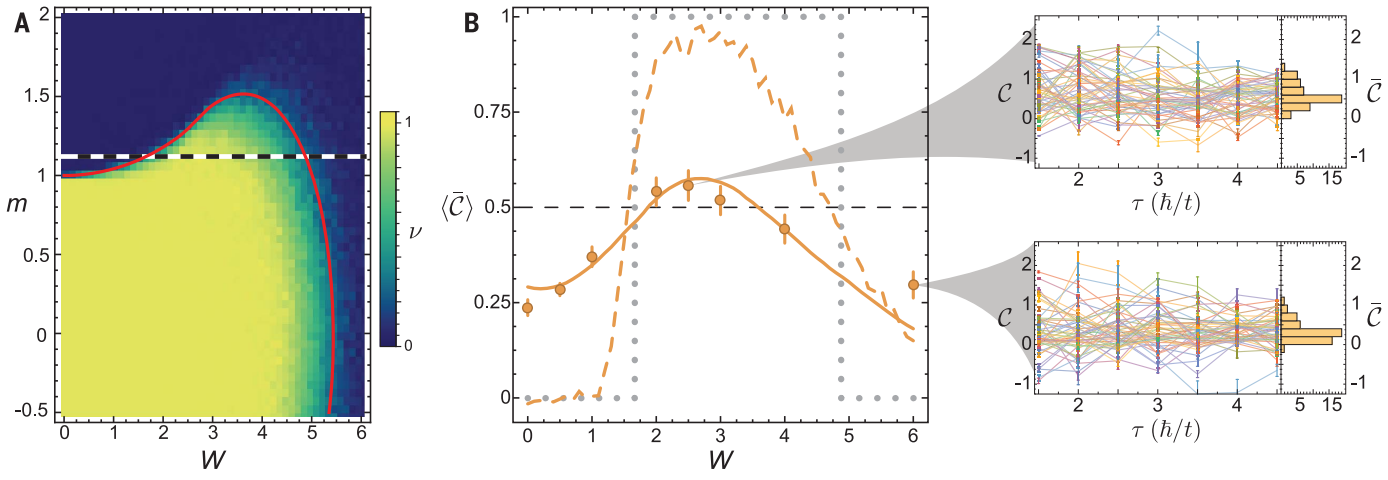


Fig. 3. Observation of the TAI phase. (**A**) Calculated topological phase diagram of the AIII wire model described in Eq. 1, showing the computed winding number (color scale at right) as a function of disorder strength W and tunneling ratio m with tunneling disorder strengths $W \equiv W_2$ ($W_1 = 0$). The striped black and white line at $m = 1.12$ indicates the region explored experimentally in (B). The solid red curve indicates the critical boundary (i.e., the set of points where the localization length Λ diverges for an infinite chain) (*23*, *30*). (**B**) $\langle \bar{C} \rangle$ as a function of W for $m = 1.12(2)$. The data are averaged over 50 independent disorder configurations and are averaged in time over the range 1.5 $\hbar/t$ to 4.5 $\hbar/t$ in steps of 0.5 $\hbar/t$. The solid gold line shows a numerical simulation according to Eq. 1 for 200 disorder configurations but with the same finite-time sampling as the data (*30*). The dashed gold line is based on the same simulation as the solid gold line but sampled to much longer times ($\tau = 1000\ \hbar/t$) in a 250–unit cell system (*30*). The dotted gray curve shows the topological index in the thermodynamic limit (*23*), which changes value at the same point as the red line position in (A). This topological index takes a value of 0.5 at the critical points, as indicated by the horizontal dashed line. $\mathcal{C}$ as a function of time for all 50 disorder realizations (distinguished by color) is shown at the right for $W = 2.5$ and 6; histograms of $\bar{\mathcal{C}}$ are shown to the right of each plot. All error bars in (B) denote one standard error of the mean.

of the tunneling time $\hbar/t \approx 130\ \mu s$) and after the discrete momentum states separate according to their momenta during a time-of-flight period (*24*). From the data shown in Fig. 2B, we calculate $\mathcal{C}$ as a function of τ as shown in Fig. 2C, along with the time average $\bar{\mathcal{C}}$. We additionally obtain the disorder-averaged topological characterization of the system $\langle \bar{\mathcal{C}} \rangle$ by averaging $\bar{\mathcal{C}}$ over many independent disorder configurations.

The dependence of $\langle \bar{\mathcal{C}} \rangle$ on the strength of applied disorder W is summarized in Fig. 2D. The inset of Fig. 2D depicts the determination of $\langle \bar{\mathcal{C}} \rangle$ (shown for the case $W = 3$), first from the time average of $\mathcal{C}$ over 16 values of τ evenly spaced between $0.5\hbar/t$ and $8\hbar/t$, followed by an average over 20 unique realizations of disorder. We observe that $\langle \bar{\mathcal{C}} \rangle$ is robust to weak disorder, maintaining a nearly quantized value close to 1. For strong disorder, $W \gtrsim 2$, we observe a relatively steep drop in $\langle \bar{\mathcal{C}} \rangle$, with it falling below $\langle \bar{\mathcal{C}} \rangle = 0.5$ for $W \gtrsim 3$. Our observed decrease of $\langle \bar{\mathcal{C}} \rangle$ with increasing disorder is in reasonably good agreement with a numerical simulation (solid gold line in Fig. 2D) of the Hamiltonian in Eq. 1 for experimental time scales. The observed decay of $\langle \bar{\mathcal{C}} \rangle$ is associated with a disorder-driven transition between topological ($W \lesssim 4$) and trivial wires ($W \gtrsim 4$).

On an infinitely long chain, we would expect to observe a sharp phase transition in the infinite-time limit of our $\langle \bar{\mathcal{C}} \rangle$ measurement, yielding quantized values of the invariant for all disorders, and half-integer values at the critical phase boundary (*30*, *34*). However, we instead observe a smooth crossover owing to broadening from our finite period of quench dynamics and the corresponding finite number of sites. The observation of a moderately sharper transition, such as that of the dashed-line numerical simulation in Fig. 2D, would require that we measure at extremely long time scales (shown for 1000 tunneling times) and for very large systems (shown for 250 unit cells), which at the moment is beyond the capabilities of our experimental technique (*30*). The slow convergence of this transition with increasing measurement time and system size is a characteristic feature of random-singlet transitions (*14*), such as those found in chiral symmetric wires at strong disorder.

Having demonstrated a disorder-driven change of topology in BDI-class wires, we now turn our attention to AIII-class wires, for which we investigate the surprising feature that an initially clean, trivial system can be driven topological through the addition of disorder. This phenomenon is manifest in the calculated phase diagram of AIII-class wires shown in Fig. 3A for m just exceeding 1. The value $|m| = 1$ is the critical point between the topological and trivial phase in the clean limit, and values of $|m| > 1$ are in the trivial phase in the absence of disorder. However, we see that random tunneling disorder induces the TAI phase over a broad range of weak to moderate W values, eventually giving way to a trivial Anderson insulator phase again for very large disorder. Beyond numerics, a mechanism for the formation of a TAI phase was first elaborated in (*17*) for 2D systems. In that work, disorder is taken into account perturbatively using the self-consistent Born approximation and was shown to effectively renormalize the parameters in the Hamiltonian [including the parameter(s) that tune between the topological and trivial phases]. The TAI phase arises because, as disorder is added to the trivial phase tuned near the clean critical point, the effective Hamiltonian is renormalized through the critical point and into the topological phase. This type of reasoning was adapted, with strong support from numerical evidence, and extended to describe the TAI phase in 1D systems, including both the BDI- and AIII-class wires that we consider here (*19*, *23*, *34*).

Here we probe the influence of tunneling disorder on atomic wires of the AIII class. Because we are interested in the TAI phase, we start with a slight dimerization [$m = 1.12(2)$] that places the system in a trivial phase in the clean limit. We note that being so near the critical point at $m = 1$ causes the bandgap in the clean limit to be much smaller than in the previous experimental setup. The choice of disorder we consider here differs from the previous case: We add disorder only to the intracell hopping terms, that is, setting $W_1 = 0$ and $W \equiv W_2$. The difference in the topological phase diagrams for the BDI case (Fig. 2A) and the AIII case (Fig. 3A) comes entirely from this change in disorder configuration. In one dimension, the topological phase diagrams for BDI and AIII systems are identical when exposed to equivalent disorder configurations (W_1 and W_2 values) and with only nearest-neighbor tunnelings present. From (*17*, *19*, *34*), we expect that, for weak disorder of this form, the intracell hopping m should be renormalized toward the topological phase, resulting in a TAI. Thanks to the smaller bandgap in this case of reduced dimerization, the effects of off-resonant driving (*24*, *30*) in our system become more pronounced. To mitigate these effects, we reduce our tunneling

energy to $t/\hbar \approx 2\pi \times 600$ Hz, resulting in a correspondingly lessened experimental time range of $\tau = 1.5\ \hbar/t$ to $4.5\ \hbar/t$.

Figure 3B shows the dependence of $\langle \bar{C} \rangle$ on the strength of added disorder in the AIII-class wire. The measured $\langle \bar{C} \rangle$ values are obtained, as in Fig. 2, through the nonequilibrium bulk dynamics of the atoms after a quench of the tunneling. Because of the restricted range of τ, we include many more disorder configurations (50) to allow for stable measures of $\langle \bar{C} \rangle$. For weak disorder, $\langle \bar{C} \rangle$ rises and reaches a pronounced maximum at $W \approx 2.5$. This is consistent with the expected change in the renormalized m parameter, that is, given the negative sign of the lowest-order correction to m, for weak disorder (*17*, *19*, *34*). $\langle \bar{C} \rangle$ then decays for very strong applied disorder. This observation of an initial increase of $\langle \bar{C} \rangle$ followed by a decrease is indicative of two phase transitions, first from trivial wires to the TAI phase and then to a trivial Anderson insulator at strong disorder, broadened by our finite interrogation time.

Despite the effects of finite-time broadening, we see that our measured $\langle \bar{C} \rangle$ rises to a value greater than 0.5 (the infinite-time $\langle \bar{C} \rangle$ value associated with the critical point) for $W \approx 2.5$, lending further evidence to our observation of the TAI phase. Based on the measurements of $\langle \bar{C} \rangle$ for $W = (2.0, 2.5, \text{and } 3.0)$ and their statistical errors, there is only a 0.3% chance for all three measurements to fall below 0.5, the critical value that indicates a change in topology (*30*). Further, the excellent agreement of our experimental $\langle \bar{C} \rangle$ data with a short-time sampled numerical simulation (solid gold line in Fig. 3B), combined with the sharper transitions expected for long-time measurements based on the same simulations (dashed gold line in Fig. 3B for 1000 tunneling times in a 250–unit cell system), provide strong evidence for the observation of disorder-driven topology in an otherwise trivial band structure (*30*).

Unlike condensed-matter systems and photonic simulators, where carrier mobility or lattice parameters may vary from sample to sample, the spectroscopic control of our atomic physics platform has allowed us to engineer many different, precisely tuned realizations of disorder. This level of control will also enable future studies of quantum criticality in disordered topological systems (*9*, *13*). By simple extension to longer evolution times, the interesting physics of logarithmic delocalization at the random-singlet transition may be studied (*14*). Combined with the ability to engineer tunneling phases (*24*) and artificial gauge fields, our technique may be extended to study disordered quantum Hall systems (*13*). And although our present study has been restricted to a regime where interactions are relatively unimportant, the presence of strong interactions in synthetic momentum-space lattices (*35*) will enable future studies of strongly interacting topological fluids.

Note added in proof: After completion and submission of this work, a related work presented complementary evidence for the TAI in photonic waveguide arrays (*36*).

REFERENCES AND NOTES

1. Y. E. Kraus, Y. Lahini, Z. Ringel, M. Verbin, O. Zilberberg, *Phys. Rev. Lett.* **109**, 106402 (2012).
2. D. R. Hofstadter, *Phys. Rev. B* **14**, 2239–2249 (1976).
3. A. Altland, M. R. Zirnbauer, *Phys. Rev. B* **55**, 1142–1161 (1997).
4. S. Ryu, A. P. Schnyder, A. Furusaki, A. W. W. Ludwig, *New J. Phys.* **12**, 065010 (2010).
5. L. Sanchez-Palencia, M. Lewenstein, *Nat. Phys.* **6**, 87–95 (2010).
6. N. Goldman, J. C. Budich, P. Zoller, *Nat. Phys.* **12**, 639–645 (2016).
7. P. W. Anderson, *Phys. Rev.* **109**, 1492–1505 (1958).
8. K. von Klitzing, G. Dorda, M. Pepper, *Phys. Rev. Lett.* **45**, 494–497 (1980).
9. F. Evers, A. D. Mirlin, *Rev. Mod. Phys.* **80**, 1355–1417 (2008).
10. X.-L. Qi, S.-C. Zhang, *Rev. Mod. Phys.* **83**, 1057–1110 (2011).
11. Y. L. Chen *et al.*, *Science* **325**, 178–181 (2009).
12. E. Abrahams, P. W. Anderson, D. C. Licciardello, T. V. Ramakrishnan, *Phys. Rev. Lett.* **42**, 673–676 (1979).
13. H. Aoki, *J. Phys. C Solid State Phys.* **16**, L205–L208 (1983).
14. D. S. Fisher, *Phys. Rev. B* **50**, 3799–3821 (1994).
15. J. Li, R.-L. Chu, J. K. Jain, S.-Q. Shen, *Phys. Rev. Lett.* **102**, 136806 (2009).
16. H. Jiang, L. Wang, Q. Sun, X. C. Xie, *Phys. Rev. B* **80**, 165316 (2009).
17. C. W. Groth, M. Wimmer, A. R. Akhmerov, J. Tworzydło, C. W. J. Beenakker, *Phys. Rev. Lett.* **103**, 196805 (2009).
18. H.-M. Guo, G. Rosenberg, G. Refael, M. Franz, *Phys. Rev. Lett.* **105**, 216601 (2010).
19. A. Altland, D. Bagrets, L. Fritz, A. Kamenev, H. Schmiedt, *Phys. Rev. Lett.* **112**, 206602 (2014).
20. P. Titum, N. H. Lindner, M. C. Rechtsman, G. Refael, *Phys. Rev. Lett.* **114**, 056801 (2015).
21. C. Liu, W. Gao, B. Yang, S. Zhang, *Phys. Rev. Lett.* **119**, 183901 (2017).
22. W. P. Su, J. R. Schrieffer, A. J. Heeger, *Phys. Rev. Lett.* **42**, 1698–1701 (1979).
23. I. Mondragon-Shem, T. L. Hughes, J. Song, E. Prodan, *Phys. Rev. Lett.* **113**, 046802 (2014).
24. E. J. Meier, F. A. An, B. Gadway, *Phys. Rev. A* **93**, 051602 (2016).
25. M. Lohse, C. Schweizer, O. Zilberberg, M. Aidelsburger, I. Bloch, *Nat. Phys.* **12**, 350–354 (2016).
26. S. Nakajima *et al.*, *Nat. Phys.* **12**, 296–300 (2016).
27. C. G. Velasco, B. Paredes, *Phys. Rev. Lett.* **119**, 115301 (2017).
28. J. Billy *et al.*, *Nature* **453**, 891–894 (2008).
29. G. Roati *et al.*, *Nature* **453**, 895–898 (2008).
30. Full details on the experimental and theoretical procedures are available as supplementary materials.
31. E. J. Meier, F. A. An, B. Gadway, *Nat. Commun.* **7**, 13986 (2016).
32. F. Cardano *et al.*, *Nat. Commun.* **8**, 15516 (2017).
33. M. Maffei, A. Dauphin, F. Cardano, M. Lewenstein, P. Massignan, *New J. Phys.* **20**, 013023 (2018).
34. A. Altland, D. Bagrets, A. Kamenev, *Phys. Rev. B* **91**, 085429 (2015).
35. F. A. An, E. J. Meier, J. Ang'ong'a, B. Gadway, *Phys. Rev. Lett.* **120**, 040407 (2018).
36. S. Stützer *et al.*, *Nature* **560**, 461–465 (2018).
37. E. J. Meier *et al.*, Replication data for: Observation of the topological Anderson insulator in disordered atomic wires, Version 1, Harvard Dataverse (2018); https://doi.org/10.7910/DVN/7BMFTD.

ACKNOWLEDGMENTS

We thank N. Goldman, M. Lewenstein, H. Shapourian, and I. Mondragon-Shem for helpful discussions. **Funding:** This material is based upon work supported by the National Science Foundation under grant no. PHY1707731 (E.J.M., F.A.A., and B.G.). A.D., M.M., and P.M. acknowledge Spanish MINECO (Severo Ochoa SEV-2015-0522, FisicaTeAMO FIS2016-79508-P, and SWUQM FIS2017-84114-C2-1-P), the Generalitat de Catalunya (SGR874 and CERCA), the EU (ERC AdG OSYRIS 339106 and H2020-FETProAct QUIC 641122), the Fundació Privada Cellex, a Cellex-ICFO-MPQ fellowship, and the "Ramón y Cajal" program. T.L.H. was supported by the ONR YIP Award N00014-15-1-2383. **Author contributions:** E.J.M. and F.A.A. performed the experiments. E.J.M. analyzed the data. B.G. supervised the experiments and data analysis. A.D., M.M., P.M., and T.L.H. developed the theoretical framework. All authors discussed the results and contributed to the preparation of the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data shown in this work can be found in tables S1 to S3 in the supplementary materials or in an online database (*37*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/929/suppl/DC1
Methods
Supplementary Text
Figs. S1 to S7
Tables S1 to S3
References (*38*–*44*)

19 February 2018; accepted 27 September 2018
Published online 11 October 2018
10.1126/science.aat3406

METALLURGY

Multicomponent intermetallic nanoparticles and superb mechanical behaviors of complex alloys

T. Yang[1,2], Y. L. Zhao[2], Y. Tong[2], Z. B. Jiao[3], J. Wei[2], J. X. Cai[4], X. D. Han[4], D. Chen[2], A. Hu[2], J. J. Kai[2], K. Lu[5], Y. Liu[6], C. T. Liu[1,2]*

Alloy design based on single–principal-element systems has approached its limit for performance enhancements. A substantial increase in strength up to gigapascal levels typically causes the premature failure of materials with reduced ductility. Here, we report a strategy to break this trade-off by controllably introducing high-density ductile multicomponent intermetallic nanoparticles (MCINPs) in complex alloy systems. Distinct from the intermetallic-induced embrittlement under conventional wisdom, such MCINP-strengthened alloys exhibit superior strengths of 1.5 gigapascals and ductility as high as 50% in tension at ambient temperature. The plastic instability, a major concern for high-strength materials, can be completely eliminated by generating a distinctive multistage work-hardening behavior, resulting from pronounced dislocation activities and deformation-induced microbands. This MCINP strategy offers a paradigm to develop next-generation materials for structural applications.

High-performance materials with gigapascal strengths and large ductility are highly desirable for enhancing engineering reliability and energy efficiency, as well as reducing CO_2 emissions for material production. However, developing advanced materials with a substantial improvement of both strength and ductility is highly challenging because of strength and ductility's mutually exclusive relationship (*1*). Single-phase alloys usually display a good ductility but relatively low strengths. The introduction of nanotwins and transformation-induced martensites has shown their capability of giving a combined increase of strength (*2*, *3*). However, the yield strength obtained from these approaches remains limited, which is generally insufficient for structural applications. Second-phase intermetallic compounds (IMCs) provide an efficient approach for enhancing alloy strengths (*4*, *5*); however, most IMCs with atomically ordered structures are intrinsically brittle. The gain of gigapascal strengths by introducing high-density IMCs invariably leads to a reduced resistance to fracture. The highly reactive elements, such as Al in the Ni_3Al intermetallic phase, will also increase the susceptibility of these intermetallics to moisture-induced environmental embrittlement and further lower their tensile ductility (*6*, *7*). Moreover, the microstructural heterogeneity of IMCs tends to introduce a localized stress-strain concentration and trigger microcracks under loading (*8*). Consequently, the early onset of plastic instability leads to a catastrophic failure of these materials.

Conventional alloy design based on single–principal-element alloy systems cannot break through this thorny dilemma, because of the limited abilities for further optimizing alloy chemistries and microstructures. Recently proposed metallurgical design in multi–principal-element alloy systems offers a promising pathway to alleviate these concerns (*2*, *9*–*12*). Nevertheless, the results obtained so far have been disappointing at ambient temperatures, at which high-strength

[1]Department of Material Science and Engineering, College of Science and Engineering, City University of Hong Kong, Hong Kong, China. [2]Center for Advanced Structural Materials/ Department of Mechanical Engineering, College of Science and Engineering, City University of Hong Kong, Hong Kong, China. [3]Department of Mechanical Engineering, Hong Kong Polytechnic University, Hong Kong, China. [4]Institute of Microstructure and Properties of Advanced Materials, Beijing University of Technology, Beijing 100124, China. [5]Shenyang National Laboratory for Materials Science, Institute of Metal Research, Chinese Academy of Sciences, Shenyang 110016, China. [6]The State Key Laboratory of Powder Metallurgy, Central South University, Changsha 410083, Hunan, China.
*Corresponding author. Email: chainliu@cityu.edu.hk

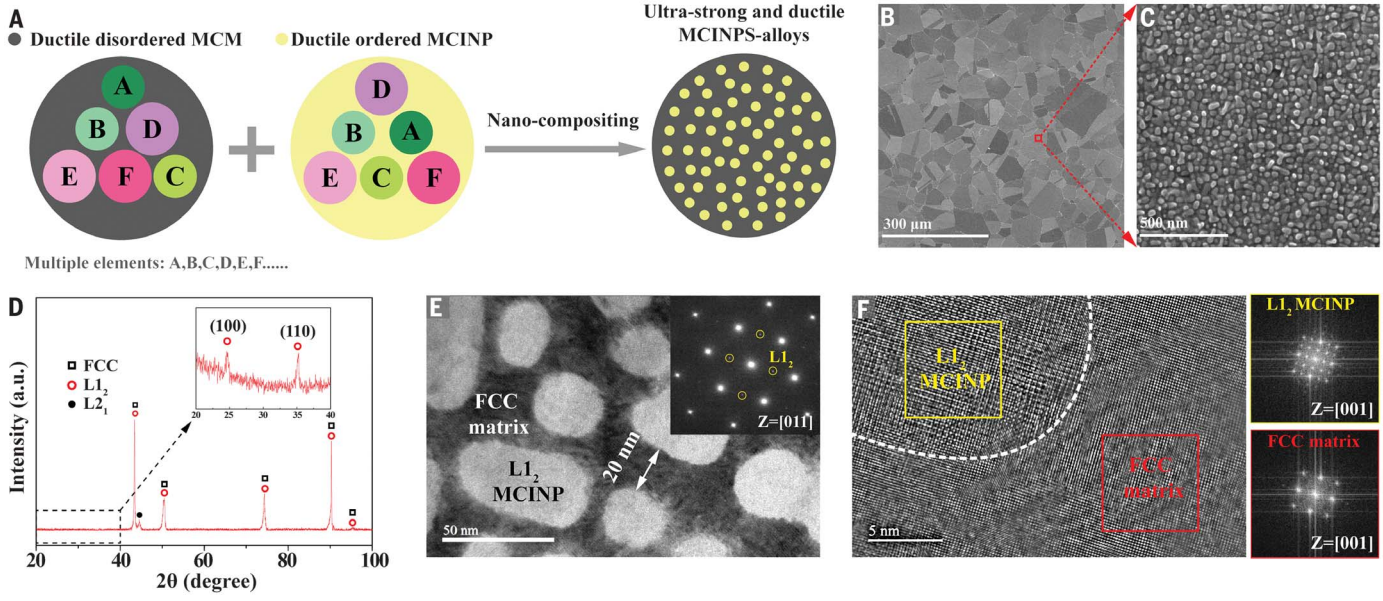


Fig. 1. Conceptual design and microstructural characterizations of the MCINPS alloys. (**A**) Schematic of the design concept of the MCINPS alloys. MCM, multicomponent matrix. (**B**) Scanning electron microscopy (SEM) image of the Al7Ti7 alloy exhibiting the typical equiaxed grain structures. (**C**) SEM image of the Al7Ti7 alloy revealing the uniform distribution of high-density $L1_2$ MCINP within the grain interior. (**D**) XRD patterns showing the phase compositions of the Al7Ti7 alloy. a.u., arbitrary units. (**E**) TEM image of the Al7Ti7 alloy showing the nanostructured morphology. The inset shows the corresponding SAED pattern. (**F**) Representative high-resolution TEM image confirming the interfacial coherency.

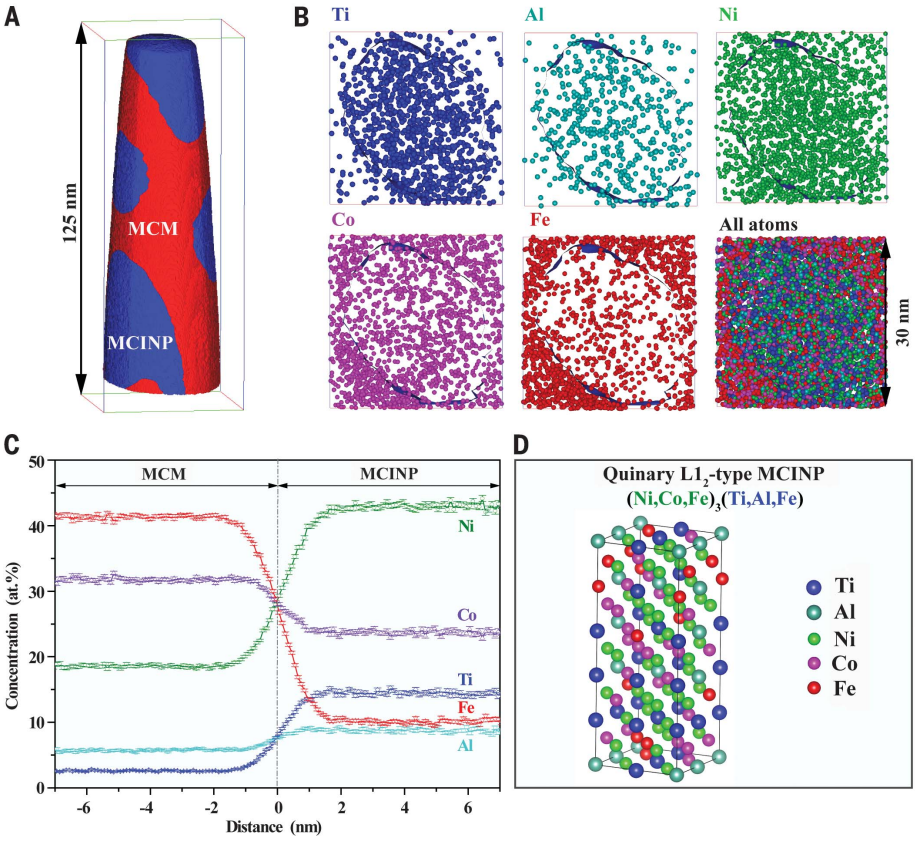


Fig. 2. Spatial morphology and multicomponent nature of the MCINPs. (**A**) 3D reconstruction map of an APT needle tip confirming the nanocomposited microstructure of the Al7Ti7 alloy. (**B**) High-resolution atom maps showing the atomistic distribution within the $L1_2$ MCINP of the Al7Ti7 alloy. (**C**) Proximity histogram across the matrix and nanoparticles revealing the multicomponent nature of the MCINPs of the Al7Ti7 alloy. (**D**) Ordering crystallographic structure and site occupancy of the $L1_2$ MCINP by density functional theory (DFT) calculations of the Al7Ti7 alloy.

ductility enhancement has not been realized yet. A typical example is the IMC hardening in face-centered cubic (fcc) high-entropy alloys (HEAs), in which the fcc-HEA matrix with a high work-hardening capability can suppress crack propagation and promote a retention of ductility only to a limited extent (*13*–*15*). Thus, when the yield strength of these alloys reaches up to 1 GPa, their tensile elongation is reduced substantially, far below what can be achieved in the freestanding fcc-type HEAs (*14*).

In this study, we developed an innovative design strategy to eliminate ductility loss in gigapascal-strength alloys. Our design concept aims to controllably create the ductile multicomponent intermetallic nanoparticles (MCINPs) for coherent strengthening in the fcc-type HEA systems (Fig. 1A), in which we achieved in situ ductilization with a nanoscale precipitation of MCINPs by controlling the order-disorder phase transformation and elemental partition. Such conceptual design enables us not only to fully exert the strengthening effect of intermetallic nanoparticles but also to maintain a high work-hardening rate and plastic deformation stability. As a result, our MCINP-strengthening (MCINPS) alloys have achieved an exceptional strength-ductility combination without encountering the common problems of early local necking and limited uniform ductility.

We designed a series of MCINPS alloys, but, because of the page limitation, we only focus on two model alloys, $(FeCoNi)_{86}$-Al_7Ti_7 (Al7Ti7) and $(FeCoNi)_{86}$-Al_8Ti_6 (Al8Ti6), to illustrate the proof of concept based on systematic thermodynamic calculations (fig. S1 and table S1). By highly alloying with Ti and Al additions, we successfully introduced high-density $L1_2$ intermetallic nanoparticles in the FeCoNi-base alloy systems. In this approach, we selected the FeCoNi for the matrix to achieve a large "FCC+$L1_2$" dual-phase region for a dense precipitation. The partial partitioning of Fe and Co atoms into the $L1_2$ phase (table S1) helped to improve the intrinsic ductility of the $L1_2$ intermetallic phase (*6*, *7*). Also, we reduced the propensity for environmental embrittlement (*6*, *7*) of the alloys by adding Ti to reduce the Al content in the $L1_2$ intermetallic phase (*16*). These bulk nanocomposited MCINPS alloys were prepared by arc melting and followed by thermomechanical treatments. The typical polycrystalline structures we identified were composed of uniform equiaxed grains (about 40~50 μm) (Fig. 1B). Within the grain interior, the near-spherical MCINPs (about 30~50 nm) were uniformly distributed in the matrix with a high volume fraction up to 50 to 55% (Fig. 1C). We characterized the fcc-$L1_2$ dual-phase nanostructures, as shown in Fig. 1D, with x-ray diffraction (XRD). We detected only a very small amount (~0.3%) of submicrometer $L2_1$-phase precipitates at grain boundaries (fig. S2). We conducted electron backscattered diffraction (EBSD), which indicated a homogeneous and fully recrystallized microstructure with a random distribution of various oriented grains in both Al7Ti7 and Al8Ti6 alloys (figs. S3 to S5). We found that the MCINPs were perfectly coherent with the matrix from transmission electron microscopy (TEM), selected-area electron diffraction (SAED), high-resolution TEM, and fast Fourier transformation (Fig. 1E). We determined a small lattice mismatch of the MCINP-matrix interface of ~0.21% for the Al7Ti7 alloy and ~0.16% for Al8Ti6 (fig. S6), dependent on the alloy composition. The small values of the lattice mismatch effectively enhance the nanoscale stabilization of the MCINPs without any heterogeneous coarsening (*8*, *17*, *18*).

We observed typical microstructural features (Fig. 2A) with a clear elemental partition between the matrix and MCINPs (Fig. 2, B and C, and table S2). The Ti and Al atoms occupied the B site of the $L1_2$ phases with a close-packed A_3B-type crystal structure. The Ni atoms occupied the A site (*18*, *19*), whereas the Fe and Co atoms could occupy both sites, depending on the chemical composition (*6*, *20*). It is possible that 2 atomic % (at %) of Fe or Co atoms can enter the B sublattice, as the sum of Ti and Al atoms is only ~23 instead of 25 at % (Fig. 2C and table S2). We performed first-principles calculations in 80-atom supercells (35 Ni, 7 Al, 11 Ti, 19 Co, and 8 Fe atoms) on the basis of the atom probe tomography (APT) compositional analysis. We constructed two $L1_2$-ordered models based on the $(Ni_{35}Co_{17}Fe_8)(Al_7Ti_{11}Co_2)$ composition with

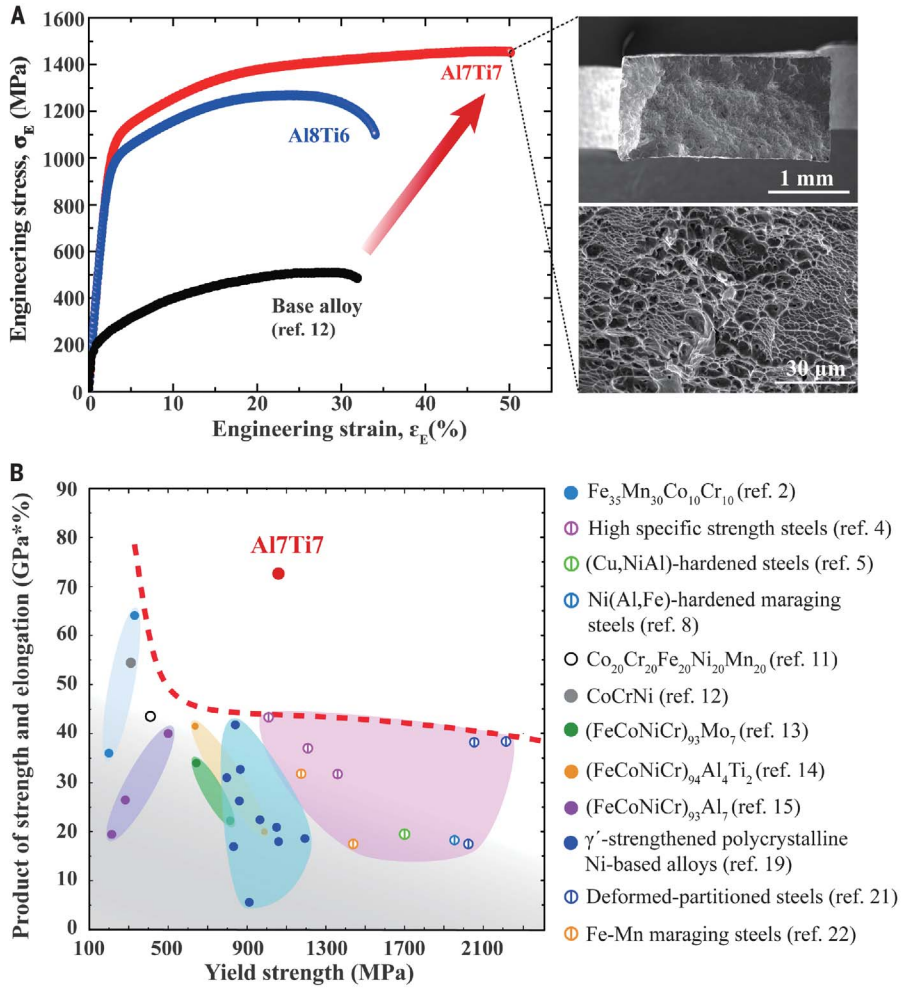


Fig. 3. Exceptional strength-ductility combination achieved in the MCINPS alloys at ambient temperature. (**A**) Engineering stress-strain curves of the MCINPS alloys compared with the FeCoNi base alloy (*12*), showing a significant increase of strength without ductility reduction. The Al7Ti7 alloy exhibits ductile dimpled structures without macroscopic necking. (**B**) Yield strength versus the product of strength and ductility of the MCINPS alloys compared with those of other high-performing materials (*2*, *4*, *5*, *8*, *11*, *13*, *14*, *19*, *21*, *22*).

2 at % Co occupying the B sublattice and the $(Ni_{35}Co_{19}Fe_6)(Al_7Ti_{11}Fe_2)$ composition with 2 at % of Fe occupying the B sublattice. The model with Fe in the B site had a lower formation energy of −0.025 eV atom^{-1}, supporting that Fe occupies on the B sublattice (fig. S7). Thus, we identified the $L1_2$-type MCINPs as the quinary $(Ni_{43.3}Co_{23.7}Fe_8)_3(Ti_{14.4}Al_{8.6}Fe_2)$ phase (Fig. 2D), which had a major impact on the mechanical response of the present alloys.

Furthermore, we measured the engineering stress-strain curves of our MCINPS alloys at ambient temperature (Fig. 3A and table S3). The main concern regarding the traditional alloy design with a high density of IMCs is severe embrittlement. Our MCINPS alloys had a tensile yield strength (σ_Y) as high as 1 GPa and an ultimate tensile strength (σ_{UTS}) of ~1.5 GPa while having a ductility of up to ~50% as the tensile elongation. The local necking in the Al7Ti7 alloy was suppressed, and the strength was five times higher than that of the single-phase FeCoNi-based alloy (*12*) (Fig. 3A). The Al7Ti7 alloy still showed a superb work-hardening ability ($\sigma_{UTS} - \sigma_Y$ = 440 MPa, σ_Y/σ_{UTS} = 0.7), even at such a high yield strength in the post-yield region. All of these ensure a large safety margin against fracture, which is vital for reliability in engineering applications. The Al7Ti7 alloy has an extremely high value of $\sigma_{UTS} \times EL_T$ (72 GPa %) when compared other high-performance alloys at room temperature (Fig. 3B) (*2*, *4*, *5*, *8*, *11*, *13*, *14*, *19*, *21*, *22*).

We attribute the pronounced increase in yield strengths to the precipitation hardening by these high-density $L1_2$-type MCINPs (*16*). The superb work-hardening capacity accounts for the large ductility (Fig. 4). Unlike traditional alloys, our MCINPS alloys exhibit the distinctive multiple-stage work-hardening behaviors responsible for different deformation stabilities. Both alloys have a gradual decrease in the work-hardening rate during the early strain up to 22% because of dislocation-controlled plastic deformation processes (*13*). Plastic instability onset generates necking in the Al8Ti6 alloy upon further straining. The Al7Ti7 alloy, by contrast, has a markedly different deformation behavior as the work hardening continues at larger strains. The instantaneous work-hardening exponent n in this stage shows a linear increase with a high value of 0.43 (Fig. 4B), which implies the onset of an additional deformation mode, enabling an extended uniform deformation to be sustained.

With the aim of deciphering the origin of the unusual work-hardening behavior of the Al7Ti7 alloy, we carefully investigated the dynamic evolution of deformation substructures at different strains with TEM (Fig. 4C). At the true strain of ~10%, the deformation was dominated by the planar slip of dislocations along the {111} primary slip planes, similar to that observed in most fcc-type alloys (*23*–*25*). As the true strain increases to ~22%, we observed well-developed HDDWs along the primary slip systems. The formation of these directional dislocation substructures (dislocation arrays and HDDWs) produced a long-range back stress that makes dislocations in the interwall space difficult to move through them, leading to an increased work-hardening response (*26*–*28*). Moreover, we observed wavy slips in the interwall space, indicating that the dislocation cross-slip is effectively activated at this stage. The intensive polyslips of dislocations are helpful for relieving the stress concentration on the {111} primary slip planes. Meanwhile, the progressive accumulation of these in-directional dislocations and their mutual interactions produced a pronounced forest dislocation hardening, i.e., the short-range effective stress hardening (*29*–*32*). We conducted tensile load-unload-reload tests to further quantify the dynamic evolutions of the work-hardening responses and associated flow stress partitioning behaviors (fig. S14). We observed a large increase in short-range effective stress, the strengthening contribution of which is almost identical to that of the long-range back stress hardening. Therefore, the high work hardening of the Al7Ti7 alloy within this stage originates from both the back stress hardening and forest dislocation hardening, which enables this alloy to maintain an unusually higher work-hardening state without local necking. By contrast, as for the true strain of ~22%, the dislocation structure of the Al8Ti6 alloy is mainly dominated by the planar-slip dislocations along the {111} primary slip planes (fig. S8). The resultant long-range back stress hardening is comparable to that of the Al7Ti7 alloy, accompanied with a lower increase in effective stress. We ascribe the inferior work-hardening response of the Al8Ti6 alloy to an insufficient effective stress hardening, leading to a relatively earlier onset of plastic instability.

With the further deformation to the true strain in the range of 28%~38% for the Al7Ti7 alloy, the main deformation characteristics are deformation-induced microbands. We detected neither mechanical twinning nor transformation-induced martensite, which are generally observed in other ductile-disordered alloys (*2*, *11*). We believe

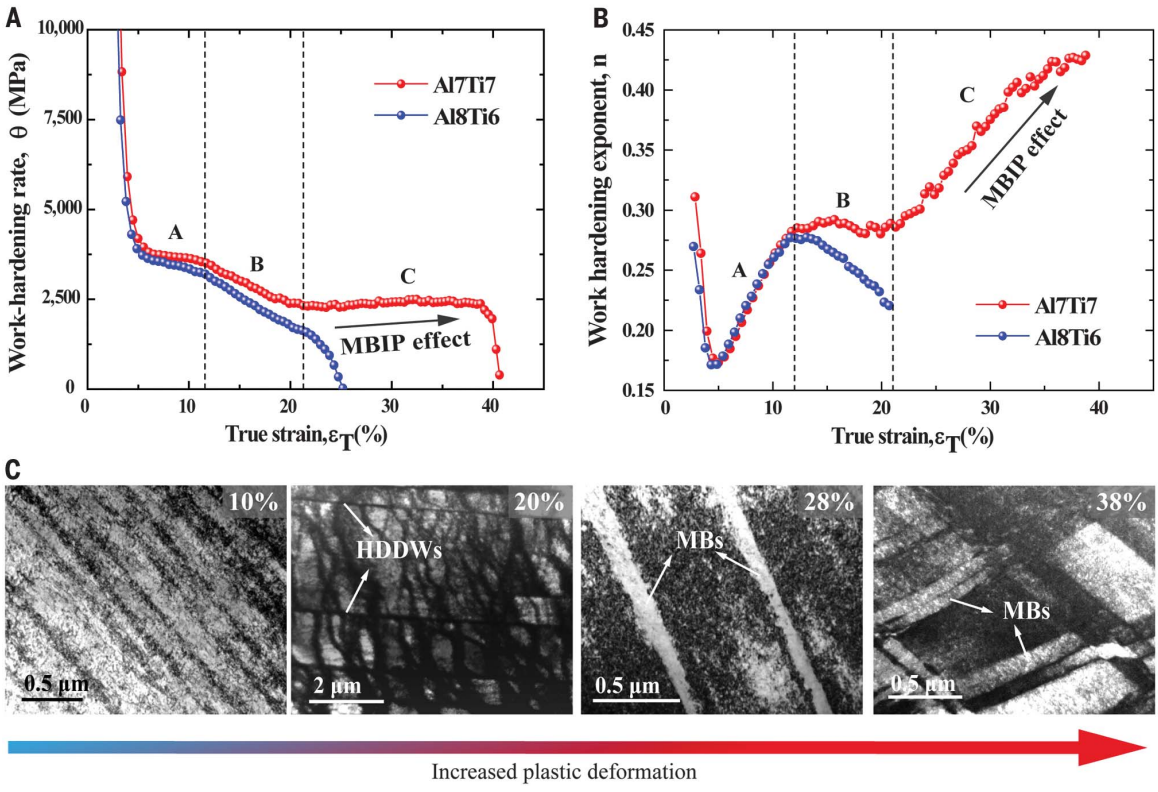


Fig. 4. Multistage work-hardening behaviors and deformation micromechanisms of the MCINPS alloys at ambient temperature. (**A**) Work-hardening rate curves of the MCINPS alloys. (**B**) Dynamic change of the work-hardening exponent (i.e., the instantaneous work-hardening exponent) within the uniform deformation process of the MCINPS alloys. (**C**) Dynamic evolution of the deformation substructures of the Al7Ti7 alloy with increasing tensile loading.

their absence is due to the medium to high stacking fault energies of the matrix alloy (*33*, *34*). Moreover, the formation of such high-density MCINPs will greatly reduce the spacing size of the matrix channel down to a much finer scale of only several nanometers, which in turn further increases the resolved stress for twin formation (*35*). Under these restrictions, the microbands act as another important deformation mode to provide an additional work-hardening source to accommodate the imposed strain, leading to an exceptional plastic stability at the high-strength level. Different from the shear-bands' softening effects observed in metallic glasses and nanotwinned alloys (*36*, *37*), the deformation-induced microbands, similar to the low-angle grain boundaries (fig. S9), are an important ductilizing mechanism to control the deformation stability of high-Mn alloys, namely the so-called microband-induced plasticity (MBIP) effect (*38*–*40*). By introducing a high-density MCINP in the Al7Ti7 alloy, we engineered the MBIP effect into the Mn-free nanostructured alloys. The formation of these microbands can generate a large increase in back stress hardening (fig. S14D), which is conducive to maintain the work-hardening capacity in the Al7Ti7 alloy for a continuous and stabilized plastic deformation. Moreover, the grain subdivision and refinement induced by these dislocation substructures further contribute to the strength enhancement because of the dynamic Hall-Petch effect (*38*, *41*). The relatively weaker microband-forming ability of the Al8Ti6 alloy might be ascribed to its lower dislocation accumulation in the early stage, in which the internal stress is insufficient to activate the cross-slip of dislocations as well as the subsequent formation of microbands (*25*, *41*–*44*). Although we observed some microbands in the necking region (fig. S15), the resulting strain hardening in Al8Ti6 comes too late to compensate for the decrease of load-carrying capability caused by the geometric softening.

The coherent strengthening by the $L1_2$-type Ni_3Al (namely, the γ′ phase) has been extensively investigated in Ni-base superalloys (*19*). However, most previous Ni_3Al phases are compositionally simple and contain a lower level of ternary elements, and the resulting superalloys show serious ductility reductions with increasing strengths (fig. S10). The multicomponent nature associated with the improved intrinsic mechanical properties of the MCINPs is the structural origin that primarily accounts for the specific mechanical properties of our MCINPS alloys. Several major contributions from the MCINPs account for the surprising ductility improvement for the alloys. First, according to the physical-metallurgy principle of ordered Ni_3Al alloys, macro-alloying with Fe and Co decreases the ordering energy and improves the intrinsic ductility (*6*, *45*, *46*), whereas a high level of the Al content leads to a serious environmental embrittlement when tested in air (*6*, *7*). For instance, an addition of 15 at % Fe improved the ductility to ~9% of normally brittle Ni_3Al (*6*). A notable example of environmental embrittlement is the decrease in ductility of boron-doped Ni_3Al from 50% to 5% in air if the Al content is increased from 24 to 25 at %. Ti doping in our alloys results in decreased Al content and largely reduces the environmental embrittlement. The higher Ti/Al ratio also promotes a substantial increase in the antiphase boundaries energy and a higher work-hardening because of the generation of dislocations cross-slip (*47*), which in turn favors the activation of microbands and produces an extended uniform deformation (*43*, *44*). The optimum level of Ti in these alloys is determined in the range of 6.5 to 7.5 at %, and the excessive Ti additions will promote the formation of more brittle $L2_1$ particles at grain boundaries, resulting in an intergranular fracture. The potential grain-boundary embrittlement can even be eliminated by further tailoring of boron additions and/or processing conditions. Second, the low lattice misfit enables the MCINP to be stabilized at the nanoscale and uniformly distributed without heterogeneous coarsening, which effectively reduces the stress-strain concentrations at local regions and suppresses the nucleation of microcracks during deformation (*8*, *18*). Moreover, from the perspective of electronic structure, small valence electron concentration values (~8.0) of the MCINP prevent the transformation from the ductile ordered cubic phase to the brittle hexagonal phase (such as η-Ni_3Ti) (*48*). For more clarity, we carefully evaluated the bulk mechanical properties of our present multicomponent $L1_2$ alloys, in which the chemical compositions are based on the APT analysis of the Al7Ti7 alloy (Fig. 2C). As expected, the present multicomponent $L1_2$ alloy is substantially stronger and more ductile than the simple Ni_3Al (*49*), exhibiting a ductility of ~40%, a yield strength of ~600 MPa, and also an excellent stain-hardening ability of about ~2.1 GPa (fig. S11). The incorporation of these strong and ductile MCINPs

allows us to not only effectively impede the dislocation motion for strengthening but also increase the damage tolerance and dislocation storage of the alloys.

In conclusion, we proposed an innovative alloy design strategy by engineering high-density MCINPs in complex alloy systems to achieve superb mechanical properties at ambient temperature. We demonstrated experimentally that the MCINPS alloys are simultaneously ultrastrong and ductile, with no strength-ductility trade-off and plastic instability. This alloy design strategy can also be feasibly applied to many other alloy systems, such as nanostructured alloys, steels, superalloys, and also HEAs, to achieve desired and enhanced properties for specific applications. The resulting new-generation complex alloys could lead to superior structural properties, which are of both great fundamental and applied importance for advanced engineering applications involving automobiles, bullet trains, cryogenic devices, and aircraft and aeronautic systems.

REFERENCES AND NOTES

1. R. O. Ritchie, *Nat. Mater.* **10**, 817–822 (2011).
2. Z. Li, K. G. Pradeep, Y. Deng, D. Raabe, C. C. Tasan, *Nature* **534**, 227–230 (2016).
3. R. Liu, Z. J. Zhang, L. L. Li, X. H. An, Z. F. Zhang, *Sci. Rep.* **5**, 9550 (2015).
4. S. H. Kim, H. Kim, N. J. Kim, *Nature* **518**, 77–79 (2015).
5. Z. B. Jiao, J. H. Luan, M. K. Miller, C. T. Liu, *Acta Mater.* **97**, 58–67 (2015).
6. N. Stoloff, *Int. Mater. Rev.* **34**, 153–184 (1989).
7. E. P. George, C. T. Liu, D. Pope, *Phys. Status Solidi* **160**, 517–529 (1997).
8. S. Jiang *et al*., *Nature* **544**, 460–464 (2017).
9. J. W. Yeh *et al*., *Adv. Eng. Mater.* **6**, 299–303 (2004).
10. O. N. Senkov, J. D. Miller, D. B. Miracle, C. Woodward, *Nat. Commun.* **6**, 6529 (2015).
11. B. Gludovatz *et al*., *Science* **345**, 1153–1158 (2014).
12. Z. Wu, H. Bei, G. M. Pharr, E. P. George, *Acta Mater.* **81**, 428–441 (2014).
13. W. H. Liu *et al*., *Acta Mater.* **116**, 332–342 (2016).
14. J. Y. He *et al*., *Acta Mater.* **102**, 187–196 (2016).
15. B. Gwalani *et al*., *Mater. Des.* **121**, 254–260 (2017).
16. Materials and methods are available as supplementary materials.
17. K. Lu, *Nat. Rev. Mater.* **1**, 16019 (2016).
18. R. C. Reed, C. M. F. Rae, in *Physical Metallurgy of the Nickel-Based Superalloys* (Elsevier, 2014), pp. 2215–2290.
19. B. Geddes, H. Leon, X. Huang, *Superalloys: alloying and performance*. (Asm International, 2010).
20. Y. I. Ustinovshikov, I. N. Shabanova, N. V. Lomova, *J. Adv. Microsc. Res.* **8**, 27–32 (2013).
21. B. B. He *et al*., *Science* **357**, 1029–1032 (2017).
22. D. Raabe, D. Ponge, O. Dmitrieva, B. Sander, *Scr. Mater.* **60**, 1141–1144 (2009).
23. B. Gludovatz *et al*., *Nat. Commun.* **7**, 10602 (2016).
24. D. Kuhlmann-Wilsdorf, *Philos. Mag. A Phys. Condens. Matter Struct. Defects Mech. Prop.* **79**, 955–1008 (1999).
25. B. Ma, C. Li, J. Zheng, Y. Song, Y. Han, *Mater. Des.* **92**, 313–321 (2016).
26. X. Hu *et al*., *Metall. Mater. Trans., A Phys. Metall. Mater. Sci.* **48**, 3943–3950 (2017).
27. X. Wu *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14501–14505 (2015).
28. M. Yang, Y. Pan, F. Yuan, Y. Zhu, X. Wu, *Mater. Res. Lett.* **4**, 145–151 (2016).
29. C. Keller, E. Hug, X. Feaugas, *Int. J. Plast.* **27**, 635–654 (2011).
30. C. Keller, M. Margulies, Z. Hadjem-Hamouche, I. Guillot, *Mater. Sci. Eng. A* **527**, 6758–6764 (2010).
31. Y. Tang *et al*., *Metals (Basel)* **8**, 444 (2018).
32. A. Vattré, B. Devincre, A. Roos, *Intermetallics* **17**, 988–994 (2009).
33. S. Zhao, G. M. Stocks, Y. Zhang, *Acta Mater.* **134**, 334–345 (2017).
34. X. Xie, G. Chen, P. McHugh, J. Tien, *Scr. Metall.* **16**, 483–488 (1982).
35. L. Murr, A. Ayala, C. Niou, *Mater. Sci. Eng. A* **216**, 69–79 (1996).
36. G. Wu, K. C. Chan, L. Zhu, L. Sun, J. Lu, *Nature* **545**, 80–83 (2017).
37. C. Hong, N. Tao, X. Huang, K. Lu, *Acta Mater.* **58**, 3103–3116 (2010).
38. J. D. Yoo, K.-T. Park, *Mater. Sci. Eng. A* **496**, 417–424 (2008).
39. W. Song, T. Ingendahl, W. Bleck, *Acta Metall. Engl. Lett.* **27**, 546–556 (2014).
40. Z. Wang *et al*., *Acta Mater.* **120**, 228–239 (2016).
41. D. Hughes, *Acta Metall. Mater.* **41**, 1421–1430 (1993).
42. I. Gutierrez-Urrutia, D. Raabe, *Scr. Mater.* **69**, 53–56 (2013).
43. P. Jackson, *Scr. Metall.* **17**, 199–202 (1983).
44. J. Huang, G. Gray III, *Acta Metall.* **37**, 3335–3347 (1989).
45. A. Chiba, S. Hanada, S. Watanabe, *Mater. Sci. Eng. A* **152**, 108–113 (1992).
46. C. Meng, J. Guo, Z. Hu, *J. Mater. Sci. Technol.* **10**, 279–284 (1994).
47. D. Raynor, J. Silcock, *Met. Sci. J.* **4**, 121–130 (1970).
48. C. T. Liu, *Int. Metals Rev.* **29**, 168–194 (1984).
49. C. T. Liu, C. L. White, *MRS Proceedings* **39**, 365 (1984).

ACKNOWLEDGMENTS

We thank L. H. Wang, Z. P. Li, and L. B. Fu (Beijing University of Technology) for technical assistance with TEM characterizations. **Funding:** The authors from City University of Hong Kong (CityU) are grateful for the financial support from the Hong Kong Government through the GRF funds with CityU grant nos. 11209314, 11205515, and C1027-14E, together with the Internal Research fund from CityU. Y.L. is grateful for the financial support from the National Natural Science Foundation of China (no. 51625404). X.D.H. is grateful for the financial support from the National Natural Science Foundation (no.11327901). **Author contributions:** T.Y. and C.T.L. designed the research; T.Y., Y.L.Z., and Y.T. prepared and characterized the samples. Y.T., Y.L.Z., and D.C. carried out TEM; Z.B.J. conducted 3D-APT. J.W. and A.H. conducted DFT calculations. J.X.C. and X.D.H. performed EBSD analysis. Y.L. conducted the high-temperature tests. T.Y., C.T.L., K.L., X.D.H., Z.B.J., Y.L., and J.J.K. analyzed the data and discussed the results. T.Y., Y.L.Z., Y.T., Z.B.J., and C.T.L. wrote the manuscript. All authors reviewed and contributed to the final manuscript. **Competing interests:** None declared. **Data and materials availability:** All data are reported in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/933/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S15
Tables S1 to S3
References (*50*–*57*)

30 December 2017; resubmitted 31 July 2018
Accepted 10 October 2018
10.1126/science.aas8815

PALEOECOLOGY

Plio-Pleistocene decline of African megaherbivores: No evidence for ancient hominin impacts

J. Tyler Faith[1,2]*, John Rowan[3,4], Andrew Du[5], Paul L. Koch[6]

It has long been proposed that pre-modern hominin impacts drove extinctions and shaped the evolutionary history of Africa's exceptionally diverse large mammal communities, but this hypothesis has yet to be rigorously tested. We analyzed eastern African herbivore communities spanning the past 7 million years—encompassing the entirety of hominin evolutionary history—to test the hypothesis that top-down impacts of tool-bearing, meat-eating hominins contributed to the demise of megaherbivores prior to the emergence of *Homo sapiens*. We document a steady, long-term decline of megaherbivores beginning ~4.6 million years ago, long before the appearance of hominin species capable of exerting top-down control of large mammal communities and predating evidence for hominin interactions with megaherbivore prey. Expansion of C_4 grasslands can account for the loss of megaherbivore diversity.

Africa is home to more species of large-bodied mammalian herbivores than anywhere else today (*1*). Because most of the world's large-bodied vertebrates became extinct toward the end of the Pleistocene (*2*), present-day African faunas serve as model systems for understanding the ecology of large mammal communities (*3*) and the impact of massive megaherbivores (>1000 kg) on ecosystems (*4*). Such knowledge feeds directly into conservation biology on a global scale by highlighting the ecological consequences of ongoing large mammal diversity loss (*1*) and illustrating the potential consequences of their rewilding (*5*).

There is a perception that Africa's exceptional large herbivore diversity is due to the continent being spared the extinctions that occurred elsewhere as modern humans (*Homo sapiens*) dispersed across the world in the past 100,000 years (*2*). This anomaly has been thought to reflect coevolution of hominin hunters and their prey (*6*) or perhaps a long history of extinctions precipitated by pre-modern hominins (*7*) in Africa. Indeed, for decades it has been suggested that hominins drove extinctions and shifts in the functional structure of large mammal communities throughout the Pleistocene (*7*–*13*). Many of these hypotheses posit that top-down control of mammal communities by tool-bearing, meat-eating hominins contributed to the demise of large-bodied herbivores (e.g., the formerly diverse Proboscidea) long before the emergence of *Homo sapiens* (*7*, *9*, *11*, *12*). Other versions of the "ancient impacts" hypothesis propose that encroachment of Early Pleistocene *Homo* into the carnivore guild led to the demise of several carnivoran lineages (*8*, *10*), perhaps leading to environmental changes through relaxed predation pressure on large-bodied herbivores (*10*). Both scenarios imply that ancient hominins played a key role in shaping African ecosystems, and by extension the environmental settings that influenced our own evolutionary history.

Despite decades of literature asserting ancient hominin impacts on African faunas, there have been few attempts to test this scenario or to explore alternatives. Here, we tested the hypothesis of top-down hominin impacts on mammal communities through the analysis of megaherbivore diversity over the past 7 million years (Ma) in eastern Africa. Our focus on this region reflects its rich and well-dated late Cenozoic fossil record, coupled with the fact that the earliest known members of the hominin clade are from eastern Africa, which therefore provides the longest well-documented history of hominin-mammal community interactions in the world. On the basis of previous hypotheses (*7*–*13*), we expect declines in megaherbivore community richness to follow or temporally coincide with hominin expansion

[1]Natural History Museum of Utah, University of Utah, Salt Lake City, UT 84108, USA. [2]Department of Anthropology, University of Utah, Salt Lake City, UT 84112, USA. [3]Institute of Human Origins and School of Human Evolution and Social Change, Arizona State University, Tempe, AZ 85282, USA. [4]Department of Anthropology, University of Massachusetts, Amherst, MA 01003, USA. [5]Department of Organismal Biology and Anatomy, University of Chicago, Chicago, IL 60637, USA. [6]Department of Earth and Planetary Sciences, University of California, Santa Cruz, CA 95064, USA.
*Corresponding author. Email: jfaith@nhmu.utah.edu

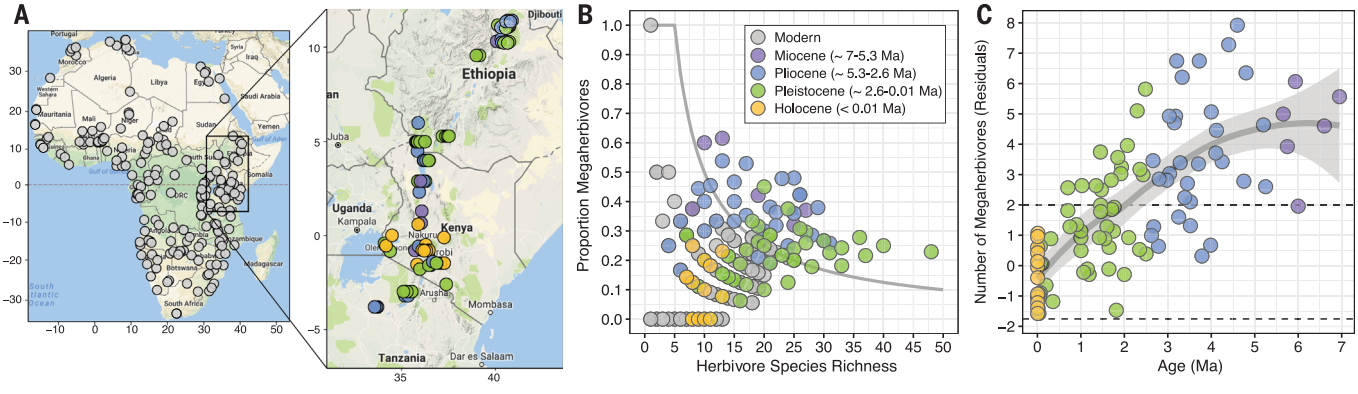


Fig. 1. Megaherbivore richness in modern and fossil communities. (**A**) Geographic distribution of the 203 modern (continental map) and 101 fossil (inset map) herbivore communities. (**B**) Relationship between the total number of herbivore species and the proportion of megaherbivore species in modern African communities and eastern African fossil assemblages. The solid line represents the maximum proportion of megaherbivores that could coexist today, based on the empirical observation of at most five sympatric megaherbivore species. Fossil assemblages falling above the line are non-analog because they include a greater proportion of megaherbivores than is observed today. (**C**) Megaherbivore richness residuals over the past 7 Ma, illustrating the long-term decline of megaherbivores starting ~4.6 Ma ago. Data points represent residuals from the least-squares regression modeling the relationship of megaherbivore richness as a function of total community richness in the modern communities (fig. S2). The solid gray line represents LOESS (locally estimated scatterplot smoothing) regression with smoothing factor = 0.75; 95% confidence limits are shown in light gray. Horizontal dashed lines encompass the middle 95% range of variation in the modern communities.

into carnivore niche space. Specifically, proponents of the ancient impacts hypothesis typically place the onset of anthropogenic diversity decline between 2 and 1 Ma ago (*7*, *8*, *10*–*12*). This encompasses the earliest evidence for systematic hominin predation upon large-bodied mammals (~2 Ma ago) (*14*) and megaherbivores (~1.95 Ma ago) (*15*), as well as the appearance of *Homo erectus* (~1.9 Ma ago), the first hominin species whose paleobiology is similar to later representatives of our genus and that consumed large amounts of animal tissue (*16*). These evolutionary changes are thought to account for an unprecedented reduction of megaherbivore diversity coupled with collapse of the large carnivore guild (*7*, *8*, *10*–*12*).

We used present-day and fossil herbivore community data to quantify long-term changes in the richness of eastern African megaherbivores (*17*). A dataset of more than 200 modern communities from protected areas across Africa allowed us to establish a baseline for present-day variability in megaherbivore community richness (Fig. 1A, table S1, and data S1). In addition, we compiled a fossil dataset that includes the presence of herbivore taxa in 101 eastern African fossil assemblages spanning the past ~7 Ma (table S2 and data S2), a period that encompasses the earliest probable and definitive hominin species in eastern Africa. Our focus on individual fossil assemblages within a single region provides a pertinent spatiotemporal scale for our research question because it allows a more direct assessment of hominin impacts on ancient herbivore communities than is possible from analyses of species occurrences at continental scales [e.g., (*9*)]. Because hominin impacts need not be the only driver of change in the eastern African megaherbivore community, we also examined trends in megaherbivore community richness relative to independent records of climatic and environmental change. These include global atmospheric partial pressure of CO_2 (pCO_2) (*18*), the percentage of C_4 biomass (e.g., tropical grasses) inferred from the stable carbon isotope ($\delta^{13}C$) composition of soil carbonates in eastern Africa (*19*), estimates of paleo-aridity derived from the stable oxygen isotopes ($\delta^{18}O$) of eastern African fossil herbivores (*20*), and the percentage of C_4 grazers among ungulate taxa in eastern African fossil assemblages (*20*). The eastern African proxy data come from many of the same sites examined in our analysis of megaherbivore diversity.

Our compilation of the eastern African fossil record reveals substantial megaherbivore extinctions through time. Over the past 7 Ma, 28 megaherbivore lineages became extinct (table S3), leading to present-day communities that are depauperate in megaherbivores. For example, modern African herbivore communities include only up to five sympatric megaherbivores (data S1), including all of the extant species (*Giraffa camelopardalis*, *Hippopotamus amphibius*, *Ceratotherium simum*, *Diceros bicornis*, and *Loxodonta africana*). In contrast, the fossil record documents paleocommunities that were considerably richer in megaherbivores, with some assemblages documenting the co-occurrence of as many as 10 megaherbivore species (data S2). Although time-averaging of fossil assemblages can produce associations of species that never spatiotemporally co-occurred, it cannot account for the exceptional richness of megaherbivores here (*17*) (fig. S1). Controlling for community richness, a variable that increases as a function of time-averaging and sampling effort, many fossil assemblages over the past ~7 Ma fall outside the modern range of variation because they include both a greater proportion and a greater absolute number of megaherbivore species (Fig. 1, B and C).

To illustrate temporal trends (Fig. 1C), we calculated residuals for fossil assemblages as deviations from expected megaherbivore richness

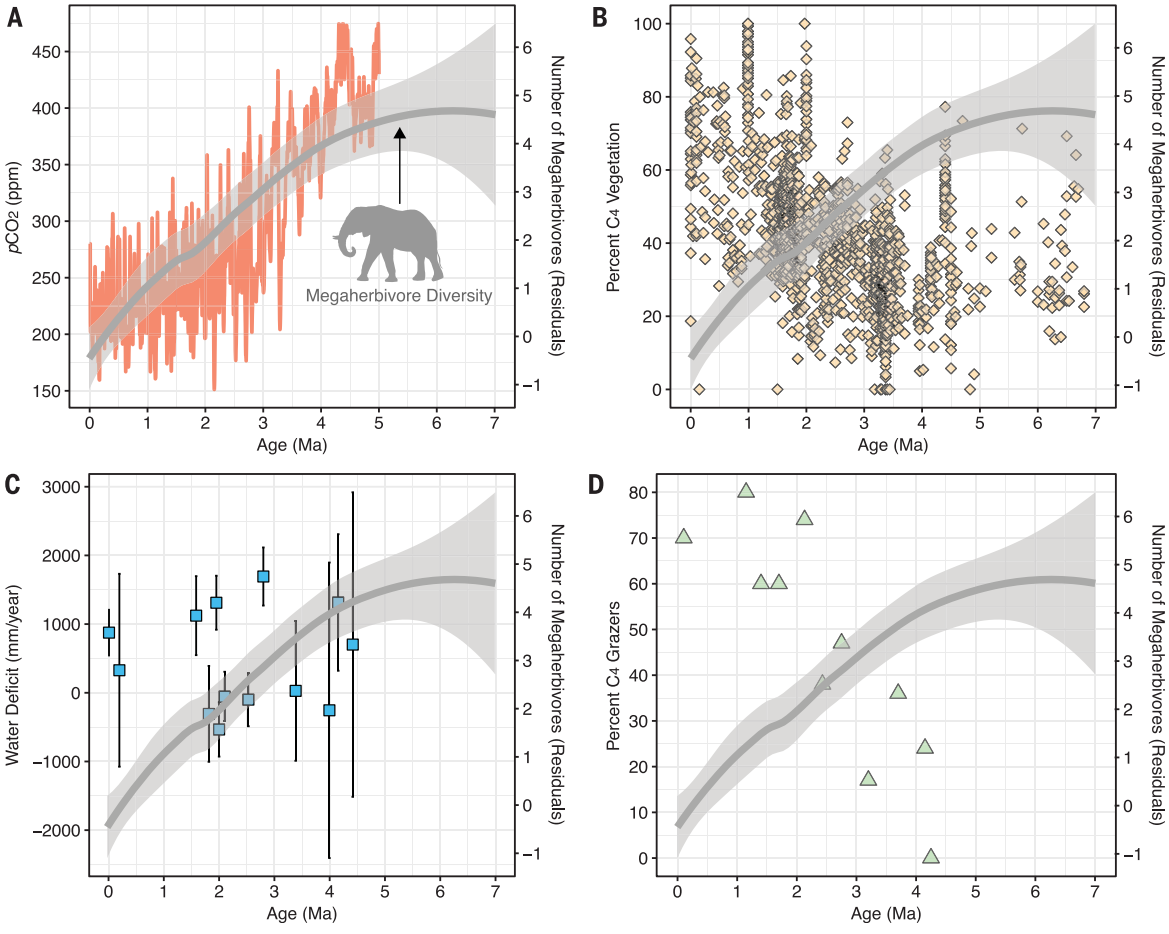


Fig. 2. The decline of megaherbivore richness relative to climatic and environmental proxies. (**A**) Global pCO_2. (**B**) Percentage of C_4 vegetation inferred from $\delta^{13}C$ of eastern African paleosol carbonates. (**C**) Estimates of water deficit (aridity) for eastern African fossil sites based on $\delta^{18}O$ of herbivore tooth enamel. Error bars represent SE of the mean water deficit estimates. (**D**) Percentage of C_4 grazers among herbivore taxa (Artiodactyla-Perissodactyla-Proboscidea) from eastern African fossil sites, calculated as the percentage of taxa with mean enamel $\delta^{13}C$ values > –1 per mil. The gray line in all panels represents the LOESS regression (95% confidence limits in light gray) for megaherbivore richness residuals, as in Fig. 1C.

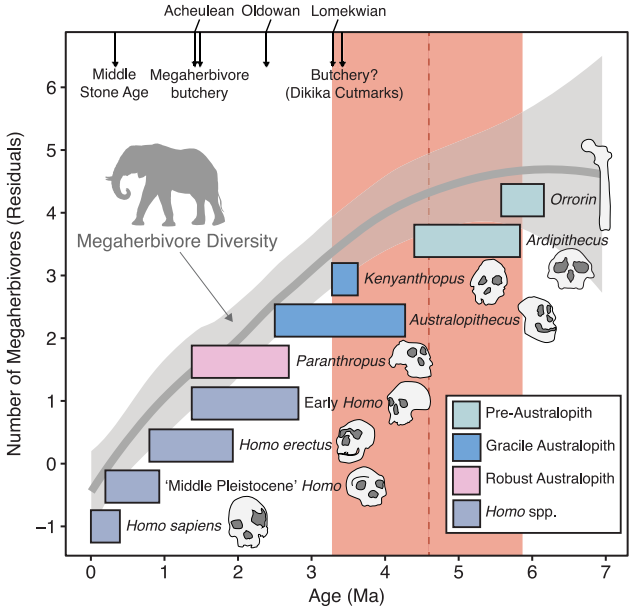


Fig. 3. The decline of megaherbivore richness relative to milestones in hominin evolution. These include the appearance of novel technologies and behaviors, as well as the observed temporal ranges of eastern African hominin taxa. The vertical dashed line indicates the breakpoint denoting onset of megaherbivore decline (4.6 Ma ago), with red shading representing the 95% confidence interval (3.3 to 5.9 Ma ago).

modeled as a linear function of herbivore community richness based on our modern community data (fig. S2). Residual analysis highlights the exceptional richness of fossil megaherbivore communities—many are well outside the modern range of variation—and documents a steady and long-term decline in richness beginning in the early Pliocene, with fossil assemblages consistently falling within the modern range of variation only after ~0.7 Ma ago (Fig. 1C). Breakpoint analysis places the onset of the megaherbivore decline at ~4.6 Ma ago (95% confidence interval, 3.3 to 5.9 Ma ago). After the ~4.6 Ma breakpoint, the rate of diversity decline through time does not change following the appearance of *Homo erectus* (1.9 Ma ago) or at either end of the hypothesized window of accelerated anthropogenic diversity decline, 2 to 1 Ma ago (table S4). Instead, the long-term decline in megaherbivore richness closely tracks global variation in atmospheric pCO_2 as well as the expansion of C_4 grasslands and C_4 grazers in eastern Africa, although there is no association with paleo-aridity (Fig. 2).

The loss of large-bodied mammals is thought to be a hallmark of anthropogenic extinctions (*6*, *7*, *9*), in part because large-bodied species are likely to have been preferentially targeted by hominin hunters and because their slower life history profiles (e.g., delayed reproductive maturity, prolonged gestation, low population growth rates) render them more susceptible to extinction. Our analyses show that the diversity of megaherbivores in eastern African fossil assemblages has undergone a steady decline since ~4.6 Ma ago (Fig. 1), beginning long before the proposed timing of anthropogenic impacts (2 to 1 Ma ago) linked to encroachment of *Homo erectus* into the carnivore guild (Fig. 3). The antiquity of the megaherbivore decline effectively eliminates top-down hominin impacts as a plausible mechanism for setting it in motion, as it would have had to involve small-bodied australopiths or pre-australopiths (e.g., *Australopithecus*, *Ardipithecus*), which were functionally equivalent to bipedal apes. Like extant chimpanzees, these hominin taxa may have preyed upon vertebrate species smaller than themselves, but they almost certainly did not hunt megaherbivore prey (*21*). Although we cannot rule out the possibility that the subsequent appearance of more derived hominin species, new stone-tool technologies, and increased carnivory may have incidentally contributed to the demise of megaherbivores, the steady decline of megaherbivores beginning ~4.6 Ma ago (Fig. 3 and table S4)—in contrast to proposed extinction pulses linked to major shifts in hominin evolution (*7*)—implies that the primary driver was decoupled from hominin evolution.

We propose that the expansion of C_4 grasslands (Fig. 2B) played a critical role in megaherbivore decline. Analysis of our fossil dataset indicates that the long-term decline of megaherbivores is primarily due to the loss of megaherbivore browsers and mixed feeders that consumed C_3 plants, including trees, shrubs, and herbs (fig. S3). Unlike megaherbivore grazers, temporal declines in the richness of megaherbivore browsers and mixed feeders are in close agreement with the overall megaherbivore diversity loss (fig. S3). Loss of C_3 consumers is also evident in a compilation of $\delta^{13}C$ data from tooth enamel of eastern African megaherbivores (fig. S3 and data S3). Their decline in fossil assemblages is likely related to the Plio-Pleistocene expansion of C_4 grasslands (Fig. 2B), which reduced the availability of C_3 forage on the landscape. Because larger-bodied species are more strongly limited by forage availability than smaller-bodied species (*4*), a decline in megaherbivore species dependent on C_3 forage is expected. At the same time, decreasing atmospheric CO_2 concentrations (Fig. 2A) would have diminished the advantage of massive body size, which allows megaherbivores to consume lower-quality foods than their smaller-bodied counterparts (*4*). Especially among C_3 plant species, plant tissue grown at lower CO_2 concentrations provides higher-quality forage because such plants have higher N content (lower C:N ratios) and fewer secondary compounds (*22*). This would allow smaller-bodied species—which are restricted to high-quality forage (*3*)—to exploit a greater range of increasingly rare C_3 resources, with the outcome being less food available to megaherbivore browsers and mixed feeders. Some or all of these mechanisms are likely to have played a role in driving large herbivore extinctions elsewhere. For example, although the timing of C_4 expansion and the nature of its ecological consequences varied globally as a result of differences in climate and historical contingencies (*23*), the C_4 expansion in the Siwalik Group (Pakistan) from ~8.5 to 6.0 Ma ago is associated with considerable losses among C_3 consumers and the last appearances of many massive herbivores, including five rhino species, a proboscidean, a chalicothere, and a sivathere (*24*).

There is a long history of debate concerning the mechanisms responsible for the expansion of C_4 grasslands in eastern Africa (*19*). This phenomenon is often linked to increased aridity after the onset of Northern Hemisphere Glaciation, ~2.8 Ma ago (*25*). However, dust flux records from marine sedimentary archives that reflect eastern African climate indicate substantial long-term increases in aridity only after ~1.5 Ma ago (*26*), long after the onset of C_4 expansion (Fig. 2B). In addition, recent analyses show that terrestrial proxies for aridity and vegetation in eastern Africa vary independently through the Plio-Pleistocene, implying that other abiotic or biotic mechanisms likely underpin habitat change (*20*). Because C_4 grasses are favored at lower CO_2 concentrations (*27*), the Plio-Pleistocene CO_2 decline (Fig. 2A) is likely an important abiotic mechanism (*19*). For example, at high temperatures, such as those inferred from Turkana Basin paleosol carbonates (>30°C) (*28*), C_4 grasses are expected to expand when atmospheric CO_2 concentrations fall below ~450 parts per million (*27*). The expansion of C_4 grasslands and associated decline of megaherbivores are consistent with long-term decline of CO_2 (Fig. 2), likely facilitated by episodes of aridity (i.e., positive water deficit) that occurred across the interval (Fig. 2C). The persistent expansion of C_4 grasslands is remarkable in

light of the decline of megaherbivore diversity. Megaherbivore browsers and mixed feeders, especially elephants (*L. africana*) and black rhinoceros (*D. bicornis*), promote grassland expansion through consumption of woody (C_3) vegetation, toppling and ringbarking of trees, and synergistic effects with fire (*4*, *29*). With all other factors held constant, we would expect the considerable decline of such taxa—which implies a reduction in megaherbivore biomass (fig. S4)—to drive an expansion of woody cover through the Plio-Pleistocene. That the exact opposite occurred (Fig. 2B) suggests a dominance of abiotic mechanisms in driving the C_4 expansion.

We note that the CO_2-driven expansion of C_4 grasslands and the associated extinction of megaherbivores can account for other changes in eastern African mammal communities that have been attributed to ancient hominin impacts. As noted earlier, the loss of richness and functional diversity among eastern African carnivorans through the Pleistocene has been attributed to the encroachment of *Homo* into the large carnivore guild (*8*, *10*–*12*). However, some extinct Pleistocene carnivorans lacking modern analogs (e.g., sabertooth felids) are known to have specialized on juvenile megaherbivores (*30*). Given the close association between large carnivore diversity and herbivore diversity (*31*), the loss of megaherbivore prey likely contributed directly to the extinction of attendant carnivores. Thus, in the absence of detailed consideration of eastern African herbivores, it is premature to invoke hominin impacts as an important driver of carnivoran diversity loss and associated ecosystem change. Together with our observations indicating that bottom-up processes can account for the loss of megaherbivores, it follows that in the search for anthropogenic impacts on ancient African ecosystems, we must focus our attention on the one species known to be capable of causing them: *Homo sapiens* over the past 300,000 years.

REFERENCES AND NOTES

1. W. J. Ripple *et al.*, *Sci. Adv.* **1**, e1400103 (2015).
2. P. L. Koch, A. D. Barnosky, *Annu. Rev. Ecol. Evol. Syst.* **37**, 215–250 (2006).
3. R. H. V. Bell, *Sci. Am.* **225**, 86–93 (July 1971).
4. R. N. Owen-Smith, *Megaherbivores: The Influence of Very Large Body Size on Ecology* (Cambridge Univ. Press, 1988).
5. J.-C. Svenning *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 898–906 (2016).
6. P. S. Martin, in *Quaternary Extinctions: A Prehistoric Revolution*, P. S. Martin, R. G. Klein, Eds. (Univ. of Arizona Press, 1984), pp. 354–403.
7. S. K. Lyons, F. A. Smith, J. H. Brown, *Evol. Ecol. Res.* **6**, 339–358 (2004).
8. L. Werdelin, M. E. Lewis, *PLOS ONE* **8**, e57944 (2013).
9. F. A. Smith, R. E. Elliott Smith, S. K. Lyons, J. L. Payne, *Science* **360**, 310–313 (2018).
10. M. Fortelius *et al.*, *Philos. Trans. R. Soc. B* **371**, 20150232 (2016).
11. C. N. Johnson *et al.*, *Science* **356**, 270–275 (2017).
12. Y. Malhi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 838–846 (2016).
13. P. S. Martin, *Nature* **212**, 339–342 (1966).
14. J. V. Ferraro *et al.*, *PLOS ONE* **8**, e62174 (2013).
15. D. R. Braun *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10002–10007 (2010).
16. S. C. Antón, R. Potts, L. C. Aiello, *Science* **345**, 1236828 (2014).
17. See supplementary materials.
18. L. B. Stap *et al.*, *Earth Planet. Sci. Lett.* **439**, 1–10 (2016).
19. N. E. Levin, *Annu. Rev. Earth Planet. Sci.* **43**, 405–429 (2015).
20. S. A. Blumenthal *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7331–7336 (2017).
21. C. Stanford, *The Hunting Apes: Meat Eating and the Origins of Human Behavior* (Princeton Univ. Press, 1999).
22. C. E. Owensby, R. C. Cochran, L. M. Auen, in *Carbon Dioxide, Populations, and Communities*, C. Körner, F. Bazzaz, Eds. (Academic Press, 1996), pp. 363–371.
23. C. Strömberg, *Annu. Rev. Earth Planet. Sci.* **39**, 517–544 (2011).
24. C. Badgley *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12145–12149 (2008).
25. P. B. deMenocal, *Earth Planet. Sci. Lett.* **220**, 3–24 (2004).
26. M. H. Trauth, J. C. Larrasoaña, M. Mudelsee, *Quat. Sci. Rev.* **28**, 399–411 (2009).
27. J. R. Ehleringer, T. E. Cerling, B. R. Helliker, *Oecologia* **112**, 285–299 (1997).
28. B. H. Passey, N. E. Levin, T. E. Cerling, F. H. Brown, J. M. Eiler, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11245–11249 (2010).
29. R. M. Holdo, R. D. Holt, J. M. Fryxell, *Ecol. Appl.* **19**, 95–109 (2009).
30. C. W. Marean, C. L. Ehrhardt, *J. Hum. Evol.* **29**, 515–547 (1995).
31. C. Sandom *et al.*, *Ecology* **94**, 1112–1122 (2013).

ACKNOWLEDGMENTS

We thank J. O'Connell and members of the University of Utah Archaeological Center for helpful discussions and comments on previous drafts of this manuscript, S. Blumenthal for providing data used in Fig. 2C, and G. Hempson for providing data used in fig. S4. **Funding:** J.T.F.'s early contributions to this project were supported by an Australian Research Council DECRA Fellowship (DE160100030). **Author contributions:** J.T.F., J.R., and A.D. conceived the project; J.R. compiled the modern and fossil data; J.T.F., J.R., and A.D. analyzed the data; J.T.F., J.R., A.D., and P.L.K. interpreted the data; J.T.F. wrote the first draft and all co-authors provided feedback. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** Data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/938/suppl/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S4
Data S1 to S3
References (*32*–*149*)

23 May 2018; accepted 28 September 2018
10.1126/science.aau2728

EPIDEMIOLOGY

Social network plasticity decreases disease transmission in a eusocial insect

Nathalie Stroeymeyt[1]*†, Anna V. Grasse[2], Alessandro Crespi[3], Danielle P. Mersch[1]‡, Sylvia Cremer[2]†, Laurent Keller[1]†

Animal social networks are shaped by multiple selection pressures, including the need to ensure efficient communication and functioning while simultaneously limiting disease transmission. Social animals could potentially further reduce epidemic risk by altering their social networks in the presence of pathogens, yet there is currently no evidence for such pathogen-triggered responses. We tested this hypothesis experimentally in the ant *Lasius niger* using a combination of automated tracking, controlled pathogen exposure, transmission quantification, and temporally explicit simulations. Pathogen exposure induced behavioral changes in both exposed ants and their nestmates, which helped contain the disease by reinforcing key transmission-inhibitory properties of the colony's contact network. This suggests that social network plasticity in response to pathogens is an effective strategy for mitigating the effects of disease in social groups.

Social insects are an ideal system to study the potential role of social network plasticity in disease defense. The networks of social and physical interactions of insect, vertebrate, and human societies share many properties that are known to influence disease spread (*1*–*11*). In addition, because the high genetic similarity and frequent physical contacts between individuals heighten the risk of infection transmission within colonies, social insects have evolved a variety of collective mechanisms to prevent infection and limit the spread of pathogens through the colony [social immunity, (*12*, *13*)]. In particular, the organization of the colony into distinct age-and-task groups has been predicted to confer a double disease-related benefit by inhibiting global transmission dynamics and by disproportionately protecting high-value individuals (i.e., the queen and young workers) from pathogen exposure (*12*–*16*). We first tested these predictions in the absence of disease (constitutive organizational immunity) by using an automated ant tracking system to detect all physical contacts in 22 *Lasius niger* colonies and quantify transmission-relevant properties of their social networks (Fig. 1, A and B, and fig. S1). We focused on global network properties known to either inhibit or facilitate disease spread (see Table 1) and compared them to those of null model networks, obtained by randomizing the identity of interacting ants while preserving the temporal properties of the contact sequence and the total number of contacts for each ant. The observed networks displayed transmission-inhibiting rather than transmission-enhancing properties: They had higher modularity, lower density, larger diameter, and lower mean and maximum degree centrality than the random networks (fig. S2 and table S1). To further test whether the ant network's overall topology does inhibit disease transmission, we developed a stochastic epidemiological model simulating the spread of a pathogen through the colony and parameterized it using experimental data on the transmission of conidiospores (hereafter "spores") of the fungus *Metarhizium brunneum*, a natural pathogen of *L. niger* ants (*12*, *17*, *18*). Our simulation model had high predictive power and outperformed simpler models in predicting the intensity and biological consequences of transmission across individuals (Fig. 2A and table S2). Simulations confirmed that, relative to the random networks, the topology of the observed networks results in slower pathogen transmission, smaller amounts of pathogen transferred to nestmates, and a more-pronounced right skew in the distribution of final pathogen loads, which corresponds to a greater proportion of host individuals receiving a low, presumably harmless load (*19*) and a lower proportion of individuals receiving a high, potentially disease-inducing load (fig. S3 and table S1).

In addition to the overall network topology, the relative positions of individuals in the network also contributed to decrease disease impact. In social insects, new infections are more likely to be picked up by foragers outside the nest than by indoor workers (hereafter referred to as "nurses") (*12*, *13*), because the nest is subject to effective preventative hygienic treatments (*20*). Importantly, there was a strong negative relationship between degree centrality and time spent outside in the observed networks [general linear mixed model (GLMM) with type III Wald chi-square test: $\chi^2 = 402.9$, df = 1, $P < 0.0001$; fig. S4 and table S3], indicating that the ants that have a higher chance of encountering pathogens (i.e., frequent foragers) also have fewer connections within the network. This should confer important disease-related benefits because epidemic risk is reduced when the disease originates from nodes with few connections (*4*, *10*). In agreement with this prediction, simulations indicate that pathogens spread significantly more slowly and less broadly when they originate from frequent foragers than when they originate from occasional foragers, nurses, or random workers (fig. S5), showing that the ant social network provides strong colony-wide protection against the most likely source of disease. Moreover, analysis of the social organization of the 22 colonies in our main experiment and an additional 11 age-marked colonies ("age experiment") showed that *L. niger* colonies display marked segregation between potential disease sources (foragers) on one hand and high-value individuals (young workers acting as nurses and the queen) on the other (fig. S6). Simulations indicated that this organization disproportionately protects high-value individuals from the most likely source of disease (i.e., pathogens carried back to the nest by frequent foragers; fig. S7).

After describing the constitutive transmission-inhibitory characteristics of the ant social network, we tested whether colonies change their patterns of social contacts in an adaptive way when confronted with disease [induced organizational immunity (*14*, *21*)]. To do so, we randomly selected 10% of the colony's workers among foragers and exposed them either to a control solution (sham-treated colonies; $n = 11$) or to a suspension of *M. brunneum* spores (pathogen-exposed colonies; $n = 11$) and then returned them to the foraging arena. After a recovery period of 2 hours, the colony's behavior was tracked for an additional 24 hours, at which time we determined the spore load of each individual from pathogen-exposed colonies using quantitative, real-time PCR (qPCR; fig. S8). *M. brunneum* spores are transferred by physical contacts between ants, which can lead to the transmission of disease to healthy nestmates regardless of the future fate of the originally exposed ants (*19*). The spores typically penetrate the host's body to cause infection from 1 to 2 days after initial exposure (*22*), that is, after the end of the experiment. As a consequence, any behavioral changes induced by pathogen exposure should reflect an active response by the host rather than a manipulation by the pathogen or a side effect of disease because no infection could take place during the experiment. Additionally, because the pathogen did not start replicating during the experiment, the measured spore loads should closely reflect the contamination level of each ant depending on its position within the colony's contact

[1]Department of Ecology and Evolution, University of Lausanne, CH-1015 Lausanne, Switzerland. [2]Institute of Science and Technology Austria (IST Austria), Am Campus 1, A-3400 Klosterneuburg, Austria. [3]Biorobotics Laboratory, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.
*Present address: Laboratory of Intelligent Systems, Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland.
†Corresponding author. Email: laurent.keller@unil.ch (L.K.); nathalie.stroeymeyt@normalesup.org (N.S.); sylvia.cremer@ist.ac.at (S.C.) ‡Present address: Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel.

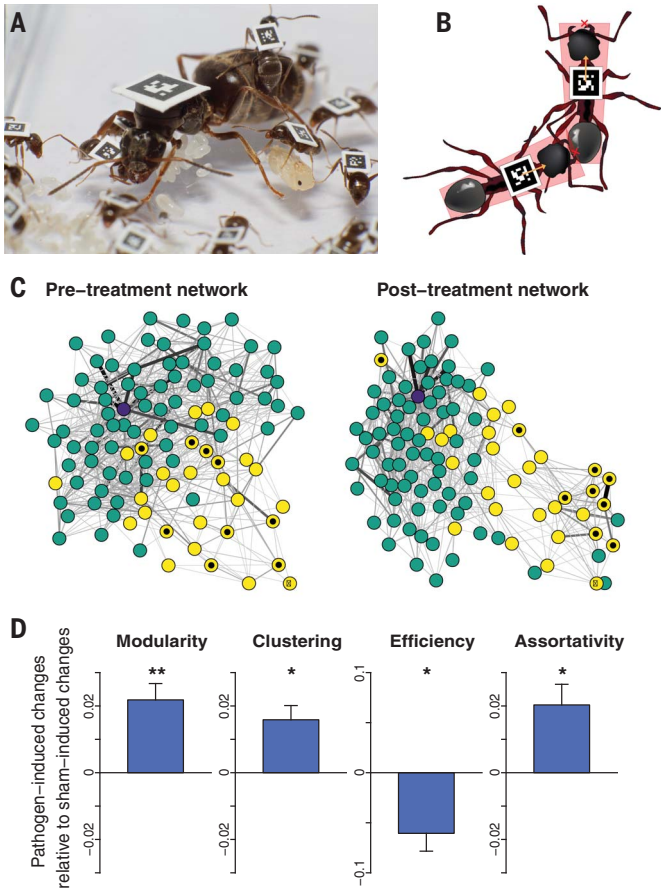


Fig. 1. Pathogen-induced changes in social networks. (**A**) Individually tagged *L. niger* queen (1.6-mm tag) and workers (0.7-mm tags). [Photo credit: T. Brütsch, University of Lausanne] (**B**) Geometrical approach used to infer physical contacts. Each ant is overlaid with a trapezoid (red), and a contact is recorded if the front end of the trapezoid of one ant overlaps with the trapezoid of the other ant. Red X's indicate the front end of the ants' trapezoids; orange arrows indicate the orientation of the ants' bodies. (**C**) Pre- and posttreatment networks from a pathogen-exposed colony. Each node represents an ant (queen, purple; nurses, green; foragers, yellow). Weighted edges represent the cumulated duration of contact between each pair of ants. Experimentally treated foragers are indicated with a black dot. (**D**) Pathogen-induced changes in network properties. For visualization of the results, bars and whiskers show the means and standard errors of the difference between post- and pretreatment values in pathogen-exposed colonies after subtraction of the mean difference in sham-treated colonies. Statistical analyses were performed on all raw data. *$P < 0.05$; **$P < 0.01$ (GLMM, interaction$_{period \times treatment}$). Sample sizes: 11 pathogen-exposed and 11 sham-treated colonies.

Table 1. Global network properties influencing disease transmission. Whenever relevant, network properties were calculated using weighted network edges. + indicates enhancement; − indicates inhibition.

Property	Definition	Effect on transmission
Modularity (*8*–*11*)	Degree of compartmentalization of the network into distinct communities of well-connected nodes	−
Clustering (*10*, *11*)	Tendency of neighboring nodes to form fully connected sets	−
Density (*1*)	Proportion of existing edges among all possible edges	+
Diameter (*6*)	Maximum length of the shortest paths between all pairs of nodes	−
Network efficiency (*28*)	Mean connection efficiency (i.e., inverse of the shortest path length) across all pairs of nodes	+
Degree centrality (*4*, *10*)	Number of edges connecting a node to other network nodes	+
Assortativity	Degree of preferential association between nodes of similar properties	Node-specific

network. Accordingly, we found a positive association between qPCR-measured pathogen transmission and network density, efficiency, and mean degree centrality on one hand and a negative association between measured pathogen transmission and network diameter, modularity, and clustering on the other (fig. S9), in agreement with theoretical predictions (table S1). Furthermore, qPCR results showed that untreated foragers received significantly higher amounts of spores from the experimentally treated foragers than either nurses or the queen [GLMM, $|z\ score| \geq 4.56$, $P < 0.0001$ in post hoc comparisons with Benjamini-Hochberg (BH) correction; Fig. 2B and table S3] and that the pathogen load received was significantly negatively correlated with network distance to treated ants (fig. S10). An additional experiment in which we exposed another 11 colonies to *M. brunneum* and allowed the disease to develop for 9 days ("survival experiment") further showed that pathogen-induced mortality was higher among untreated foragers than among nurses and that all queens were still alive at the end of the experiment (fig. S11). These results altogether confirm that the queen and nurses are effectively protected from disease carried back to the nest by foragers.

Comparisons of the network's structure before and after treatment revealed that pathogen exposure induced changes that should both reduce overall disease spread and reinforce the protection of high-value individuals. Indeed, relative to the sham treatment, pathogen exposure induced a strengthening of three of the network's transmission-inhibiting properties (increased modularity and clustering and decreased transmission efficiency), an increase in social segregation between task groups (increased task assortativity and increased network distances between the queen and workers), and a decrease in the spreading influence (degree centrality) of pathogen-exposed foragers [GLMM, global network properties, interaction between period and treatment (interaction$_{period \times treatment}$, where period is pre- versus posttreatment and treatment is sham versus pathogen-exposed): $\chi^2 \geq 4.25$, df = 1, $P \leq 0.039$ in all tests; individual node properties, post hoc comparisons with BH correction: $|z\ score| \geq 5.10$, $P < 0.0001$; Fig. 1, C and D, fig. S12, and table S3]. Simulations confirmed that these pathogen-induced changes in the social network protected the colony against disease by significantly reducing the probability of contamination and the number of spores received by untreated ants (GLMM, treatment-induced changes in sham-treated versus pathogen-exposed colonies: $\chi^2 \geq 6.26$, df = 1, $P \leq 0.012$; fig. S13 and table S3), which is consistent with our empirical findings (fig. S9). Simulations also confirmed that pathogen-induced network changes reinforced the protection of high-value individuals by significantly decreasing the probability of the queen and nurses receiving a high load, while simultaneously increasing their probability of receiving a low load (GLMM, queens: $\chi^2 \geq 5.08$, $P < 0.024$;

nurses: |z score| ≥ 8.38, $P < 0.0001$; foragers: |z score| ≤ 0.58, $P \geq 0.56$ in post hoc comparisons with BH correction; Fig. 3, A and B, and table S3). These findings are important because the consequences of contamination with *M. brunneum* have been shown to be highly dose-dependent in ants and termites: Whereas high-level contamination can lead to disease-induced death, low-level contamination is beneficial because it does not induce mortality and allows individuals to develop active immunization that protects them against future challenges with the same pathogen (*19*, *23*, *24*). Interestingly, converting the threshold used to distinguish high from low simulated dose (Fig. 3A) to a real exposure dose (Fig. 2A) revealed that it is in the same order of magnitude as a dose causing 2% mortality (lethal dose 2%, LD2) (table S4), which was identified as a dose inducing immunization in a related species (*19*). Thus, pathogen-induced network changes should confer a dual advantage for the colony because they simultaneously decrease the probability that the queen and nurses receive a potentially mortal load and increase the probability that they receive a harmless load leading to beneficial immunization. To test whether the high versus low

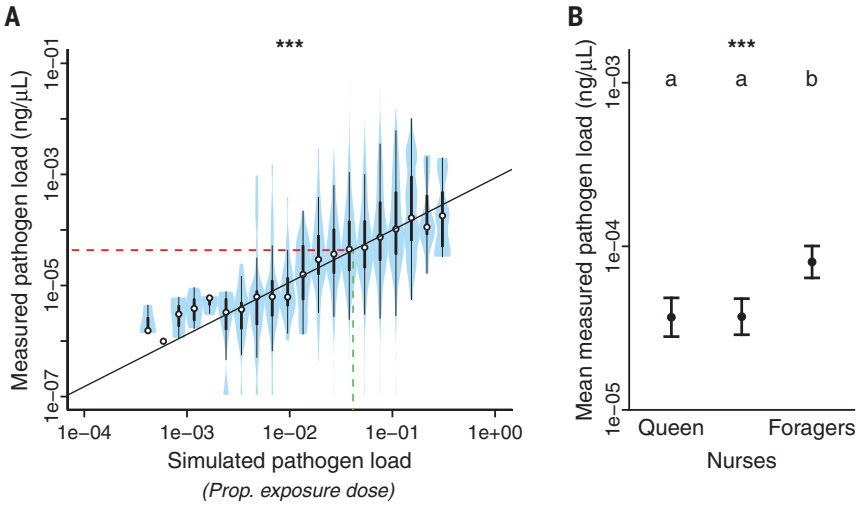


Fig. 2. Experimentally measured and simulated *M. brunneum* transmission in pathogen-exposed colonies. (**A**) Relationship between qPCR-measured pathogen loads and simulated pathogen loads (i.e., the total amount of pathogen received by each ant, averaged across 500 simulations of disease transmission and expressed as a proportion of the original load of treated foragers) in untreated ants. Simulations were run over the posttreatment networks using experimentally treated foragers as disease origin. Points represent the medians, thick black lines represent the interquartile ranges, thin black lines represent 1.5× the interquartile range, and violin shadings represent the density of individual data points. ***$P < 0.0001$ (GLMM). The dashed green line represents the threshold value identified from simulation results in Fig. 3A, and the dashed red line its corresponding value of measured pathogen load calculated using the fitted GLMM model. (**B**) Mean measured pathogen loads in untreated ants. Points represent the means, and whiskers represent the standard errors. ***$P < 0.0001$ (GLMM). Same letters (a and a) indicate $P = 0.96$ and different letters (a and b) indicate $P < 0.0001$ in post hoc comparisons (BH corrected). Sample size: 11 colonies from the main experiment and 11 colonies from the age experiment.

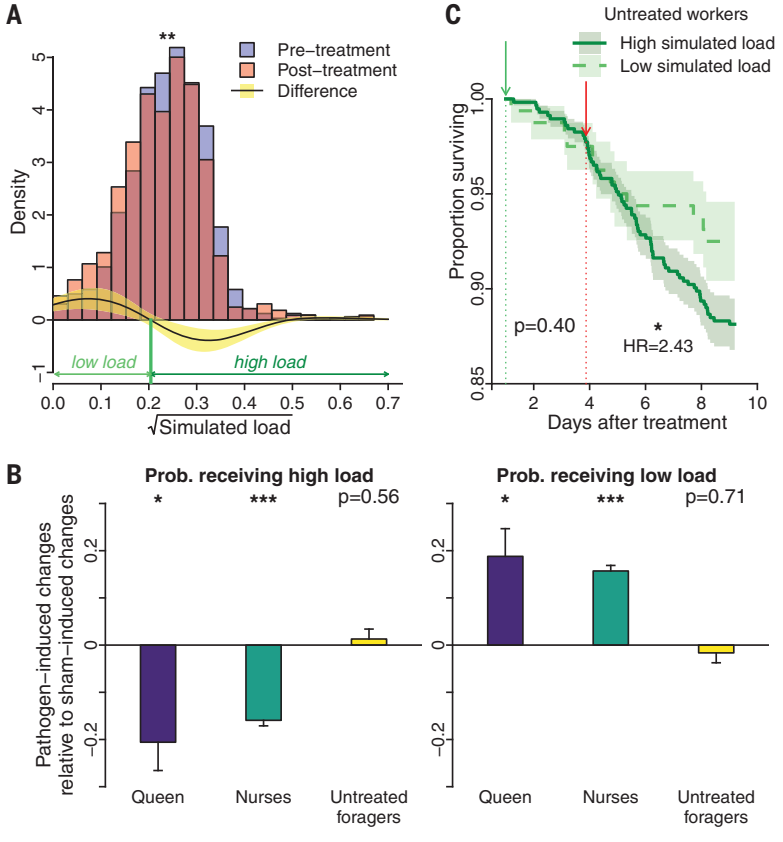


Fig. 3. Pathogen-induced changes in simulated disease transmission. (**A**) Density distributions of the simulated pathogen loads of untreated ants in simulations run over pretreatment and posttreatment networks (purple and peach histograms; overlap in mauve) in pathogen-exposed colonies. The black line and yellow shading represent the mean and standard error of the difference between post- and pretreatment distributions. The green line highlights the value at which there is an inversion in the sign of the density difference. This value was used as a threshold to distinguish high from low simulated loads in all subsequent analyses (e.g., in Fig. 3, B and C). Kolmogorov-Smirnov test, D statistic = 0.073, **$P = 0.006$. (**B**) Pathogen-induced changes in the probability of individuals receiving a high load or a nonzero low load in simulations. Visual representation as in Fig. 1F; statistical analysis on all raw data *$P < 0.05$; ***$P <$ 0.0001 (GLMM; post hoc comparisons between pathogen-induced and sham-induced changes, BH corrected). Sample sizes, pathogen-exposed/sham-treated colonies: 11/11 queens, 862/685 nurses, 214/284 untreated foragers and 105/94 treated workers. (**C**) Survival (lines) ± 95% confidence intervals (shaded areas) of untreated workers predicted to receive either a high load (dark green, solid line) or a low load (light green, dashed line) on the basis of simulations run over the first 24 hours posttreatment in the survival experiment. The green arrow represents the time of load prediction (1 day after treatment), and the red arrow represents the time at which an increase in mortality was detected among untreated ants (around 4 days after treatment; fig. S11). Mixed-effects Cox proportional hazard model comparing high- versus low-load workers, first 4 days: hazard ratio (HR) = 0.57, $\chi^2 = 0.71$, df = 1, $P = 0.40$; after day 4: HR = 2.43, $\chi^2 = 6.43$, df = 1, *$P = 0.011$.

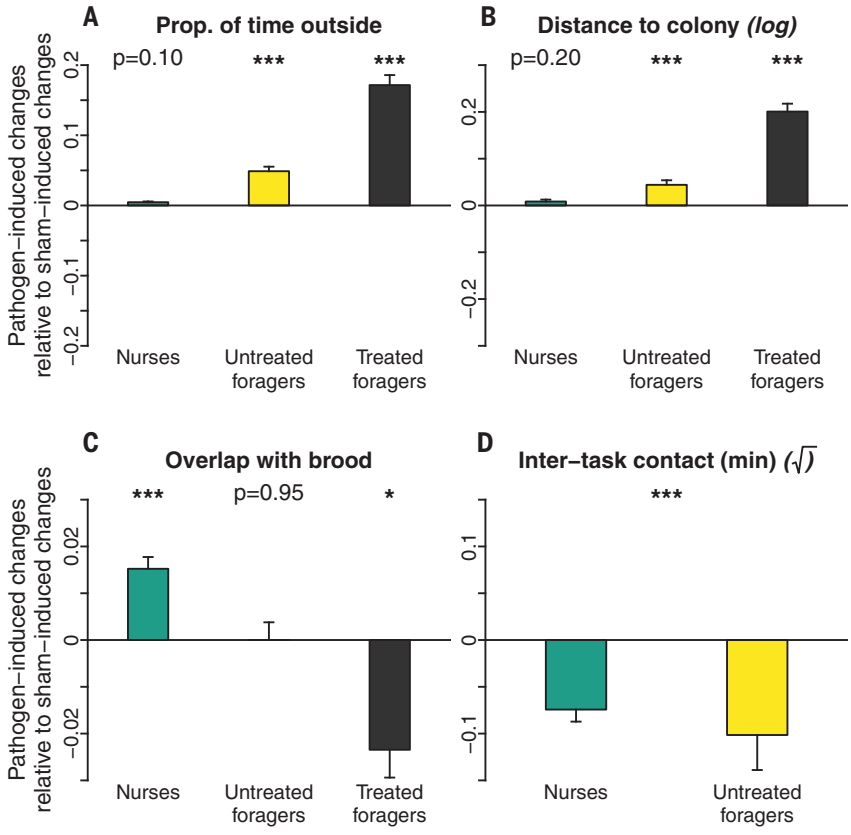


Fig. 4. Pathogen-induced changes in individual behavior. (**A** to **D**) Pathogen-induced changes in the proportion of time individual ants spent outside the nest (A), their distance to colony members (in millimeters measured between the center of gravity of the workers of interest and the rest of the colony) (B), their overlap with the brood (Bhattacharyya's affinity index) (C), and the duration of contact between untreated workers from a given task group (nurses or foragers) and untreated workers from the other task group (D). Visual representation as in Fig. 1F. Statistical analysis on all raw data. For (A) to (C), $*P < 0.05$ and $***P < 0.0001$ (GLMM, post hoc comparisons between pathogen-induced and sham-induced changes, BH corrected). For (D), $***P < 0.0001$ (GLMM, $\text{interaction}_{\text{period}\times\text{treatment}}$). Sample sizes are as in Fig. 3.

contamination loads identified in our simulations have the expected effects on individual survival, we used data from the survival experiment to compare the 9-day mortality of workers predicted to have received either a high or a low load in the simulations. As expected, workers predicted to have received a low load did not experience any detectable change in mortality, whereas workers predicted to have received a high load experienced a greater than twofold increase in mortality around 4 days after treatment (Fig. 3C). This load-dependent difference in mortality was notably observed among workers that had few direct contacts with pathogen-exposed foragers (fig. S11), underscoring the value of using a whole-network approach rather than only considering direct contacts with contaminated individuals to study disease transmission.

Further analyses revealed that the pathogen-induced changes in network properties resulted from behavioral changes not only among pathogen-exposed foragers but also among their untreated nestmates. First, pathogen-exposed foragers actively isolated themselves: They spent more time outside, increased their average distance to the rest of the colony, and reduced their area of movement while inside the nest (GLMM: $|z\ \text{score}| \geq 5.82$, $P < 0.0001$ in post hoc comparisons with BH correction; Fig. 4, A and B, fig. S14, and table S3). This led to a decrease in contact time between pathogen-exposed foragers and all untreated workers ($|z\ \text{score}| \geq 3.61$, $P \leq 0.00047$; fig. S14 and table S3). Untreated foragers also actively isolated themselves: They spent more time outside and increased their average distance to the rest of the colony ($|z\ \text{score}| \geq 3.54$, $P < 0.00061$; Fig. 4, A and B, and table S3). By contrast, nurses increased their spatial overlap with the brood ($z\ \text{score} = -4.43$, $P < 0.0001$; Fig. 4C and table S3) and moved the brood deeper inside the nest (GLMM: $\chi^2 = 22.68$, df = 1, $P < 0.0001$; fig. S14 and table S3). This led to a decrease in contact time between untreated foragers and nurses in pathogen-exposed colonies ($\chi^2 = 68.85$, df = 1, $P < 0.0001$; Fig. 4D and table S3). These pathogen-induced changes in the behavior of untreated workers contributed at least in part to the changes in overall network structure described above, as similar, lower-magnitude effects were observed in networks that excluded experimentally treated foragers (modularity and task assortativity: nonsignificant trends, $\chi^2 \leq 2.13$, df = 1, $P \geq 0.14$; clustering and efficiency: significant effects, $\chi^2 \geq 4.35$, df = 1, $P \leq 0.037$; fig. S15).

Overall, this study shows that both pathogen-exposed and untreated workers are able to rapidly detect the presence of a pathogen and immediately adjust their behavior to reinforce the disease-inhibitory effects of the colony's social network, thus reducing individual contamination risk. Such an early response, occurring before pathogen-exposed workers develop an infection, could potentially be triggered by chemical or mechanical cues associated with the fungal spores (*19*, *25*), though little is currently known about the detection mechanisms involved. Social network plasticity may represent a widespread strategy for collective resilience in the face of environmental hazards, conferring groups with an effective and easy-to-implement mechanism of response to external and internal disturbance.

REFERENCES AND NOTES

1. M. Barthélemy, A. Barrat, R. Pastor-Satorras, A. Vespignani, *J. Theor. Biol.* **235**, 275–288 (2005).
2. K. T. D. Eames, M. J. Keeling, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13330–13335 (2002).
3. R. K. Hamede, J. Bashford, H. McCallum, M. Jones, *Ecol. Lett.* **12**, 1147–1157 (2009).
4. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, *Nature* **438**, 355–359 (2005).
5. D. P. Mersch, A. Crespi, L. Keller, *Science* **340**, 1090–1093 (2013).
6. T. O. Richardson, T. E. Gorochowski, *J. R. Soc. Interface* **12**, 20150705 (2015).
7. T. O. Richardson *et al.*, *PLOS Comput. Biol.* **13**, e1005527 (2017).
8. P. Sah, S. T. Leu, P. C. Cross, P. J. Hudson, S. Bansal, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 4165–4170 (2017).
9. M. Salathé, J. H. Jones, *PLOS Comput. Biol.* **6**, e1000736 (2010).
10. E. M. Volz *et al.*, *PLOS Comput. Biol.* **7**, 10.1371/annotation/85b99614-44b4-4052-9195-a77d52dbdc05 (2011).
11. C. L. Nunn *et al.*, *Philos. Trans. R. Soc. London B Biol. Sci.* **370**, 20140111 (2015).
12. S. Cremer, C. D. Pull, M. A. Fürst, *Annu. Rev. Entomol.* **63**, 105–123 (2018).
13. S. Cremer *et al.*, *Curr. Biol.* **17**, R693–R702 (2007).
14. N. Stroeymeyt *et al.*, *Curr. Opin. Insect Sci.* **5**, 1–15 (2014).
15. D. Naug, S. Camazine, *J. Theor. Biol.* **215**, 427–439 (2002).
16. D. Naug, B. Smith, *Proc. Biol. Sci.* **274**, 61–65 (2007).
17. C. D. Pull, W. O. H. Hughes, M. J. F. Brown, *Naturwissenschaften* **100**, 1125–1136 (2013).
18. S. Angelone, M. J. Bidochka, *J. Invertebr. Pathol.* **156**, 73–76 (2018).
19. M. Konrad *et al.*, *PLOS Biol.* **10**, e1001300 (2012).
20. T. Brütsch *et al.*, *Ecol. Evol.* **7**, 2249–2254 (2017).
21. P. C. Lopes *et al.*, *Sci. Rep.* **6**, 31790 (2016).
22. S. Vestergaard *et al.*, *J. Invertebr. Pathol.* **73**, 25–33 (1999).
23. R. B. Rosengaus *et al.*, *Naturwissenschaften* **86**, 588–591 (1999).
24. L. Liu *et al.*, *Sci. Rep.* **5**, 15106 (2015).
25. W. O. H. Hughes *et al.*, *Proc. Biol. Sci.* **269**, 1811–1819 (2002).
26. N. Stroeymeyt *et al.*, Social network plasticity decreases disease transmission in a eusocial insect. Zenodo (2018); https://doi.org/10.5281/zenodo.1322669
27. N. Stroeymeyt *et al.*, Social network plasticity decreases disease transmission in a eusocial insect. Zenodo (2018); https://doi.org/10.5281/zenodo.1322676
28. V. Latora, M. Marchiori, *Phys. Rev. Lett.* **87**, 198701 (2001).

ACKNOWLEDGMENTS

We thank F. Bienert, A. Pagano, T. Studer, and F. Rein for their help with ant marking; U. Behrmann and D. Segalla for their help with data curation; B. Kheradmand and K. Seif for running preliminary trials that facilitated the study; S. Metzler and M. Konrad for providing dose-survival data in the *L. neglectus*–*M. brunneum* host-pathogen

system; J. Eilenberg for providing the *M. brunneum* fungal strain; and M. Chapuisat, D. Farine, R. Poulin, T. Richardson, B. Casillas-Perez, and three anonymous reviewers for commenting on previous versions of the manuscript. **Funding:** This project was funded by two European Research Council Advanced Grants (Social Life, 249375, and resiliANT, 741491) and two Swiss National Science Foundation grants (CR32I3_141063 and 310030_156732) to L.K. and a European Research Council Starting Grant (SocialVaccines, 243071) to S.C. **Author contributions:** N.S., L.K., and S.C. designed the research. A.C. and D.P.M. developed hardware and software for the production and curation of tracking data. N.S. and A.V.G. performed the research. N.S. analyzed the data and wrote the manuscript with extensive feedback from L.K. and S.C. All authors commented on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data are available at (*26*) and codes are available at (*27*) and https://github.com/laurentkeller/anttrackingUNIL.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/941/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S5
References (*29*–*57*)

2 March 2018; accepted 22 October 2018
10.1126/science.aat4793

NEUROSCIENCE

Egocentric coding of external items in the lateral entorhinal cortex

Cheng Wang[1]*, Xiaojing Chen[1]*, Heekyung Lee[1], Sachin S. Deshmukh[2], D. Yoganarasimha[3], Francesco Savelli[1], James J. Knierim[1,4]†

Episodic memory, the conscious recollection of past events, is typically experienced from a first-person (egocentric) perspective. The hippocampus plays an essential role in episodic memory and spatial cognition. Although the allocentric nature of hippocampal spatial coding is well understood, little is known about whether the hippocampus receives egocentric information about external items. We recorded in rats the activity of single neurons from the lateral entorhinal cortex (LEC) and medial entorhinal cortex (MEC), the two major inputs to the hippocampus. Many LEC neurons showed tuning for egocentric bearing of external items, whereas MEC cells tended to represent allocentric bearing. These results demonstrate a fundamental dissociation between the reference frames of LEC and MEC neural representations.

The hippocampus is critical for the formation and retrieval of episodic memories (*1*–*3*). Episodic memories preserve a person's original, first-person (i.e., egocentric) perspective (*4*). Cognitive map theory proposes that the hippocampus provides an objective spatial framework that is used to organize the various components of an experience, binding them so that they can be stored and later retrieved as a re-experience of the original event (*5*). Current theories propose that the various allocentric spatial correlates of cells in the hippocampus and MEC (*6*) provide the objective spatial framework of an experience. In contrast, representations of the specific components of an experience (i.e., the content) have been hypothesized to be provided by the LEC (*7*, *8*), in an egocentric framework (*9*).

We thus examined the behavior of 275 LEC cells while seven rats performed a foraging task in two-dimensional (2D) open arenas in which the walls or boundaries were the most prominent features of the local environment (Fig. 1A and fig. S1A). Although the firing rate maps of LEC cells showed weak spatial selectivity (*10*, *11*), some LEC cells showed activity patterns that suggested an egocentric representation of the nearby boundaries (*12*) or, equivalently, the center of the apparatus (Fig. 1, B to E, and figs. S2 and S3). We constructed tuning curves of the egocentric bearing (the angle of the vector from the rat to an external item, referenced from the allocentric head direction of the rat) of the nearest boundary ($Bearing_{Boundary}$) (Fig. 1B) (*13*, *14*) and compared them with classic allocentric head direction tuning curves (curves on right, Fig. 1, C, and E). LEC cells tended to have more sharply tuned egocentric bearing selectivity than head direction selectivity (Wilcoxon signed-rank test, z = 8.46, $P = 2.56 \times 10^{-17}$, n = 275) (Fig. 1F). Across the population with significant tuning, the preferred $Bearing_{Boundary}$ covered all angles, albeit nonuniformly, with an underrepresentation of 180° (fig. S4A). From these cells, it was possible to decode the $Bearing_{Boundary}$ of the rats throughout the session (fig. S4) (*15*). Egocentric bearing tuning was not an artifact of directional sampling biases, body-centered self-motion, or movement direction (fig. S5), and it was stable within a session (fig. S6). Some LEC cells showed tuning for distance of the animal to the boundary, and 23% of these cells showed $Bearing_{Boundary}$ selectivity as well (fig. S7 and table S1).

In contrast, $Bearing_{Boundary}$ tuning of MEC neurons recorded in the small-large box task (fig. S1B) was significantly weaker than that of LEC cells, as quantified by the mean vector length of the tuning curve ($MVL_{Boundary}$) (LEC median = 0.31, MEC median = 0.11; Wilcoxon rank-sum test, z = 6.84, $P = 8.0 \times 10^{-12}$; LEC, n = 72, MEC, n = 117) (Fig. 1, G and H, and fig. S8). Because some MEC neurons show conjunctive tuning for location and head direction (*6*), which could be confounded with $Bearing_{Boundary}$ tuning, we applied a generalized linear model (GLM) framework (*15*) (fig. S9) to dissociate an allocentric model (with a spatial location component conjunctive with allocentric bearing, i.e., classic head direction) from an egocentric model (with a spatial location component conjunctive with egocentric bearing). The goodness of fit was assessed with a Bayesian information criterion measure ($\Delta BIC_{Boundary}$)(*15*). A negative value of the $\Delta BIC_{Boundary}$ indicated that the egocentric model was a better fit than the allocentric model, and a positive value indicated the opposite. LEC and MEC cells were significantly better fit by the egocentric and the allocentric models, respectively (LEC median = −0.0076, MEC median = 0.0016; Wilcoxon signed-rank test; LEC: z = −6.44, $P = 1.22 \times 10^{-10}$, n = 72; MEC: z = 5.68, $P = 1.34 \times 10^{-8}$, n = 117; LEC vs. MEC: Wilcoxon rank-sum test, z = −8.88, $P = 6.46 \times 10^{-19}$) (Fig. 1I, fig. S10, and tables S1 to S3), revealing a double dissociation in the preferred reference frames for angular coordinates of LEC and MEC representations.

We constructed mean vector length (MVL) spatial maps to search for the location that generated the sharpest egocentric tuning curve [i.e., the largest MVL (MVL_{Max})] (Fig. 1J). For cells with tuning for the egocentric bearing of the boundary, the location that produces the MVL_{Max} will be at the center of the arena (fig. S2). The MVL_{Max} locations of LEC cells but not of MEC cells were significantly clustered around the center (Monte Carlo test for nonrandom distribution; LEC: $P < 0.001$, n = 72; MEC: P = 0.16, n = 117), with those of MEC cells significantly more scattered than those for LEC cells (bootstrap test, $P < 0.001$) (Fig. 1K and figs. S3 and S8). Moreover, the values of MVL_{Max} were significantly larger for LEC than MEC (LEC median = 0.35, MEC median = 0.22; Wilcoxon rank-sum test, z = 4.73, $P = 2.21 \times 10^{-6}$, LEC, n = 72, MEC, n = 117) (Fig. 1L).

We next investigated whether LEC cells could encode the egocentric bearing of discrete, external items when a rat foraged in an arena that contained four objects (Fig. 2A and fig. S1, C and D) (*16*). We first determined which one of the four objects for each cell generated the strongest egocentric bearing tuning (largest MVL). LEC cells had a significant preference for egocentric representation of those objects ($\Delta BIC_{Object} < 0$), whereas MEC cells preferred the allocentric representation ($\Delta BIC_{Object} > 0$) (LEC median = −0.0026; MEC median = 0.0011; Wilcoxon signed-rank test; LEC: z = −8.13, $P = 4.41 \times 10^{-16}$, n = 140; MEC: z = 3.97, $P = 7.08 \times 10^{-5}$, n = 88; LEC vs. MEC: Wilcoxon rank-sum test, z = −8.35, $P = 6.85 \times 10^{-17}$) (Fig. 2, B and C; figs. S11 and S12; and table S4).

To investigate whether LEC cells represented the locations of each object in the rat's immediate vicinity as it explored, we partitioned the box into sections based on the closest object to each location in the box (Fig. 2D and fig. S11B). Using the data in each partition, we constructed the MVL map to search for the MVL_{Max} location in the entire arena (Fig. 2D) and computed the egocentric bearing tuning of the object in that partition ($Bearing_{Object}$) (Fig. 2E and fig. S13). An object-representing partition was defined as a partition in which the MVL of $Bearing_{Object}$ was >0.25 and the MVL map showed a small, localized peak (less than 20% of the arena with values >70% of the peak value) close to the local object (<25 cm). For the cells with two or more object-representing partitions, the preferred $Bearing_{Object}$ values across the partitions were significantly more similar to each other for LEC cells than for MEC cells (Fig. 2F) (Rayleigh test: LEC, z = 12.15, $P = 1.42 \times 10^{-7}$, n = 15; MEC, z = 0.51, P = 0.62, n = 8; absolute value: LEC median = 18.26°, MEC median = 114.35°; LEC vs. MEC, Wilcoxon rank-sum test, z = −3.58, $P = 3.40 \times 10^{-4}$), indicating a more

[1]Zanvyl Krieger Mind/Brain Institute, Johns Hopkins University, Baltimore, MD, USA. [2]Centre for Neuroscience, Indian Institute of Science, Bangalore, India. [3]National Brain Research Centre, Haryana, India. [4]Solomon H. Snyder Department of Neuroscience, Johns Hopkins University, Baltimore, MD, USA.
*These authors contributed equally to this work. †Corresponding author. Email: jknierim@jhu.edu

reliable egocentric code for objects in LEC than MEC (figs. S11D and S12B).

To determine whether behaviorally salient locations were represented in an egocentric framework, we recorded LEC cells in a goal-oriented task (Fig. 3A and fig. S1E), in which the goal (a food well) was shifted between sessions. Food pellets were thrown sporadically into the food well and other places in the box, encouraging both exploration of the entire box and frequent visits to the goal (Fig. 3B). Some cells encoded the egocentric bearing of the goal [current goal locations (Fig. 3C) and standard goal location (Fig. 3D); see also fig. S14]. The standard goal quadrant (southeast) was overrepresented compared to the northeast quadrant, which never contained a goal location (Monte Carlo test against a uniformly distributed random distribution, $P < 0.002$ for all three sessions) (Fig. 3E). Between sessions, there was a significant change in the number of cells that represented the standard versus displaced goal locations (bootstrap test, $P < 0.01$ for both S1 vs. S2 and S2 vs. S3) (Fig. 3E), which could not be explained by behavioral confounds at the goal location (fig. S15). As in the first experiment, some cells also showed distance selectivity to the MVL_{Max} location (fig. S16 and table S5).

These results demonstrate that LEC represents the bearing of external items in an egocentric framework. They confirm a key prediction of the hypothesis, that the LEC provides the hippocampus with the sensory input that constitutes the content of an episodic memory (*7*, *8*, *16*, *17*). They are also consistent with an earlier hypothesis that, in contrast to the MEC representation of "self" location, LEC mainly represents information of "other" items in an egocentric framework (*9*). Findings in rats, bats, and primates have highlighted the involvement of the hippocampus in such representations of "other" (*14*, *18*–*22*). We suggest that the LEC provides information about items in the environment, which is used for building comprehensive maps of what is "out there" (*9*, *21*, *22*). This functional segregation of "self" information in MEC and "other" information in LEC may reflect a fundamental organizing principle of the medial temporal lobe memory system (*7*, *9*).

Spatial information in the hippocampal formation is generally considered to be represented in an allocentric framework. However, the present findings revealed that LEC cells displayed a robust spatial representation in an egocentric framework. Thus, LEC provides an integration of spatial and nonspatial information, in the form of a spatial relationship between the observer

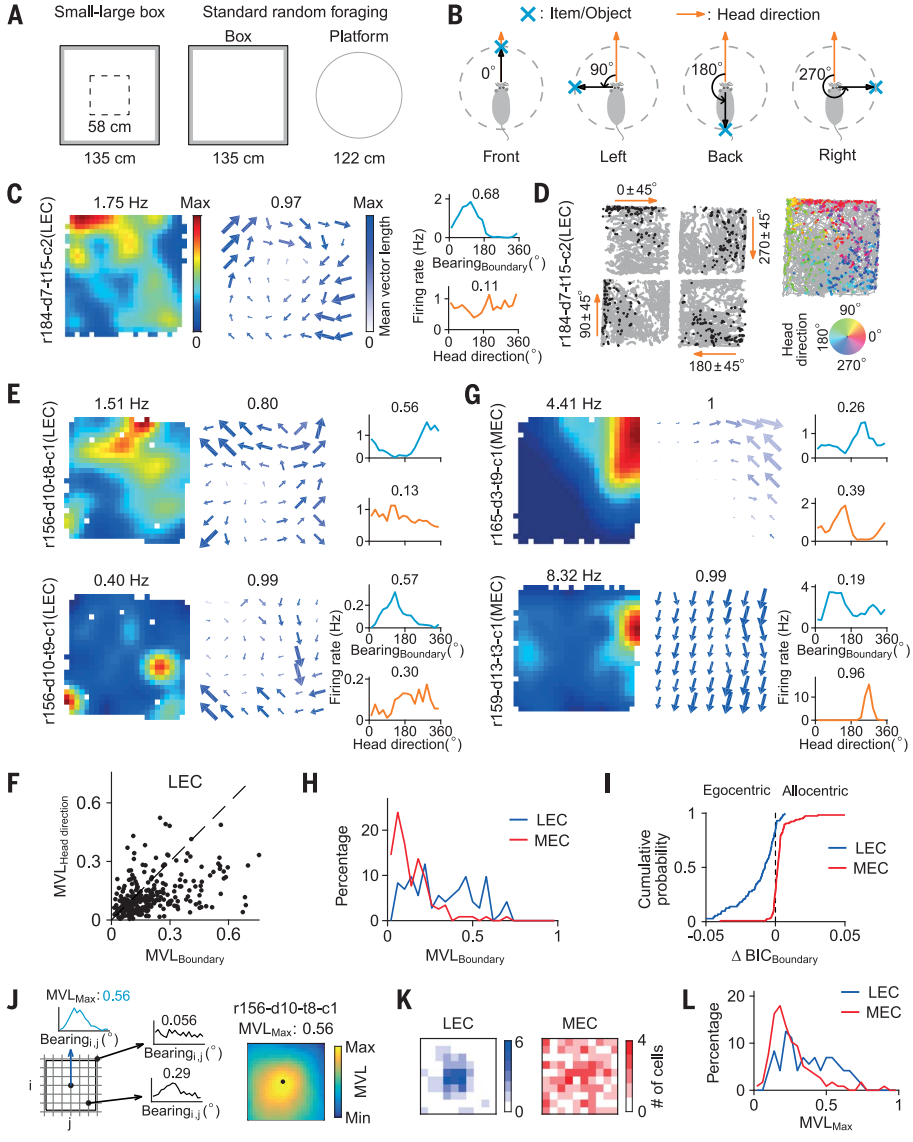


Fig. 1. LEC but not MEC cells show egocentric bearing tuning of nearby boundaries in open arenas. (**A**) Experimental setups. In some experiments (small-large box), rats foraged first in a small box and then the inner walls were removed to allow the rat to forage in the large box. In other experiments, rats foraged either in a square box or a circular platform. (**B**) Definition of egocentric bearing, the angle of the vector (black) from the rat to an external item (cyan), in reference to the rat's allocentric head direction (orange). (**C**) Example of an LEC cell. (Left) Spatial rate map (number on top denotes the peak firing rate). (Middle) Head direction tuning field plot showing preferred head direction (arrow direction) at different locations. Arrow size denotes firing rate. Color saturation denotes the MVL of the tuning curve. Number on top, maximum MVL. (Right) Tuning curves of egocentric bearing of the boundary (top) and allocentric bearing (i.e., classic head direction) (bottom). Number on top is the MVL. (**D**) Same cell as shown in panel C. (Left four panels) The trajectory (gray line) and position of the rat when the cell fired (black dots), separated based on the current head direction. (Right) Overall trajectory; dot color signifies head direction. (**E**) A pair of simultaneously recorded cells showing opposite preferred $Bearing_{Boundary}$ values. (**F**) Scatter plot of MVLs for allocentric head direction tuning vs. $Bearing_{Boundary}$ tuning. (**G**) Two examples of MEC cells (see additional description in the caption for fig. S8B). (**H**) Distributions of MVLs of $Bearing_{Boundary}$ tuning curves of LEC and MEC cells. (**I**) Cumulative distributions of $\Delta BIC_{Boundary}$ values for LEC and MEC cells. (**J**, left) Schematic of MVL map construction. Tuning curves of egocentric bearing of each location (coordinates *i*, *j*) were constructed; the values of the MVL for each location were plotted. (Right) Example of an MVL map with the location of the peak (MVL_{Max} location) denoted with a black dot. (**K**) 2D histograms of the distributions of MVL_{Max} locations for LEC and MEC cells in the small-large box task. Overrepresentation of the arena center corresponds to overrepresentation of the environmental boundaries (fig. S2A). (**L**) Distributions of MVL_{Max} values for LEC and MEC cells.

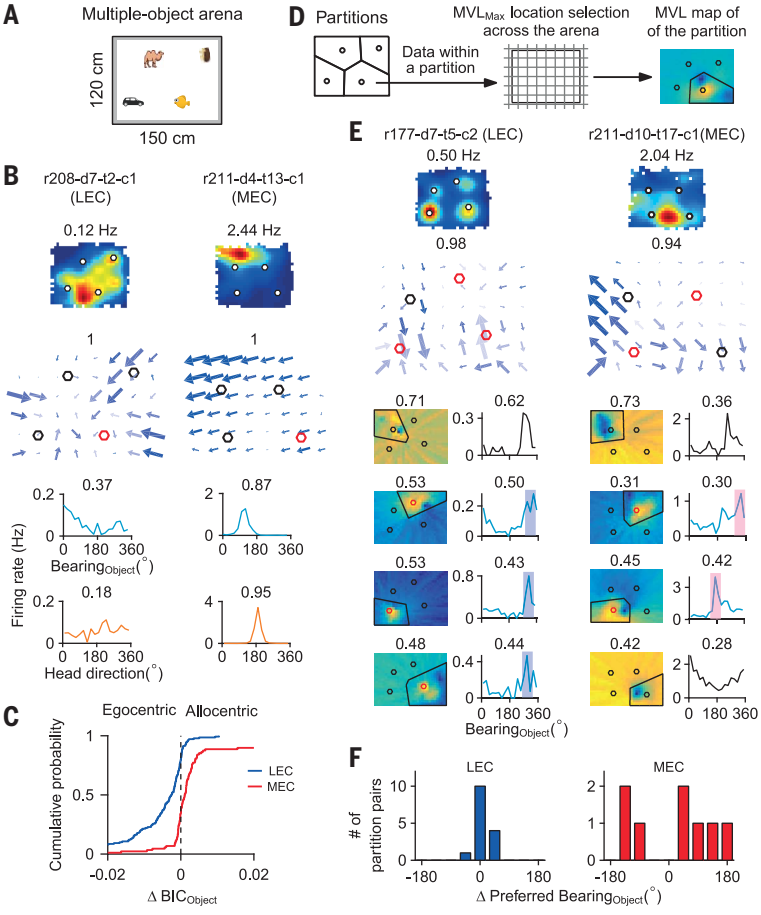


Fig. 2. LEC but not MEC cells show egocentric bearing tuning of 3D objects. (**A**) The multiple-object arena. (**B**) Example cells from LEC (left column) and MEC (right column). Top to bottom, spatial rate map (circles denote the objects), head direction tuning field plot (red circle denotes the object with the highest MVL_{Object}), egocentric bearing tuning curve for the object with the highest MVL_{Object}, head direction tuning curve. (**C**) Cumulative distributions of best ΔBIC_{Object} values for LEC and MEC cells. (**D**) Schematic showing construction of an MVL map with data in a local partition. (**E**) Examples of cells from LEC (left) and MEC (right) with multiple object-representing partitions. (Top to bottom) Spatial rate map, head direction tuning field plot, and MVL maps, together with tuning curves of egocentric bearing of the local object for four partitions. Colored tuning curves were constructed from object-representing partitions (the partitions with red circles denoting the object), which are the ones with a sharp peak near the local object on the MVL map. Note that the preferred bearings (shaded regions) of the LEC cell were consistent, whereas the MEC cell had different preferred bearings for each object. (**F**) Distributions of the differences in preferred $Bearing_{Object}$ values between object-representing partition pairs for LEC and MEC cells with more than one object-representing partition.

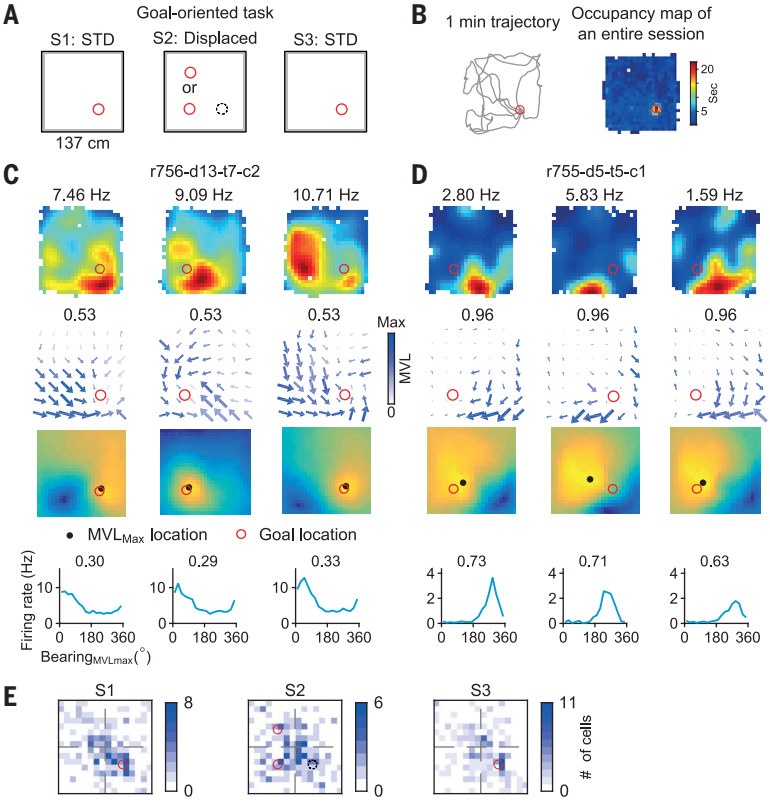


Fig. 3. LEC cells show goal-related egocentric tuning. (**A**) The goal-oriented task. A single food well (red circle) was shifted from the standard goal location (STD) in session 1 to one of two displaced locations in session 2, and then back to the original standard location in session 3. (**B**, left) Typical trajectory of rats in this task for 1 min. (Right) Typical occupancy map in an entire session. The goal location is overoccupied. (**C**) An example cell in three sessions (columns). (Top to bottom) Spatial rate maps, head direction tuning field plots, MVL maps (red circle, food well), and egocentric bearing tuning curves for MVL_{Max} locations. The MVL_{Max} locations for this neuron coincided with the current goal location in each session, as shown in the MVL maps. (**D**) Another example cell. The standard goal location was represented on the MVL maps regardless of the current goal locations. (**E**) 2D histograms of the distributions of MVL_{Max} locations in three sessions.

and specific items in the world, which deviates from the conventional view that spatial ("where") information is conveyed only through the postrhinal (parahippocampal)-MEC pathway (*7*, *8*). These results also clarify the interpretation of recent lesion studies of the entorhinal cortex. First, LEC is critical for performance of an object-place association task that promotes the use of an egocentric strategy (*23*); second, hippocampal cells in rats with a complete lesion of MEC still had place fields, which suggests that input from LEC alone can support hippocampal spatial selectivity (*24*) through transformation of egocentric direction and distance information, as proposed in some of the earliest models of place cells (*25*, *26*). It is thus likely that LEC contributes to the process of abstraction of egocentric information into an allocentric framework for memory storage (*5*, *12*). The parietal cortex is often viewed as an essential node in providing limbic system structures with egocentric representations of cues (*13*) and actions (*27*, *28*) for memory and navigation (*29*), but it is assumed that the information must undergo a transformation from egocentric to allocentric coordinates (*29*) before reaching the hippocampus. However, at least in rats, the LEC receives direct projections from the parietal cortex (*30*), thus allowing the hippocampus to combine directly these egocentric representations from the parietal cortex with the MEC allocentric representations of self-position.

REFERENCES AND NOTES

1. W. B. Scoville, B. Milner, *J. Neurol. Neurosurg. Psychiatry* **20**, 11–21 (1957).
2. L. R. Squire, *Neurobiol. Learn. Mem.* **82**, 171–177 (2004).
3. F. Vargha-Khadem *et al.*, *Science* **277**, 376–380 (1997).
4. M. A. Conway, *Neuropsychologia* **47**, 2305–2313 (2009).
5. J. O'Keefe, L. Nadel, *The Hippocampus as a Cognitive Map* (Oxford: Clarendon Press, 1978).
6. E. I. Moser, M.-B. Moser, B. L. McNaughton, *Nat. Neurosci.* **20**, 1448–1464 (2017).
7. J. J. Knierim, J. P. Neunuebel, S. S. Deshmukh, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130369 (2013).
8. H. Eichenbaum, A. P. Yonelinas, C. Ranganath, *Annu. Rev. Neurosci.* **30**, 123–152 (2007).
9. J. E. Lisman, *Prog. Brain Res.* **163**, 615–625 (2007).
10. E. L. Hargreaves, G. Rao, I. Lee, J. J. Knierim, *Science* **308**, 1792–1794 (2005).
11. D. Yoganarasimha, G. Rao, J. J. Knierim, *Hippocampus* **21**, 1363–1374 (2011).
12. A. Peyrache, N. Schieferstein, G. Buzsáki, *Nat. Commun.* **8**, 1752 (2017).
13. A. A. Wilber, B. J. Clark, T. C. Forster, M. Tatsuno, B. L. McNaughton, *J. Neurosci.* **34**, 5431–5446 (2014).
14. A. Sarel, A. Finkelstein, L. Las, N. Ulanovsky, *Science* **355**, 176–180 (2017).
15. Materials and methods are available as supplementary materials.
16. S. S. Deshmukh, J. J. Knierim, *Front. Behav. Neurosci.* **5**, 69 (2011).
17. A. Tsao, M.-B. Moser, E. I. Moser, *Curr. Biol.* **23**, 399–405 (2013).
18. E. T. Rolls, S. M. O'Mara, *Hippocampus* **5**, 409–424 (1995).
19. N. J. Killian, M. J. Jutras, E. A. Buffalo, *Nature* **491**, 761–764 (2012).
20. L. Acharya, Z. M. Aghajan, C. Vuong, J. J. Moore, M. R. Mehta, *Cell* **164**, 197–207 (2016).
21. T. Danjo, T. Toyoizumi, S. Fujisawa, *Science* **359**, 213–218 (2018).
22. D. B. Omer, S. R. Maimon, L. Las, N. Ulanovsky, *Science* **359**, 218–224 (2018).
23. D. I. G. Wilson, S. Watanabe, H. Milner, J. A. Ainge, *Hippocampus* **23**, 1280–1290 (2013).
24. J. B. Hales *et al.*, *Cell Reports* **9**, 893–901 (2014).
25. J. O'Keefe, *Hippocampus* **1**, 230–235 (1991).
26. P. E. Sharp, *Psychobiology* **19**, 103–115 (1991).
27. B. L. McNaughton *et al.*, *Cereb. Cortex* **4**, 27–39 (1994).
28. J. R. Whitlock, G. Pfuhl, N. Dagslott, M.-B. Moser, E. I. Moser, *Neuron* **73**, 789–802 (2012).
29. P. Byrne, S. Becker, N. Burgess, *Psychol. Rev.* **114**, 340–375 (2007).
30. G. M. Olsen, S. Ohara, T. Iijima, M. P. Witter, *Hippocampus* **27**, 335–358 (2017).

ACKNOWLEDGMENTS

The authors thank G. Rao, A. Smolinsky, and M. Lee for technical assistance and A. Tsao for critical reading of the manuscript. **Funding:** This work was supported by NIH grant R01 MH094146, NIH grant R01 NS039456, and the International Human Frontier Science Program Organization LT00683/2006C (F.S.). **Author contributions:** C.W. and J.J.K. conceived and designed the study. C.W., X.C., H.L., S.S.D., Y.D., and F.S. performed the experiments. C.W. and X.C. performed the data analyses. J.J.K. supervised the experiments and analyses. C.W., X.C., and J.J.K. wrote the paper with inputs from all authors. **Competing interests:** None declared. **Data and materials availability:** Data and code are accessible at https://sites.krieger.jhu.edu/knierim-lab/resources/data-for-wang-chen-et-al/.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/945/suppl/DC1
Materials and Methods
Figs. S1 to S16
Tables S1 to S5
References (*31*–*39*)

15 June 2018; accepted 19 October 2018
10.1126/science.aau4940

DEVELOPMENTAL BIOLOGY

Molecular to organismal chirality is induced by the conserved myosin 1D

G. Lebreton[1]*, C. Géminard[1]†‡, F. Lapraz[1]‡, S. Pyrpassopoulos[2]‡, D. Cerezo[1], P. Spéder[1]§, E. M. Ostap[2], S. Noselli[1]¶

The emergence of asymmetry from an initially symmetrical state is a universal transition in nature. Living organisms show asymmetries at the molecular, cellular, tissular, and organismal level. However, whether and how multilevel asymmetries are related remains unclear. In this study, we show that *Drosophila* myosin 1D (Myo1D) and myosin 1C (Myo1C) are sufficient to generate de novo directional twisting of cells, single organs, or the whole body in opposite directions. Directionality lies in the myosins' motor domain and is swappable between Myo1D and Myo1C. In addition, Myo1D drives gliding of actin filaments in circular, counterclockwise paths in vitro. Altogether, our results reveal the molecular motor Myo1D as a chiral determinant that is sufficient to break symmetry at all biological scales through chiral interaction with the actin cytoskeleton.

Asymmetry is ubiquitous in living organisms and plays essential roles at all biological scales from molecular to behavioral (e.g., polarization of neurons, asymmetric division of stem cells, location of organs, hand preference, etc.). Chirality is a particular kind of asymmetry. A molecule or organ is said to be chiral if it is not superimposable on its mirror image, like our left and right hands.

A fundamental feature of biological systems lies in the chiral uniformity (or homochirality) of the building blocks (L-amino acids, D-sugars, chiral polymers including DNA, microtubules, F-actin, etc.) from which they are assembled. Whether macroscopic asymmetries (e.g., the location of the human heart on the left side of the body or the handedness of snail shell coiling) of living organisms are directly related to their molecular chirality remains an open question. Such a link was suggested by the "F-molecule model," involving a hypothetical F-shaped chiral molecule whose three arms would orient cells in space and differentiate left from right (*1*). The occurrence of asymmetries at all biological scales further raises the question of their origin and mode of propagation. In other words, does a single asymmetry-determining process propagate to higher levels, or do multiple independent processes occur? To understand the role of chirality in biological systems, we examined the establishment of left-right (LR) asymmetry in *Drosophila* (*2*–*5*). In this organism, the conserved *myosin 1D* (*myo1D*) gene is essential for dextral looping of all native LR organs (*6*–*9*). *myo1D* is a situs inversus gene, as its absence leads to the full reversal of organ positioning along the LR axis, with organs adopting a mirror-image orientation (referred to as sinistral). *Drosophila* has several independent, tissue-specific LR organizers in which *myo1D* is expressed and necessary (*10*–*14*).

To determine whether *myo1D* is sufficient to drive LR asymmetry, we ectopically expressed the protein in different naïve tissues (i.e., tissues devoid of LR asymmetry). We found that Myo1D expression in the larval epidermis induces dextral twisting of the whole larval body (Fig. 1, A and B). This phenotype is 100% penetrant and is specific to Myo1D (fig. S1, A to E). The larval body can rotate up to 180°, flipping the mouth parts toward the dorsal side of the larvae (Fig. 1B) and misaligning adjacent denticle belts by 17.4° [0° in wild type (WT)] (fig. S1, F, G, and I). Of note, twisted posture alters locomotion behavior: the larvae move through directional barrel rolling rather than through normal crawling (movie S1). Dextral twisting can also be induced in pupae and adult abdomens (fig. S2, A to D).

To test whether *myo1D* can induce asymmetry at the single-organ level, we ectopically expressed it in tracheal precursors. In this condition, the whole trachea undergoes pronounced dextral twisting (see materials and methods), adopting a spiraling ribbon shape with multiple turns, instead of the smooth and linear conformation of WT trachea (Fig. 1, C and D, and fig. S3). We next investigated twisting at the cellular level, through quantification of the geometry and polarity of epidermal cells expressing Myo1D ectopically. In control conditions, cell membrane orientation shows a Gaussian distribution centered on 0° [i.e., cells are perpendicular to the anterior-posterior (AP) axis] (Fig. 1E). In contrast, Myo1D-expressing cells show elongation (fig. S4) and a clear shift in membrane orientation toward one side (Fig. 1F), indicating that Myo1D also induces directional polarization at the cell level.

The Myo1D protein contains head (motor), neck, and tail domains that are all essential for normal Myo1D function (*15*). To assess their requirement for *myo1D* gain-of-function phenotypes, we expressed Myo1D proteins with point mutations or truncations (*15*). Results show that the integrity of the protein is essential, with all domains being required (Fig. 1G). Point mutations in actin- or adenosine triphosphate (ATP)–binding sites indicate that Myo1D actin-based motor activity is crucial for its function in establishing both native (*15*) and de novo LR asymmetry (Fig. 1G).

To test Myo1D specificity in generating de novo asymmetry, we ectopically expressed seven other *Drosophila* myosins (Fig. 2, A to C, and fig. S5, A to G). Only Myo1C (*16*) overexpression led to twisted larvae (Fig. 2C). Myo1C-induced twisting is opposite (sinistral) to that of Myo1D (Fig. 2, A to C) and less pronounced (90° versus 180° for Myo1D), with a denticle belt angle of −9.9° instead of 17.4° for Myo1D (fig. S1, F, H, and I). Myo1C also induces twisting of trachea (Fig. 2, E and F), pupae, and adults (fig. S2, E and F), as well as spiraling locomotion (movie S2), all with sinistral orientation. Additionally, Myo1C-expressing cells show opposite lateralized polarity to Myo1D-expressing cells (Fig. 2, G and H). Finally, Myo1C requires the same domain integrity as Myo1D for sinistral twisting (Fig. 2I), in particular a fully functional motor domain. We found that both myosins are antagonistic, cancelling out each other's gain-of-function phenotype (Fig. 2, B to D, and fig. S2, G and H), reminiscent of previous data showing that Myo1C can antagonize Myo1D in native LR organs (*6*, *17*).

The fact that two paralogous myosins can polarize the whole larvae in opposite ways provides a framework for determining the molecular basis of directionality. Thus, we performed structure-function analysis by swapping the head, neck, and tail domains from each protein (Fig. 3). Expression of each chimera in the larvae showed that the determinant for directionality in these myosins lies in their motor-head domain (Fig. 3). Keeping the neck and tail domains from the same protein increases protein activity, whereas swapping the sole neck reduces it markedly (Fig. 3), suggesting an important coupling between head and neck for full myosin activity.

Myosin head domains are known to directly interact with actin filaments. Chickadee, the *Drosophila* homolog of profilin, is involved in actin polymerization. Reducing chickadee protein expression led to the suppression or strong reduction of the body-twisting phenotypes induced by either of the myosins (fig. S6, A to G), pointing to actin formation and/or dynamics for myosin chiral activity. Next, we assessed the ability of myosins to produce motion of actin filaments while bound to fluid-supported lipid bilayers in vitro (*18*). Both myosins bind to phosphatidylinositol 4,5-bisphosphate [PI(4,5)P2] (fig. S7) and propel actin filament motility on supported lipid bilayers (movie S3), with Myo1D having a speed 1.7 times that of Myo1C (table S1). Myo1D drove counterclockwise, circular motility of actin, whereas Myo1C did not show any turning bias in these conditions (Fig. 4, A and B, and table S1). Circular motility persisted when Myo1D was attached to a fluid bilayer by means of a biotin-streptavidin linkage, confirming the finding that asymmetric motility is a property of the motor domain. Mixing the two proteins and increasing the Myo1C/Myo1D ratio reduces circular F-actin movement, thus recapitulating the Myo1C antagonism toward

[1]Université Côte D'Azur, CNRS, Inserm, Institut de Biologie Valrose, Nice, France. [2]Pennsylvania Muscle Institute and the Center for Engineering Mechanobiology, University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA.
*Present address: Centre de Biologie Intégrative, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cedex, France. †Present address: Institut de Recherche en Cancérologie de Montpellier (IRCM), Inserm U1194, 208 Avenue des Apothicaires, 34298 Montpellier Cedex 5, France. ‡These authors contributed equally to this work. §Present address: Institut Pasteur, 25 Rue du Dr Roux, 75015 Paris, France.
¶Corresponding author. Email: noselli@unice.fr

Myo1D observed in vivo for normal and de novo LR asymmetry (Figs. 2D and 4C) (*6*, *17*). Altogether, these results reveal a chiral interaction between Myo1D and F-actin in vitro, providing evidence for a molecular origin of LR asymmetry induced by this myosin at the cell, organ, and whole-body levels.

In conclusion, Myo1D represents a chiral determinant that is necessary for native handedness and sufficient to create de novo LR asymmetry from the molecular to the behavioral level, with chiral information being encoded within the motor domain itself. We propose that the multiscale property of Myo1D emerges from its molecular chiral interaction with F-actin (Fig. 4D). This model is in accordance with the so-called F-molecule model for LR asymmetry establishment (*1*), predicting that chiral factors (in particular, molecular motors) set axis direction at the molecular level through vector information and then propagate across organization scales.

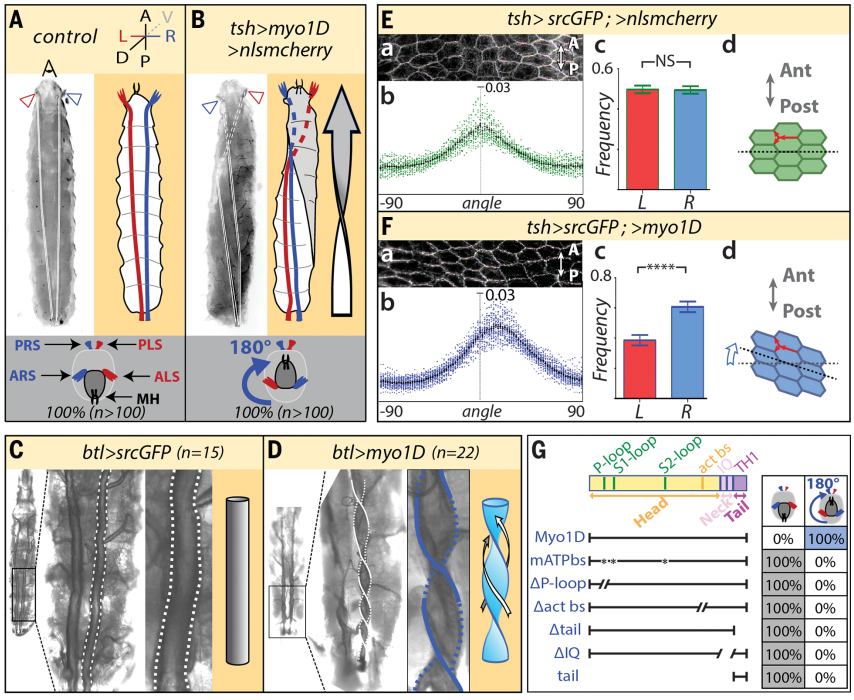


Fig. 1. Myo1D induces de novo LR asymmetry. (**A**) (Top) Control larvae (white, left panel; schematized, right panel) showing bilateral symmetry. The trachea is indicated, with anterior tracheal spiracles indicated by a red (left) or blue (right) open triangle. (Bottom) Front view of the larvae (ventral down, dorsal up). MH, mouth hooks; ALS, anterior left spiracle; ARS, anterior right spiracle; PLS, posterior left spiracle; PRS, posterior right spiracle. Left and right spiracles are shown in red and blue, respectively. A, anterior; P, posterior; L, left; R, right; D, dorsal; V, ventral. (**B**) *myo1D* expression in the epidermis (*tsh>myo1D*) is sufficient to induce dextral twisting of the whole larval body by 180°. (**C**) Linear morphology of a control (*btl>srcGFP*) third instar larvae trachea. (**D**) *myo1D* expression in the tracheae (*btl>myo1D*) induces their dextral twisting with multiple loops. (**E** and **F**) Morphometric analysis of the larval epidermis of WT (Ea) and *myo1D*-expressing epidermal cells (Fa). Graphic plot showing distribution of cellular angles relative to the AP axis (Eb and Fb) and plot of the sum of rightward (−90° to 0°:R)− against leftward (0° to 90°:L)−oriented angles (Ec and Fc), showing that *myo1D* overexpression induces polarized reorganization of epithelial cells toward the dextral orientation (Ed and Fd). *n* = 24 (*n*, sample size) for *tsh>srcGFP>nls-mCherry*; *n* = 39 for *tsh>myo1D-RNAi>srcGFP*. Error bars indicate SD. ****$P < 0.0001$; NS, nonsignificant. (**G**) Structure-function analysis of Myo1D twisting activity. Integrity of the protein as well as its ATP- and actin-binding sites are essential for its function. *n* = 25 for each.

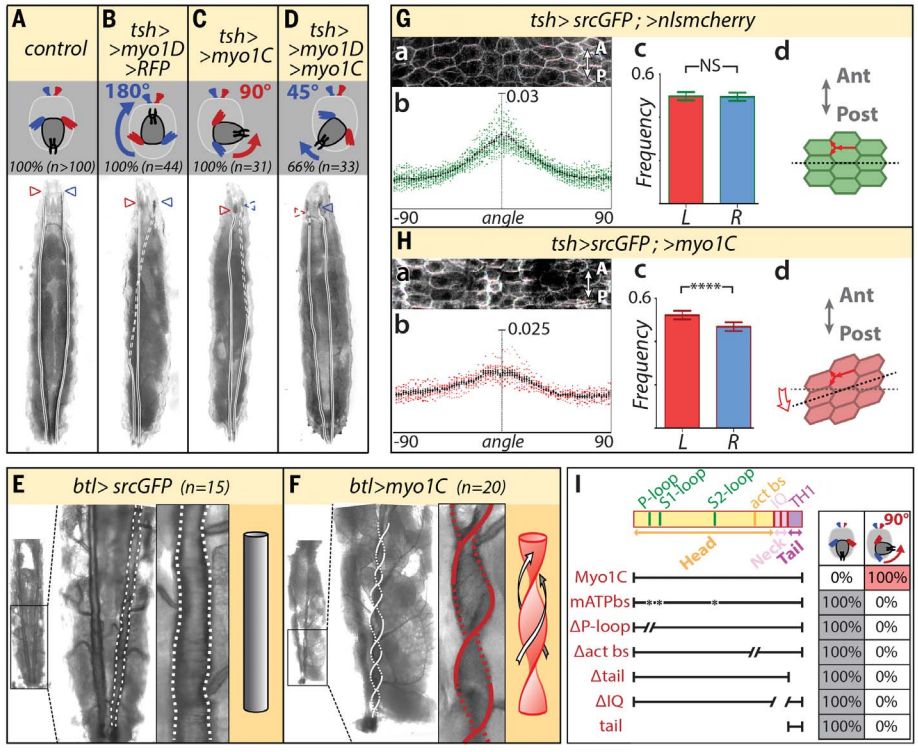


Fig. 2. Myo1C is a sinistral myosin antagonist to Myo1D. (**A** to **D**) Control (white) larvae (A) and larval-twisting phenotype induced by overexpression of *myo1D* (B) or *myo1C* (C) or by coexpression of *myo1D* and *myo1C* (D). (**E**) Linear morphology of a control (*btl>srcGFP*) third instar larvae trachea. (**F**) *myo1c* expression in the tracheae (*btl>myo1C*) induces their sinistral twisting. (**G** and **H**) Morphometric analysis of the larval epidermis of WT (Ga) and *myo1C*-expressing epidermal cells (Ha). Graphic plot showing distribution of cellular angles relative to the AP axis (Gb and Hb) and plot of the sum of rightward (−90° to 0°:R)− against leftward (0° to 90°:L)−oriented angles (Gc and Hc), showing that *myo1C* overexpression induces polarized reorganization of epithelial cells toward sinistral (Gd and Hd). *n* = 24 for *tsh>srcGFP>nls-mCherry*; *n* = 12 for *tsh>myo1C-RNAi>srcGFP*. Error bars indicate SD. ****$P < 0.0001$; NS, nonsignificant. (**I**) Structure-function analysis of Myo1C twisting activity. Integrity of the protein, as well as its ATP- and actin-binding sites, is essential for its function. *n* = 25 for each.

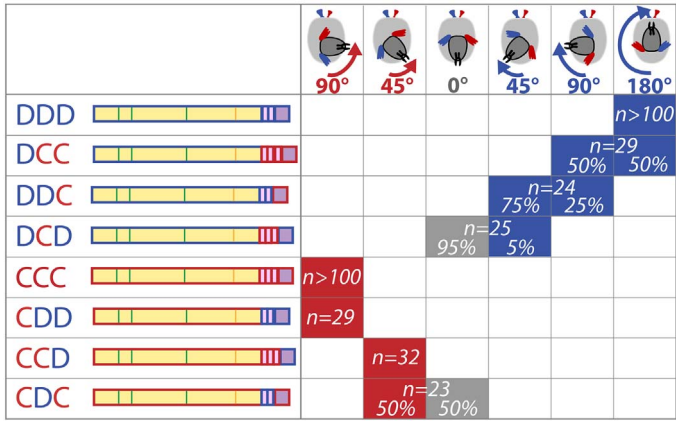


Fig. 3. The motor domain confers Myo1D and Myo1C chirality. Twisting activity of all Myo1D-Myo1C chimeras swapping head (motor), neck, and tail domains. The head domain provides directionality to the proteins. Domains designated D and C pertain to Myo1D and Myo1C, respectively.

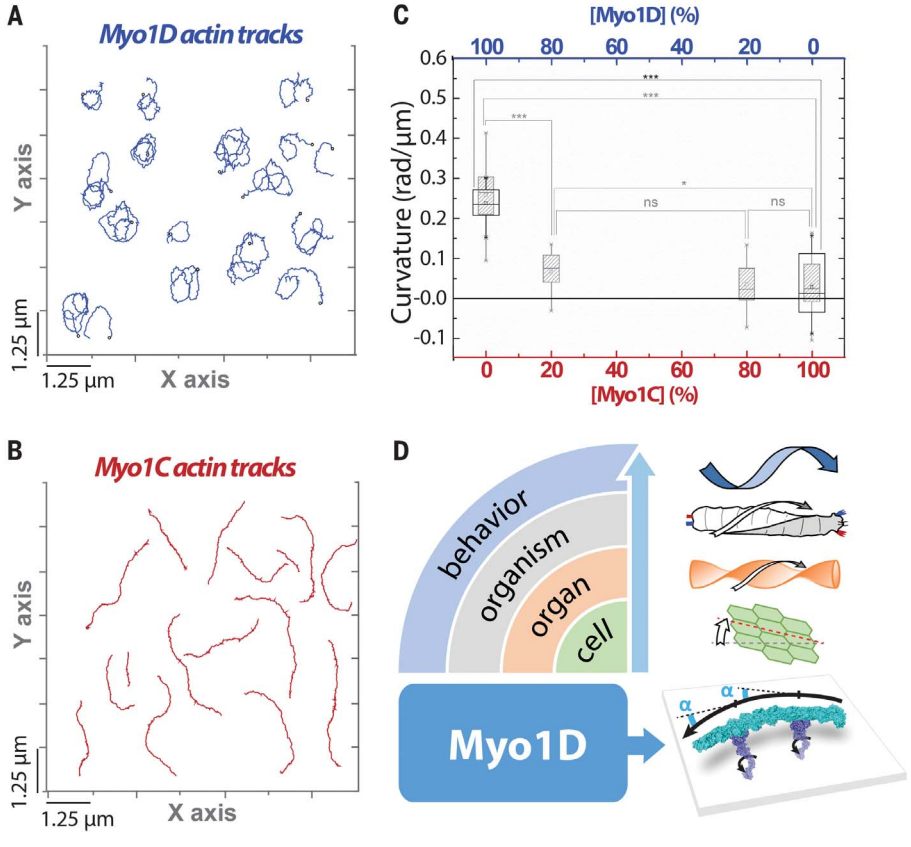


Fig. 4. Chiral interaction between Myo1D and F-actin in vitro. (**A** and **B**) Representative tracks (track origin indicated by open circle) of actin filaments from actin gliding assays on 2% PI(4,5)P2, 98% dioleoylphosphatidylcholine (DOPC)–supported lipid bilayers for Myo1D (A) and Myo1C (B). (**C**) Curvature of actin filament tracks on 2% PI(4,5)P2, 98% DOPC (black boxes)– and 2% biotinPE, 98% DOPC (gray boxes)–supported lipid bilayers for different relative Myo1D and Myo1C concentrations. Error bars correspond to the minimum and maximum values of each dataset. ***$P < 0.001$; *$P < 0.05$; ns, nonsignificant. (**D**) Model of Myo1D chiral activity across multiple organization scales. Bottom cartoon represents actin filament (cyan) gliding powered by Myo1D (purple) on lipid bilayers (gray). The black arrow indicates the actin gliding direction. For most myosins, the power stroke occurs in the direction of the actin filament axis. Mechanisms for actin turning (angle α) include (i) myosin lever-arm translation no longer along the long axis of the actin filament, (ii) myosin lever-arm circular rotation during the power stroke (arrows), or (iii) myosin-induced actin conformational change that results in leftward bending of the actin filament.

Myo1D's ability to generate de novo asymmetry has two important meanings. First, it indicates that some tissues are competent for asymmetry but lack chiral determinant activity, with Myo1D exposing an intrinsic yet only partly unfolded polarized LR axis. Second, the large-scale changes in organ and/or body shape and in posture and/or behavior triggered by simple misregulation of Myo1D fit the fundamental property of so-called toolkit genes proposed to control morphological evolution (*19*). Hence, our findings provide clues to understand the origin of torsion in evolution, proposed to rely on a single genetic event (macromutation) for a gastropod's 180° torsion (*20*). Accordingly, recent results have shown the essential role of actin regulators in snail coiling (*21*, *22*).

REFERENCES AND NOTES

1. N. A. Brown, L. Wolpert, *Development* **109**, 1–9 (1990).
2. M. Blum, K. Feistel, T. Thumberger, A. Schweickert, *Development* **141**, 1603–1613 (2014).
3. J.-B. Coutelis, N. González-Morales, C. Géminard, S. Noselli, *EMBO Rep.* **15**, 926–937 (2014).
4. T. Nakamura, H. Hamada, *Development* **139**, 3257–3262 (2012).
5. A. Raya, J. C. Izpisúa Belmonte, *Nat. Rev. Genet.* **7**, 283–293 (2006).
6. S. Hozumi *et al.*, *Nature* **440**, 798–802 (2006).
7. P. Spéder, G. Adám, S. Noselli, *Nature* **440**, 803–807 (2006).
8. M. Tingler *et al.*, *Curr. Biol.* **28**, 810–816.e3 (2018).
9. T. Juan *et al.*, *Nat. Commun.* **9**, 1942 (2018).
10. J. B. Coutelis, A. G. Petzoldt, P. Spéder, M. Suzanne, S. Noselli, *Semin. Cell Dev. Biol.* **19**, 252–262 (2008).
11. C. Géminard, N. González-Morales, J. B. Coutelis, S. Noselli, *Genesis* **52**, 471–480 (2014).
12. P. Spéder, S. Noselli, *Curr. Opin. Cell Biol.* **19**, 82–87 (2007).
13. N. González-Morales *et al.*, *Dev. Cell* **33**, 675–689 (2015).
14. M. Suzanne *et al.*, *Curr. Biol.* **20**, 1773–1778 (2010).
15. S. Hozumi *et al.*, *Dev. Dyn.* **237**, 3528–3537 (2008).
16. N. S. Morgan, D. M. Skovronsky, S. Artavanis-Tsakonas, M. S. Mooseker, *J. Mol. Biol.* **239**, 347–356 (1994).
17. A. G. Petzoldt *et al.*, *Development* **139**, 1874–1884 (2012).
18. S. Pyrpassopoulos, E. A. Feeser, J. N. Mazerik, M. J. Tyska, E. M. Ostap, *Curr. Biol.* **22**, 1688–1692 (2012).
19. S. B. Carroll, *Cell* **134**, 25–36 (2008).
20. W. F. Ponder, D. R. Lindberg, *Zool. J. Linn. Soc.* **119**, 83–265 (1997).
21. R. Kuroda *et al.*, *Sci. Rep.* **6**, 34809 (2016).
22. A. Davison *et al.*, *Curr. Biol.* **26**, 654–660 (2016).

ACKNOWLEDGMENTS

We thank K. Matsuno, J. Casanova, and T. Lin for reagents; the Bloomington *Drosophila* Stock Center, the National Institute of Genetics Fly (NIG-Fly), and Vienna *Drosophila* RNAi Center (VDRC) for providing *Drosophila* fly lines; the iBV PRISM platform; M. Gettings (Getting Published) for comments; and members of the S.N. laboratory for discussions. **Funding:** F.L. is supported by Agence Nationale pour la Recherche (ANR; ANR-13-BSV2-0006) and Université Côte d'Azur (UCA); G.L. is supported by ANR (ANR-13-BSV2-0006). E.M.O. is supported by NIH Grant R37GM057247. Work in the S.N. laboratory is supported by ANR (ANR-13-BSV2-0006; ANR-17-CE13-0024), UCA, Centre National pour la Recherche Scientifique (CNRS), Institut National pour la Recherche Médicale (Inserm), and the LABEX SIGNALIFE (ANR-11-LABX-0028-01). **Author contributions:** G.L. and S.N. designed the experiments. G.L., C.G., and F.L. performed the experiments (except gliding assays), with the help of D.C. P.S. constructed the chimeras. S.P. performed and S.P. and E.M.O. designed and analyzed the gliding assays. G.L., E.M.O., and S.N. wrote the manuscript with input from all of the authors. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/949/suppl/DC1
Materials and Methods
Figs. S1 to S7
Table S1
References (*23*–*30*)
Movies S1 to S3

13 April 2018; accepted 4 October 2018
10.1126/science.aat8642

IMMUNOLOGY

Structural basis of latent TGF-β1 presentation and activation by GARP on human regulatory T cells

Stéphanie Liénart[1]*, Romain Merceron[2,3]*, Christophe Vanderaa[1], Fanny Lambert[1], Didier Colau[4], Julie Stockis[1,5], Bas van der Woning[6], Hans De Haard[6], Michael Saunders[6], Pierre G. Coulie[1,5], Savvas N. Savvides[2,3]†, Sophie Lucas[1,5]†

Transforming growth factor–β1 (TGF-β1) is one of very few cytokines produced in a latent form, requiring activation to exert any of its vastly diverse effects on development, immunity, and cancer. Regulatory T cells (T_{regs}) suppress immune cells within close proximity by activating latent TGF-β1 presented by GARP (glycoprotein A repetitions predominant) to integrin αVβ8 on their surface. We solved the crystal structure of GARP:latent TGF-β1 bound to an antibody that stabilizes the complex and blocks release of active TGF-β1. This finding reveals how GARP exploits an unusual medley of interactions, including fold complementation by the amino terminus of TGF-β1, to chaperone and orient the cytokine for binding and activation by αVβ8. Thus, this work further elucidates the mechanism of antibody-mediated blockade of TGF-β1 activation and immunosuppression by T_{regs}.

Cytokines of the transforming growth factor–β (TGF-β) family exert widespread and diverse effects on metazoan cells (*1*). Owing to their potency and pleiotropy, family members such as the closely related TGF-β1, -β2, and -β3 isoforms are produced as latent, inactive cytokines, which require a tightly regulated step of extracellular activation to acquire the ability to bind their cognate heterotetrameric receptor (*2*, *3*). The mechanisms of TGF-β activation appear to be isoform- and cell type–specific, but none have been fully elucidated on a structural basis.

TGF-β1 is the predominant isoform expressed by immune cells, including the immunosuppressive FOXP3$^+$ CD4$^+$ regulatory T cells known as T_{regs} (*4*). Immunosuppression by T_{regs} prevents autoimmunity but is detrimental in cancer or chronic infections (*5*). T_{regs} suppress other immune cells within close proximity by activating latent TGF-β1 on their surface (*6*). This results from the following multistep process. Newly synthesized pro-TGF-β1 homodimers form disulfide bonds with GARP (also known as LRRC32), a transmembrane protein retained at the surface of T_{regs} and a few other cell types (*7*–*11*). Pro-TGF-β1 is then cleaved by furin to produce latent TGF-β1, in which the C-terminal dimer, or mature TGF-β1 (mTGF-β1), remains noncovalently associated to the N-terminal dimer known as LAP (latency associated peptide). Latent TGF-β1 is inactive because the two LAP monomers, by means of their arm and "straightjacket" domains, form a ring around mTGF-β1, masking the interaction sites with the TβRI and TβRII receptor chains (*12*). TGF-β1 activation entails the release of mTGF-β1 from LAP. Tensile force exerted by the cytoskeleton and transmitted by integrin αVβ6 to LAP is hypothesized to unfasten and elongate the straightjacket, leading to the release and activation of mTGF-β1 (*13*). However, this mechanism of activation is restricted to cells expressing αVβ6. T_{regs} lack αVβ6. Instead, latent TGF-β1 presented by GARP on the surface of T_{regs} is activated by integrin αVβ8. This was shown using antibodies against either the GARP:(latent)TGF-β1 complex or the β8 integrin subunit, which block TGF-β1 activation and immunosuppression by T_{regs} (*6*, *14*). Despite these advances in our understanding of TGF-β1 activation, the mechanism by which GARP structurally contributes to activation via integrin αVβ8 has eluded the field. Thus, we sought to determine the crystal structure of GARP bound to latent TGF-β1 by using a specific helper antibody to stabilize the complex.

We first coexpressed the extracellular domain of human GARP ($GARP_{ECD}$) and latent TGF-β1 and purified a soluble form of the $GARP_{ECD}$:TGF-β1 binary complex for x-ray crystallography. As this complex proved recalcitrant to crystallization, we reasoned that a ternary assembly with a Fab fragment might be more amenable to crystallization. We chose to use the Fab derived from the monoclonal antibody MHG-8 because it binds GARP:TGF-β1 complexes but not free GARP and blocks TGF-β1 activation by T_{regs} (*6*). Resolving such a complex would also provide key insights into therapeutically relevant epitopes.

Preparative reconstitution of $GARP_{ECD}$:TGF-β1:MHG-8_{Fab} complexes yielded crystals that enabled determination of a structure to 3.1-Å resolution (fig. S1 and table S1). The structure features a tripartite assembly obeying a 1:1:1 stoichiometry, comprising one molecule of $GARP_{ECD}$, one complex of latent TGF-β1 (i.e., one LAP_A/LAP_B and one mTGF-$β1_A$/mTGF-$β1_B$ dimer), and one copy of MHG-8_{Fab} (Fig. 1A and table S2). The structure provided a snapshot of how latent TGF-β1 is sequestered by GARP via an intricate combination of covalent and noncovalent interactions. Tethering of latent TGF-β1 by GARP occurred on the opposite side from the RGD integrin-binding motifs in LAP (*9*, *13*), in an orientation compatible with integrin αVβ8 binding (Fig. 1A).

$GARP_{ECD}$ adopted a solenoid horseshoe structure with a nearly perfect circular curvature and featured 20 leucine-rich repeats (LRRs, R1 to R20) capped by N- and C-terminal motifs (Fig. 1B). Each repeat, comprising a consensus sequence of 21 to 29 residues, manifested a hydrophobic core running through the horseshoe solenoid (Fig. 1C). The concave side of the horseshoe was lined by an extended β sheet of parallel β strands contributed by each of the 20 repeats and the two capping motifs (Fig. 1, B to D). This side also displayed N-linked glycosylation at four of the five predicted sites (Fig. 1B). The convex side displayed an irregular mix of secondary structures connected by loops of various lengths (Fig. 1B and fig. S2). The GARP scaffold was further stabilized by two intrachain disulfide bonds: one (Cys^{26}-Cys^{35}) anchoring a β hairpin in the N-terminal cap and a second (Cys^{425}-Cys^{436}) anchoring the ends of an otherwise disordered loop protruding from the convex side of R15 (Fig. 1B). This structure of GARP may serve as a structural prototype for LRRC33, the only other LRR protein known to bind and present latent TGF-β1 on cell surfaces and that shares high levels of sequence identity with GARP (fig. S2) (*15*, *16*).

The most mechanistically relevant insights from our structure center on the $GARP_{ECD}$:TGF-β1 interface (Fig. 2 and figs. S2 to S4). Half (28 of 56) of the GARP residues interacting with latent TGF-β1 were conserved in LRRC33 (fig. S2). Thus, LRRC33 may interact with latent TGF-β1 in a manner similar to GARP. Most residues (22 of 33) of latent TGF-β1 involved in the interaction with GARP were conserved in TGF-β2 and -β3, which also associate with GARP (figs. S4 and S5, A and B) (*17*). Thus, GARP is poised to actively regulate the bioavailability of all three TGF-β isoforms.

The $GARP_{ECD}$:TGF-β1 complex unmasked rarely seen combinations of structural features, not only due to the diversity and extent of its interaction landscape but also because it involved a twofold-symmetric growth factor (latent TGF-β1), bound by a molecular partner (GARP), that lacks any element of symmetry. Our structure revealed that latent TGF-β1 uses both copies of LAP and only one of the two mTGF-β1 modules to saddle the side of the GARP horseshoe. This results in three main interfaces, namely GARP:mTGF-$β1_B$, GARP:LAP_A, and GARP:LAP_B (Fig. 2A). The binding footprint of latent TGF-β1 to GARP comprised the ascending lateral and convex surfaces roughly halfway into the horseshoe (R4 to R11) and buried a total of ~1750 Å^2 of molecular surface area. Latent TGF-β1 interacted with a

[1]de Duve Institute, UCLouvain, 1200 Brussels, Belgium. [2]Department of Biochemistry and Microbiology, Ghent University, 9052 Ghent, Belgium. [3]VIB Center for Inflammation Research, 9052 Ghent, Belgium. [4]Ludwig Cancer Research, Brussels, Belgium. [5]Walloon Excellence in Lifesciences and Biotechnology, 1300 Wavre, Belgium. [6]argenx, 9052 Zwijnaarde, Belgium.
*These authors contributed equally to this work.
†Corresponding author. Email: sophie.lucas@uclouvain.be (S.Lu.); savvas.savvides@ugent.be (S.N.S.)

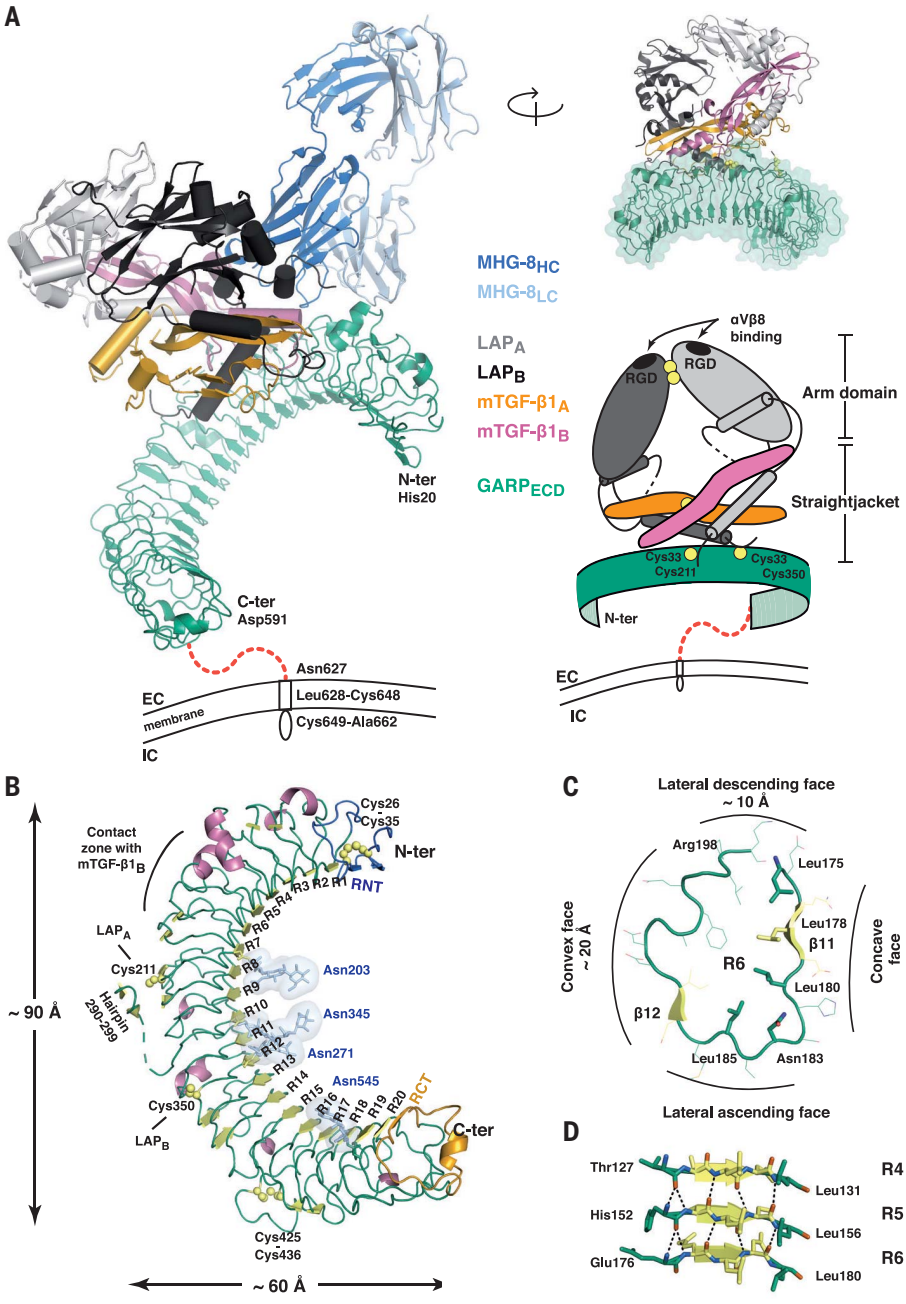


Fig. 1. Structure of the $GARP_{ECD}$:TGF-β1:MHG-8_{Fab} complex. (**A**) (Left) Cartoon representations of the $GARP_{ECD}$:TGF-β1:MHG-8_{Fab} structure as determined by x-ray crystallography. Latent TGF-β1 is composed of two LAP molecules (LAP_A and LAP_B) and two mTGF-β1 molecules (mTGF-$β1_A$ and mTGF-$β1_B$). The red dashed line denotes the linker between the extracellular domain and the transmembrane domain of GARP. (Top right) Cartoon representation of a different view of the GARP:TGF-β1 complex, with MHG-8_{Fab} not illustrated for clarity. (Bottom right) Schematic representation of protein components participating in the $GARP_{ECD}$:TGF-β1 complex. Yellow circles represent interchain disulfide bonds. (**B**) Cartoon representation and close-up view of the $GARP_{ECD}$. Secondary structure elements, modeled N-glycans, and intramolecular disulfide bonds are depicted. (**C**) Cartoon representation of a typical LRR fragment (R6) in GARP. By convention, the loop fragments connecting the concave to the convex side of the repeats form the lateral ascending face, whereas the loops connecting the convex to the concave side form the lateral descending face. (**D**) Cartoon representation and β-strand complementation observed in the GARP structure. The hydrogen bonding network is represented with black dashed lines. IC, intracellular; EC, extracellular; HC, heavy chain; LC, light chain; N-ter, N terminus; C-ter, C terminus; R, LRR (leucine-rich repeat); RNT, LRR N terminus; RCT, LRR C terminus; RGD, RGDLXX(I/L) binding motif for RGD-binding integrins.

predominantly hydrophobic cushion in GARP via van der Waals contacts mediated by the N-terminal loop of LAP_A and LAP_B, by the loop comprising residues 323 to 337 in mTGF-$β1_B$, and by helix α3 in mTGF-$β1_B$ (Fig. 2A, i).

Two structural features were particularly notable in the interfaces of GARP with the LAP modules. First, latent TGF-β1 was covalently tethered to GARP via two disulfide bonds: one between Cys^{33} in LAP_A and Cys^{211} in GARP, the other between Cys^{33} in LAP_B and Cys^{350} in GARP (Fig. 2A, ii and iii) (*9*). Covalent linkage between distinct proteins is an unusual mode of interaction when redox regulation is not at play. The LAP_A–GARP disulfide was nested in a hydrophobic pocket capped by Trp^{297} of GARP (Fig. 2A, ii). Second, the N-terminal loop of LAP_A inserted into a hydrophobic pocket formed by strands β12 in R6 and β14 in R7 of GARP (Fig. 2A, ii). This unanticipated structural feature led to complementation of the GARP horseshoe fold by a β-strand addition contributed by LAP_A (strand β1A), supporting a role of GARP as a molecular chaperone for latent TGF-β1 biogenesis. GARP sealed this unusual interaction by inserting a β hairpin formed by residues 290 to 299. Notably, this hairpin was the only ordered part within a long segment (residues 280 to 317) that originates at R10 and was predicted to be intrinsically disordered (Fig. 1B and fig. S2). A binding mode of this nature was proposed to partially compensate for the high entropic cost associated with the binding of intrinsically disordered segments to partner molecules. However, structural snapshots of this phenomenon have only rarely been experimentally observed (*18*).

Our structure unveiled the structural plasticity of latent TGF-β1 and conformational changes that underlie its sequestration by GARP. The structure of latent TGF-β1 bound to GARP was globally similar to that of unbound porcine latent TGF-β1 (*12*), with structural disorder in the bow-tie region that can associate with αVβ6 integrin (*13*). However, it displayed three drastic local conformational changes compared with all TGF-β1 structures known to date. First, helix α3 in mTGF-$β1_B$ became a 3_{10} helix element (η1) and was rotated upward by 45° (Fig. 2B). Second, the preceding loop with residues 323 to 337 was substantially restructured toward the GARP concavity (Fig. 2B). Such extensive conformational

switching resolved otherwise severe steric clashes and introduced Tyr^{336} of mTGF-$\beta1_B$ into a new structural context stabilized by a stacking interaction with Tyr^{137} of GARP and by hydrogen bonding between its main-chain atoms and Ser^{160} of GARP (Fig. 2A). Third, helix α1 in LAP_B, which mediates interactions between LAP_B and GARP, underwent marked unwinding while remaining physically tethered to GARP via the LAP_B-GARP disulfide (Fig. 2, A and C). Unwinding of helix α1 was also seen in LAP_A. Collectively, such extensive structural plasticity in latent TGF-β1 in a region far from the integrin binding site may explain why it can associate with two structurally very distinct classes of proteins. Namely, GARP or LRRC33 on one hand and LTBPs (latent TGF-β binding proteins) on the other (*15*).

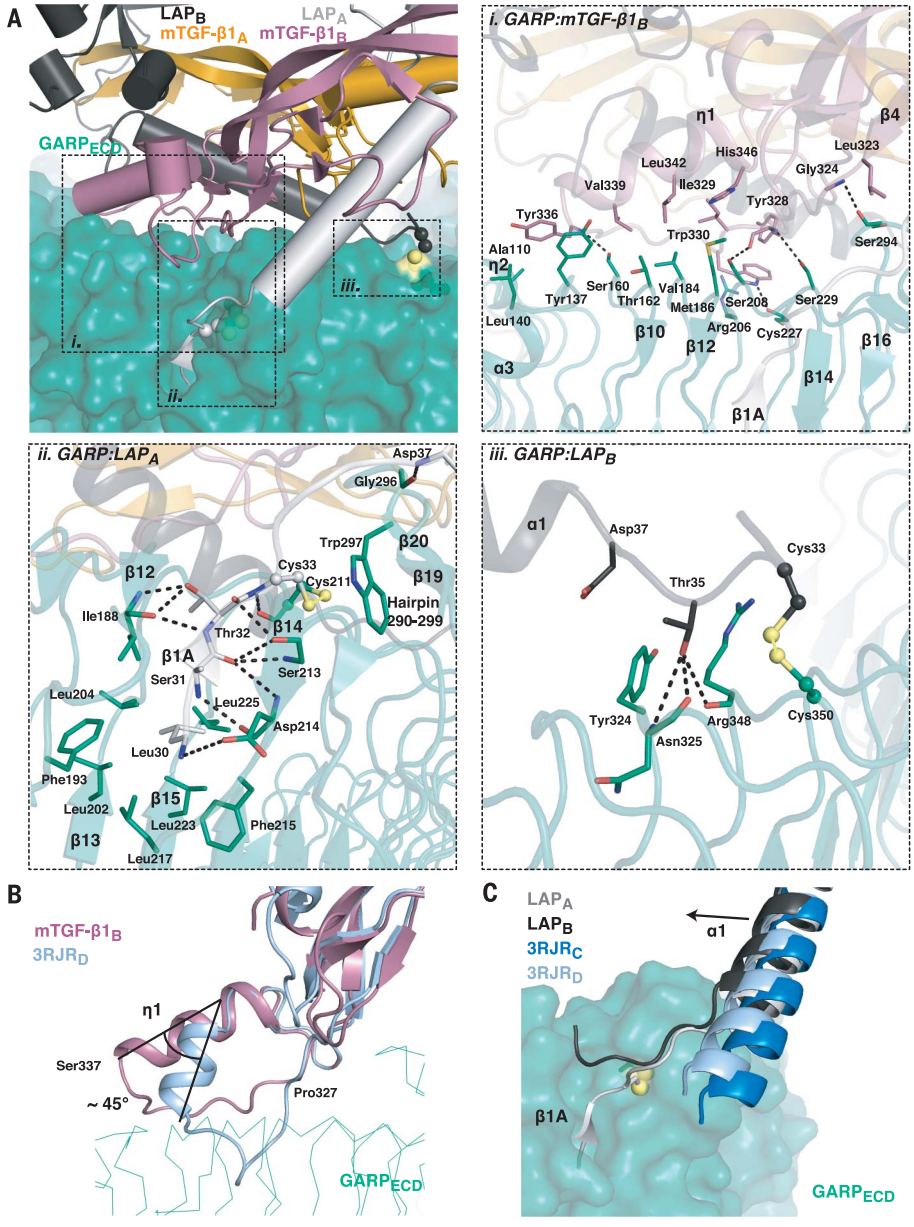


Fig. 2. GARP:TGF-β1 interfaces and plasticity of latent TGF-β1. (**A**) Cartoon representation and close-up views on the three main GARP:TGF-β1 interfaces, outlined by dashed boxes. Details of the GARP:mTGF-$\beta1_B$, GARP:LAP_A, and GARP:LAP_B interfaces are depicted in panels (i), (ii), and (iii), respectively. Proteins are shown in cartoon representation, and secondary structure elements are indicated. The main interfacing residues are shown as sticks, and polar interactions are depicted with black dashed lines. Only contacts observed in both copies of the complex in the asymmetric unit are shown. Disulfide bonds are represented by sticks and spheres, and sulfur atoms are colored yellow. (**B**) Close-up view on the GARP:mTGF-$\beta1_B$ interface, with structural superposition of the porcine mTGF-β1 from porcine latent TGF-β1 (chain D, PDB code 3RJR) showing the main conformational change observed in our structure. (**C**) Close-up view on the GARP:LAP_A interface, with structural superposition of LAP_B in our structure and the two LAP in porcine latent TGF-β1 (chains C and D, PDB code 3RJR) showing a second main conformational change in our structure. This is observed as an unwinding of the LAP helix α1. The disulfide bond between LAP_A and GARP is depicted.

The MHG-8_{Fab} used to generate the crystals blocked TGF-β1 activation by human T_{regs} as efficiently as the full MHG-8 antibody (Fig. 3A). Thus, our crystal structure also revealed how an antibody can block TGF-β1 activation from GARP:TGF-β1 complexes and antagonize T_{reg} suppression (*6*). MHG-8 bound a mixed conformational epitope, concomitantly contacting residues from GARP and latent TGF-β1 (Figs. 1A and 3, B to D). Specifically, CDR1 and CDR3 of the MHG-8 light chain and CDR3 of the heavy chain contacted the convex side of GARP, whereas the heavy chain also contacted the C-terminal arm domain of LAP_B and the 3_{10} helix η1 of mTGF-$\beta1_B$ via its CDR2, CDR3, and framework regions (Fig. 3, C and D, and figs. S6 to S8). We confirmed involvement of certain residues by measuring MHG-8 binding to GARP:TGF-β1 complexes including hemagglutinin (HA)–tagged mutant forms of GARP or latent TGF-β1 expressed in human embryonic kidney (HEK) 293T cells. Mutations at three sites in GARP (Leu^{140}→Lys and Glu^{142}→Leu, Arg^{143}→Ala, or Arg^{163}→Glu) and one site in mTGF-β1 (Lys^{338}→Glu) substantially decreased binding of MHG-8 (Fig. 3B). A majority (9 of 17) of LAP or mTGF-β1 residues in contact with MHG-8 were not conserved in the corresponding positions of TGF-β2 and -β3. This explains why MHG-8 does not bind GARP:TGF-β3 complexes and only binds GARP:TGF-β2 with very low affinity (figs. S4 and S5C).

The antagonistic activity of MHG-8 on T_{regs} may result from its ability to impair the release of mTGF-β1 induced by combined pulling by integrin αVβ8 on one side and resistance or synergistic pulling by GARP on the other side of LAP (Figs. 3C and 4). By simultaneously contacting amino acids in the arm of LAP_B and in GARP, MHG-8 may reinforce the fastening of the straightjacket, preventing the elongation required to release mTGF-β1 (Figs. 3C and 4). Alternatively, by cross-linking amino acids in GARP with the 3_{10} helix η1 of mTGF-$\beta1_B$, MHG-8 may prevent the release of mTGF-β1 and immobilize its 3_{10} helix η1 in a position that is incompatible with binding to the TβRI chain of its receptor, even if the straightjacket is elongated and relaxed. In either scenario, MHG-8 sterically prevents mTGF-β1 release by engaging both GARP and latent TGF-β1 while still allowing binding of αVβ8, as demonstrated previously (*14*).

Thus, we elucidate the role of GARP in TGF-β1 biogenesis and activation, as well as the mechanistic principles underlying antibody-mediated blockade of TGF-β1 activation and immunosuppression by human T_{regs}. Visualization of this large molecular assembly illustrates the feasibility of blocking TGF-β1 activity emanating from a precisely defined and restricted cellular source, such as the surface of T_{regs}. This work also suggests how other TGF-β1–presenting proteins, such as LTBPs deposited in the extracellular matrix (*2*, *19*), may operate as alternative

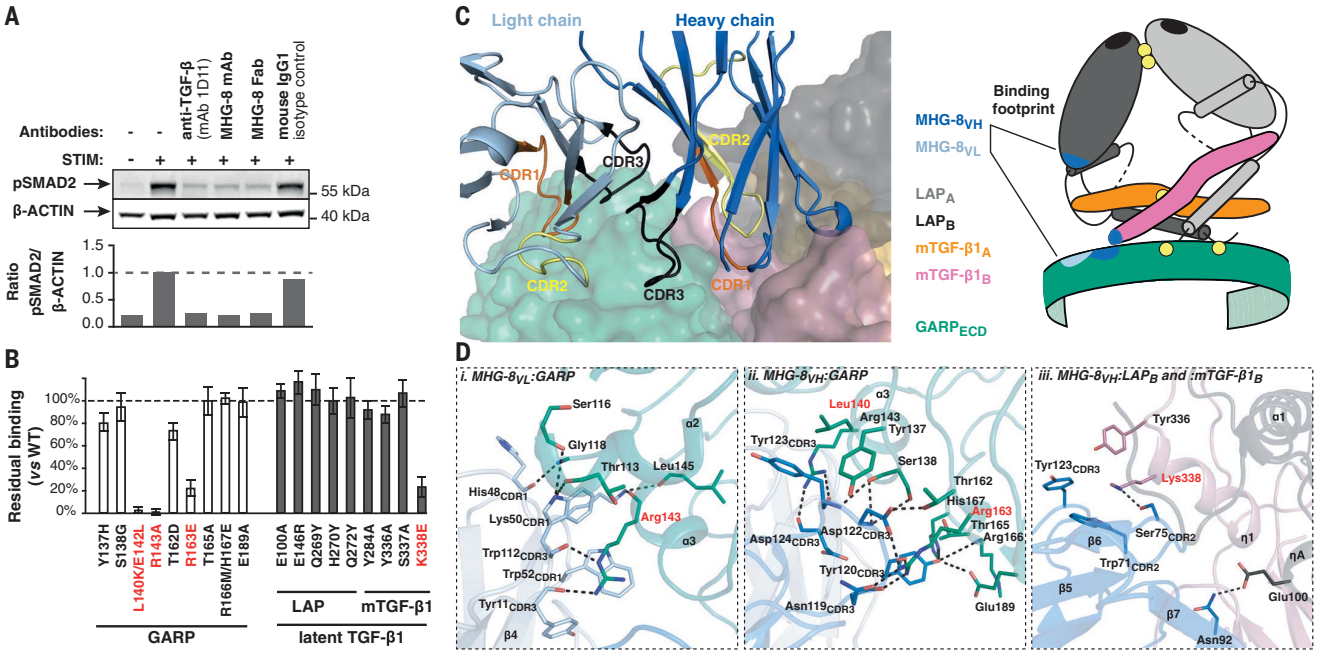


Fig. 3. Structural description of the MHG-8$_{Fab}$:GARP and MHG-8:TGF-β1 interfaces. (**A**) Immunoblot analysis of SMAD2 phosphorylation (pSMAD2) as a readout for TGF-β1 activation in human T_{regs} stimulated in the presence or absence of the indicated antibodies (20 μg/ml) or Fab (13 μg/ml). STIM, stimulation with anti-CD3– and anti-CD28–coated beads; mAb, monoclonal antibody; IgG1, immunoglobulin G1. (**B**) Binding of MHG-8 to transfected HEK293T cells expressing GARP:TGF-β1 complexes comprising the indicated N-terminally HA-tagged mutant forms of GARP or TGF-β1. Bar graphs indicate residual binding to a given mutant complex by comparison with the wild-type (WT) complex taken as 100% (mean of three to five experiments). Error bars represent the 99% confidence interval around the mean. Mutations in red are associated with significantly reduced binding compared with the WT, as their confidence interval does not contain the 100% binding threshold indicated by the dashed horizontal line. (**C**) (Left) Close-up view on the MHG-8$_{Fab}$:GARP and MHG-8:TGF-β1 interfaces. GARP and TGF-β1 are shown in surface representation, and the variable region of MHG-8$_{Fab}$ is depicted as a cartoon. CDR1, -2, and -3 loops are displayed in orange, yellow, and black, respectively. (Right) Schematic representation of the overall structure and contact area by MHG-8$_{VH}$ (dark blue) and MHG-8$_{VL}$ (light blue). (**D**) Details of the MHG-8$_{VL}$:GARP, MHG-8$_{VH}$:LAP$_A$, and MHG-8$_{VH}$:LAP$_B$:mTGF-β1$_B$ interfaces are depicted in panels (i), (ii), and (iii), respectively. Proteins are represented as cartoons, and secondary structure elements are indicated. Main interfacing residues are shown as sticks, and polar interactions are indicated by black dashed lines. Only contacts observed in both copies of the complex in the asymmetric unit are shown. CDR, complementarity determining region; VH, variable domain of the heavy chain; VL, variable domain of the light chain. Single-letter abbreviations for amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

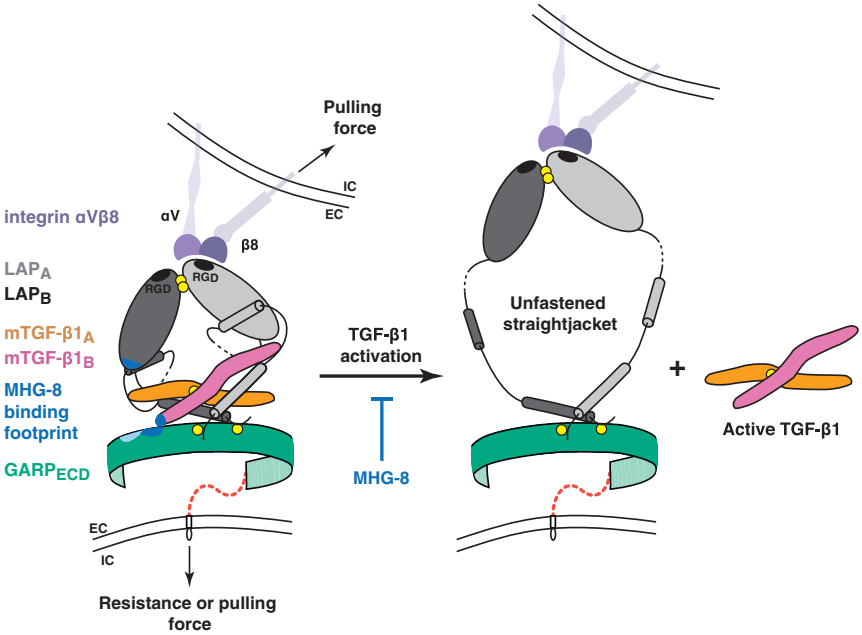


Fig. 4. Model for the integrin- and GARP-dependent mechanical activation of TGF-β1 and its antagonism by monoclonal antibody MHG-8. A schematic representation of the GARP$_{ECD}$:TGF-β1:MHG-8$_{Fab}$ complex, with the MHG-8 binding footprint depicted in blue, is shown. Visualization of the covalent sequestration of latent TGF-β1 by GARP suggests how GARP may facilitate TGF-β1 activation by integrin αVβ8 on the T_{reg} surface. Structural analyses of a pro-TGF-β1 monomer bound by the integrin αVβ6 head show how cytoskeletal forces elicit reshaping of latent TGF-β1 to release mTGF-β1 from LAP (*13*). Maximal harnessing of such forces depends on the direction of the force vector (*20*). Thus, GARP may provide strong anchoring to the cell surface, with its juxtamembrane region providing the necessary flexibility that allows for the optimal alignment of pulling-force vectors. Our crystal structure also suggests how an antibody can block TGF-β1 activation from GARP:TGF-β1 complexes on T_{regs} by cross-linking GARP to one LAP and one mTGF-β1 monomer, which sterically prevents mTGF-β1 release while still allowing integrin αVβ8 binding.

sources of active TGF-β1 and may also be specifically targeted by antibody- or small-molecule–based reagents. Such structural and mechanistic insights may facilitate the design of particularly specific approaches to treat various diseases associated with altered TGF-β1 activity and/or dysfunctional T_{reg} responses, most notably for cancer immunotherapy.

REFERENCES AND NOTES

1. C. J. David, J. Massagué, *Nat. Rev. Mol. Cell Biol.* **19**, 419–435 (2018).
2. A. P. Hinck, T. D. Mueller, T. A. Springer, *Cold Spring Harb. Perspect. Biol.* **8**, a022103 (2016).
3. I. B. Robertson, D. B. Rifkin, *Cytokine Growth Factor Rev.* **24**, 355–372 (2013).
4. M. A. Travis, D. Sheppard, *Annu. Rev. Immunol.* **32**, 51–82 (2014).
5. H. Nishikawa, S. Sakaguchi, *Curr. Opin. Immunol.* **27**, 1–7 (2014).
6. J. Cuende *et al.*, *Sci. Transl. Med.* **7**, 284ra56 (2015).
7. E. Gauthy *et al.*, *PLOS ONE* **8**, e76186 (2013).
8. J. Stockis, O. Dedobbeleer, S. Lucas, *Mol. Biosyst.* **13**, 1925–1935 (2017).
9. R. Wang *et al.*, *Mol. Biol. Cell* **23**, 1129–1139 (2012).
10. J. Stockis, D. Colau, P. G. Coulie, S. Lucas, *Eur. J. Immunol.* **39**, 3315–3322 (2009).
11. D. Q. Tran *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 13445–13450 (2009).
12. M. Shi *et al.*, *Nature* **474**, 343–349 (2011).
13. X. Dong *et al.*, *Nature* **542**, 55–59 (2017).
14. J. Stockis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E10161–E10168 (2017).
15. Y. Qin *et al.*, *Cell* **174**, 156–171.e16 (2018).
16. A. C. Ng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108** (suppl. 1), 4631–4638 (2011).
17. B. X. Wu *et al.*, *J. Biol. Chem.* **292**, 18091–18097 (2017).
18. T. Flock, R. J. Weatheritt, N. S. Latysheva, M. M. Babu, *Curr. Opin. Struct. Biol.* **26**, 62–72 (2014).
19. I. B. Robertson, D. B. Rifkin, *Cold Spring Harb. Perspect. Biol.* **8**, a021907 (2016).
20. Y. Chen, S. E. Radford, D. J. Brockwell, *Curr. Opin. Struct. Biol.* **30**, 89–99 (2015).

ACKNOWLEDGMENTS

We thank S. Depelchin for editorial assistance and J. Cuende for early technical assistance. Proteros biostructures GmbH performed the crystallization and synchrotron radiation experiments. **Funding:** This work was supported by grants from WELBIO (grant BF-2014-01) and the ERC (Consolidator grant TARG-SUP 682818) to S.Lu., and from the VIB to S.N.S. S.Li. was supported by a Télévie fellowship (F.R.S.-FNRS). **Author contributions:** S.Li., R.M., B.v.d.W., H.D.H., M.S., P.G.C., S.N.S., and S.Lu. analyzed the data; S.Li., R.M., S.N.S., and S.Lu. wrote the manuscript; S.Lu. conceived of the study; S.Li., C.V., F.L., D.C., and J.S. performed the mutagenesis experiments, binding experiments and experiments with T_{regs} and produced and purified the mAb; B.v.d.W., H.D.H., and M.S. produced the GARP:latent TGF-β1 complexes; and R.M. performed crystallographic refinement and structure analysis with contributions from S.N.S. **Competing interests:** The authors declare the following financial competing interests: B.v.d.W., H.D.H., and M.S. are full-time employees of argenx BVBA. Patents pertaining to the results presented in the paper have been filed under the Patent Cooperation Treaty (International Publication Number WO 2016/125017 A1), with S.Lu., S.Li., B.v.d.W., H.D.H., P.G.C., and M.S. as inventors and UCLouvain and argenx as applicants. S.Lu. has received research support from argenx and owns stock options in argenx. **Data and materials availability:** Atomic coordinates and structure factor files have been deposited in the Protein Data Bank (PDB) with accession code 6GFF. The MHG-8 antibody is available from S.Lu. under a material transfer agreement with the UCLouvain. All other data needed to evaluate the conclusions of this study are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/952/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S3
References (*21*–*24*)

25 May 2018; accepted 10 October 2018
Published online 25 October 2018
10.1126/science.aau2909

IMMUNOLOGY

ESCRT-dependent membrane repair negatively regulates pyroptosis downstream of GSDMD activation

Sebastian Rühl[1,2], Kateryna Shkarina[3]*, Benjamin Demarco[3]*, Rosalie Heilig[3]*, José Carlos Santos[3]*, Petr Broz[1,3]†

Pyroptosis is a lytic form of cell death that is induced by inflammatory caspases upon activation of the canonical or noncanonical inflammasome pathways. These caspases cleave gasdermin D (GSDMD) to generate an N-terminal GSDMD fragment, which executes pyroptosis by forming membrane pores. We found that calcium influx through GSDMD pores serves as a signal for cells to initiate membrane repair by recruiting the endosomal sorting complexes required for transport (ESCRT) machinery to damaged membrane areas, such as the plasma membrane. Inhibition of the ESCRT-III machinery strongly enhances pyroptosis and interleukin-1β release in both human and murine cells after canonical or noncanonical inflammasome activation. These results not only attribute an anti-inflammatory role to membrane repair by the ESCRT-III system but also provide insight into general cellular survival mechanisms during pyroptosis.

Gasdermin D (GSDMD) is a pore-forming protein that induces pyroptosis, a necrotic form of cell death that is initiated after inflammasome activation (*1*–*3*). The detection of pathogen- or host-derived danger signals by inflammasomes triggers the activation of inflammatory caspases (caspase-1 and -11 in mice and caspase-1 and -4 in humans), which cleave GSDMD to release autoinhibition on its N-terminal domain ($GSDMD^{NT}$) (*1*, *2*). The $GSDMD^{NT}$ targets the plasma membrane and organelles, where it forms large pores to initiate pyroptosis (*4*–*7*). Damage to the plasma membrane does not necessarily result in cell death, as it has been observed that the influx of Ca^{2+} ions from the extracellular milieu triggers repair programs involving either the endocytosis of a damaged membrane or its shedding in the form of ectosomes (*8*–*11*). This latter mechanism relies on components of endosomal sorting complexes required for transport (ESCRT-0 and -III and associated factors that control ESCRT recruitment and disassembly) and has been specifically implicated in the restoration of plasma membrane integrity during necroptosis (*12*) or upon chemical- or laser-induced damage (*10*, *11*).

To investigate if cells repair membranes damaged by GSDMD pores, we imaged the disruption of the electrochemical gradient (by using the Ca^{2+} dye Fluo-8) and loss of membrane integrity [by using propidium iodide (PI)] in mouse bone marrow–derived macrophages (BMDMs). These cells were either untreated or transfected with lipopolysaccharide (LPS) to activate caspase-11. Within 2 hours, up to 50% of LPS-transfected wild-type (WT) BMDMs underwent pyroptosis, as demonstrated by a strong PI signal (PI^{hi}) (fig. S1A). A marked spike of Fluo-8 signal preceded the PI^{hi} signal, indicating a loss of membrane integrity and Ca^{2+} influx (Fig. 1, A and B; fig. S1, B to D; and movies S1 and S2). In contrast, Fluo-8 or PI signals did not change in *Casp11*$^{-/-}$ or *Gsdmd*$^{-/-}$ BMDMs (Fig. 1B; fig. S1, C and D; and movies S1 and S2). Unexpectedly, we found that Ca^{2+} influx–negative WT BMDMs that maintained the electrochemical gradient gradually acquired low levels of PI (PI^{lo}) (Fig. 1C). This PI^{lo} signal reached only 10% of the PI^{hi} signal observed in the WT BMDMs that had lost membrane integrity upon LPS transfection. Furthermore, neither unstimulated WT BMDMs (Fig. 1A) nor LPS-transfected *Casp11*$^{-/-}$ or *Gsdmd*$^{-/-}$ cells showed

[1]Focal Area Infection Biology, Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland. [2]Department of Immunology, St. Jude Children's Research Hospital, Memphis, TN 38105, USA. [3]Department of Biochemistry, University of Lausanne, Chemin des Boveresses 155, 1066 Epalinges, Switzerland.
*These authors contributed equally to this work.
†Corresponding author. Email: petr.broz@unil.ch

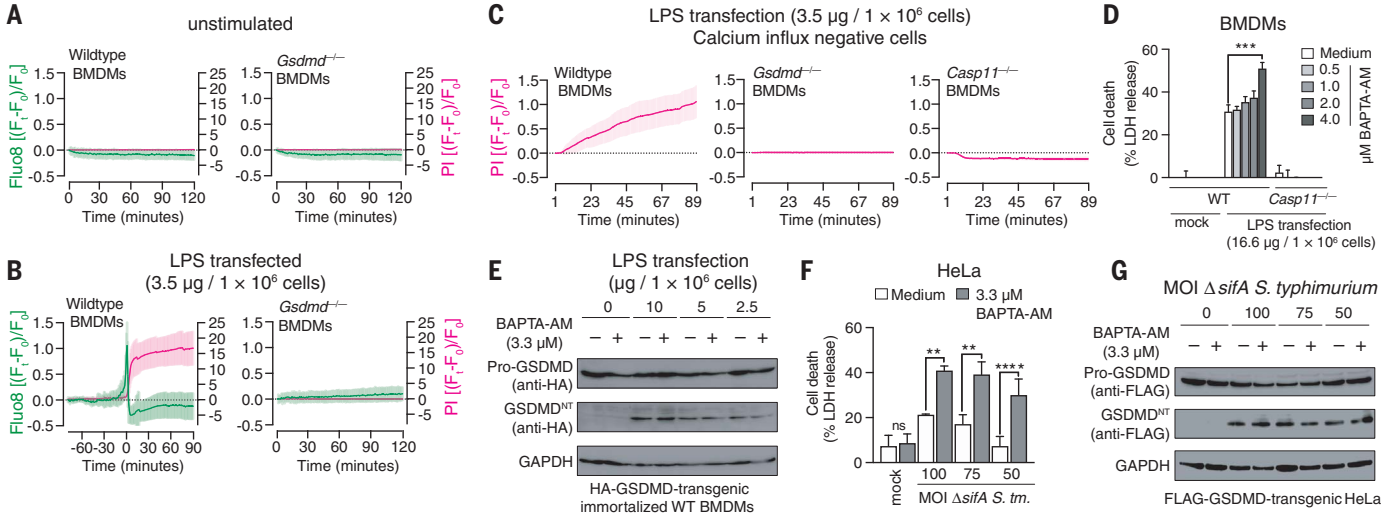


Fig. 1. GSDMD-induced calcium flux negatively regulates pyroptosis. (**A** and **B**) Fluo-8 (Ca^{2+} indicator) and PI signals in unstimulated or LPS-transfected WT and *Gsdmd*$^{-/-}$ BMDMs. Graphs show average data from n = 22, 22, 28, and 30 cells. F_t, Fluorescence at time t; F_0, Fluorescence at time 0. (**C**) Acquisition of PI^{lo} signal in Ca^{2+} influx–negative WT, *Casp11*$^{-/-}$, and *Gsdmd*$^{-/-}$ BMDMs treated as described for panel (A). Graphs show average data from n = 29, 29, and 26 cells. (**D**) LDH release from BAPTA-AM–treated BMDMs mock transfected or transfected with LPS for 2 hours. (**E**) GSDMD processing in hemagglutinin (HA)–GSDMD–transgenic iBMDMs 2 hours post-LPS transfection. GAPDH, glyceraldehyde-3-phosphate dehydrogenase. (**F** and **G**) LDH release and GSDMD processing in untreated or BAPTA-AM–treated HeLa cells infected with Δ*sifA S. typhimurium* for 5 hours. MOI, multiplicity of infection. The graphs in panels (D) and (F) show means ± SD. $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$; ns, not significant (Student's t test). Results are representative of at least three independent experiments.

this PI^{lo} signal (Fig. 1C). These findings confirmed that the low-level PI influx was caused by GSDMD pores and suggested that the PI^{lo} cells had repaired their plasma membrane to prevent lysis, given the absence of a Fluo-8 signal peak.

We then used the Ca^{2+} chelators BAPTA-AM and EDTA to block membrane repair, as previously reported (*10*, *11*). Ca^{2+} chelators increased the levels of cell lysis and lactate dehydrogenase (LDH) release from WT BMDMs after LPS transfection in a dose-dependent manner (Fig. 1D and figs. S1A and S2). Notably, Ca^{2+} chelation was not cytotoxic (Fig. 1D and fig. S2A) and did not enhance LPS transfection or caspase-11 activation, as assessed by immunoblotting for $GSDMD^{NT}$ (Fig. 1E and fig. S3A). BAPTA-AM also increased caspase-4–dependent pyroptosis without significantly altering GSDMD processing in HeLa cells infected with a Δ*sifA* deletion strain of *Salmonella enterica* serovar Typhimurium (hereafter *S. typhimurium*) (Fig. 1, F and G, and fig. S3B) (*13*). Thus, Ca^{2+} chelation enhanced pyroptosis in human and mouse cells, potentially by preventing the initiation of plasma membrane repair.

ESCRT proteins target membranes during membrane repair to form a punctate pattern (*10*, *12*). Consistently, CHMP4 puncta were observed in HeLa cells expressing CHMP4 fused to green fluorescent protein (CHMP4-GFP) from an endogenous promoter (*14*) after *S. typhimurium* infection, $GSDMD^{NT}$ expression, or perforin treatment (control), in contrast with untreated cells (fig. S4 and movie S3). Moreover, Ca^{2+} chelation significantly reduced the number of cells with CHMP4-GFP–positive puncta (Fig. 2A), indicating that ESCRT assembly requires Ca^{2+} influx via GSDMD pores. Next, we transfected CHMP4-mCherry or CHMP4-GFP fusion proteins into human embryonic kidney (HEK) 293T cells stably expressing caspase-1 fused with a modified FKBP domain (DmrB–Casp-1). After dimerization and activation of DmrB–Casp-1 with B/B homodimerizer, cells underwent pyroptosis as demonstrated by morphological changes and plasma membrane annexin-V staining (Fig. 2, B and C, and movies S4 to S6). CHMP4 relocalized to a punctate pattern in DmrB–Casp-1-transgenic HEK293T cells expressing either high [Fig. 2, B and C, and fig. S5, A to C (arrows)] or low (fig. S5B) levels of CHMP4.

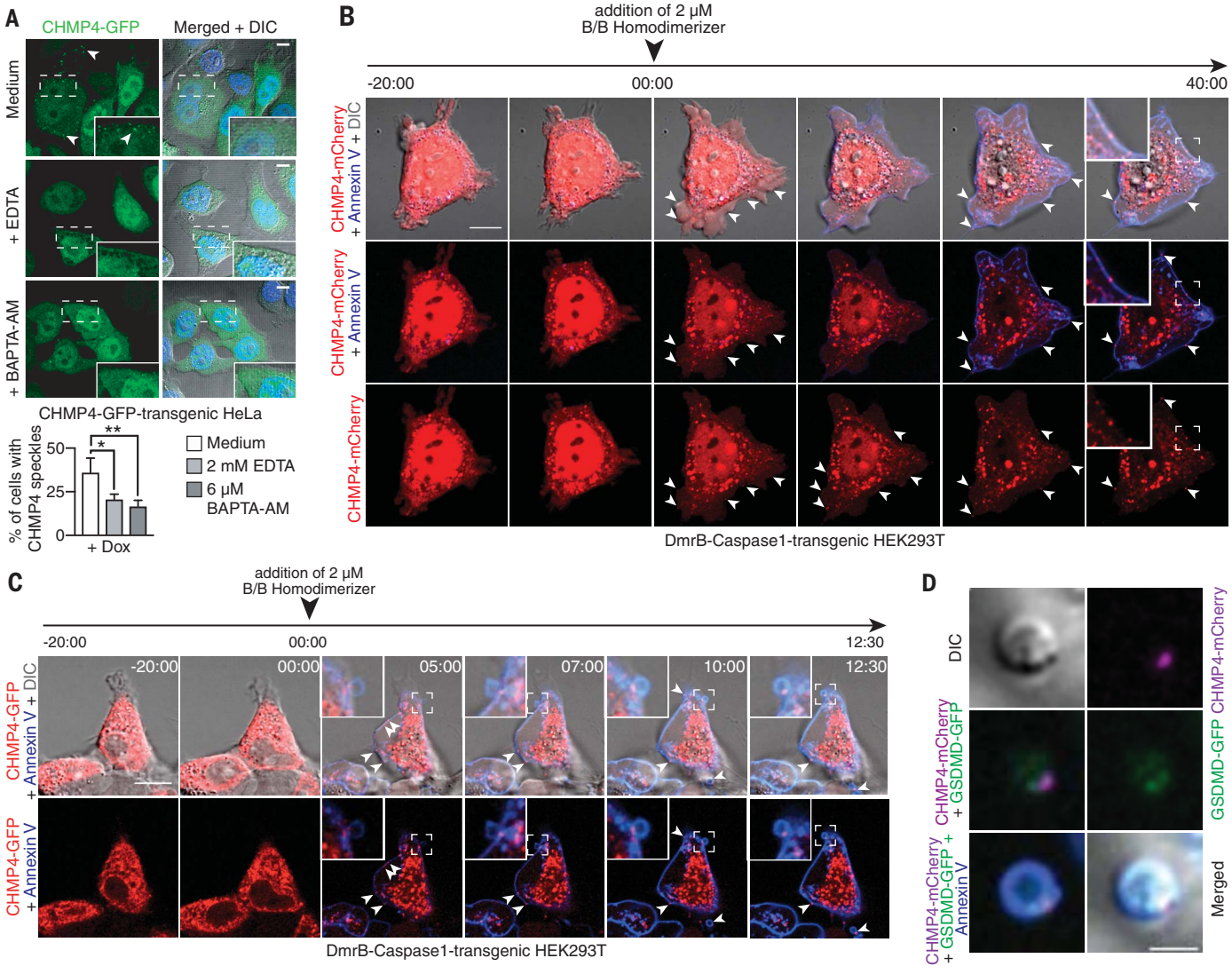


Fig. 2. The ESCRT machinery translocates to the plasma membrane during pyroptosis. (**A**) Fixed-cell microscopy images and CHMP4-GFP puncta quantification of HeLa cells expressing CHMP4-GFP and Dox-inducible $GSDMD^{NT}$ treated with 1 μg/ml of Dox for 4.5 hours. DIC, differential interference contrast. (**B** and **C**) Time-lapse confocal images of either CHMP4-mCherry–expressing (B) or CHMP4-GFP–expressing (C) DmrB–Casp-1–transgenic HEK293T cells stained with annexin-V as membrane marker. Insets show CHMP4 localizing to the plasma membrane or necks of budding vesicles. (**D**) Close-up of annexin-V^+ vesicle released from CHMP4-mCherry–expressing and $GSDMD\text{-}GFP^{internal}$–expressing DmrB–Casp-1–transgenic HEK293T cells after a 1-hour homodimerizer treatment. Graphs show means ± SD. $^*P < 0.05$; $^{**}P < 0.01$ (Student's *t* test). Results are representative of at least three independent experiments. Arrowheads indicate CHMP4 puncta. Scale bars, 5 μm (A), 10 μm [(B) and (C)], or 1 μm (D).

Similar CHMP4 puncta were detected after activation of DmrB–Casp-11 (fig. S6A). Although many CHMP4 puncta clearly localized to the plasma membrane (Fig. 2, B and C; figs. S4 to S6; and movie S6), puncta also formed in the cytoplasm, presumably on intracellular organelles. During necroptosis, the ESCRT machinery promotes the formation of annexin-V$^+$ plasma membrane vesicles, which have been proposed to remove damaged areas of the plasma membrane (*12*). Similar annexin-V$^+$ vesicles were detected budding from HEK293T cells after activation of DmrB–Casp-1 (Fig. 2, C and D; fig. S5D; and movie S5) or DmrB–Casp-11 (fig. S6B). Notably, these vesicles showed clear CHMP4 puncta at the neck (Fig. 2C and figs. S5D and S6B). When GSDMD-GFP^{internal} (fig. S7A) was coexpressed, it localized close to CHMP4-mCherry puncta on the cell periphery (fig. S7B) and on the membrane of annexin-V$^+$ vesicles (Fig. 2D and fig. S7C), suggesting that ESCRT-induced ectosomes remove GSDMD pores from the plasma membrane.

We then examined the impact of ESCRT-III inactivation on inflammasome effector functions. We transiently expressed either WT or dominant-negative CHMP3 or VPS4A in immortalized mouse BMDMs (iBMDMs) or HeLa cells (fig. S8, A to C), because prolonged ESCRT depletion is cytotoxic (*10*, *12*). The expression of dominant-negative VPS4A^{E228Q} mutant protein (Glu^{228}→Gln) or the CHMP3$^{1-179}$ fragment inhibited ESCRT disassembly, as verified by a punctate pattern (fig. S8D), and reduced cell viability after 20 to 24 hours of expression (fig. S8E). This prompted us to restrict our experiments to a maximum of 15 hours postinduction with doxycycline (Dox). The expression of VPS4A^{E228Q} in WT iBMDMs resulted in two- to fourfold-higher levels of cell death after LPS transfection compared with VPS4A^{WT}-expressing controls (Fig. 3, A and B; fig. S9; and movies S7 and S8) and also decreased recovery after LPS transfection (fig. S10). ESCRT inactivation also strongly enhanced interleukin-1β (IL-1β) release after LPS transfection (Fig. 3C), but notably, it enhanced neither LDH nor cytokine release from untreated, *Casp11*$^{-/-}$, or *Gsdmd*$^{-/-}$ cells (fig. S11). IL-1β release after noncanonical inflammasome activation depends on GSDMD-induced K$^+$ efflux and subsequent activation of the NLRP3–caspase-1 pathway (*15*, *16*). Because LPS-transfected WT iBMDMs expressing VPS4A^{E228Q} had higher levels of processed, activated caspase-1 and released more mature IL-1β than VPS4A^{WT}-expressing control cells (Fig. 3, C and D), we hypothesized that this was due to enhanced K$^+$ efflux via increased GSDMD pore

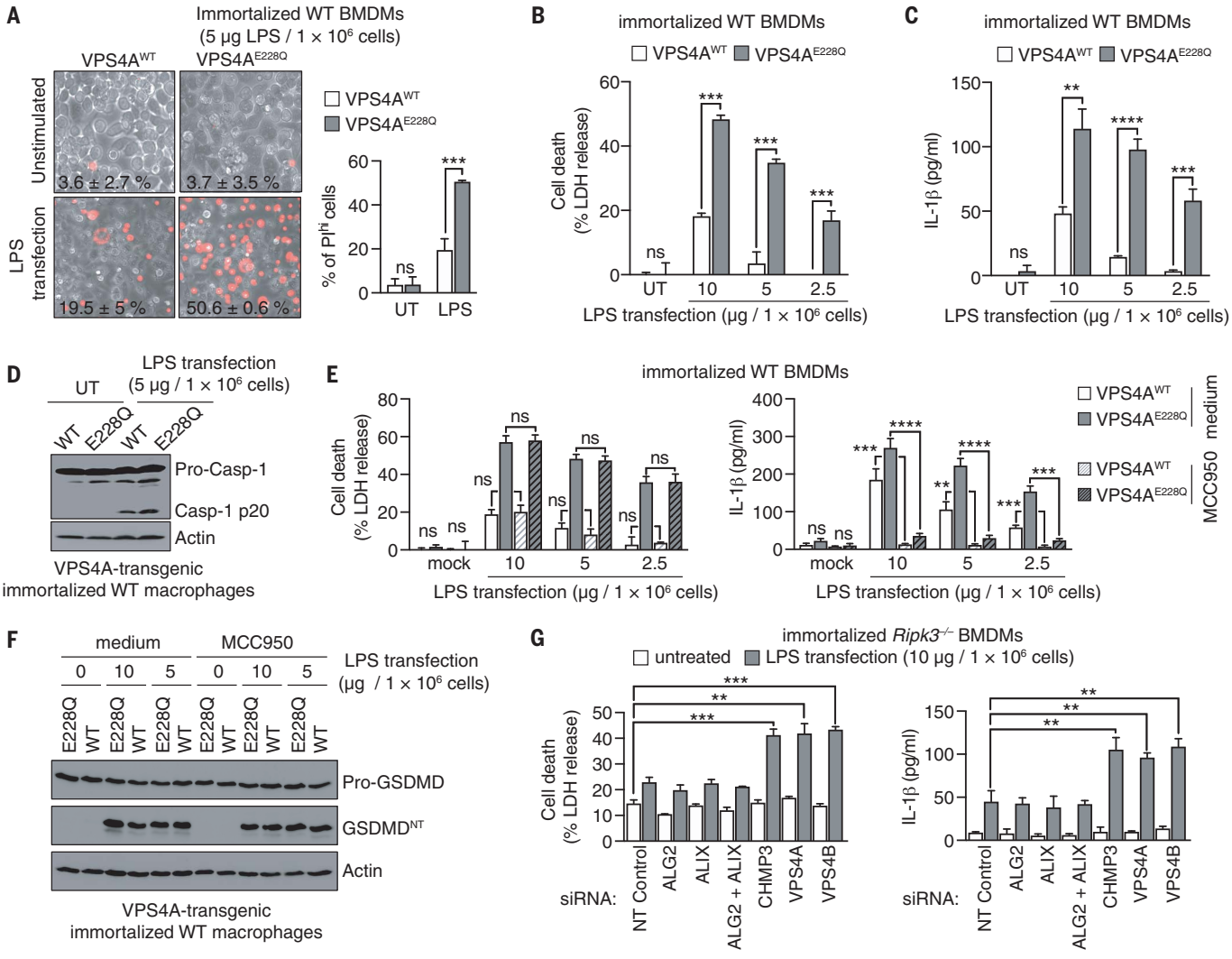


Fig. 3. ESCRT inactivation enhances noncanonical inflammasome-induced pyroptosis. (**A** to **D**) PI staining, LDH release, IL-1β release, and caspase-1 processing from VPS4A^{WT}-transgenic or VPS4A^{E228Q}-transgenic iBMDMs 2 hours after LPS transfection. UT, untreated. (**E** and **F**) LDH release, IL-1β release, and GSDMD processing from VPS4A^{WT}-transgenic or VPS4A^{E228Q}-transgenic iBMDMs 2 hours post-LPS transfection in the presence or absence of 2.5 μM MCC950. (**G**) LDH and IL-1β release 4 hours post-LPS transfection from *Ripk3*$^{-/-}$ iBMDMs treated with the corresponding small interfering RNA. NT, nontargeting siRNA. Graphs show means ± SD. $^{**}P < 0.01$; $^{***}P < 0.001$, $^{****}P < 0.0001$; ns, not significant (Student's *t* test). Results are representative of at least three independent experiments. VPS4A$^{WT/E228Q}$ expression in panels (A) to (F) was induced with 0.5 μg/ml of Dox for 6 hours.

formation. Consistently, and in agreement with previous reports (*15*), cytokine release but not cell death was attenuated by treatment with the NLRP3 inhibitor MCC950 or high extracellular K^+ (Fig. 3E and fig. S12A).

We next assayed GSDMD processing and found that $GSDMD^{NT}$ levels were slightly enhanced in $VPS4A^{E228Q}$-transgenic iBMDMs compared with $VPS4A^{WT}$-transgenic iBMDMs at high LPS concentrations (Fig. 3F). We postulated that the enhanced GSDMD processing was caused by enhanced NLRP3 inflammasome activation (Fig. 3E and fig. S12A) and was thus caspase-1 dependent. Indeed, GSDMD processing did not differ between $VPS4A^{E228Q}$- and $VPS4A^{WT}$-expressing iBMDMs in the presence of MCC950 (Fig. 3F and fig. S12B), confirming that ESCRT negatively regulates pyroptosis downstream of caspase-11 activation and GSDMD processing.

To study the role of ESCRT-associated proteins, we knocked down the expression of ESCRT-I and -III proteins and of two proteins, ALG2 and ALIX, that have been implicated in the Ca^{2+}-dependent recruitment of ESCRTs (*8*). As spontaneous necroptosis occurs upon ESCRT depletion (*12*), we performed all knockdown experiments in *Ripk3*$^{-/-}$ iBMDMs (*12*), which remained viable despite successful protein depletion (fig. S13, A to E). The knockdown of CHMP3, VPS4A, and VPS4B significantly enhanced pyroptosis and IL-1β release in *Ripk3*$^{-/-}$ iBMDMs transfected with LPS (Fig. 3G). By contrast, single or double knockdown of ALIX and ALG2 had no impact on pyroptosis (Fig. 3G). ALG2 and ALIX may act redundantly with TSG101 (*17*), but simultaneous depletion was not possible due to the cytotoxicity of TSG101 depletion in BMDMs (fig. S13F).

ESCRT inactivation also enhanced pyroptosis in human cells infected with Δ*sifA S. typhimurium* (fig. S14, A to D). In this experiment, ESCRTs appeared to have an additional function upstream of inflammasomes, as ESCRT inactivation resulted in higher levels of ruptured, galectin-3$^+$ *Salmonella*-containing vacuoles (fig. S15, A to C), despite equal levels of bacterial invasion (fig. S14, E and F). This implied that ESCRTs repair damaged vacuoles and endosomes, as suggested recently (*17*), and thereby control bacterial escape into the cytosol. However, the elevated levels of cytosolic bacteria in ESCRT-depleted cells showed only a negligible impact on GSDMD processing at the time points examined (fig. S15, D to F).

We next asked if ESCRTs also regulate cell death after canonical inflammasome activation. Ca^{2+} chelation significantly enhanced pyroptosis after DNA transfection of BMDMs (fig. S16, A and B), without changing caspase-1 processing (fig. S16, C and D). Furthermore, ESCRT depletion significantly enhanced cell death and IL-1β release in iBMDMs after NLRC4 activation with log-phase *S. typhimurium* bacteria but did not enhance ASC speck formation or processing of caspase-1 or GSDMD (Fig. 4, A and B, and fig. S17). This indicated that ESCRTs restrict pyroptosis downstream of GSDMD. To further exclude any effects of ESCRT depletion on the upstream signals that control caspase-1 activation, we used DmrB–Casp-1–transgenic HEK293T cells. Activation of DmrB–Casp-1 induced pyroptosis with faster kinetics in cells expressing dominant-negative $VPS4A^{E228Q}$ or $CHMP3^{1-179}$ than in cells expressing the WT proteins, particularly if DmrB–Casp-1 activity was limited by adding an inhibitory washout compound (Fig. 4C and fig. S18A). B/B homodimerizer enhanced pyroptosis in ESCRT-deficient cells compared with ESCRT-proficient cells across a variety of concentrations (Fig. 4D and fig. S18B). Notably, ESCRT inactivation did not affect GSDMD processing (Fig. 4E and fig. S18C) and enhanced pyroptosis only in DmrB–Casp-1–transgenic HEK293T cells when

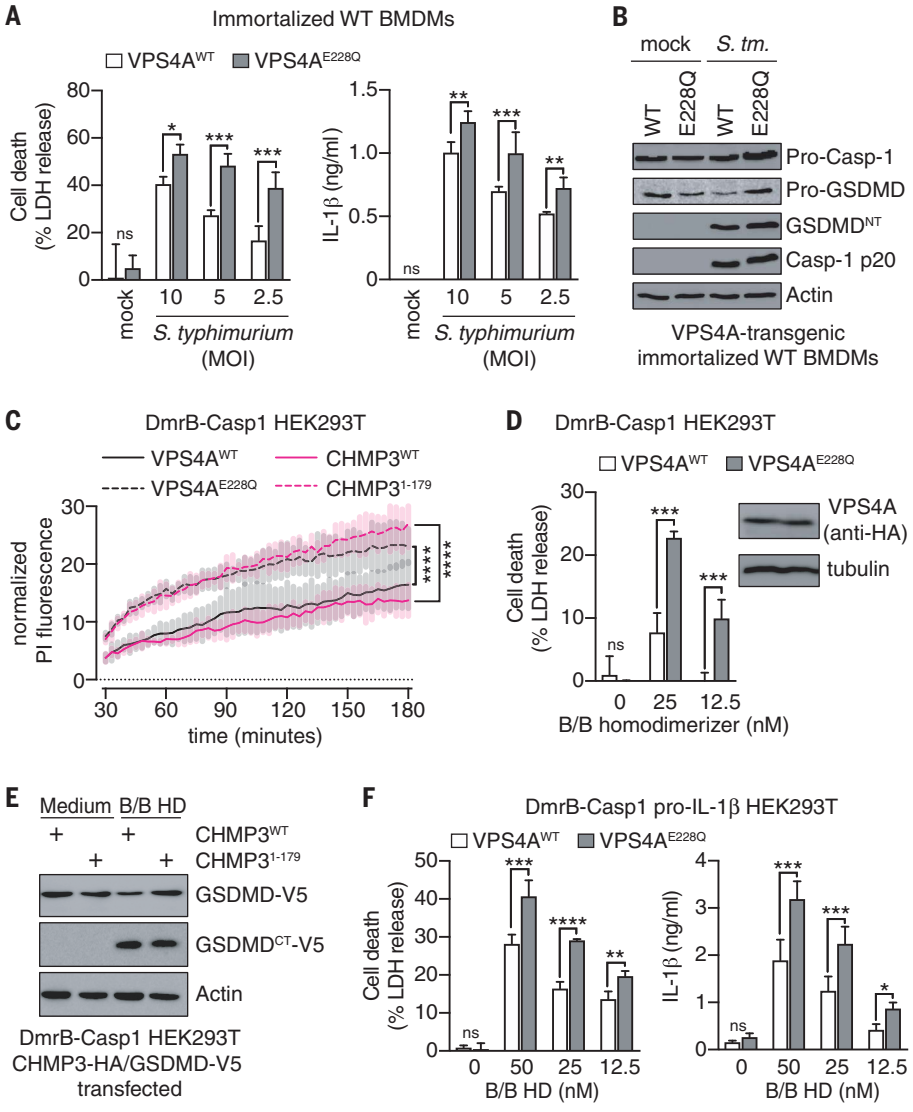


Fig. 4. ESCRT negatively regulates pyroptosis and cytokine release downstream of caspase-1. (**A** and **B**) LDH release, IL-1β release, and processing of caspase-1 and GSDMD from $VPS4A^{WT/E228Q}$-transgenic iBMDMs infected with log-phase *S. typhimurium* cells for 1 hour. (**C**) Time course of PI staining in DmrB–Casp-1–transgenic HEK293T cells expressing FLAG-hGSDMD-V5 and Dox-inducible VPS4A or CHMP3 constructs. Cells were treated with 25 nM B/B homodimerizer (B/B HD) at t = 0 and with a 50-fold excess of washout compound at t = 30 min. (**D**) LDH release at t = 120 min by DmrB–Casp-1–transgenic HEK293T cells treated as described for panel (C) except with different B/B HD concentrations. Immunoblots show equal expression levels of VPS4A proteins. (**E**) GSDMD processing at t = 60 min of DmrB–Casp-1–transgenic HEK293T cells treated as described for panel (C) except 50 nM B/B HD was used. (**F**) LDH and IL-1β release at t = 90 min from DmrB–Casp-1–transgenic and proIL-1β double-transgenic HEK293T cells expressing FLAG-hGSDMD-V5 and Dox-inducible VPS4A constructs. Cells were treated with B/B HD at t = 0 and with a 50-fold excess of washout compound at t = 30 min. Graphs show means ± SD. *P < 0.05; **P < 0.01; ***P < 0.001, ****P < 0.0001 (Student's t test) [(A), (D), and (F)] or a two-way analysis of variance (C). Results are representative of at least three independent experiments. Protein expression was induced for 6 hours [0.5 μg/ml of Dox for (A) and (B)] or 16 hours [1 μg/ml of Dox for (C) to (F)].

GSDMD was coexpressed (fig. S18, D and E). Finally, ESCRT inactivation also enhanced IL-1β release, as shown in pro–IL-1β–transgenic HEK293T cells (Fig. 4F and fig. S18F).

Thus, in a variety of cell lines and systems, the ESCRT machinery dampens GSDMD pore–mediated cell death and negatively regulates IL-1β secretion after canonical or noncanonical inflammasome activation. We propose that ESCRTs play a central role in removing GSDMD pores from the plasma membrane in the form of ectosomes and that membrane repair could allow cells to restrict pyroptosis while permitting limited GSDMD-dependent cytokine release (*18*, *19*). Our results place ESCRTs downstream of caspase activation and GSDMD processing, but they do not exclude the possibility that ESCRTs also control the activation of inflammasomes (e.g., by repairing damaged bacteria-containing vacuoles and thus restricting cytosolic access of bacteria-derived ligands). Further experiments are needed to assess this aspect of ESCRTs and to determine if ESCRT-induced vesicles transport cytokines and danger signals.

REFERENCES AND NOTES

1. N. Kayagaki *et al.*, *Nature* **526**, 666–671 (2015).
2. J. Shi *et al.*, *Nature* **526**, 660–665 (2015).
3. P. Broz, V. M. Dixit, *Nat. Rev. Immunol.* **16**, 407–420 (2016).
4. L. Sborgi *et al.*, *EMBO J.* **35**, 1766–1778 (2016).
5. R. A. Aglietti *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 7858–7863 (2016).
6. J. Ding *et al.*, *Nature* **535**, 111–116 (2016).
7. X. Liu *et al.*, *Nature* **535**, 153–158 (2016).
8. N. W. Andrews, P. E. Almeida, M. Corrotte, *Trends Cell Biol.* **24**, 734–742 (2014).
9. C. Tam *et al.*, *J. Cell Biol.* **189**, 1027–1038 (2010).
10. A. J. Jimenez *et al.*, *Science* **343**, 1247136 (2014).
11. L. L. Scheffer *et al.*, *Nat. Commun.* **5**, 5646 (2014).
12. Y. N. Gong *et al.*, *Cell* **169**, 286–300.e16 (2017).
13. J. Shi *et al.*, *Nature* **514**, 187–192 (2014).
14. I. Poser *et al.*, *Nat. Methods* **5**, 409–415 (2008).
15. N. Kayagaki *et al.*, *Nature* **479**, 117–121 (2011).
16. S. Rühl, P. Broz, *Eur. J. Immunol.* **45**, 2927–2936 (2015).
17. M. L. Skowyra, P. H. Schlesinger, T. V. Naismith, P. I. Hanson, *Science* **360**, eaar5078 (2018).
18. R. Heilig *et al.*, *Eur. J. Immunol.* **48**, 584–592 (2018).
19. C. L. Evavold *et al.*, *Immunity* **48**, 35–44.e6 (2018).

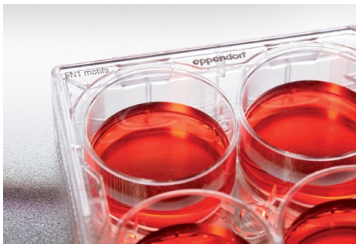
ACKNOWLEDGMENTS

We thank A. Hyman (MPI-CBG, Dresden, Germany) for CHMP4-GFP HeLa cells, A. Oberst (University of Washington, Seattle, WA, USA) for the DmrB–caspase-1 construct, J. Pieters (Biozentrum, Basel, Switzerland) for support and discussions, and C. Ramon-Barros for technical assistance. Microscopy analysis was performed at the Biozentrum and UNIL Imaging Core Facilities. **Funding:** This work was supported by grants from the Swiss National Science Foundation (310030_175576) and the ERC (ERC-2017-CoG –770988, InflamCellDeath) to P.B. and a Werner Siemens Fellowship (Fellowships for Excellence, University of Basel) to S.R. **Author contributions:** S.R. and P.B. designed the study. S.R., K.S., B.D., R.H., J.C.S., and P.B. performed experiments, analyzed data, and wrote the manuscript. **Competing interests:** We declare no competing interests. **Data and materials availability:** All data are available in the article or in the supplementary materials. Cell lines and plasmids are available from P.B. upon request.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6417/956/suppl/DC1
Materials and Methods
Figs. S1 to S18
Movies S1 to S8

14 December 2017; accepted 8 October 2018
10.1126/science.aar7607



Stem-Cell Cultureware

Eppendorf's CCCadvanced FN 1 motifs cell cultureware uses synthetic fibronectin (FN)-derived motifs to support cell attachment by mimicking native extracellular matrix (ECM) proteins. The major challenge of induced pluripotent stem cell (iPSC) cultivation is to preserve stem-cell properties until differentiation is induced. To support these properties, an appropriate culture medium and a defined ECM-mimicking growth surface/coating are crucial. Biological coatings commonly used to expand stem cells are inherently complex, nondefined growth surfaces that frequently reduce experimental reproducibility. Poor growth from these coatings is exacerbated by lot-to-lot variations between coating media; the use of undefined growth factors and ECM components known to sustain cell adhesion and pluripotency; and potential pathogen contamination risk during preparation and storage. Eppendorf's cultureware supports long-term expansion of human iPSCs (hiPSCs) over 25 passages, and is also suitable for human mesenchymal stem cells and other ECM-dependent cells. Furthermore, it allows expansion of stem cells in xeno-free and restrictive culture conditions.

Eppendorf
For info: 800-645-3050
www.eppendorf.com/us-en/seed-proceed-reproduce

3D-iPSC Matrix

MyEZGel 3D-iPSC, a powerful new tool for in vitro 3D human induced pluripotent stem cell (hiPSC) culture, is a synthetic, xeno-free, hydrogel matrix inspired by muscle and spider-silk proteins. It is proven to overcome basic problems associated with traditional hydrogel-based matrices, such as complicated scaffold synthesis, temperature/pH sensitivity, low harvest/yield, and cellular toxicity from materials used. 3D culture systems for iPSCs are gaining in popularity, as matrix-grown cells have more in vivo–like gene expression; better differentiation potential, proliferation, and cellular function; and are more physiologically relevant in vitro models. Provided as a patented nanofiber solution, a 3D microenvironment can be easily formulated to specific stem-cell growth conditions at room temperature or 37°C. By using MyEZGel, cells can be maintained in a pluripotent state or differentiated for advanced disease modeling and drug discovery research.

AMS Biotechnology
For info: 617-945-5033
www.amsbio.com/myezgel-3d-stem-cell-culture-matrix.aspx

Lab Safety Shields

Asynt Lab Safety Shields fit neatly around all major stirrer hotplate brands, providing lab scientists with enhanced safety while maintaining easy access to hotplate controls and good all-around visibility of the reaction or process. Built in robust 5-mm polycarbonate, the new Large Lab Safety Shield offers additional user protection for applications such as heated parallel reactions, especially those requiring user intervention when the fume hood sash is open. The device also shields reactions from the cooling effect of fume hood ventilation. Compactly designed (29.5 cm x 16 cm x 17 cm) to minimize use of valuable fume hood space, Asynt's budget-priced Standard Lab Safety Shield fits neatly around the stirrer hotplate to provide robust, convenient protection from splashes, aerosols, and spills. All Asynt safety shields are resistant to a wide range of laboratory solvents, and temperatures up to 170°C. They can also be optimized to your experimental setup.

Asynt
For info: +44-(0)-1638-781709
www.asynt.com/product/compact-lab-safety-shield-large

Extracellular Matrix Assay

Keratin Orange is an extracellular matrix assay designed to measure α-keratin in mammalian hair samples. α-keratin is a fibrous, structural protein found in all vertebrates in hair, wool, horns, nails, claws, and hooves as well as the stratum corneum of the skin. This simple colorimetric assay is based on the binding of the dye Orange G to α-keratin from digested hair samples. It complements Biocolor's other extracellular matrix assays: the Sircol assay for soluble and insoluble collagen, the Blyscan assay for sulfated glycosaminoglycans, the Fastin assay for elastin, and the Purple-Jelley assay for hyaluronan.

Biocolor Life Science Assays
For info: +44-(0)-2893-350-258
www.biocolor-assays.com

Shotgun Metagenomic and Strain Sequencing Tests

CoreBiome, a genomics company offering innovative solutions for microbiome discovery, offers USD 99 shotgun metagenomic and strain sequencing tests. These new services provide affordable, high-quality species and gene profiling for the human microbiome, and deep characterization of microbial isolate libraries. BoosterShot, an alternative to amplicon sequencing for large-scale microbiome experimentation, starts at a sequencing depth of 2 million sequences per sample, with a turnaround time of 3 weeks from sample arrival to delivery of a sophisticated, interactive online report, along with raw data. BoosterShot is recommended for human-stool and oral microbiomes, which are low in host-DNA contamination, and are well represented in whole-genome reference databases. The second service, StrainView, is for whole-genome shotgun sequencing of hundreds or thousands of isolated bacterial genomes simultaneously. CoreBiome combines its proprietary, efficient library-preparation protocols with optimal, alignment-based database searches, to deliver genome assemblies for single-isolate samples and visualizations of shared genome content across your strains.

CoreBiome
For info: 612-888-2466
www.corebiome.com

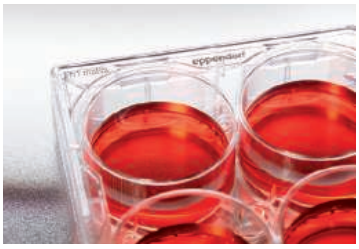
IAA Measurement System

The IAA Physiolyzer is the world's first complete research system for measuring endogenous and exogenous indole-3-acetic acid (IAA). The system uses the noninvasive microtest (NMT) technique to measure not just the concentration of IAA, but also the direction of transport (i.e., influx and efflux) from live samples of any size. IAA flux can be measured from concentrated organelles, single cells, cells, tissue, and small intact organisms, all of which are undamaged by testing. The flux data unfolds on-screen in real time with femtomole sensitivity, and can continue hands-off operation for hours, allowing researchers to see both short- and long-term IAA flux responses to genetic modifications, and to biotic and/or abiotic stimuli. The IAA Physiolyzer is perfect for labs researching development, physiology, signal transduction, and more. 3D flux measurement is also available in advanced versions.

YoungerUSA
For info: 413-992-2003
www.youngerusa.com/index.php/products/iaa-physiolyzer

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.



Stem-Cell Cultureware

Eppendorf's CCCadvanced FN 1 motifs cell cultureware uses synthetic fibronectin (FN)-derived motifs to support cell attachment by mimicking native extracellular matrix (ECM) proteins. The major challenge of induced pluripotent stem cell (iPSC) cultivation is to preserve stem-cell properties until differentiation is induced. To support these properties, an appropriate culture medium and a defined ECM-mimicking growth surface/coating are crucial. Biological coatings commonly used to expand stem cells are inherently complex, nondefined growth surfaces that frequently reduce experimental reproducibility. Poor growth from these coatings is exacerbated by lot-to-lot variations between coating media; the use of undefined growth factors and ECM components known to sustain cell adhesion and pluripotency; and potential pathogen contamination risk during preparation and storage. Eppendorf's cultureware supports long-term expansion of human iPSCs (hiPSCs) over 25 passages, and is also suitable for human mesenchymal stem cells and other ECM-dependent cells. Furthermore, it allows expansion of stem cells in xeno-free and restrictive culture conditions.

Eppendorf
For info: 800-645-3050
www.eppendorf.com/us-en/seed-proceed-reproduce

3D-iPSC Matrix

MyEZGel 3D-iPSC, a powerful new tool for in vitro 3D human induced pluripotent stem cell (hiPSC) culture, is a synthetic, xeno-free, hydrogel matrix inspired by muscle and spider-silk proteins. It is proven to overcome basic problems associated with traditional hydrogel-based matrices, such as complicated scaffold synthesis, temperature/pH sensitivity, low harvest/yield, and cellular toxicity from materials used. 3D culture systems for iPSCs are gaining in popularity, as matrix-grown cells have more in vivo–like gene expression; better differentiation potential, proliferation, and cellular function; and are more physiologically relevant in vitro models. Provided as a patented nanofiber solution, a 3D microenvironment can be easily formulated to specific stem-cell growth conditions at room temperature or 37°C. By using MyEZGel, cells can be maintained in a pluripotent state or differentiated for advanced disease modeling and drug discovery research.

AMS Biotechnology
For info: 617-945-5033
www.amsbio.com/myezgel-3d-stem-cell-culture-matrix.aspx

Lab Safety Shields

Asynt Lab Safety Shields fit neatly around all major stirrer hotplate brands, providing lab scientists with enhanced safety while maintaining easy access to hotplate controls and good all-around visibility of the reaction or process. Built in robust 5-mm polycarbonate, the new Large Lab Safety Shield offers additional user protection for applications such as heated parallel reactions, especially those requiring user intervention when the fume hood sash is open. The device also shields reactions from the cooling effect of fume hood ventilation. Compactly designed (29.5 cm x 16 cm x 17 cm) to minimize use of valuable fume hood space, Asynt's budget-priced Standard Lab Safety Shield fits neatly around the stirrer hotplate to provide robust, convenient protection from splashes, aerosols, and spills. All Asynt safety shields are resistant to a wide range of laboratory solvents, and temperatures up to 170°C. They can also be optimized to your experimental setup.

Asynt
For info: +44-(0)-1638-781709
www.asynt.com/product/compact-lab-safety-shield-large

Extracellular Matrix Assay

Keratin Orange is an extracellular matrix assay designed to measure α-keratin in mammalian hair samples. α-keratin is a fibrous, structural protein found in all vertebrates in hair, wool, horns, nails, claws, and hooves as well as the stratum corneum of the skin. This simple colorimetric assay is based on the binding of the dye Orange G to α-keratin from digested hair samples. It complements Biocolor's other extracellular matrix assays: the Sircol assay for soluble and insoluble collagen, the Blyscan assay for sulfated glycosaminoglycans, the Fastin assay for elastin, and the Purple-Jelley assay for hyaluronan.

Biocolor Life Science Assays
For info: +44-(0)-2893-350-258
www.biocolor-assays.com

Shotgun Metagenomic and Strain Sequencing Tests

CoreBiome, a genomics company offering innovative solutions for microbiome discovery, offers USD 99 shotgun metagenomic and strain sequencing tests. These new services provide affordable, high-quality species and gene profiling for the human microbiome, and deep characterization of microbial isolate libraries. BoosterShot, an alternative to amplicon sequencing for large-scale microbiome experimentation, starts at a sequencing depth of 2 million sequences per sample, with a turnaround time of 3 weeks from sample arrival to delivery of a sophisticated, interactive online report, along with raw data. BoosterShot is recommended for human-stool and oral microbiomes, which are low in host-DNA contamination, and are well represented in whole-genome reference databases. The second service, StrainView, is for whole-genome shotgun sequencing of hundreds or thousands of isolated bacterial genomes simultaneously. CoreBiome combines its proprietary, efficient library-preparation protocols with optimal, alignment-based database searches, to deliver genome assemblies for single-isolate samples and visualizations of shared genome content across your strains.

CoreBiome
For info: 612-888-2466
www.corebiome.com

IAA Measurement System

The IAA Physiolyzer is the world's first complete research system for measuring endogenous and exogenous indole-3-acetic acid (IAA). The system uses the noninvasive microtest (NMT) technique to measure not just the concentration of IAA, but also the direction of transport (i.e., influx and efflux) from live samples of any size. IAA flux can be measured from concentrated organelles, single cells, cells, tissue, and small intact organisms, all of which are undamaged by testing. The flux data unfolds on-screen in real time with femtomole sensitivity, and can continue hands-off operation for hours, allowing researchers to see both short- and long-term IAA flux responses to genetic modifications, and to biotic and/or abiotic stimuli. The IAA Physiolyzer is perfect for labs researching development, physiology, signal transduction, and more. 3D flux measurement is also available in advanced versions.

YoungerUSA
For info: 413-992-2003
www.youngerusa.com/index.php/products/iaa-physiolyzer

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.


RESEARCH ON THE EDGE
28.57050°N | 80.6490°W
KENNEDY SPACE CENTER, FL
ONE DAY WE'RE LAUNCHING SATELLITES. THE NEXT WE'RE LAUNCHING CAREERS.
At the University of New Hampshire, we find answers where you wouldn't expect. Like on more than 20 NASA satellites, which carry research from our Space Science Center as far as the edge of our solar system. And because students are essential research partners on these projects, we've managed to launch some stellar careers along the way.
Welcome to the edge of possible: where intellectual curiosity is rewarded with exciting research opportunities and challenging academic programs. Here, students thrive under the mentorship of Guggenheim and Carnegie fellows and winners of the Pulitzer and other honors. Whether it's blasting research into space or mapping the bottom of the ocean, we exist to help our students create possibilities for our world.
unh.edu/edge
University of New Hampshire

SCIENCE TRANSCENDING BOUNDARIES

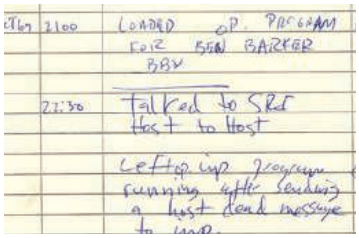
CELEBRATING SCIENTIFIC MILESTONES

In 1869,

Dmitri Mendeleev presented the Periodic Table, which documented the atomic weights and valence of the many elements that make up Earth, to the Russian Chemical Society.

One hundred years later, in 1969,

history was made by NASA with the first moon walk by Neil Armstrong, made possible by Apollo 11's successful landing on the moon.



On the ground in California that same year,

the first ARPANET message was sent from the University of California, Los Angeles to the Stanford Research Institute. This early network became the basis for today's Internet.

MEETING PROGRAM AND REGISTRATION ARE AVAILABLE ONLINE: AAAS.ORG/MEETINGS

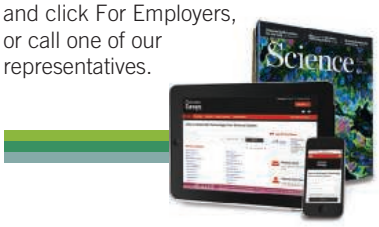
AAAS | ANNUAL MEETING
Washington, DC | Feb. 14–17, 2019

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Science Careers

SCIENCE CAREERS ADVERTISING

For full advertising details, go to ScienceCareers.org and click For Employers, or call one of our representatives.



AMERICAS

+1 202 326-6577
+1 202 326-6578
advertise@sciencecareers.org

EUROPE, INDIA, AUSTRALIA, NEW ZEALAND, REST OF WORLD

+44 (0) 1223 326527
advertise@sciencecareers.org

CHINA, KOREA, SINGAPORE, TAIWAN, THAILAND

+86 131 4114 0012
advertise@sciencecareers.org

JAPAN

+81 3-6459-4174
advertise@sciencecareers.org

CUSTOMER SERVICE

AMERICAS

+1 202 326-6577

REST OF WORLD

+44 (0) 1223 326528

advertise@sciencecareers.org

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. *Science* reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. *Science* encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCareers.org

UT Health San Antonio
Department of Molecular Medicine

The Department of Molecular Medicine in the Joe R. & Teresa Lozano Long School of Medicine at the UT Health San Antonio invites applications from outstanding scientists and physician-scientist in the areas of tumor microenvironment, oncogenic signaling, cell metabolism, integrative (synthetic) biology, genomics/epigenomics or any combination thereof. The successful candidate will have an appointment in the Department of Molecular Medicine and extremely qualified individuals may be considered to Co-lead the Cancer Development and Progression program in the Mays Cancer Center, UT Health San Antonio and MD Anderson Cancer Center. Competitive candidates will have an outstanding history of novel contributions to cancer research with a record of current and sustained extramural grant support. The candidate will be eligible for a generous start up package and open space in a new, modern, and fully equipped research facility conducive to multi discipline research. In addition, the successful candidate may have an appointment to a prestigious endowment from the department.

For full consideration, interested candidates should submit to http://uthscsa.edu/hr/employment.asp preferably by January 1 a cover letter, curriculum vitae, a summary of research accomplishments and outline of future plans, and contact information for at least three professional references.

Zelton Dave Sharp, Ph.D., Chair
Faculty Search Committee
UT Health San Antonio
Department of Molecular Medicine
Joe R. & Teresa Lozano Long School of Medicine

UT Health San Antonio is an Equal Employment Opportunity/Affirmative Action Employer including protected veterans and persons with disabilities. All faculty appointments are designated as security sensitive positions.

Department of Geosciences
PRINCETON UNIVERSITY

Assistant Professor

The Department of Geosciences at Princeton University is seeking applications for a tenure-track assistant professor faculty position in geology, broadly defined. We are particularly interested in interdisciplinary scientists who could interact productively with existing faculty working in geophysics and/or climate. Possible fields of specialty include, but are not limited to, petrology, volcanology, tectonics, glaciology, rock deformation, earth surface processes, and paleontology.

Applicants should send a curriculum vitae, including a publication list, a statement of research and teaching interests, and contact information for three references to **https://www.princeton.edu/acad-positions/position/9581**. Evaluation of applications will begin as they arrive; for fullest consideration, apply by **December 21, 2018**, but applications will be accepted until the position is filled.

Princeton is especially interested in candidates who can contribute to the diversity and excellence of our academic community. For general information about applying to Princeton and how to self-identify, please link to http://web.princeton.edu/sites/dof/ApplicantsInfo.htm.

Princeton University is an equal opportunity employer and all qualified applicants will receive consideration for employment without regard to race, color, religion, sex, sexual orientation, gender identity, national origin, disability status, protected veteran status, or any other characteristic protected by law.

This position is subject to the University's background check policy.

Information about the research activities of the Department of Geosciences may be viewed at http://geoweb.princeton.edu.

AUS | الجامعة الأميركية في الشارقة
American University of Sharjah

American University of Sharjah, one of the Middle East's leading universities, has set itself an ambitious research agenda, pushing the limits of science and engineering and providing exciting opportunities for those seeking to realize the university's research goals.

LEADING A NEW ERA IN RESEARCH

AUS is a preeminent and top-ranked university in the region, currently embarking on a massive new era in research, founding seven new research entities:

Biosciences and Bioengineering Research Institute (BBRI)
Materials Science and Engineering Research Institute (MSERI)
Gulf Environments Research Institute (GERI)
Smart Cities Research Institute (SCRI)
High-Performance Computing Center (HPCC)
Geospatial Analysis Center (GAC)
Genomics Research Laboratory (GRL)

To learn more about the AUS research agenda, and the opportunities available, visit: www.aus.edu/research

10 ways that *Science* Careers can help advance your career

1. Register for a free online account on **ScienceCareers.org**.
2. Search thousands of job postings and find your perfect job.
3. Sign up to receive e-mail alerts about job postings that match your criteria.
4. Upload your resume into our database and connect with employers.
5. Watch one of our many webinars on different career topics such as job searching, networking, and more.
6. Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
7. Complete an interactive, personalized career plan at "my IDP."
8. Visit our Career Forum and get advice from career experts and your peers.
9. Research graduate program information and find a program right for you.
10. Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

By **Camila H. Coelho**

Finding my inner Wonder Woman

It was an early blow to my self-confidence. I was attending my first group meeting in the lab where I had just started as a postdoc, and I was pleased that I had managed to follow most of the discussion. Then, in front of everyone, my supervisor turned to me and asked about my previous accomplishments. I froze. As a Ph.D. student, I had achieved lots to be proud of. But all of that was in my home country of Brazil. Now I was in Washington, D.C.—and I didn't know what the word "accomplishment" meant. When I looked it up later in my English-Portuguese dictionary, I realized what I should have said: presenting my research at international conferences, publishing, teaching, and more. But at the time, all I could utter in response was "I don't know."

I had decided to do a postdoc abroad because I thought the training would help me secure a faculty position. I applied for and received a Brazilian government fellowship to spend 18 months working abroad. A professor I had met at a conference agreed to give me a position in his lab. It all seemed so easy—until I actually started.

In a new lab and a new country, I struggled. After the mortifying lab meeting incident, my confidence took another hit when some of my experiments and analyses didn't work. The paper I intended to publish did not materialize. I used to be invited to give talks; now, I was asked to speak more slowly because of my accent. My confidence was shattered. Could I only be successful in my own country?

"Could I only be successful in my own country?"

After months of self-doubt, I reminded myself that I had potential. I needed to do something to regain my lost confidence.

I saw that many of the Latin American scientists I had connected with were thriving at large institutions that had more resources to support training and development. Although I liked my lab and had a good relationship with my supervisor, my institution didn't offer enough of these opportunities. I thought a change of environment might be what I needed.

I contacted a principal investigator at the National Institutes of Health (NIH). He said that, if my current supervisor approved, I could join his lab for a 6-month trial period, which would take me to the end of my fellowship. My original supervisor agreed that changing institutions would be best for me, and I made the move.

Once again in an unfamiliar environment, I still felt insecure. I was certain that my new supervisor didn't actually think I was good. I thought he was just being generous and giving me a chance to get some training. I doubted that he would renew my contract when it expired. Nonetheless, I was determined to make the most of the time I had.

I started to see the benefits of my new environment surprisingly quickly. There were more opportunities to interact with others—for example, in meetings, in the hallways, and in chats with administrative staff—which forced me to talk more. It was intimidating at first, but with practice I began to feel more confident in my English. The many training programs that NIH offers—covering topics such as leadership and writing as well as cutting-edge science—helped, too. I proposed new projects and collaborations, including one with my original supervisor. I began to receive positive feedback on my presentations. My accent is still the same, so I assume that this is due to my restored confidence.

I've now been in the lab for a little more than 2 years, and I'm so proud of how far I've come. I've presented my research at international conferences, sometimes winning travel awards to attend them. I review and edit manuscripts for journals, and I've published my own papers. My confidence is back. I feel a little bit like Wonder Woman. She was a strong warrior in her homeland. When she left, she experienced obstacles and failures—only to become even stronger than before. ■

Camila H. Coelho is a visiting postdoctoral fellow at the National Institute of Allergy and Infectious Diseases in Bethesda, Maryland. Send your career story to SciCareerEditor@aaas.org.