An unexpected ally against an invasive stinkbug *p. 542*

Targeting lacteal junctions to impede obesity *pp. 551 & 599*

Where's the evidence about stakeholder engagement? *p. 554*

Science

$15
10 AUGUST 2018
sciencemag.org

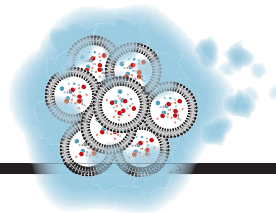
AAAS

MIGHTY WATERS

A global map of river area

pp. 546 & 585

CONTENTS



548 & 604
Condensing synaptic vesicles in neurotransmission

10 AUGUST 2018 • VOLUME 361 • ISSUE 6402

542
Pests in the making: young brown marmorated stinkbugs and eggs

NEWS

IN BRIEF

IN DEPTH

FEATURE

558

INSIGHTS

PERSPECTIVES

POLICY FORUM

BOOKS *ET AL.*

Published by AAAS

CONTENTS

552
A bigger picture through human phenotype studies

10 AUGUST 2018 • VOLUME 361 • ISSUE 6402

ON THE COVER

Iguazú Falls, Earth's largest waterfall system, on the Iguazú River between Argentina and Brazil. Earth's turbulent rivers and streams are natural hotspots for atmospheric exchange. Thousands of satellite images have revealed that rivers and streams cover nearly 45% more of the world's surface than previous spatial estimates, indicating that these water bodies contribute more to atmospheric concentrations of greenhouse gas than previously thought. See pages 546 and 585. *Photo: Inge Johnsson/Alamy Stock Photo*

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. *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

EDITORIAL

Health of the Hajj

For 5 days, starting 19 August, Saudi Arabia will host the Hajj, the world's largest annual religious pilgrimage, where people from over 180 countries will converge on Mecca. Infectious disease transmission associated with this mass transnational movement of people is well known—malaria in 632 CE, meningitis in 1987 and 2000, polio in 2004, and pandemic influenza in 2009. As the former Deputy Minister of Public Health for Saudi Arabia, I know how immense the challenge is to ensure that the country is prepared to contain the spread of infectious disease and maintain public well-being during this event. Although major progress has been made over the past 30 years in Saudi Arabia and in pilgrimage countries, there is still much more to do to avoid a health catastrophe, given that the pilgrim quota may grow to 2.2 million by 2020.

Nearly half of Hajj pilgrims are 56 years of age or older, and about half have preexisting health conditions. Many pilgrims originate from countries with low income (66%) or that are involved in conflict (18%), with suboptimal health care, disease surveillance, or prevention education. These factors allow outbreaks to go undetected before reaching epidemic thresholds, and Hajj pilgrims can become unsuspecting disease carriers. The intense and exhausting nature of the rituals, extreme desert temperatures, proximity between pilgrims during congregation and prayers, and commitment to share facilities all collectively create an ideal environment for infectious disease transmission.

This year's Hajj will be held in the shadows of recently documented outbreaks of cholera, Ebola virus disease, polio, Middle East respiratory syndrome, measles, meningitis, diphtheria, Lassa hemorrhagic fever, yellow fever, hepatitis, and Nipah virus infection across the countries of pilgrim origin. Hepatitis virus infections and tuberculosis are endemic in many of the pilgrims' home countries. Alkhurma virus is endemic in Saudi Arabia. Potential for zoonotic disease transmission exists through the animal sacrifice rituals, including of imported livestock. Yet only three of these conditions require mandatory vaccination as an entry requirement for participation in the Hajj (meningococcal meningitis, polio, and yellow fever).

In 2009, when the H1N1 influenza pandemic coincided with the Hajj, Saudi Arabia convened stakeholders to create a discipline called "mass gathering medicine." Saudi Arabia strengthened real-time Hajj surveillance and set up a reporting system and emergency operations center to monitor pilgrim safety and infectious disease events. However, if diseases of high transmission potential and case fatality such as Nipah virus or Ebola virus should emerge at the Hajj, Saudi Arabia's response capacities may not be sufficient to stall their impact.

"This year's Hajj will be held in the shadows of recently documented outbreaks..."

Surveillance at entry and exit points could quickly improve disease detection. Likewise, travel services could ensure compliance with international health regulations at all departure and arrival ports. The Hajj pilgrim venues and home country preparatory venues could actually become one-stop points for multiple disease surveillance and education for prevention. It will also be important to create a stockpile of medicines and vaccines in the event of a disease outbreak. Such an endeavor will require investment by the World Health Organization and a multinational strategy that is coordinated by the Saudi Ministry of Health.

Performing the Hajj is a deeply emotional wish of all Muslims, and many pilgrims save their lifelong earnings to participate. Thus, restrictions on participating in Hajj to reduce the risk of disease spread is a geopolitical challenge for governments. Vision 2030 is a national development initiative of Saudi Arabia that aims to support the expected growth in the international Hajj pilgrimage, which may reach 4.5 million by 2030. The initiative should involve pilgrim countries and international stakeholders to develop public health safeguards for mass gatherings, which for many, may be a once-in-a-lifetime event.

–Ziad A. Memish

Ziad A. Memish
is a professor at the College of Medicine, Alfaisal University & Prince Mohammed Bin Abdulaziz Hospital, Ministry of Health, Riyadh, Kingdom of Saudi Arabia and a professor in the Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA. memishz@pmah.med.sa

PHOTOS: (INSET) BANDAR AL-DANDANI/AFP/GETTY IMAGES; (TOP RIGHT) ZIAD MEMISH

10.1126/science.aau9617

NEWS

IN BRIEF

Edited by **Jeffrey Brainard**

CLIMATE SCIENCE

Sizzling summer is linked to climate warming

July may not have been the hottest on record—that distinction goes to July 2016, according to the European Union's Copernicus Climate Change Service. For many people suffering the effects, however, this summer's raging forest fires and blistering heat waves stand out as exemplars of the impacts that human-driven warming is having worldwide. Across the western United States, thousands of wildfires have burned 2 million hectares since January, far above the average for the previous decade; in California, the Mendocino Complex fire this week became the largest in state history. Although warming is far from the only reason for these conflagrations in the western United States, which are also fueled by a legacy of fire suppression, scientists believe that in the past 3 decades human-induced warming has dried out more fuel, doubling the amount of forest burned compared with that from natural climate variability alone. Meanwhile, an unrelenting heat wave has gripped Europe, especially in the north; human-driven warming doubled the odds of it happening, according to a rapid study released in late July by scientists at the University of Oxford in the United Kingdom.

Ebola emerges in conflict zone

PUBLIC HEALTH | Just days after officials in the Democratic Republic of the Congo (DRC) celebrated the defeat of an Ebola outbreak, a new one began last week in a different part of the country, an armed conflict zone where health workers will have difficulty working safely. As of 6 August, 16 confirmed and 27 probable cases of the disease were reported in six health districts in North Kivu province, and 34 people had died. The DRC's health ministry says there's no evidence they are linked to the previous outbreak in Équateur province, on the opposite site of the country, which the ministry declared over on 24 July, after 54 cases and 33 deaths. The DRC has now had 10 outbreaks since Ebola's virus was first discovered there in 1976, more than any other country. North Kivu province has seen fighting between government forces and rebels, and health teams may have to travel with armed escorts. "On the scale of degree of difficulty, trying to extinguish an outbreak of a deadly, high-threat pathogen in a war zone reaches the top of any of our scales," said Peter Salama, head of the World Health Organization's Health Emergencies Programme in Geneva, Switzerland. An experimental Ebola vaccine that was used during the Équateur outbreak will likely be used again during the new one.

Physicist faulted for groping

WORKPLACE | Arizona State University (ASU) in Tempe concluded last week that Lawrence Krauss, an outspoken physics popularizer, grabbed a woman's breast at a 2016 conference, violating ASU's sexual harassment policy. "Responsive action is being taken to prevent any further recurrence of similar conduct," ASU

PHOTO: MARK RALSTON/AFP/GETTY IMAGES

Firefighters use a back burn to battle the Carr fire near Redding, California, which has been fueled by dry conditions.

Executive Vice President and Provost Mark Searle wrote in a 31 July letter to Melanie Thomson, a microbiologist based in Ocean Grove, Australia, who provided the letter and the accompanying investigative report to *Science*. In 2017, Thomson filed a complaint about the incident, which she witnessed during a convention in Melbourne, Australia. The university's probe found that Krauss grabbed the unnamed woman's breast after she took a selfie with him. He denied this to ASU investigators, but they found the preponderance of the evidence favored the accounts of the woman and multiple witnesses. An ASU spokesperson said Krauss remains on paid administrative leave, which ASU instituted on 6 March. In July, the university declined to renew his position as director of ASU's Origins Project, which explores the origins of the universe, life, disease, and social systems. In an email, Krauss wrote, "The process at ASU is ongoing, therefore I believe it would be inappropriate to comment at this time."

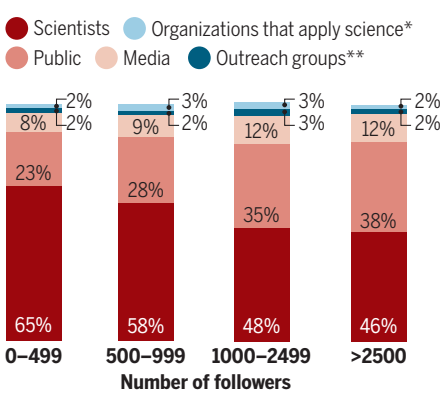
NSF wants to pick your brain

RESEARCH PRIZES | Got an idea for research that could transform the world? The National Science Foundation (NSF) in Alexandria, Virginia, is all ears. The $7.8 billion agency is giving the public a chance to win glory—and some money— in its first-ever online contest, dubbed the NSF 2026 Idea Machine. "We don't want [the idea] to be something NSF is already doing," says Suzi Iacono, head of NSF's Office of Integrative Activities. "We want it to be exciting, and original, and important." Anyone can enter, from individual scientists to professional societies to a high school science class. And the grander the grand challenge, the better. "We don't want single projects," Iacono says, "but rather big umbrella themes, with lots of community engagement and involving all units at NSF." The contest website is scheduled to go live on 31 August and will be up for 2 months; winners will be announced next summer.

Tweets can aid science outreach

SCIENCE COMMUNICATION | Some scientists start Twitter accounts to share their work and findings with audiences outside of their labs and field stations. But does this form of public outreach succeed? To find out, two scholars examined the Twitter followers of 110 academic ecologists and evolutionary biologists who tweet mostly about science, rather than personal matters. These researchers, mostly in the United States and the United Kingdom, had varying degrees of success reaching nonscientists. Those with fewer than 1000 followers attracted scientists as the large majority of their audience, the study found. But among researchers with bigger followings, members of the public and the media made up most of their followers, Isabelle Côté of Simon Fraser University in Burnaby, Canada, and Emily Darling of the University of Toronto in Canada report. Nonscientists amplified a scientist's reach because their Twitter accounts typically

Who follows scientists on Twitter



*Nonprofit conservation groups, management agencies, etc.
**Museums, zoos, science educators
Note: Percentages are rounded.

CREDITS: (GRAPHIC) N. DESAI/*SCIENCE*; (DATA) I. CÔTÉ AND E. DARLING

THREE QS

Here, have some mosquitoes

To combat the spread of dengue fever, caused by a mosquito-borne virus, researchers recently persuaded some 7000 households in the northeastern Australian city of Townsville to accept an unusual request: Would they host a tub of mosquito eggs in their yards? The mosquitoes were infected with the bacterium *Wolbachia*, which reduces their ability to harbor certain viruses and spread them to humans. The sprawling effort, covering 66 square kilometers, wasn't designed to test whether the mosquitoes prevented dengue outbreaks—although they appeared to have that effect. It was instead a trial of how to deploy these winged disease fighters, which spread *Wolbachia* when they mate with wild mosquitoes, on a city-wide scale. The program's director, medical entomologist Scott O'Neill of Monash University in Melbourne, Australia, told *Science* about the project.

Q: What was unique about this effort?
A: Firstly, the size of it: Most other releases have been done on the scale of 1 or 2 square kilometers. And here, for most of the land area covered ... the community actually deployed the mosquitoes on its own behalf.

Q: Do people ever have a problem with putting mosquitoes in their backyard?
A:: If the community is feeling like there's too many mosquitoes around, then we would usually pull back or stop. But there was no organized opposition to the project in Townsville. If you're living in a place where there's regular, annual transmission of a disease like dengue, and there's nothing actually working to prevent it, then people can be quite fearful of it.

Q: Is it different trying to sell suburban Australians on this technology from people in poor countries?
A: When we talk to the communities—whether it's Townsville or a subsistence fishing village in central Vietnam—the top two concerns are nearly always "Is it safe for me and my children if the mosquitoes bite us?" and "Could there be some environmental impact that's not really understood by releasing the mosquito?" I think that we probably put much more emphasis on community engagement than any of the other teams I've seen working with other technologies.

had larger followings than scientist-run accounts, the pair notes. They say their research supports the view that building audiences on Twitter requires persistence and effective strategies. Their study appeared online on 28 June in *Facets*.

Canada to unmuzzle scientists

SCIENCE POLICY | Canada unveiled on 30 July the country's first science integrity policy, intended to reinforce the right of government scientists to speak about their work to the media. The guidelines address scientists' complaints that they were muzzled under the previous Conservative government. Under the national policy, scientists can speak publicly without prior approval by their agencies, although they may be required to notify supervisors. It also lays out a mix of other protections and expectations for government scientists, including avoiding conflicts of interest and making scientific findings transparent. Any department or agency that employs more than 10 scientists will be required to develop its own version of the policy by the end of the year and provide annual reports. The government's chief science adviser developed the policy in cooperation with the Ottawa-based Professional Institute of the Public Service of Canada, a union that represents federal scientists. It "represents the culmination of over 3 years' work by federal scientists in promoting scientific research and evidence-based policy within government," said Debi Daviau, the union's president.

BIOANTHROPOLOGY

Stonehenge remains hail from Wales

Archaeologists have identified the origin of mysterious cremated remains found under Stonehenge, near Amesbury, U.K.: Some of these people probably came from western Wales. Previous research established that several of the ancient monument's huge stones came from quarries in that area, suggesting people from Wales helped build Stonehenge. The remains were interred during the earliest stages of Stonehenge's construction, from 3000 to 2480 B.C.E., according to earlier research. Because their bodies had been cremated, DNA analysis was not possible, scientists wrote last week in *Scientific Reports*. Instead, they analyzed an isotope of the element strontium in the bones. Strontium varies depending on local geography and so can reveal where the deceased had lived in the decade or so before death. Bones usually blur such analyses by absorbing strontium from the soil, but in this case the high temperatures of cremation modified the bone, effectively sealing in the strontium as it was at death.

Chemical analysis yielded new clues about the origins of Stonehenge's builders.

Four medals awarded, one stolen

SCIENTIFIC PRIZES | The Fields Medal, often called the Nobel Prize of mathematics, was awarded last week to four prominent young mathematicians: Caucher Birkar of the University of Cambridge in the United Kingdom, Alessio Figalli of ETZ Zurich in Switzerland, Peter Scholze of the University of Bonn in Germany, and Akshay Venkatesh of the Institute for Advanced Study in Princeton, New Jersey. The International Mathematical Union, which gives out the awards every 4 years to mathematicians age 40 or younger, recognized these winners for groundbreaking work in categorizing equations, optimal transport theory, arithmetic geometry, numbers theory, and other mathematical subfields. Minutes after Birkar received his medal, he noticed that his briefcase, where he had placed it, was missing. The briefcase was later recovered outside the ceremony venue in Rio de Janeiro, Brazil, but the medal was gone.

Review calls for Mars sequel

PLANETARY SCIENCE | NASA needs to prepare for future explorations of Mars beyond an upcoming mission to collect rock samples, the National Academies of Sciences, Engineering, and Medicine writes in a report this week. It provides a midpoint assessment of the agency's performance on planetary science projects since the academies' 2011 decadal survey, which recommended priorities for 2013 to 2022. In 2020, NASA plans to address one priority by launching a $2.4 billion rover to drill samples from Mars and store the rocks there; a separate mission would return the samples to Earth. The report warns, though, "There is currently no vision for a program beyond sample return, either for scientific investigation or to prepare for future human exploration." Otherwise the report largely endorses NASA's efforts, although it notes the agency seems likely to fall short of meeting the recommended cadence for its competitively awarded Discovery and New Frontiers mission programs; NASA would need to pick three additional Discovery missions and one more New Frontiers to meet the decadal report's recommendation.

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

PHOTO: RICHARD T. NOWITZ/GETTY IMAGES

Published by AAAS

Drought records in stalagmites from Mawmluh Cave in India's Meghalaya state are ambiguous.

IN DEPTH

EARTH SCIENCE

New geological age comes under fire

Timing and extent of ancient drought used to define the Meghalayan are uncertain

By **Paul Voosen**

Last month, the International Commission on Stratigraphy (ICS), the bureaucracy that governs geological time, declared we are living in a new geological age. No, it's not the Anthropocene, the much-debated proposal for a geological division defined by human impact on Earth. The new age anointed by ICS is called the Meghalayan, based on signs in the rock record of a global drought that began about 4200 years ago. It is one of three newly named subdivisions of the Holocene, the geological epoch that began 11,700 years ago with the retreat of ice age glaciers. And the name will now filter its way into textbooks.

Many scientists say, however, that the "4.2-kiloyear event" was neither a global drought nor fixed to that moment in time. "The whole idea of defining the subdivision of the Holocene with a break at 4.2 seems a bit baseless," says Raymond Bradley, a climatologist at the University of Massachusetts in Amherst. Jessica Tierney, a paleoclimatologist at the University of Arizona in Tucson, says ICS, following the lead of some paleoclimate scientists, mistakenly lumped together evidence of other droughts and wet periods, sometimes centuries away from the 4200-year-old event, to mark the beginning of the Meghalayan. This is a "paleoclimate white whale," she says.

The first clues to the Meghalayan came from archaeology. In the early 1990s, Harvey Weiss, an archaeologist at Yale University, was excavating a compelling story of drought-induced collapse in Mesopotamia. At Tell Leilan, an ancient city of the Akkadian Empire in northeastern Syria, he found evidence that drought pushed people out of the city 4200 years ago. The signal repeated across much of Mesopotamia.

It was a good starting point for Mike Walker, a geologist at the University of Wales in Lampeter who a decade ago began the effort to divide the Holocene. Scientists commonly talk about an early, middle, and late Holocene—tracking the glaciers' retreat and partial return—but with wildly different time spans in mind. ICS asked Walker to standardize those divisions for the sake of clear scientific communication. But although abrupt changes in the rock record mark earlier geological divisions, such changes are scarce in the relatively calm Holocene.

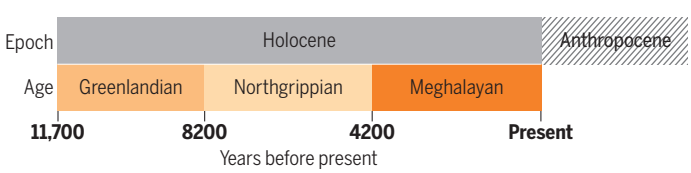
Finding a date to divide the early and middle Holocene, now dubbed the Greenlandian and Northgrippian, was easy, Walker says. About 8200 years ago, an outburst of freshwater from naturally dammed glacial lakes poured into the North Atlantic Ocean. The floods are believed to have disrupted a conveyor belt of ocean currents, leading to signals of global cooling that can be reliably found in ice cores, lakebeds, and cave rocks.

But the second division proved harder. "We were struggling," Walker says. Fortunately, some paleoclimate scientists were picking up where Weiss's archaeological work left off. Signs of a 4200-year-old drought were emerging in the Mediterranean, the Americas, and Asia, where researchers linked it to a weakened monsoon. In 2012, paleoclimatologists reported an analysis of a stalagmite from Mawmluh Cave, a limestone complex in Meghalaya state, a wet part of northeastern India. Stalagmites, calcium carbonate formations on the floors of caves, grow drip by drip as mineral-rich rainwater seeps in. In the Mawmluh stalagmite, a shift in oxygen isotopes seemed to show a stark drying around 4200 years ago, a clear signal of centuries-long drought. Walker thought the stalagmite signal could serve as the perfect geological exemplar, or "golden spike," marking the beginning of the Meghalayan. In June, a few dozen geologists from ICS and its parent body ratified Walker's proposal for the new ages with little dissent.

Paleoclimatologist Ashish Sinha is surprised that ICS used

Rocks of ages

Geologists have divided the Holocene into three ages. One, the Meghalayan, is based on controversial evidence. A new epoch, the Anthropocene, is still under debate.



CREDITS: (PHOTO) ABHIJEET KHEDGIKAR/SHUTTERSTOCK.COM; (GRAPHIC) N. DESAI/*SCIENCE*

the stalagmite for its golden spike—and few know it better, as it was his lab at California State University in Dominguez Hills that found and analyzed it. His team could date only a few of the stalagmite's layers, and water had partially dissolved the rock close to the drying event, potentially blurring the record. An unpublished analysis of other Meghalayan stalagmites by paleoclimate scientists from Xi'an Jiaotong University in China adds to the doubts: It found a steady weakening of the monsoon over more than 600 years, rather than a sudden drought 4200 years ago. The closest thing to a sharp drought can be seen 4000 years ago, in a few decades-long events. These excursions could be said to match the golden spike "to an extent," says Gayatri Kathayat, who led the research, "but not entirely."

Elsewhere in the world, the 4200-year-old event is even less apparent, according to a team at Northern Arizona University (NAU) in Flagstaff. Over the past few years, the NAU team has amassed 550 published paleoclimate records of temperature and moisture change during the Holocene, based mainly on stalagmites, lake sediments, and ice cores. Graduate student Hannah Kolus scrutinized the records in vain for significant changes in global temperature or moisture about 4200 years ago. "You don't see that signal at all," Kolus says.

The archaeological evidence is also far from definitive, adds Mark Altaweel, an archaeologist at University College London. He says political collapse, not drought, may have doomed some settlements in Mesopotamia. And in ancient Egypt, Greece, and elsewhere the evidence of a global drought is even murkier, adds Guy Middleton, an archaeologist at Charles University in Prague. "Nothing happened as suddenly or as synchronously as made out." The drought makes no sense as a marker, he says. "It is new mythmaking."

Walker wishes Kathayat's new stalagmite records had been published in time for their proposal. But he thinks that, though scattered in time and space, the signs of drought are good enough to define a new division that geologists can use to clarify their discussions of the Holocene. "The fact that this is extremely damned close is encouraging for us," Walker adds. For Bradley, it shows the stark division between ICS, which studies Earth's deep history, and scientists who study the recent past. "[We're] on totally different pages, really totally different books," Bradley says.

Critics of the Meghalayan will have plenty of time to bolster their arguments because for now, debate is over. To prevent continual spats, ICS freezes discussion for a decade after it ratifies a boundary. "It gives time for new ideas to bed down," Walker says. ■

RESEARCH FUNDING

March of Dimes curtails support for researchers

Cuts at historic nonprofit leave grantees fuming

By **Kelly Servick**

March of Dimes, the 80-year-old nonprofit organization that helped develop the polio vaccine and has funded pioneering studies of birth defects, is abruptly scaling back some research grants amid financial struggles. Scientists have been caught by surprise.

In recent weeks, the group has told 37 of the 42 recipients of its individual investigator awards that it is ending their grants, which average $300,000 over 3 years. It plans to maintain the remaining awards at reduced levels; all five are focused on understanding and preventing premature birth. The group, based in White Plains, New York, is also trimming grants to its prematurity research centers, which are housed at academic institutions around the United States. It will not award any new grants this year, but still plans to give out eight to 10 of its 2-year, $150,000 awards for young scientists in 2019.

> ***"The lack of transparency has been nothing short of appalling."***
> **Andrew Holland**, Johns Hopkins University

March of Dimes has trimmed about $3 million from its annual research budget of roughly $20 million this year, says Kelle Moley, the group's chief scientific officer. She attributes the group's struggles to declining donations, particularly from the organization's signature March for Babies. "The walks were our main funding source … and now there's a million different kinds of walks," Moley says. "They're just not getting the donations that we used to get 10 or 20 years ago." Tax filings show that the group's expenses have exceeded revenues each year from 2012 to 2016. It announced last year that it would sell its headquarters in White Plains.

The cuts also reflect a strategic shift to focus research spending on preterm birth, which Moley calls "the biggest threat right now facing newborn babies."

Researchers are lamenting the loss of a key funding source for an often overlooked area of research. "These sort of basic research grants in developmental biology are hard to come by," says developmental biologist Maria Jasin of Memorial Sloan Kettering Cancer Center in New York City, who had a $250,000 award from March of Dimes to study a protein that influences DNA rearrangement during sperm and egg formation. "It's really a shame that there will now be this hole."

Jasin and other grantees say they were blindsided by the cuts, and now are scrambling to find ways to support graduate students and activities funded by the grants. "The way they've approached this has been completely inhumane," says molecular biologist Andrew Holland of Johns Hopkins University School of Medicine in Baltimore, Maryland, who had a 3-year, $250,000 grant from March of Dimes to study the role of a genetic pathway in microcephaly. A late July email from the group informed him that he would not be receiving the remaining $160,000 on the grant. "The lack of transparency has been nothing short of appalling," he says.

Chromosome biologist Andreas Hochwagen of New York University in New York City, who had a March of Dimes grant to probe chromosomal mishaps during the creation of sperm and egg cells, was startled to learn in late July that the group would pay only for project expenses incurred through June. "For the whole month of July … they didn't tell me that I wasn't being supported anymore," he says. "That I find a little outrageous."

Moley acknowledged that the cuts have created "a little bit of a gap" for some researchers. "I understand this is difficult for them, and I know the March of Dimes has been very generous in the past," she says. She also says the organization has been seeking more corporate donors. "I'm hoping that as we go forward, and as our plan is working … we can go back to funding a more broad focus."

March of Dimes has changed course before. It was founded in 1938 by then-President Franklin Roosevelt as the National Foundation for Infantile Paralysis and funded pioneering efforts to develop a polio vaccine. When such vaccines became widely available, the group shifted its focus to birth defects. ■

Lake Victoria is home to hundreds of cichlid species, diverse in both appearance and behavior.

EVOLUTION

Hybrids spawned Lake Victoria's rich fish diversity

Cichlid speciation helped by pre-existing genetic variation

By **Elizabeth Pennisi**, *in Waimea, Hawaii*

In the shallow waters of Lake Victoria, the world's largest tropical lake, swim some 500 species of cichlid fish with a dizzying variety of appearances, habitats, and behaviors. Genomic studies have shown they arose from a few ancestral species in just 15,000 years, a pace that has left researchers baffled about how so much genetic variation could have evolved so quickly. Now, extensive sequencing of cichlids from around Lake Victoria suggests much of it was there at the start, in the cichlids' ancestors. Ancient and more recent dallying between cichlid species from multiple watersheds apparently led to genetically diverse hybrids that could quickly adapt to life in the lake's many niches.

Reported last week at the Origins of Adaptive Radiation meeting here, the work is "a tour de force, with many lines of evidence," says Marguerite Butler, a functional morphologist at the University of Hawaii in Honolulu. It joins other research suggesting that hybridization is a powerful force in evolution (*Science*, 18 November 2016, p. 817). "What hybridization is doing is allowing the good stuff to be packed together," Butler says.

Some of Lake Victoria's cichlids nibble plants; others feed on invertebrates; big ones feast on other fish; lake bottom lovers consume detritus. Species vary in length from a few centimeters to about 30 centimeters; come in an array of shapes, colors, and patterns; and dwell in different parts of the lake. Mutations don't usually happen fast enough to produce such variety so quickly. "It's been really hard to figure out what's going on," says Rosemary Gillespie, an evolutionary biologist at the University of California, Berkeley.

Ole Seehausen, an evolutionary biologist at the University of Bern who has studied cichlids for more than 25 years, wondered whether hybridization could have generated the genetic raw material. In earlier research, his team collected cichlids from the rivers and lakes surrounding Lake Victoria and partly sequenced each species's DNA to build a family tree. Its branching pattern indicated that Lake Victoria's cichlids are closely related to a species from the Congo River and one from the Upper Nile River watershed, the group reported last year in *Nature Communications*.

A close look at all their genomes suggested the two river species hybridized with each other long ago. Seehausen proposed that during a warm spell about 130,000 years ago, water from tributaries of the Malagarasi River, itself a tributary of the Congo, temporarily flowed into Lake Victoria, bringing Congo fish into contact with Upper Nile fish.

To explore the cichlids' genetic history in more detail, Seehausen and postdocs Matt McGee, Joana Meier, and David Marques have now sequenced 450 whole cichlid genomes, representing many varieties of 150 species from the area's lakes, and from the Congo, Upper Nile, and other nearby watersheds. Clues in the genomes suggest multiple episodes of mixing took place. Periods of drying have repeatedly caused Lake Victoria to disappear, and Seehausen and his team propose that fish in the remaining waterways evolved independently until wetter periods reunited them. This "fission-fusion-fission" process restored genetic diversity each time.

About 15,000 years ago, three groups of fish, themselves products of the ancient hybridizations, came together in Lake Victoria as it filled again. Their ancestry provided the "standing variation" that natural selection could pick from to help the fish adapt to a vast range of niches, producing the cichlid bounty seen today. "Hybridization may turn out to be the most powerful engine of new species and new adaptations," Seehausen says.

"It's mind-blowing," says Dolph Schluter, an evolutionary biologist at The University of British Columbia in Vancouver, Canada. "All the variation required for speciation is already there" in the hybrids.

Studies of other species also suggest standing variation can speed evolution. Biologists trying to understand how marine sticklebacks adapted so quickly to living in freshwater have discovered that a crucial gene variant was already present—in low percentages—in the fishes' marine ancestors. At the meeting, researchers offered similar stories of standing variation jump-starting diversification, for example enabling long-winged beetles to evolve into short-winged ones on the Galápagos Islands. "I've never seen so many talks where you have evidence that genes are borrowed from old variation and further evolution is somehow facilitated by that," Schluter says.

Andrew Hendry, an evolutionary biologist at McGill University in Montreal, Canada, cautions colleagues not to completely dismiss new mutations in rapid species diversification: "What's not clear to me is whether [the role of ancient hybridization] is a general phenomenon," he says.

Regardless, "The implications for conservation are blaring," says Oliver Ryder, who heads conservation genetics efforts at the San Diego Zoo in California. Endangered species are currently managed as reproductively isolated units, and conservationists are reluctant to bolster populations by breeding the threatened animal with related species or populations. Eight years ago, however, a controversial program that mated Florida panthers with Texas cougars helped rescue the former from extinction. Studies such as Seehausen's, says Ryder, suggest that in the long run, hybridization is important to preserving a species's evolutionary potential. ■

PHOTOS (CLOCKWISE FROM LEFT): TIM ALEXANDER (2); MORITZ MUSCHICK

'Ōhi'a are tough enough to thrive on new lava but can be downed by invasive fungi.

CONSERVATION

Hope blooms for Hawaii's iconic native tree

Some 'ōhi'a trees resist foreign fungi—and fencing out animals can also protect the trees from fatal infections

By **Elizabeth Pennisi**

Hawaii's red-blossomed 'ōhi'a is tough enough to colonize recent lava flows, but until this summer the iconic native tree seemed doomed. Four years ago, an invasive fungus began to kill 'ōhi'a (*Metrosideros polymorpha*) on the island of Hawaii; by now, the blight has spread across 800 square kilometers. The news got worse in May, when dying trees tested positive for the fungus on the neighboring island of Kauai, fueling fears that rapid 'ōhi'a death (ROD) would span the state.

But the picture brightened at a meeting on Oahu late last month. Aerial surveys and studies on land and in the lab now suggest that some 'ōhi'a will survive. The killer fungus turns out to be two distantly related species, one of them less deadly to 'ōhi'a, and some trees seem to have a native resistance to both strains. Management practices such as fencing out animals also appear to slow the spread of the fungus. "We are not going to see an extinction of 'ōhi'a," predicts Flint Hughes of the U.S. Forest Service's Pacific Southwest Research Station in Hilo, who is coordinating ROD research. "As we understand it more, our management tools are improving and we are learning about the potential weak points of the fungus and the strengths of 'ōhi'a."

At the annual Hawaiian Conservation Conference in Honolulu from 24 to 26 July, Lisa Keith, a research plant pathologist with the U.S. Department of Agriculture's Agricultural Research Service in Hilo, described key differences between the two 'ōhi'a killers: *Ceratocystis lukuohia*, which slays trees in weeks by clogging their circulatory systems, and *C. huliohia*, which causes canker sores and seems less deadly. Related fungi cause rot in sweet potatoes and pineapples, and affect cacao and mango trees, among other plants. Keith's genetic studies determined that the 'ōhi'a's deadlier foe comes from Latin America, whereas the less aggressive fungus, which is infecting trees on Kauai, may have been imported from Asia or Australia.

Some trees are also more resilient, Keith and her colleagues found when they infected each of Hawaii island's five varieties of 'ōhi'a with the two fungi. "There's definitely variability" in how the seedlings respond, she says, with some still alive after 1.5 years, even after being infected with the deadlier species.

Aerial drone surveys confirm the lab findings. For the past several years, Ryan Perroy and Timo Sullivan from the University of Hawaii (UH) in Hilo have flown cameras above the treetops in four 40-hectare sites on the island. The progress of the disease is easy to track in the drone images: Affected trees turn dark red, then silvery gray as they die. At two of the sites, Sullivan reported at the meeting, ROD's spread has greatly slowed, and the number of infected trees has hovered between 30% and 40%, compared with nearly 100% in the initial outbreak.

Larger trees seem most susceptible, possibly because they are more injury-prone and provide a greater target for windborne sawdust carrying the pathogens. In addition, Sullivan says, "Fences appear to make a difference," with more trees surviving in enclosed areas. Because the fungi seem to infect only wounded trees—such as those on which animals have rubbed, munched, or rooted around the base—those behind fences are less vulnerable, he suggests.

Gregory Asner, an ecologist with the Carnegie Institution for Science in Palo Alto, California, has seen the same pattern in aircraft surveys of Hawaii island. He detects early signs of the blight with an onboard spectrometer sensitive to changes in 'ōhi'a leaf water, sugars, and defensive chemical content. By overlaying those data on land-use maps, Asner says he, too, has turned up "some really strong evidence that feral animals are spreading the fungus," again by wounding trees and making them more susceptible to infection. Domestic livestock can also do damage. Studies on the ground provide further support: Hughes's team set up 200 0.1-hectare plots across the island of Hawaii, fenced and unfenced, and found the protected trees were healthier. "Knowing that fences can have a very positive impact is very exciting," he says.

Ironically, another Hawaiian icon—the Kilauea volcano—may aid the 'ōhi'a. The fungus has done the most damage in the Puna district of the island of Hawaii. That area happens to be where this spring lava began spewing out of a giant fissure, destroying hundreds of homes. "It's absolutely terrible," says Rebecca Ostertag, a tropical forest ecologist at UH. "But there is a small sliver of a silver lining." In many areas devastated by ROD, foresters are worried 'ōhi'a won't be able to return because fast growing invasive trees, such as strawberry guava, will crowd them out. But that won't happen on Puna's fresh lava fields. 'Ōhi'a is always the first—and often, the only—tree to get a foothold in such landscapes. So, Hughes says, "Those new lava flows will give 'ōhi'a a chance to recover." ■

PHOTO: ISTOCK.COM/PHOTO75

CELL BIOLOGY

Infrared method could safely identify cells

Researchers use powerful synchrotron radiation to profile cell types

By **Mitch Leslie**

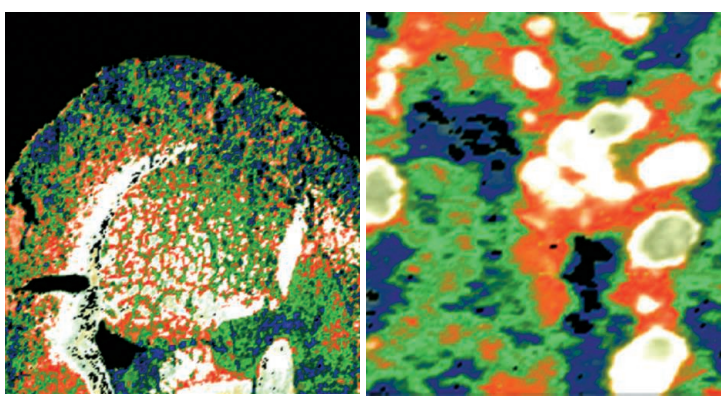
To identify a cell, researchers often have to abuse it—rip it from its home, douse it with toxic fixatives, doctor its DNA, or coerce it into making exotic proteins that could upset its biochemistry. Even if the cell survives, it may never be the same again. But a strong yet gentle beam of light could one day allow researchers to classify cells while leaving them unharmed and alive for additional study.

A team led by biophysicist Cynthia McMurray and physicist Michael Martin of the Lawrence Berkeley National Laboratory (LBNL) in California has found that by scanning cells with an intense beam of infrared radiation produced by a synchrotron, a type of particle accelerator, they can capture a biochemical signature that reveals cells' identities.

The researchers presented early results from the method in June at a meeting in the United Kingdom, and they are now evaluating it with a 1-year pilot grant from the Chan Zuckerberg Initiative (CZI). If it works, the team's spectral phenotyping technique could provide a tool for another endeavor backed by CZI: the Human Cell Atlas, an international project that aims to chart the type and location of every cell in the body. And if the synchrotron-driven method can be adapted to more modest infrared instruments available to other labs and hospitals, spectral phenotyping might one day also help diagnose illnesses, probe the cellular changes that lead to disease, and delve into embryonic development. "The tools we are putting together will blow open this field," McMurray predicts.

Scientists who are familiar with the still unpublished results call the approach promising. "I'm looking forward to seeing the research that's going to come out," says spectroscopist Peter Gardner of The University of Manchester in the United Kingdom. Chemical physicist Hugh Byrne of the Dublin Institute of Technology is impressed by how thoroughly the group is testing its approach. "It's a concerted program to demonstrate the capabilities of the technique."

Martin and McMurray like to contrast their approach with a widely used cell-identification technique: fluorescent labeling. To spur cells of a specific type to produce a label such as green fluorescent protein (GFP), scientists have to equip them with the molecule's gene. The techniques for adding DNA can alter the cells, and because GFP is foreign to them—it's originally from a jellyfish—it could also modify their physiology. Moreover, McMurray notes, researchers typically have to zap fluorescent labels with a laser to induce them to light up, which can harm or kill cells. Other techniques are no less invasive. "If you are doing labeling or staining, you are changing the true chemistry" of cells, Martin says. "We want to explore what the chemistry is, not alter it to do the measurements."



In tissue from a mouse's brain, seen at lower (left) and higher magnification (right), infrared spectroscopy can distinguish astrocytes (red) from neurons (green).

IMAGES: ARIS POLYZOS & LILA LOVERGNE

That's where infrared spectroscopy comes in. "Infrared is not invasive, so it can be used on intact tissues and living cells," McMurray says. When a sample is exposed to different wavelengths of infrared radiation, how much light of each wavelength it absorbs indicates the kinds of chemical groups it contains. Unlike fluorescent labeling, the absorption pattern usually can't reveal whether a cell is producing a specific molecule—for example, the immune receptors CD4 or CD8, which are often used to define two classes of T cells. But a cell's infrared spectrum does reveal broad types of molecules—such as fats and proteins—providing a biochemical fingerprint. As a result, "You get a much more holistic picture of the cell," Byrne says.

Martin and McMurray say standard infrared sources don't provide the sensitivity they needed, so the team turned to LBNL's Advanced Light Source synchrotron, whose infrared beam is one of the brightest in the world. It "allows us to get better resolution and fidelity," Martin says. At the June SPEC2018 conference in Glasgow, U.K., McMurray and Martin revealed they could discriminate two types of brain cells—neurons and astrocytes—in slices of brain from mice. In brain tissue from rodents with a condition mimicking Huntington disease, they could also detect an increase in lipids that indicates degeneration. In the future, the researchers plan to automate cell identification by enlisting machine learning algorithms to pick out distinguishing features of each cell type.

McMurray and her colleagues still need to determine whether a cell's spectral signature remains constant or varies with its location in the body. For potential medical uses, they also want to find out whether a human cell's infrared signature changes when a person becomes ill. So far, however, the researchers have analyzed only mouse tissues. "We wanted to make sure the method is robust," McMurray says.

One limitation of the new technique is obvious—synchrotrons are huge, expensive, and rare, and often have months-long waiting lists. "You aren't going to be taking your synchrotron into the hospital," Gardner says. But lab machines are rapidly approaching the infrared-generating power of particle accelerators, he notes. McMurray adds that after using the synchrotron to pinpoint distinctive spectral patterns for a variety of cell types, the researchers plan to publish a catalog that would allow other scientists to compare results from their own samples, even ones captured with less discerning lab devices.

Gardner expects the project to have an impact. "They have the tools, the expertise, and the personnel to accelerate this work," he says. ■

FEATURES

CONTROL FREAKS

Importing an exotic species for pest control takes years of preparation. What happens when it arrives on its own?

By **Kelly Servick**, *in Bridgeton, New Jersey*

The samurai wasp (*Trissolcus japonicus*) arrived by accident in the United States before scientists were ready to release it.

Published by AAAS

In a peach orchard down a rural road here, an uninvited guest has run amok. The brown marmorated stinkbug (*Halyomorpha halys*) has been gorging on the unripe fruit. The bugs have sunk their needle-sharp stylets into the peaches, creating wounds that ooze a clear, sugary goo; form corky brown blemishes; and leave the trees more vulnerable to infection.

In this orchard, managed by the Rutgers University Agricultural Research and Extension Center, the mottled, shield-shaped stinkbug is a research subject. In surrounding farms and homes, however, it's a despised invasive pest known for its indiscriminate appetite, its tendency to escape cold weather by crowding into homes—sometimes by the thousands—and the pungent, cilantrolike odor it releases when crushed. (Exterminators often recommend that homeowners vacuum up the insects instead.) Native to Asia, the bug was first spotted in the United States in 1998; it has since reached 43 states and Washington, D.C., attacking fruit trees, corn, soybeans, berries, tomatoes, and other crops. Statistics are scarce, but an industry group estimates that Mid-Atlantic apple growers alone lost $37 million to stinkbug damage in 2010.

In the peach orchard, however, another surprise invader also is on the march—and it may prove to be the stinkbug's nemesis.

Like many invasive species, the brown marmorated stinkbug has no major enemies in its new home to keep its population in check. So in 2005, entomologist Kim Hoelmer and his team at the U.S. Department of Agriculture's (USDA's) Agricultural Research Service (ARS) in Newark, Delaware, turned to a strategy known as classical biological control: They traveled to Asia to find natural enemies of the stinkbug that they might release in the United States.

Fanning out to agricultural fields and botanical gardens, the team searched for the bug's tiny clusters of barrel-shaped eggs. They checked whether any had been invaded by parasitoid wasps, which inject their own eggs into the stinkbug's, leaving larvae that eat the developing bugs before chewing their way out. By far the most pervasive parasite they found was the samurai wasp (*Trissolcus japonicus*), which, despite its fearsome name, is stingerless and smaller than a sesame seed. The ARS team imported several strains of the wasp to a quarantined facility in Newark and began painstaking tests to decide whether it was a good biocontrol candidate.

Then in 2014, Hoelmer got an unexpected phone call. Elijah Talamas, a taxonomist at the Florida Department of Agriculture and Consumer Services in Gainesville, had been helping another ARS team identify native wasps parasitizing stinkbug eggs in Maryland. Talamas, an expert on *Trissolcus* species, had recognized that some were samurai wasps.

"It was stunning news," Hoelmer recalls. He had spent years studying the wasp in the lab to make sure that, if released, it would do its job without harming native species. But the insect was already here. Genetic tests confirmed that the wasps in Maryland hadn't escaped from any of his quarantined strains. Somehow, they had immigrated on their own.

Over the decades, a variety of uninvited biocontrol candidates have popped up on new continents, including a fungus that kills forest-stripping gypsy moths and a beetle that devours allergy-inducing ragweed. "The examples definitely are piling up," says Donald Weber, an ARS entomologist in College Park, Maryland, whose team found the first U.S. samurai wasps. "We've had this mindset that natural enemies would be less likely to establish" than invasive pests, he says. But sometimes, "It might be fairly easy."

Those unexpected arrivals can unsettle scientists and regulators. Rules aimed at carefully controlling insect releases can seem nonsensical when the species in question is already happily spreading on its own. And the arrival of the samurai wasp—a promising biocontrol agent against a high-profile pest, with a formal proposal for release already in the works—has prompted a fresh look at some U.S. regulations.

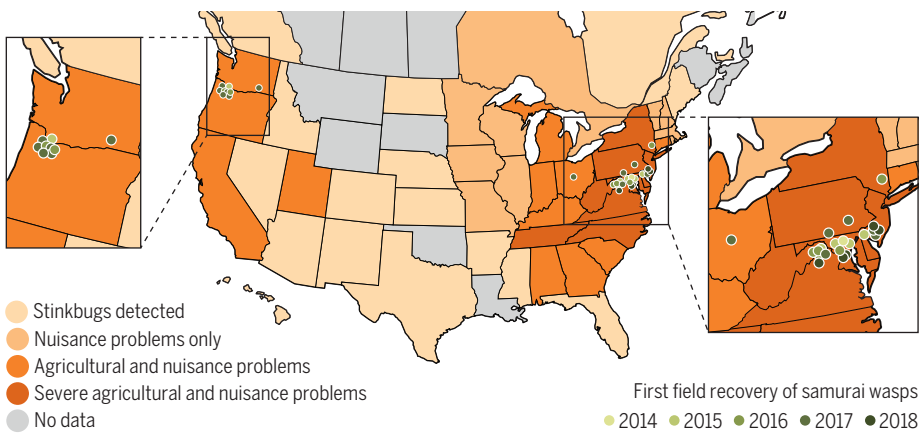
But unplanned introductions also free researchers from some of the usual constraints, allowing them to explore key questions about a biocontrol agent's impact in field experiments rather than just the lab. The team at the peach orchard, for example, is one of about a dozen U.S. groups now releasing the samurai wasp into fields and orchards to see whether it will be an ally in fighting the exotic stinkbugs—or yet another problematic invader.

CLASSICAL BIOLOGICAL CONTROL has logged some undeniable successes, such the release of the South American wasp *Apoanagyrus lopezi* in Africa in the 1980s to control the cassava mealybug (*Phenacoccus manihoti*). That project preserved a staple crop and saved an estimated 20 million lives, earning its architect, Swiss entomologist Hans Rudolf Herren, the 1995 World Food Prize. But many of the best-known biocontrol efforts are the historical disasters: the mongooses unleashed for rat control in Hawaii in 1883 that devastated native birds and turtles, and the cane toads sent to Australia in 1935 that failed to control sugarcane-destroying beetles but—because the toads themselves are poisonous—killed native reptiles, frogs, birds, and mammals that ate the toads.

As the field matured, many nations began to strictly regulate the release of biocontrol agents—which can include insects, fungi, and bacteria—and required studies to predict potential "nontarget effects." As Weber puts it, "People are a lot more responsible now than when they were running around releasing mongooses." In the United States, researchers must submit a proposal to USDA's Animal and Plant Health Inspection Service (APHIS). That proposal must include data from lab experiments gauging whether their candidate is likely to eat or parasitize species other than the targeted pest. Three groups then vet the evidence of its safety: a scientific review panel with representatives from

Ancient enemies reunited

Since the samurai wasp's first U.S. appearance in 2014, surveys have turned up at least three genetically distinct populations in areas affected by the brown marmorated stinkbug.



CREDITS: (MAP) STOP BROWN MARMORATED STINK BUG. ADAPTED BY N. DESAI/*SCIENCE*; (DATA) KIM HOELMER AND JIM WALGENBACH; (PHOTO, OPPOSITE PAGE) ELIJAH TALAMAS

Canada and Mexico, an APHIS official, and sometimes the U.S. Fish and Wildlife Service. The process can take years.

But organisms have a way of sidestepping bureaucracy. Over and over, potentially beneficial species have popped up uninvited, likely reaching new continents by the same shadowy routes of international trade and travel that spread pests. Entomologist Paul DeBach of the University of California, Riverside, in a 1971 essay, called this phenomenon fortuitous biological control. And it can be a boon. Recently, the North American leaf beetle (*Ophraella communa*), which feeds on the invasive ragweed *Ambrosia artemisiifolia*, sneaked into Europe, potentially saving many people from pollen-induced misery.

The brown marmorated stinkbug is a versatile pest that attacks valuable crops and plagues homeowners.

And when an exotic fungus—*Entomophaga maimaiga*, which kills the gypsy moth in its caterpillar stage—started to spread around New England in the late 1980s, "it was very exciting," says entomologist Ann Hajek of Cornell University. Her team had tried without success to establish the fungus, but once it arrived, she spent years making sure it wasn't harming local caterpillar species. "We were lucky," she says: The natives were mostly unaffected.

Unplanned arrivals can bring heartache, however. Entomologist Tim Haye of the Centre for Agriculture and Bioscience International in Delémont, Switzerland, and his collaborators spent a decade developing plans to release the European parasitoid wasp *Trichomalus perfectus* to control the invasive weevil *Ceutorhynchus obstrictus*, a canola plant pest, in Canadian prairies. Then in 2009, the wasp appeared on its own, in Quebec in Canada. "I was very disappointed," Haye says.

Moreover, such invaders don't always prove fortuitous. A team led by ARS entomologist Keith Hopper investigated the parasitoid wasp *Aphelinus certus* as a biocontrol agent against the soybean aphid (*Aphis glycines*), but lab tests revealed it had a broad range of aphid hosts, including some native ones. "I would never have brought that thing in," says ecologist George Heimpel of the University of Minnesota in St. Paul, a collaborator on that project. But in 2005, the wasp showed up in Pennsylvania. Now that it's established, Heimpel's group is studying whether native aphid populations are in danger. "A lot of people don't care about native aphids," he says. "I am one of those that do."

Yet such unplanned introductions also hand researchers an opportunity, creating what Haye calls a gigantic field trial for testing predictions about the control agent's effects. For the samurai wasp, researchers can now get permission from state regulators to launch experiments that involve breeding and releasing accidentally introduced wasp strains in the wild—work that wouldn't be allowed with intentionally imported insects, which must stay quarantined until federal regulators are convinced they're safe.

ONE NIGHT LAST MAY, 3600 samurai wasps streamed from mesh cages into the stinkbug-infested New Jersey orchard. A team led by Rutgers entomologist Anne Nielsen and entomologist Kevin Rice of the University of Missouri in Columbia had strung up yellow sticky cards baited with stinkbug egg clusters among the peaches and along the adjacent forest edge. They planned to wait a few days, collect the cards, and count the wasps to see whether they had ventured into the orchard to pursue the peach-destroying bugs.

The insects were descendants of wasps that Nielsen first discovered in a nearby New Jersey orchard in 2017—the first find in a U.S. agricultural crop. Since 2014, the samurai wasp has turned up in 10 states and Washington, D.C. Researchers have identified at least three genetically distinct strains, suggesting multiple introductions. Maybe wasp embryos were hiding in stinkbug eggs on plants aboard a cargo ship. An adult wasp may even have hitched a ride with an unsuspecting airline passenger. (While awaiting a flight from New York City to Russia, Talamas once watched a different parasitoid wasp species, native to the United States, land on a page of his book. "All it had to do was fly down the walkway … next stop: Russia." He trapped the hitchhiker in his contact lens case and, on arrival, preserved it in vodka.)

Now that the wasps are in the United States, research questions abound, Nielsen says. In their native range, they parasitize up to 90% of brown marmorated eggs. But will their behavior be different here? Where will they congregate and forage? Will they dramatically reduce stinkbug populations? Could farmers support the wasp by adjusting their practices—for example, not spraying pesticides where the insects are most concentrated? The chance to probe basic questions about a little-studied exotic species, Rice says, is "fabulously exciting."

For U.S. regulators, however, the wasp's unexpected arrival poses an conundrum. "This is a good chance for us to codify policy and decide, 'How are we going to handle these circumstances?'" says Robert Pfannenstiel, an APHIS entomologist in Riverdale, Maryland, who reviews release applications. "Will we allow changes from our policies and processes that are already in place, or not?"

Studies so far suggest the samurai wasp is a promising biocontrol agent. Although in laboratory tests it has parasitized some eggs laid by native species, it has shown a strong preference for brown marmorated stinkbug eggs. Scientists can release the accidental strains in states where they've already been discovered, but for now they can't spread the wasps indiscriminately. APHIS prohibits moving exotic species that haven't been formally cleared for release into new states.

PHOTO: MATT ROURKE/AP PHOTO

A samurai wasp emerges from the brown marmorated stinkbug egg in which it hatched. Scientists hope the wasp will reduce populations of the pest in U.S. crops.

PHOTO: CHRIS HEDSTROM

(Nielsen and Rice, for example, couldn't legally perform their same experiment if they drove 4 hours north to Connecticut, where samurai wasps haven't been found—so far.)

The wasp needs to go through regulatory review, just like any other candidate, Pfannenstiel says. "The danger, in one case, of saying, 'Oh, we can tell it's not a risk,' and then releasing it [is that] there's pressure to do that repeatedly, and start making judgment calls rather than determinations" based on data. Field studies of the accidental strains could speed the evaluation and help the wasp's chances of approval—or reveal new reasons not to release it, he says. "I go into these evaluations with no preconceptions."

Hoelmer's team at ARS is still preparing a petition to APHIS to release one deliberately imported strain, which he hopes could serve as a backup if the accidentally introduced strains spread slowly. He also intends to include the accidentally introduced strains in his release petition, because their biology is so similar to that of the strain he has studied extensively. He plans to submit his petition by the end of this year, and hopes for a decision next year. For now, he says, growers and researchers in states where the wasp hasn't been detected will simply "have to wait until it crosses the border."

The caution can seem excessive, given the wasp's steady spread. "Should we really wait?" asks Haye, who also believes the wasp is a good biocontrol agent. "Or should we speed up the spread and prevent damage—and help farmers—knowing that it's not 100% specific and that eradicating it is not an option anyways?"

Those are just some of the questions likely to come up at a special symposium on the implications of accidentally introduced biocontrol candidates, set for the annual meeting of the Entomological Society of America in Vancouver, Canada, in November. Weber, co-organizing the symposium, hopes that regulators weighing the benefits and risks of a biocontrol proposal will factor in the potential for surprises: "The longer the regulatory process, the more likely the species will accidentally show up," he says. "There might be a fairy tale world where you can regulate everything that comes in," but in reality, "that horse leaves the barn on a regular basis."

THE RESEARCHERS AT WORK in the Rutgers orchard aren't yet endorsing the samurai wasp as a biocontrol agent. First, they'd like more evidence that it won't harm native species. "They're invasive," Rice says. "They're not in a different bucket" from the stinkbugs.

The restrictions on spreading the wasp "can be frustrating, and it can seem arbitrary, but the regulations are there for a reason," Nielsen adds. Still, she says, the wasp "is likely our best hope of controlling the brown marmorated." Examining the sticky cards this summer, her team found a roughly equal distribution of wasps in the peaches and the nearby woods. That finding suggests the wasps are perfectly happy foraging for stinkbug eggs among the fruit, which bodes well for the wasps' ability to control the pests. The team plans to run a similar experiment soon to see how the wasps spread into another crop, soybeans.

Meanwhile, researchers in California have sent Talamas another surprise: a new, accidentally introduced *Trissolcus* parasitoid wasp, this one native to India and Pakistan, which emerged from the egg of another exotic stinkbug pest, *Bagrada hilaris*. "I think that these introductions are happening constantly," Talamas says, but come to light only when taxonomists bother to take a close look. He'll publish the new finding later this month in the *Journal of Hymenoptera Research*.

Such arrivals are "humbling," Weber says—a reminder of the limits that humans face in shaping their environment. "We have less control over things than we think." ■

INSIGHTS

The Slave River flows from a confluence of rivers in northeastern Alberta, Canada, to the Northwest Territories.

PERSPECTIVES

GEOLOGY

Measuring Earth's rivers

Satellite images enable global tally of freshwater ecosystems and resources

By **Margaret Palmer**[1] *and* **Albert Ruhi**[2]

A recent communication to corporate leaders warned that "Water is connected to every global risk we face" (*1*). For the past 7 years, water crises have ranked among the top five global risks according to the World Economic Forum (*2*). The reason for this concern is that water is linked to environmental, societal, and economic risks—from biodiversity loss and ecosystem collapse to food crises, and from failures of urban planning to energy price shocks. Mitigating water-related risks to humans and ecosystems requires reliable accounting of freshwater resources. An essential first step is knowing where to find Earth's fresh water. On page 585 of this issue, Allen and Pavelsky (*3*) make an important step in this direction by improving estimates of the global surface area of rivers using satellite observations.

Most fresh water in the biosphere exists as polar ice, glaciers, permanent snow, and groundwater; liquid surface waters only represent ~0.3% of total fresh waters. This small fraction is disproportionally concentrated in lakes and wetlands, which hold around 50 times as much water as rivers. Nevertheless, rivers and streams make major contributions to regional biodiversity. These contributions stem from the high spatiotemporal variability in habitats and flows that characterize running waters (see the second photo) and from the tight connections between rivers and adjacent terrestrial and marine habitats, which allow for rich, intertwined food webs.

Furthermore, site-level measurements have shown that running-water systems are also larger sources of methane and carbon dioxide emissions than previously thought (*4*). This observation challenges the long-held assumption that rivers are largely conduits of organic matter that link lands to oceans. Thus, in addition to their role in supporting biodiverse assemblages, running waters may be important sites of carbon storage and critical components of global biogeochemical cycles.

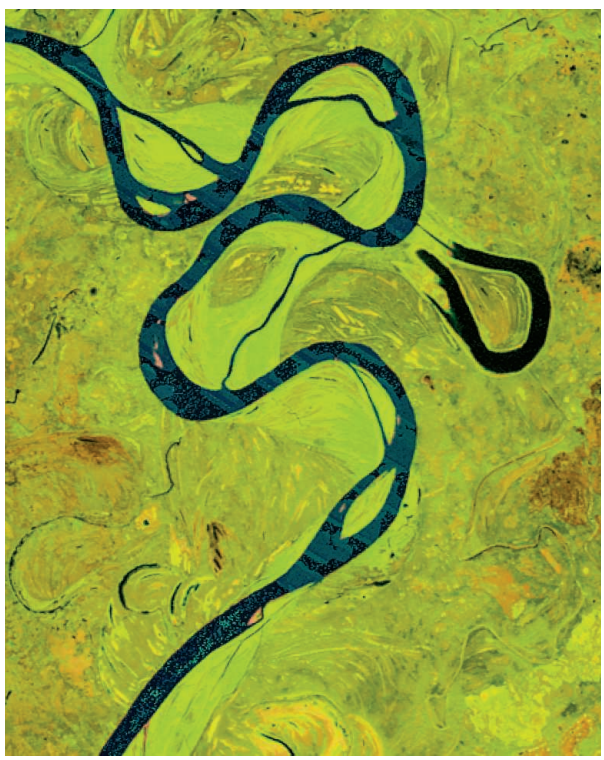
[1]*National Socio-Environmental Synthesis Center, University of Maryland, Annapolis, MD 21401, USA.* [2]*Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA 94720, USA. Email: mpalmer@sesync.org*

To scale up site-level measurements for estimating the global contribution of rivers to biodiversity or gas fluxes, scientists must not only contend with their variability (*5*) but also estimate water surface area precisely. Allen and Pavelsky tackle this problem in a novel way. Using satellite-derived data and a combination of statistical tools, they argue that global river surface area is 29 to 59% larger than the best previous figure (*6*). Past global estimates of river surface area relied on statistical properties of branching networks and/or flow-routing algorithms rather than on direct observations (*6*, *7*). Allen and Pavelsky instead used Landsat images to create horizontal river shapes. They validated their river width estimates with available in situ measurements across the United States and Canada.

The work is important for at least three reasons. First, it delivers global coverage at a resolution that is useful for river system models and for estimating biogeochemical fluxes, including greenhouse gas emissions. Second, it is relevant from a biodiversity conservation perspective because the results show that river surface area had been underestimated in highly biodiverse regions such as Central America, New Zealand, and most of Southeast Asia. Third, use of empirical data allows more accurate surface area estimates for places where geomorphic scaling relationships do not hold. River networks shrink or expand often in response to human activities. In highly developed regions, river channels may be drained and flow augmented or reconfigured, all of which influence their surface area. Previous approaches did not account for such possibilities—an important omission, given the pervasiveness of river ecosystem modification.

Quantifying the distribution and temporal variability of fresh waters by use of remote sensing is becoming a vibrant field of research (*8*), and scientists' ability to detect and understand changes in the global water cycle will increase further as the spatial and temporal resolution of remote-sensing applications continue to improve (*9*). In their validation process, Allen and Pavelsky found that river widths estimated from 30-m-resolution Landsat data were most accurate and complete for rivers wider than 90 m. For smaller waterways, the authors relied on Pareto frequency distributions (for data-rich basins) and on power-law distributions (for data-poor basins) to include up to first-order streams (small, perennial streams at the tops of watersheds) in the estimates of total river surface area. Recently, automated methods were developed for estimating subpixel water fraction with 30-m-resolution Landsat data, suggesting that it may be possible in the future to map streams as narrow as 5 m (*10*).

Future mapping should also focus on enhancing the temporal resolution of observations. The Sentinel-2 satellites of the European Space Agency (ESA) now collect observations at 10-m resolution at 5-day intervals (*11*), capturing the temporal dynamics of small water bodies. Further, the image acquisition modes of ESA's Sentinel-1 satellites allow global collection of high-resolution synthetic aperture radar (SAR) data on a 12-day basis or better and can do so day or night and regardless of cloud cover (*12*). As new methods for mapping water from SAR imagery continue to emerge (*13*), use of Sentinel-1 alone combined with Landsat and Sentinel-2 will allow global mapping of surface water on a near-daily basis for many regions. Outside the world of open-access data, private companies are pushing the game even further. For instance, the San Francisco–based Planet has launched a group of small satellites that gather data globally every 24 hours or less at a spatial resolution of 3.7 meters—specifications that make these data precious for a wide range of environmental applications (*14*).



A temporal composite image of Slave River in Northwest Territories, Canada, shows how river meanders have created oxbows. The methods used to integrate Landsat time series are available in (*15*).

Allen and Pavelsky's study may stimulate further research that uses freely available data to improve global estimates of inundated surface areas of wetlands and streams and to better understand how these flooded areas change over time. This should contribute to major advances in regional and global ecological and Earth system models. Fresh waters are a large source of greenhouse gas emissions yet represent the greatest source of uncertainty in global methane emissions. This is mainly because the surface area of fresh waters and its temporal variability are poorly constrained. Measurements of gas emissions from ephemeral streams are in their infancy but may collectively represent nearly 20,000 km^2 when flowing—more than the main stem of the Amazon (*7*). Including these smaller water bodies in future area estimates, and understanding how their contraction-expansion phases influence global biogeochemical processes, will be major scientific advances. Many of these findings could have direct policy relevance, given that the size and variability in surface water inundation influences the protection that water bodies receive under law.

Water-related risks will likely continue to exist in the future (*2*), but remote-sensing technologies are providing increasingly precise estimates of global freshwater distribution and trends. These data can benefit freshwater scientists and managers alike, opening new avenues to study the global water cycle and to manage it sustainably. ■

PHOTO: MIKE WULDER

REFERENCES AND NOTES

1. C. Kammerer, "U.N. Global Climate and Pacific Institute Initiative" (2018); https://ceowatermandate.org/posts/water-connected-every-global-risk-face.
2. World Economic Forum, *The Global Risks Report 2018* (World Economic Forum, ed. 13, 2018).
3. G. H. Allen, T. M. Pavelsky, *Science* **361**, 585 (2018).
4. A. V. Borges *et al.*, *Nat. Geosci.* **8**, 637 (2015).
5. J. T. Crawford *et al.*, *J. Geophys. Res.* **122**, 1036 (2017).
6. P. A. Raymond *et al.*, *Nature* **503**, 355 (2013).
7. J. A. Downing *et al.*, *Inland Waters* **2**, 229 (2012).
8. C. Huang, Y. Chen, S. Zhang, J. Wu, *Rev. Geophys.* 10.1029/2018RG000598 (2018).
9. M. K. Vanderhoof, H. E. Distler, M. W. Lang, L. C. Alexander, *Wetl. Ecol. Manag.* **26**, 63 (2018).
10. B. DeVries *et al.*, *Remote Sens.* **9**, 807 (2017).
11. M. Drusch *et al.*, *Remote Sens. Environ.* **120**, 25 (2012).
12. R. Torres *et al.*, *Remote Sens. Environ.* **120**, 9 (2012).
13. W. Huang *et al.*, *Remote Sens.* **10**, 797 (2018).
14. M. Strauss, *Science* **355**, 782 (2017).
15. T. Hermosilla *et al.*, *Remote Sens. Environ.* **158**, 220 (2015).

ACKNOWLEDGMENTS

This work was supported by a grant from the National Science Foundation (DBI#1639145). Both authors contributed equally to the writing.

10.1126/science.aau3842

SIGNALING

Phase changes in neurotransmission

Synaptic vesicles cluster by partitioning into phase-separated condensates

By **Edgar E. Boczek** *and* **Simon Alberti**

Clusters of synaptic vesicles (SVs) are distinctive features of axon terminals in neurons. SVs contain neurotransmitters that are released upon electrical activity. The discovery of SVs as defined packages of neurotransmitters provided the molecular basis for the quantized theory of synaptic transmission in discrete units of information (*1*, *2*). It has remained enigmatic what holds these vesicles together in clusters. On page 604 of this issue, Milovanovic *et al.* (*3*) show that synapsin 1, a protein that is essential for SV cluster formation, condenses into liquid droplets in vitro. Lipid vesicles and other SV cluster components specifically partition into these condensates, suggesting that SV cluster formation occurs through protein phase separation.

Synapsin 1 is a major component of SV clusters, in which it reaches high local concentrations (*4*). It harbors a folded adenosine triphosphate (ATP)–binding domain of unknown function and a long intrinsically disordered region (IDR) at its C terminus. The IDR contains numerous proline-rich SRC homology 3 (SH3)–domain binding sites. Milovanovic *et al.* found that purified synapsin 1 phase-separates out of solution in physiological conditions. The condensates had characteristics of a liquid phase because the droplets fused and material was rapidly exchanged with the surroundings. Other components of SV clusters, such as the SH3 domain–containing proteins growth factor receptor–bound 2 (GRB2) and intersectin 1, specifically partitioned into the synapsin 1 droplets. Thus, synapsin 1 phase separation may generate a reservoir for various synaptic proteins close to their site of action, as suggested by the authors previously (*5*).

One recently introduced framework proposes that condensates contain scaffolds that are required for condensate integrity and scaffold-binding clients that are dispensable for condensate formation but partition into the condensed phase once it has formed (*6*). That synapsin 1 can phase-separate on its own implies that it may be a scaffold protein, whereas the partitioning of SH3 domain–containing proteins into the synapsin 1 droplets indicates that they may be clients. At low concentrations, the SH3 domain–containing clients decrease the threshold concentration for phase separation. By contrast, excess client proteins prevent droplet formation. This observation is reminiscent of previous studies showing that client RNAs regulate the phase separation behavior of scaffold RNA-binding proteins (*7*, *8*).

What is the role of condensed synapsin 1? Fluorescently labeled lipid vesicles (liposomes) partitioned into synapsin 1 droplets and then formed clusters, a remarkable observation that is reminiscent of SV clusters in vivo. To date, such a specific interaction between lipid vesicles and phase-separated protein condensates has not been observed.

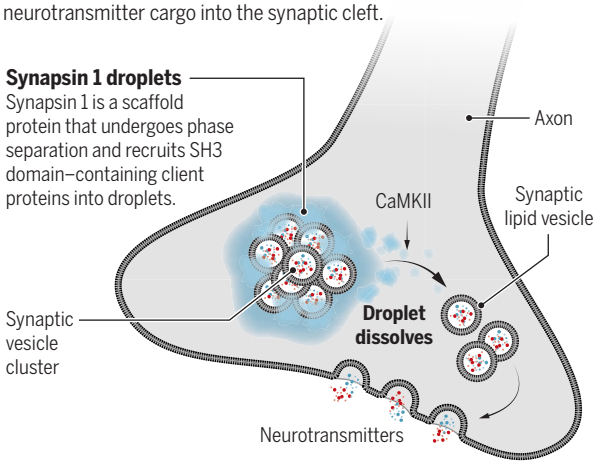
Neurotransmission requires the release of SVs from their clusters and fusion with the plasma membrane to discharge their neurotransmitter contents. It was previously shown that SV cluster release is controlled by calcium/calmodulin-dependent protein kinase II (CaMKII)–mediated phosphorylation of synapsin 1 (*9*). Consistently, the synapsin 1 condensates dissolve when CaMKII and ATP are added. This is also true for liposome-containing synapsin 1 droplets, thus providing further evidence for the idea that SV clusters are condensed droplets that are formed and dissolved in response to physiological cues (see the figure). Together with previous studies showing regulation of phase separation by kinases (*10*, *11*), this suggests that phosphorylation of scaffold proteins is an important general principle for controlling condensation. What remains to be determined is how the kinase interacts with the condensate and how droplet dissolution is achieved at the molecular level.

Previous models of SV cluster formation suggested a network of protein-protein and protein-SV interactions. In these models, synapsin 1 and the SH3 domain–containing proteins form a solid matrix that was proposed to specifically bind to SVs via BAR (Bin-amphiphysin-Rvs) domains (*12*). Alternatively, SVs were suggested to be connected by nondynamic networks of SH3 or BAR domain–containing proteins that are linked together by dynamin or vesicular glutamate transporter 1 (VGLUT1)–endophilin, both of which are involved in SV clustering (*12*). Although these models can explain the clustering of SVs, they do not explain the observation of dynamic SV rearrangements within the clusters (*13*). The phase-separation model presented by Milovanovic *et al.* can account for this observation: A liquid phase containing SVs has the advantage over a solid matrix or network in that it allows gradual SV recruitment and release without the need for full matrix or network disassembly (*5*).

The situation in axon terminals is more complex than the in vitro system used in this study, and additional factors may play important roles. Nonetheless, this study paves the way for a better understanding of the molecular basis of synapse organization. It also provides another beautiful example of the power of in vitro reconstitution biochemistry. A stepwise increase of complexity using this in vitro system will

Clustering of synaptic vesicles

Synaptic lipid vesicles are proposed to be clustered together by partitioning into synapsin 1 droplets. Upon synapsin 1 phosphorylation by CaMKII, the droplets dissolve, leading to the release of synaptic vesicles to the membrane for fusion and delivery of their neurotransmitter cargo into the synaptic cleft.



GRAPHIC: C. BICKEL/*SCIENCE*

Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. Email: alberti@mpi-cbg.de

provide valuable insights into the interplay between the different factors. Moreover, mutant proteins with altered phase-separation behavior can now be developed using the authors' in vitro system and then tested in living cells. It will be fascinating to see how these mutants influence SV cluster formation, neurotransmitter release, and endplate potentials in the synapse.

Numerous important questions still need to be answered: What are the molecular interactions driving the formation of synapsin 1 condensates? That the synapsin 1 IDR phase-separates without the folded ATP-binding module and in the absence of SH3 domain–containing proteins suggests that a multivalent interaction network mediated by the IDRs drives phase separation. Synapsin 1 does not contain typical aromatic amino acids like other phase-separating proteins, suggesting an unusual mechanism that requires further investigation. This question of specific molecular interactions is particularly important because of the likely presence of other phase-separated compartments in the axon terminals. Furthermore, what controls the size of the SV clusters and keeps the condensates from coalescing? What role does the ATP-binding domain play? What drives the interaction between the protein condensate and the lipid vesicles, and how is specificity achieved in this context? A specific mechanism mediated by lipid-binding domains or a less specific physical principle seems conceivable. More generally, clustering of particles and vesicles by phase separation may emerge as a principle of fundamental importance in biology. ■

"...clustering of particles and vesicles by phase separation may emerge as a principle of fundamental importance in biology."

REFERENCES

1. J. del Castillo, B. Katz, *J. Physiol.* **124**, 560 (1954).
2. G. J. Augustine, H. Kasai, *J. Physiol.* **578**, 623 (2007).
3. D. Milovanovic *et al.*, *Science* **361**, 604 (2018).
4. B. G. Wilhelm *et al.*, *Science* **344**, 1023 (2014).
5. D. Milovanovic, P. De Camilli, *Neuron* **93**, 995 (2017).
6. S. F. Banani *et al.*, *Cell* **166**, 651 (2016).
7. S. Maharana *et al.*, *Science* **360**, 918 (2018).
8. P. R. Banerjee *et al.*, *Angew. Chem. Int. Ed.* **56**, 11354 (2017).
9. F. Benfenati *et al.*, *Nature* **359**, 417 (1992).
10. F. Wippich *et al.*, *Cell* **152**, 791 (2013).
11. J. B. Woodruff *et al.*, *Science* **348**, 808 (2015).
12. O. Shupliakov *et al.*, *Semin. Cell Dev. Biol.* **22**, 393 (2011).
13. Y. Wu *et al.*, *eLife* **3**, e01621 (2014).

10.1126/science.aau5477

THERMAL CONDUCTIVITY

Ultrahigh thermal conductivity confirmed in boron arsenide

High-quality crystals minimize conductivity losses caused by phonon scattering at defects

By **Chris Dames**[1,2]

Few materials possess ultrahigh thermal conductivity κ at room temperature, defined here as exceeding that of copper ($\kappa_{Cu} \approx 400$ W/m·K), yet such materials are vital for dissipating the heat loads from modern electronics (see the figure). Diamond is the highest-κ bulk material known, with $\kappa_{Diamond}$ as large as 2300 W/m·K, but it is costly to synthesize in sufficient quality and quantity, which limits its impact on applications. On the basis of widely accepted phenomenological theory (*1*), for decades boron arsenide (BAs) was expected to have an unremarkable κ_{BAs} of 200 W/m·K, although this value was never measured. In 2013, a stunning prediction was made that κ_{BAs} might be 10 times as high and approach $\kappa_{Diamond}$ (*2*). Testing such predictions (*2*, *3*) required BAs crystals of unprecedented quality, which launched a flurry of experimental efforts. Reports by Kang *et al.* (*4*) on page 575, Tian *et al.* (*5*) on page 582, and Li *et al.* (*6*) on page 579 of this issue describe the synthesis of small yet superb crystals of BAs and the measurement of κ_{BAs} values of ~1100 W/m·K. These results verify the essence of the predictions (*2*, *3*) and earn BAs long-overdue recognition as an ultrahigh-κ material.

In insulating and semiconductor materials, heat is transported by phonons, which in the purest crystals are limited only by phonon-phonon scattering. If the interatomic bonding forces are well known, κ can in principle be calculated directly from the Boltzmann transport equation (*7*). Such calculations were intractable for more than 80 years because of their computational complexity and a lack of highly accurate interatomic potentials. However, great progress was still made through approximate models that yield intuitive insights, such as Slack's classic guidelines (*1*) for high κ—simple crystals, light atoms, and stiff bonds with low anharmonicity.

[1]*Department of Mechanical Engineering, University of California, Berkeley, CA 94720, USA.* [2]*Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. Email: cdames@berkeley.edu*

Such insights have proved very effective, as exemplified by the trends shown in the figure for the group IV materials: $\kappa_{Ge} < \kappa_{Si} < \kappa_{Diamond}$, i.e., increasing phonon κ moving up a column of the periodic table. A similar trend was expected for the BX compounds, where X is a group V element: $\kappa_{BAs} < \kappa_{BP} < \kappa_{BN}$. This trend is consistent with the experimental values for BN and BP in the figure, and so for more than 40 years, no one questioned (or bothered to measure) the Slack estimate of 200 W/m·K (*1*).

An accumulation of advances in the algorithms for solving the Boltzmann equation (*8*), the availability of force constants calculated from first principles, and computing power converged to enable direct calculations of κ with high accuracy, a breakthrough that is changing this field (*9*). Such calculations still invoke some kind of approximation, typically either the relaxation-time approximation or truncating the calculation at three- rather than four-phonon interactions. The first such calculation for BAs (*2*) was stunning, predicting (at three-phonon accuracy) a κ_{BAs} approaching $\kappa_{Diamond}$. Perhaps just as important, that paper highlighted two new phenomenological insights based on energy and momentum conservation: Suppression of three-phonon scattering can occur when the energy gap between acoustic and optical phonon branches is large, and also when the three acoustic branches are bunched tightly together.

The essence of the prediction (*2*) was widely accepted and launched efforts at experimental verification. The first measurements of κ_{BAs} followed soon thereafter (*10*), although point-defect scattering limited the thermal conductivity to only 200 W/m·K. A burst of other work followed. Measurement of the BAs dispersion relation (*11*) verified the predicted acoustic-optical energy gap and acoustic bunching (*2*). A four-phonon accurate calculation (albeit under the relaxation-time approximation) (*3*) revised the expected κ_{BAs} downward, to a value of 1400 W/m·K which was still very impressive. Finally, a major multidisciplinary research initiative was launched that supported two of the current teams, and also provides a

nice example of positive feedback between academia and funding agencies.

However, realizing ultrahigh κ_{BAs} would require extraordinary materials synthesis efforts to minimize phonon scattering by grain boundaries and point defects. Furthermore, BAs poses special synthesis challenges because of the constituent elements' high melting points, the thermodynamic preference for an undesired $B_{12}As_2$ phase below ~920°C, and arsenic's high volatility and toxicity (*10*). Determination, patience, and time-consuming growth protocols that ran for weeks to even months led to successful syntheses by all three groups. For the culminating measurements of κ itself, the three studies relied mainly on a laser-based method with ~10-μm spatial resolution capable of surface mapping. Such microscopic measurements were invaluable for providing timely feedback to the material growers on numerous tiny samples, some of which were smaller than a grain of sand.

The three studies are consistent in their main conclusions. As shown in the figure, the κ_{BAs} values from their best "champion samples" all lie within ±15% of the cross-study average of 1140 W/m·K, which also agrees reasonably with a new calculation (1250 W/m·K) that is fourth-order accurate without invoking the relaxation-time approximation (*5*). Furthermore, the studies all found a temperature (T) dependence of κ steeper than T^{-1}, confirming that κ_{BAs} is indeed limited mainly by four-phonon scattering (*3, 5*), rather than by scattering at grain boundaries or point defects, which would have given a flatter T dependence. Studying their κ_{BAs} values across a broader range of sample quality, Li *et al.* (*6*) also identified an intriguing correlation with the Raman spectra, which suggests that free carriers (electrons and holes) might also play an important role in scattering the phonons in BAs.

Heat spreaders for thermal management of high-power electronics rely primarily on synthetic diamond, BN, SiC, Ag, Al, and Cu (see the figure). Graphite is less suited for these applications because of its very low κ value normal to the basal planes (~10 W/m·K), as well as its poor mechanical properties. A critical consideration is the spreader's coefficient of thermal expansion (CTE), which should match that of the heat source that it contacts. For typical devices based on Si, GaAs, or GaN, with CTEs ≈3 to 6 parts per million (ppm) per K, BAs at 3 ppm/K is a better match than diamond at 1 ppm/K.

Cooling applications also require a very high thermal boundary conductance between the heat spreader and surrounding metals. In this regard, the available measurements for interfaces of Al/BAs [130 MW/m²·K (*6*)] and Au/Ti/BAs [40 MW/m²·K (*5*)] are promising, and in line with trends for other high-quality interfaces (*12*). Still, if BAs materials are to realize their potential for electronics cooling applications, the material growers will have to develop much lower cost and more scalable synthesis methods while maintaining outstanding quality. Developing chemical vapor deposition methods for high-quality BAs films—similar to those already used for some industrial diamond chips—is one obvious though challenging direction. Another application direction might be to develop BAs powders for metal matrix composites, which would require only small particles (say, 50 μm diameter).

It is natural to wonder what other doors may be opened by this new understanding of ultrahigh phonon κ. The link is firmly established between ultrahigh κ and acoustic phonon bunching combined with large acoustic-optical gaps (*2*), but there may still be room to complement this with improved chemical intuition about the most promising types of materials and bonding, beyond simply increasing the mass ratio in a binary compound. Perhaps these ideas can be generalized fruitfully for highly anisotropic materials, with the potential to overcome some of the shortcomings of graphite.

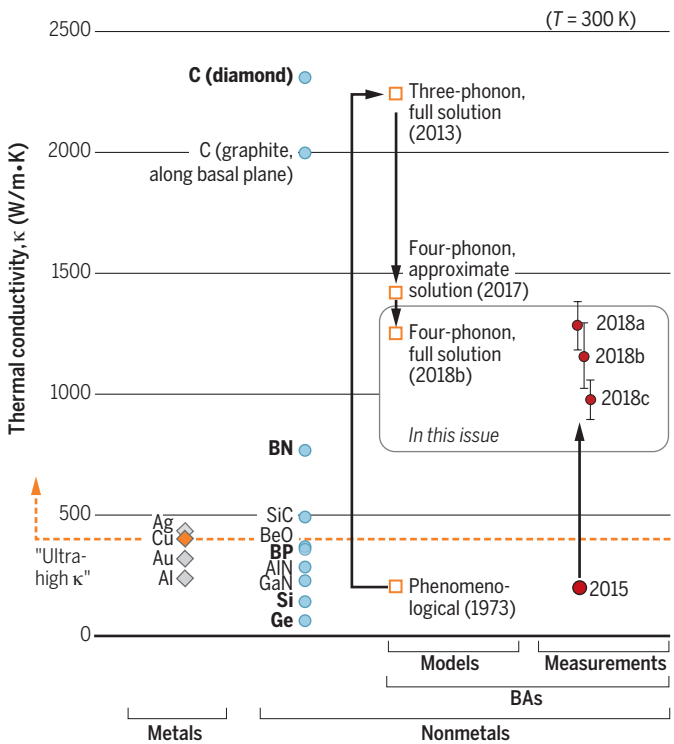
We can also appreciate how the phenomena underlying the ultrahigh κ of BAs are a special case of the more general strategy of restricting the phase space for phonon-phonon scattering, which has also inspired recent efforts at κ engineering in coherent phononic metamaterials, although these efforts have so far proved elusive (*13*). Related scattering selection rule concepts help explain the ultrahigh κ in some two-dimensional materials (*14*). Looking forward, it is exciting to imagine what other breakthroughs in thermal materials may lie in wait, by fusing the new generation of powerful tools for prediction, synthesis, and measurement with the best traditions of phenomenological insight. ■

High thermal conductivity materials in context

As confirmed in this issue (*4*–*6*), crystals of boron arsenide (BAs) of unprecedented quality have a thermal conductivity κ_{BAs} about five times higher than traditionally thought (*1*). This finding verifies the essence of recent first-principles predictions (*2, 3*) and belatedly promotes BAs to its rightful place among the higher ranks of this group.

Nailing down a moving target

Arrows indicate the evolution of κ_{BAs} by year. For the models (squares), the year and reference are 1973 (*1*), 2013 (*2*), 2017 (*3*), and 2018b (*5*). For experimental measurements (circles), the year and reference are 2015 (*10*), 2018a (*4*), 2018b (*5*), and 2018c (*6*).



GRAPHIC: ADAPTED BY N. CARY/*SCIENCE*

REFERENCES

1. G. A. Slack, *J. Phys. Chem. Solids* **34**, 321 (1973).
2. L. Lindsay, D. A. Broido, T. L. Reinecke, *Phys. Rev. Lett.* **111**, 025901 (2013).
3. T. L. Feng, L. Lindsay, X. L. Ruan, *Phys. Rev. B* **96**, 161201 (2017).
4. J. S. Kang *et al.*, *Science* **361**, 575 (2018).
5. F. Tian *et al.*, *Science* **361**, 582 (2018).
6. S. Li *et al.*, *Science* **361**, 579 (2018).
7. R. Peierls, *Ann. Phys.* **395**, 1055 (1929).
8. M. Omini, A. Sparavigna, *Phys. Rev. B* **53**, 9064 (1996).
9. D. A. Broido, M. Malorny, G. Birner, N. Mingo, D. A. Stewart, *Appl. Phys. Lett.* **91**, 231922 (2007).
10. B. Lv *et al.*, *Appl. Phys. Lett.* **106**, 074105 (2015).
11. H. Ma *et al.*, *Phys. Rev. B* **94**, 220303 (2016).
12. C. Monachon, L. Weber, C. Dames, *Annu. Rev. Mater. Res.* **46**, 433 (2016).
13. J. Lee *et al.*, *Nat. Commun.* **8**, 14054 (2017).
14. J. H. Seol *et al.*, *Science* **328**, 213 (2010).

10.1126/science.aau4793

OBESITY

Tighter lymphatic junctions prevent obesity

Zippering of cellular junctions in intestinal lacteals prevents fat uptake

By **Donald M. McDonald**

Dietary fats take a circuitous route from the intestine to the bloodstream, where they serve as nutrients or are stored in adipose tissue. Fats processed in the digestive tract are packaged by the intestinal epithelium into tiny lipid-protein particles called chylomicrons. Chylomicrons are taken up by lymphatic channels (lacteals) within villi that line the small intestine and are transported by lymphatic vessels to the blood circulation. How chylomicrons enter lacteals has long been a mystery, as the particles are too large to cross the endothelial cells that line lymphatics. Moreover, scattered openings observed in lacteals seemed too sparse to explain chylomicron entry, and evidence for vesicular transport was limited (*1*, *2*). The discovery of specialized, discontinuous button-like junctions between lacteal endothelial cells raised another possibility (*3*). Button junctions have open regions and closed regions and are strikingly different from zipper junctions that tightly seal endothelial cells in blood vessels and collecting lymphatics. On page 599 of this issue, Zhang *et al.* (*4*) show that chylomicron entry is dependent on button junctions. Importantly, experimental manipulations that led to transformation of button junctions into zippers prevented chylomicron uptake and protected mice from developing obesity while on a high-fat diet.

Intestinal lymphatic vessels not only transport chylomicrons but are also routes for the recirculation of fluid into blood. According to the conventional view, fluid flows into lacteals along hydrostatic pressure gradients when their endothelial cells are pulled apart by interstitial forces (*5*). The extent of separation of the cells at openings is governed by forces within villi transmitted to the lacteal wall by anchoring filaments. Intake of dietary fat promotes chylomicron production and water absorption by the intestinal epithelium that lead to higher interstitial pressure, lymph flow, and chylomicron transport through lacteals (*5*). Lacteal contractility, which is driven by longitudinal smooth muscle cells in intestinal villi, can propel fluid and cells in lymph at velocities up to 150 μm/s into downstream lymphatics en route to the bloodstream (*6*).

Zhang *et al.* sought to determine the contributions of the endothelial cell receptors, vascular endothelial growth factor receptor–1 (VEGFR1) and neuropilin-1 (NRP1), to the regulation of fat transport. By deleting the corresponding genes, *Vegfr1* (also known as *Flt1*) and *Nrp1*, in endothelial cells of mice, they unexpectedly found that button junctions in lacteals transformed into zipper junctions and that this conversion was accompanied by resistance to obesity when mice were given a high-fat diet (see the figure).

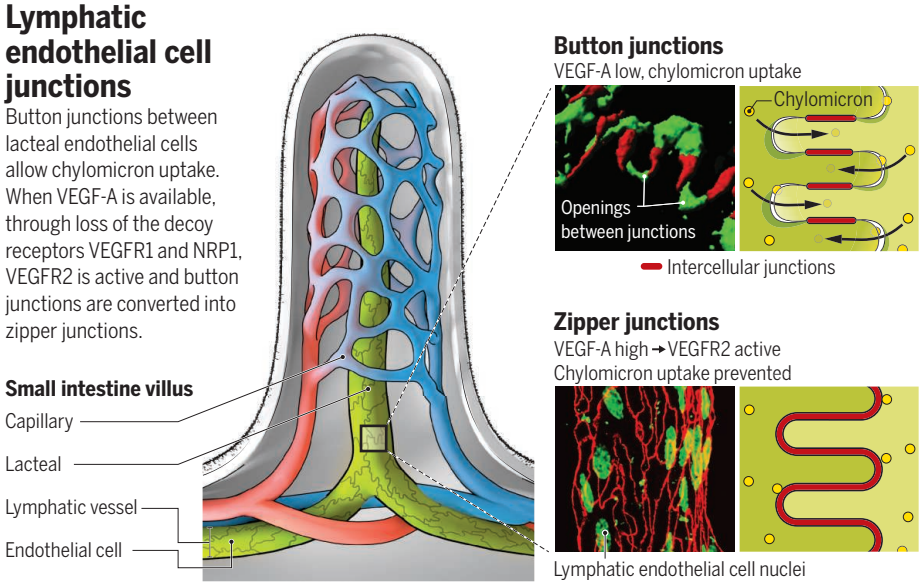
This work was initiated to resolve a controversy over the involvement of VEGFR1 and NRP1 in fat transport in obesity and type 2 diabetes. According to one concept, VEGF-B, a ligand of VEGFR1 and NRP1, increases expression of fatty acid transporters in blood vessels through binding and activating VEGFR1 and NRP1, which consequently promote fat transport into tissues (*7*). A contrasting interpretation is that VEGF-B improves vascular and metabolic health by displacing another ligand, VEGF-A, which can then activate VEGFR2 signaling (*8*). VEGFR2 signaling is essential for maintaining vascular function, which is compromised in metabolic syndrome and type 2 diabetes. In the absence of VEGF-B, VEGFR1 and NRP1 can act as decoy receptors that compete for VEGF-A binding to VEGFR2. In support of this idea, Zhang *et al.* provide evidence that VEGFR1 and NRP1 bind VEGF-A and prevent excessive VEGFR2 signaling that would otherwise promote button-to-zipper transformation and thereby reduce chylomicron uptake into lacteals.

Zhang *et al.* add to the evidence that endothelial cell junctions are dynamic structures. Button junction plasticity is evident during development: Initial lymphatics have zipper junctions until late gestation, when they change into button junctions through a process that requires the growth factor angiopoietin-2 (ANGPT2) (*9*, *10*). Alternatively, button junctions change into zippers in growing lymphatics in inflamed tissues (*9*). This transformation is reversible by treatment with the corticosteroid dexamethasone (*9*). Further evidence of plasticity is provided by the dependence of lacteal function on VEGF-C, a ligand for VEGFR3, and on delta-like ligand 4 (DLL4), a NOTCH ligand. Mice lacking *Vegfc* or *Dll4* have reduced dietary lipid absorption and are resistant to obesity (*11*, *12*). Genetic inactivation of *Dll4* in lymphatic endothelial cells results in transformation of button junctions into zippers (*12*). In

Department of Anatomy, University of California, San Francisco, CA 94143-0452, USA. Email: donald.mcdonald@ucsf.edu

Lymphatic endothelial cell junctions

Button junctions between lacteal endothelial cells allow chylomicron uptake. When VEGF-A is available, through loss of the decoy receptors VEGFR1 and NRP1, VEGFR2 is active and button junctions are converted into zipper junctions.



GRAPHICS: V. ALTOUNIAN/*SCIENCE*; CONFOCAL IMAGES REPRODUCED WITH PERMISSION FROM BALUK *ET AL.* (*3*)

addition to the gatekeeper function of cellular junctions, chylomicron entry into lacteals is controlled by the expression of the transcription factor pleomorphic adenoma gene-like 2 (PLAGL2) in intestinal epithelial cells. *Plagl2*-deficient mice have impaired lacteal uptake of chylomicrons and die postnatally from lack of fat absorption (*13*).

The findings by Zhang *et al.* raise the question of how VEGF-A can have opposite effects on lymphatics, where it reduces permeability by promoting zipper junctions, and on blood vessels, where it increases leakage by opening normally closed zipper junctions in endothelial cells. Although the basis of this dichotomy is unclear, mechanistic insights come from studies of Rho-associated protein kinases (ROCKs). Pharmacologic inhibition of ROCKs promote zippering of endothelial cell junctions and suppress chylomicron uptake into lacteals. ROCK inhibitors also reduce leakage in blood vessels (*14*). In both cases, ROCK inhibitors increase the barrier function of endothelial cells by stabilizing proteins at intercellular junctions.

These advances highlight the importance of button junctions in intestinal lacteals for the uptake of dietary fat. Although button junctions were identified in lymphatics more than 10 years ago (*3*), the finding that the junctions are essential for chylomicron uptake by lacteals and can be genetically and pharmacologically regulated point to new strategies for management of obesity. However, understanding the mechanism and consequences of the transformation of button junctions into zippers is at an early stage. Zippering junctions in lymphatics could impair uptake of essential nutrients in chylomicrons or have adverse effects by compromising fluid drainage and immune cell trafficking. Nonetheless, the discovery that the uptake of intestinal fat is determined by the structure of junctions in lymphatic endothelial cells opens new avenues for investigating the regulation of body weight. ■

REFERENCES

1. J. R. Casley-Smith, *J. Cell Biol.* **15**, 259 (1962).
2. W. O. Dobbins 3rd, E. L. Rollins, *J. Ultrastruct. Res.* **33**, 29 (1970).
3. P. Baluk *et al.*, *J. Exp. Med.* **204**, 2349 (2007).
4. F. Zhang *et al.*, *Science* **361**, 599 (2018).
5. P. R. Kvietys, D. N. Granger, *Ann. N. Y. Acad. Sci.* **1207** (suppl. 1), E29 (2010).
6. K. Choe *et al.*, *J. Clin. Invest.* **125**, 4042 (2015).
7. C. Hagberg *et al.*, *Physiology* **28**, 125 (2013).
8. M. R. Robciuc *et al.*, *Cell Metab.* **23**, 712 (2016).
9. L.-C. Yao *et al.*, *Am. J. Pathol.* **180**, 2561 (2012).
10. W. Zheng *et al.*, *Genes Dev.* **28**, 1592 (2014).
11. H. Nurmi *et al.*, *EMBO Mol. Med.* **7**, 1418 (2015).
12. J. Bernier-Latmani *et al.*, *J. Clin. Invest.* **125**, 4572 (2015).
13. F. Van Dyck *et al.*, *Cell Metab.* **6**, 406 (2007).
14. C. M. Mikelis *et al.*, *Nat. Commun.* **6**, 6725 (2015).

10.1126/science.aau5583

ESSAY

The future of humans as model organisms

A "human phenomic science" approach could accelerate personalized medicine

By **Garret FitzGerald[1], David Botstein[2], Robert Califf[3,4], Rory Collins[5], Keith Peters[6], Nick Van Bruggen[2], Dan Rader[1]**

Ten years ago, Nobel laureate Sydney Brenner remarked, "We don't have to search for a model organism anymore. Because we are the model organisms" (*1*). Indeed, over the past decade, we have deepened our understanding not only of how the genomic blueprint for human biology manifests physical and chemical characteristics (phenotype), but also of how traits can change in response to the environment. A better grasp of the dynamic relationship between genes and the environment may truly sharpen our ability to determine disease risk and response to therapy. A collection of human phenotypic data, and its integration with "omic" information (genomic, proteomic, transcriptomic, epigenomic, microbiomic, and metabolomic, among others), along with remote-sensing data, could provide extraordinary opportunities for discovery. A comprehensive "human phenomic science" approach could catalyze this effort through both large-scale "light" phenotyping studies and "deep" phenotyping studies performed in smaller groups of individuals.

Data integration is already advancing medicine at the individual patient level. The identification of unexpected disease associations with genes (*2*), the linkage of gene variants to unanticipated human phenotypes (*3*), and the use of Mendelian randomization (a method to estimate causal effects) to predict biomarker validity (*4*) or drug response (*5*) are some examples. Efforts are underway to broaden the diversity of populations studied and to establish standards for phenotyping (*6*), but the depth and quality of characterization at scale are limited by cost and feasibility.

[1]*The Institute for Translational Medicine and Therapeutics, The Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.* [2]*Calico Life Sciences, South San Francisco, CA 94080, USA.* [3]*Duke Forge, Duke University, Durham, NC 27710, USA.* [4]*Verily Life Sciences, South San Francisco, CA 94080, USA.* [5]*Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK.* [6]*University of Cambridge School of Clinical Medicine, Addenbrooke's Hospital, Cambridge CB2 2SP, UK. Email: garret@upenn.edu*

Complementary to large-scale, comparatively lighter basal phenotype studies are deeper phenotyping studies of smaller groups of individuals in settings where phenotypic responses can be evoked under well-controlled conditions (*7*). Such studies can address and refine hypotheses generated from data at scale, enabling investigators interested in mechanisms and drug development to seek proof of concept in focused, prospective clinical trials. Deep phenotyping is ideally suited to elucidate gene function and comes close to conducting a "human gene knockout" study (*8*). The approach can also connect the modifiable causes of common diseases to patients with rare diseases, enabling a better understanding of the pathophysiology of common diseases. For example, detailed characterization of "outlier" patients can direct mechanistic interrogations of common diseases at population scale. One example is the gene encoding proprotein convertase subtilisin/kexin type 9 (PCSK9) and cardiovascular disease (*9*). The cardioprotective phenotype that was revealed through light phenotyping at scale inferred the validity of PCSK9 as a drug target; this was confirmed in deep phenotyping studies and then in randomized clinical trials. Another example is the autoimmune phenotype of patients with atypical presentations of the myalgic encephalomyelitis/chronic fatigue syndrome (*10*). Deep phenotyping revealed a mechanistic basis for a symptom complex that is sometimes miscategorized as an affective disorder. Furthermore, the two approaches to phenotyping might be integrated to enable the identification of patients with rare diseases from large phenotypic datasets. This could lead to more efficient recruitment for deep phenotyping studies. A major challenge will be harmonizing protocols for deep phenotyping. However, while moving toward standardization, it may be possible to inte-

> ***"Deep phenotyping… comes close to…a 'human gene knockout' study."***

grate data when differing protocols are used, as suggested in a recent quantitative analysis of merged yeast proteome datasets (*11*).

What would this approach mean for animal studies? Understanding the larger picture at the organism level has been restricted by the practical limitation of performing bench studies. By necessity, the sample size is small and the phenotypical variance inherent to the system must be controlled. Despite the availability of resources such as the Collaborative Cross and Diversity Outbred mice (which recapitulate levels of genetic heterozygosity seen in humans), the majority of preclinical studies are still performed with inbred strains of mice housed in artificial environments, with every effort made to obtain phenotypic unity. The "noise" in the system is managed to enable reductionist hypotheses to be tested. However, modeling in mice (even with outbred strains) may insufficiently predict the human condition. With deep genotypic and phenotypic evaluation of large human cohorts, the "physiological noise" due to variability in genetics and environmental exposures is measurable and becomes meaningful. A true big-picture systems biology approach to discovery is feasible, and can be facilitated by techniques that provide the quantity and quality of data required to enable physiological measurement at scale (imaging, monitoring devices, etc.).

Human phenomic science would put a collective tag on a variety of experimental approaches. Examples include harvesting induced pluripotent stem cells to parse mechanistic distinctions in patients with varied syndromes of pain (*12*), or characterizing distinctions in the phenotypes evoked by vascular, inflammatory, or metabolic stimuli to gain insight into time-dependent disease expression (*13*). Such studies could be integrated with data from the same patients "in the wild" using remote-sensing devices. This facilitates an intrinsic step in developing precision medicine strategies—comparing interindividual differences in exposure, behavior, and drug response. This might be useful in deciphering "nondipping" hypertension, where there is little information on the efficacy of conventionally formulated antihypertensive drugs, the mechanism involved, and the stability of the phenotype at the individual level, despite the linkage of this syndrome at scale to poor cardiovascular outcomes.

As for drug development, heterogeneous information (physiologic, multi-omics, and imaging data) collected under basal and evoked conditions could be used to interrogate dose-dependent differences in drug response in human phenotypic studies. This information could be integrated with analogous data from other model organisms, allowing more direct linkage to functional outcomes relevant to humans. Some examples include the development of algorithms that predict rare but serious adverse effects of drugs, and the identification of well-defined subgroups that have favorable therapeutic responses to specific medicines. For instance, in assessing nonsteroidal anti-inflammatory drugs (NSAIDs) that differ in selectivity for inhibiting cyclooxygenase-2, one could compare dose-dependent perturbation of the transcriptome, proteome, metabolome, and microbiome across species, with blood pressure as a quantitative surrogate for cardiovascular risk in humans and mice, together with thrombogenesis assays in mice and fish. With machine learning and kinetic and structure-based modeling approaches, one could then identify signatures of NSAID-induced cardiovascular risk and test their predictive efficacy at the individual level prospectively in clinical trials. An extension of this concept is to complement Mendelian randomization for predicting drug efficacy and risk in an individualized approach. New cancer drugs highlight the need to elucidate mechanisms that underlie variability of drug response in therapeutic efficacy as well as in toxicity. Chimeric antigen receptor (CAR)–T cells may cure a life-threatening disease but may also cause a lethal cytokine storm (*14*). Similarly, checkpoint blockade with programmed cell death–1 inhibitors may restrain, leave unaltered, or advance tumor progression (*15*).

Developing a suite of evoked phenotypes might improve the screening of newly approved drugs for unanticipated risks and efficacies. In the United States, the Food and Drug Administration could provide a safe harbor for the emergence of such knowledge, much as it did to encourage studies of pharmacogenetics. Too often, understanding the spectrum of drug action—both efficacy and toxicity—is delayed by risk-averse strategies of drug development focused only on the planned initial indication for approval. Consent to detailed phenotyping might also be linked to drugs still in development but under compassionate use, thus accelerating the acquisition of knowledge relevant to mechanism of action, potential toxicity, clinical trial design, and ultimate clinical utility.

The value of large, prospective datasets and associated biobanks to drug development and precision medicine is already apparent. This could be enhanced by the detailed evoked phenotyping that is only possible in relatively smaller numbers of individuals. As the cost of data recording and analysis declines, deeper phenotyping can be applied at scale and the clear distinction between phenotyping approaches is likely to erode.

What is the advantage of labeling this type of research as human phenomic science? After all, phenotyping studies in humans have long existed. However, investigators are few and resources are scattered. Shifting from the detection of large average effects to a more precise approach to medicine requires great investment and skilled investigators (who, in this case, are in short supply). Within the spectrum of clinical research, the naming of "health services research" and "clinical epidemiology" helped to shape these disciplines with defined skill sets, training programs, and collective activities (meetings, departments). The success of this tactic is reflected in the growth of, and investments in, these clinical areas. Giving human phenome–related research a name can serve a similar purpose of defining, maturing, and refining the endeavor. If we want to move from the detection of large average effects of therapies to a more precise approach to medicine, then we ought to consider a strategic effort to recruit investigators and develop infrastructure to further support human phenomic science. ■

REFERENCES

1. www.genomeweb.com/archive/sydney-brenner-urges-cancer-researchers-consider-bedside-bench-approach
2. W. Haynes *et al.*, bioRxiv 214833 [Preprint]. 16 February 2018. https://doi.org/10.1101/214833.
3. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium, *Mol. Autism* **8**, 21 (2017).
4. B. F. Voight *et al.*, *Lancet* **380**, 572 (2012).
5. J. C. Cohen *et al.*, *N. Engl. J. Med.* **354**, 1264 (2006).
6. www.nationalacademies.org/hmd/Activities/Research/GenomicBasedResearch/Innovation-Collaboratives/Global_Genomic_Medicine_Collaborative.aspx
7. G. A. FitzGerald, *Sci. Transl. Med.* **7**, 284fs15 (2015).
8. D. Saleheen *et al.*, *Nature* **544**, 235 (2017).
9. J. C. Cohen, H. H. Hobbs, *Science* **340**, 689 (2013).
10. M. Hornig *et al.*, *Transl. Psychiatry* **7**, e1080 (2017).
11. B. Ho *et al.*, *Cell Syst.* **6**, 192 (2018).
12. T. Grosser *et al.*, *Science* **355**, 1026 (2017).
13. C. Skarke *et al.*, *Sci. Rep.* **7**, 17141 (2017).
14. M. Kalaitsidou *et al.*, *Immunotherapy* **7**, 487 (2015).
15. A. Ribas, J. D. Wolchok, *Science* **359**, 1350 (2018).

10.1126/science.aau7779

ILLUSTRATION: TAIJI/SHUTTERSTOCK

POLICY FORUM

Members of the Havasupai tribe pray over blood samples at Arizona State University. Disregard for how the research could undermine the tribe's interests led to a lawsuit and out-of-court settlement.

SCIENCE AND SOCIETY

Building an evidence base for stakeholder engagement

The private sector provides lessons and models

By **James V. Lavery**

Science is a social enterprise. Many scientific programs interact with a wide range of communities and stakeholders to secure various types of access and permission, to seek cooperation and collaboration for scientific studies, to fulfill regulatory and ethical requirements, and to try to shape research strategies and to improve the translation of their findings into policy or practice. But these interactions are motivated disproportionately by the interests and goals of the scientific programs and less by the need to elicit and understand their implications for stakeholders. However, there is increasing recognition that substantive community and stakeholder engagement (CSE) can improve the performance, and even make or break the success, of some science programs by providing a means of navigating, and responding to, the complex social, economic, cultural, and political settings in which science programs are conducted. For CSE to become more widely accepted by funders and researchers, and to contribute more conspicuously to the success of science programs and policy, it will have to establish a more coherent and convincing body of evidence about the nature of CSE strategies and their specific contributions to the performance of science programs.

The zeal that drives scientists in their quest for discovery and their deep-rooted faith in the scientific enterprise can sometimes lead them to underestimate, or disregard, the potential for their actions to negatively affect the interests of stakeholders beyond the immediate frame of reference of their scientific protocols. For example, Ashkenazi Jews faced stigmatization and discrimination on the basis of findings of population-genetics research (*1*), and unauthorized research on historical human migration patterns damaged the collective cultural identity of the Havasupai tribe of Arizona (*2*). Despite the importance of such harms, the dominant ethics paradigms in science—scientific integrity and human-subject research protections—provide little guidance about how to anticipate and avoid them.

A common intuition is that these harms can be mitigated by CSE. The idea has attracted interest in a wide range of disciplines, including sustainable development (*3*), regulation of new biotechnologies (*4*), and humanitarian emergencies (*5*), along with long-established practices in community-based participatory research (*6*), patient engagement in clinical research (*7*), and global health (*8*).

Yet, as one recent commentary about CSE noted, "there is limited empirical evidence on the best practices for stakeholder engagement and even less on evaluation of engagement demonstrating the association between the quality and quantity of engagement and research outcomes" (*9*). This lack of evidence about CSE could be the sustaining force for a self-fulfilling prophecy, because those with the authority to make budget decisions for science programs lack clarity about the circumstances under which CSE is necessary, its appropriate scope and form, and a clear and coherent value proposition for how CSE improves the ethics of research and enhances the impact of their investments. The result is often expressed as skepticism or indifference to the potential value of CSE.

KEY CHALLENGES

First, the generation of useful and comparable evidence for CSE is complicated by the absence of an agreed theory of CSE. What are its constituent elements? What mechanisms are involved? What programmatic and ethical outcomes does it produce and under what circumstances? And how do these vary according to the nature of the science and the specific settings of application? Answers to these individual questions would not only provide insights about how CSE works in various contexts but would also facilitate the development of useful theory, which will be essential to move CSE beyond a static and critically unexamined set of practice conventions.

Second, the coherence and comparability of the evidence is undermined by the extraordinary degree of variability in the working language for CSE. Concepts, such as "engagement," "sensitization," "mobilization," "empowerment," and "trust-building," are often conflated and interchanged casually, even though the goals and outcomes they imply differ substantially. Similarly, assumptions about what constitutes the relevant "community," or who should be counted as a legitimate "stakeholder," are often poorly stipulated or specified. This conceptual ambiguity and heterogeneity compounds the problem of insufficient precision and explicitness in the reporting

Hubert Department of Global Health, Rollins School of Public Health, Center for Ethics, Emory University, Atlanta, GA 30322, USA. Email: jlavery@emory.edu

PHOTO: JIM WILSON/*THE NEW YORK TIMES*/REDUX PICTURES

of the social processes and outcomes associated with the prevalent concepts, which frustrates efforts to compare study findings.

Third, there is a tendency to think narrowly about CSE, or to emphasize or exaggerate some aspects relative to others. For example, CSE strategies tend to rely heavily on mechanisms such as community advisory boards, which typically provide limited and uncertain representation of the full range of relevant stakeholder interests and perspectives. They also emphasize communications and various strategies for developing and delivering "key messages" to educate host communities about the goals and merits of the science program. The provision of information is a necessary aspect of CSE but is often emphasized at the expense of listening to and acknowledging the interests of stakeholders. For example, a recent HIV pre-exposure prophylaxis trial for women was critically undermined when it was discovered that many of the participants were simply not using the study product (*10*). The women's interests in the trial—their reasons for participating—were at odds with the researchers' expectations, but these were not identified through the conventional engagement mechanisms.

Fourth, the circumstances described above exacerbate an emerging tension between moves to standardize CSE practices and measurement strategies in science programs (*5*, *11*) and the need to customize them to account for unique social, economic, political, and cultural complexities that shape the contexts within which CSE strategies are executed. Even relatively simple CSE strategies involve multiple interacting components; long, nonlinear implementation chains; and complex sets of human interests, relationships, and associations. In many cases, the motivating interests, reasoning, and behavior of stakeholders are active mechanisms in the performance of the science program itself (*12*). The development of generic approaches to CSE in the form of core principles (*4*), best practices (*5*), and key metrics and indicators (*11*) provides useful momentum for the development of CSE as a legitimate domain of knowledge generation but is unlikely to provide reliable guidance precisely when effective CSE might be most valuable, that is, when the science is controversial or when the human contexts are most complex and/or contentious. Undue confidence in standardized approaches to CSE could inadvertently weaken the force of appeals for better evidence.

IS EVIDENCE WORTH THE INVESTMENT?

In light of these challenges, the broader question for science policy and programming is whether funders should accept that CSE has sufficient potential value for the performance of certain science programs—in addition to the ethical rationales that likely motivate the majority of investments in CSE—to warrant greater investment in an appropriate evidence base. The uncertainty rests on some unresolved, but fundamental, questions about the relationship between science programs and those who might have legitimate grounds to be considered stakeholders. Under what circumstances is CSE necessary? How much and what kind of CSE is necessary? What standing do stakeholders have to assert their interests in any science program and by what processes? What factors should determine the weight that any set of stakeholder interests should carry? What specific obligations should the scientific program be required to acknowledge and accept, including obligations to make changes to their protocols or practices to avoid setting back—that is, harming—stakeholder interests? More broadly, how should science programs be planned, designed, and managed—including the necessary flexibility in budgets and protocols—to allow them to act on valuable stakeholder insights to improve their protocols, practices, and impact and to make changes to avoid harm? These questions have deep normative implications, but, without adequate empirical evidence, they are destined to remain marginal curiosities. Table S1 provides an overview of some of the key components of CSE strategies and a selection of some of the current gaps in evidence.

DESIGN AND MANAGEMENT

Despite the linguistic and conceptual variability described above, CSE has a relatively stable core logic. Science programs are often designed, and final protocol and budget decisions made, remotely from the settings and populations in which they will be conducted. Although scientists tend to emphasize the potential benefits of their work, their programs can also feel like an imposition for some stakeholders and, in some cases, have negative implications for them. For example, comparative trials of new agricultural biotechnologies versus conventional crops can disrupt local market dynamics for small-holder farmers and create discord among neighbors. Therefore, some process is required to identify these potential implications and ensure that the appropriate understandings, agreements, and authorizations are sought from those whose interests may be at stake.

Analogous engagement challenges in other sectors—in particular, private-sector product and service design and supply-chain management—have fueled enormous investments, innovation, evidence generation, and industry-wide adoption of practices in customer-relationship management, customer-experience management, and a wide range of human-centered design strategies. These developments have resulted in paradigm shifts in the way the interests and insights of consumers are elicited and incorporated into product and service design and development processes. For example, human-centered design strategies have produced acclaimed products and services, from the first Apple computer mouse to award-winning educational programs to innovations in financial technology. These innovations share three main features. First, they are built on insights from consumers about their experiences with the products or services in question. Second, their development was motivated by the absence of appropriate management strategies that could reliably deliver consumer insights to inform product and service design. And third, they are highly transferable models, applicable wherever consumers' reasoning and behavior play a role in the performance of the product or service.

The relevance and importance of these features for CSE cannot be overstated. Insights about stakeholder interests and perspectives provide the ethical foundation for CSE (*13*). But they also offer unique value as a means of critically examining, and refining, the design of science programs in response to the specific circumstances of a given research setting. As such, these insights about stakeholder interests also provide an important unit of analysis for a great deal of the necessary empirical research on CSE: What interests do stakeholders have in our science programs? How can we best address them? How do stakeholders want to be engaged? And how can engaging with stakeholders improve the performance of science programs?

Like customer-relationship management, or human-centered design, CSE involves the design and management of social processes. An early experience of the Eliminate Dengue program (now called the World Mosquito Program) (*14*) offers a useful illustration. In its initial open-release trials of *Wolbachia*-infected mosquitoes for population replacement in pursuit of a scalable dengue virus–transmission control strategy, Eliminate Dengue engaged a wide range of stakeholders in Queensland, Australia (*14*). The aims were to facilitate successful trials and to avoid imposing the technology on the population against its will. The CSE effort was widely considered to be successful. Though, for the reasons elaborated above, there was no established standard of "success" to guide an evaluation. Our case study of the Eliminate Dengue CSE strategy sought to critically examine the perceptions of success through in-depth interviews with a wide range of stakeholders, including Eliminate Dengue team members, regulators, a local federal cabinet minister, and local business owners

and residents. On the basis of this study (*14*), I highlight six specific features that appear to have contributed to an impressively uniform perception of success among stakeholders: (i) consistent support (enabling conditions) from the funder and implementation partner to prioritize CSE activities, (ii) clear and consistent leadership to establish CSE as a key priority within the program, (iii) an inclusive view of stakeholders, (iv) a proactive approach to eliciting stakeholder interests and insights and a willingness to be flexible with the design and conduct of the trials in response to them, (v) a clear and coherent set of guiding principles and ethical commitments to stakeholders, and (vi) an explicit management strategy, effectively integrated with the general day-to-day program management, that operationalized the program's guiding principles and ethical commitments and adopted stakeholder interests as a central focus of overall program management (*14*).

WHAT CAN FUNDERS DO?

A number of major funders of science programs have already made substantial investments in research on CSE, including the Wellcome Trust, the National Institutes of Health, and the Bill & Melinda Gates Foundation. These investments have helped to illuminate some of the latent potential of CSE, but the limited evidence they have generated remains largely unknown among funders and generally insufficient to overturn a seemingly common view of CSE as simply another administrative requirement. Funders have a unique power to reframe this narrative, and better justify investments in evidence about CSE, by emphasizing its potential to improve the performance, as well as the ethics, of science programs.

In addition, although a central premise of CSE is that "feedback" from stakeholders is important, science programs are usually not structured in ways that permit meaningful revisions or refinements to programs—particularly to their protocols and budgets—in response to insights and feedback from stakeholders. Funders could substantially advance the mutual value of CSE for researchers and stakeholders by experimenting with more flexible and responsive management strategies, including innovations in protocol and budget processes, and studying the implications for various aspects of program performance. A timely example of this type of innovation comes from the Canadian International Development Research Centre (IDRC), which has recently published results of the implementation of a new tool it has developed to assess the quality of the research it funds, called Research Quality Plus (RQ+) (*15*). The development of the tool reflects IDRC's acknowledgment of "the crucial role of stakeholders and users in determining whether research is salient and legitimate. It focuses attention on how well scientists position their research for use, given the mounting understanding that uptake and influence begins during the research process, not only afterwards" (*15*).

In many scientific fields, a lack of agreement on nomenclature and conceptualization has presented obstacles to progress and has required extensive negotiations and deliberations, often in the form of specific conferences or consensus-building processes. In many cases, these initial deliberations have been critical for the advancement of the discipline and have given rise to some enduring governance structures, such as the International Union of Pure and Applied Chemistry and the *Diagnostic and Statistical Manual of Mental Disorders* of the American Psychiatric Association. For CSE, those funders who are most committed to building an appropriate evidence base might form a consortium to shape a working consensus on basic concepts and nomenclature for CSE to ensure that evidence is built on a sound conceptual architecture, before endorsing and adopting specific "standards" of practice for CSE.

> ***"...science programs are usually not structured in ways that permit meaningful revisions...in response to...stakeholders."***

At a minimum, funders should examine their current investments in CSE associated with science programs and ask whether these investments are contributing to the evidence base for CSE. Some of the necessary insights about how CSE works might be achieved simply by encouraging better reporting and scrutiny of the CSE strategies already being implemented in many science programs: What were the aims of these strategies? Did they work as expected? How and why did they work, or not work, in various contexts? What outcomes were attributable to the CSE? And how, if at all, were these outcomes conceptualized and measured? More production and reporting of this type of evidence should eventually reduce unproductive conceptual and linguistic variability and could provide valuable insights to improve theories of change for how CSE works and identify what tailoring and scaling might be required by different contexts. Improved reporting on these questions could also provide momentum for a broader research agenda for CSE, which could prove to be valuable across a wide range of scientific disciplines. Table S1 offers a point of departure for such a research agenda.

Because there is a self-evident sense in which stakeholders ought to have some say in what is done to them, with them, or on their behalf, funders have already incorporated CSE into many of their science programs. Perfect agreement about the best ways to conceptualize and design CSE strategies is not necessary to improve funding and implementation practices. An empirical evidence–based approach will eventually sort out how, and under what circumstances, CSE adds both ethical and practical value to science programs. A management-oriented approach to evidence generation that focuses on the interests and experiences of stakeholders may yield important insights about how CSE functions in various contexts, analogous to strategies used in customer-relationship and -experience management and human-centered design strategies in product and service development. Such an approach may offer critical insights about how the deeper ethical goals of CSE (*13*) might be more reliably accomplished. Research funders and implementation partners can play a critical role in enabling and establishing this evidence base to guide the appropriate utilization of CSE strategies in science programs. ■

REFERENCES AND NOTES

1. C. Weijer, E. J. Emanuel, *Science* **289**, 1142 (2000).
2. M. M. Mello, L. E. Wolf, *N. Engl. J. Med.* **363**, 204 (2010).
3. C. C. Hicks *et al.*, *Science* **352**, 38 (2016).
4. C. Emerson *et al.*, *Science* **358**, 1135 (2017).
5. D. Pedi *et al.*, *J. Health Commun.* **22**, 39 (2017).
6. Department of Health and Human Services, "Principles of community engagement, Second edition" (Publication 11-7782, National Institutes of Health, 2011).
7. Joint United Nations Programme on HIV/AIDS (UNAIDS), "Good participatory practice: Guidelines for biomedical HIV prevention trials" (Publication JC1853E, UNAIDS, 2011).
8. J. P. Domecq *et al.*, *BMC Health Serv. Res.* **14**, 89 (2014).
9. M. S. Goodman, V. L. Sanders Thompson, *Transl. Behav. Med.* **7**, 486 (2017).
10. L. Van Damme *et al.*, FEM-PrEP Study Group, *N. Engl. J. Med.* **367**, 411 (2012).
11. K. M. MacQueen *et al.*, *BMC Med. Ethics* **16**, 44 (2015).
12. J. Shoveller *et al.*, *Crit. Public Health* **26**, 487 (2015).
13. K. F. King *et al.*, *BMC Med. Ethics* **15**, 84 (2014).
14. P. A. Kolopack, J. A. Parsons, J. V. Lavery, *PLOS Negl. Trop. Dis.* **9**, e0003713 (2015).
15. J. Lebel, R. McLean, *Nature* **559**, 23 (2018).

ACKNOWLEDGMENTS

This work was funded by the Bill & Melinda Gates Foundation, grant OPP1183715, Stakeholder and Community Engagement in Global Health and Global Development. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/554/suppl/DC1

10.1126/science.aat8429

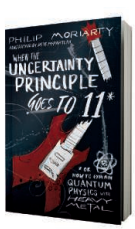
QUANTUM PHYSICS

Turn it up

Science meets metal in a musical introduction to modern physics

By **David J. Reilly**

With increasing numbers of students entering tertiary education, maintaining interest in the physical sciences requires new approaches to engagement that marry course content with students' passions and interests beyond the classroom. In Philip Moriarty's book, *When the Uncertainty Principle Goes to 11: Or How to Explain Quantum Physics with Heavy Metal*, introductory physics comes to life with the energy and enthusiasm of a Metallica concert.



When the Uncertainty Principle Goes to 11
Philip Moriarty
BenBella Books, 2018.
350 pp.

Attempting to explain quantum mechanics by using heavy metal music may at first sound like a challenge worthy of a game show. And yet at closer inspection, it becomes clear that music is an ideal platform for introducing concepts including the physics of waves, harmonics, resonance phenomena, energy conservation, and Fourier analysis in accessible and relatable ways.

Moriarty blends humor with pop-culture references while explaining Heisenberg's uncertainty principle, the cosmic microwave background, and the statistical mechanics of crowds in a concert mosh pit. The book zigzags between physics, blog references, links to YouTube videos, and commentary of the author's own life as a musician and nanoscientist.

Paired with quirky illustrations by Pete McPartlan, *When the Uncertainty Principle Goes to 11* makes for a refreshing and accessible introduction to nanoscience for the curious metalhead. The book features some deep dives into the techniques of scanning tunneling microscopy and its use in arranging and probing single atoms for the construction of atomic electronic devices and somehow connects these concepts back to heavy metal.

Readers raised on a diet of shred-guitar chops will find Moriarty's endless analogies amusing and useful on-ramps to a repertoire of modern physics. His analysis of Eddie Van Halen's guitar effects pedals and the quantitative comparison between the sound of the Big Bang and the loudest heavy metal concert are worth the price of admission alone.

Despite its title, though, the book turns down the volume on quantum mechanics, entirely skipping references to quantum entanglement or the new quantum information science that is currently propelling the quest for technologies such as quantum computers, sensors, and secure communications systems. Rather, Moriarty's perspective is that quantum mechanics is mostly just wave mechanics, devoid of enigmatic concepts and thus not so mind-bending after all.

"Ultimately, wave interference is the source of all the weirdness of the quantum world," he states. Pushing back against the new-age peddling of Deepak Chopra's "quantum woo" is one thing, but in his effort to make the quantum world accessible to the many, Moriarty also squeezes out most of the mystery that has perplexed the best minds for the past 100 years.

Still, as a platform for science engagement, musings about the deep connections between science, technology, and culture make for compelling reading, and Moriarty's book is no exception. As the historian Patrick McCray underscores (see his engaging blog *Leaping Robot*), our understanding of scientific concepts and their technological importance is deeply entrenched and shaped by cultural forces. McCray and other scholars have argued that we owe much of today's technology to yesteryear's science-fiction writers, who, in their imagination, promoted a dazzling vision for the future and articulated much of the technological roadmap, long before science knew how to get there. In this order, culture leads and technology follows.

In music, as in film and other forms of artistic expression, it is often the imperfec-

The reviewer is at the ARC Centre of Excellence for Engineered Quantum Systems and Microsoft Quantum–Sydney, The University of Sydney, NSW 2006, Australia. Email: david.reilly@sydney.edu.au

Researchers have shown that the same mathematical function can be used to describe the movement of molecules in a gas and moshers at a metal gig.

tions imposed by the technologies of the day that establish and define cultural epochs. These technological limitations come first, shape our human experiences, and create subcultures that follow.

Distorted electric guitar sounds that define heavy metal music were once a nuisance derived from the limitation of guitar amplifier technology. Now, we have software tools, running on processors built from billions of transistors, that emulate the sound of a single vacuum tube, because tubes sound better.

Entire cultural epochs get wrapped up into layers of these technological limitations. To a discerning ear, for instance, the sound of recorded music of the 1980s is distinctive. With a little practice, an opening chord or drum lick is all it takes to instantly reference the particular spectral character of iron-oxide tape that was prevalent throughout recordings from this era. And how about that filter Huji, which makes your Instagram photos look like they were shot with a disposable camera?

In Moriarty's book, the technology-culture feedback comes full circle. Initially shaped and defined by the technological limits of making music louder (much, much louder), the subculture of heavy metal music now provides an opportunity to engage with, and comprehend, fundamental physics. ■

10.1126/science.aau3244

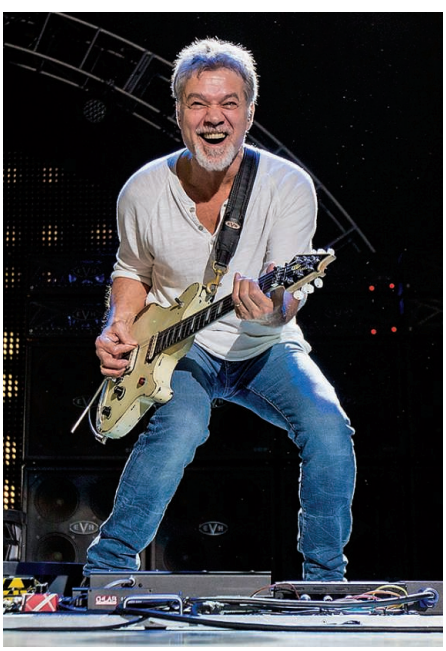


PHOTO: DANIEL KNIGHTON/GETTY IMAGES

Waveform interference is a key feature of the double-slit experiment and Eddie Van Halen's "brown sound."

GENETICS

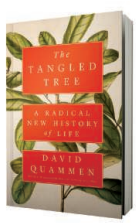
Lateral moves

An engaging history reveals the scientific struggle to understand horizontal gene transfer

By **Ivor T. Knight**

Eight percent of the human genome originated in virus genomes, including genes now essential to human life. This is just one insight gleaned from the current deluge of genome data that is providing ever more evidence for what we have long known to be true about microorganisms: The transfer of genetic material from one organism to another, or horizontal gene transfer, is an important source of variation and a factor driving evolution.

In David Quammen's new page turner, *The Tangled Tree: A Radical New History of Life*, the author reveals how new molecular techniques have come to revolutionize the way we understand evolutionary processes and how we classify life into coherent groups. In an accessible style that has won him accolades in the past, Quammen does a marvelous job of weaving together the scientific and human story of this revolution.



The Tangled Tree
David Quammen
Simon & Schuster, 2018. 487 pp.

In the first chapters, Quammen lays the groundwork, recounting how a tree came to be the preferred graphic illustrating order in the complexity of life. Here, he includes anecdotes about the historical figures involved and colorful descriptions of their trees. The geologist Edward Hitchcock's tree, for example, is described as looking "more like a windbreak of tightly placed Lombardy poplars grown to maturity along a roadway."

Early in the book, Quammen introduces Carl Woese, the pioneering scientist who used ribosomal RNA differences among microorganisms to discover a third kingdom of life (Archaea). Quammen meticulously researched Woese's life, conducting countless interviews with those who knew him and poring over his papers in the archives at the University of Illinois. He provides a particularly detailed account of the development of Woese's famous 1980 paper with George Fox that described the evidence for three kingdoms (1), even delving into the contentious discussion of authorship order among the two.

The reviewer is at the Department of Biology, Behrend College, Pennsylvania State University, Erie, PA, USA. Email: itk2@psu.edu

The stories of other well-known scientists weave in and out, including Lynn Margulis, with whom Woese battled bitterly over whether there are five or three kingdoms of life. We learn that, ironically, it was Woese's ribosomal RNA evidence that showed bacteria to be the ancestor of mitochondria in eukaryotes, consistent with Margulis's once controversial endosymbiotic theory.

The primary evidence for horizontal gene transfer, the reason for the tangles in the tree, comes from comparative molecular analysis. Here again, Quammen uses the stories of scientists to contextualize the technical explanations of bacterial gene transfer mechanisms, describing the exciting discoveries of Oswald Avery, Frederick Griffith, Tsutomu Watanabe, and Esther and Joshua Lederberg.

Like Quammen's previous works, this book is meant to be widely read and communicates the process of scientific inquiry as effectively as it does the results of the inquiry. The bitterness and grudges that existed between scientists passionately competing in the discovery process are described along with the selfless gestures and lifelong loyalties that often existed in the same relationships. On occasion, Quammen takes the reader aside to provide a direct insight, such as the fact that relationships among scientists are always shaped by personal chemistry: "It's a smallish world these scientists live in, much interconnected."

My only criticism is that some of the book's shorter sections strayed into topics that seemed a bit out of place for a primarily historical account. For example, there is a short section on CRISPR that did not seem especially relevant to the book's central themes. This, however, is a minor concern. Such diversions certainly did not detract from the overall reading experience.

In *The Tangled Tree*, David Quammen has once again crafted a delightful read on a complex and important subject. This story about the revolution in our understanding of the history of life is a tale well told—and one that has far-reaching implications. ■

REFERENCE

1. G. E. Fox *et al.*, *Science* **209**, 457 (1980).

10.1126/science.aau4096

Podolskie Tovtry National Park, Ukraine, is home to the city of Kamianets-Podilskyi and subject to intense human pressure.

PHOTO: ISTOCK.COM/YURIYVLASENKO

LETTERS

Edited by **Jennifer Sills**

Protected land: Many factors shape success

The design of effective protected areas (PAs) is widely debated. In their Report "One-third of global protected land is under intense human pressure" (18 May, p. 788), K. R. Jones *et al.* show that the "human footprint" inside the majority of the world's protected areas has increased and suggest that PA performance can be improved by "upgrading" PAs with increased restrictions on human activities. They link their metric to biodiversity conservation goals, but different components of the human footprint will vary widely in terms of impacts. For example, maintenance of pasture lands may be compatible with conservation in some PAs (*1*). Increased access to PAs (by way of roads and rivers, for example) also leads to diverse outcomes: Wildlife poachers have a different impact than visitors to sacred sites (*2*). Some components of the human footprint (such as roads) also affect indigenous lands, local resource rights, and cultural diversity, as areas of high biodiversity often correspond with high cultural diversity (*3*, *4*).

Conservation policies require an understanding of relevant drivers, not only the IUCN category and PA size as suggested by Jones *et al.* Recent syntheses of theory [e.g., (*5*, *6*)] point to a diverse set of factors that influence the effectiveness and ethical standing of conservation. Use of adaptive approaches, diverse and nested systems of governance, respect for indigenous rights and responsibilities, and the integration of diverse worldviews and knowledge systems have all been shown to influence conservation outcomes (*5*, *6*).

Prescribing changes to PA design and management requires outcome indicators that are linked to policy targets and accurately depict local realities (*7*). These indicators will be more just and effective if they incorporate priorities, worldviews, caretaking practices, and knowledge systems of stakeholders as components of the social-ecological system in which the PA is embedded (*7*–*9*). How we measure success matters. Lumping all human activity into one metric of "human pressure" may not help biodiversity conservation in the long run.

Michael C. Gavin,[1,2]* Joe McCarter,[3,4] Fikret Berkes,[5] Eleanor J. Sterling,[3] Nancy J. Turner[6]

[1]Human Dimensions of Natural Resources, Colorado State University, Fort Collins, CO 80521, USA. [2]Max Planck Institute for the Science of Human History, D-07745 Jena, Germany. [3]Center for Biodiversity and Conservation, American Museum of Natural History, New York, NY 10024, USA. [4]Melanesia Program, Wildlife Conservation Society, Suva, Fiji. [5]Natural Resources Institute, University of Manitoba, Winnipeg, MB R3T 2M6, Canada. [6]School of Environmental Studies, University of Victoria, Victoria, BC V8W 2Y2, Canada.
*Corresponding author.
Email: michael.gavin@colostate.edu

REFERENCES

1. Z. Molnár, F. Berkes, in *Reconnecting Natural and Cultural Capital*, M. L. Paracchini, P. C. Zingari, C. Blasi, Eds. (Publications Office of the European Union, 2018), pp. 183–194.
2. B. Verschuuren, N. Furuta, *Asian Sacred Natural Sites: Philosophy and Practice in Protected Areas and Conservation* (Routledge, 2016).
3. J. Pretty *et al.*, *Conserv. Soc.* **7**, 100 (2009).
4. UNESCO-SCPD Programme, Linking Biological and Cultural Diversity (2018); www.cbd.int/lbcd.
5. M. C. Gavin *et al.*, *Trends Ecol. Evol.* **30**, 140 (2015).
6. M. C. Gavin *et al.*, *Sustainability* **10**, 1 (2018).
7. E. J. Sterling *et al.*, *Nature Ecol. Evol.* **1**, 1798 (2017).
8. E. Sterling *et al.*, *Environ. Soc.* **8**, 63 (2017).
9. J. McCarter *et al.*, *Ecol. Soc.* **23** (2018).

10.1126/science.aau5168

Protected land: Threat of invasive species

In the Report "One-third of global protected land is under intense human pressure," (18 May, p. 788), K. R. Jones *et al.* conclude that 6 million square kilometers (32.8%) of protected land is under intense human pressure. Jones *et al.* assessed the intensity of human pressure with the human footprint index, which quantifies the proportional representation of agricultural land use, urbanization, and human infrastructure at a 1-km^2 resolution across the globe (*1*). However, this index underestimates the considerable impact of invasive alien species on the biodiversity of protected areas distant from any evidence of human activity or infrastructure.

Many of the protected areas identified by Jones *et al.* as experiencing low human pressure are in fact undergoing major ecosystem changes as a result of invasive alien species. These include extensive felling of native forest in southern Chile by North American beavers (*2*), introduced European mustelids jeopardizing the survival of endemic flightless birds in isolated mountain areas of southern New Zealand (*3*), and high densities of North African camels altering the remote outback ecosystems of western Australia (*4*). Even far-flung sub-Antarctic islands, with no resident human population, have seen degradation of biodiversity as a result of introduced rats and cats (*5*).

Current understanding of invasive alien species impacts in protected areas remains poor (*6*). Although species inventories may give an indication of the pressure posed by invasive alien species in protected areas, greater effort is required to quantify the impacts of invasive alien species in

Camels are an invasive species in the remote ecosystems of western Australia.

remote protected areas, control existing populations, and prevent further spread. Biological invasions are undoubtedly abetted by human infrastructure and land use, but they are able to spread far away from such locations remarkably rapidly (*7*). The human footprint index does not capture the intensity of invasive alien species impacts distant from human infrastructure and thus may provide a false sense of protected area integrity where, in fact, species and ecosystem functions are in serious decline.

Philip E. Hulme
The Bio-Protection Research Centre, Lincoln University, Canterbury, New Zealand.
Email: philip.hulme@lincoln.ac.nz

REFERENCES

1. O. Venter *et al.*, *Sci. Data* **3**, 160067 (2016).
2. D. A. Wardle, R. Bardgett, R. M. Callaway, W. H. Van der Putten, *Science* **332**, 1273 (2011).
3. C. F. J. O'Donnell, K. A. Weston, J. M. Monks, *NZ J. Ecol.* **333**, 31 (2017).
4. A. P. Woolnough, G. Edwards, Q. Hart, *Rangelands J.* **38**, 109 (2016).
5. J. Dawson, *Conserv. Biol.* **29**, 143 (2015).
6. P. E. Hulme *et al.*, *Conserv. Lett.* **7**, 459 (2014).
7. P. E. Hulme, *Science* **335**, 537 (2012).

10.1126/science.aau3784

Response

To respond to Gavin *et al.*, we emphasize that when evaluating protected area performance, it is important at the outset to have a clear definition of success. Under the International Union for Conservation of Nature's definition, successful protected areas "conserve the composition, structure, function and evolutionary potential of biodiversity" (*1*). Although some protected areas are clearly important for the preservation of cultural values or ecosystem services, these goals only apply if they "do not interfere with the conservation outcome" (*1*). For a protected area to successfully conserve the full range of biodiversity, it must therefore be free from those human pressures that negatively affect species and ecosystems.

Gavin and colleagues question whether the human footprint adequately represents human pressure on biodiversity within protected areas. The human footprint uses fine scale (1-km^2) data on eight different human pressures, including land use, infrastructure, population density, and human access, all of which have been shown to drive biodiversity declines, increase species extinction risk, and impede species movement (*2*, *3*). Although the human footprint does not directly map some important pressures, such as invasive species and climate change, and hence may underestimate total pressure, it includes many anthropogenic pressures that are known to seriously affect biodiversity (*4*). Given the overwhelming scientific consensus showing that human pressure is detrimental to biodiversity, it is sensible to assume that if pressures occur inside protected area boundaries, they are likely to compromise the biodiversity that those sites have been designated to conserve.

We agree with Gavin *et al.* that there is substantial local variation in the importance of different human pressures and that a diverse range of approaches is required to stop those pressures from threatening biodiversity within protected areas. We also agree that it is critical to consider local realities and different governance systems, to respect indigenous rights, and to promote social and ecological justice. However, regardless of the site-based approach used, effective conservation outcomes can only occur when the human pressures causing biodiversity declines are halted. Allowing humans to hunt and farm within protected areas in ways that degrade species populations and their habitat will, by definition, undermine the ability of protected areas to achieve their fundamental goal (*1*). For Earth's remaining biodiversity to have any chance, big changes in how we plan for nature conservation are now crucial (*5*), and a critical part of any future plan will be ensuring those places set aside to secure biodiversity actually work.

Hulme correctly points out that the human footprint does not explicitly map invasive species, and therefore likely underestimates their impacts within protected areas. In our Report, we included the caveat that "the human footprint does not account for all pressures affecting biodiversity, such as poaching or climate change." Indeed, a comprehensive map of human pressures would likely reveal no truly impact-free land within protected areas, or anywhere on Earth, as close inspection in even the most remote areas reveals trace pollutants (*6*), plastics (*7*), and signs of anthropogenic climate change (*8*).

We recognize that invasive species are a key threat for almost one-quarter of endangered species (*9*), and their exclusion from the human footprint is not a minor omission. However, the human footprint may implicitly include many pressures that are not explicitly mapped. In particular, human population density closely explains invasive species prevalence in protected areas (*10*), and roads provide a major conduit for the spread of exotic plants (*11*). Moreover, poaching is strongly correlated with roads and other forms of access (*12*). Through these relationships, the human footprint does, in part, act as a proxy for unmapped threats such as invasive species and poaching.

Nevertheless, a revised human footprint directly capturing a more comprehensive range of threats to biodiversity is clearly a future research priority. Beyond this, it is crucial to project future risks to biodiversity from threats such as climate change (*8*) and determine the most appropriate actions to help biodiversity survive these impacts. Importantly, one of the fundamental ways to manage global scale threats such as climate change is to stop more easily abatable threats, such as those considered in the human footprint (*13*).

In the era of big data and bottom-up, user-driven mapping approaches [such as OpenStreetMap (*14*)], developing a more comprehensive map of threats to biodiversity, including the presence of invasive species, is becoming increasingly possible. Given the accelerating rate of human encroachment on the natural environment (*15*), it is also becoming increasingly important.

Kendall R. Jones,[1,2*] Oscar Venter,[3] Richard A. Fuller,[2,4] James R. Allan,[1,2] Sean L. Maxwell,[1,2] Pablo Jose Negret,[1,2] James E. M. Watson[1,2,5]
[1]School of Earth and Environmental Sciences, The University of Queensland, St Lucia, QLD 4072,

PHOTO: SHERALEE STOLL/ALAMY STOCK PHOTO

Australia. [2]Centre for Conservation and Biodiversity Science, The University of Queensland, St Lucia, QLD 4072, Australia. [3]Ecosystem Science and Management Program, University of Northern British Columbia, Prince George, BC V2N 4Z9, Canada. [4]School of Biological Sciences, University of Queensland, Brisbane, QLD 4072, Australia. [5]Wildlife Conservation Society, Global Conservation Program, Bronx, NY 10460, USA.
*Corresponding author. Email: k.jones10@uq.edu.au

REFERENCES

1. N. Dudley, S. Stolton, P. Shadie, *Guidelines for Applying Protected Area Management Categories* (IUCN, Gland, Switzerland, 2008).
2. M. D. Marco *et al.*, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **369**, 20130198 (2014).
3. M. A. Tucker *et al.*, *Science* **359**, 466 (2018).
4. P. J. K. McGowan, *Nature* **537**, 172 (2016).
5. J. E. M. Watson, O. Venter, *Nature* **550**, 48 (2017).
6. A. S. Koziol, J. A. Pudykiewicz, *Chemosphere* **45**, 1181 (2001).
7. D. K. A. Barnes, F. Galgani, R. C. Thompson, M. Barlaz, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **364**, 1985 (2009).
8. B. R. Scheffers *et al.*, *Science* **354**, aaf7671 (2016).
9. S. L. Maxwell, R. A. Fuller, T. M. Brooks, J. E. M. Watson, *Nat. News* **536**, 143 (2016).
10. D. Spear, L. C. Foxcroft, H. Bezuidenhout, M. A. McGoch, *Biol. Conserv.* **159**, 137 (2013).
11. J. L. Gelbard, J. Belnap, *Conserv. Biol.* **17**, 420 (2003).
12. M. J. Shaffer, J. A. Bishop, *Trop. Conserv. Sci.* **9**, 525 (2016).
13. D. B. Segan, K. A. Murray, J. E. M. Watson, *Glob. Ecol. Conserv.* **5**, 12 (2016).
14. C. Barrington-Leigh, A. Millard-Ball, *PLOS One* **12**, e0180698 (2017).
15. J. E. M. Watson *et al.*, *Curr. Biol.* **26**, 2929 (2016).

10.1126/science.aau7317

TECHNICAL COMMENT ABSTRACTS

Comment on "Unexpected reversal of C_3 versus C_4 grass response to elevated CO_2 during a 20-year field experiment"

Julie Wolf and Lewis Ziska

Reich *et al.* (Reports, 20 April 2018, p. 317) assert that the responses of C_3 and C_4 grass biomass to elevated CO_2 "challenge the current C_3-C_4 [elevated CO_2] paradigm," but these responses can be explained by the natural history of the experimental plants and soils without challenging this paradigm.
Full text: dx.doi.org/10.1126/science.aau1073

Response to Comment on "Unexpected reversal of C_3 versus C_4 grass response to elevated CO_2 during a 20-year field experiment"

Peter B. Reich, Sarah E. Hobbie, Tali D. Lee, Melissa A. Pastore

Wolf and Ziska suggest that soil and species attributes can explain an unexpected 20-year reversal of C_3-C_4 grass responses to elevated CO_2. This is consistent with our original interpretation; however, we disagree with the assertion that such explanations somehow render our results irrelevant for questioning a long-standing paradigm of plant response to CO_2 based on C_3-C_4 differences in photosynthetic pathway.
Full text: dx.doi.org/10.1126/science.aau1300

TECHNICAL COMMENT

PLANT ECOLOGY

Comment on "Unexpected reversal of C_3 versus C_4 grass response to elevated CO_2 during a 20-year field experiment"

Julie Wolf* and Lewis Ziska

Reich *et al*. (Reports, 20 April 2018, p. 317) assert that the responses of C_3 and C_4 grass biomass to elevated CO_2 "challenge the current C_3-C_4 [elevated CO_2] paradigm," but these responses can be explained by the natural history of the experimental plants and soils without challenging this paradigm.

Reich *et al*. (*1*) explain that positive responses of plant biomass to elevated CO_2 have disappeared in C_3 grasses and appeared in C_4 grasses over the 20 years of the BioCON experiment. They assert, as do the authors of the associated Perspective (*2*), that these results challenge current expectations of C_3 and C_4 plant responses to elevated CO_2. Additional context should be made available to qualify this assertion. The pattern documented by Reich *et al*. can be explained by considering the natural history of the experimental plants and soils, without challenging general expectations of C_3 and C_4 grass responses to elevated CO_2 in the absence of other limitations.

The soil at the BioCON experimental field, which was not described in the paper or its supplement, was an excessively drained outwash sand, originally described as a Typic Udipsamment (*3*). When the experiments at the Cedar Creek Ecosystem Science Reserve (including BioCON) were initiated, topsoil was bulldozed away from the experimental field to remove existing savannah vegetation and seedbank. The field was then fumigated with methyl bromide (*4*). Remaining subsoil would have been composed of >90% sand, with little organic matter aside from coatings on sand mineral surfaces. Therefore, despite its 20-year duration, the BioCON experiment documents responses in a disturbed, developing soil. Although results from this experiment might be relevant to agricultural or urbanized soils (*5*), extrapolating to plant communities in mature, undisturbed soils worldwide is problematic (*6*, *7*).

Several publications from BioCON have demonstrated the importance of plant species identity, species richness, and functional group diversity in moderating responses to CO_2 and N enrichment [e.g., (*8*–*10*)], yet Reich *et al*. have used results from monocultures and four-species assemblages of only C_3 or C_4 grasses to make a broad statement about the general responses of C_3 and C_4 grasses to elevated CO_2. Despite variation among species, the C_4 grasses as a group tend to have higher nitrogen use efficiencies than C_3 grasses, reflecting their relatively smaller investment of N in photosynthetic carboxylation enzymes (*11*). Given the individual characteristics of the eight experimental grass species (Table 1) (*12*) and the initial seeding rate of 12 g seed/m^2 for all

Adaptive Cropping Systems Laboratory, USDA Agricultural Research Center, Beltsville, MD 20705, USA.
***Corresponding author. Email: julie.wolf2@ars.usda.gov**

Table 1. Traits of the grass species grown in the BioCON experiment (*12*).

Species	Common name	Active growth period	Life span	Growth form	Growth rate	Minimum root depth (inches)	Height at maturity (feet)	Low-growing grass?	Water usage	Drought tolerance	Fertility requirement
					C_4 species						
Andropogon gerardii	Big bluestem	Summer	Long	Bunch	Moderate	20	6	No	Low	High	Low
Bouteloua gracilis	Blue grama	Summer and fall	Moderate	Bunch	Rapid	16	1	No	Medium	High	Low
Schizachyrium scoparium	Little bluestem	Summer and fall	Long	Bunch	Moderate	14	3	No	Low	High	Low
Sorghastrum nutans	Indiangrass	Summer and fall	Long	Bunch	Moderate	24	6	No	Medium	Medium	Low
					C_3 species						
Elymus (Agropyron) repens	Quackgrass	Spring and summer	Moderate	Rhizomatous	Rapid	14	2.6	Yes	Medium	Low	Medium
Bromus inermis	Smooth brome	Spring, summer, and fall	Long	Rhizomatous	Moderate	12	2.5	Yes	Medium	Medium	High
Koeleria cristata (macrocantha)	Junegrass	Spring and fall	Short	Bunch	Rapid	20	1.5	No	High	High	Medium
Poa pratensis	Kentucky bluegrass	Spring, summer, and fall	Long	Rhizomatous	Moderate	10	1.5	Yes	High	Low	High

plots (*8*), the C_3 grasses would be expected to fill their plots faster than the C_4 grasses, which they did; the C_3 grasses grew greater overall biomass per plot than the C_4 grasses in the first few years (*1*). Short-lived positive responses of the C_3 plant biomass to elevated CO_2 might also be expected, because their higher overall leaf N contents allow for some dilution of N; increased aboveground biomass with diluted N under elevated CO_2 was indeed observed in early years (*13*).

The low fertility and water-holding capacity of the experimental soils, however, would favor the experimental C_4 grass species over time, because their fertility requirements—and, in some cases, their water requirements—are lower than in the C_3 species grown (Table 1). This advantage would not be obvious in the earlier years of the experiment, because of the slower growth rates and longer lifespans of the C_4 species relative to the C_3 plants grown (Table 1), but C_4 biomass would be expected to increase relative to C_3 plants over time in these conditions, with associated increases in organic matter additions to the soil from roots and litter. Eventually, the C_4 plots would accumulate more organic matter, providing carbon substrate for N-mineralizing microbes, as well as increased soil nutrient and water-holding capacity. These changes would alleviate N and H_2O limitations in the C_4 plots relative to the C_3 plots, leading to further enhancements in annual biomass accumulation and nitrogen mineralization rates. Therefore, the observed shifts in relative response to elevated CO_2 over time relate to the differential nutritional requirements also inherent in C_3 and C_4 photosynthetic metabolism, as well as to experimental conditions. Consequently, the observations do not disagree with general expectations of C_3-C_4 dynamics under elevated CO_2 when no other limitations are present.

In addition to methods used to prepare the site before treatment application, the statistical design of the free-air CO_2 enrichment (FACE) arrangement is also important. The authors state that the 88 one- and four-species, C_3-only and C_4-only plots analyzed constitute a fully factorial experiment. In fact, these plots are a subset of a broader experiment where three ambient and three elevated FACE rings provide blocked CO_2 treatments, and N, functional group, and species richness treatments are applied as fully factorial split-plot treatments within the blocks (*10*). The monoculture and four-species plots analyzed in this paper are unevenly distributed among three ambient and three elevated FACE rings. This unbalanced design usually means that the model sum of squares for overall treatment effects is not equal to the sum of individual treatment sums of squares, which precludes straightforward repeated-measures analysis (*14*). The authors do not describe how their statistical analysis addresses these limitations, nor do they mention any multiple-test correction to the *P* values obtained in this and earlier reports of nonindependent response variables over the years of the experiment.

We recognize that long-term (20-year) experiments such as BioCON are invaluable and provide unique information; however, before extrapolating to a broader, ubiquitous inference, attention should be given to both the statistical details and the broader context of the environmental limitations associated with the location. In low-*n* FACE experiments, as described here, underlying variability in soils, particularly nutrient availability, could have an outsized impact on result interpretation [e.g., (*15*)].

The general theory of C_3-C_4 dynamics under elevated CO_2, and its use in the Earth System Models that encode it, is a fundamental aspect of plant biological responses to rising carbon dioxide. Questioning this aspect should be encouraged. However, we would caution that additional research is necessary before the C_3-C_4 dynamic in response to CO_2 is invalidated.

REFERENCES

1. P. B. Reich, S. E. Hobbie, T. D. Lee, M. A. Pastore, *Science* **360**, 317–320 (2018).
2. M. Hovenden, P. Newton, *Science* **360**, 263–264 (2018).
3. D. Grigal, L. Chamberlain, H. Finney, D. Wroblewski, E. Gross, "Soils of the Cedar Creek Natural History Area" (Miscellaneous Report 123, University of Minnesota Agricultural Experiment Station, 1974).
4. J. Wolf, N. C. Johnson, D. L. Rowland, P. B. Reich, *New Phytol.* **157**, 579–588 (2003).
5. T. O. West, W. M. Post, *Soil Sci. Soc. Am. J.* **66**, 1930–1946 (2002).
6. P. Smith, *Global Change Biol.* **10**, 1878–1883 (2004).
7. J. Six, K. Paustian, *Soil Biol. Biochem.* **68**, A4–A9 (2014).
8. J. M. Craine, P. B. Reich, *New Phytol.* **150**, 397–403 (2001).
9. P. B. Reich *et al.*, *Nature* **410**, 809–812 (2001).
10. P. B. Reich *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10101–10106 (2004).
11. R. H. Brown, *Crop Sci.* **18**, 93–98 (1978).
12. Natural Resources Conservation Service, PLANTS Database (U.S. Department of Agriculture, 2018); http://plants.usda.gov.
13. P. B. Reich *et al.*, *New Phytol.* **150**, 435–448 (2001).
14. J. Gurevitch, S. T. Chester Jr., *Ecology* **67**, 251–255 (1986).
15. N. H. Oh, M. Hofmockel, M. L. Lavine, D. D. Richter, *Global Change Biol.* **13**, 2626–2641 (2007).

7 May 2018; accepted 11 July 2018
10.1126/science.aau1073

TECHNICAL RESPONSE

PLANT ECOLOGY

Response to Comment on "Unexpected reversal of C_3 versus C_4 grass response to elevated CO_2 during a 20-year field experiment"

Peter B. Reich[1,2]*, Sarah E. Hobbie[3], Tali D. Lee[4], Melissa A. Pastore[3]

Wolf and Ziska suggest that soil and species attributes can explain an unexpected 20-year reversal of C_3-C_4 grass responses to elevated CO_2. This is consistent with our original interpretation; however, we disagree with the assertion that such explanations somehow render our results irrelevant for questioning a long-standing paradigm of plant response to CO_2 based on C_3-C_4 differences in photosynthetic pathway.

In a thoughtful consideration of the mechanisms responsible for the unexpected reversal of C_3 versus C_4 grass community responses to elevated CO_2 observed over a 20-year period (*1*), Wolf and Ziska (*2*) make many excellent points. However, they inaccurately represent the interpretations and conclusions of our paper, include at least one key factual error, and come to several conclusions that we believe the evidence does not support.

The thesis of Wolf and Ziska is that our results (*1*) can be explained by considering the natural history of the experimental plants and soils "…without challenging general expectations of C_3 and C_4 grass responses to elevated CO_2 in the absence of other limitations." We agree that consideration of the natural history of these plants and soils can help to illuminate the mechanisms and patterns we observed, but nonetheless our long-term results do challenge predictions one would make purely from photosynthetic pathway implications. Moreover, we find curious the authors' assumption that general expectations of C_3 versus C_4 responsiveness to elevated CO_2 have historically been framed as relevant only when no other limitations are present (which in any case never or almost never occurs); instead, such differences between C_3 and C_4 species are considered to hold generically, across the full spectrum of limitations. Additionally, we dispute that we extrapolated these results to grasslands globally; instead, we stated (*1*) in the penultimate sentence that "Our results thus serve as a reminder that even the best-predicted short-term ecosystem responses to global change can yield mid-term (decades) to long-term (centuries) surprises, as complex responses and interactions may occur over time." Our final sentence further stated that "Determining whether the mid- to long-term responses demonstrated here are themselves broadly predictable represents a major unmet challenge for experimental and observational studies." Both statements suggest that extrapolating from short-term physiology to long-term biogeochemistry may be problematic; however, they do not suggest that the specific responses of our experimental communities are repeatable elsewhere.

Wolf and Ziska introduce three topics they think did not receive needed attention in our original paper: (i) the level of soil disturbance during establishment of the experiment, and thus the relevant ecological context for our study; (ii) the nature of the plant species used in the experiment, and what that might mean for interpreting the results; and (iii) whether elements of the statistical design and/or analyses were of concern. We address each in turn.

As Wolf and Ziska note, the soil at the experimental site is sandy, nutrient-poor and very well drained (*3*). It is not true, however, that "topsoil was bulldozed away from the experimental field to remove existing savannah vegetation and seedbank." The experiment was established in an old field and topsoil was not removed; prior to planting, the experimental area was tilled to a depth of 25 cm, fumigated with methyl bromide to eliminate seeds of undesired plant species, and the soil reinoculated with unfumigated soil suspended in water (*4*, *5*). It is true that the experiment is situated on disturbed soil (initially to institute agriculture long ago, and 20 years ago to establish the experiment). But much of the world's grasslands are also disturbed, having experienced changes in grazer identity, fire regime, woody species control, cropping, and/or pasture management that also directly or indirectly influence soil properties. More important in our view is asking whether and how CO_2-mediated feedbacks to the nitrogen cycle that occurred at this site might occur elsewhere in poor (or, for that matter, rich) soils. At present, we have little basis for knowing the answer. Thus, although we agree that "extrapolating to plant communities in mature, undisturbed soils worldwide is problematic," we believe that our results are highly relevant for soils with chronic nitrogen limitations, whether disturbed or undisturbed, akin to much of the world's grasslands, and much less relevant to agricultural systems, where nutrient inputs can easily swamp plant-driven feedbacks.

Wolf and Ziska discuss differences in perennial C_3 and C_4 grasses that may well be related to the different temporal patterns shown by the two functional groups (in both CO_2 levels) and suggest that these differences led to the shifting responses to CO_2 over time. We agree that a number of aspects of physiology and life history likely led to these C_3 grasses having high biomass in early years of the experiment, which then declined, and to the C_4 grasses slowly (if erratically) ramping up their biomass over time. What is unclear is whether these successional differences between the C_3 and C_4 plants influenced their responses to CO_2. Indeed, in future work, we plan to test a suite of interrelated hypotheses regarding whether plant-soil feedbacks and differences in resource use efficiency among C_3 and C_4 species contributed to the differing temporal patterns of effects

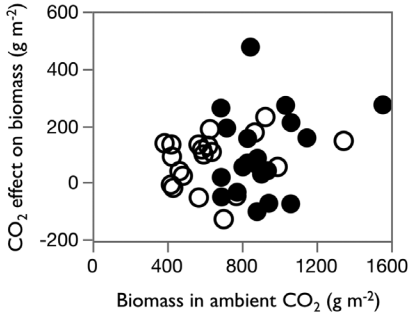


Fig. 1. The CO_2 effect on biomass in plots comprising C_3 grasses and C_4 grasses in relation to total biomass in ambient CO_2. Biomass (aboveground + belowground, 0 to 20 cm) in ambient CO_2 is shown for each year from 1998 to 2017 for plots comprising C_3 grasses (open circles) and C_4 grasses (solid circles); the CO_2 effect on biomass is expressed as the difference in biomass between elevated and ambient CO_2. Each point represents data pooled across N treatments and across monoculture and four-species plots (equally weighted) for each functional group (n = 22 plots for each functional group at each CO_2 level) in each year. There was no relationship ($P > 0.10$) between the CO_2 effect size and the ambient biomass for all data pooled or for either functional group alone.

[1]Department of Forest Resources, University of Minnesota, St. Paul, MN 55108, USA. [2]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, NSW 2753, Australia. [3]Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN 55108, USA. [4]Department of Biology, University of Wisconsin, Eau Claire, WI 54701, USA.
*Corresponding author. Email: preich@umn.edu

of elevated CO_2. Here, we evaluated one related question that we pursued in the early stages of analyzing the data prior to publication: whether the relative "vigor" (as measured by biomass in ambient treatment) of the C_3 versus C_4 grasses influenced their response to elevated CO_2. One might plausibly hypothesize that when a community is growing most vigorously, it might have the greatest capacity to use extra resources, such as elevated CO_2. The data, though, provide no evidence that in years with high biomass accumulation under ambient CO_2, the biomass response to elevated CO_2 was higher (for either the C_3 or C_4 group, alone or together) (Fig. 1). Nonetheless, we share the belief of Wolf and Ziska that ecosystem-scale changes driven by differences among C_3 and C_4 species could play a role in the observed and future responses to elevated CO_2. It makes no sense, however, for Wolf and Ziska to lodge this as a criticism of our paper, as the mechanisms they propose (involving leaf nitrogen status and feedbacks to soil nitrogen biogeochemistry) are consistent with our observations and interpretations. Thus, if ecosystem-scale temporal changes do play a role, they would likely help to explain and support our unexpected result, rather than suggest that we can continue to rely on the notion that C_4 grassland communities would be generally unresponsive to CO_2 while those dominated by C_3 grasses would generally show biomass enhancement. We hope that by elucidating the underlying mechanisms in the future, we will be in a better position to evaluate the generality of the responses we observed.

Wolf and Ziska also question three aspects of our statistical approach. First, we wish to clarify that despite their implication to the contrary, the 88 plots do in fact constitute a complete factorial of species number, functional group, CO_2 treatment, and N treatment. Second, it is true that plot-level replication is uneven among species richness levels (there are a total of three four-species plots and eight one-species plots per functional group at each unique CO_2 and N level), and Wolf and Ziska expressed concern that the unbalanced design produced sums of squares that are not straightforward to interpret. We did not explain explicitly in the original paper that the analyses we made are insensitive to this. The analyses were made in a statistical program (JMP 13.1) using a mixed model (with both fixed and random effects) that deploys a maximum likelihood approach; it does not calculate sums of squares, but partitions the variance and then uses that to calculate *F* statistics. It is robust to unbalanced sample sizes; hence, this concern is unwarranted. The third statistical concern involved our lack of multiple-test corrections to the *P* values. This concern is not applicable, as we only focused on the main effects and interactions from the analysis shown in table 1 of (*1*) and did not compare specific levels of factors. Regardless, visual examination and further analyses of the data shown in figure S2 of (*1*) support our conclusion. We tested whether the effect size (for both biomass and net N mineralization) was significantly related to year (a continuous variable), functional group, or their interaction. We found no significant main effects of year or functional group, but we did find significant year × functional group interactions ($P < 0.0001$ and $P = 0.0035$ for biomass and net N mineralization, respectively), supporting the interpretation that the response to elevated CO_2 changed over time and did so in opposing fashion for the C_3 and C_4 grass groups.

In summary, the questions raised about statistics (*2*) are not of concern, and we disagree with Wolf and Ziska about the appropriate context within which to view the experiment, but we agree that finding the appropriate context for field experiments is always challenging and should be done carefully. Perhaps most intriguing, Wolf and Ziska highlight important aspects of the way in which these North American grassland ecosystems change over time. Such changes might plausibly contribute to the reversal of responses to elevated CO_2 of these two functional groups over time such as we observed; but if they did, this would be entirely consistent with our interpretation that over time, plant-soil feedbacks can outweigh photosynthetic pathway differences in driving ecosystem responses to CO_2. Ecosystems change over time in complex ways that we are only beginning to understand. Even if we could fully explain why our experimental communities responded as they did over 20 years, which we cannot yet do, it would remain a post hoc explanation of an unpredicted and unexpected result. We hope for more unexpected results among the thousands of observational, experimental, and modeling studies researchers are engaged in globally, as such results often do a better job of illuminating gaps in our thinking and data than results that conform to expected outputs, and which thus may hide errors in our thinking, data, and modeling regarding underlying mechanisms (*6*).

REFERENCES AND NOTES

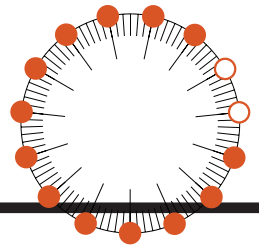
1. P. B. Reich, S. E. Hobbie, T. D. Lee, M. A. Pastore, *Science* **360**, 317–320 (2018).
2. J. Wolf, L. Ziska, *Science* **361**, eaau1073 (2018).
3. D. Grigal, L. Chamberlain, H. Finney, D. Wroblewski, E. Gross, "Soils of the Cedar Creek Natural History Area" (Miscellaneous Report 123, University of Minnesota Agricultural Experiment Station, 1974).
4. J. M. Craine, D. A. Wedin, P. B. Reich, *Global Change Biol.* **7**, 947–953 (2001).
5. P. B. Reich *et al.*, *New Phytol.* **150**, 435–448 (2001).
6. P. B. Reich, R. L. Rich, X. Lu, Y.-P. Wang, J. Oleksyn, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13703–13708 (2014).

28 May 2018; accepted 11 July 2018
10.1126/science.aau1300

RESEARCH

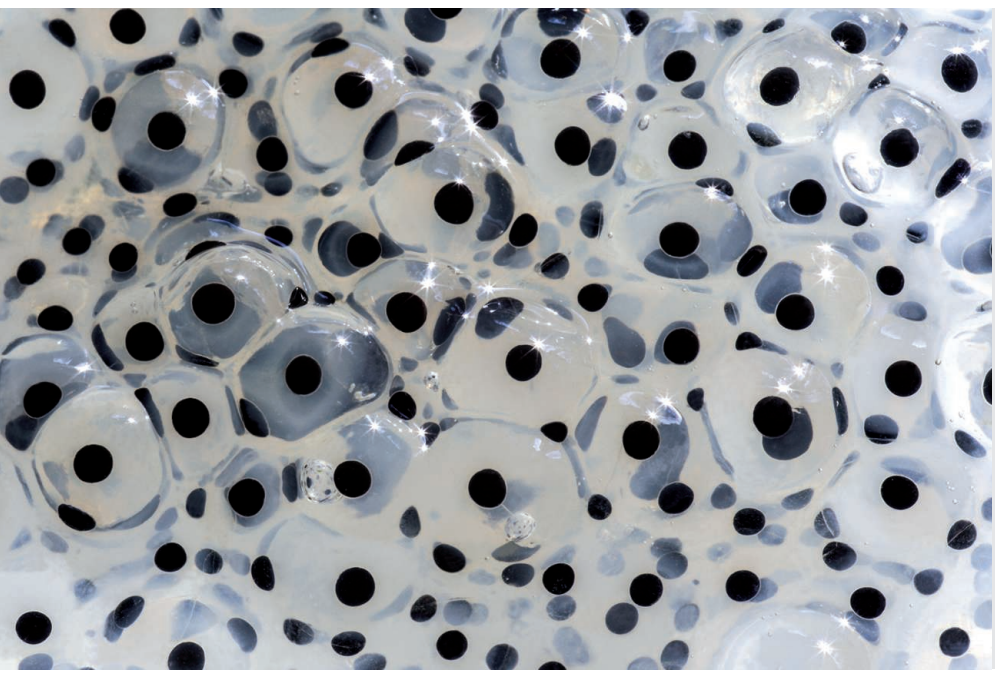
Kerr solitons for optical combs

Kippenberg et al., *p. 567*



IN *SCIENCE* JOURNALS

Edited by **Caroline Ash**



SIGNAL TRANSDUCTION

Visualizing a traveling wave of cell death

When diffusion is too slow for communication over long distances, cells can use waves of chemical activity. By using fluorescent probes and microscopy, Cheng and Ferrell show that in frog eggs (which are very large cells), waves of apoptotic signals can be seen passing through the egg cytoplasm. The pathways that trigger cell death have positive feedback loops that lead to self-regenerating waves. The speed of the waves (~30 micrometers per minute) is too fast to be explained by diffusion. —LBR

Science, this issue p. 607

Frog eggs are used to observe cell signals.

ANTHROPOLOGY

Peopling the Americas: Which way?

For much of the 20th century, archaeologists firmly believed that the initial pathway into the Americas was via the "ice-free corridor," a route from interior Alaska that snaked into the high plains of North America between two massive sheets of ice. Over the past 20 years, this consensus has been challenged and, for many, overturned, by data that support the "kelp highway" hypothesis—i.e., migration along the coastline of western North America. In a comprehensive review of genetic, archaeological, and paleoecological data, Potter *et al*. argue that both routes should be considered as viable pathways into the Americas. —MSA

Sci. Adv. 10.1126/sciadv.aat5473 (2018).

NEUROSCIENCE

Feedback reduces opioid prescriptions

Most people addicted to opioids began taking them because they were legally prescribed. Little attention has been paid to changing physicians' prescribing behavior. Using a randomized controlled trial format, Doctor *et al*. monitored the effect of notifying physicians who had a patient die of opioid overdose within 12 months of a prescription. The physicians received an injunction to prescribe safely from their county's medical examiner. This intervention led to reductions in high-intensity prescribing, reductions in the likelihood that an opioid-naïve patient received a prescription, and a reduction in overall cumulative opioid intake. —PRS

Science, this issue p. 588

RIVER NETWORKS

Expanding the role of rivers

The surfaces of rivers and streams are interfaces for a host of chemical exchanges with the atmosphere and biosphere. For instance, carbon dioxide outgassing from rivers is estimated to be equivalent to one-fifth of combined emissions from fossil fuel combustion and cement production. Allen and Pavelsky used satellite imagery to estimate the surface area of rivers and streams (see the Perspective by Palmer and Ruhi). The stunning map that they generated results in an upward revision, by about one-third, to the total surface area of rivers and streams on Earth. —BG

Science, this issue p. 585; see also p. 546

2D MATERIALS

Computers tease out interaction effects

Although graphene is often thought of as a material in which electron-electron interactions are negligible, some of its properties cannot be explained by such a simple picture. Tang *et al*. undertook comprehensive quantum Monte Carlo numerical calculations that consider both long-range and contact interactions in systems that, like graphene, have two-dimensional (2D) Dirac electrons. Different 2D Dirac materials systems, such as topological insulators and graphene on various substrates, reside in different parts of the resulting phase diagram. —JS

Science, this issue p. 570

CREDITS: (FROM TOP) T. J. KIPPENBERG *ET AL.*; ROSEMARY CALVERT/GETTY IMAGES

Published by AAAS

KIDNEY CANCER

Pediatric and adult kidney tumors differ

Understanding tumor origins and the similarities and differences between organ-specific cancers is important for determining treatment options. Young *et al*. generated more than 72,000 single-cell transcriptomes from healthy and cancerous human kidneys. From these data, they determined that Wilms tumor, a pediatric kidney cancer, originates from aberrant fetal cells, whereas adult kidney cancers are likely derived from a specific subtype of proximal convoluted tubular cell. —LMZ

Science, this issue p. 594

EVOLUTION

Adaptive conflicts with the modern world

Mammals evolved in terrestrial environments. Those that now live in the marine environment have had to adapt to the particular selective pressures that this environment imposes. Meyer *et al*. surveyed the genomes of several marine mammal species to identify regions of convergent change. Multiple losses of the *Paraoxonase 1* gene are evident in marine mammals, likely resulting from remodeling of lipid metabolism or antioxidant networks. The multiple occurrences of this loss of function across taxa indicate an evolutionary benefit. However, *Paraoxonase 1* is the primary mammalian defense against organophosphorus toxicity. Marine mammals may be at a great disadvantage in the Anthropocene if run-off of this agricultural product into the marine environment continues. —SNV

Science, this issue p. 591

Marine mammals are genetically vulnerable to some pollutants.

CREDITS: (FROM LEFT) ROBERT K. BONDE/USGS – GAINESVILLE; DAVE WALSH/VWPICS/ALAMY STOCK PHOTO

AUTOIMMUNITY

Pancreatic perturbation

Autoimmune pancreatitis (AIP) is difficult to diagnose and can sometimes be confused with pancreatic cancer, which presents with similar symptoms. AIP is an inflammatory disease involving elevated immunoglobulin G4 (IgG4); however, the target autoantigens are unknown. Previous work pointed to the extracellular matrix providing the target for IgG4. Shiokawa *et al*. now show that a truncated form of laminin 511 may be a major autoantigen in AIP. Half of the AIP patients that they investigated had antibodies against laminin 511, which were absent in healthy controls. Patient pancreatic tissues were positive for laminin 511, and immunization of mice with this protein induced AIP-like symptoms. —LP

Sci. Transl. Med. **10**, eaaq0997 (2018).

CELL BIOLOGY

Going through a phase

Neuronal communication at synapses relies on regulated neurotransmitter secretion. Neurotransmitters are stored in small vesicles that are organized in clusters within nerve terminals. On stimulation, the vesicles fuse with the presynaptic plasma membrane, but despite their tight packing, replacement synaptic vesicles are rapidly recruited. Vesicles newly reformed by membrane recycling randomly intermix with the clusters. Milovanovic *et al*. show that synapsin, an abundant synaptic vesicle–associated protein, organizes these vesicle clusters by liquid-liquid phase separation—like oil in water. —SMH

Science, this issue p. 604

IN OTHER JOURNALS

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

A melt river on Petermann Glacier in northwestern Greenland

ICE SHEETS

A responsive past

One of the ways that we might develop more accurate projections of the Greenland Ice Sheet's response to global warming is to investigate analogous periods in the past and use them to establish constraints for its future behavior. Reusche *et al.* investigated the history of two large outlet glaciers in the northwestern sector of the ice sheet across the Holocene by measuring surface-exposure ages of associated moraines. Their data shed light on the competing effects of generally warming surface air temperatures and discrete climate cooling events. These results help show how quickly the ice sheet can respond to both positive and negative centennial atmospheric temperature fluctuations. —HJS

Geophys. Res. Lett. 10.1029/2018GL078266 (2018).

MICROBIOTA

Establishing bad host relations

The human microbiota is a mixture of microorganisms that are maintained in symbiosis with the host. However, sometimes this symbiosis goes awry, causing pathogen outgrowth and disease. For example, periodically, *Staphylococcus aureus* emerges from the skin-resident microbiome as a disease-causing

A malaria-free childhood improves educational outcomes in Tanzania.

NEURODEVELOPMENT

Malaria challenges learning

Owing to interventions, the prevalence of malaria has declined in Tanzania. Klejnstrup *et al*. analyzed 15 years' worth of data on malaria rates, along with school achievement data for more than 200,000 children. They found that children born in periods with higher prevalences of malaria struggled more with numeracy and English literacy than counterparts at less risk of malaria. The effect of a malaria-free childhood on educational outcomes rivaled the effects of smaller class sizes and better-trained teachers. —PJH

PLOS ONE 10.1371/journal.pone.0199542 (2018).

pathogen. Boldock *et al.* show that such irruptions can be mediated by particulate peptidoglycan (PTG) expressed from the cell walls of nonpathogenic (commensal) skin bacteria. PTG promotes *S. aureus* survival in innate immune cells such as macrophages and neutrophils, and this facilitates systemic infection. This phenomenon is not mediated by established receptor pathways such as Nod1, Nod2, Myd88, or the NLPR3 inflammasome. —GKA

Nat. Microbiol. **3**, 881 (2018).

METABOLISM

High fat promotes overeating

Disagreement abounds over what constitutes a healthy diet and which components might provoke overeating and obesity in human populations. Hu *et al.* took a systematic approach to the latter question by feeding mice diets that differed in their proportion of fat (from 8 to 80%), protein (from 5 to 30%), carbohydrate (from 10 to 80%), or sucrose (from 5 to 30%). The only mice that ate more than they needed and became overweight were those eating a diet high in fat content. Furthermore, high-fat diets stimulated hedonic or reward pathways in the brain, as measured by gene expression. Because an equivalent comparison in humans would require a decade of study, we are unlikely to know how closely the results would be replicated in people. —LBR

Cell Metab. 10.1016/j.cmet.2018.06.010 (2018).

PLANT GENOMICS

Decoding parasitic plant genomes

Parasitism has evolved multiple times in plants and resulted in some major agricultural pests, including relatives of the morning glory family called dodder or strangleweed. To examine the effects of parasitism on the genome, Vogel *et al.* and Sun *et al.* respectively sequenced the genomes of field dodder (*Cuscuta campestris*) and Australian dodder (*Cuscuta australis*). Both studies identified major gene losses, likely facilitating the transformation into leafless, rootless plants unable to photosynthesize. Vogel *et al.* documented more than 50 examples of gene transfer into field dodder from their hosts. Sun *et al*. examined transcriptomes of the haustoria, which are specialized organs that allow dodder to extract water and nutrients from host plants. —LMZ

Nat. Commun. 10.1038/s41467-018-04344-z, 10.1038/s41467-018-04721-8 (2018).

GEOPHYSICS

Gravity tracking of a great earthquake

Changes in local gravity are connected to changes in subsurface mass. Panet *et al.* used GRACE (Gravity Recovery and Climate Experiment) satellite data to track large-scale deformation before and after the magnitude-9 Tohoku-Oki earthquake in Japan. Their analysis of the gravity data indicates that deformation occurred deep in the subducting slab starting several months before the earthquake. After the earthquake, mass transferred to the Pacific and Philippine Sea plate interiors. The large-scale and long-time-period gravity observations provide a distinctive perspective on this devastating and well-studied earthquake. —BG

Nat. Geosci. 10.1038/s41561-018-0099-3 (2018).

OPTICS

Lighting the path to AI

Artificial intelligence (AI) explores different architectures that strive to exploit the powerful information-processing capability of the brain. Artificial neural networks use connected artificial components (mimicking the function of neurons and synapses) to process information and perform complex tasks such as written and spoken language and image recognition from vast datasets. All the networks require training, however, which usually has been done by computer, and the process can be very time-consuming. Hughes *et al.* developed an optical method in which the training process is done with laser light propagating through a complex network of paths patterned into an optical chip. The results bring the prospect of an optical chip–based AI platform operating at the speed of light a step closer. —ISO

Optica **5**, 864 (2018).

PHOTO: JAKE LYELL/ALAMY STOCK PHOTO

ALSO IN *SCIENCE* JOURNALS

Edited by **Caroline Ash**

OPTICS

Shrinking optical metrology

The ability to generate laser frequency combs—light sources comprising equidistant laser lines spanning a large range of wavelengths—has revolutionized metrology and precision spectroscopy. The past decade has seen frequency combs being generated in optical microresonator circuits, offering the prospect of shifting precision metrology applications from the realm of national laboratories to that of everyday devices. Kippenberg *et al*. review the development of microresonator-generated frequency combs and map out how understanding and control of their generation is providing a new basis for precision technology. —ISO

Science, this issue p. 567

STRUCTURAL BIOLOGY

The first step in Hedgehog signaling

The Hedgehog (Hh) signaling pathway is important in embryogenesis; overactivation is associated with cancer. Central to the pathway is the membrane receptor Patched 1 (Ptch1), which indirectly inhibits a G protein–coupled receptor called Smoothened. This inhibition is relieved when Ptch1 binds the secreted protein Hh. Gong *et al*. report the cryo–electron microscopy structures of human Ptch1 alone and in complex with its Hh ligand at 3.9 and 3.6 Å, respectively. Both structures include two steroid-shaped densities, and mutational analysis indicates that the interaction between Ptch1 and Hh is steroid-dependent. —VV

Science, this issue p. 568

ORGANIC CHEMISTRY

A rapid screen for complex reactants

Chemists engaged in reaction discovery tend to report outcomes involving a few, relatively simple reactants. It remains a major challenge to fine-tune reported conditions when the reactants become more structurally complex, as often happens in pharmaceutical research. Lin *et al*. developed a protocol for rapidly screening different catalytic conditions for C–N coupling across a wide range of complex substrates. The product detection scheme relies on mass spectrometry of nanomole-scale reaction mixtures without any need for intervening chromatography. —JSY

Science, this issue p. 569

THERMAL CONDUCTIVITY

Moving the heat aside with BAs

Thermal management becomes increasingly important as we decrease device size and increase computing power. Engineering materials with high thermal conductivity, such as boron arsenide (BAs), is hard because it is essential to avoid defects and impurities during synthesis, which would stop heat flow. Three different research groups have synthesized BAs with a thermal conductivity around 1000 watts per meter-kelvin: Kang *et al*., Li *et al*., and Tian *et al*. succeeded in synthesizing high-purity BAs with conductivities half that of diamond but more than double that of conventional metals (see the Perspective by Dames). The advance validates the search for high-thermal-conductivity materials and provides a new material that may be more easily integrated into semiconducting devices. —BG

Science, this issue p. 575, p. 579, p. 582; see also p. 549

METABOLISM

Zipping up obesity

Chylomicrons are specialized particles that carry dietary fats from the intestine to the bloodstream for absorption into the body. Lacteals are lymphatic vessels that act as the highway for chylomicron transport, but it is unclear how passage occurs. Zhang *et al*. report that two endothelial cell receptors, neuropilin-1 (NRP1) and vascular endothelial growth factor receptor 1 (VEGFR1, also known as FLT1), are required to convert the entry spaces between lacteals from open junctions to closed, zipped structures (see the Perspective by McDonald). Mice that were fed a high-fat diet were subsequently rendered resistant to weight gain if NRP1 and FLT1 were inactivated. —PNK

Science, this issue p. 599; see also p. 551

ASTHMA

Nefarious neutrophil cytoplasts

In addition to DNA release, neutrophil extracellular trap (NET) formation can result in enucleated cells called cytoplasts. Krishnamoorthy *et al*. examined how neutrophil cytoplasts contribute to asthmatic inflammation in mouse models of allergic lung inflammation and in asthmatic patients. In mice, airway exposure to bacterial lipopolysaccharide with house dust mite allergen induced NET formation in the lung, which was associated with interleukin-17 (IL-17) production on subsequent exposure to allergen. Cytoplasts, rather than neutrophil DNA released in NETosis, triggered neutrophilia on allergen exposure, and cytoplasts alone were sufficient to induce IL-17 production by antigen-specific T cells. Cytoplasts also correlated with IL-17 levels in bronchoalveolar lavage fluid from severe asthmatics. —CNF

Sci. Immunol. **3**, eaao4747 (2018).

REVIEW SUMMARY

OPTICS

Dissipative Kerr solitons in optical microresonators

Tobias J. Kippenberg*, Alexander L. Gaeta, Michal Lipson, Michael L. Gorodetsky*

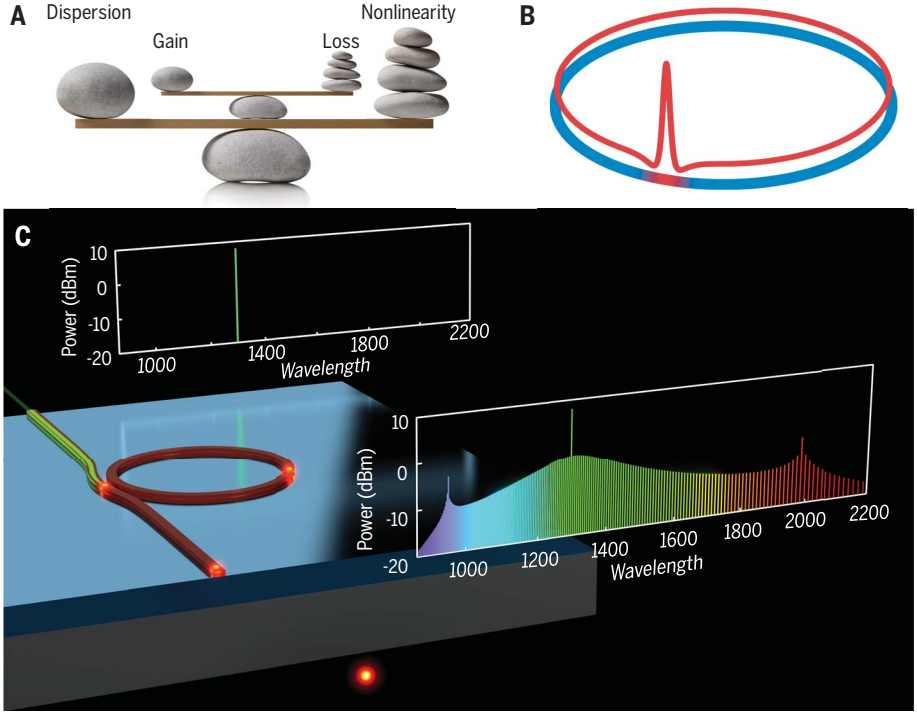
BACKGROUND: Laser frequency combs, which consist of equidistant laser lines, have revolutionized time-keeping, metrology, and spectroscopy. Conventional optical frequency combs based on mode-locked lasers are still mostly confined to scientific laboratories. In recent years, there has been progress in the development of optical frequency combs based on compact, chip-scale microresonators ("microcombs"), with such microcombs operating in the dissipative soliton regime. Dissipative solitons rely on a double balance of nonlinearity and dispersion as well as dissipation and gain and are an example of self-organization in driven dissipative nonlinear systems. Dissipative temporal solitons are providing the long-sought-for regime of coherent, broadband microcomb spectra and in addition provide an experimental setting in which to study dissipative soliton physics. Microcombs are now capable of producing coherent, octave-spanning frequency combs, with microwave to terahertz repetition rates, at low pump power, and in chip-scale devices and have been used in a wide variety of applications, owing to bandwidth and coherence provided by the dissipative temporal soliton states.

ADVANCES: Underlying these recent advances has been the observation of temporal dissipative Kerr solitons (DKSs) in microresonators, which represent self-enforcing stationary localized solutions of a damped, driven, and detuned nonlinear Schrödinger equation, which was originally used to describe spatial self-organization phenomena. They correspond to solitons (localized patterns) in "open" systems—that is, systems that exhibit dissipation. DKSs, opposite to fiber solitons, therefore rely on a double balance of nonlinearity and dispersion as well as parametric gain and loss and depend on the laser-cavity detuning as a control parameter. Methods have been established that enable the reliable generation of such DKSs in a wide range of microresonator platforms, including platforms based on silicon nitride (Si_3N_4) that are compatible with photonic integration. In addition, a variety of previously unknown and nonanticipated soliton effects have been observed, such as soliton crystallization, Raman-Stokes solitons, and previously unseen soliton breather dynamics. Moreover, dissipative solitons have been shown to be suprisingly robust against imperfections in the resonator mode structure. Dissipative soliton states in microresonators have triggered a large number of applications, including in terabit-coherent optical communications, atomic clocks, ultrafast distance measurements, dual-comb spectroscopy, photonic integrated frequency synthesizers, and the calibration of astrophysical spectrometers for exoplanet searches.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aan8083

OUTLOOK: The reliable generation of DKSs in microresonators has established a nascent research field at the interface of soliton physics, frequency metrology, and integrated photonics and provided impetus to microcomb sources. Emerging frontiers include using advances in nanofabrication and materials synthesis to realize ultralow-propagation-loss photonic nanostructures with unusual dispersion properties, which could explore dissipative solitons in new and unexplored parameter regimes and allow the synthesis of even broader spectra that in time domain could correspond to single-cycle pulses and whose spectral bandwidth could be extended to the mid-infrared or visible range. Beyond the narrow class of materials used for DKSs so far (Si, MgF_2, SiO_2, and Si_3N_4), many other materials exist with distinct advantages, such as III-V semiconductors, which are already widely used in light-emitting or laser diodes. Beyond existing applications, DKSs could be applied to optical coherence tomography, Raman spectral imaging, chip-scale atomic clocks based on optical transitions, or ultrafast photonic analog-to-digital conversion and have a potential to make frequency metrology and spectroscopy ubiquitous. ■



CREDIT: SIMON HOENL, EPFL

DKSs in optical microresonators. (**A**) Principle of DKSs, representing a double balance of dispersion and nonlinearity as well as (parametric) gain and cavity loss. (**B**) Optical field envelope of a single DKS, containing the localized soliton on top of a continuous-wave (CW) background field. (**C**) Graphic image of dissipative soliton formation in a CW laser–driven photonic chip–based microresonator, generating a continuously circulating DKS, which in frequency space corresponds to a coherent optical frequency comb. The optical frequency comb is shown with two dispersive waves that arise from higher-order dispersion.

This list of author affiliations is available in the full article online.
*Corresponding author. Email: tobias.kippenberg@epfl.ch (T.J.K.); mg@rqc.ru (M.L.G.)
Cite this article as T. J. Kippenberg *et al.*, *Science* **361**, eaan8083 (2018). DOI: 10.1126/science.aan8083

REVIEW

OPTICS

Dissipative Kerr solitons in optical microresonators

Tobias J. Kippenberg[1]*, Alexander L. Gaeta[2], Michal Lipson[3], Michael L. Gorodetsky[4,5]*

The development of compact, chip-scale optical frequency comb sources (microcombs) based on parametric frequency conversion in microresonators has seen applications in terabit optical coherent communications, atomic clocks, ultrafast distance measurements, dual-comb spectroscopy, and the calibration of astophysical spectrometers and have enabled the creation of photonic-chip integrated frequency synthesizers. Underlying these recent advances has been the observation of temporal dissipative Kerr solitons in microresonators, which represent self-enforcing, stationary, and localized solutions of a damped, driven, and detuned nonlinear Schrödinger equation, which was first introduced to describe spatial self-organization phenomena. The generation of dissipative Kerr solitons provide a mechanism by which coherent optical combs with bandwidth exceeding one octave can be synthesized and have given rise to a host of phenomena, such as the Stokes soliton, soliton crystals, soliton switching, or dispersive waves. Soliton microcombs are compact, are compatible with wafer-scale processing, operate at low power, can operate with gigahertz to terahertz line spacing, and can enable the implementation of frequency combs in remote and mobile environments outside the laboratory environment, on Earth, airborne, or in outer space.

Solitons are robust waveforms that preserve their shape upon propagation in dispersive media and can be found in a variety of nonlinear systems, ranging from hot plasma (*1*), ferrite films, fiber optics (*2*), cold atoms (*3*), hydrodynamics, and biology to cloud and sand dune formation. Although the initial promises of optical solitons for increased bandwidth of telecommunication were not implemented, dissipative solitons circulating in cavities of mode-locked lasers (*4*) are actively used. Similar to the physics of open quantum systems, the "open" soliton systems—dissipative solitons (*5*)—are highly relevant to actual experimental systems, in which dissipation cannot be neglected. This is in contrast to the mathematical treatment of solitons that have focused on conservative, integrable systems. Dissipative solitons (*5*) balance loss and gain in an active media, along with the balance of nonlinearity and dispersion (Fig. 1). Although dissipative optical solitons are known to occur in mode-locked lasers (*4*), dissipative optical solitons can also occur in systems that have parametric gain—gain from four-wave mixing that is the underlying process that enables Kerr frequency comb (*6*, *7*) formation in microresonators. Such dissipative solitons sustained by parametric gain were first studied in fiber cavities (*8*). In 2014, it was observed that such dissipative Kerr solitons (DKSs) can spontaneously form in parametrically driven Kerr frequency combs in optical microresonators (*9*). In these systems, optical sidebands are generated by means of four-wave mixing processes (*10*, *11*) and undergo a self-organization process that leads to the emergence of a soliton pulse train, which can mathematically and rigorously be mapped (*12*) to the Lugiato-Lefever equation (LLE) (*13*–*15*), an equation extensively studied in applied mathematics for decades.

The equation was originally developed to describe spatially localized pattern formation in driven nonlinear systems, in which it can lead to the formation of "dissipative structure." Since their observation in microresonators, such DKSs have been generated in a wide variety of microresonators, ranging from bulk crystalline (*9*, *16*) and silica microdisks (*17*) to photonic chip-scale devices (*17*–*20*), under both continuous-wave (CW) and pulsed excitation (*21*). In the frequency domain, the pulse train constitutes an optical frequency comb (*22*–*24*). Such optical frequency combs, pioneered by using mode-locked lasers based on pulse trains, are key ingredients of optical atomic clocks and have proven invaluable for a wide range of scientific and technological applications [for example, reviewed in (*25*)]. Although optical comb technology has matured since its first introduction two decades ago based on mode-locked lasers, and is commercially available, such devices are typically not amenable to wafer-scale integration and have comparatively narrow line spacing in the hundreds of megahertz range. DKSs in microresonators have provided a route to compact low power in chip-frequency comb generation. Although Kerr combs have been known for more than one decade (*7*) [and have been reviewed in (*6*, *26*–*28*)], the discovery of the DKS regime (*9*) has unlocked their full potential by providing a route to broadband and fully coherent microresonator-based frequency combs, overcoming earlier challenges of low coherence (*29*–*31*). Such soliton-based microcombs in chip-integrated microresonators have achieved low-power, octave-spanning frequency combs in various spectral windows that now encompass the near-infrared (*32*, *33*), telecommunication (*34*, *35*), and mid-infrared spectral window (*18*, *36*), with repetition rates from only a few gigahertz (*37*) to terahertz. The observation of DKS in microresonators yields a merging of soliton physics and high-precision frequency comb applications, stimulating a renaissance in dissipative soliton research and enabling many technological applications. Soliton microcombs have already been applied succesfully to dual-comb spectroscopy in the near- (*38*–*40*) and mid-infrared (*36*), scanning comb spectroscopy (*41*), as well as the demonstration of a self-referenced comb for the counting of optical frequencies [both with (*42*) and without external broadening (*43*)]. Likewise, soliton microcombs have been used in massively parallel communication (*44*), in pairs at both the transmitter and receiver side, for low-noise microwave generation (*45*) as well as for chip-scale dual-comb–based light detection and ranging (LIDAR) (*46*, *47*). Moreover, soliton microcombs have been used to realize an all integrated photonics-based chip-frequency synthesizer (*48*) and been used as a microphotonic astrocomb for exoplanet searches (*49*, *50*). Soliton microcombs exhibit surprisingly rich nonlinear dynamics and have led to the observation of a variety soliton dynamics effects, such as bright soliton formation, dark pulses, flat-topped "platicons" formation (*51*), soliton Cherenkov radiation (*20*), Raman self-frequency shift (*52*–*54*), Stokes soliton (*55*), and breather solitons (*56*–*60*). Moreover, the inherently multimode nature of microresonators gives rise to new and unanticipated dynamics, such as avoided-crossing–induced dispersive wave formation (*61*, *62*) or soliton-intermode breathing dynamics. Likewise, soliton crystallization (*63*) and switching (*64*) have been observed to occur. Today, a growing understanding of the nonlinear dynamics in soliton microcombs has been established, and new dynamics have been observed whose exploration continues to develop.

Dissipative Kerr solitons

The topic of this Review is a soliton that has long been theoretically discussed (*13*–*15*) but only recently experimentally observed: temporal DKSs, in which losses in a passive nonlinear optical cavity are compensated by a parametric interaction via a CW external laser pump in a resonator containing a Kerr nonlinearity (*10*). Such DKS states, first studied in fiber cavities (*8*), have

[1]École Polytechnique Fédérale de Lausanne (EPFL), Institute of Physics, Lausanne, CH-1015, Switzerland. [2]Department of Applied Physics and Applied Mathematics, Columbia University, NY 10027, USA. [3]Department of Electrical Engineering, Columbia University, NY 10027, USA. [4]Faculty of Physics, Lomonosov Moscow State University, Moscow, 119991, Russia. [5]Russian Quantum Center, Moscow, 143025, Russia.
*Corresponding author. Email: tobias.kippenberg@epfl.ch (T.J.K.); mg@rqc.ru (M.L.G.)

recently been discovered (*9*) as states in Kerr frequency combs (*7*). Although DKS formation has similarities to soliton mode-locking in femtosecond lasers, it does not require additional saturable absorbers to stabilize them, and they differ fundamentally because the pump laser frequency is a part of the soliton spectrum. The external coherent pump provides a central control parameter of the soliton and in addition constitutes one of the comb lines—which has no counterpart in conventional mode-locked lasers. Early interest in optical cavities filled with a nonlinear medium and pumped by means of CW source was initially stimulated by optical bistability and its application to optical switching (*65*), demonstrated by using iterative discrete-time input-output relations (known today as Ikeda map) so that such resonators exhibit multistability. However, it was found afterward (*66*) that this system can also support very localized optical pulses. If the intensity in the ring resonator changes only slightly per round trip, then the system can be described by a mean-field equation with periodic boundary conditions (*67*). If the medium has Kerr nonlinearity and quadratic dispersion, the mean field equation has the form of a nonlinear Schrödinger Equation (NLS) (*68*) with dissipative and driving terms. DKSs were first studied and observed with externally injected pulses (*8*). Although these experiments conjectured (*8*) that soliton formation may play a role, only later experiments revealed first evidence of such a regime, in terms of a transition from chaotic to low-phase-noise Kerr comb states (*30*), as well as evidence of pulse formation (which is consistent with a parameter regime allowing for soliton formation) (*69*). The first experiments that unambiguously identified DKSs in microresonators found DKS states by recording the transmission during Kerr comb formation. Strikingly, a series of steps on the red-detuned side of the resonance was observed. These steps, as is now well understood, are a sign of soliton formation and correspond to the continued reduction of the number of solitons in the cavity from $N, N-1, \ldots$. Stably accessing soliton states requires overcoming the thermal instabilities associated with the drop in intracavity power and led to the development of a wide range of techniques, so that now stably accessing Kerr soliton states can occur in a wide range of microresonator platforms.

The formation of DKSs in a crystalline microresonator are shown in Fig. 1 (*9*). A fundamental analytical property of DKSs is that they require a red-detuned pump laser from resonance, as also shown experimentally by using a Pound-Drever-Hall error signal. Although a red-detuned laser excitation in microresonators is conventionally unstable because of the thermal nonlinearity, the presence of a DKS can in turn self-stabilize the system for red detuning. This, however, necessitates overcoming transient thermal effects that are associated with the discrete steps in cavity transmission. Understanding and overcoming the thermal instability in the red-detuned soliton-formation regime, as first accomplished with optimized frequency tuning schemes (*9*), has been a pivotal experimental technique to enable stable DKS formation in microresonators. Single DKSs can equally be accessed in fully planar architectures (*70*), notably in photonic-chip–based silicon nitride (Si_3N_4) resonators (*71*), despite the fact that this platform has a significantly lower quality (Q) factor and can exhibit dispersion imperfections and strong thermal effects. Photonic-chip–based microresonators based on Si_3N_4 are amenable to wafer-scale manufacturing and integration and, because of their higher nonlinearity (compared with that of crystals or silica), enable broadband DKS (*72*) with high repetition rates (>100 GHz). Although tuning into soliton states in integrated Si_3N_4 microresonators has also been achieved with a slow laser-tuning method (*73*), the strong thermal effects have lead to the development of different methods, such as "power kicking" (*74*), fast tuning by use of heaters (*19*), and most recently, use of single-sideband modulator schemes (*75*). Using these techniques, DKSs have been generated in a wide variety of microresonators, ranging from bulk crystalline (*9*, *45*) and silica microdisks (*17*) and microspheres (*76*) to photonic chip-scale devices in Si and Si_3N_4 (*20*, *77*) and fiber cavities (*21*), under both CW and pulsed excitation (*21*). Moreover, schemes have been devised for soliton feedback stabilization, based on the soliton power (*78*), or by using the effective laser detuning of the soliton state (*79*).

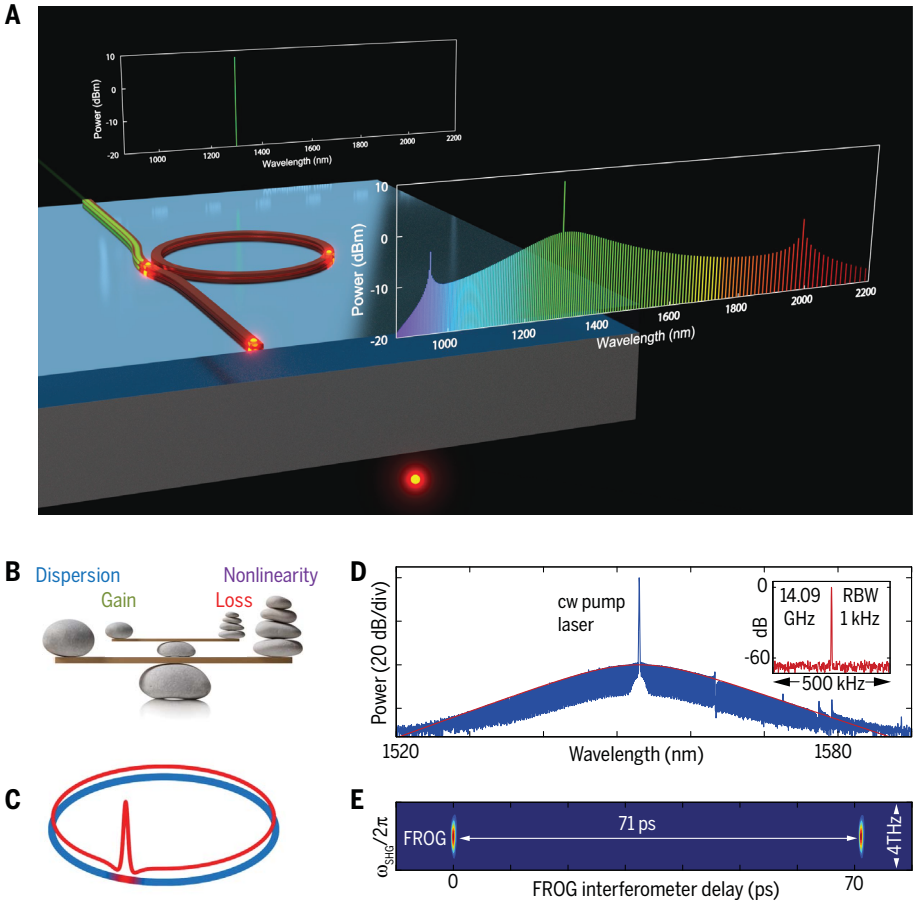


Fig. 1. Soliton microcombs technology. (**A**) Graphic image of dissipative soliton formation in a CW laser-driven crystalline WGM microresonator, enabling conversion of a CW laser to a train of DKS pulses. (**B**) Principle of DKSs, representing a double balance of dispersion and nonlinearity as well as (parametric) gain and cavity loss. (**C**) Field envelope of a temporal soliton. (**D**) Experimental optical spectrum of a DKS in a crystalline MgF_2 resonator. (Inset) Detected microwave beatnote of the soliton pulse train. (**E**) Frequency-resolved optical gating spectrum showing localized optical pulses, separated by the soliton-microcomb line spacing. [Images are adapted from (*9*)]

Dissipative soliton regime in Kerr frequency combs

We briefly review the physics of the parametric process in microresonators, discussed in detail in (*30*, *80*). Kerr frequency combs were initially discovered in silica microtoroids, and experiments proved that the parametrically generated (*11*, *81*) sidebands were equidistant to at least one part in 10^{-17} as compared with the optical carrier. In these early experiments, the combs repetition rate was in the terahertz range, and a femtosecond-laser frequency comb was used to bridge and verify the equidistant nature of the teeth spacing. It is today understood that such highly coherent combs only exist in certain regimes.

Physics of Kerr comb formation

The resonance frequencies of a microresonator can be Taylor-expanded around the pumped mode ω_0 so that

$$\omega_\mu = \omega_0 + \sum_j D_j \mu^j / j! \qquad (1)$$

where $D_1/2\pi$ is the mean free spectral range (FSR), D_2 is the group velocity dispersion (GVD) (positive for anomalous GVD), and $j \in Z$. If the FSR (because of D_2 in Eq. 1) increases with frequency, the resonator has anomalous GVD, as required for parametric oscillations and soliton formation. It is often useful to introduce the integrated dispersion, which describes the deviation of a given resonance from an equidistant frequency grid,

$$D_{\mathrm{int}}(\mu) = \omega_\mu - (\omega_0 - D_1\mu)$$

For a microresonator with anomalous GVD, the third-order Kerr nonlinearity leads to a nonlinear coupling between different modes. The nonlinear coupling coefficient $g = \hbar\omega_0^2 c n_2 / n^2 V_0$ is defined as a Kerr frequency shift per photon; n and n_2 are the refractive and nonlinear optical indices, respectively; V_0 is the effective (nonlinear) volume of the pumped mode; and c is the speed of light.

Parametric oscillation (*11*, *81*)—the emergence of symmetrically spaced signal and idler sidebands around the pump—occurs when the parametric gain exceeds the cavity decay rate. The threshold condition is equivalently understood as a Kerr-induced shift that is comparable with the cavity decay rate and thus the onset of cavity bistability ($g \cdot \bar{n}_c \sim \kappa/2$, where $\bar{n}_c$ is the number of photons in the mode). Unlike conventional lasers in which the threshold (*11*, *81*) scales as $\propto V_0/Q$, the threshold for parametric oscillation scales with $\propto V_0/Q^2$, which highlights the dramatic reduction of parametric threshold possible for ultrahigh-Q microresonators. When scanning the laser into resonance from the blue-detuned side (Fig. 2C), the first pair of sidebands that oscillate are those that are closest to the maximum of the parametric gain curve. The sideband number is approximately

$$\mu_i \approx \sqrt{\frac{\kappa}{D_2}}$$

If this distance corresponds to one single FSR, the subsequent cascade leads to fully coherent frequency combs. Such combs are referred to as "natively spaced comb," "Turing rolls," or (coherent) modulation instability (MI) combs. This coherent operating regime was initially observed in optical silica microtoroids (*7*). The coherence in these systems is achieved because the comb formation leads to a native comb, whose spacing corresponds to a single FSR. In this scenario, one can create fully coherent and relatively broadband-frequency combs, which has been the case for toroid resonators in the telecommunication band or crystalline resonators in the mid-infrared spectra range (*82*), and generally integrated resonators with a large FSR (*31*), which lead to a large accumulated dispersion parameter D_2. Not all platforms, however, yield intrinsically low phase noise. It was observed that the comb formation can give rise to low coherence states—in particular, in integrated platforms based on, for example, Si_3N_4, Hydex glass, Si, AlN, or AlGaAs. Counterintuitively, however, even for ultrahigh-Q resonators, such as crystalline resonators or silica disks, low coherence was observed, which is associated with subcomb formation.

Subcomb formation and chaotic comb states

In the case of resonators that have low Q (such as Si_3N_4 and other integrated resonators) or low dispersion (such as MgF_2 crystalline resonators

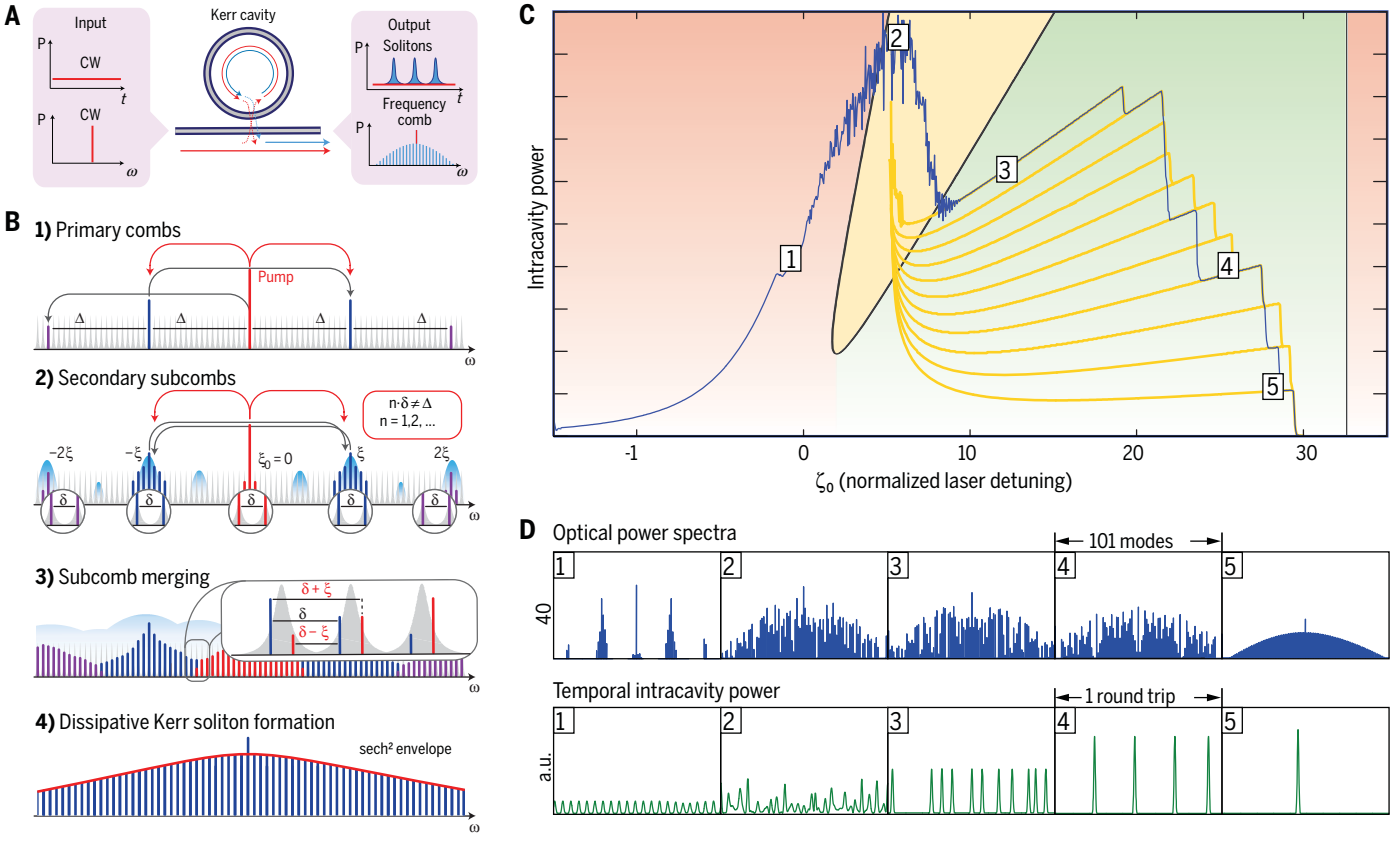


Fig. 2. Numerical simulations of DKSs in optical microresonators. (**A**) Temporal and spectral input and output from a CW laser–driven resonator supporting DKS. (**B**) The mode proliferation in the case of a resonator that exhibits subcomb formation, with eventual transition to DKS. (**C**) Intracavity field as a function of the laser detuning. Shown are the regions of modulation instability (marked "1"), breather soliton (yellow, marked "2") and stable soliton formation (green, marked "3" to "5"). Different trajectories corresponding to multiple simulations are shown in yellow (bold line). The different steps designate transitions between different soliton states. (**D**) intracavity waveform corresponding to the different chaotic MI, breather solitons, and stable DKS states. [Images are adapted from (*9*)]

in the telecommunication band), the first pair of sidebands (generated when scanning a pump laser into resonance from the blue-detuned side) are separated by many FSRs, which leads to the formation of subcombs (because $\mu_i = \sqrt{\frac{\kappa}{D_2}} \gg 1$).

In this process, depicted in Fig. 2, first a primary comb is generated with a line spacing Δ. Upon increasing pump power, secondary sidebands around the primary sidebands are generated, leading to the formation of subcombs. The initial subcombs all share the same repetition rate δ. Importantly, the primary and secondary spacing do not need to be commensurate. Therefore, once the pumping power is further increased, the subcombs merge and lead to the counterintuitive situation in which more than one single comb tooth can occupy a cavity resonance (*30*). This scenario leads the comb to exhibit beatnotes that can exhibit multiplets, which for sufficiently strong pump power merge into broad beatnotes. This comb state therefore exhibits low coherence, making it unsuitable for metrology. Although the comb spectral envelope of these combs is recorded with an optical spectrum analyzer to be spectrally smooth, the underlying intracavity waveform in the chaotic MI regime is vastly varying. Many early reported comb spectra (including photonic integrated resonator platforms or crystalline resonators in the telecommunication band) are chaotic in origin and exhibit low coherence. It was first demonstrated in Si_3N_4 microresonators (*30*) that transitions to low noise also exist. Such a transition to low noise state was shortly thereafter also observed in silica resonators (*83*), and evidence was found that the coherent regime is concomitant with pulse trains (*69*). These results indicated that the major challenge and roadblock of high noise could be overcome, unlocking the full potential of chip-scale frequency combs. Today, it is understood that these transitions are likely associated with DKSs (*9*, *13*, *15*), which give rise to coherent comb states. Indeed, although the LLE provides an accurate description of soliton states, many phenomena require corrections beyond the established models and are to be explored in Kerr microresonators.

Transition to the DKS regime

An unusual and unexpected observation was experimentally made in crystalline resonators (*9*) when analyzing the scan across the cavity resonance during the comb formation process: an unusual set of discrete step–like behavior in the transmitted power occurred, which exhibited a remarkably regular and repeatable step height. Transient measurements of the comb's beatnote in this regime showed low phase noise, indicating a coherent comb-formation regime. This observation confirmed an earlier theoretical prediction (*84*) that used numerical simulation of the laser-pump scan based on the coupled-mode equations (CMEs) (*85*). These simulations (*9*) equally revealed that during the laser scan across the resonance, unexpectedly sharp and discrete step–like transitions to low-noise states appear (Fig. 2). The numerical simulation predicted that the regions with discrete steps are associated with solitons inside the cavity, and the discrete steps in the transmission trace are associated with the annihilation of individual solitons, one by one. The numerical simulation uses the bare cavity detuning as a parameter, which does not correspond to the actual (effective) detuning from the resonance because of the Kerr frequency shift caused by the pump. Actual soliton formation can only occur when the laser transitions to the red-detuned side of the cavity resonance, where the intracavity field is bistable (Fig. 3A). The simulation reveals the primary sidebands, subcomb formation, and chaotic MI, followed by the formation of stable DKS inside the cavity. The numerical simulation was repeated multiple times, and the yellow curves in Fig. 3A show the evolution of all possible trajectories, whereas blue denotes the numerical simulation of a single laser scan trajectory. Although numerical simulations of dissipative solitons using a mean field model equation (the LLE) have been carried out extensively, simulations of the actual laser scanning process relevant to the microresonator case are a new development and pivotal in the understanding of DKS. The simulations also predicted the existence of a single soliton state with a sech^2 spectral envelope (Fig. 2D). These predictions are in agreement with experiments. Although MI occurs for the effectively blue-detuned region, the regime of soliton formation occurs in the bistable regime where two solutions exist—where the laser is effectively red-detuned (Fig. 3A). Experimentally, this can be verified by using a Pound-Drever-Hall error signal, which can differentiate between effective red- and blue-detuning (Fig. 3A).

Stably accessing DKS in the presence of thermal nonlinearity

Accessing the DKS is in practice compounded by thermal effects. A fundamental consequence of the existence of bright solitons is the operation in the bistable regime, which causes the optical pump to be effectively red-detuned from the (Kerr and thermally shifted) cavity resonance. Although blue-detuned excitation is thermally stable—the resonator and laser form an autonomous feedback loop that stabilizes the laser cavity detuning—the opposite is true for a red-detuned operation (required for soliton formation) (*86*). Moreover, the series of discrete steps in resonator transmission laser scan due to soliton formation compound the stable access of solitons because concomitant with a discrete step is a change in intracavity power and, as a consequence, temperature change. This in turn, via

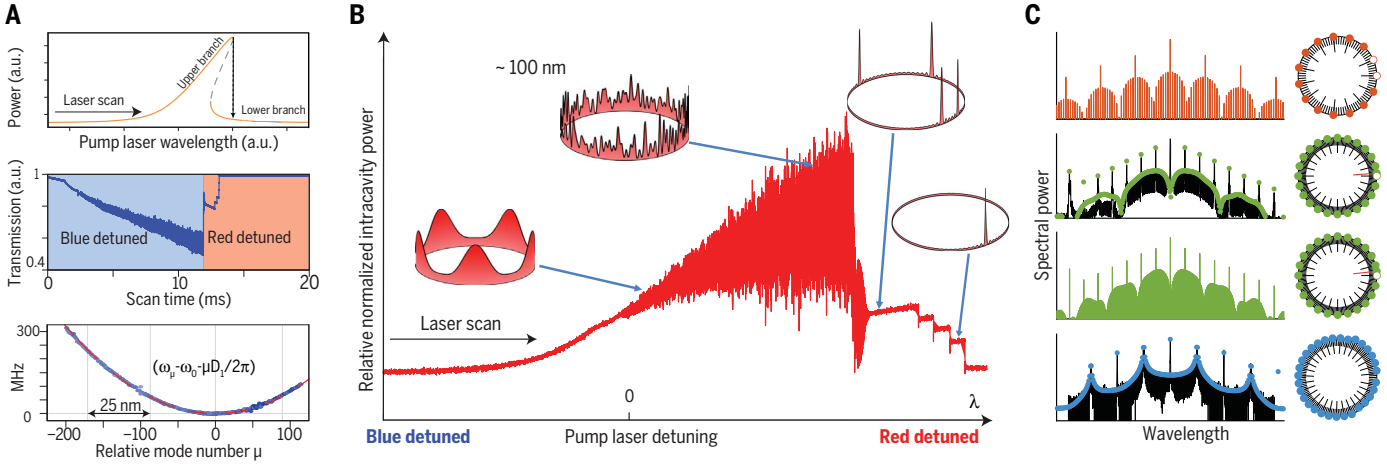


Fig. 3. Transition to the dissipative soliton states and soliton crystals. (**A**) (Top)The cavity bistability in the intra-cavity power due to the Kerr nonlinearity. (Bottom) The integrated dispersion profile [$D_{\mathrm{int}}(\mu)$] of a measured progression of resonances exhibiting anomalous dispersion (quadratic variation of the integrated dispersion). (Middle) A series of steps on the red-detuned side of the resonance, indicative of dissipative soliton formation. (**B**) The evolution of the intracavity power as a function of laser detuning, revealing in particular a series of discrete steps associated with soliton formation. (**C**) More complex arrangements of a large number of solitons that are ordered in crystals (*63*, *134*) but that contain defects, such as vacancies or defects. [Images are adapted from (*9*, *63*)]

the thermal effect (temperature-dependent refractive index), changes the relative laser cavity detuning. In early work, this challenge was overcome by using an optimized laser scan (*9*). Since then, several techniques have been developed, from "power kicking" (*74*) to very fast laser modulation using single sideband modulators (*75*), as well as fast thermal on-chip tuning (*19*) and carrier injection (*77*). All developed techniques have in common that once the DKS state is reached, the thermal response of the cavity is dominated by the DKS, causing the system to be self-stabilized in the presence of a red-detuned (typically several linewidth) strong pump field. This is due to the fortuitous circumstance that thermal effects make accessing the soliton state more difficult; once the solitons are generated, the thermal nonlinearity induced by the soliton self-stabilizes the laser-cavity detuning, constituting an autonomous feedback loop. Once accessed, DKS in microresonators can be passively stable for hours. Another useful feature of the DKS regime is that if probed with a modulated laser, the soliton state (*64*) exhibits two resonances: a $\mathcal{C}$ and $\mathcal{S}$ resonance, which correspond to the modulation response of the cavity and solitonic solution and reveal the effective laser detuning. Experimentally, the detuning of the laser determines the soliton duration, in which detuning further from resonance decreases the soliton duration, until the point at which the soliton existence range limit is reached (Fig. 2C, green region). In addition to continous wave pumping schemes, pulsed pumping can be used. Although the generation of the Kerr soliton microcombs can occur at record low power [tens of milliwatts have been reported for octave spanning spectra (*48*)], the efficiency of the process is low because of the detuned nature of the pump during the DKS formation. Periodic pulsed pumping at a frequency similar to the cavities' inverse round-trip time can enhance this efficiency by approximately the ratio of the round-trip time to the pulse duration. This scheme was recently shown for DKSs in fiber cavities (*21*), exhibiting a locking behavior to the drive pulse and leading to a substantial increase in the conversion efficiency. This scheme also has been demonstrated for Si_3N_4 microresonators for efficient and broadband comb generation (*49*). Because of the higher efficiency of pulsed pumping, thermal effects are suppressed, making tuning to the DKS state possible by using slow laser scans.

Theoretical modeling and numerical simulations of DKS in microresonators

Kerr comb formation can, in addition to CMEs described above, also be described by a mean field. The two descriptions are equivalent, but only the latter enables a connection to the soliton to be established. The internal optical field envelope $A(\phi, t)$ in a microresonator with a self-focusing Kerr nonlinearity and only second-order GVD may be described with the LLE (*12*, *13*, *16*, *72*), where ϕ is the angular coordinate in a ring resonator in a frame copropagating the envelope with the group velocity, and t is the slow time (slower than the round trip)

$$\frac{\partial A}{\partial t} - i\frac{D_2}{2}\frac{\partial^2 A}{\partial \phi^2} - ig|A|^2 A + (\kappa/2 + i\Delta)\,A = \sqrt{\eta\kappa s} \tag{2}$$

Here, the input laser power is given by $P = \hbar\omega_0|s|^2$, $\eta = \kappa_{ex}/\kappa$ is the coupling efficiency determined as a ratio of the output coupling rate κ_{ex} to the total loss rate κ ($\eta = 1/2$ corresponds to critical coupling, and $\eta \approx 1$ corresponds to strong overcoupling), $\hbar$ is the reduced Planck's constant, and ω_0 is the pumped optical frequency. D_1 falls out from the equation in transforming to a rotating frame $\phi = \varphi - D_1 t$, consideration of terms with $j > 2$ leads to appearance of higher-order derivatives in Eq. 2, and $\Delta = \omega_0 - \omega$ is the pump detuning. It is convenient to switch to a dimensionless equation

$$i\frac{\partial \Psi}{\partial \tau} + \frac{1}{2}\frac{\partial^2 \Psi}{\partial \theta^2} + |\Psi|^2\Psi = \zeta_0\Psi - i\Psi + if \tag{3}$$

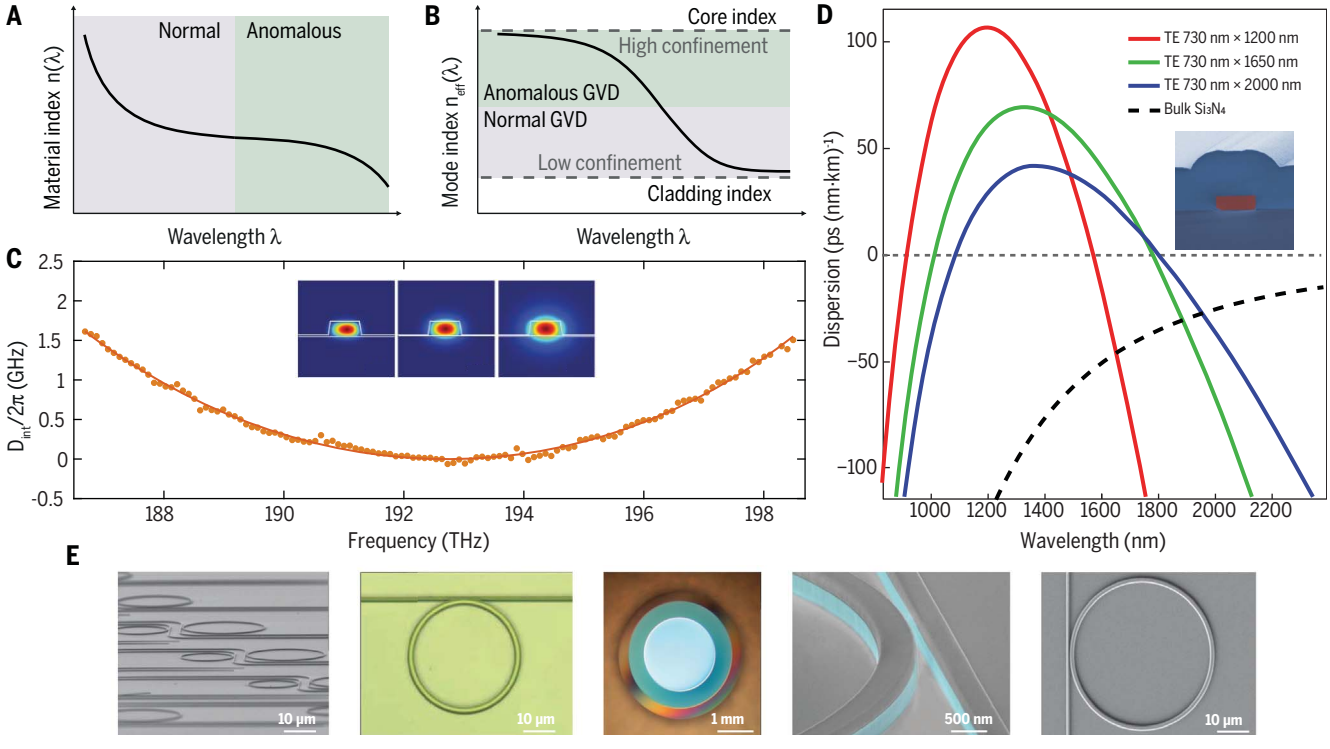


Fig. 4. Dispersion engineering in photonic chip–based integrated waveguides and resonators. (**A** and **B**) Effective refractive index of a photonic integrated waveguide as a function of wavelength, revealing that tight confinement waveguides give rise to anomalous group velocity dispersion (positive curvature of the effective index curve). By contrast, weak confinement leads to normal GVD. (**C**) Shown is the measured integrated dispersion parameter $D_{int}(\mu)$ of a Si_3N_4 microresonator by use of frequency comb-assisted spectroscopy, revealing purely anomalous GVD. (**D**) Engineering of the GVD of a Si_3N_4 waveguide cladded with silica, opening an anomalous GVD window in the telecommunication band (*98*). (**E**) Integrated microresonator platforms showing Kerr comb formation to date, including diamond, AlN, AlGaAs, silica on silicon, and Si_3N_4. DKSs have been generated in silica, Si_3N_4, and Si.

$(\theta = \phi\sqrt{\frac{\kappa}{2D_2}}, f = \sqrt{\frac{8g\eta P_{in}}{\kappa^2 \hbar\omega_0}}, \Psi = \sqrt{\frac{2g}{\kappa}}A, \zeta_0 = 2\Delta/\kappa, \tau = \kappa t/2)$. Without the right part, Eq. 3—known as the nonlinear Schrödinger (NLS) equation—is integrable with a sech-shaped soliton solution (*87*). An exact stationary solution of Eq. 3 in the form of solitonic pulses on CW background is also known (*14*, *88*, *89*), when only a driving term without losses is considered. Although this solution provides a good approximation in the limit of large detuning ζ_0, it gives little insight in the understanding of the DKS dynamics without numerical simulations and asymptotic approximations. The damped driven NLS (Eq. 3) is frequently referred to as the LLE (*13*), an important equation first introduced to describe two-dimensional spatial transversal solitonic field patterns in nonlinear Fabry-Perot etalons and later reformulated for longitudinal temporal solitons (*15*). The same longitudinal equation was, however, earlier analyzed in application to plasma physics (*90*, *91*), where solitons on background and existence boundaries were found.

In (*85*), an alternative approach for Kerr comb simulation was proposed on the basis of discrete analysis of each comb line nonlinearly coupled with all others. The system of equations in this way may consist of hundreds of CMEs with millions of nonlinear terms that nevertheless can be efficiently numerically integrated by using the fast Fourier transform (*92*). The CME description is equivalent to LLE (*11*) and may be considered as its discrete Fourier transform. The CME approach is useful to analyze dynamics for each

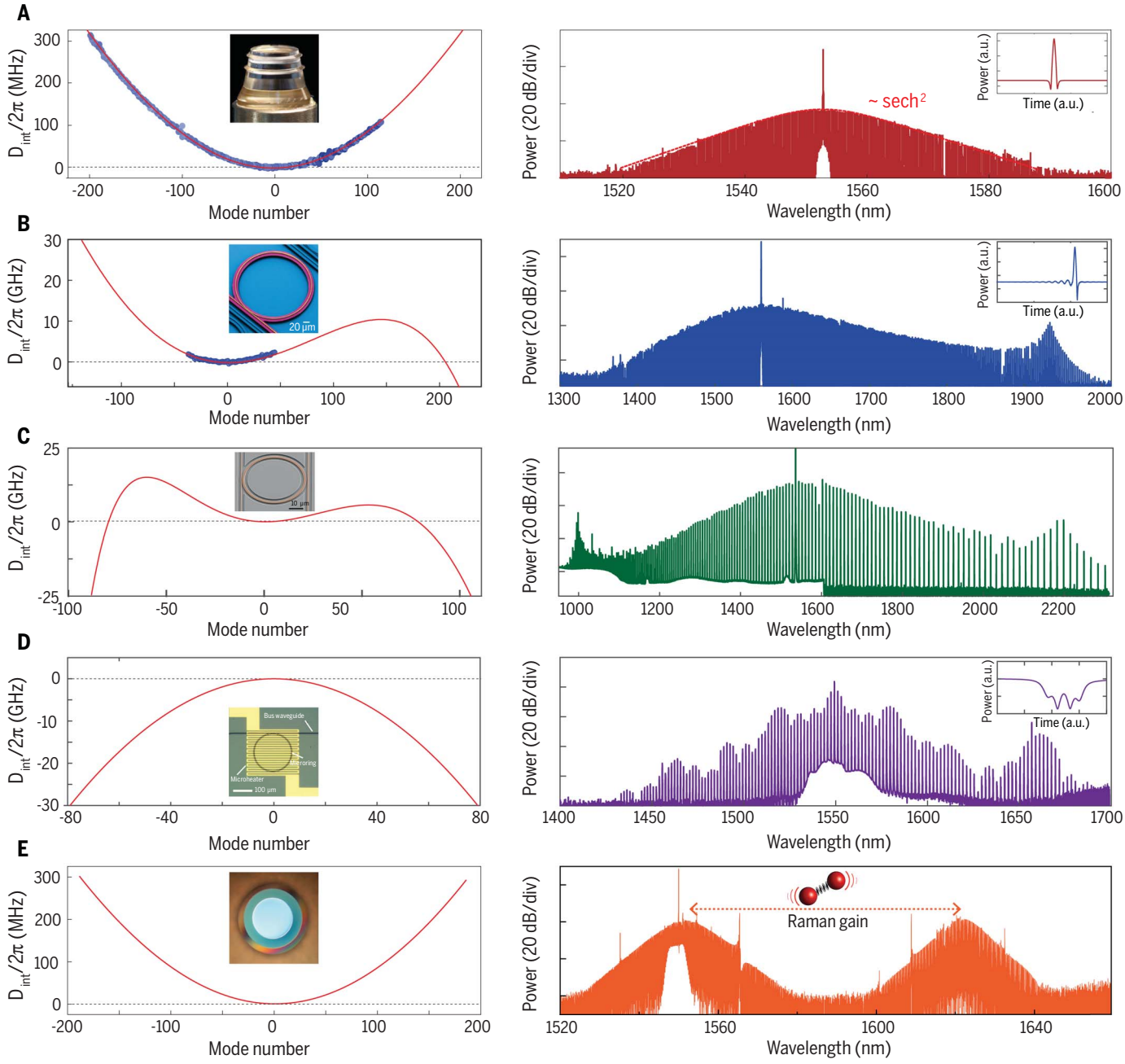


Fig. 5. The influence of dispersion and Raman effect on DKS. (**A**) A nearly purely quadratic dispersion (anomalous GVD)–dominated DKS generated in a crystalline microresonators. (**B**) Soliton-induced Cherenkov radiation—a soliton whose spectral envelope encompasses the normal GVD regime. A dispersive wave is emitted where the phase matching condition $[D_{int}(\mu) = 0]$ is satisfied. (**C**) A spectrum with dual dispersive waves, covering a full octave in a Si_3N_4 microresonator with terahertz mode spacing. (**D**) A platicon state (dark pulses) generated in the normal dispersion regime. (**E**) A Raman soliton, generated from the single soliton state in a silica microdisk.

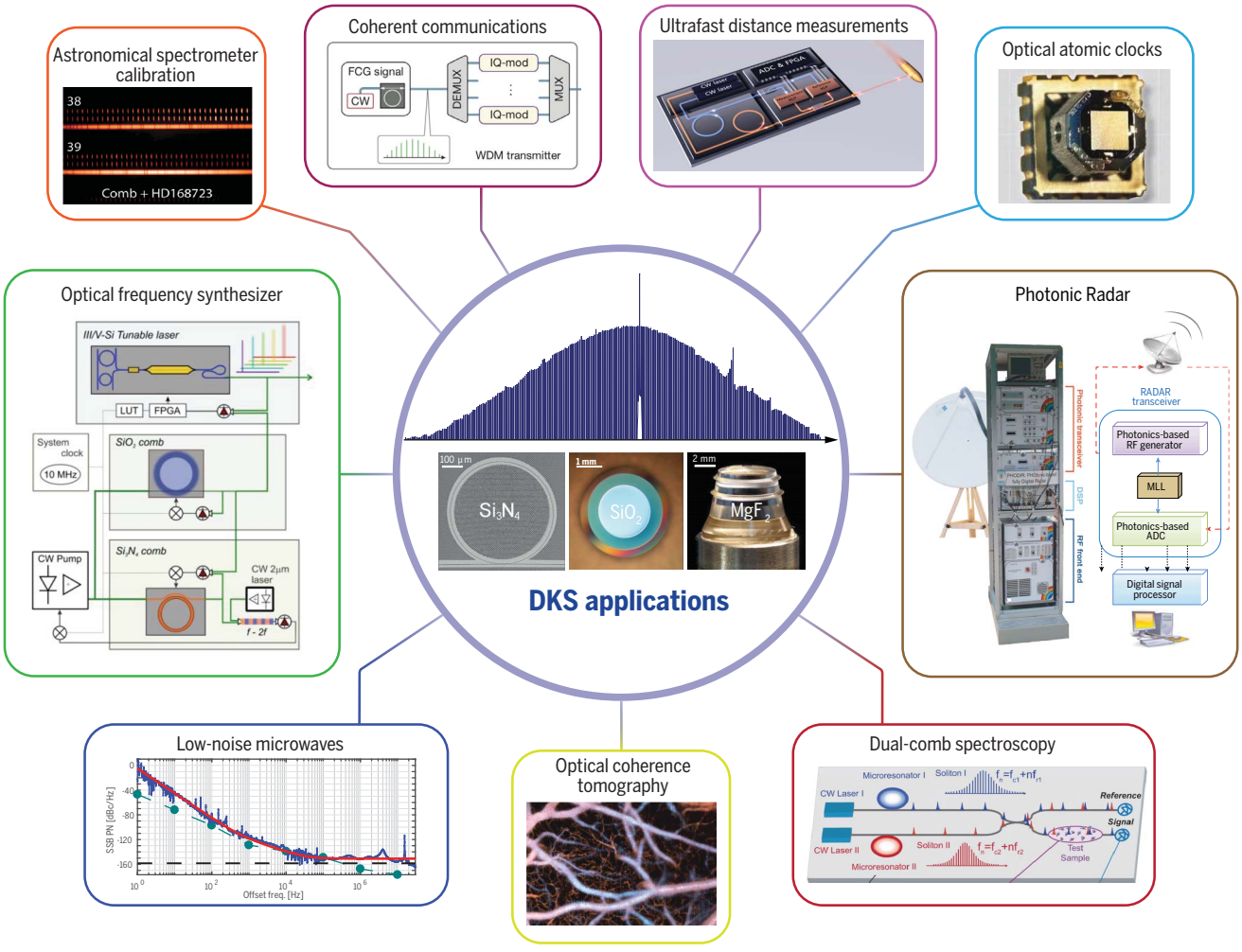


Fig. 6. Applications of soliton microcombs. Applications of areas of microcombs (from top, clockwise): ultrafast dual soliton microcomb distance measurements (LIDAR) (*46*), chip-scale atomic clocks, photonic Radar (*135*), dual-microcomb spectroscopy (*38*), optical coherence tomography, low-noise microwave generation (*45*), integrated optical frequency synthesizer (*136*), and astrophysical spectrometer calibration for exoplanet detection (*49*).

comb line and also in cases in which local dispersion for particular lines is perturbed and not easily described by the Taylor series (*61*).

Microresonator platforms for soliton formation

Soliton Kerr frequency combs in microresonators were first observed (*9*) in crystalline MgF_2 resonators (*45*, *93*) with ultrahigh-*Q* factor, leading to a low threshold for nonlinear effects (*94*). The platforms in which solitons can be generated have been extended significantly and include platforms that are amenable to wafer-scale processing, such as silica microdisk resonators (*17*) and photonic integrated platforms, notably Si_3N_4 (*19*, *20*, *95*, *96*). Such photonic chip–based resonators based on Si_3N_4 enable integration of waveguides and resonators, enable further functionality such as heaters, and are space compatible (*70*). Moreover, solitons have been generated in Si microresonators in the mid-infrared (*77*), where two-photon absorption is negligible. A challenge in this case is attaining sufficiently high *Q* for efficient nonlinear parametric frequency conversion.

One platform that is particularly well suited is based on Si_3N_4, a material that is already a part of the complementary metal-oxide semiconductor (CMOS) process, for the strain engineering of transistors and as a capping layer. Its high refractive index of $n \approx 2$ enables tight optical confinement waveguides and microring resonators, and the high bandgap (~3 eV) mitigates multiphoton absorption in the telecommunication band. Recent measurements have revealed that Si_3N_4 has intrinsic material limited *Q*-factors that can exceed 10^7 (*97*). Although the *Q*-factors attained in Si_3N_4 are still substantially below those of crystalline resonators, the significantly tighter optical confinement and higher Kerr-nonlinearity have allowed DKS generation. Moreover, DKSs have been observed in fiber Fabry-Perot microcavities (*21*). Although more platforms exist in which Kerr comb generation has been shown—including, in particular, compound semiconductors such as AlGaAs or AlN—soliton formation has to date not been demonstrated in these materials. Irrespective of the platform or material used, a requirement for soliton formation is attaining anomalous group velocity dispersion. As one approaches the bandgap of the material, the normal GVD will start to dominate. This limitation has been overcome with waveguide geometry–based dispersion engineering (*98*).

Dispersion engineering

Optical microresonators exhibit variation of the FSR because of the effect of material and geometric (resonator) dispersion. Material dispersion, commonly expressed via the Sellmeier equations, poses stringent limits on the wavelength range in which anomalous GVD is possible. By contrast, the resonator dispersion can be engineered via the resonator geometry and can be engineered to be anomalous GVD, overcoming the inherent material dispersion (*98*). For a nanophotonic waveguide, the dependence of the effective refractive index on the wavelengths leads inherently

to normal GVD when the mode is weakly confined and to anomalous GVD for tight confinement. This is illustrated in Fig. 4 and has been applied to achieve Kerr comb generation in Si_3N_4, despite the material having normal material dispersion (*71*). Moreover, coatings can be used to further engineer dispersion properties (*99*). A limitation arises from the whispering-gallery mode (WGM) potential induced normal GVD for small waveguide radii (*99*). The comb bandwidth can be significantly extended when making use of higher-order soliton broadening effects. If higher-order dispersion terms are significant, and the comb extends into regions where the dispersion changes sign, the comb bandwidth will be significantly extended. (*20*, *35*, *72*, *100*–*103*). Comb lines near zeros in the integrated dispersion profile (Fig. 3A) form dispersive waves coupled to solitons also known as soliton Cherenkov radiation (*103*–*106*). An analogous effect may occur because of dispersion perturbation induced by mode coupling (*62*, *106*). Proper design of the resonators (*73*) also allows for suppressing the nonfundamental modes that, because of mode couplings with multiple avoided mode crossings (*61*), may prevent generation of solitons. Because of the very short duration of the DKS in microresonators, the LLE needs to be generalized to include higher-order dispersion (*72*, *105*), Raman scattering, and self-steepening terms (*107*–*109*) to describe the soliton microcombs.

Dispersive waves in DKS

For solitons whose spectral bandwidth extends into the normal GVD regime, the LLE has to be extended with higher-order dispersion terms. This gives rise in particular to the effect of soliton-induced Cherenkov radiation, also known as dispersive wave formation. This process was first described for supercontinuum generation in optical fiber (*104*) and applies also to DKSs in optical microresonators (*72*). In this case, the soliton develops an oscillatory tail, on the trailing or leading edge of the pulse. The spectral location of this dispersive wave corresponds approximately to the condition where $D_{\mathrm{int}}(\mu) = 0$, where the resonator FSR matches that of the dispersionless soliton. Dispersive waves offer the practical advantage that they enable extension of the spectral coverage of the soliton Kerr comb to the normal dispersion regime and increase the power in the spectral ends of the comb. Suitable quartic dispersion enables also the creation of DKSs with two dispersive waves that can cover a full octave (Fig. 5C). The soliton Cherenkov radiation also entails a spectral recoil (*105*, *110*).

Raman effect in DKS and the Stokes soliton

The Raman effect is particularly relevant in amorphous media such as silica or Si_3N_4 but has been observed to be negligible in crystalline material of MgF_2 because of the narrow Raman gain bandwidth. The Raman effect leads to a Raman self-frequency shift: a soliton that is shifted with respect to the pump [also called frequency-locked Raman soliton (*53*)]. The Raman effect can also lead to a new type of soliton, the Stokes soliton, which is a soliton generated by the Raman gain and locked to the pump soliton (Fig. 5E) (*55*). Recently, it was also shown that DKSs experience a limitation on their temporal duration and bandwidth because of the stimulated Raman scattering (*111*), once the comb spacing is less than the Raman gain bandwidth. In addition, it was shown that in the presence of a large Raman gain (such as in diamond), comb generation can only occur if the FSR is sufficiently large to suppress Raman lasing.

Counterpropagating solitons and spatial multiplexing

WGM microresonators support both counterpropagating traveling modes, which allows simultaneous excitation of DKS in both clockwise and counterpropagating directions (*112*, *113*). The two combs, however, have been observed to lock to each other, an effect likely associated with Rayleigh backscattering. Such locking was recently demonstrated and enables the creation of two phase-locked pulse streams of soliton Kerr combs, which can be used for dual-comb distance measurements (*47*). Such dual-comb soliton pulse streams can also be generated by using spatial multiplexing of solitons—the generation of DKSs in several spatial mode families of one and the same resonator (*114*).

Normal dispersion solitonic pulses

The LLE also exhibits solutions for normal GVD, a solution known as dark soliton (short dip in CW background). However, because of periodic boundary conditions, purely dark solitons with opposite sign of phase on both sides are impossible in microresonators, whereas low-contrast gray solitons have a quite narrow, hardly achievable range of existence (*115*, *116*). Nevertheless, mode-locked Kerr frequency combs in the normal dispersion regime were experimentally demonstrated in different materials (*117*–*120*). It was revealed by using numerical simulations that these experimental results may be interpreted by use of a type of solitonic pulses called "platicons," flat-topped bright or dark (depending on the duration) pulses that can be softly excited and stably exist in microresonators with normal dispersion in case of bichromatic or amplitude-modulated pump (*120*) as well as with a local dispersion perturbation (*51*). In real microresonators, such a perturbation may occur because of the normal mode coupling between different mode families (*119*, *122*). Platicons represent bound states of opposing switching waves (*123*) that connect upper and lower branches of bistable resonance to satisfy periodic boundary conditions. They may be identified in experiments because of their characteristic optical spectrum with pronounced wings (Fig. 5). In (*124*), it was demonstrated that the dynamics of platicons in the presence of the third-order dispersion is quite peculiar and drastically different from bright solitons. In (*125*), a possibility of stable coexistence of dark and bright solitons in case of nonzero third-order dispersion was revealed.

Applications of DKSs

Accessing DKSs in microresonators has unlocked in a short timeframe several promising applications. The ability of DKSs to generate broadband-coherent combs with gigahertz repetition rates and short pulse streams has a number of applications in which either compactness is required or the high repetition rates are advantageous (Fig. 6).

Microwave-to-optical link

Soliton Kerr microcombs have enabled the counting of the cycles of light by creating a direct microwave-to-optical link, by using self-referencing assisted with external fiber-based broadening (*42*, *126*) and by directly exploiting the broadband nature of DKS (*43*).

Optical frequency synthesizers

DKSs based on a dual-comb approach have been used to create an integrated optical frequency synthesizer that allows synthesizing an optical frequency from a radio-frequency reference (*48*).

Massively parallel coherent communications

Kerr combs with hundreds of equally spaced lines can serve as a reference for wavelength division multiplexing in massively parallel coherent telecommunications, with demonstrated data rates exceeding 50 terabits per second, at both the receiver and transmitter ends (*44*).

Dual-comb spectroscopy

Two slightly different frequency combs can form on a photodetector a radio-frequency comb (*25*), with the repetition rate equal to both combs' optical repetition rates difference. DKSs have been used for such dual-comb measurements in both near- and mid-infrared (*37*–*40*) and allow even full integration on chip.

LIDAR

Dual-microcomb may be implemented for ultrafast distance measurements with submicrometer resolution (*46*), using counterclockwise solitons (*47*).

Low-noise microwave generation

Soliton Kerr microcombs can also be used for compact ultrastable microwave generation, especially if augmented with compact atomic cell references (*126*, *128*).

Astrophysical spectrometer calibration

The large line spacing makes DKSs ideal for calibrating astrophysical spectrometers for exoplanet searches. In contrast to conventional laser frequecy combs that require filtering, this is not necessary for DKSs. Recent work demonstrated that DKSs provide calibration sufficient for search of Earth-like planets (*49*, *50*).

Beyond these already demonstrated applications, there is application potential in several domains, in particular for the visible range. Optical coherence tomography, dual-comb coherent anti-Stokes Raman spectroscopy (*129*), or photonic microwave signal processing (*130*) are obvious

new directions, but qualitatively and quantitatively new nonlinear physics and devices are expected when advancing the ability to engineer dispersion over broad bandwidth in photonic chips and, when investigated, more complex photonic devices and geometries, such as photonic band-gap waveguides and cavities.

Further developments and outlook

The fast progress in recent years—in the understanding of soliton formation, the broad range of platforms in which solitons in driven dissipative nonlinear microresonators have been generated, and the rapid rise of applications and proof of concepts—indicate a fruitful playground for future progress. Yet challenges exist, in particular toward making the technology viable: increasing bandwidth of solitons, increasing the efficiency, and increasing the photonic integration. Moreover, despite decades of research, integrated photonic resonators presently still exhibit propagation losses that are more than three orders of magnitude higher than those of standard optical fibers. Overcoming this challenge would allow dramatic improvements in power efficiency and likely requires further research in photonic materials and fabrication processes. Several areas can be identified that may offer advances and new avenues both technological in nature as well as fundamental, such as new and improved materials and new platforms. Although the advances cannot be predicted, several developments can be anticipated.

Advanced dispersion engineering and new nonlinear materials

The coherent synthesis of an even broader spectrum, possibly corresponding to single-cycle pulses, depends critically on dispersion engineering. Dispersion engineering may open single or even subcycle waveforms inside dielectric resonators, as well as new coupled soliton complexes. Such advanced dispersion engineering is in its infancy at present and could use multilayer coatings, more complex waveguides that use coupled resonators or photonic crystal waveguide geometries, or combinations of existing materials. To date, solitons have been generated in a small class of materials (MgF_2, Si_3N_4, Si, and SiO_2), and evident potential exists when using III-V semiconductors in particular, which offer atomic-scale deposition processes and tailored band gap (*131*). Likewise, solitons are surprisingly stable attractors—and are robust to dispersion disorder. This could enable machine-learning algorithms to design complex highly multimode systems, whose composite mode coupling induced a broadband desired dispersion landscape that supports solitons. If achieved, it may become possible to engineer arbitrary and much more complex and flat dispersion landscapes that could host multiple a-GVD–hosted soliton complexes.

Quantum technology

A further area under exploration is the use of parametric interactions, a process that is capable of generating correlated pairs of photons for the generation of quantum states. One application lies in generating quantum random number generation using optical parametric oscillations (*132*). Moreover, the parametric fluorescence in Kerr nonlinear microresonators has already been successfully used for correlated photon states and higher-order correlated photon states (*133*). Beyond technological advances outlined above, there are also fundamental open questions. An emerging question is the behavior of coupled parametric soliton arrays. The behavior of such driven, dissipative, nonlinear systems is an emergent question that is also experimentally accessible and may give rise to unanticipated effects. These possible advances reveal that DKS-based microcombs may offer advances to a wide range of areas, from timekeeping and spectroscopy to sensing and telecommunications, and are emerging as candidate building blocks in tomorrow's photonic technologies.

REFERENCES AND NOTES

1. N. J. Zabusky, M. D. Kruskal, Interaction of "solitons" in a collisionless plasma and the recurrence of initial states. *Phys. Rev. Lett.* **15**, 240–243 (1965). doi: 10.1103/PhysRevLett.15.240
2. L. F. Mollenauer, R. H. Stolen, J. P. Gordon, Experimental observation of picosecond pulse narrowing and solitons in optical fibers. *Phys. Rev. Lett.* **45**, 1095–1098 (1980). doi: 10.1103/PhysRevLett.45.1095
3. K. E. Strecker, G. B. Partridge, A. G. Truscott, R. G. Hulet, Formation and propagation of matter-wave soliton trains. *Nature* **417**, 150–153 (2002). doi: 10.1038/nature747; pmid: 11986621
4. P. Grelu, N. Akhmediev, Dissipative solitons for mode-locked lasers. *Nat. Photonics* **6**, 84–92 (2012). doi: 10.1038/nphoton.2011.345
5. N. Akhmediev, A. Ankiewicz, *Dissipative Solitons*, Lecture Notes in Physics (Springer-Verlag, 2005).
6. T. J. Kippenberg, R. Holzwarth, S. A. Diddams, Microresonator-based optical frequency combs. *Science* **332**, 555–559 (2011). doi: 10.1126/science.1193968; pmid: 21527707
7. P. Del'Haye *et al*., Optical frequency comb generation from a monolithic microresonator. *Nature* **450**, 1214–1217 (2007). doi: 10.1038/nature06401; pmid: 18097405
8. F. Leo *et al*., Temporal cavity solitons in one-dimensional Kerr media as bits in an all-optical buffer. *Nat. Photonics* **4**, 471–476 (2010). doi: 10.1038/nphoton.2010.120
9. T. Herr *et al*., Temporal solitons in optical microresonators. *Nat. Photonics* **8**, 145–152 (2014). doi: 10.1038/nphoton.2013.343
10. A. B. Matsko, A. A. Savchenkov, D. Strekalov, V. S. Ilchenko, L. Maleki, Optical hyperparametric oscillations in a whispering-gallery-mode resonator: Threshold and phase diffusion. *Phys. Rev. A* **71**, 033804 (2005). doi: 10.1103/PhysRevA.71.033804
11. T. J. Kippenberg, S. M. Spillane, K. J. Vahala, Kerr-nonlinearity optical parametric oscillation in an ultrahigh-*Q* toroid microcavity. *Phys. Rev. Lett.* **93**, 083904 (2004). doi: 10.1103/PhysRevLett.93.083904; pmid: 15447188
12. Y. K. Chembo, C. R. Menyuk, Spatiotemporal Lugiato-Lefever formalism for Kerr-comb generation in whispering-gallery-mode resonators. *Phys. Rev. A* **87**, 053852 (2013). doi: 10.1103/PhysRevA.87.053852
13. L. A. Lugiato, R. Lefever, Spatial dissipative structures in passive optical systems. *Phys. Rev. Lett.* **58**, 2209–2211 (1987). doi: 10.1103/PhysRevLett.58.2209; pmid: 10034681
14. I. V. Barashenkov IV, Y. S. Smirnov, Existence and stability chart for the ac-driven, damped nonlinear Schrödinger solitons. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **54**, 5707–5725 (1996). doi: 10.1103/PhysRevE.54.5707; pmid: 9965759
15. M. Haelterman, S. Trillo, S. Wabnitz, Dissipative modulation instability in a nonlinear dispersive ring cavity. *Opt. Commun.* **91**, 401–407 (1992). doi: 10.1016/0030-4018(92)90367-Z
16. A. B. Matsko *et al*., Mode-locked Kerr frequency combs. *Opt. Lett.* **36**, 2845–2847 (2011). doi: 10.1364/OL.36.002845; pmid: 21808332
17. X. Yi, Q.-F. Yang, K. Y. Yang, M.-G. Suh, K. Vahala, Soliton frequency comb at microwave rates in a high-*Q* silica microresonator. *Optica* **2**, 1078 (2015). doi: 10.1364/OPTICA.2.001078
18. A. G. Griffith *et al*., Silicon-chip mid-infrared frequency comb generation. *Nat. Commun.* **6**, 6299 (2015). doi: 10.1038/ncomms7299; pmid: 25708922
19. C. Joshi *et al*., Thermally controlled comb generation and soliton modelocking in microresonators. *Opt. Lett.* **41**, 2565–2568 (2016). doi: 10.1364/OL.41.002565; pmid: 27244415
20. V. Brasch *et al*., Photonic chip-based optical frequency comb using soliton Cherenkov radiation. *Science* **351**, 357–360 (2016). doi: 10.1126/science.aad4811; pmid: 26721682
21. E. Obrzud, S. Lecomte, T. Herr, Temporal solitons in microresonators driven by optical pulses. *Nat. Photonics* **11**, 600–607 (2017). doi: 10.1038/nphoton.2017.140
22. T. Udem, R. Holzwarth, T. W. Hänsch, Optical frequency metrology. *Nature* **416**, 233–237 (2002). doi: 10.1038/416233a; pmid: 11894107
23. J. L. Hall, Defining and measuring optical frequencies: The optical clock opportunity—and more (Nobel lecture). *Rev. Mod. Phys.* **78**, 1279–1295 (2006). doi: 10.1103/RevModPhys.78.1279; pmid: 17086589
24. S. T. Cundiff, J. Ye, *Colloquium* : Femtosecond optical frequency combs. *Rev. Mod. Phys.* **75**, 325–342 (2003). doi: 10.1103/RevModPhys.75.325
25. I. Coddington, N. Newbury, W. Swann, Dual-comb spectroscopy. *Optica* **3**, 414 (2016). doi: 10.1364/OPTICA.3.000414
26. Y. K. Chembo, Kerr optical frequency combs: Theory, applications and perspectives. *Nanophotonics* **5**, 214 (2016). doi: 10.1515/nanoph-2016-0013
27. A. A. Savchenkov, A. B. Matsko, L. Maleki, On Frequency Combs in Monolithic Resonators. *Nanophotonics* **5**, 363–391 (2016). doi: 10.1515/nanoph-2016-0031
28. A. Pasquazi *et al*., Micro-combs: A novel generation of optical sources. *Phys. Rep.* **729**, 1–81 (2018). doi: 10.1016/j.physrep.2017.08.004
29. P. Del'Haye *et al*., Octave spanning tunable frequency comb from a microresonator. *Phys. Rev. Lett.* **107**, 063901 (2011). doi: 10.1103/PhysRevLett.107.063901; pmid: 21902324
30. T. Herr *et al*., Universal formation dynamics and noise of Kerr-frequency combs in microresonators. *Nat. Photonics* **6**, 480–487 (2012). doi: 10.1038/nphoton.2012.127
31. F. Ferdous *et al*., Spectral line-by-line pulse shaping of on-chip microresonator frequency combs. *Nat. Photonics* **5**, 770–776 (2011). doi: 10.1038/nphoton.2011.255
32. S. H. Lee *et al*., Towards visible soliton microcomb generation. *Nat. Commun.* **8**, 1295 (2017). doi: 10.1038/s41467-017-01473-9; pmid: 29101367
33. M. Karpov, M. H. Pfeiffer, T. J. Kippenberg, Photonic chip-based soliton frequency combs covering the biological imaging window. arXiv:1706.06445 [physics.optics] (2017).
34. Q. Li *et al*., Stably accessing octave-spanning microresonator frequency combs in the soliton regime. *Optica* **4**, 193–203 (2017). doi: 10.1364/OPTICA.4.000193; pmid: 28603754
35. M. H. P. Pfeiffer *et al*., Octave-spanning dissipative Kerr soliton frequency combs in Si_3N_4 microresonators. *Optica* **4**, 684 (2017). doi: 10.1364/OPTICA.4.000684
36. M. Yu *et al*., Silicon-chip-based mid-infrared dual-comb spectroscopy. arXiv:1610.01121 [physics.optics] (2016).
37. M. G. Suh, K. Vahala, Gigahertz-repetition-rate soliton microcombs. *Optica* **5**, 65–66 (2018). doi: 10.1364/OPTICA.5.000065
38. M.-G. Suh, Q.-F. Yang, K. Y. Yang, X. Yi, K. J. Vahala, Microresonator soliton dual-comb spectroscopy. *Science* **354**, 600–603 (2016). doi: 10.1126/science.aah6516; pmid: 27738017
39. N. G. Pavlov *et al*., Soliton dual frequency combs in crystalline microresonators. *Opt. Lett.* **42**, 514–517 (2017). doi: 10.1364/OL.42.000514; pmid: 28146515
40. A. Dutt *et al*., On-chip dual comb source for spectroscopy. arXiv:1611.07673 [physics.optics] (2016).
41. M. Yu, Y. Okawachi, A. G. Griffith, M. Lipson, A. L. Gaeta, Microresonator-based high-resolution gas spectroscopy. *Opt. Lett.* **42**, 4442–4445 (2017). doi: 10.1364/OL.42.004442; pmid: 29088183
42. J. D. Jost *et al*., Counting the cycles of light using a self-referenced optical microresonator. *Optica* **2**, 706 (2015). doi: 10.1364/OPTICA.2.000706
43. V. Brasch, E. Lucas, J. D. Jost, M. Geiselmann, T. J. Kippenberg, Self-referenced photonic chip soliton Kerr

frequency comb. *Light Sci. Appl.* **6**, e16202 (2016). doi: 10.1038/lsa.2016.202

44. P. Marin-Palomo *et al*., Microresonator-based solitons for massively parallel coherent optical communications. *Nature* **546**, 274–279 (2017). doi: 10.1038/nature22387; pmid: 28593968
45. W. Liang *et al*., High spectral purity Kerr frequency comb radio frequency photonic oscillator. *Nat. Commun.* **6**, 7957 (2015). doi: 10.1038/ncomms8957; pmid: 26260955
46. P. Trocha *et al*., Ultrafast optical ranging using microresonator soliton frequency combs. *Science* **359**, 887–891 (2018). doi: 10.1126/science.aao3924; pmid: 29472477
47. M.-G. Suh, K. Vahala, Soliton microcomb range measurement. arXiv:1705.06697 [physics.optics] (2017).
48. D. T. Spencer *et al*., An integrated-photonics optical-frequency synthesizer. arXiv:1708.05228 [physics.app-ph] (2017).
49. E. Obrzud *et al*., A microphotonic astrocomb. arXiv:1712.09526 [physics.optics] (2017).
50. M.-G. Suh *et al*., Searching for exoplanets using a microresonator astrocomb. arXiv:1801.05174 [physics.optics] (2018).
51. V. E. Lobanov, G. Lihachev, T. J. Kippenberg, M. L. Gorodetsky, Frequency combs and platicons in optical microresonators with normal GVD. *Opt. Express* **23**, 7713–7721 (2015). doi: 10.1364/OE.23.007713; pmid: 25837109
52. M. Karpov *et al*., Raman self-frequency shift of dissipative Kerr solitons in an optical microresonator. *Phys. Rev. Lett.* **116**, 103902 (2016). doi: 10.1103/PhysRevLett.116.103902; pmid: 27015482
53. C. Milián, A. V. Gorbach, M. Taki, A. V. Yulin, D. V. Skryabin, Solitons and frequency combs in silica microring resonators: Interplay of the Raman and higher-order dispersion effects. *Phys. Rev. A* **92**, 033851 (2015). doi: 10.1103/PhysRevA.92.033851
54. X. Yi, Q.-F. Yang, K. Y. Yang, K. Vahala, Theory and measurement of the soliton self-frequency shift and efficiency in optical microcavities. *Opt. Lett.* **41**, 3419–3422 (2016). doi: 10.1364/OL.41.003419; pmid: 27472583
55. Q. F. Yang, X. Yi, K. Y. Yang, K. Vahala, Stokes solitons in optical microcavities. *Nat. Phys.* **13**, 53–57 (2017). doi: 10.1038/nphys3875
56. E. Fermi, J. Pasta, S. Ulam, *Los Alamos Report LA-1940* (1955); reproduced in A. C. Newell, *Nonlinear Wave Motion* (AMS, 1974).
57. A. B. Matsko, A. A. Savchenkov, L. Maleki, On excitation of breather solitons in an optical microresonator. *Opt. Lett.* **37**, 4856–4858 (2012). doi: 10.1364/OL.37.004856; pmid: 23202069
58. C. Y. Bao *et al*., *Phys. Rev. Lett.* **117**, 5 (2016).
59. M. Yu *et al*., Breather soliton dynamics in microresonators. *Nat. Commun.* **8**, 14569 (2017). doi: 10.1038/ncomms14569; pmid: 28232720
60. E. Lucas, M. Karpov, H. Guo, M. L. Gorodetsky, T. J. Kippenberg, Breathing dissipative solitons in optical microresonators. *Nat. Commun.* **8**, 736 (2017). doi: 10.1038/s41467-017-00719-w; pmid: 28963496
61. T. Herr *et al*., Mode spectrum and temporal soliton formation in optical microresonators. *Phys. Rev. Lett.* **113**, 123901 (2014). doi: 10.1103/PhysRevLett.113.123901; pmid: 25279630
62. Q.-F. Yang, X. Yi, K. Y. Yang, K. Vahala, Spatial-mode-interaction-induced dispersive waves and their active tuning in microresonators. *Optica* **3**, 1132 (2016). doi: 10.1364/OPTICA.3.001132
63. D. C. Cole, E. S. Lamb, P. DelHaye, S. A. Diddams, S. B. Papp, Soliton crystals in Kerr resonators. *Nat. Photonics* **11**, 671–676 (2017). doi: 10.1038/s41566-017-0009-z
64. H. Guo *et al*., Universal dynamics and deterministic switching of dissipative Kerr solitons in optical microresonators. *Nat. Phys.* **13**, 94–102 (2017). doi: 10.1038/nphys3893
65. K. Ikeda, H. Daido, O. Akimoto, Optical turbulence: Chaotic behavior of transmitted light from a ring cavity. *Phys. Rev. Lett.* **45**, 709–712 (1980). doi: 10.1103/PhysRevLett.45.709
66. D. W. M. Laughlin, J. V. Moloney, A. C. Newell, Solitary waves as fixed points of infinite-dimensional maps in an optical bistable ring cavity. *Phys. Rev. Lett.* **51**, 75–78 (1983) PRL. doi: 10.1103/PhysRevLett.51.75
67. T. Hansson, S. Wabnitz, Dynamics of microresonator frequency comb generation: Models and stability. *Nanophotonics* **5**, 231 (2016). doi: 10.1515/nanoph-2016-0012
68. A. Hasegawa, F. Tappert, Transmission of stationary nonlinear optical pulses in dispersive dielectric fibers. I. Anomalous dispersion. *Appl. Phys. Lett.* **23**, 142–144 (1973). doi: 10.1063/1.1654836
69. K. Saha *et al*., Modelocking and femtosecond pulse generation in chip-based frequency combs. *Opt. Express* **21**, 1335–1343 (2013). doi: 10.1364/OE.21.001335; pmid: 23389027
70. V. Brasch, Q.-F. Chen, S. Schiller, T. J. Kippenberg, Radiation hardness of high-Q silicon nitride microresonators for space compatible integrated optics. *Opt. Express* **22**, 30786–30794 (2014). doi: 10.1364/OE.22.030786; pmid: 25607027
71. J. S. Levy *et al*., CMOS-compatible multiple-wavelength oscillator for on-chip optical interconnects. *Nat. Photonics* **4**, 37–40 (2010). doi: 10.1038/nphoton.2009.259
72. S. Coen, H. G. Randle, T. Sylvestre, M. Erkintalo, Modeling of octave-spanning Kerr frequency combs using a generalized mean-field Lugiato-Lefever model. *Opt. Lett.* **38**, 37–39 (2013). doi: 10.1364/OL.38.000037; pmid: 23282830
73. A. Kordts, M. H. P. Pfeiffer, H. Guo, V. Brasch, T. J. Kippenberg, Higher order mode suppression in high-*Q* anomalous dispersion SiN microresonators for temporal dissipative Kerr soliton formation. *Opt. Lett.* **41**, 452–455 (2016). doi: 10.1364/OL.41.000452; pmid: 26907395
74. V. Brasch, M. Geiselmann, M. H. P. Pfeiffer, T. J. Kippenberg, Bringing short-lived dissipative Kerr soliton states in microresonators into a steady state. *Opt. Express* **24**, 29312–29320 (2016). doi: 10.1364/OE.24.029312; pmid: 27958591
75. J. Stone *et al*., Thermal and nonlinear dissipative-soliton dynamics in Kerr microresonator frequency combs. arXiv:1708.08405 [physics.optics] (2017).
76. K. E. Webb, M. Erkintalo, S. Coen, S. G. Murdoch, Experimental observation of coherent cavity soliton frequency combs in silica microspheres. *Opt. Lett.* **41**, 4613–4616 (2016). doi: 10.1364/OL.41.004613; pmid: 28005849
77. M. Yu, Y. Okawachi, A. G. Griffith, M. Lipson, A. L. Gaeta, Mode-locked mid-infrared frequency combs in a silicon microresonator. *Optica* **3**, 854 (2016). doi: 10.1364/OPTICA.3.000854
78. X. Yi, Q.-F. Yang, K. Youl Yang, K. Vahala, Active capture and stabilization of temporal solitons in microresonators. *Opt. Lett.* **41**, 2037–2040 (2016). doi: 10.1364/OL.41.002037; pmid: 27128068
79. E. Lucas, H. Guo, J. D. Jost, M. Karpov, T. J. Kippenberg, Detuning-dependent properties and dispersion-induced instabilities of temporal dissipative Kerr solitons in optical microresonators. *Phys. Rev. A* **95**, 043822 (2017). doi: 10.1103/PhysRevA.95.043822
80. T. Herr, M. L. Gorodetsky, T. J. Kippenberg, in *Dissipative Kerr Solitons in Optical Microresonators* (Wiley, 2015), pp. 129–162.
81. A. A. Savchenkov *et al*., Low threshold optical oscillations in a whispering gallery mode CaF_2 resonator. *Phys. Rev. Lett.* **93**, 243905 (2004). doi: 10.1103/PhysRevLett.93.243905; pmid: 15697815
82. C. Y. Wang *et al*., Mid-infrared optical frequency combs at 2.5 μm based on crystalline microresonators. *Nat. Commun.* **4**, 1345 (2013). doi: 10.1038/ncomms2335; pmid: 23299895
83. J. Li, H. Lee, T. Chen, K. J. Vahala, Low-pump-power, low-phase-noise, and microwave to millimeter-wave repetition rate operation in microcombs. *Phys. Rev. Lett.* **109**, 233901 (2012). doi: 10.1103/PhysRevLett.109.233901; pmid: 23368202
84. M. L. Gorodetsky, WE-Heraeus Seminar on Micro and Macro-cavities in Classical and Non-classical Light (2011); http://goo.gl/tmniGm.
85. Y. K. Chembo, D. V. Strekalov, N. Yu, Spectrum and dynamics of optical frequency combs generated with monolithic whispering gallery mode resonators. *Phys. Rev. Lett.* **104**, 103902 (2010). doi: 10.1103/PhysRevLett.104.103902; pmid: 20366426
86. T. Carmon, L. Yang, K. Vahala, Dynamical thermal behavior and thermal self-stability of microcavities. *Opt. Express* **12**, 4742–4750 (2004). doi: 10.1364/OPEX.12.004742; pmid: 19484026
87. V. E. Zakharov, A. B. Shabat, Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media. *Sov. Phys. JETP* **34**, 62 (1972).
88. T. S. Raju, C. N. Kumar, P. K. Panigrahi, On exact solitary wave solutions of the nonlinear Schrödinger equation with a source. *J. Phys. Math. Gen.* **38**, L271–L276 (2005). doi: 10.1088/0305-4470/38/16/L02
89. W. H. Renninger, P. T. Rakich, Closed-form solutions and scaling laws for Kerr frequency combs. *Sci. Rep.* **6**, 24742 (2016). doi: 10.1038/srep24742; pmid: 27108810
90. D. J. Kaup, A. C. Newell, Solitons as Particles, Oscillators, and in slowly changing media: A singular perturbation theory. *Proc. R. Soc. London Ser. A* **361**, 413–446 (1978). doi: 10.1098/rspa.1978.0110
91. K. Nozaki, N. Bekki, Solitons as attractors of a forced dissipative nonlinear Schrödinger equation. *Phys. Lett. A* **102**, 383–386 (1984). doi: 10.1016/0375-9601(84)91060-0
92. T. Hansson, D. Modotto, S. Wabnitz, On the numerical simulation of Kerr frequency combs using coupled mode equations. *Opt. Commun.* **312**, 134–136 (2014). doi: 10.1016/j.optcom.2013.09.017
93. I. S. Grudinin *et al*., High-contrast Kerr frequency combs. *Optica* **4**, 434 (2017). doi: 10.1364/OPTICA.4.000434
94. V. B. Braginsky, M. L. Gorodetsky, V. S. Ilchenko, Quality-factor and nonlinear properties of optical whispering-gallery modes. *Phys. Lett. A* **137**, 393–397 (1989). doi: 10.1016/0375-9601(89)90912-2
95. P.-H. Wang *et al*., Intracavity characterization of micro-comb generation in the single-soliton regime. *Opt. Express* **24**, 10890–10897 (2016). doi: 10.1364/OE.24.010890; pmid: 27409909
96. M. H. P. Pfeiffer *et al*., Photonic Damascene process for integrated high-*Q* microresonator based nonlinear photonics. *Optica* **3**, 20 (2016). doi: 10.1364/OPTICA.3.000020
97. X. Ji *et al*., Ultra-low-loss on-chip resonators with sub-milliwatt parametric oscillation threshold. *Optica* **4**, 619 (2017). doi: 10.1364/OPTICA.4.000619
98. M. A. Foster *et al*., Broad-band optical parametric gain on a silicon photonic chip. *Nature* **441**, 960–963 (2006). doi: 10.1038/nature04932; pmid: 16791190
99. J. Riemensberger *et al*., Dispersion engineering of thick high-*Q* silicon nitride ring-resonators via atomic layer deposition. *Opt. Express* **20**, 27661–27669 (2012). doi: 10.1364/OE.20.027661; pmid: 23262714
100. L. Zhang *et al*., Generation of two-cycle pulses and octave-spanning frequency combs in a dispersion-flattened micro-resonator. *Opt. Lett.* **38**, 5122–5125 (2013). doi: 10.1364/OL.38.005122; pmid: 24281525
101. Y. Okawachi *et al*., Bandwidth shaping of microresonator-based frequency combs via dispersion engineering. *Opt. Lett.* **39**, 3535–3538 (2014). doi: 10.1364/OL.39.003535; pmid: 24978530
102. I. S. Grudinin, N. Yu, Dispersion engineering of crystalline resonators via microstructuring. *Optica* **2**, 221 (2015). doi: 10.1364/OPTICA.2.000221
103. C. Bao *et al*., High-order dispersion in Kerr comb oscillators. *J. Opt. Soc. Am. B* **34**, 715 (2017). doi: 10.1364/JOSAB.34.000715
104. N. Akhmediev, M. Karlsson, Cherenkov radiation emitted by solitons in optical fibers. *Phys. Rev. A* **51**, 2602–2607 (1995). doi: 10.1103/PhysRevA.51.2602; pmid: 9911876
105. A. V. Cherenkov, V. E. Lobanov, M. L. Gorodetsky, Dissipative Kerr solitons and Cherenkov radiation in optical microresonators with third-order dispersion. *Phys. Rev. A* **95**, 033810 (2017). doi: 10.1103/PhysRevA.95.033810
106. X. Yi *et al*., *Nat. Commun.* **8**, 9 (2017). doi: 10.1038/s41467-017-00020-w; pmid: 28377584
107. M. R. E. Lamont, Y. Okawachi, A. L. Gaeta, Route to stabilized ultrabroadband microresonator-based frequency combs. *Opt. Lett.* **38**, 3478–3481 (2013). doi: 10.1364/OL.38.003478; pmid: 24104792
108. H. Kumar, F. Chand, Dark and bright solitary wave solutions of the higher order nonlinear Schrödinger equation with self-steepening and self-frequency shift effects. *J. Nonlinear Opt. Phys. Mater.* **22**, 135001 (2013). doi: 10.1142/S021886351350001X
109. C. Bao *et al*., Nonlinear conversion efficiency in Kerr frequency comb generation. *Opt. Lett.* **39**, 6126–6129 (2014). doi: 10.1364/OL.39.006126; pmid: 25361295
110. M. Erkintalo, Y. Q. Xu, S. G. Murdoch, J. M. Dudley, G. Genty, Cascaded phase matching and nonlinear symmetry breaking in fiber frequency combs. *Phys. Rev. Lett.* **109**, 223904 (2012). doi: 10.1103/PhysRevLett.109.223904; pmid: 23368122
111. Y. Wang, M. Anderson, S. Coen, S. G. Murdoch, M. Erkintalo, Stimulated Raman scattering imposes fundamental limits to the duration and bandwidth of temporal cavity solitons. *Phys. Rev. Lett.* **120**, 053902 (2018). doi: 10.1103/PhysRevLett.120.053902; pmid: 29481150

112. Q. F. Yang, X. Yi, K. Y. Yang, K. Vahala, Counter-propagating solitons in microresonators. *Nat. Photonics* **11**, 560–564 (2017). doi: 10.1038/nphoton.2017.117
113. C. Joshi *et al*., Counter-rotating cavity solitons in a silicon nitride microresonator. *Opt. Lett.* **43**, 547–550 (2018). doi: 10.1364/OL.43.000547; pmid: 29400837
114. E. Lucas *et al*., Spatial multiplexing of soliton microcombs. arXiv:1804.03706v2 [physics.optics] (2018). doi: arxiv.org/abs/1804.03706v2
115. A. B. Matsko, A. A. Savchenkov, L. Maleki, Normal group-velocity dispersion Kerr frequency comb. *Opt. Lett.* **37**, 43–45 (2012). doi: 10.1364/OL.37.000043; pmid: 22212785
116. C. Godey, I. V. Balakireva, A. Coillet, Y. K. Chembo, Stability analysis of the spatiotemporal Lugiato-Lefever model for Kerr optical frequency combs in the anomalous and normal dispersion regimes. *Phys. Rev. A* **89**, 063814 (2014). doi: 10.1103/PhysRevA.89.063814
117. W. Liang *et al*., Generation of a coherent near-infrared Kerr frequency comb in a monolithic microresonator with normal GVD. *Opt. Lett.* **39**, 2920–2923 (2014). doi: 10.1364/OL.39.002920; pmid: 24978237
118. S. W. Huang *et al*., Mode-locked ultrashort pulse generation from on-chip normal dispersion microresonators. *Phys. Rev. Lett.* **114**, 053901 (2015). doi: 10.1103/PhysRevLett.114.053901; pmid: 25699441
119. X. Xue *et al*., Mode-locked dark pulse Kerr combs in normal-dispersion microresonators. *Nat. Photonics* **9**, 594–600 (2015). doi: 10.1038/nphoton.2015.137
120. J. K. Jang *et al*., Dynamics of mode-coupling-induced microresonator frequency combs in normal dispersion. *Opt. Express* **24**, 28794–28803 (2016). doi: 10.1364/OE.24.028794; pmid: 27958523
121. V. E. Lobanov, G. Lihachev, M. L. Gorodetsky, Generation of platicons and frequency combs in optical microresonators with normal GVD by modulated pump. *Europhys. Lett.* **112**, 54008 (2015). doi: 10.1209/0295-5075/112/54008
122. X. X. Xue *et al*., Normal-dispersion microcombs enabled by controllable mode interactions. *Laser Photonics Rev.* **9**, L23–L28 (2015). doi: 10.1002/lpor.201500107
123. P. Parra-Rivas, E. Knobloch, D. Gomila, L. Gelens, Dark solitons in the Lugiato-Lefever equation with normal dispersion. *Phys. Rev. A* **93**, 063839 (2016). doi: 10.1103/PhysRevA.93.063839
124. V. E. Lobanov, A. V. Cherenkov, A. E. Shitikov, I. A. Bilenko, M. L. Gorodetsky, Dynamics of platicons due to third-order dispersion. *Eur. Phys. J. D* **71**, 185 (2017). doi: 10.1140/epjd/e2017-80148-0
125. P. Parra-Rivas, D. Gomila, L. Gelens, Coexistence of stable dark- and bright-soliton Kerr combs in normal-dispersion resonators. *Phys. Rev. A* **95**, 053863 (2017). doi: 10.1103/PhysRevA.95.053863
126. S. B. Papp *et al*., Microresonator frequency comb optical clock. *Optica* **1**, 10 (2014). doi: 10.1364/OPTICA.1.000010
127. P. Del'Haye *et al*., *Nat. Photonics* **10**, 516–520 (2016).
128. W. Liang *et al*., *IEEE Photonics J.* **9**, 11 (2017).
129. T. Ideguchi *et al*., Coherent Raman spectro-imaging with laser frequency combs. *Nature* **502**, 355–358 (2013). doi: 10.1038/nature12607; pmid: 24132293
130. V. Ataie, D. Esman, B. P.-P. Kuo, N. Alic, S. Radic, Subnoise detection of a fast random event. *Science* **350**, 1343–1346 (2015). doi: 10.1126/science.aac8446; pmid: 26659052
131. M. Pu, L. Ottaviano, E. Semenova, K. Yvind, Efficient frequency comb generation in AlGaAs-on-insulator. *Optica* **3**, 823 (2016). doi: 10.1364/OPTICA.3.000823
132. Y. Okawachi *et al*., Quantum random number generator using a microresonator-based Kerr oscillator. *Opt. Lett.* **41**, 4194–4197 (2016). doi: 10.1364/OL.41.004194; pmid: 27628355
133. M. Kues *et al*., On-chip generation of high-dimensional entangled quantum states and their coherent control. *Nature* **546**, 622–626 (2017). doi: 10.1038/nature22986; pmid: 28658228
134. H. Taheri, A. B. Matsko, L. Maleki, Optical lattice trap for Kerr solitons. *Eur. Phys. J. D* **71**, 153 (2017). doi: 10.1140/epjd/e2017-80150-6
135. P. Ghelfi *et al*., A fully photonics-based coherent radar system. *Nature* **507**, 341–345 (2014). doi: 10.1038/nature13078; pmid: 24646997
136. D. T. Spencer *et al*., An optical-frequency synthesizer using integrated photonics. *Nature* **557**, 81–85 (2018). doi: 10.1038/s41586-018-0065-7; pmid: 29695870

ACKNOWLEDGMENTS

The authors thank M. Karpov and A. Lukashchuk for assistance in preparing the figures, as well as S. Höhnle from IBM for rendering Fig. 1. **Author contributions:** This Review was written by T.J.K. and M.L.G. with critical input and support of A.L.G. and M.L. All authors contributed and discussed the content of the Review. **Funding:** This material is based on work supported by the Defense Advanced Research Projects Agency (DARPA), Defense Sciences Office (DSO) and (MTO), grants HR0011-15-C-0055 (DODOS), W31P4Q-16-1-0002 (SCOUT), and W31P4Q-14-C-0050 (PULSE); The European Space Technology Centre with ESA (4000116145/16/NL/MH/GM); the Swiss National Science Foundation via BRIDGE (176563); a Swiss National Science Foundation Grant and Ministry of Education and Science of the Russian Federation (project RFMEFI58516X0005); and the Air Force Office of Scientific Research, Air Force Material Command, USAF under awards FA9550-15-1-0099 and FA9550-15-1-0303. **Competing interests:** T.J.K. is a cofounder and shareholder of LiGenTec SA, a start-up company offering Si_3N_4 photonic integrated circuits as a foundry service. T.J.K., A.L.G., M.L.G., and M.L., are co-inventors of patents in the field of publication. Patents are owned by EPFL, Cornell, and the Max Planck Society.

10.1126/science.aan8083

RESEARCH ARTICLE SUMMARY

STRUCTURAL BIOLOGY

Structural basis for the recognition of Sonic Hedgehog by human Patched1

Xin Gong*, Hongwu Qian*, Pingping Cao*, Xin Zhao, Qiang Zhou, Jianlin Lei, Nieng Yan†

INTRODUCTION: The Hedgehog (Hh) pathway is critical for embryogenesis and tissue regeneration. Hh signaling is activated by binding of the secreted and lipid-modified protein Hh to the membrane receptor Patched (Ptch). In the absence of Hh, Ptch suppresses the downstream G protein–coupled receptor Smoothened (Smo) via an unknown and indirect mechanism. Binding of Hh to Ptch relieves the inhibition on Smo and turns on the signaling events that lead to the transcriptional activation of the Hh pathway. Aberrations of Hh signaling are associated with birth defects or tumorigenesis. Despite rigorous investigations, the molecular basis for the interplay among Hh, Ptch, and Smo remains unclear, and the structural basis for the recognition between Ptch and Hh is yet to be elucidated.

The 1447-residue human Ptch1 protein is predicted to contain 12 transmembrane segments (TMs) and to share structural similarity with the bacterial resistance-nodulation-division (RND) family transporters. TMs 2 to 6 of Ptch1 constitute the sterol-sensing domain (SSD), which has been found in several proteins involved in sterol transport and metabolism. The molecular mechanism for potential sterol binding or transport activity of these SSD-containing proteins remains elusive.

RATIONALE: To obtain a sample suitable for structural study, we generated several constructs of human Ptch1 based on sequence conservation and functional characterizations. Eventually, the truncated human Ptch1 containing residues 1 to 1305, which was transiently expressed in human embryonic kidney (HEK) 293F cells, exhibited a sufficient expression level and good solution behavior after affinity and size exclusion chromatography purification. Both oligomeric and monomeric states of Ptch1 were observed. The monomeric form was used for single-particle cryo–electron microscopy (cryo-EM) analysis because of its excellent behavior under cryo conditions.

ON OUR WEBSITE

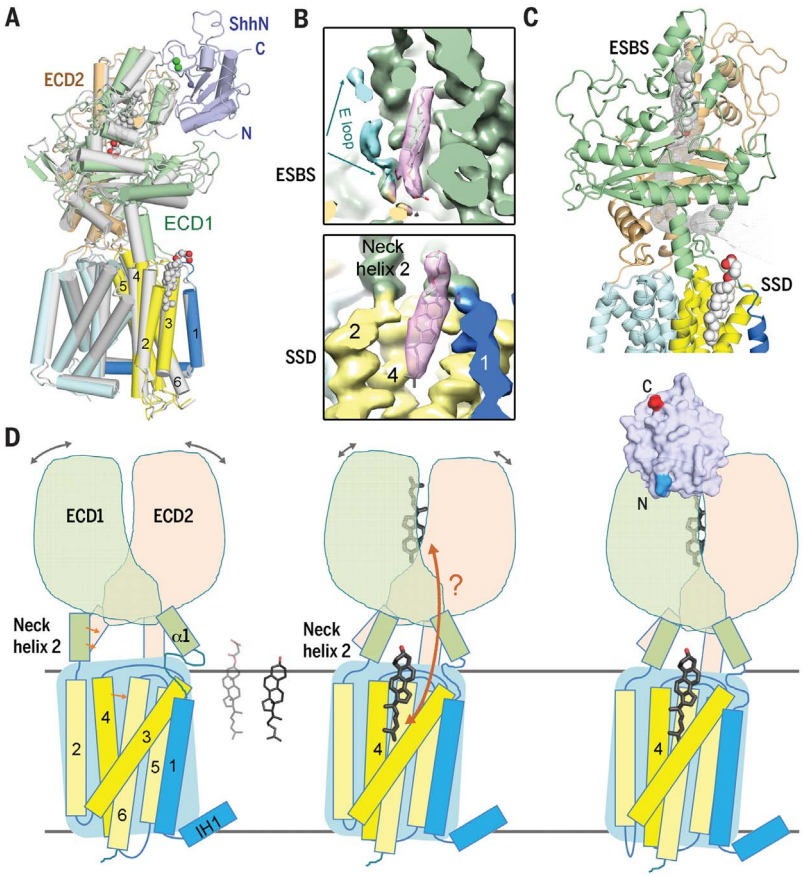
Read the full article at http://dx.doi.org/10.1126/science.aas8935

Among the three mammalian Hh homologs Sonic (Shh), Desert (Dhh), and Indian (Ihh), Shh has been the prototype for functional and mechanistic investigations. The N-terminal domain of human Sonic Hh (ShhN, residues 24 to 197) expressed and purified in *Escherichia coli* was able to form a stable complex with the detergent-solubilized Ptch1 protein in the presence of cholesteryl hemisuccinate (CHS).

RESULTS: The cryo-EM structures of human Ptch1 alone and in complex with ShhN were determined at overall resolutions of 3.9 Å and 3.6 Å, respectively. Two interacting extracellular domains, ECD1 and ECD2, and 12 TMs are resolved for Ptch1. ECD1 and ECD2 move toward each other upon arrival of ShhN and together constitute the docking site for ShhN. Detailed recognition between ShhN and Ptch1 was analyzed and biochemically confirmed.

Two elongated densities, both consistent with CHS, are observed in Ptch1 with or without ShhN, one in a pocket enclosed by the ECD domains and the other in a membrane-facing cavity of the SSD. Structure-guided biochemical analyses revealed steroid-dependent interaction between ShhN and Ptch1. The structure of a steroid binding–deficient Ptch1 mutant displays pronounced conformational rearrangements relative to the wild-type protein.

CONCLUSION: The structures of human Ptch1 and its complex with ShhN reveal the molecular basis for the recognition between Ptch1 and ShhN. The identification of two steroid-binding sites in Ptch1 establishes an important framework for future investigations of Hh signaling and provides critical insight into sterol perception by SSD-containing proteins. ■



Structural basis for sterol perception by the sterol-sensing domain (SSD) and steroid-dependent binding between Ptch1 and ShhN. (**A**) Structural comparison of Ptch1 alone (silver) and in complex with ShhN (domain-colored). (**B**) Two CHS binding sites on Ptch1, one enclosed by extracellular domains [the extracellular steroid binding site (ESBS)] and one on the SSD. (**C**) Potential sterol transfer between the two binding sites. (**D**) Functional implications of the structures.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: nyan@princeton.edu
Cite this article as X. Gong *et al.*, *Science* **361**, eaas8935 (2018). DOI: 10.1126/science.aas8935

RESEARCH ARTICLE

STRUCTURAL BIOLOGY

Structural basis for the recognition of Sonic Hedgehog by human Patched1

Xin Gong[1]*†, Hongwu Qian[1]*†, Pingping Cao[1]†, Xin Zhao[1], Qiang Zhou[1], Jianlin Lei[2], Nieng Yan[1]*‡

The Hedgehog (Hh) pathway involved in development and regeneration is activated by the extracellular binding of Hh to the membrane receptor Patched (Ptch). We report the structures of human Ptch1 alone and in complex with the N-terminal domain of human Sonic hedgehog (ShhN) at resolutions of 3.9 and 3.6 angstroms, respectively, as determined by cryo–electron microscopy. Ptch1 comprises two interacting extracellular domains, ECD1 and ECD2, and 12 transmembrane segments (TMs), with TMs 2 to 6 constituting the sterol-sensing domain (SSD). Two steroid-shaped densities are resolved in both structures, one enclosed by ECD1/2 and the other in the membrane-facing cavity of the SSD. Structure-guided mutational analysis shows that interaction between ShhN and Ptch1 is steroid-dependent. The structure of a steroid binding–deficient Ptch1 mutant displays pronounced conformational rearrangements.

The Hedgehog (Hh) signaling pathway, which has been investigated extensively in *Drosophila* and vertebrates, plays a pivotal role in embryogenesis and postnatal tissue maintenance and regeneration (*1*–*5*). Activation of Hh signaling is initiated by binding of the secreted protein Hh to the membrane-embedded receptor Patched (Ptch) in Hh-responsive cells (*6*–*8*). In the absence of Hh, Ptch inhibits Smoothened (Smo), a class F GPCR (G protein–coupled receptor), although the mechanism remains unclear (*9*–*11*). Binding of Hh to Ptch relieves the inhibition on Smo and turns on the signaling events that lead to the transcriptional activation of the Hh pathway (*12*–*17*). Compromised Hh pathway activity may result in birth defects, whereas aberrant activation of Hh signaling by suppressing Ptch or activating Smo has been implicated in tumorigenesis of several tissues, as exemplified by basal cell carcinoma and medulloblastoma (*18*–*22*).

In mammals, three Hh homologs have been identified—Sonic (Shh), Desert (Dhh), and Indian (Ihh)—among which Shh represents the prototype for functional and mechanistic elucidation (*5*). A ~450-residue Shh precursor undergoes autocatalytic cleavage and yields an N-terminal domain (ShhN) of ~20 kDa that is responsible for all known signaling activities. ShhN is modified with N-terminal palmitoyl and C-terminal cholesteryl moieties (*23*–*27*). Although unpalmitoylated ShhN exhibits decreased activity (*26*, *28*), these modifications are dispensable for the high-affinity binding to Ptch (*29*).

Because abnormal activation of the Hh pathway is associated with tumorigenesis, different types of inhibitors have been developed that target Hh signaling (*30*, *31*) (fig. S1). Most of these are antagonists for Smo or inhibitors of downstream components in the Hh pathway. However, some of them are designed to disrupt the interaction between Ptch1 and ShhN, such as Robotnikinin (*32*) and HL2-m5 macrocyclic peptide (*33*) (fig. S1). Structural information on the ShhN-Ptch1 complex may provide important insight into design or optimization of ligands to disrupt the formation of a complex between ShhN and Ptch1. Whereas a number of crystal structures of ShhN segments alone and in complex with different binding proteins have been reported (*34*), there has been a lack of structural information on Ptch.

Patched (*ptc*) was cloned as a segment-patterning gene in *Drosophila* (*35*, *36*). The human homolog was later identified as a tumor suppressor for basal cell nevus syndrome (also known as Gorlin syndrome) (*37*, *38*). There are two Ptch homologs in mammals, Ptch1 and Ptch2, both of which, despite their divergent expression patterns and physiological functions, bind to the three types of mammalian Hh ligands with similar affinity (*39*, *40*). The full-length human Ptch1 (hPtch1), consisting of 1447 amino acids, is predicted to contain 12 transmembrane segments (TMs), two extracellular domains (ECDs), and two intracellular domains (*37*). Ptch proteins share sequence similarity with the bacterial resistance-nodulation-division (RND) family transporters exemplified by the proton-driven multidrug resistance pump AcrB (*41*, *42*). TMs 2 to 6 of Ptch constitute the sterol-sensing domain (SSD) that has been found in a number of sterol transport and metabolism-related proteins, such as 3-hydroxy-3-methylglutaryl–coenzyme A reductase (HMGCR), SREBP cleavage-activating protein (SCAP), Niemann-Pick type C1 (NPC1), NPC1-like 1 (NPC1L1), and another Hh signaling component, Dispatched (Disp) (*43*). Although the structures of human NPC1 have recently been determined, the molecular mechanism of potential sterol binding or transport mediated by the SSD-containing proteins remains enigmatic (*44*, *45*).

Here, we report the cryo–electron microscopy (cryo-EM) structures of human Ptch1 alone and in complex with ShhN at overall resolutions of 3.9 Å and 3.6 Å, respectively. The structures reveal the detailed recognition between ShhN and Ptch1. Unexpectedly, two steroid molecules are observed, one in a pocket enclosed by the ECDs and one in the membrane-facing cavity of the SSD. Structure-guided biochemical analyses suggest that the interaction between ShhN and Ptch1 is steroid-dependent—an observation corroborated by additional structural evidence.

Structural determination of human Ptch1 and the Ptch1-ShhN complex

To obtain a sample suitable for structural study, we generated several constructs of human Ptch1 based on sequence conservation and functional characterizations (fig. S2) (*8*, *46*, *47*). Consistent with a previous observation in a cell-based binding assay that the C terminus–truncated mouse Ptch1 binds to ShhN with affinity similar to that of wild-type Ptch1 (Ptch1-WT) (*8*), the C terminal intracellular domain of mouse Ptch1 was shown to be dispensable for Ptch1-dependent regulation of canonical Hh signaling (*46*, *47*). We therefore tested a number of C-terminus truncations and identified an optimal construct for human Ptch1 (residues 1 to 1305) that exhibited sufficient expression level and good solution behavior (fig. S3). [See below for details of the subcloning, transient human embryonic kidney (HEK) 293F cell expression, purification, and cryo-EM data acquisition of Ptch1.]

Consistent with the reported oligomerization of Ptch1 (*47*, *48*), recombinant Ptch1 eluted in two peaks upon size-exclusion chromatography (SEC), suggesting distinct oligomerization states (fig. S3B). Preliminary assessment using cryo-EM showed that protein particles in the earlier peak were heterogeneous, posing a major challenge for structural determination. In contrast, the proteins in the later peak, despite its smaller size, appeared homogeneous and displayed multiple orientations in the thin layer of vitreous ice. We therefore focused on the monomeric form of Ptch1 for cryo-EM analysis.

An EM reconstruction was obtained at 3.9 Å resolution from 94,445 selected particles (fig. S3C) according to the gold-standard Fourier shell correlation 0.143 criterion. Although the construct used for structural determination contains 1305 residues, all the intracellular segments—including the 72 residues on the N terminus, residues 608 to 730 that connect TMD1 (TMs 1 to 6)

[1]State Key Laboratory of Membrane Biology, Beijing Advanced Innovation Center for Structural Biology, Tsinghua-Peking Joint Center for Life Sciences, School of Life Sciences and School of Medicine, Tsinghua University, Beijing 100084, China. [2]Technology Center for Protein Sciences, Ministry of Education Key Laboratory of Protein Sciences, School of Life Sciences, Tsinghua University, Beijing 100084, China.
*Present address: Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA. †These authors contributed equally to this work.
‡Corresponding author. Email: nyan@princeton.edu

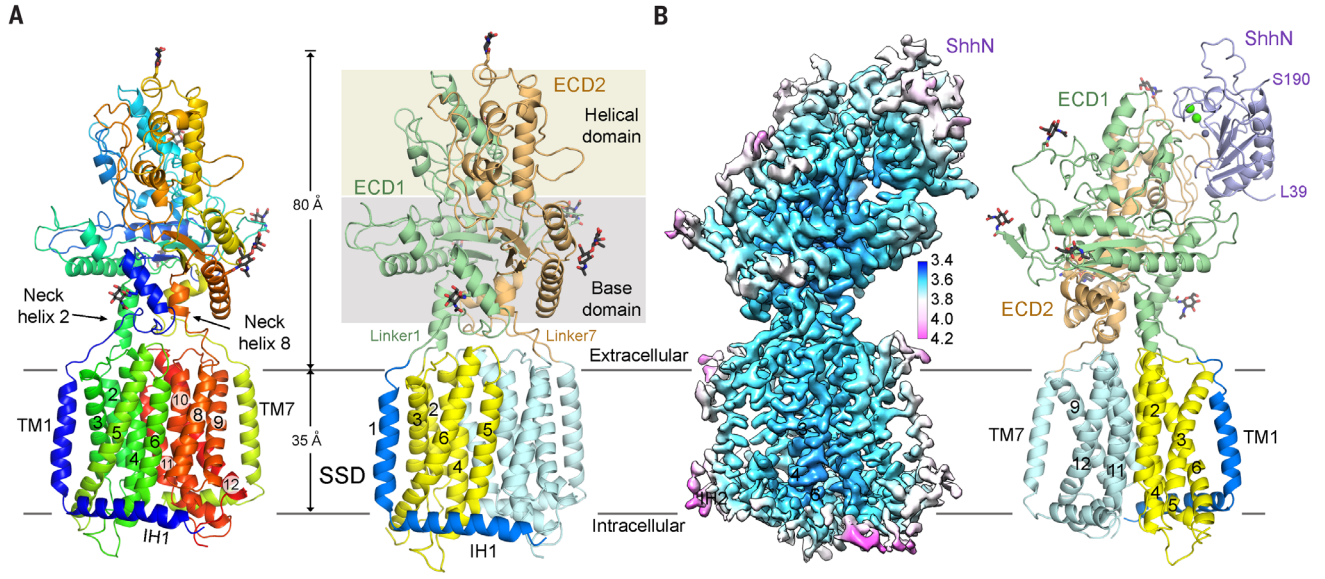


Fig. 1. Cryo-EM structures of human Patched1 (Ptch1) alone and in complex with ShhN. (**A**) Overall structure of Ptch1. The structures shown on the left and right are rainbow and domain-colored, respectively. The sterol-sensing domain (SSD) that consists of TMs 2 to 6 is colored yellow. The sugar moieties are shown as sticks. See fig. S7 for the detailed structural illustration of individual domains. Unless otherwise indicated, the same color code is applied to all figures. (**B**) Structure of Ptch1 in complex with ShhN. Left: The resolution map for the Ptch1-ShhN complex. The color code for resolutions, shown with the unit Å, is calculated using Relion 2.0 (*73*) and generated in Chimera (*80*). Right: Overall structure of the Ptch1-ShhN complex. The Ca^{2+} and Zn^{2+} ions are shown as green and purple spheres, respectively. All structure figures were prepared using PyMol (*81*).

and TMD2 (TMs 7 to 12), and the C-terminal segment after Ser^{1185}—are invisible, likely owing to intrinsic flexibility. In total, 990 residues were built with 969 side chains assigned (Fig. 1A, Table 1, and figs. S4A, S5A, and S6A).

The well-resolved regions include two ECDs and the 12 TMs that display a typical RND family architecture with a two-fold pseudosymmetry around an axis that is perpendicular to the membrane (Fig. 1A and fig. S7A) (*44*, *49*). The two TMDs, each preceded by an amphiphilic helix that lies on the intracellular boundary of the membrane (designated IH), can be superimposed with a root mean square deviation (RMSD) of 3.7 Å over 186 Cα atoms (fig. S7B).

The two ECDs in Ptch1, ECD1 between TM1 and TM2 and ECD2 between TM7 and TM8, correspond to domain C and domain I in NPC1, respectively (*44*). Each ECD is linked to the N-terminal TM by a long linker (linker 1 for ECD1 and linker 7 for ECD2) and to the C-terminal TM by a Neck helix (Fig. 1A and fig. S7C). ECD1 and ECD2 both contain a topologically similar core region right above the TMD, which we term the "base domain" (Fig. 1A and fig. S7, C and D). Despite similar folds of the two base domains, ECD1 and ECD2 display divergent spatial arrangements and cannot be well superimposed (fig. S7D).

On top of the base domains, ECD1 and ECD2 each contain an extra helical domain that is absent in NPC1 (Fig. 1A). Notably, both ECDs are enriched in lengthy loops (fig. S7C). For instance, the segment (residues 943 to 969) that connects β2 and β3 in the base domain of ECD2 forms an upward-extending loop hairpin (designated the H loop) that represents part of the helical domain (figs. S2 and S7, A and C). The two ECDs interact with each other through an extended

Table 1. Summary of data collection and model statistics.

Data collection	Ptch1-WT	Ptch1-ShhN	Ptch1-3M
EM		Titan Krios (Thermo Fisher)	
Voltage (kV)		300	
Detector		K2 Summit (Gatan)	
Pixel size (Å/pixel)		1.091	
Electron dose (e^-/Å^2)		50	
Number of micrographs	3983	4221	6680
Reconstruction			
Software		RELION 2.0	
Number of used particles	94,445	137,823	154,721
Symmetry		C1	
Resolution	3.9 Å	3.6 Å	4.1 Å
Map sharpening *B*-factor (Å^2)	–170	–130	–197
Model building			
Software		COOT	
Refinement			
Software		Phenix	
Model composition			
Protein residues	990	1144	982
Side chain	969	1132	969
Sugar moieties	7	7	7
Ligands	2	2	0
Metal ions	0	3	0
Validation			
RMS deviations			
Bond length (Å)	0.01	0.01	0.01
Bond angle (degrees)	1.55	1.61	1.47
Ramachandran plot statistics (%)			
Preferred	86.21	87.42	88.63
Allowed	13.18	11.61	11.07
Outlier	0.61	0.97	0.31

yet loose interface and both project into the extracellular space with similar heights of approximately 80 Å (Fig. 1A). Six glycosylation sites were observed, four on ECD1 and two on ECD2 (fig. S7C).

For structural determination of the complex between Ptch1 and ShhN, we overexpressed human ShhN (residues 24 to 197) in *Escherichia coli* and incubated the purified protein with the C terminus–truncated human Ptch1 for cryo-EM analysis. (See below for details of the reconstitution, sample preparation, and cryo-EM data acquisition of the Ptch1-ShhN complex.) A total of 137,823 selected particles yielded a reconstruction at 3.6 Å resolution, which contains density for ShhN (Fig. 1B, Table 1, and figs. S4B and S5B). The crystal structure of ShhN (PDB code 4C4M) was docked into the density with manual adjustment. Residues 39 to 190 of ShhN were resolved (Fig. 1B). The improved resolution allows more accurate structural modeling of Ptch1 (Fig. 1B and fig. S6, B and C). Most of the structural descriptions hereafter are based on the complex structure if not otherwise noted.

Interface between Ptch1 and ShhN

ShhN binds to the Ptch1 surface constituted by the upper helical domains of the two ECDs (Fig. 2A). An extended loop (residues 206 to 213) on ECD1 designated the E loop (because it contains the conserved Glu^{212}) (fig. S2), the ensuing helix α3 on ECD1, and the H loop on ECD2 constitute the primary binding site for ShhN (Fig. 2A).

In addition to Ptch and Smo, Hh signaling is subject to regulation by cell surface or extracellular proteins, such as the co-receptors Cdon/Boc (equivalent of Ihog/Boi in invertebrates), the extracellular Hh antagonist Hhip (Hh interacting protein), and the antagonist antibody 5E1 (*34*). ShhN, which shares structural homology to a bacterial carboxypeptidase but lacks hydrolytic activity, contains a Lys/Arg-enriched pseudo–active site groove that represents the primary surface for many binding proteins (*8*). Examination of the structure of the Ptch1-ShhN complex reveals that the Ptch1-interacting surface, which involves the pseudo–active site groove, overlaps with that for other binding proteins, including the FNIII repeat of Cdo (CdoFn3), Hhip, and 5E1 Fab (fig. S8). The structure thus provides the basis for the observed competitive binding for Hh by Ptch1 and several binding proteins (*50*–*53*).

The ShhN-Ptch1 interface is primarily mediated by polar and charged residues (Fig. 2B). ECD1 and ECD2 each contribute multiple acidic residues for interaction with the Lys/Arg patch that lines the pseudo–active site groove of ShhN. The reported resolution for the interface is approximately 3.8 Å (Fig. 1B), insufficient for accurate distance measurement between atoms that may form electrostatic interactions. Nonetheless, the backbones are well defined, supporting analysis of the interface residues. Glu^{212} on the E loop of ECD1 appears to be a coordination center, interacting with Lys^{87}, Arg^{123}, and Arg^{153} on ShhN. Glu^{221} on helix α3 interacts with Tyr^{44} and Lys^{45} of ShhN, and the preceding Asp^{217} is hydrogen-bonded to Ser^{177} (Fig. 2B, left). ECD2 mainly engages Glu^{947}, Asp^{951}, Asp^{954}, and Glu^{958} on the H loop for ShhN binding (Fig. 2B, right).

The structural observation was verified by mutational analysis. We used wild-type or mutated ShhN proteins that were fused to MBP as bait to pull down Ptch1 variants (Fig. 2C). Single point mutations of the aforementioned ShhN or Ptch1 residues to opposite charges or double mutations to Ala all resulted in compromised complex formation. As a control, Ala substitution of ShhN-Thr^{77} (T77A), which is not engaged in Ptch1 interactions, had no effect on Ptch1 binding (Fig. 2C).

An extracellular steroid-binding site in Ptch1

Structural comparison of Ptch1 in the absence and presence of ShhN reveals marked shifts of

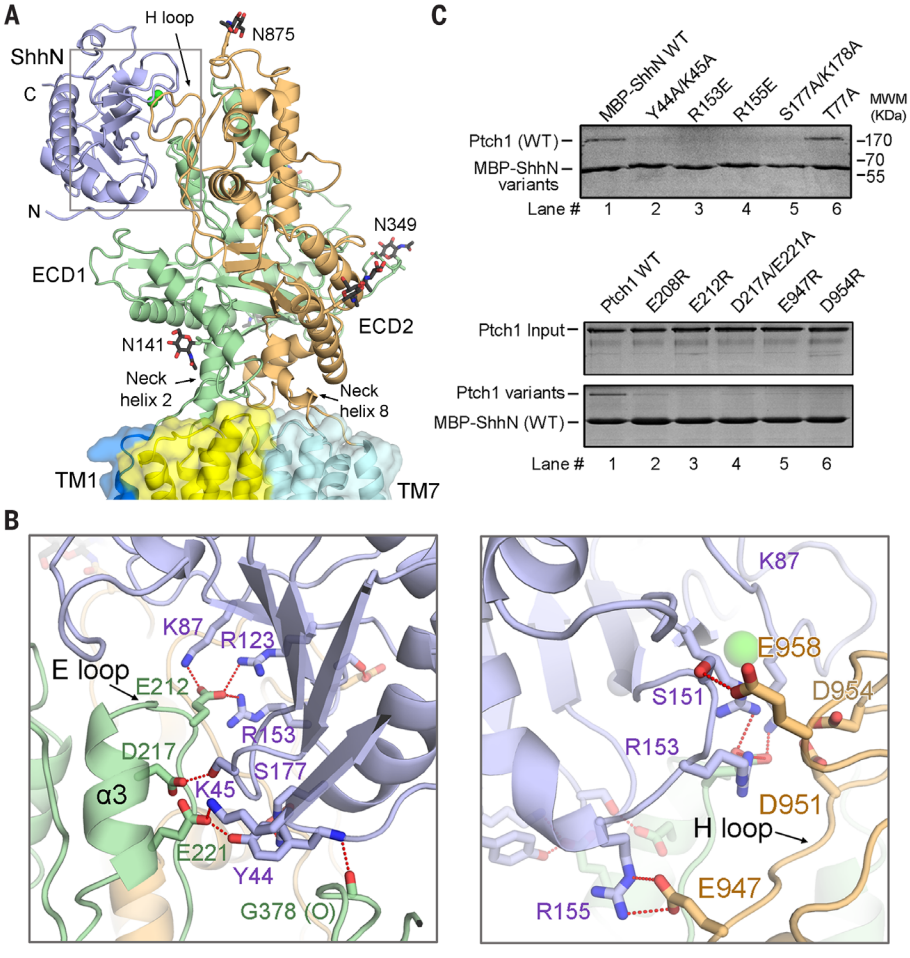
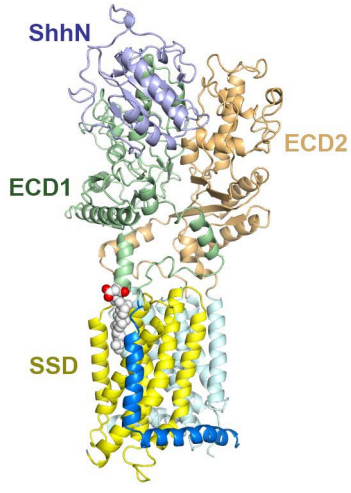


Fig. 2. Specific recognition between Ptch1 and ShhN. (**A**) The two extracellular domains, ECD1 and ECD2, of Ptch1 together constitute the docking site for ShhN. (**B**) The ECDs of Ptch1 recognize ShhN through extensive polar interactions. The left and right panels highlight the specific recognition of ShhN by ECD1 and ECD2, respectively. The potential electrostatic interactions are represented by dotted red lines. (**C**) Biochemical verification of the observed interface.



Movie 1. Conformational changes of Ptch1 upon ShhN binding. The morph was generated using the ShhN-free and -bound structures of Ptch1 as the first and end frames. The structures are shown as cartoon and domain-colored. The morph was generated by aligning the two structures and generating 60 intermediates in Chimera. The morphs were merged and displayed in PyMol using the "mset" command.

the ECDs, whereas the transmembrane segments remain unchanged (Movie 1). The two ECDs, which exhibit limited intradomain rearrangements except for some loop regions, move toward each other upon ShhN binding, resulting in increased interface between ECD1 and ECD2 from ~3830 Å^2 to ~4370 Å^2 (Fig. 3A). Helix α1, which connects to TM1 via linker 1, swings together with ECD2 (Fig. 3A, left). The lengthy linker 1 and the C-terminal short loop that connect the Neck helix to the TMD may confer the structural plasticity for the observed domainwise swing motions. Note that whereas the residues in the E loop in ECD1 are well resolved in the Ptch1-ShhN complex, even the backbone of the corresponding segment is broken in the map for Ptch1 alone, which suggests that this segment is stabilized by ShhN (Fig. 3B).

In addition to accommodating ShhN, the E loop covers a pocket on the interface of ECD1 and ECD2. Unexpectedly, a stretch of density is observed in this elongated pocket in both EM reconstructions with or without ShhN (Fig. 3, C and D, and fig. S6D). In the EM map for the Ptch1-ShhN complex (3.6 Å resolution), the density sealed between ECD1 and ECD2 is better defined and has a size and shape consistent with a cholesteryl hemisuccinate (CHS) molecule. The CHS molecule is localized on the interface between the upper helical domains of the two ECDs, where ECD1 provides the primary coordination. Nearly 20 hydrophobic residues on the E loop and helices α3, α6, and α7 from ECD1 form the contour of the pocket for CHS, whereas ECD2 only engages two small hydrophobic residues, Val^{932} and Ala^{935}, on its helix α7 (Fig. 3D).

Although CHS appears to be sealed within the ECDs, it was in the buffer used for protein extraction, suggesting dynamic access to this extracellular steroid-binding site (ESBS). As the E loop and helix α3 on ECD1 are also the major constituents of the docking site for ShhN (Fig. 2B),

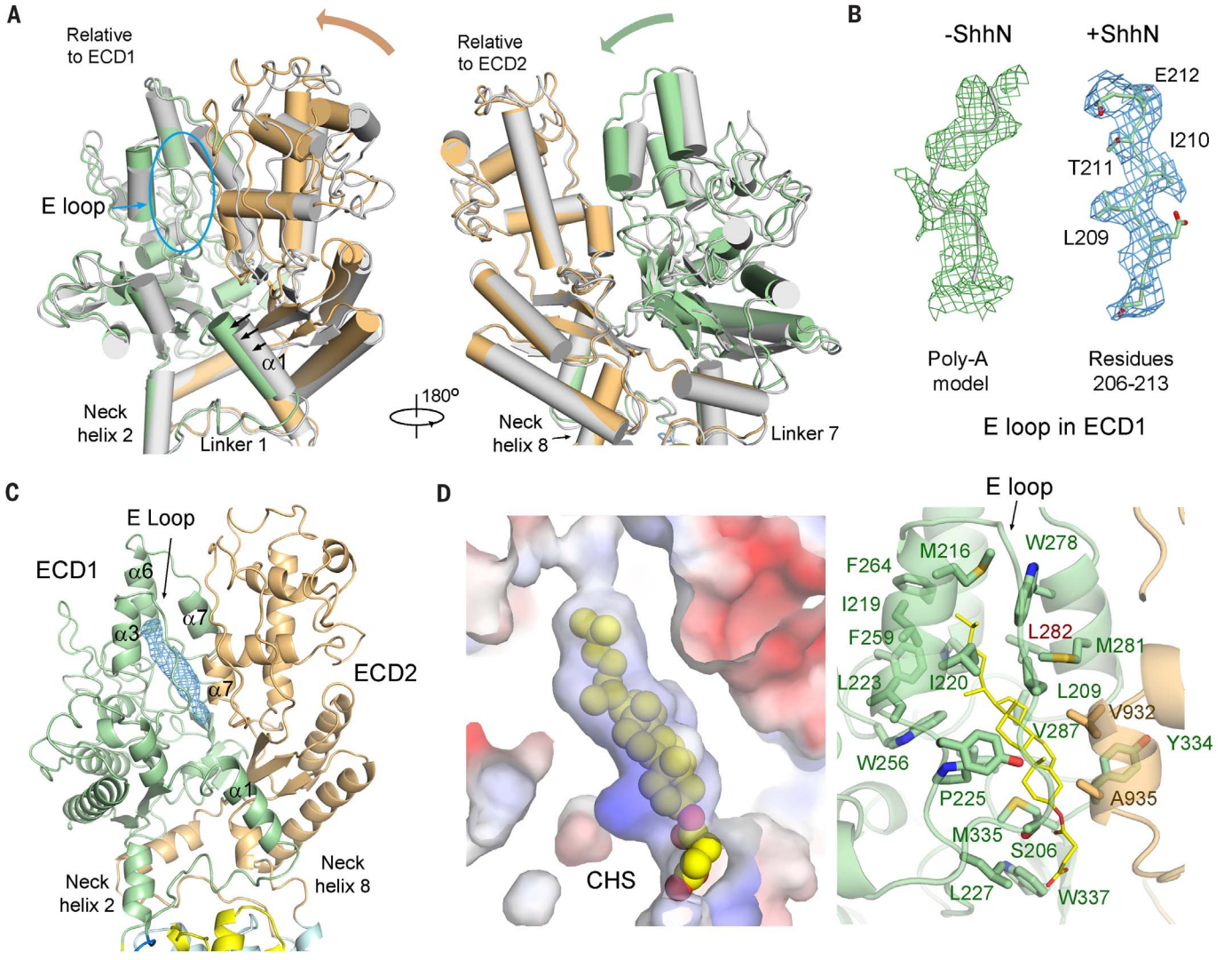


Fig. 3. The extracellular steroid-binding site (ESBS) is sealed upon ShhN binding. (**A**) The two ECDs of Ptch1 close up upon ShhN binding. The structural comparison of Ptch1 in the absence (silver) and presence (domain-colored) of ShhN is made relative to ECD1 (left) and ECD2 (right). See Movie 1 for the morph illustrating the conformational changes of Ptch1 upon ShhN association. (**B**) The E loop in ECD1 appears to be stabilized by ShhN. Shown here are the local EM maps, both contoured at 5σ, for the E loop in Ptch1 alone (left) and in the presence of ShhN (right). (**C**) A CHS-like density is embedded in the cavity enclosed by ECD1 and ECD2 in Ptch1. Shown here is the density in the Ptch1-ShhN complex. For the corresponding density in Ptch1 alone, see fig. S6D. (**D**) ECD1 provides the primary accommodation site for the CHS molecule. Left: The contour and composition of the cavity define the orientation of the bound CHS. Shown here is a cut-open view of the electrostatic surface potential of Ptch1 in the complex. CHS is shown as yellow spheres. Right: The E loop and helix α3 in ECD1 and helix α7 in ECD2 serve as the lid that shields the ligand. The residues that constitute the pocket are shown as sticks. Note that a cholesterol molecule can be accommodated in the hydrophobic pocket the same as CHS.

the different quality of the EM maps for this region with or without ShhN suggests that binding of ShhN may restrict the motion of these gating elements, hence potentially limiting the access to or exit from the ESBS.

Coordination of the steroidal moiety by the SSD

A similar stretch of density is found in a surface cavity on the membrane-facing side of the SSD. Although the size and shape of the density are also consistent with a CHS, we cannot exclude the possibility that it may belong to an endogenous molecule or a digitonin (Fig. 4A). The hydrophobic composition and the V shape of the SSD cavity define the orientation of the bound ligand. We modeled a CHS molecule into the density. A number of conserved hydrophobic residues on TMs 2 to 4 of the SSD shape the cavity that accommodates the cholesteryl moiety (Fig. 4A, insets, and fig. S9).

Although the ligand seems to bind to a similar location as the computationally predicted docking site on the SSD of NPC1 (*45*), the exact position and detailed coordination may be different. The SSDs exhibit considerable conformational divergence between Ptch1 and NPC1 (Fig. 4B). In both proteins, a well-characterized Pro residue (Pro^{691} in NPC1 and Pro^{504} in Ptch1) maps to the bottom of the SSD pocket and directly participates in the ligand interaction (Fig. 4C and fig. S9A). The single point mutation $Pro^{691} \rightarrow$ Ser, which is a disease mutation associated with Niemann-Pick disease type C, was shown to eliminate the cross-linking of NPC1 with the compound U18666A, a steroid blocker for NPC1-mediated cholesterol egress (*54*). Ptch1-P504L ($Pro^{504} \rightarrow$ Leu) is associated with Gorlin syndrome (*55*). Structural analysis of Ptch1 suggests that substitution of the Pro with a polar residue (Ser) or a bulky residue (Leu) may alter the chemical environment or change the contour of the pocket, hence making it less favorable for accommodating a cholesteryl moiety (Fig. 4, A and C).

The interaction between ShhN and Ptch1 is ligand-dependent

To validate the structural observations and to investigate the functional relevance of the bound steroid ligands, we generated three Ptch1 variants with point mutations designed to disrupt

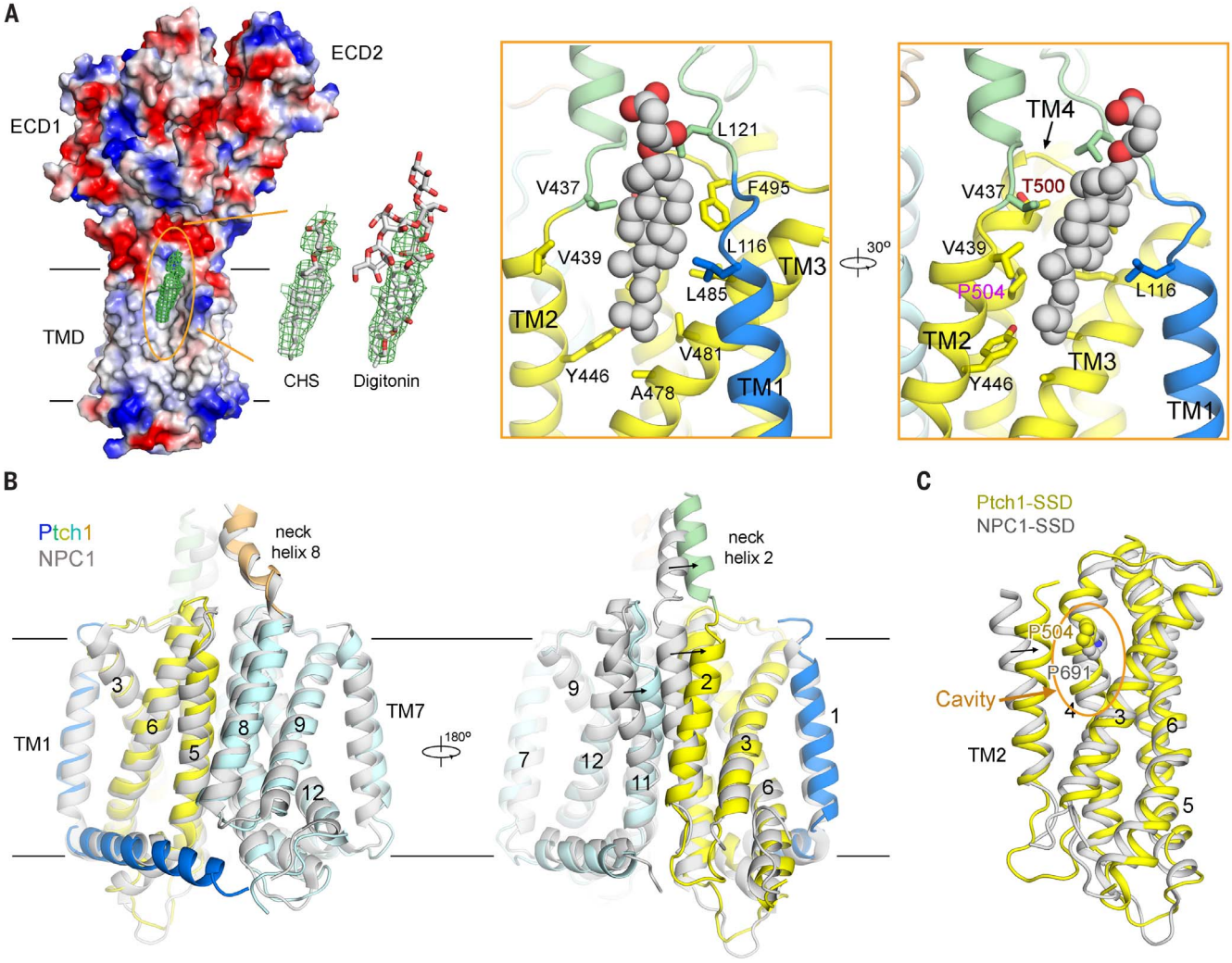


Fig. 4. Coordination of the cholesteryl moiety by the sterol-sensing domain (SSD). (**A**) The cavity on the membrane-facing side of SSD can accommodate a cholesterol-like molecule. The density is consistent with a CHS molecule, although the steroidal backbone of digitonin can fit in the density in a less favorable manner. Insets: Coordination of the cholesteryl moiety by SSD residues. Hydrophobic residues on TMs 2 to 4 of SSD constitute a V-shaped cavity, defining the orientation of the ligand should it be a cholesterol or cholesteryl derivative. Note that the hemisuccinate group of CHS is not specifically coordinated. Therefore, the observed interactions would be the same for a cholesterol molecule. (**B**) Structural comparison of the TMDs of Ptch1 and NPC1 (PDB code 5U74). TM2, TM11, and Neck helix 2 exhibit pronounced conformational changes between Ptch1 and NPC1. (**C**) Conformational divergence of the SSD between Ptch1 and NPC1. Ptch1-TM2 moves toward TM3, resulting in a narrower SSD cavity relative to that in NPC1-SSD. Within the cavity, a conserved and functionally important Pro (Pro^{691} in NPC1, Pro^{504} in Ptch1) is mapped to the side wall.

steroid binding: Ptch1-L282Q (Leu^{282} → Gln), where the hydrophobic residue within the ESBS is replaced by a polar one (Fig. 3D); Ptch1-T500F/P504L (Thr^{500} → Phe and Pro^{504} → Leu), in which the substitution with bulky residues may interfere with steroid binding to the SSD (Fig. 4A, right inset); and Ptch1-3M (with all three of these point mutations). It is difficult to directly measure the affinities between steroid molecules and the Ptch1 variants because of the interference by the detergent micelles. We therefore sought to solve the structures of Ptch1 variants using protocols identical to those for wild-type proteins.

Because the Ptch1-ShhN complex exhibits better resolution, we had planned to solve the structure of the complex. However, complex formation was markedly compromised between ShhN and all three Ptch1 variants. In particular, Ptch1-3M could be pulled down only minimally by MBP-ShhN (Fig. 5A). To semiquantitatively compare the affinity with ShhN between these variants and Ptch1-WT, we performed a concentration gradient titration of Ptch1 proteins (Fig. 5B). The affinity for ShhN of the ESBS mutant L282Q and the SSD mutant T500F/P504L appeared to be weakened by factors of approximately 16 and 4, respectively, whereas that of Ptch1-3M was reduced by more than a factor of 64. This observation seems to suggest a synergistic effect on the interaction with ShhN by the two steroid-binding sites (Fig. 5B).

To validate that the reduced ShhN binding was indeed related to steroid binding, we next examined the effect of CHS on complex formation (Fig. 5C). Ptch1 purified with *n*-dodecyl-β-D-maltopyranoside (DDM) barely interacted with ShhN, whereas addition of CHS restored complex formation (Fig. 5C, lanes 1 and 2). CHS was also required for ShhN binding when Ptch1 was purified with the detergents $C_{12}E_8$ or Cymal-6, which have distinct polar and hydrophobic moieties (Fig. 5C, lanes 3 to 6). In contrast, digitonin was sufficient to support complex formation (Fig. 5C, lane 7).

Taken together, our biochemical analyses suggest that complex formation between ShhN and Ptch1 may depend on the steroid ligand. To elucidate the molecular basis for this biochemical discovery, we focused on the structural determination of Ptch1-3M.

Pronounced conformational changes of Ptch1 upon steroid binding

The protocol for purification and data collection of Ptch1-3M is identical to that of Ptch1-WT. The numbers of the initial (>3.3 million) and selected (154,721) particles for 3D map reconstruction were both higher than that for Ptch1-WT alone and in complex with ShhN (Table 1); however, the final resolution was 4.1 Å for the overall structure of Ptch1-3M (figs. S4C, S5A, and S10A). As anticipated, the densities corresponding to CHS in the ESBS and SSD diminished, confirming the disrupted steroid binding by these point mutations (Fig. 6A).

Surprisingly, the overall structure of Ptch1-3M undergoes striking rearrangements from that of the wild-type protein (Movie 2). When the two structures are overlaid relative to the SSD, both TMD1/2 and ECD1/2 undergo marked twists (Fig. 6B). Individual domain comparison shows minor interdomain shifts between the two ECDs (Fig. 6C and Movie 2). Within ECD1, the E loop, whose backbone exhibits broken density in Ptch1-WT alone (Fig. 3B), is invisible in Ptch1-3M, indicating increased flexibility in the absence of CHS (Fig. 6A). Meanwhile, helix α3 moves toward the pocket (Fig. 6C). The resolution of the ECDs in Ptch1-3M is also considerably lower than that in Ptch1-WT (fig. S10), suggesting increased overall mobility. Because ECD1 and ECD2 need to move closer to form the ShhN docking site (Fig. 3A), the increased flexibility of the ECDs, particularly of the E loop and helix α3, may account for the reduced binding affinity with ShhN when the ESBS is not occupied by a steroid ligand.

Structural comparison of Ptch1-3M and Ptch1-WT affords an opportunity to investigate how the SSD senses sterol. Because none of the three residues Leu^{282}, Thr^{500}, or Pro^{504} is involved in intraprotein interaction, the structural shifts from Ptch1-3M (steroid-free) to Ptch1-WT (steroid-bound) may be interpreted as the conformational changes induced by arrival of the steroid ligand.

Binding of the steroid molecule to the SSD causes Neck helix 2, TM2, and TM4 to tilt toward the ligand (Fig. 6, D and E, and Movie 2). Such conformational shifts are similar to those observed in the structural comparison between ligand-free NPC1 and Ptch1-WT (Fig. 4, B and C). However, the pronounced relative rotation of the two TMDs around an axis that is perpendicular to their interface is unexpected (Fig. 6D, left, and Movie 2). Detailed structural comparison of Ptch1-3M and Ptch1-WT reveals the interdomain interactions that underlie the coupled conformational changes of TMD1/2 and ECD1/2 (Fig. 6, B and D, and Movie 3).

The short linker between TM5 and TM6 shifts concertedly with TM4 upon steroid binding (Fig. 6E) and pulls along linker 1 and helix α1 in ECD1. The motion of helix α1 induces the domainwise motion of ECD2 (Fig. 6C and Movie 3). The two base domains and the short linker between TM11 and TM12 move concomitantly with Neck helix 2 through extensive interactions (Movie 3).

The coupled movements of the four domains, TMD1/2 and ECD1/2, predict that mutations in one domain may alter an event in another domain through allosteric action. Supporting this notion, lack of steroid binding to SSD may result in increased flexibility of the two ECDs, hence leading to reduced ShhN binding (Fig. 5, A and B). In addition, Ptch1 variants containing a single point mutation—G509V (Gly^{509} → Val) or D513Y

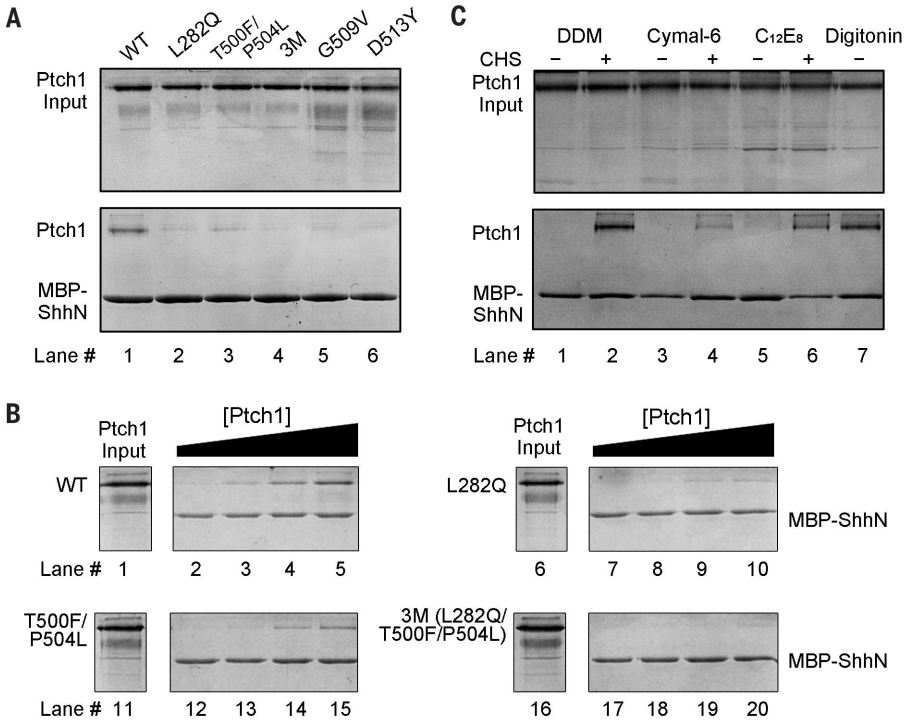


Fig. 5. SSD and ESBS mutations lead to reduced ShhN binding. (**A**) Point mutations in SSD or ESBS that are designed to disrupt steroid binding lead to reduced ShhN affinity. 3M denotes the Ptch1 variant containing L282Q, T500F, and P504L point mutations. (**B**) Concentration titration of Ptch1 variants in the pull-down assay semiquantitatively reveals the distinct effect of SSD and ESBS mutations on ShhN binding. (**C**) The interaction between purified Ptch1 and ShhN appears to be steroid-dependent. Ptch1 was purified in distinct detergents as indicated. CHS is required to retain the interaction when DDM, Cymal-6, or $C_{12}E_8$ was used for Ptch1 purification. Ptch1 purified in digitonin forms a complex with ShhN in the absence of CHS.

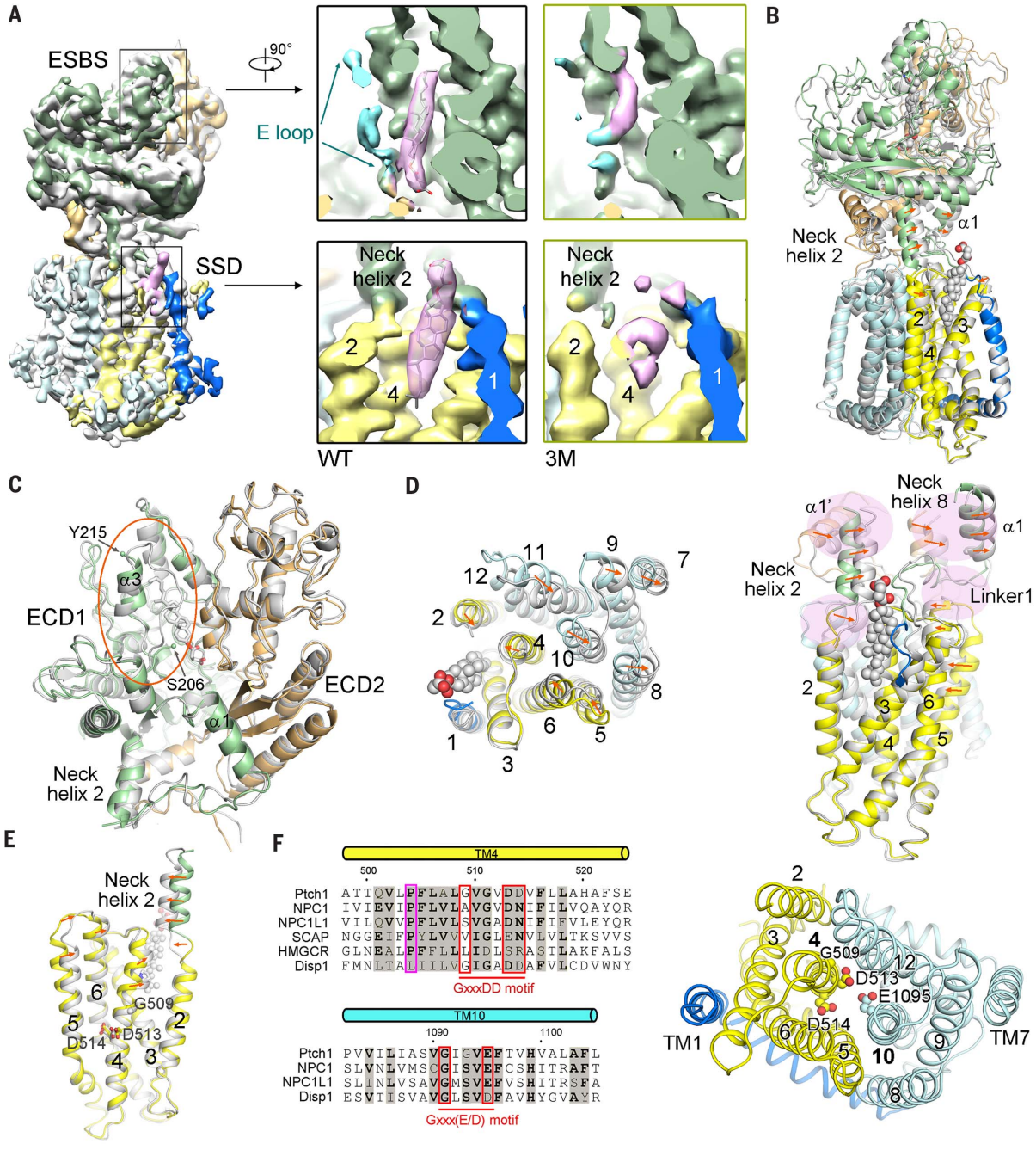


Fig. 6. Marked conformational changes of Ptch1 upon CHS binding. (**A**) Comparison of the EM maps of Ptch1-WT (silver) and Ptch1-3M (domain-colored). Both EM maps are low-pass filtered to 4.1 Å. The densities corresponding to the ligands in ESBS and SSD largely disappeared in the EM map of Ptch1-3M. Insets: The densities for CHS in ESBS (upper) and SSD (lower) are shown as semi-transparent contour in the low-pass filtered map of Ptch1-WT (left panels). Only shattered density debris remains at the corresponding positions in the map of Ptch1-3M (right panels). The E loop in ECD1 is no longer visible in Ptch1-3M. (**B**) Pronounced structural changes between Ptch1-WT (silver) and Ptch1-3M (domain-colored). The structural superimposition is made relative to SSD. See Movie 2 for the morph illustrating the overall conformational changes between the two structures. (**C**) ECDs undergo minor interdomain shifts between Ptch1-WT and Ptch1-3M. The two structures are superimposed relative to ECD1. The E-loop (residues 207 to 214) is invisible and helix α3 inclines toward the ESBS pocket. (**D**) Structural changes of Ptch1-TMD upon CHS binding. An extracellular view and a tilted side view are shown. The orange arrows indicate the shifts of the corresponding segments from Ptch1-3M to Ptch1-WT when the structural superimposition is made relative to SSD. The interactions that may be responsible for the coupled motions of different domains are highlighted by semitransparent pink patches. See Movie 3 for the morph illustrating the concerted movements of TMD1/2 and ECD1/2. (**E**) Structural changes of each SSD segment upon CHS binding. TM4, which moves toward the ligand, harbors the GxxxDD motif. (**F**) The acidic residues that are important for cotransport activity of RND members are conserved in Ptch1. Left: Sequence alignment of TM4 and TM10 from the indicated human SSD-containing proteins. The conserved and functionally crucial Pro in TM4 is highlighted by magenta box. The GxxxDD motif on TM4 and the Gxxx(E/D) motif on TM10, whose corresponding residues are essential for ion flux in the bacterial RND family, are highlighted by red boxes. Right: The conserved acidic residues on TM4/10 of Ptch1 are clustered in the center of the interface between TMD1 and TMD2. Shown here is an extracellular view of the Ptch1-TMD.

(Asp^{513} → Tyr), both of which are severe oncogenic mutations associated with Gorlin syndrome (*56*–*58*)—also exhibited reduced ShhN binding (Fig. 5A, lanes 5 and 6). Gly^{509} and Asp^{513} are localized in the middle segment of TM4 and mapped to the interface with TMD2 (Fig. 6F). As TM4 provides the primary site for steroid binding to SSD, these point mutations on the TMD2-facing side of TM4 may trap Ptch1 in a conformation that is incompatible for steroid binding to its SSD.

Discussion

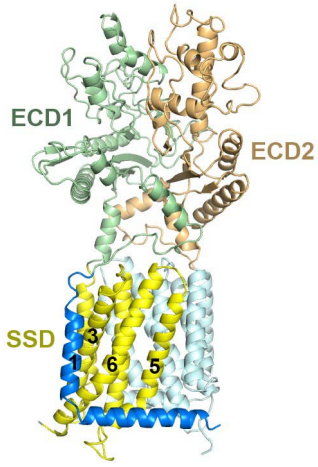
Structural implications for steroid perception by SSD

SSD-containing proteins are associated with cholesterol metabolism and transport (*43*). For instance, SCAP is a key player in the sterol regulatory element–binding protein (SREBP) pathway that controls the cellular homeostasis of sterol (*59*), HMGCR catalyzes the rate-limiting step of cholesterol synthesis (*60*), and NPC1 is essential for cholesterol egress from the lysosomes and late endosomes to other cellular compartments (*61*). However, direct evidence for sterol binding by the SSD has been missing. We find that Ptch1 possesses two sites that can accommodate steroid molecules, one within the ECDs and the other on the SSD.

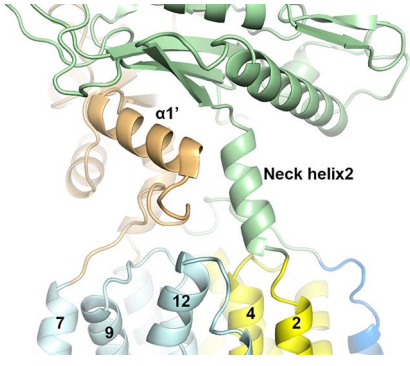
Structural comparison of the ligand-free Ptch1-3M and CHS-bound Ptch1-WT reveals the conformational changes of the SSD as well as the overall structure upon CHS binding (Fig. 6). Also considering the structural differences of the SSDs between the ligand-free NPC1 and CHS-bound Ptch1 (Fig. 4C), the CHS-induced conformational changes observed here may represent a paradigm that SSD-containing proteins use for sterol perception.

Potential cross-talk between the ESBS and SSD

Although the steroid binding to the SSD was anticipated, the extracellular steroid-binding site in Ptch1 and the steroid-dependent binding between Ptch1 and ShhN were not. An immediate question concerns the source of this trapped steroid. Because of its highly hydrophobic nature, a free cholesterol would be unlikely to diffuse into this site directly from the aqueous milieu, unless it is a cholesterol derivative with increased water solubility. There are then three possible sources for a hydrophobic steroid molecule: (i) It may be delivered by a NPC2-like lipid transfer



Movie 2. Conformational changes of Ptch1 upon CHS binding. The structures of Ptch1-3M and Ptch1-WT represent the starting and end frames. The procedure for morph generation was the same as for Movie 1.



Movie 3. The coupled motions of distinct domains in Ptch1 upon CHS binding. The morph is the same as in Movie 2. The zoomed-in views highlight the interactions that may underlie the concerted motions of TMD1/2 and EMD1/2 upon CHS binding.

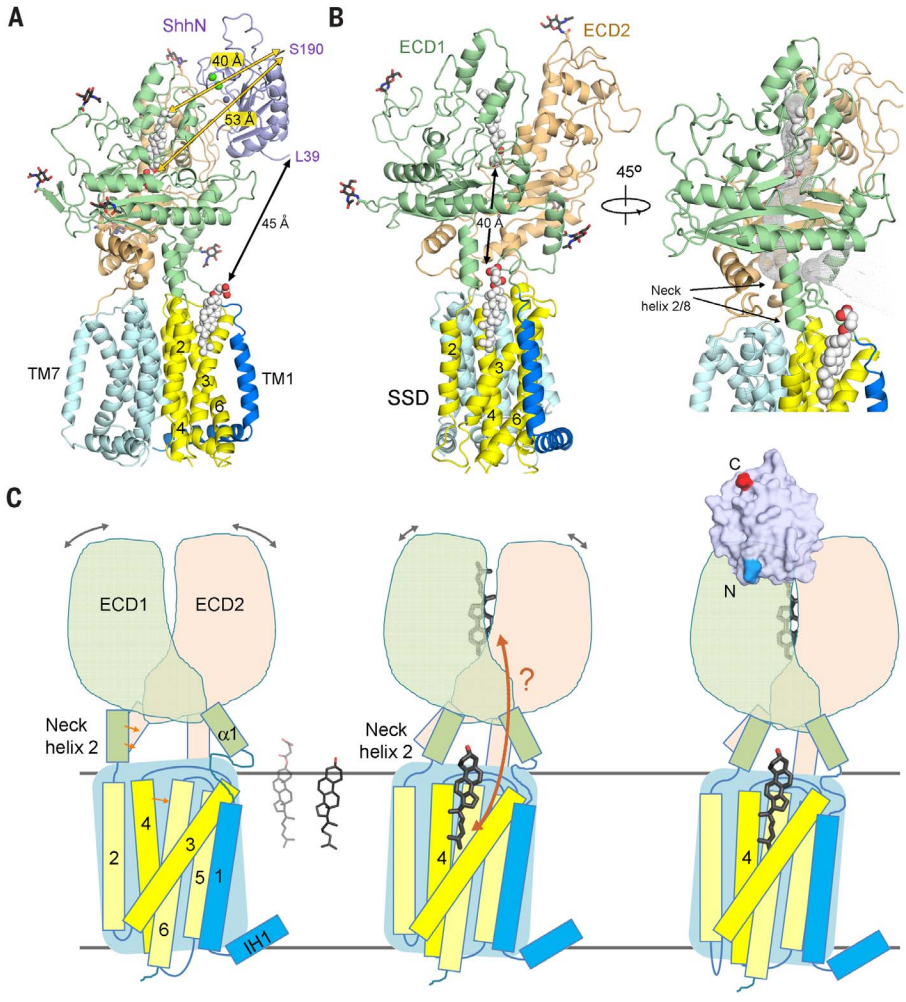


Fig. 7. Functional implications of the structures. (**A**) The relative positions of the two CHS molecules with respect to the N and C termini of ShhN in the complex. (**B**) The relative positions of the two steroid-binding sites on Ptch1. Right: ECD1 and ECD2 enclose a tunnel that can access to both ESBS and SSD. The tunnel is calculated by HOLE (*82*). (**C**) Summary of the structures reported in this study. Cartoons highlight the differences in the structures of Ptch1-3M (left), Ptch1-WT alone (center), and Ptch1-ShhN complex (right). The gray double-headed arrows indicate the structural flexibility of the ECDs. The same views of a CHS (gray) and a cholesterol (black) are shown to indicate their identical backbone. The physiologically relevant cholesterol is shown in the structural model. See text for discussion of functional implications of the structures.

protein, (ii) it may come from a steroidal moiety on a posttranslationally modified protein, or (iii) it may be extracted from the membrane and transferred to the ESBS by Ptch1 itself.

Although option 1 is formally possible, there has been no report on extracellular lipid-transfer proteins for Ptch1. With respect to option 2, the C-terminal Gly^{197} of ShhN is modified by a cholesteryl moiety, which can be a potential source. However, the last visible residue on the C terminus of ShhN is Ser^{190}, and a seven-residue peptide, even in a completely extended conformation, is not long enough to deliver the cholesteryl moiety to the ESBS in cis (Fig. 7A). In addition, in the structure of Ptch1 alone, the ESBS may allow a peptide to be connected to the steroid ligand; however, the sealed ESBS can hardly accommodate such a peptide in the Ptch1-ShhN complex (Fig. 7A). Therefore, we do not favor possibility 2.

Ptch1 has been speculated to be a transporter given its structural similarity with the RND pumps (*9*). A recent study discovered that Ptch1-mediated inhibition of Smo depends on extracellular Na^+ (*62*). Furthermore, the GxxxD(D) motif on TM4 and similar Gxxx(E/D) motif on TM10, which are required for the cotransport activity of the bacterial RND transporters, are conserved in Ptch1 (Fig. 6, E and F) (*9*, *63*). Consistent with steroid transport, ECD1 and ECD2 of Ptch1 encompass a tunnel that connects the ESBS and SSD in the absence of ShhN (Fig. 7B). It remains to be investigated whether a steroid molecule can be transferred between these two sites, and if so, whether this tunnel provides the diffusion path (Fig. 7C).

Structural implications for Hh signaling

Despite decades of rigorous characterization, the molecular mechanism for the inhibition of Smo by Ptch1 and the release of the inhibition by ShhN remains enigmatic. There are several outstanding questions. For instance, if Ptch1 functions as a sterol transporter, the transport activity is yet to be established. In addition, the abundance of cholesterol or its derivatives in primary cilia where Ptch1 inhibits Smo is unclear (*64*). Despite these and other outstanding questions, the three structures presented here reveal that Ptch1 undergoes steroid-induced conformational changes while binding of ShhN locks Ptch1 in one conformation (Fig. 7C). These findings lead us to speculate that the steroid-dependent structural shifts of Ptch1 may be required for its suppression of Smo, and that ShhN may inhibit Ptch1 by blocking this conformational change.

Many fundamental questions remain with regard to the function and mechanism of Ptch1 and ShhN. For instance, the unpalmitoylated ShhN used in this study is less active than the palmitoylated form in a cell-based assay by a factor of 30 to 40, although it retains high affinity with Ptch1 (*26*, *65*). What is more intriguing, a palmitoylated N-terminal 22-residue peptide was recently shown to activate Hh signaling in a cell-based assay (*66*), which cannot be explained by the three structures. Palmitoylate may interfere with the conformational changes or steroid-binding ability of Ptch1. Another question concerns the endogenous ligand or substrate for Ptch1. The composition and contour of the ESBS and the SSD cavity both allow accommodation of cholesterol with orientations similar to that of the CHS in the structure (Fig. 7A). Cholesterol derivatives with small-size modifications at C3-OH may also be compatible. The endogenous ligand(s) for Ptch1 may also act as agonists or antagonists for Smo. The structures reported here provide the basis for docking or simulation analysis that may facilitate the identification of such ligands. Finally, the structures of Ptch1 and ShhN presented here are both of monomers. The functional relevance of their oligomeric form remains unclear.

Despite these questions, the structures of human Ptch1 and its complex with ShhN set an important framework for future investigation of Hh signaling initiation and provide insight into the functional mechanism of SSD-containing proteins.

Materials and methods

Protein expression and purification

The complementary DNA of human Ptch1 (Uniprot: Q13635; residues 1 to 1305) was subcloned into the pCAG vector with an N-terminal FLAG tag and a C-terminal His_{10} tag. The Ptch1 mutants were generated with a standard 2-step PCR-based strategy. When the HEK293F cell density reached 2.0×10^6 cells per ml, the cells were transiently transfected with the expression plasmids and polyethylenimines (PEIs) (Polysciences). Approximately 1 mg of plasmids were premixed with 3 mg of PEIs in 50 ml of fresh medium for 15 to 30 min before application. For transfection, 50 ml of mixture was added to 1 liter of cell culture and incubated for 30 min. Transfected cells were cultured for 48 hours before harvest. For purification of WT Ptch1 and its variants, the HEK293F cells were collected and resuspended in the buffer containing 25 mM Tris pH 8.0, 150 mM NaCl, and protease inhibitor cocktails (Amresco). After sonication on ice, the membrane fraction was solubilized at 4°C for 2 hours with 1% (w/v) DDM (Anatrace) and 0.2% CHS (Anatrace). For biochemical assays reported in Fig. 5C, the protein was extracted using 1% (w/v) indicated detergents in the absence or presence of 0.2% CHS. After centrifugation at 20,000*g* for 1 hour, the supernatant was collected and applied to anti-Flag M2 affinity resin (Sigma). The resin was rinsed with wash buffer (W1 buffer) containing 25 mM Tris pH 8.0, 150 mM NaCl, 0.06% digitonin (w/v) (Sigma) or two critical micelle concentrations (CMC) of the other detergents with/without CHS as indicated in Fig. 5C, and protease inhibitors. The protein was eluted with W1 buffer plus FLAG peptide (200 μg/ml). The eluent was then applied to the nickel affinity resin (Ni-NTA, Qiagen). After three times of rinsing with W1 buffer plus 20 mM imidazole, the protein was eluted from the nickel resin with W1 buffer plus 250 mM imidazole. The eluent was then concentrated and further purified by SEC (Superose 6 10/300 GL, GE Healthcare) in buffer containing 25 mM Tris pH 8.0, 150 mM NaCl, and 0.06% digitonin or two CMC of the other detergents with/without CHS as indicated in Fig. 5C. Ptch1 proteins were eluted in two peaks, which were separately pooled and concentrated for EM and biochemical analysis. The later peak corresponding to monomer was pooled and concentrated to ~15 mg/ml for cryo sample preparation. The former peak corresponding to oligomer was used for the pull-down assay for different detergents as shown in Fig. 5C, whereas the latter peak was used for all the other pull-down assays.

The complementary DNA of human Shh (Uniprot: Q15465) N-terminal domain (ShhN, residues 24 to 197) was subcloned into the pET15 vector with N-terminal His_6 or MBP tags. The ShhN mutants were generated with a standard PCR-based strategy. The expression plasmids were transformed into *E. coli* BL21 (DE3) strain and protein overexpression was induced by 0.2 mM isopropyl β-D-thiogalactoside at an OD_{600nm} of 1.0. After growing at 20°C for 16 hours, the cells were collected and homogenized in buffer containing 25 mM Tris pH 8.0 and 150 mM NaCl. After sonication and centrifugation, the supernatant was applied to Ni-NTA affinity resin or amylose resin (New England Biolabs) and further purified by ion-exchange chromatography (Source 15S, GE Healthcare) and SEC (Superdex 200 10/300 GL, GE Healthcare). The peak fractions were pooled and concentrated for structural and biochemical analysis.

The Ptch1-ShhN complex was reconstituted by incubating the purified and concentrated Ptch1 and ShhN at a mass ratio of approximately 2:1 at 4°C for 1.5 hours before cryo-sample preparation.

Cryo-EM sample preparation and data collection

The cryo grids were prepared using Thermo Fisher Vitrobot Mark IV. The Quantifoil R1.2/1.3 Cu grids were glow-discharged with air for 30 s at medium level in Plasma Cleaner (Harrick Plasma, PDC-32G-2). Aliquots of 3.5 μl of purified Ptch1-WT, Ptch1-3M, or Ptch1-ShhN complex were applied to glow-discharged grids. After being blotted with filter paper for 4.0 s, the grids were plunged into liquid ethane cooled with liquid nitrogen. A total of 3983 micrograph stacks for Ptch1-WT, 6680 micrograph stacks for Ptch1-3M, and 4221 micrograph stacks for Ptch1-ShhN complex were automatically collected with AutoEMation II (*67*, *68*) on Titan Krios at 300 kV equipped with K2 Summit direct electron detector (Gatan), Quantum energy filter (Gatan), Cs corrector (Thermo Fisher), and Volta phase plate (Thermo Fisher) at a nominal magnification of 105,000× with a fixed defocus value of −0.7 μm. Each stack was exposed in super-resolution mode for 5.6 s with an exposing time of 0.175 s per frame, resulting in 32 frames per stack. The total dose rate was about 50 $e^-/Å^2$ for each stack. The stacks were first motion-corrected with MotionCorr (*69*) and binned 2-fold, resulting in a pixel size of 1.091 Å/pixel.

The output stacks from MotionCorr were further motion-corrected with MotionCor2 (*70*); meanwhile, dose weighting was performed (*71*). The defocus values were estimated with Gctf (*72*).

Cryo-EM data processing

For the Ptch1-WT dataset, a total of 1,710,446 particles were automatically picked with RELION 2.0 (*73*). After 2D classification, a 3D initial model was built from typical 2D class averages with e2initialmodel.py (*74*). A total of 736,934 particles were selected from 2D classification and subjected to a global angular search 3D classification with one class and 40 iterations. The outputs of the 33rd to 40th iterations were subjected to local angular search 3D classification with four classes separately. A total of 516,223 particles were selected by combining the good classes of the local angular search 3D classification, yielding a 3D reconstruction with an overall resolution of 6.1 Å after 3D auto-refinement. The handedness of the 3D reconstruction was checked and corrected. Then 156,723 particles were selected after 3D classification against one good reference and four bad references, giving rise to improved resolution at 4.4 Å after 3D auto-refinement. Finally, a subset of 94,445 particles were further selected by "skip alignment" 3D classification with an adapted mask, yielding overall resolution of 3.9 Å after 3D auto-refinement. The procedure for data processing of the Ptch1-3M was the same as that for Ptch1-WT. Eventually a subset of 154,721 particles was selected to generate a 3D reconstruction with overall resolution of 4.1 Å.

The procedure for data processing of the Ptch1-ShhN complex was similar with that for Ptch1 alone. To be brief, 1,518,460 particles were automatically picked. 2D classification resulted in 716,512 good particles that were subjected to subsequent global angular search 3D classification and local angular search 3D classification. After combining the good classes, 435,942 particles yielded a 3D reconstruction with overall resolution of 5.6 Å. After 3D classification against one good reference and four bad references, 137,823 particles were selected and gave rise to a 3D reconstruction of 4.4 Å resolution, which was further improved to 3.6 Å by applying an adapted mask during 3D auto-refinement.

All 2D classification, 3D classification, and 3D auto-refinement were performed with RELION 2.0. Resolutions were estimated with the gold-standard Fourier shell correlation 0.143 criterion (*75*) with high-resolution noise substitution (*76*).

Model building and refinement

The Ptch1-ShhN complex map at 3.6 Å was used for de novo model building of Ptch1. A poly-Ala model was derived using the NPC1 structure as the initial model (PDB code 3JD8) in COOT (*77*). Sequence assignment was guided mainly by bulky residues such as Phe, Tyr, Trp, and Arg. The chemical properties of amino acids were taken into consideration to facilitate model building. The glycosylation sites predicted in UniProt were used for model validation. One or two N-acetylglucosamine moieties were built to each site based on the densities. The crystal structure of ShhN (PDB code 4C4M) was docked into the map for Ptch1-ShhN complex and adjusted manually in COOT. For the Ptch1-WT and Ptch1-3M, the structural model of Ptch1 from the complex was fitted into the corresponding maps followed by manual adjustment.

All structure refinements were carried out by PHENIX (*78*) in real space with secondary structure and geometry restraints. Overfitting of the models was monitored by refining the model against one of the two independent half-maps and testing the refined model against the other map (*79*).

MBP-mediated pull-down assay

For pull-down assays, the Ptch1 and MBP-ShhN, WT or indicated variants, were incubated with amylose resin resuspended in 200 μl of buffer containing 25 mM Tris pH 8.0, 150 mM NaCl, and 0.06% digitonin or the indicated detergents at concentration of their respective 2× CMC at 4°C for ~1 hour. Then the mixture was spun down at 500*g* for 3 min. After removing the supernatant, the resin was extensively rinsed with the assay buffer to remove unbound proteins. Finally the resin was resuspended in 200 μl of buffer, out of which 20 μl was applied to SDS-PAGE analysis followed by Coomassie blue staining.

In the assay reported in Fig. 3C, all the indicated Ptch1 and MBP-ShhN variants were applied at concentrations of ~0.15 mg/ml. In the titration pull-down assay reported in Fig. 5B, ShhN was used at ~0.15 mg/ml, while Ptch1 was serially diluted fourfold at concentrations of ~0.60, ~0.15, ~0.04, and ~0.01 mg/ml, respectively. In the assay to test the effects of different detergents on the complex formation, the molar ratio of the two proteins was approximately 1:1 with Ptch1 and MBP-ShhN at concentrations of ~0.15 and 0.08 mg/ml, respectively.

REFERENCES AND NOTES

1. J. Briscoe, P. P. Thérond, The mechanisms of Hedgehog signalling and its roles in development and disease. *Nat. Rev. Mol. Cell Biol.* **14**, 416–429 (2013). doi: 10.1038/nrm3598; pmid: 23719536
2. E. Pak, R. A. Segal, Hedgehog Signal Transduction: Key Players, Oncogenic Drivers, and Cancer Therapy. *Dev. Cell* **38**, 333–344 (2016). doi: 10.1016/j.devcel.2016.07.026; pmid: 27554855
3. P. W. Ingham, Y. Nakano, C. Seger, Mechanisms and functions of Hedgehog signalling across the metazoa. *Nat. Rev. Genet.* **12**, 393–406 (2011). doi: 10.1038/nrg2984; pmid: 21502959
4. L. Lum, P. A. Beachy, The Hedgehog response network: Sensors, switches, and routers. *Science* **304**, 1755–1759 (2004). doi: 10.1126/science.1098020; pmid: 15205520
5. P. W. Ingham, A. P. McMahon, Hedgehog signaling in animal development: Paradigms and principles. *Genes Dev.* **15**, 3059–3087 (2001). doi: 10.1101/gad.938601; pmid: 11731473
6. V. Marigo, R. A. Davey, Y. Zuo, J. M. Cunningham, C. J. Tabin, Biochemical evidence that patched is the Hedgehog receptor. *Nature* **384**, 176–179 (1996). doi: 10.1038/384176a0; pmid: 8906794
7. D. M. Stone *et al*., The tumour-suppressor gene patched encodes a candidate receptor for Sonic hedgehog. *Nature* **384**, 129–134 (1996). doi: 10.1038/384129a0; pmid: 8906787
8. N. Fuse *et al*., Sonic hedgehog protein signals not as a hydrolytic enzyme but as an apparent ligand for patched. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10992–10999 (1999). doi: 10.1073/pnas.96.20.10992; pmid: 10500113
9. J. Taipale, M. K. Cooper, T. Maiti, P. A. Beachy, Patched acts catalytically to suppress the activity of Smoothened. *Nature* **418**, 892–897 (2002). doi: 10.1038/nature00989; pmid: 12192414
10. R. Rohatgi, L. Milenkovic, M. P. Scott, Patched1 regulates hedgehog signaling at the primary cilium. *Science* **317**, 372–376 (2007). doi: 10.1126/science.1139740; pmid: 17641202
11. K. C. Corbit *et al*., Vertebrate Smoothened functions at the primary cilium. *Nature* **437**, 1018–1021 (2005). doi: 10.1038/nature04117; pmid: 16136078
12. P. Aza-Blanc, F. A. Ramírez-Weber, M. P. Laget, C. Schwartz, T. B. Kornberg, Proteolysis that is inhibited by hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* **89**, 1043–1053 (1997). doi: 10.1016/S0092-8674(00)80292-5; pmid: 9215627
13. C. H. Chen *et al*., Nuclear trafficking of Cubitus interruptus in the transcriptional regulation of Hedgehog target gene expression. *Cell* **98**, 305–316 (1999). doi: 10.1016/S0092-8674(00)81960-1; pmid: 10458606
14. N. Méthot, K. Basler, Hedgehog controls limb development by regulating the activities of distinct transcriptional activator and repressor forms of Cubitus interruptus. *Cell* **96**, 819–831 (1999). doi: 10.1016/S0092-8674(00)80592-9; pmid: 10102270
15. J. Kim, M. Kato, P. A. Beachy, Gli2 trafficking links Hedgehog-dependent activation of Smoothened in the primary cilium to transcriptional activation in the nucleus. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21666–21671 (2009). doi: 10.1073/pnas.0912180106; pmid: 19996169
16. J. W. Yoon *et al*., Gene expression profiling leads to identification of GLI1-binding elements in target genes and a role for multiple downstream pathways in GLI1-induced cell transformation. *J. Biol. Chem.* **277**, 5548–5555 (2002). doi: 10.1074/jbc.M105708200; pmid: 11719506
17. X. Wen *et al*., Kinetics of hedgehog-dependent full-length Gli3 accumulation in primary cilia and subsequent degradation. *Mol. Cell. Biol.* **30**, 1910–1922 (2010). doi: 10.1128/MCB.01089-09; pmid: 20154143
18. C. C. Hui, S. Angers, Gli proteins in development and disease. *Annu. Rev. Cell Dev. Biol.* **27**, 513–537 (2011). doi: 10.1146/annurev-cellbio-092910-154048; pmid: 21801010
19. J. Taipale, P. A. Beachy, The Hedgehog and Wnt signalling pathways in cancer. *Nature* **411**, 349–354 (2001). doi: 10.1038/35077219; pmid: 11357142
20. P. A. Beachy, S. S. Karhadkar, D. M. Berman, Tissue repair and stem cell renewal in carcinogenesis. *Nature* **432**, 324–331 (2004). doi: 10.1038/nature03100; pmid: 15549094
21. L. L. Rubin, F. J. de Sauvage, Targeting the Hedgehog pathway in cancer. *Nat. Rev. Drug Discov.* **5**, 1026–1033 (2006). doi: 10.1038/nrd2086; pmid: 17139287
22. J. Xie *et al*., Activating Smoothened mutations in sporadic basal-cell carcinoma. *Nature* **391**, 90–92 (1998). doi: 10.1038/34201; pmid: 9422511
23. R. K. Mann, P. A. Beachy, Novel lipid modifications of secreted protein signals. *Annu. Rev. Biochem.* **73**, 891–923 (2004). doi: 10.1146/annurev.biochem.73.011303.073933; pmid: 15189162
24. J. A. Porter, K. E. Young, P. A. Beachy, Cholesterol modification of hedgehog signaling proteins in animal development. *Science* **274**, 255–259 (1996). doi: 10.1126/science.274.5285.255; pmid: 8824192
25. J. A. Porter *et al*., Hedgehog patterning activity: Role of a lipophilic modification mediated by the carboxy-terminal autoprocessing domain. *Cell* **86**, 21–34 (1996). doi: 10.1016/S0092-8674(00)80074-4; pmid: 8689684
26. R. B. Pepinsky *et al*., Identification of a palmitic acid-modified form of human Sonic hedgehog. *J. Biol. Chem.* **273**, 14037–14045 (1998). doi: 10.1074/jbc.273.22.14037; pmid: 9593755
27. P. M. Lewis *et al*., Cholesterol modification of sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by Ptc1. *Cell* **105**, 599–612 (2001). doi: 10.1016/S0092-8674(01)00369-5; pmid: 11389830
28. Z. Chamoun *et al*., Skinny hedgehog, an acyltransferase required for palmitoylation and activity of the hedgehog signal. *Science* **293**, 2080–2084 (2001). doi: 10.1126/science.1064437; pmid: 11486055
29. K. P. Williams *et al*., Functional antagonists of sonic hedgehog reveal the importance of the N terminus for activity. *J. Cell Sci.* **112**, 4405–4414 (1999). pmid: 10564658
30. B. Z. Stanton, L. F. Peng, Small-molecule modulators of the Sonic Hedgehog signaling pathway. *Mol. Biosyst.* **6**, 44–54 (2010). doi: 10.1039/B910196A; pmid: 20024066

31. S. Peukert, K. Miller-Moslin, Small-molecule inhibitors of the hedgehog signaling pathway as cancer therapeutics. *ChemMedChem* **5**, 500–512 (2010). doi: 10.1002/cmdc.201000011; pmid: 20229564
32. B. Z. Stanton *et al*., A small molecule that binds Hedgehog and blocks its signaling in human cells. *Nat. Chem. Biol.* **5**, 154–156 (2009). doi: 10.1038/nchembio.142; pmid: 19151731
33. A. E. Owens *et al*., Design and Evolution of a Macrocyclic Peptide Inhibitor of the Sonic Hedgehog/Patched Interaction. *J. Am. Chem. Soc.* **139**, 12559–12568 (2017). doi: 10.1021/jacs.7b06087; pmid: 28759213
34. P. A. Beachy, S. G. Hymowitz, R. A. Lazarus, D. J. Leahy, C. Siebold, Interactions between Hedgehog proteins and their binding partners come into view. *Genes Dev.* **24**, 2001–2012 (2010). doi: 10.1101/gad.1951710; pmid: 20844013
35. J. E. Hooper, M. P. Scott, The Drosophila patched gene encodes a putative membrane protein required for segmental patterning. *Cell* **59**, 751–765 (1989). doi: 10.1016/0092-8674(89)90021-4; pmid: 2582494
36. Y. Nakano *et al*., A protein with several possible membrane-spanning domains encoded by the Drosophila segment polarity gene patched. *Nature* **341**, 508–513 (1989). doi: 10.1038/341508a0; pmid: 2797178
37. R. L. Johnson *et al*., Human homolog of patched, a candidate gene for the basal cell nevus syndrome. *Science* **272**, 1668–1671 (1996). doi: 10.1126/science.272.5268.1668; pmid: 8658145
38. H. Hahn *et al*., Mutations of the human homolog of Drosophila patched in the nevoid basal cell carcinoma syndrome. *Cell* **85**, 841–851 (1996). doi: 10.1016/S0092-8674(00)81268-4; pmid: 8681379
39. D. Carpenter *et al*., Characterization of two patched receptors for the vertebrate hedgehog protein family. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13630–13634 (1998). doi: 10.1073/pnas.95.23.13630; pmid: 9811851
40. P. G. Zaphiropoulos, A. B. Undén, F. Rahnama, R. E. Hollingsworth, R. Toftgård, PTCH2, a novel human patched gene, undergoing alternative splicing and up-regulated in basal cell carcinomas. *Cancer Res.* **59**, 787–792 (1999). pmid: 10029063
41. T. T. Tseng *et al*., The RND permease superfamily: An ancient, ubiquitous and diverse family that includes human disease and development proteins. *J. Mol. Microbiol. Biotechnol.* **1**, 107–125 (1999). pmid: 10941792
42. A. Yamaguchi, R. Nakashima, K. Sakurai, Structural basis of RND-type multidrug exporters. *Front. Microbiol.* **6**, 327 (2015). doi: 10.3389/fmicb.2015.00327; pmid: 25941524
43. P. E. Kuwabara, M. Labouesse, The sterol-sensing domain: Multiple families, a unique role? *Trends Genet.* **18**, 193–201 (2002). doi: 10.1016/S0168-9525(02)02640-9; pmid: 11932020
44. X. Gong *et al*., Structural Insights into the Niemann-Pick C1 (NPC1)-Mediated Cholesterol Transfer and Ebola Infection. *Cell* **165**, 1467–1478 (2016). doi: 10.1016/j.cell.2016.05.022; pmid: 27238017
45. X. Li *et al*., Structure of human Niemann-Pick C1 protein. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 8212–8217 (2016). doi: 10.1073/pnas.1607795113; pmid: 27307437
46. M. C. Harvey, A. Fleet, N. Okolowsky, P. A. Hamel, Distinct effects of the mesenchymal dysplasia gene variant of murine Patched-1 protein on canonical and non-canonical Hedgehog signaling pathways. *J. Biol. Chem.* **289**, 10939–10949 (2014). doi: 10.1074/jbc.M113.514844; pmid: 24570001
47. A. Fleet, J. P. Lee, A. Tamachi, I. Javeed, P. A. Hamel, Activities of the Cytoplasmic Domains of Patched-1 Modulate but Are Not Essential for the Regulation of Canonical Hedgehog Signaling. *J. Biol. Chem.* **291**, 17557–17568 (2016). doi: 10.1074/jbc.M116.731745; pmid: 27325696
48. X. Lu, S. Liu, T. B. Kornberg, The C-terminal tail of the Hedgehog receptor Patched regulates both localization and turnover. *Genes Dev.* **20**, 2539–2551 (2006). doi: 10.1101/gad.1461306; pmid: 16980583
49. S. Murakami, R. Nakashima, E. Yamashita, A. Yamaguchi, Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* **419**, 587–593 (2002). doi: 10.1038/nature01050; pmid: 12374972
50. J. S. McLellan *et al*., The mode of Hedgehog binding to Ihog homologues is not conserved across different phyla. *Nature* **455**, 979–983 (2008). doi: 10.1038/nature07358; pmid: 18794898
51. B. Bishop *et al*., Structural insights into hedgehog ligand sequestration by the human hedgehog-interacting protein HHIP. *Nat. Struct. Mol. Biol.* **16**, 698–703 (2009). doi: 10.1038/nsmb.1607; pmid: 19561611
52. I. Bosanac *et al*., The structure of SHH in complex with HHIP reveals a recognition role for the Shh pseudo active site in signaling. *Nat. Struct. Mol. Biol.* **16**, 691–697 (2009). doi: 10.1038/nsmb.1632; pmid: 19561609
53. H. R. Maun *et al*., Hedgehog pathway antagonist 5E1 binds hedgehog at the pseudo-active site. *J. Biol. Chem.* **285**, 26570–26580 (2010). doi: 10.1074/jbc.M110.112284; pmid: 20504762
54. F. Lu *et al*., Identification of NPC1 as the target of U18666A, an inhibitor of lysosomal cholesterol export and Ebola infection. *eLife* **4**, e12177 (2015). doi: 10.7554/eLife.12177; pmid: 26646182
55. N. Boutet *et al*., Spectrum of PTCH1 mutations in French patients with Gorlin syndrome. *J. Invest. Dermatol.* **121**, 478–481 (2003). doi: 10.1046/j.1523-1747.2003.12423.x; pmid: 12925203
56. K. Fujii *et al*., Mutations in the human homologue of Drosophila patched in Japanese nevoid basal cell carcinoma syndrome patients. *Hum. Mutat.* **21**, 451–452 (2003). doi: 10.1002/humu.9132; pmid: 12655573
57. G. R. Hime *et al*., Functional analysis in Drosophila indicates that the NBCCS/PTCH1 mutation G509V results in activation of smoothened through a dominant-negative mechanism. *Dev. Dyn.* **229**, 780–790 (2004). doi: 10.1002/dvdy.10499; pmid: 15042702
58. G. Tate, M. Li, T. Suzuki, T. Mitsuya, A new germline mutation of the PTCH gene in a Japanese patient with nevoid basal cell carcinoma syndrome associated with meningioma. *Jpn. J. Clin. Oncol.* **33**, 47–50 (2003). doi: 10.1093/jjco/hyg005; pmid: 12604725
59. J. L. Goldstein, M. S. Brown, A century of cholesterol and coronaries: From plaques to genes to statins. *Cell* **161**, 161–172 (2015). doi: 10.1016/j.cell.2015.01.036; pmid: 25815993
60. K. L. Luskey, B. Stevens, Human 3-hydroxy-3-methylglutaryl coenzyme A reductase. Conserved domains responsible for catalytic activity and sterol-regulated degradation. *J. Biol. Chem.* **260**, 10271–10277 (1985). pmid: 2991281
61. M. T. Vanier, Complex lipid trafficking in Niemann-Pick disease type C. *J. Inherit. Metab. Dis.* **38**, 187–199 (2015). doi: 10.1007/s10545-014-9794-4; pmid: 25425283
62. B. R. Myers, L. Neahring, Y. Zhang, K. J. Roberts, P. A. Beachy, Rapid, direct activity assays for Smoothened reveal Hedgehog pathway regulation by membrane cholesterol and extracellular sodium. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E11141–E11150 (2017). doi: 10.1073/pnas.1717891115; pmid: 29229834
63. L. Guan, T. Nakae, Identification of essential charged residues in transmembrane segments of the multidrug transporter MexB of Pseudomonas aeruginosa. *J. Bacteriol.* **183**, 1734–1739 (2001). doi: 10.1128/JB.183.5.1734-1739.2001; pmid: 11160105
64. S. C. Goetz, K. V. Anderson, The primary cilium: A signalling centre during vertebrate development. *Nat. Rev. Genet.* **11**, 331–344 (2010). doi: 10.1038/nrg2774; pmid: 20395968
65. F. R. Taylor *et al*., Enhanced potency of human Sonic hedgehog by hydrophobic modification. *Biochemistry* **40**, 4359–4371 (2001). doi: 10.1021/bi002487u; pmid: 11284692
66. H. Tukachinsky, K. Petrov, M. Watanabe, A. Salic, Mechanism of inhibition of the tumor suppressor Patched by Sonic Hedgehog. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E5866–E5875 (2016). doi: 10.1073/pnas.1606719113; pmid: 27647915
67. J. Lei, J. Frank, Automated acquisition of cryo-electron micrographs for single particle reconstruction on an FEI Tecnai electron microscope. *J. Struct. Biol.* **150**, 69–80 (2005). doi: 10.1016/j.jsb.2005.01.002; pmid: 15797731
68. X. Fan *et al*., Near-Atomic Resolution Structure Determination in Over-Focus with Volta Phase Plate by Cs-Corrected Cryo-EM. *Structure* **25**, 1623–1630.e3 (2017). doi: 10.1016/j.jsb.2005.01.002; pmid: 15797731
69. X. Li *et al*., Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590 (2013). doi: 10.1038/nmeth.2472; pmid: 23644547
70. S. Q. Zheng *et al*., MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017). doi: 10.1038/nmeth.4193; pmid: 28250466
71. T. Grant, N. Grigorieff, Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015). doi: 10.7554/eLife.06980; pmid: 26023829
72. K. Zhang, Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016). doi: 10.1016/j.jsb.2015.11.003; pmid: 26592709
73. D. Kimanius, B. O. Forsberg, S. H. Scheres, E. Lindahl, Accelerated cryo-EM structure determination with parallelisation using GPUs in RELION-2. *eLife* **5**, e18722 (2016). doi: 10.7554/eLife.18722; pmid: 27845625
74. G. Tang *et al*., EMAN2: An extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007). doi: 10.1016/j.jsb.2006.05.009; pmid: 16859925
75. P. B. Rosenthal, R. Henderson, Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* **333**, 721–745 (2003). doi: 10.1016/j.jmb.2003.07.013; pmid: 14568533
76. S. Chen *et al*., High-resolution noise substitution to measure overfitting and validate resolution in 3D structure determination by single particle electron cryomicroscopy. *Ultramicroscopy* **135**, 24–35 (2013). doi: 10.1016/j.ultramic.2013.06.004; pmid: 23872039
77. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010). doi: 10.1107/S0907444910007493; pmid: 20383002
78. P. D. Adams *et al*., PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010). doi: 10.1107/S0907444909052925; pmid: 20124702
79. A. Amunts *et al*., Structure of the yeast mitochondrial large ribosomal subunit. *Science* **343**, 1485–1489 (2014). doi: 10.1126/science.1249410; pmid: 24675956
80. E. F. Pettersen *et al*., UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004). doi: 10.1002/jcc.20084; pmid: 15264254
81. W. L. DeLano, The PyMOL Molecular Graphics System; www.pymol.org. (2002).
82. O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, M. S. Sansom, HOLE: A program for the analysis of the pore dimensions of ion channel structural models. *J. Mol. Graph.* **14**, 354–360 (1996). doi: 10.1016/S0263-7855(97)00009-X; pmid: 9195488

ACKNOWLEDGMENTS

We thank X. Li for technical support during EM image acquisition. **Funding:** Supported by National Key Basic Research (973) Program grant 2015CB910101 (N.Y.) and National Key R&D Program grants 2016YFA0500402 (N.Y.) and 2016YFA0501100 (J.L.) from Ministry of Science and Technology of China; National Natural Science Foundation of China projects 31621092, 31630017, and 31611130036 (N.Y.); and the Shirley M. Tilghman endowed professorship from Princeton University (N.Y.). We thank the Tsinghua University Branch of China National Center for Protein Sciences (Beijing) for providing cryo-EM facility support. We thank the computational facility support on the cluster of Bio-Computing Platform (Tsinghua University Branch of China National Center for Protein Sciences Beijing) and the "Explorer 100" cluster system of Tsinghua National Laboratory for Information Science and Technology. **Author contributions:** N.Y. and X.G. conceived the project; X.G., H.Q., P.C., X.Z., Q.Z., and J.L. performed the experiments; all authors contributed to data analysis; and N.Y. and X.G. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Atomic coordinates and EM density maps of the human Ptch1-WT (PDB: 6DMB; EMDB: EMD-7963), Ptch1-3M (PDB: 6DMO; EMDB: EMD-7964), and ShhN-bound Ptch1 (PDB: 6DMY; EMDB: EMD-7968) have been deposited in the Protein Data Bank (www.rcsb.org) and the Electron Microscopy Data Bank (www.ebi.ac.uk/pdbe/emdb/).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/eaas8935/suppl/DC1
Figs. S1 to S10
References (*83*, *84*)

3 January 2018; accepted 19 June 2018
Published online 28 June 2018
10.1126/science.aas8935

RESEARCH ARTICLE SUMMARY

ORGANIC CHEMISTRY

Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS

Shishi Lin, Sergei Dikler, William D. Blincoe, Ronald D. Ferguson, Robert P. Sheridan, Zhengwei Peng, Donald V. Conway, Kerstin Zawatzky, Heather Wang, Tim Cernak, Ian W. Davies, Daniel A. DiRocco, Huaming Sheng*, Christopher J. Welch*, Spencer D. Dreher*

INTRODUCTION: The invention of new chemical reactions provides new bond construction strategies for improved access to diverse regions of structural space. However, a pervasive, long-standing bias toward reporting successful results means that the shortcomings of even mature reaction methods remain poorly defined, making practical syntheses of structurally diverse targets far from certain. Distinct tools and experimental approaches are required to expose and record the problematic structural elements that limit different synthetic methods. The experimental space required to systematically survey reaction failure is vast, and existing ultrahigh-throughput (uHT) reaction screening approaches are inadequate for exploring the diversity of conditions pertaining in modern synthetic methods. Additionally, analytical approaches must continuously improve to meet the throughput demands of this expansive reaction screening.

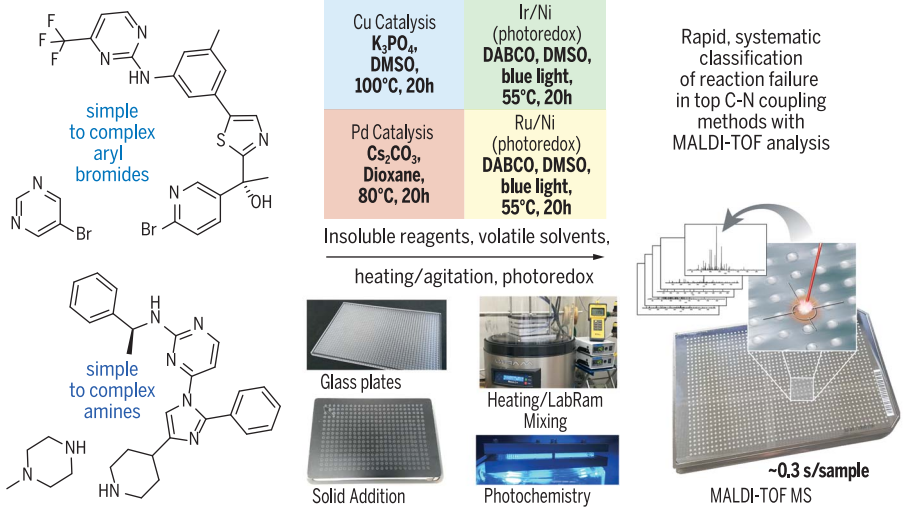
RATIONALE: We report a nanomole-scale screening protocol that can be used to execute heterogeneous reactions with heating and agitation, use of volatile solvents, and capacity for photoredox chemistry. These advances in miniaturized chemistry screening were combined with the use of matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI-TOF MS), enabling analysis of 1536 reactions in ~10 min. Together, these advances create a platform that can enable systematic reaction evaluation and data capture to survey the dark space of chemical reactions.

RESULTS: Using the Buchwald-Hartwig C–N coupling reaction to exemplify this process, an uHT Glorius fragment additive poisons diagnostic approach was first applied to demonstrate that MALDI-MS could provide adequate data quality to monitor the formation of a single product under a wide variety of different synthetic conditions. Four catalytic methods—Ir/Ni and Ru/Ni dual-metal photoredox catalysis, as well as heterogeneous and high-temperature Cu and Pd catalysis—with extended nanomole chemistry requirements were evaluated for the synthesis of a single product in the presence of 383 structurally diverse simple and complex potential poisons. Using a normalizing internal standard that was closely related to the product and optimized operating parameters, MALDI-MS provided good correlation with existing ultra performance liquid chromatography (UPLC)–MS approaches (coefficient of determination R^2 up to 0.85), allowing correct binning of "hits" and "misses" (defined as >50% product signal knockdown) up to 95% of the time. Next, the more challenging goal of exploring diverse whole-molecule C–N couplings was explored. In this case, it was not practical to have either product standards or closely related internal standards to enable analytical quantitation. A "simplest-partner test" was employed, in which 192 aryl bromides and 192 secondary amines were each coupled with a MS-active "simplest partner," guaranteeing a somewhat normalized MS response for all products. The formation of 384 different products using the four aforementioned synthetic methods was monitored by MALDI-MS, with pass-fail binning of results correlating well with UPLC-MS in the identification of common structural elements (such as functional group counts, H-bond donors and acceptors, and polar surface area) that lead to reaction failure.

CONCLUSION: In the near future, each problematic structural element that is identified through systematic dark-space exploration can be promoted for in-depth examination to precisely define the specific parameters that determine reaction outcome at the atomic and quantum molecular level. Predictive machine learning models will use this focused data to enable synthetic practitioners to select the most appropriate reactions for use in a particular synthetic setting. In addition, functionality that persistently fails across synthetic methods can sharply define important challenges for the invention of improved chemical reactions. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aar6236



Extended nanomole chemistry and MALDI-TOF MS for systematic reaction profiling. Nanomole-scale chemistry tools that can execute a wide variety of synthetic protocols are combined with rapid MALDI-TOF MS analysis to enable broad reaction profiling to map the dark space of chemical reactivity. DMSO, dimethyl sulfoxide; DABCO, 1,4-diazabicyclo[2.2.2]octane.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: huaming.sheng@merck.com (H.S.); christopher@welchinnovation.com (C.J.W.); spencer_dreher@merck.com (S.D.D.)
Cite this article as: S. Lin *et al.*, *Science* **361**, eaar6236 (2018). DOI: 10.1126/science.aar6236

RESEARCH ARTICLE

ORGANIC CHEMISTRY

Mapping the dark space of chemical reactions with extended nanomole synthesis and MALDI-TOF MS

Shishi Lin[1], Sergei Dikler[2], William D. Blincoe[3], Ronald D. Ferguson[1], Robert P. Sheridan[4], Zhengwei Peng[4], Donald V. Conway[5], Kerstin Zawatzky[3], Heather Wang[3], Tim Cernak[6]*, Ian W. Davies[7]†, Daniel A. DiRocco[7], Huaming Sheng[3]‡, Christopher J. Welch[3]‡§||, Spencer D. Dreher[1]‡

Understanding the practical limitations of chemical reactions is critically important for efficiently planning the synthesis of compounds in pharmaceutical, agrochemical, and specialty chemical research and development. However, literature reports of the scope of new reactions are often cursory and biased toward successful results, severely limiting the ability to predict reaction outcomes for untested substrates. We herein illustrate strategies for carrying out large-scale surveys of chemical reactivity by using a material-sparing nanomole-scale automated synthesis platform with greatly expanded synthetic scope combined with ultrahigh-throughput matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI-TOF MS).

The field of synthetic organic chemistry has spawned a vast array of creative reactions that can be logically combined to prepare nearly any molecule. However, efficient selection of the precise reaction sequence that leads to a particular product remains a challenge, as poorly understood substrate-specific interactions often necessitate laborious screening of combinations of catalysts, reagents, and conditions. High-throughput experimentation (HTE) chemistry facilitates these investigations by increasing the pace of problem-solving (*1*–*7*), and an emerging strategy uses machine learning (ML) data generated with these screening tools to create predictive models for specific problems (*8*, *9*). Data mining surveys of existing published or proprietary databases containing information on hundreds of millions of reactions have been somewhat encouraging (*10*–*14*), but a pervasive, long-standing bias toward reporting successful results limits the utility of this information for model building. Additionally, these data have not been collected under controlled experimental conditions, and most of the substrates are not representative of the complexity that arises in applied synthetic problems. Hence, no existing large, structured datasets meet all of the requirements for effective reaction modeling. Consequently, intentional surveys are required to systematically map reactivity patterns, with the rapid identification of inaccessible regions, the dark space of chemical reactivity, being particularly important for focusing subsequent ML investigations.

Chemical reactivity space is vast, even for a single synthetic transformation, with permutations of possible substrate structures, catalysts, and reaction conditions soon proliferating into unmanageable experimental complexity. Large-scale reactivity surveys have thus far been limited by the throughput of reaction screening and analysis technologies. Additionally, the scarcity and high cost of relevant complex substrates requires that screening be carried out on as small a scale as possible. Existing miniaturized HTE chemistry tools have worked only under a narrow range of conditions and have thus far not been applicable to the vast majority of high-quality synthetic protocols. In this work, we describe how engineering advances in automated, miniaturized reaction experimentation and the use of fast matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) for routine reaction analysis affords an opportunity to rapidly survey the chemical reactivity landscape. We demonstrate how these advances can be used to systematically reveal diverse sources of reaction failure within large substrate sets, information that is largely missing from current chemistry data sources, representing a notable advance toward broadly effective reaction profiling.

Extending nanomole synthesis for systematic reaction profiling

Our interest in reaction profiling has grown from an internal effort termed the C–N Coupling Initiative, which aims to systematically build tools to completely understand and leverage a given synthetic transformation. A recent examination of our electronic notebooks (ELNs) shows C–N coupling to be a hit-or-miss affair (~10,000 examples, ~35% failure rate) that, despite its potential to render druglike molecules, may be underused because of its higher demand for optimization than, for example, the Suzuki reaction (~50,000 example, 18% failure). Attempts to use this ELN dataset for creating predictive models have been disappointing, as the data are fractured, with many different classes of amines that would be expected to have different reactivity patterns (primary, secondary, NH-heterocyclic, amides, sulfonamides) while also containing deep channels of project-driven substrate similarity in the aryl halide coupling partners. Additionally, these data are collected under a wide array of synthetic protocols, with specific conditions seldom being repeated. Given this lack of a suitable dataset, a real and urgent goal of applied research is to increase the understanding of C–N coupling reactions by building an experimental process suitable for fast reaction mapping. Clearly, this systematic reaction profiling capability will also be of general value for surveying other reaction types where available knowledge is scarce.

Failure in metal-catalyzed reactions can result from a variety of causes, including inhibition of catalysts due to substrate binding, competitive side reactions, local steric and electronic interactions, and substrate decomposition caused by harsh reaction conditions. In addition, small perturbations in molecular structure may result in changes in reactivity, especially for complex substrates. We perceived that mapping reaction failure arising from this diverse patchwork of mechanisms would require a large collection of structurally diverse analogs as well as a tremendous number of experiments. Emerging nanomole-scale tools have the material-sparing miniaturized scale and ultrahigh-throughput (uHT) required for broad profiling; however, existing experimental strategies involve simplified engineering and are inadequate for examining most high-quality synthetic protocols. The first-generation nanomole synthesis platform that we developed (*1*) was limited to plastic-compatible, homogenous, ambient temperature reactions in low-volatility polar, aprotic solvents, encompassing only a small subset of organic transformations. The recent work by Perera *et al.* (*2*) extends nanomole synthesis capabilities in a flow chemistry format, allowing heating and the use of diverse solvents; however, this format also requires that reactions remain homogeneous, necessitating the use of very high dilution and

[1]Chemistry Capabilities Accelerating Therapeutics, Merck & Co., Inc., Kenilworth, NJ 07033, USA. [2]Bruker Daltonics, Inc., Billerica, MA 01821, USA. [3]Analytical Research and Development, Merck & Co., Inc., Rahway, NJ 07065, USA. [4]Modeling and Informatics, Merck & Co., Inc., Kenilworth, NJ 07033, USA. [5]Discovery Sample Management, Merck & Co., Inc., Kenilworth, NJ 07033, USA. [6]Discovery Chemistry, Merck & Co., Inc., Boston, MA 02115 USA. [7]Process Research and Development, Merck & Co., Inc., Rahway, NJ 07065, USA. *Present address: Department of Medicinal Chemistry, Department of Chemistry, Program in Chemical Biology, University of Michigan, Ann Arbor, MI 48109, USA. †Present address: Princeton Catalysis Initiative, Princeton University, Princeton, NJ 08544, USA. **‡Corresponding author. Email: huaming.sheng@merck.com (H.S.); christopher@welchinnovation.com (C.J.W.); spencer_dreher@merck.com (S.D.D.)** §Present address: Indiana Consortium for the Analytical Sciences, Indianapolis, IN 46202, USA. ||Present address: Welch Innovation, LLC, Cranbury NJ 08512, USA.

limiting study to extremely fast kinetics that can keep up with serial high-performance liquid chromatography (HPLC) analysis. Rather than accepting such constraints, we systematically engineered and validated new plate-based nanomole synthesis tools with a general ability to carry out a wide variety of typical synthetic reactions. We identified effective chemically compatible glass microplate reactors, performed fast 384-tip dosing of reagent solutions in volatile solvents, and designed aluminum sealing blocks that can retain volatile solvents on heating. In addition, we used LabRam resonant acoustic mixing to both agitate reactions and create milky slurries of solid inorganic bases that can be added to reactions by parallel liquid handling. Finally, we created a nanomole-scale photochemistry tool to enable reaction evaluation in the rapidly growing field of photoredox catalysis. We describe the use of these tools in reaction profiling experiments below, but they are expected to benefit extensions to nanoscale chemical and biological evaluation as well (*15*).

MALDI-TOF MS reaction analysis

Owing to the sheer number of experiments necessary, surveying chemical reactivity can be limited by the rate of analytical measurement. Despite considerable recent progress, HPLC-MS assays typically used for such screening are constrained by the fastest speed with which samples can be mechanically taken up and injected (~10 s per sample) (*16*). Spectroscopic techniques can be faster but often lack the resolving power needed to study a diverse collection of molecules. High-throughput mass spectrometry provides excellent mass-to-charge ratio separation of individual products (*17*–*20*), and MALDI-MS has been intensively optimized in recent years for rapid and robust molecule-specific MS imaging of biological tissue slices. The new generation of MALDI–time-of-flight (TOF) instruments equipped with a 10-kHz scanning beam laser, substantially faster *X*,*Y* stage, and faster plate loading-unloading cycle substantially improves sample throughput to hundreds of thousands of measurements per day. We were therefore intrigued by the possibility of applying this fast-analysis approach (several samples per second) to high-throughput reaction screening (HTS). Although MALDI is commonly used for high-throughput analysis, applications are typically limited to either quantitative biochemical assays that feature a single substrate or product within each well (*21*–*28*) or chemistry discovery using an MS-active tag or special MALDI plate to ensure MS detection of all labeled products (*29*–*31*). Neither of these approaches enabled the rapid, label-free analysis of a diverse collection of products that we required.

Direct MS analysis of small molecules can be vulnerable to interference from reaction components that would typically be removed by chromatography. To analyze samples by MALDI-TOF MS, we found a suitable ionization matrix, and instrument settings afforded excellent performance for diverse pharmaceutically relevant products, with the addition of an ionizing internal standard helping to normalize the effects of different reaction conditions on product ionization and MALDI spot formation. Screening and optimization of MALDI ionization matrices were carried out using a library of druglike informer compounds (*32*), leading to the identification of α-cyano-4-hydroxycinnamic acid (CHCA) as a generally preferred matrix [4 mg/ml in 50% H_2O/acetonitrile (ACN) with 0.1% trifluoroacetic acid (TFA)] affording an optimal signal-to-noise ratio across a wide range of protonated model compounds. A general protocol including MALDI plate preparation, automated data acquisition, and data processing was established for MALDI analysis, in which 175 nl of quenched reaction mixture in dimethyl sulfoxide (DMSO) with added internal standard was directly spotted on an HTS MALDI target plate in 1536 format using a positive-displacement liquid-handling robot with 16 channels. In our experiments, the spotting time for a single plate was ~30 min, but MALDI plate preparation time could be shortened by using a liquid-handling robot with a 1536-channel pipetting head. The resulting DMSO on the target plate could be evaporated either under vacuum at 0.01 bar (2 min) or at atmospheric pressure (1 hour). CHCA matrix solution (150 nl) was then applied onto the dried reaction spots through the same liquid-dispensing procedure. The readout speed for each 1536 MALDI plate ranged from 8 to 11 min, depending on the number of laser shots per spectrum. All automated MALDI runs were set up, triggered, and processed using dedicated uHTS software for MALDI-TOF instruments (Bruker MALDI PharmaPulse 2.0). In the whole-molecule analysis experiment, a custom processing script was used to simultaneously extract nearly 400 masses, peak intensities, and peak areas for reaction products and the internal standard into a single spreadsheet, in addition to

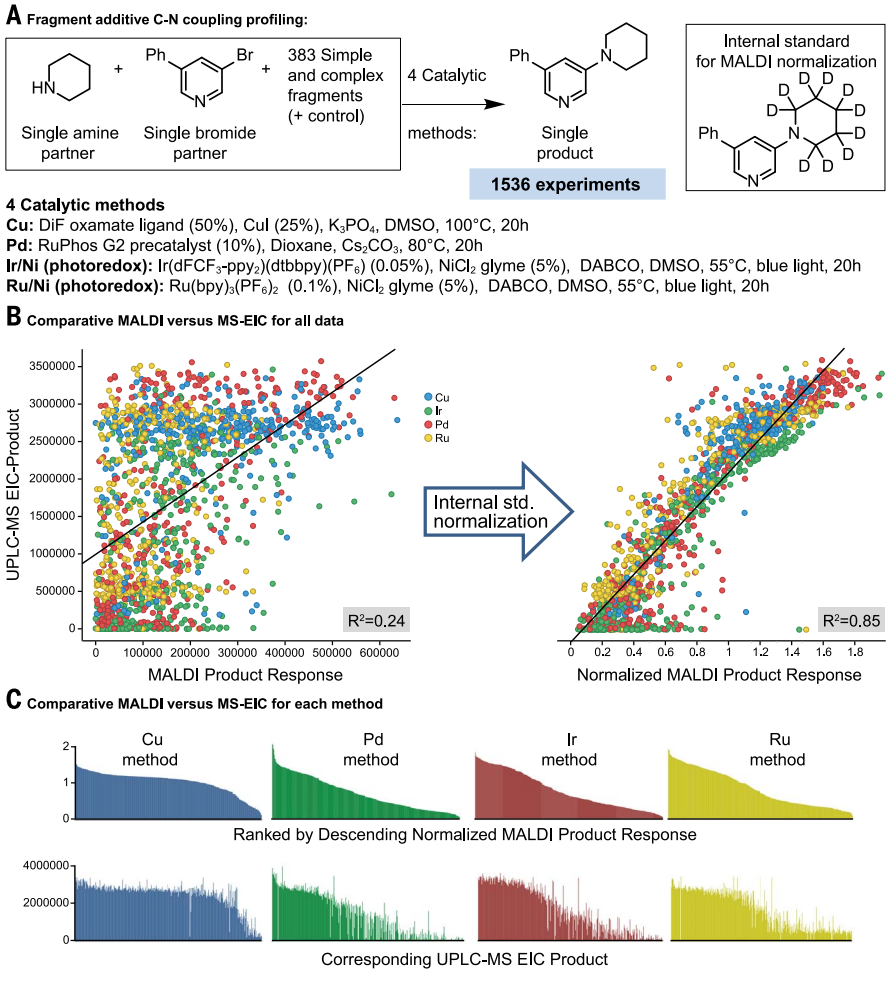


Fig. 1. uHT MALDI-MS approach to screening potential reaction poisons. (**A**) 1536 reactions were run with four leading C–N coupling methods to evaluate the impact of 383 single and polyfunctional fragments on one simple coupling reaction, requiring 11 min of MALDI acquisition time. Ph, phenyl; DiF, difluoro; RuPhos, 2-dicyclohexylphosphino-2′,6′-diisopropoxybiphenyl; ppy, 2-phenyl pyridinyl; dtbbpy, di-tert-butyl di-pyridyl; DABCO, 1,4-diazabicyclo[2.2.2]octane; bpy, bipyridinyl. (**B**) Raw MALDI product response is not well correlated with UPLC-MS or UPLC-UV measures, but normalization using an internal standard reveals very good correlation. (**C**) Ranked, normalized MALDI product response data with corresponding UPLC-MS (EIC) data below for comparison.

standard MALDI PharmaPulse processing. A signal-to-noise-ratio of 5/1 was used in the software to extract the final MALDI-MS data.

uHT fragment poisons analysis

To test our MALDI analytical approach for general reaction profiling, we began with an analytically simplified screening protocol, developed by Glorius and co-workers, for identifying reaction poisons—i.e., functional groups and reagents that interfere with a given reaction (Fig. 1) (*33*, *34*). Because the same reaction is investigated in each experimental well, a single product standard can be used throughout, greatly simplifying analysis. We realized that the throughput of the MALDI technique could enable screening of much larger sets of steric and electronic functional variants and higher-order polyfunctional molecular fragments to precisely identify specific deleterious binding motifs and cooperative interactions. At the same time, we could evaluate the ability of the MALDI system to provide accurate product quantification in the presence of different reaction conditions and a wide variety of additives.

We selected four different C–N coupling conditions (Fig. 1A)—Ir/Ni and Ru/Ni photoredox (*35*) as well as Pd and Cu conditions (*32*)—that were previously demonstrated to perform well using our chemistry informer libraries approach. Each method was evaluated using the new nanomole experimentation platforms in the presence of 383 polar molecular fragments (with one control reaction containing no additive) ranging from simple compounds containing a single functional group (such as amides, esters, and N-heterocycles) to more complex polyfunctional compounds. The resulting 1536 fragment-additive reactions were analyzed in a single MALDI experiment in 11 min, with the settings described above and using a deca-deuterated product analog, added to each well after reaction completion, as an internal standard (Fig. 1A). The reactions were also analyzed using a conventional 2-min ultra performance liquid chromatography (UPLC)–MS method. The raw MALDI product response shows only marginal correlation with the UPLC-MS or UPLC–ultraviolet absorption (UV) data for the same reactions [extracted ion count (EIC), coefficient of determination R^2 = 0.24; UV 210 nm, R^2 = 0.16]; however, normalization of the MALDI signal with the closely related internal standard signal affords notably good correlation (EIC, R^2 = 0.85), showing that chemistry-specific effects on MALDI signal intensity can be mitigated by the use of internal standards (Fig. 1B). The normalized MALDI product response metric, ranked from high to low for all four methods, is shown in Fig. 1C, with the corresponding UPLC-MS EIC data shown below for comparison. This analytical approach is similar to the application of MALDI-MS for HTS of enzyme inhibitors, but this work demonstrates that MALDI, in conjunction with an appropriate product standard, can be used for routine reaction screening in the presence of a diversity of interfering elements.

To simplify MALDI-based assessments in fragment-additive screening, we binned the normalized MALDI product response for each reaction into poisons (>50% signal knockdown) and nonpoisons (<50% signal knockdown) for each fragment and synthetic method (Fig. 2). Across all four methods, the MALDI assignment matched the UPLC-MS EIC signal ~95% of the time and the UPLC-UV 210-nm data ~88% of the time (we found that UPLC-UV 210-nm data often show peaks that interfere with the product peak). The fragment screening data showed the expected outcome that each of the synthetic methods is susceptible to poisoning by particular single functional groups (see figs. S19 and S20 for complete lists). We were also able to observe that when these single-functional poisons are incorporated as parts of polyfunctional fragments, reaction poisoning almost always occurs (>90% of the time). Notably, polyfunctional fragments composed of functional groups that are not themselves poisons often (~50% of the time) act as poisons as well. The structural details of this poisoning are very specific. For example, we observed cases in which the combination of N-heterocycle and ketone strongly poisons a given catalytic system, whereas other similar structural combinations do not (Fig. 2). This work provides some indication of how screening large permutations of multifunctional structures can identify precise motifs that lead to reaction failure. We were gratified to find that with a large number of experiments, the effects of false negatives and positives on identifying trends are, to some degree, offset by the internal validation that stems from using multiple fragments that are structurally related. The fragment-based analysis herein reveals Cu to be substantially more resilient to single and poly-functional poisoning than the other methods evaluated, which should support its wider use in complex, densely functionalized substrates. The Ir/Ni and Ru/Ni photoredox methods were very similar to each other in their inhibition profiles and also generally performed very similarly to Pd. Most importantly,

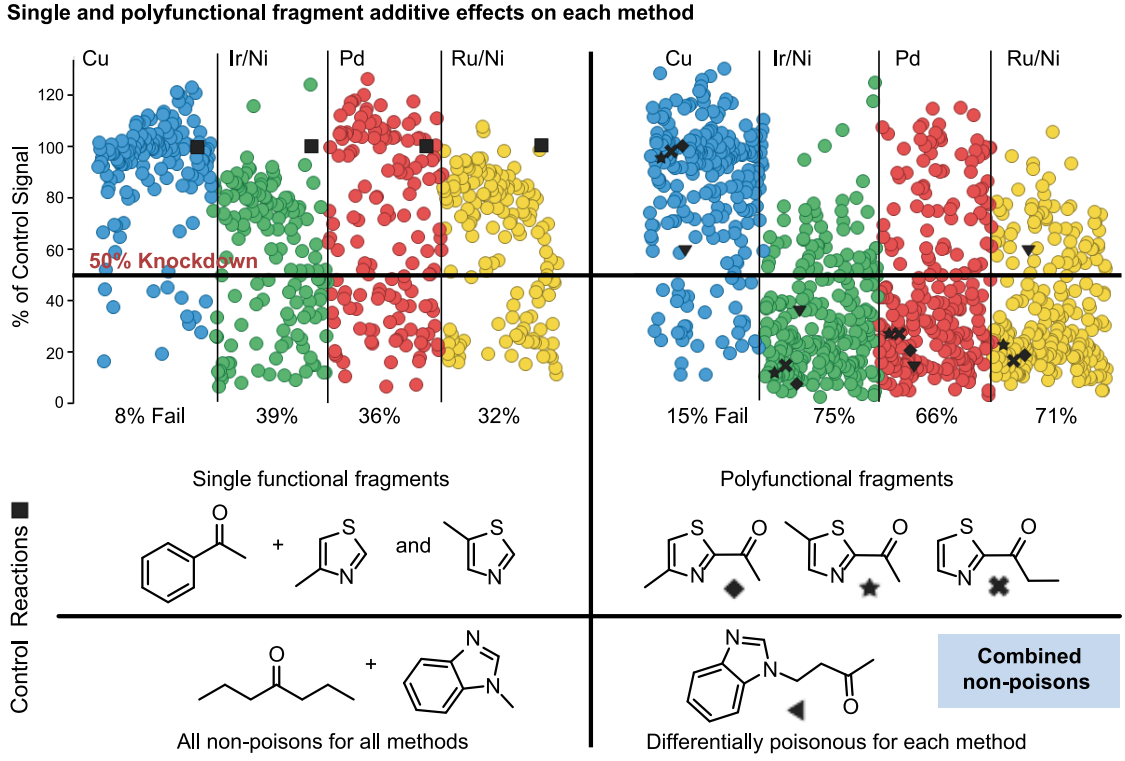


Fig. 2. Fragment additives poisoning study. Comparative rates of poisoning determined by MALDI analysis for single functional fragments and polyfunctional fragments for the four tested methods are displayed. This study reveals that Cu is well-suited to managing diverse functionality. Polyfunctional fragments that are composed of linked, single functional fragments that are not poisons themselves often lead to differential poisoning.

the specific problematic functional arrays that poison each method can be roughly mapped, and subsequent detailed nanomole chemistry exploration with structural variants can provide focused data for effective ML predictive modeling [as in (*8*)].

uHT simplest-partner whole-molecule analysis

We next investigated the much more challenging use of MALDI-MS to characterize the effects present in large, whole-molecule substrate sets (Fig. 3). Whereas the fragments approach is useful for rapid identification of problematic functionality, classifying local steric and electronic as well as bulk molecular effects such as solubility can be accomplished only by using whole-molecule substrates. Previous work in our labs explored the effects of pharma-relevant whole-molecule informer compounds (*32*) on the performance of different synthetic methods. Though useful to begin to systematically evaluate different synthetic methods, the 18 compounds in these test sets do not provide enough structural diversity to begin to assign general causes of reaction failure. In addition, the impact of amines in complex reactivity space was not previously taken into account. Hence, we envisioned a much larger virtual array of 192 N-heterocycle–containing aryl bromides crossed with 192 cyclic secondary amines (Fig. 3A), both of increasing molecular complexity, representing 36,864 distinct potential products. This array, if evaluated with the four aforementioned C–N coupling protocols, would result in 147,456 experiments. Although experimental assessment of all substrates and protocols is conceivable, we explored a systematic "simplest-partner" approach to study the isolated structural effects of individual building blocks in this space. In this protocol, each bromide is subjected to reaction with the simplest amine in the set and each amine with the simplest bromide. Applying the four different synthetic methods to this 384-substrate test set affords a total of 1536 experiments, or ~1% of the experimental space. To ensure that every reaction in the set would have a strong MALDI-MS response, the simplest coupling partners were chosen to incorporate a basic nitrogen atom, thereby ensuring MS detectability in the positive-ion mode. Although measuring products in negative-ion mode was not the focus of this work, analysis of small molecules by MALDI-TOF in negative-ion mode can be successfully achieved when appropriate sample preparation methods are used (*36*–*38*). Clearly, this general strategy would be poorly suited for the synthesis of hydrocarbons or other species that have poor MS ionization properties.

Traditional exploration of complex substrate space using HPLC analysis requires the preparation of product standards to obtain response factors for quantification and to confirm product identity for every new molecule of interest, an untenable proposition for our envisioned experiment. The prospect of adding an internal standard that is tailored to each product for MALDI normalization is equally daunting. We reasoned that in our search for causes of reaction failure, a focus on common structural reactivity trends rather than absolute product yields might afford valuable insights without the traditional need for individual product standards. Also, we hoped that the use of a single internal standard in all wells, although clearly not controlling for structure-based variations in MALDI signal intensity, might still be useful for normalizing the effects of MALDI spotting variability, thereby vastly simplifying experimental execution. Hence, we endeavored to use rapid MALDI responses for 384 different products coupled with parameterized and clustered molecular descriptors (such as functional group counts, steric hindrance, and bulk properties such as H-bond donors and acceptors and polar surface area) to identify fundamental incompatibilities between given sets of reaction conditions and the diverse structural properties within an enormous compound set. Once identified, such reactivity trends can be confirmed by conventional analytical methods. The MALDI signal for all 1536 reactions was acquired using a single internal standard for normalization across all reactions, as well as a conventional 2-min UPLC-MS analysis for comparison. Two-dimensional scatter plots for all 1536 reactions (Fig. 3B) reveal that the normalized product MALDI responses are poorly correlated compared with the UPLC-MS (EIC, R^2 = 0.33) and UPLC-UV data [total wavelength chromatogram (TWC), R^2 = 0.30; we found an averaged wavelength to be more useful in this diverse compounds set than a single wavelength, such as UV 210 nm]. However, we found that we could use the MALDI data to create binary thresholds for reaction success or failure that

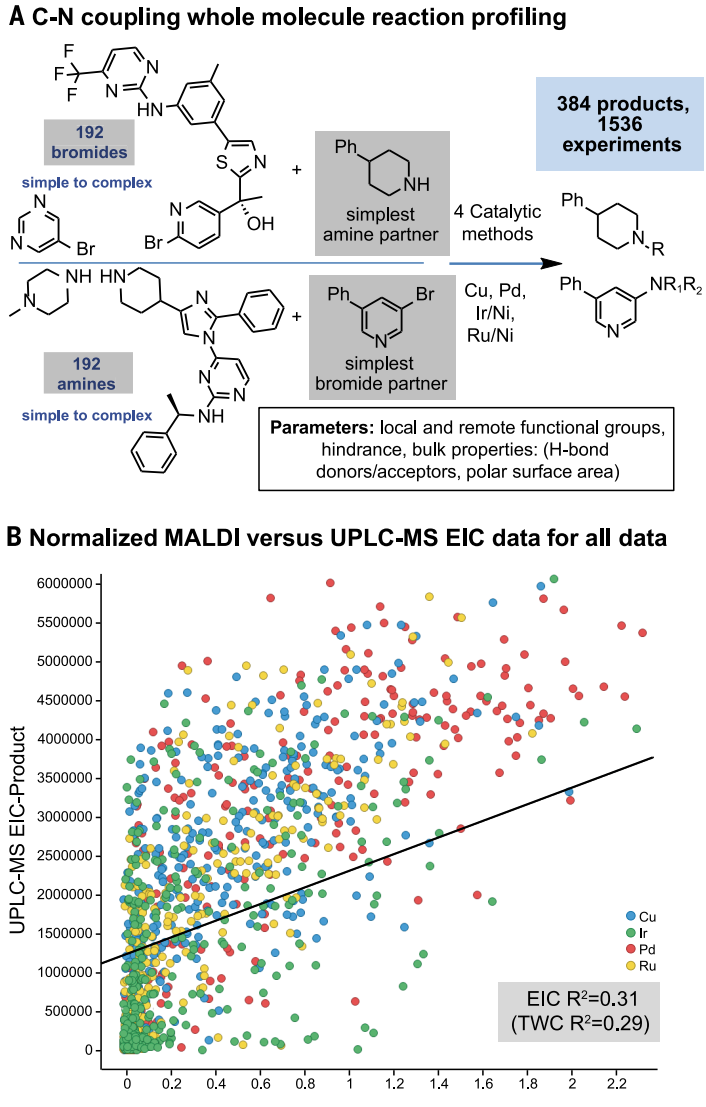


Fig. 3. Whole-molecule simplest-partner evaluation with MALDI analysis. (**A**) Bromides and amines (192 of each) are each crossed with a simple, mass-active coupling partner under the four previously described (Fig. 1A) catalytic methods. (**B**) Using a single internal standard, the normalized MALDI data show poor correlation with UPLC-MS and UPLC-UV metrics.

permitted us to identify general structural entities that correlate with reaction failure within a large, whole-molecule set. To this end, we set an arbitrary pass-fail threshold of 20% of the average MALDI value across the entire set. When the binned data for each synthetic method are clustered as a percentage of failed reactions for different structural parameters that have at least 10 examples within a set (Fig. 4A), the data from the normalized MALDI experiment show essentially the same trends revealed in the UPLC-MS data (both EIC and TWC responses), suggesting that MALDI can be used to stack-rank the structural entities that most often cause reaction failure for each set of conditions (see figs. S26 to S33 for complete lists). Again, any of these identified problematic structural parameters can be promoted for higher-order systematic ML studies.

Although lists of potential poisons for specific synthetic methods are useful, understanding the comparative reactivity of different synthetic methods for problematic functionality is arguably even more important. Figure 4B shows how the clustered binary pass-fail data for all four synthetic methods can be combined into a single graph that reveals structural entities that are generally problematic for bromides and amines

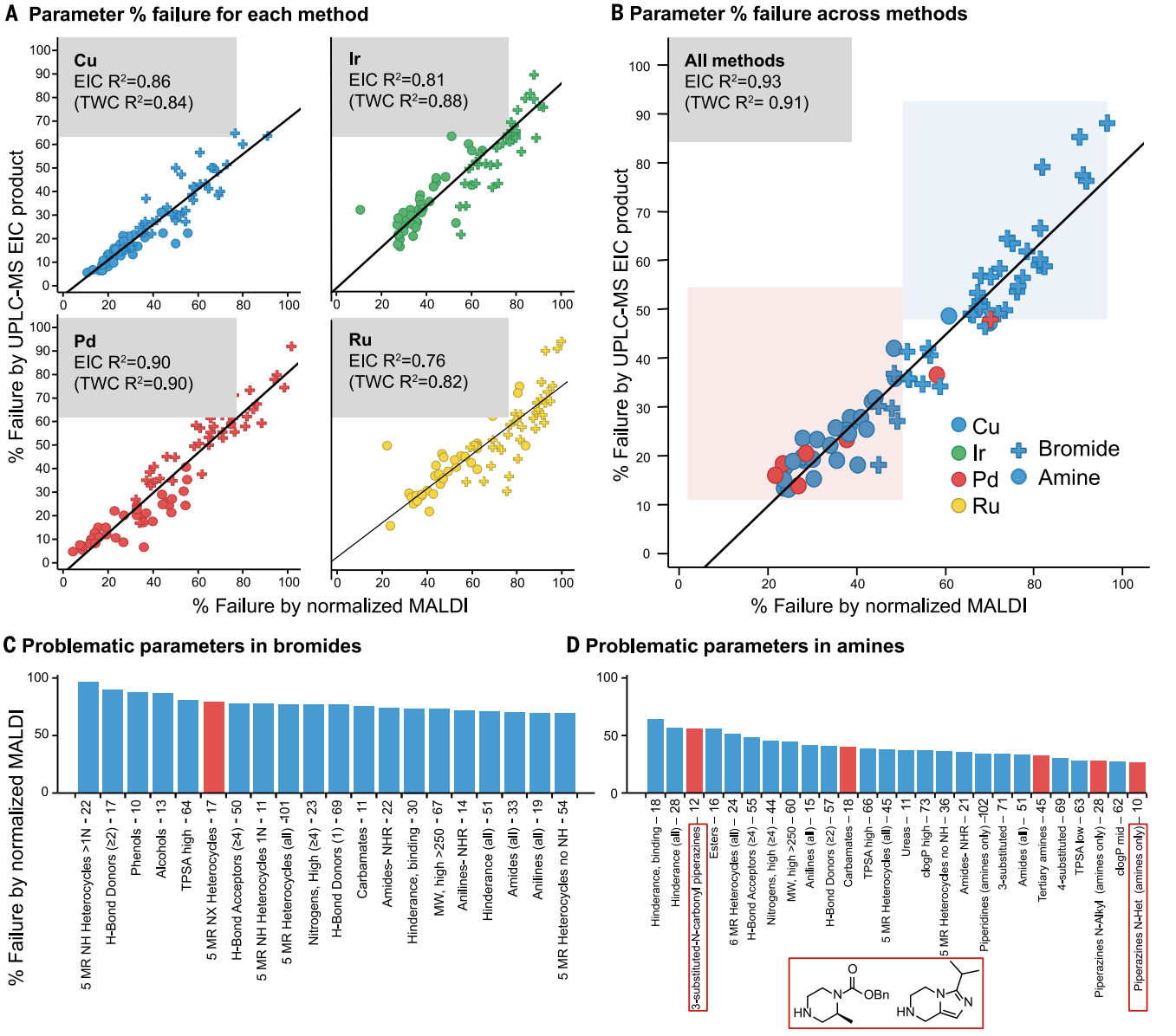


Fig. 4. Whole-molecule reactivity trends from binary thresholding analysis. (**A**) Failure percentage for different clustered parameters determined by pass-fail binary binning using thresholded MALDI data correlates very well with similarly binned UPLC-MS EIC and UPLC-UV TWC data. Each point on the graph represents the failure rate of different specific aggregated structural parameters, such as NH heterocycles or H-bond donors, for each synthetic method. Circles and crosses denote whether the trend is for amines (circles) or bromides (crosses). The symbol color of each circle or cross indicates which synthetic method was used for the data point. (**B**) The average failure rate for clustered parameters for all four methods reveals the functionality that is problematic across methods. The color of each data point reveals which method has the lowest failure rate. (**C** and **D**) The most problematic parameters for aryl bromides (blue shaded area from Fig. 4B) and amines (red shaded area in Fig. 4B) are listed in descending order of failure rates across methods, along with the number of examples in the test set. In nearly all cases, Cu is the preferred method. However, several very specific structural types in amines are found to be problematic for Cu and show greater reactivity with Pd. MR, membered ring; TPSA, total polar surface area; NHR, nitrogen with an alkyl group and a hydrogen; MW, molecular weight; clogP, calculated logP (logarithm of the partition coefficient between *n*-octanol and water); Bn, benzyl.

across all methods. The data can serve as a valuable map for academic research focused on creating new methods to overcome these common problems. This graph also reveals which method affords the lowest failure rate for each parameter. Figure 4B clearly shows the synthetic advantage of the Cu system in nearly all parameters used to describe aryl bromides (the most problematic are listed in Fig. 4C), which again substantiates its value in complex synthesis. At the same time, for a number of parameters within the amine structural space (Fig. 4B, red), Pd is the best catalyst. Diving deeper (Fig. 4D), we can identify very specific structural features within amines that are problematic for Cu and for which Pd can provide a synthetic advantage. Particularly notable is the reproducibly higher performance of Pd versus Cu for 3-substituted-4-*N*-carbonyl piperazines and 3,4-N-heterocycle–appended piperazines, revealing subtle conformational effects several bonds removed from the reactive amine that are not considered in the current understanding of C–N coupling reactivity. These systems were explored at a larger reaction scale (10 μmol, 40× scale-up) with additional structural examples to confirm the trends (see figs. S37 to S39 for further details on Cu versus Pd amine trends).

The fragment-additive and whole-molecule studies both indicate that Cu has a substantial advantage over the other methods examined in this study, with respect to single and polyfunctional group tolerance and scope. These observations hold with aryl bromides, but Cu also appears to have specific liabilities in the amine structural space. This study demonstrates that uHT fragment and whole-molecule approaches can provide complementary reactivity information, and both approaches will likely be inextricably linked in the future of predictive chemistry. Given the observed importance of precisely defined structures evident in this work, as well as a shortage of diverse test substrates, chemists will likely alternate between the two approaches to refine their maps of holistic reactivity effects. The most important message from this work is that, rather than accurately determining reaction yields, broad uHT mapping of substrate space with no isolated product standards can reveal specific molecular interactions using adequately discriminating binary analytical approaches that simply separate successes from failures. The reactivity trends identified by MALDI in these experiments were confirmed with conventional UV/MS scoring, which suggests that MALDI alone can be used to uncover results previously accessible only via slower analytical methods. Looking forward, we hope that the convenience, robustness, and speed of MALDI will facilitate creation of large sets of structural and mass response data to enable quantitative yield prediction to further increase the value of the MALDI-MS approach.

Full factorial whole-molecule space

The remaining 99% of the full factorial whole-molecule crossover substrate coupling space may contain additional information on higher-order bimolecular interactions that can also influence reaction performance. The simplest-partner approach can be used to triage the experiments required to map the enormity of the remaining space. We reasoned that if either compound in a complex pair performs poorly in the simplest-partner test, then structural poisoning will be likely in the more complex combination. Likewise, when both partners perform well with simple partners, the combination is likely to work well together. We investigated 288 experiments in this crossover space and found that we could assign >50% of the space into "hits" or "misses" (with 90% accuracy, see supplementary materials for details). The remaining space, with substrates having middling MALDI responses, could not be effectively resolved. This pruning approach allows us to chart complex bimolecular space using the minimal coverage provided by the simplest-partner experiments, thereby eliminating >70,000 experiments that do not need to be performed.

Toward predictive informatics

The miniaturized uHT reaction engineering and MALDI-TOF MS analytical advancements described in this work enable generation of large, structured datasets that pinpoint problematic structural elements within complex substrate space. Identified problems can then be promoted for more detailed mapping with the use of uHT experimentation coupled with atomic and quantum molecular physical descriptors to enable structure-based ML. In this work, we ran more than 3000 experiments studying the effects of just four synthesis conditions on a relatively narrow area of chemistry space (the coupling of cyclic secondary amines and N-heterocycle–containing aryl bromides). The real power in this approach will harness the pronounced analytical speed of MALDI analysis for the iterative evaluation of large substrate arrays against diverse catalysts, bases, solvents, reagent stoichiometries, and temperatures, enabling big data chemistry informatics in the search for general solutions to problematic areas in organic synthesis.

A persistent focus on revealing structural limitations in chemistry is a win-win scenario for academic researchers and synthetic practitioners alike. Dark space that remains inaccessible across synthetic methods becomes fodder for new experimental research of known value, and specific knowledge of limitations will lead to predictivity and understanding that will increase the speed and success of the design-make-test cycle, helping to remove synthetic problem-solving from the critical path. Even with optimally efficient tools and strategies in place, mapping the entire landscape of useful chemical reactivity is currently beyond the reach of any single organization. Given the promising enabling value of such a massive survey, a precompetitive public-private partnership to address this gap might even be an achievable and worthwhile goal.

Materials and methods summary

The specific extended nanomole-scale chemistry protocols used for the four described C–N coupling synthetic methods (Ir/Ni and Ru/Ni photoredox, Cu, and Pd), as well as the MALDI-MS and comparative UPLC-MS analysis protocols, are described in detail in the supplementary materials. A brief summary is provided below.

Description of extended nanomole-scale chemistry platform

Nanomole-scale chemistry reactions, to this point, have been run using plastic plates with low-volatility, plastic-compatible solvents (polar aprotics, DMSO, and *N*-methyl-2-pyrrolidone) and homogeneous components (bases and catalysts) at room temperature. In this work, reaction engineering advancements have enabled a much wider scope of potential reaction conditions. The use of glass 1536-well plates and rapid 384-tip pipetting enables the use of volatile, non–plastic-compatible solvents (such as dioxane, used in this study) and substrates (piperidine). Development of a resonant acoustic LabRam grinding protocol to produce long-lasting slurries that can be dosed with liquid-handling robotics has enabled the use of heterogeneous inorganic bases (Cs_2CO_3 and K_3PO_4). An aluminum reactor block was designed that can provide a high-quality seal to prevent solvent loss, and a heating mechanism was devised that can be equipped to heat the aluminum block inside the LabRam, which provides a robust mechanism for parallel reaction agitation. Finally, a nanomole-scale photochemistry tool was created using a modified aluminum reactor with an acrylic plastic bottom that allows uniform light penetration.

Description of MALDI-TOF analysis

All reaction mixtures were quenched and diluted to standard UPLC-MS concentration. These quenched reactions were then spotted using a Mosquito HTS liquid-handling robot on Bruker HTS MALDI targets with barcodes (1.0-mm thickness) in 1536 format [HTS MALDI plate 1.0 mm, BC, part 1833280]. The targets were mounted on Bruker HTS MALDI adapters (part 1847571). Reaction mixture (175 nl) was deposited and allowed to dry, followed by depositing 150 nl of a 4-mg/ml solution of α-cyano-4-hydroxycinnamic acid in 0.1% TFA, 50% ACN/H_2O. These targets were analyzed on a Bruker Rapiflex MALDI-TOF/TOF system.

Description of comparative UPLC-MS analysis

The quenched reactions were also monitored using a Waters Acquity UPLC I-Class system (Waters Corp.) equipped with a binary pump, flow-through needle sampler, column manager, photodiode array detector, SQ detector 2 with electrospray ionization source in the positive mode, and MassLynx software. Separations were performed on a Waters CORTECS UPLC C18+ column (dimensions: 30 mm by 2.1 mm; particle size: 1.6 μm).

REFERENCES AND NOTES

1. A. Buitrago Santanilla *et al.*, Nanomole-scale high-throughput chemistry for the synthesis of complex molecules.

Science **347**, 49–53 (2015). doi: 10.1126/science.1259203; pmid: 25554781
2. D. Perera *et al*., A platform for automated nanomole-scale reaction screening and micromole-scale synthesis in flow. *Science* **359**, 429–434 (2018). doi: 10.1126/science.aap9112; pmid: 29371464
3. S. W. Krska, D. A. DiRocco, S. D. Dreher, M. Shevlin, The Evolution of Chemical High-Throughput Experimentation To Address Challenging Problems in Pharmaceutical Synthesis. *Acc. Chem. Res.* **50**, 2976–2985 (2017). doi: 10.1021/acs.accounts.7b00428; pmid: 29172435
4. V. I. Martin, J. R. Goodell, O. J. Ingham, J. A. Porco Jr., A. B. Beeler, Multidimensional reaction screening for photochemical transformations as a tool for discovering new chemotypes. *J. Org. Chem.* **79**, 3838–3846 (2014). doi: 10.1021/jo500190b; pmid: 24697145
5. K. Troshin, J. F. Hartwig, Snap deconvolution: An informatics approach to high-throughput discovery of catalytic reactions. *Science* **357**, 175–181 (2017). doi: 10.1126/science.aan1568; pmid: 28706066
6. A. McNally, C. K. Prier, D. W. MacMillan, Discovery of an α-amino C-H arylation reaction using the strategy of accelerated serendipity. *Science* **334**, 1114–1117 (2011). doi: 10.1126/science.1213920; pmid: 22116882
7. M. N. Hopkinson, A. Gómez-Suárez, M. Teders, B. Sahoo, F. Glorius, Accelerated Discovery in Photocatalysis using a Mechanism-Based Screening Method. *Angew. Chem. Int. Ed.* **55**, 4361–4366 (2016). doi: 10.1002/anie.201600995; pmid: 27000485
8. D. T. Ahneman, J. G. Estrada, S. Lin, S. D. Dreher, A. G. Doyle, Predicting reaction performance in C-N cross-coupling using machine learning. *Science* **360**, 186–190 (2018). doi: 10.1126/science.aar5169; pmid: 29449509
9. A. Milo, A. J. Neel, F. D. Toste, M. S. Sigman, A data-intensive approach to mechanistic elucidation applied to chiral anion catalysis. *Science* **347**, 737–743 (2015). doi: 10.1126/science.1261043; pmid: 25678656
10. C. W. Coley, R. Barzilay, T. S. Jaakkola, W. H. Green, K. F. Jensen, Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **3**, 434–443 (2017). doi: 10.1021/acscentsci.7b00064; pmid: 28573205
11. S. Szymkuć *et al*., Computer-Assisted Synthetic Planning: The End of the Beginning. *Angew. Chem. Int. Ed.* **55**, 5904–5937 (2016). doi: 10.1002/anie.201506101; pmid: 27062365
12. A. Lavecchia, Machine-learning approaches in drug discovery: Methods and applications. *Drug Discov. Today* **20**, 318–331 (2015). doi: 10.1016/j.drudis.2014.10.012; pmid: 25448759
13. M. H. Todd, Computer-aided organic synthesis. *Chem. Soc. Rev.* **34**, 247–266 (2005). doi: 10.1039/b104620a; pmid: 15726161
14. R. Sheridan *et al*., Toward structure-based predictive tools for the selection of chiral stationary phases for the chromatographic separation of enantiomers. *J. Chromatogr. A* **1467**, 206–213 (2016). doi: 10.1016/j.chroma.2016.05.066; pmid: 27318509
15. N. J. Gesmundo *et al*., Nanoscale synthesis and affinity ranking. *Nature* **557**, 228–232 (2018). doi: 10.1038/s41586-018-0056-8; pmid: 29686415
16. C. J. Welch, Are We Approaching a Speed Limit for the Chromatographic Separation of Enantiomers? *ACS Cent. Sci.* **3**, 823–829 (2017). doi: 10.1021/acscentsci.7b00250; pmid: 28852695
17. M. Wleklinski *et al*., High throughput reaction screening using desorption electrospray ionization mass spectrometry. *Chem. Sci.* **9**, 1647–1653 (2018). doi: 10.1039/C7SC04606E; pmid: 29675211
18. S. Sun, T. R. Slaney, R. T. Kennedy, Label free screening of enzyme inhibitors at femtomole scale using segmented flow electrospray ionization mass spectrometry. *Anal. Chem.* **84**, 5794–5800 (2012). doi: 10.1021/ac3011389; pmid: 22656268
19. K. J. Boggio *et al*., Recent advances in single-cell MALDI mass spectrometry imaging and potential clinical impact. *Expert Rev. Proteomics* **8**, 591–604 (2011). doi: 10.1586/epr.11.53; pmid: 21999830
20. X. W. Diefenbach *et al*., Enabling Biocatalysis by High-Throughput Protein Engineering Using Droplet Microfluidics Coupled to Mass Spectrometry. *ACS Omega* **3**, 1498–1508 (2018). doi: 10.1021/acsomega.7b01973
21. M. S. Ritorto *et al*., Screening of DUB activity and specificity by MALDI-TOF mass spectrometry. *Nat. Commun.* **5**, 4763 (2014). doi: 10.1038/ncomms5763; pmid: 25159004
22. C. Haslam *et al*., The Evolution of MALDI-TOF Mass Spectrometry toward Ultra-High-Throughput Screening: 1536-Well Format and Beyond. *J. Biomol. Screen.* **21**, 176–186 (2016). doi: 10.1177/1087057115608605; pmid: 26428484
23. M. Winter *et al*., Establishing MALDI-TOF as Versatile Drug Discovery Readout to Dissect the PTP1B Enzymatic Reaction. *SLAS Discov.* **23**, 561–573 (2018). doi: 10.1177/2472555218759267; pmid: 29466676
24. K. Beeman *et al*., Integration of an In Situ MALDI-Based High-Throughput Screening Process: A Case Study with Receptor Tyrosine Kinase c-MET. *SLAS Discov.* **22**, 1203–1210 (2017). doi: 10.1177/2472555217727701; pmid: 28820955
25. R. E. Heap *et al*., Identifying Inhibitors of Inflammation: A Novel High-Throughput MALDI-TOF Screening Assay for Salt-Inducible Kinases (SIKs). *SLAS Discov.* **22**, 1193–1202 (2017). doi: 10.1177/2472555217717473; pmid: 28692323
26. T. Si *et al*., Profiling of Microbial Colonies for High-Throughput Engineering of Multistep Enzymatic Reactions via Optically Guided Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *J. Am. Chem. Soc.* **139**, 12466–12473 (2017). doi: 10.1021/jacs.7b04641; pmid: 28792758
27. M. W. Duncan, H. Roder, S. W. Hunsucker, Quantitative matrix-assisted laser desorption/ionization mass spectrometry. *Brief. Funct. Genomics* **7**, 355–370 (2008). doi: 10.1093/bfgp/eln041; pmid: 19106161
28. P. Wang, R. W. Giese, Recommendations for quantitative analysis of small molecules by matrix-assisted laser desorption ionization mass spectrometry. *J. Chromatogr. A* **1486**, 35–41 (2017). doi: 10.1016/j.chroma.2017.01.040; pmid: 28118972
29. J. R. Cabrera-Pardo, D. I. Chai, S. Liu, M. Mrksich, S. A. Kozmin, Label-assisted mass spectrometry for the acceleration of reaction discovery and optimization. *Nat. Chem.* **5**, 423–427 (2013). doi: 10.1038/nchem.1612; pmid: 23609094
30. T. J. Montavon, J. Li, J. R. Cabrera-Pardo, M. Mrksich, S. A. Kozmin, Three-component reaction discovery enabled by mass spectrometry of self-assembled monolayers. *Nat. Chem.* **4**, 45–51 (2012). doi: 10.1038/nchem.1212; pmid: 22169871
31. A. B. Diagne, S. Li, G. A. Perkowski, M. Mrksich, R. J. Thomson, SAMDI Mass Spectrometry-Enabled High-Throughput Optimization of a Traceless Petasis Reaction. *ACS Comb. Sci.* **17**, 658–662 (2015). doi: 10.1021/acscombsci.5b00131; pmid: 26521847
32. P. S. Kutchukian *et al*., Chemistry informer libraries: A chemoinformatics enabled approach to evaluate and advance synthetic methods. *Chem. Sci.* **7**, 2604–2613 (2016). doi: 10.1039/C5SC04751J; pmid: 28660032
33. K. D. Collins, F. Glorius, A robustness screen for the rapid assessment of chemical reactions. *Nat. Chem.* **5**, 597–601 (2013). doi: 10.1038/nchem.1669; pmid: 23787750
34. J. Richardson, J. C. Ruble, E. A. Love, S. Berritt, A Method for Identifying and Developing Functional Group Tolerant Catalytic Reactions: Application to the Buchwald-Hartwig Amination. *J. Org. Chem.* **82**, 3741–3750 (2017). doi: 10.1021/acs.joc.7b00201; pmid: 28245358
35. E. B. Corcoran *et al*., Aryl amination using ligand-free Ni(II) salts and photoredox catalysis. *Science* **353**, 279–283 (2016). doi: 10.1126/science.aag0209; pmid: 27338703
36. A. R. Korte, Y. J. Lee, MALDI-MS analysis and imaging of small molecule metabolites with 1,5-diaminonaphthalene (DAN). *J. Mass Spectrom.* **49**, 737–741 (2014). doi: 10.1002/jms.3400; pmid: 25044901
37. L. J. Soltzberg, P. Patel, Small molecule matrix-assisted laser desorption/ionization time-of-flight mass spectrometry using a polymer matrix. *Rapid Commun. Mass Spectrom.* **18**, 1455–1458 (2004). doi: 10.1002/rcm.1505; pmid: 15216505
38. R. Shi *et al*., Hydroxyl-Group-Dominated Graphite Dots Reshape Laser Desorption/Ionization Mass Spectrometry for Small Biomolecular Analysis and Imaging. *ACS Nano* **11**, 9500–9513 (2017). doi: 10.1021/acsnano.7b05328; pmid: 28850220

ACKNOWLEDGMENTS

We thank A. Donofrio (Merck & Co., Inc., Kenilworth, NJ, USA) for helpful discussions. **Funding:** S.L. and K.Z. are grateful to the MRL Postdoctoral Research Fellows program for financial support. Additional financial support was provided by Merck & Co., Inc., Kenilworth, NJ, USA. **Author contributions:** S.L., R.P.S., Z.P., I.W.D., D.A.D., and S.D.D. designed the chemistry experiments. S.L., R.D.F., H.S., and S.D.D. performed the chemistry experiments. S.L., R.D.F, D.V.C., T.C., I.W.D., D.A.D., and S.D.D. developed the extended nanochemistry tools. S.L., S.D., W.D.B., H.W., H.S., and C.J.W. developed the MALDI-MS analytical platform. S.L., K.Z., and S.D.D. analyzed comparative UPLC-MS data. S.L., R.P.S., Z.P., and S.D.D. analyzed the reactivity trends. Z.P. conducted ELN data mining. S.L., S.D., T.C., I.W.D., H.S., C.J.W., and S.D.D. wrote the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data described in this work are included in the Excel file associated with the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/eaar6236/suppl/DC1
Materials and Methods
Figures S1 to S55
Tables S1 to S8
References
Data S1 to S5

1 December 2017; accepted 15 May 2018
Published online 24 May 2018
10.1126/science.aar6236

RESEARCH ARTICLE

2D MATERIALS

The role of electron-electron interactions in two-dimensional Dirac fermions

Ho-Kin Tang[1,2], J. N. Leaw[1,2], J. N. B. Rodrigues[1,2], I. F. Herbut[3], P. Sengupta[1,4], F. F. Assaad[5], S. Adam[1,2,6]*

The role of electron-electron interactions in two-dimensional Dirac fermion systems remains enigmatic. Using a combination of nonperturbative numerical and analytical techniques that incorporate both the contact and long-range parts of the Coulomb interaction, we identify the two previously discussed regimes: a Gross-Neveu transition to a strongly correlated Mott insulator and a semimetallic state with a logarithmically diverging Fermi velocity accurately described by the random phase approximation. We predict that experimental realizations of Dirac fermions span this crossover and that this determines whether the Fermi velocity is increased or decreased by interactions. We explain several long-standing mysteries, including why the observed Fermi velocity in graphene is consistently about 20% larger than values obtained from ab initio calculations and why graphene on different substrates shows different behaviors.

In 1952, Freeman Dyson made the argument that all theoretical solutions to problems in quantum electrodynamics have the form of a perturbative asymptotic series expansion in the fine-structure constant α, where the Coulomb potential is of the form α/r, with r being the distance between two electrons. Furthermore, he showed that such solutions are uncontrolled beyond the order of perturbation theory given by the inverse fine-structure constant (*1*). For condensed matter realizations of two-dimensional (2D) Dirac fermions, the equivalent of the fine-structure constant is the long-range Coulomb coupling constant $\alpha \sim 1$, which implies that any perturbative theory is potentially uncontrolled (*2*). Indeed, first-order perturbation theory for such systems gives rather peculiar results. To first order in α, the inverse coupling constant (proportional to the Fermi velocity) itself diverges both in the infrared (long distances) and ultraviolet (small distances). Introducing a lattice scale fixes the ultraviolet divergence but not the infrared one (*3*, *4*).

This divergence is what lead Ye and Sachdev (*5*) to describe the effects of long-range electron-electron interactions as "dangerously irrelevant": Although the system flows under the renormalization group to a noninteracting theory, physical observables are strongly renormalized by the Coulomb interaction. Since then, dozens of theoretical works (*6*) have confirmed this basic picture: In the absence of disorder, and precisely at half filling, the role of long-range Coulomb interactions is to renormalize the electron Fermi velocity to infinity (limited only by the speed of light, if one includes a dynamical interaction).

In a parallel development, the Hubbard model on a honeycomb lattice provided a low-energy realization of Dirac fermions interacting through a short-range contact interaction U. Increasing the short-range interaction results in a quantum phase transition at a critical value $U = U_c$ from a semimetal to an antiferromagnetic Mott insulator. This phase transition was predicted to be of the Gross-Neveu universality class (*7*), which was recently confirmed numerically (*8*, *9*). In diagrammatic perturbation theory, typically both the Fermi velocity and quasiparticle residue vanish at a quantum phase transition. However, for the Gross-Neveu critical point, the Fermi velocity remains finite (despite the vanishing of the quasiparticle residue), and this suppression of the Fermi velocity to a finite value has been observed numerically (*10*). Renormalization group studies (*11*) have shown that the Fermi velocity is not modified by weak short-range interactions, and, in our numerics below, we verify all these features for the Hubbard model on a honeycomb lattice. However, because 2D Dirac fermions are unable to screen the long-range Coulomb potential, it is widely believed (*6*, *12*) that this pure on-site Hubbard model has limited applicability to experiments done in real materials.

Experimentally, 2D Dirac fermions can be realized in a variety of condensed matter systems, including on the surfaces of 3D topological insulators (*13*, *14*) and in artificial graphene made from quantum corrals of carbon monoxide arranged in a honeycomb lattice on a copper substrate (*15*), as well as in other systems (*16*). For concreteness, we focus our attention on graphene, the most studied and versatile realization of 2D Dirac fermions. Experiments have been unable to realize the precise configuration necessary to probe this strange interacting metallic state that features electron quasiparticles moving at the speed of light, despite their quasiparticle character smearing away; however, several probes of ultraclean graphene—including magnetotransport (*17*), infrared spectroscopy (*18*), capacitance (*19*), angle-resolved photoemission spectroscopy (*20*), tunneling spectra (*21*), and Raman scattering (*22*)—all reveal a clear breakdown of the noninteracting theory.

In this work, we address the competing effects of short-range and long-range parts of any realistic model of the Coulomb interaction. We use a nonperturbative, numerically exact, projective quantum Monte Carlo method to study the evolution of physical observables in a controllable manner. We find that, in the regime dominated by long-range interactions, there is an enhancement of Fermi velocity consistent with perturbation theory. Conversely, close to the phase transition dominated by short-range interactions, we find a suppression of Fermi velocity and a collapse of the numerical data for different values of the ratio between long-range and short-range interactions onto one curve. Our numerical results interpolate between these two limits and are valid for all interaction strengths, but are constrained by the finite system sizes in our simulations. Therefore, we use a renormalization group scheme to extrapolate the quantum Monte Carlo results to experimentally relevant energy scales, where we predict that observables will depend on both the short-range and long-range components of the Coulomb interaction as well as the energy scale of the observation (as we explain below, all these parameters can be tuned in current experiments). Moreover, we find that the lattice scale not only regularizes the ultraviolet divergence of the Fermi velocity but can also make the infrared divergence unnoticeable in the experimental window.

Theoretical model

To accomplish this, we study interacting fermions on a honeycomb lattice with competing short-range and long-range interactions. The model is described by the Hamiltonian $\hat{H} = -t\sum_{\langle ij\rangle,\sigma}(\hat{c}^{\dagger}_{i\sigma}\hat{c}_{j\sigma}+\text{h.c.}) + \frac{1}{2}\sum_{i,j}(\hat{n}_i-1)\mathcal{V}_{ij}(\hat{n}_j-1)$ where h.c. is the Hermitian conjugate, t is the tight-binding hopping, the second-quantized

[1]Centre for Advanced 2D Materials, National University of Singapore, 6 Science Drive 2, 117546 Singapore. [2]Department of Physics, Faculty of Science, National University of Singapore, 2 Science Drive 3, 117542 Singapore. [3]Department of Physics, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada. [4]School of Physical and Mathematical Sciences, Nanyang Technological University, 21 Nanyang Link, 637371 Singapore. [5]Institut für Theoretische Physik und Astrophysik, Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany. [6]Yale-NUS College, 16 College Avenue West, 138527 Singapore.
*Corresponding author: Email: shaffique.adam@yale-nus.edu.sg

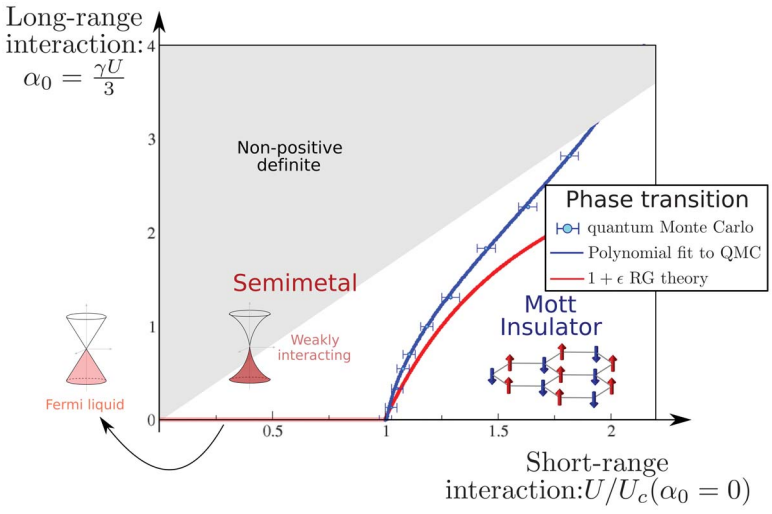


Fig. 1. Phase diagram for fermions on the honeycomb lattice with competing short-range and long-range Coulomb interactions. For any value of long-range interaction α_0, there is a critical value of the short-range interaction $U_c(\alpha_0)$ calculated by using quantum Monte Carlo (QMC) (data points), for which the system undergoes a quantum phase transition to the Mott insulator. In the presence of long-range interactions, a larger value of on-site interaction is required to reach the quantum phase transition. The phase diagram can be understood by solving the renormalization group (RG) flow equations (red curve), including both on-site and nearest-neighbor interactions, where the effective on-site interactions are reduced by the long-range Coulomb tail. The solid blue line is a quartic interpolation of the data points. The shaded portion shows the region inaccessible to our numerical method. Error bars indicate our numerical uncertainty. The Fermi liquid regime, the weakly interacting semimetal, and the strongly interacting Mott antiferromagnet are marked by illustrations.

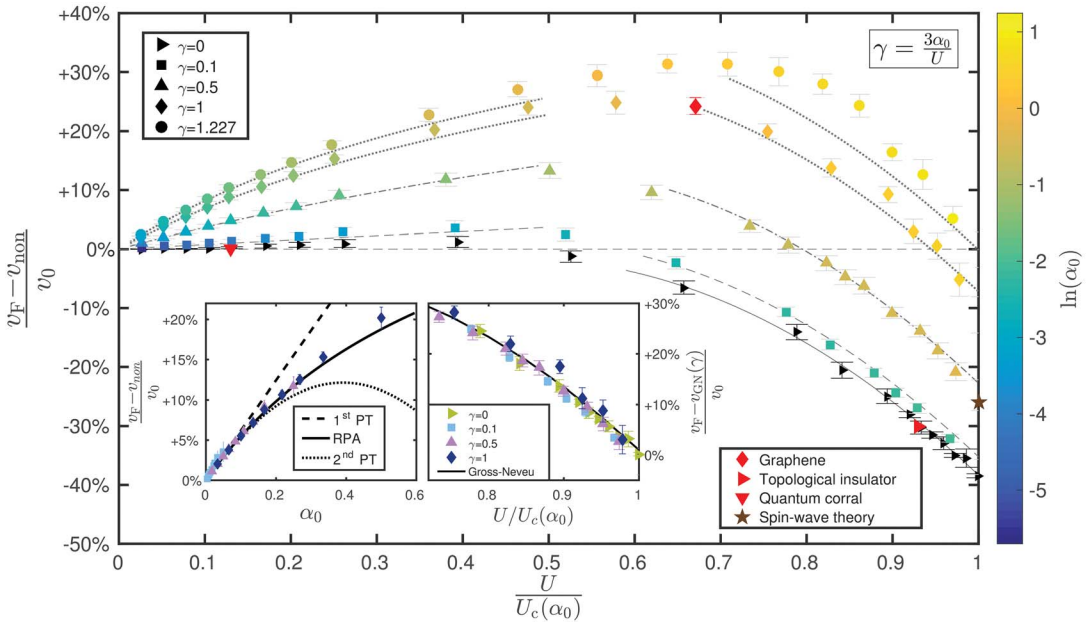


Fig. 2. Dirac fermion Fermi velocity renormalized by electron-electron interactions. Projective quantum Monte Carlo results for different short-range (U) and long-range (α_0) components of the Coulomb interaction. Plotted is the relative change of the Fermi velocity with respect to the noninteracting value at the Dirac point. Small $U/U_c(\alpha_0)$ defines the weak-coupling regime, where Monte Carlo data for different ratios γ of the long-range and short-range components collapse as a function of α_0; here electron-electron interactions enhance the Fermi velocity in agreement with the RPA (left inset). Perturbation theory (PT) results are also shown. A metal-to-Mott insulator phase transition of the Gross-Neveu universality class occurs at $U = U_c(\alpha_0)$, where a suppression of Fermi velocity can be understood as the coupling between Dirac fermions and the bosonic excitations of the nascent antiferromagnetic state (the brown star is an estimate of this Fermi velocity suppression determined by using spin-wave theory). The right inset shows quantum Monte Carlo data at various values of γ for the change in Fermi velocity from the value at the Gross-Neveu point collapse onto one curve as one moves away from the phase transition. Our numerics span the full crossover between the weak-coupling fixed point and the Gross-Neveu critical point. Estimates place topological insulators close to the phase transition, whereas quantum corral-like honeycomb lattices are in the weak-coupling limit. Graphene Dirac fermions lie somewhere in between these two regimes [see table S2 in (*16*)].

operator $\hat{c}^{\dagger}_{i\sigma}$ ($\hat{c}_{i\sigma}$) creates (annihilates) an electron of spin $\sigma = \uparrow\downarrow$ at position $\mathbf{r}_i$, and $\hat{n}_i = \sum_{\sigma} \hat{c}^{\dagger}_{i\sigma}\hat{c}_{i\sigma}$ gives the electron density at position $\mathbf{r}_i$. The interaction $\mathcal{V}_{ij}$ consists of a short-range part acting between electrons on the same site with different spins, $\mathcal{V}_{ii} = U$, and a long-range part depending on the distance between the electrons $r_{ij} = 2a|\mathbf{r}_i - \mathbf{r}_j|/3$ as $\mathcal{V}_{ij} = \alpha_0/r_{ij}$, where a is the lattice constant. We define $\gamma = 3\alpha_0/U$ as the ratio between the long-range and short-range components. The inclusion of the long-range component was made possible by recent developments in lattice quantum chromodynamics (*23*), which we adapt for our purposes here (*24*).

Including electron-electron interactions can do one of two things: The Dirac fermions can remain metallic but with a modified Fermi velocity or the interactions can gap the system, giving a Mott insulator. We use our quantum Monte Carlo method to map out the phase transition between the semimetallic phase and the Mott insulating phase by calculating the antiferromagnetic structure factor [see eq. S7 in (*16*)], which, in the thermodynamic limit, is finite for the Mott insulator and vanishing in the metallic phase. For any given strength of the long-range interaction, the structure factor shows a unique crossing point $U_c(\alpha_0)$ for

various system sizes [see fig. S3 in (*16*) for a representative example with $\gamma = 0.5$ and system size (L) = 6, 9, 12, and 15]. By tracing this crossing point for 10 different choices of γ, we can map out the phase diagram of this model shown in Fig. 1, highlighting the competing effects of the short-range and long-range parts of the Coulomb potential. We confirm an earlier finding (*25*) that the critical point remains in the Gross-Neveu universality class, even with the presence of the long-range Coulomb interaction. It is commonly believed (*6*, *12*) that increasing the strength of the Coulomb interaction will favor the Mott insulating phase. By contrast, we find that reducing the relative role of Coulomb interaction versus on-site interaction, for example, through dielectric screening (*26*) or biaxial strain (*24*), provides the most favorable route for realizing antiferromagnetic order in graphene experiments. Renormalization group calculations in $1+\epsilon$ dimensions are the preferred method to understand such phase transitions (*27*). In section 2.2 of (*16*), we extend this technique to calculate the spin-full Gross-Neveu model, from which the red curve in Fig. 1 is obtained. This gives us two intuitive ways of thinking about why the phase transition shifts to the right with increasing long-range interactions. First, as one increases the long-range Coulomb tail, the effective on-site potential decreases. (One can think of this, qualitatively, as the difference between the on-site potential and the nearest-neighbor potential.) Therefore, with the inclusion of the long-range piece, one needs a larger on-site potential to get the same effective critical Hubbard potential (*28*). Second, although the Hubbard potential favors an antiferromagnetic ground state, the nearest-neighbor potential favors instead a charge density wave ground state. By including the long-range piece, one needs a larger on-site potential to stabilize the antiferromagnetic phase.

Figure 2 shows our main numerical results on the Fermi velocity, which is the defining property of the massless Dirac spectrum. We plot the renormalization of Fermi velocity $(v_F - v_{\mathrm{non}})/v_0$ against $U/U_c(\alpha_0)$, which is the short-range component of the interactions normalized by the critical value obtained from the phase diagram in Fig. 1. Here v_F is the interaction-renormalized Fermi velocity obtained in quantum Monte Carlo at the simulation scale $\Lambda_s \equiv ka = 0.48$, where k is the momentum, v_{non} is the tight-binding Fermi velocity at the simulation scale, and v_0 is the tight-binding Fermi velocity at the Dirac point. Each data point in Fig. 2 is an extrapolation to the thermodynamic limit from four lattice sizes, each with an average of ~100,000 quantum Monte Carlo sweeps. The interacting Fermi velocity at the simulation scale is obtained by first determining the convergent ground state value of the unequal time Green function $G_k(\tau)$ for large τ, where the single exponential decay time, $\log[G_k(\tau)] \sim (E^{N+1}_{k,\mathrm{GS}} - E^{N}_{0,\mathrm{GS}})\tau$ determines the first excitation energy of the system [defined here as the energy difference between the ground state (GS) of $N + 1$ fermions with a total momentum k and the ground state of N fermions with zero total momentum]. Full details of our quantum Monte Carlo scheme (figs. S1 to S3), datasets, the analysis, and the discussion on the use of twisted boundary conditions (figs. S4 to S7) are provided in (*16*).

Our numerical data show that the Fermi velocity renormalization has notably different behavior for $U/U_c(\alpha_0) \ll 1$ (which we call the "weak-coupling regime") and $U/U_c(\alpha_0) \lesssim 1$, which is in the vicinity of the Gross-Neveu critical point. In the weak-coupling regime, we observe an increase in the Fermi velocity, and all the quantum Monte Carlo data for different ratios of short-range and long-range components γ collapse when plotted as a function of the long-range interaction α_0 (see Fig. 2, left inset). By contrast, in the vicinity of the Gross-Neveu critical point, the Hubbard model ($\alpha_0 = 0$) shows a 40% decrease in Fermi velocity. Even after including the long-range component of the Coulomb interaction, by subtracting the intercept of the Fermi velocity at the Gross-Neveu critical point, all the numerical data collapse to the Hubbard model function form (see Fig. 2, right inset). This shows that interacting fermions on a honeycomb lattice are governed by two very different fixed points: one controlled by the long-range interaction, giving an enhancement in Fermi velocity, and the other governed by the short-range interaction, giving a suppression of the Fermi velocity. As discussed in table S2, estimates for the realistic Coulomb potential in graphene place it in the crossover between these two regimes, whereas topological insulator Bi_2Se_3 is close to the Mott transition and artificial graphene using quantum corrals is in the weak-coupling regime (*16*). Our numerical results span the full crossover between these two regimes.

The emergence of a stable weak-coupling fixed point and an unstable Gross-Neveu fixed point is anticipated by renormalization group studies. In the weak-coupling regime, all the quantum Monte Carlo data (Fig. 2, lines plotted on the left-hand side of the main panel) can be reproduced by using a one-parameter random phase approximation (RPA) theory $v^{\mathrm{RPA}}(k) = v_0\left\{1 + [F_1(\alpha_0) - F_0(\alpha_0)]\ln\left(\frac{\Lambda}{|\Lambda_s|}\right)\right\}$, where Λ_s is our numerical scale and $\Lambda = 6.2 \pm 0.2$ is obtained from fitting the data for small α. The

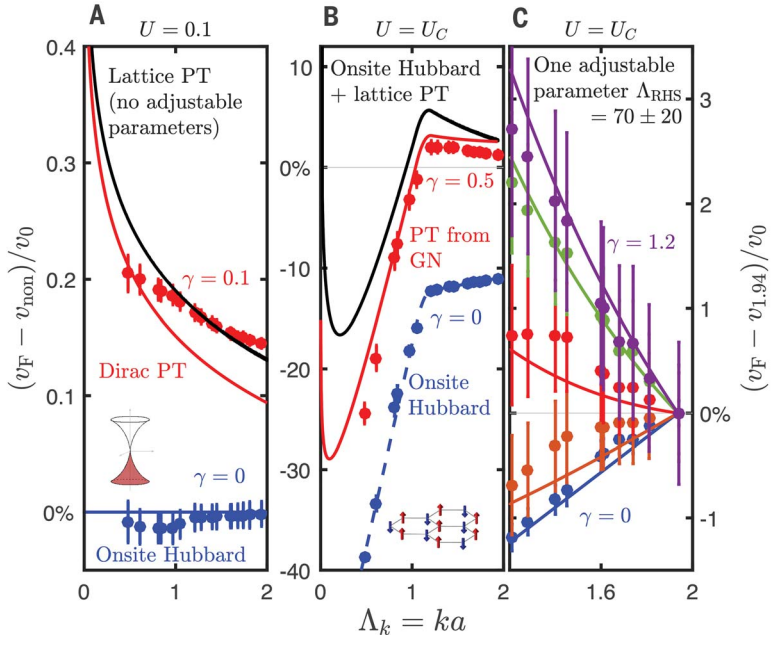


Fig. 3. Determining the renormalization group flow parameters from our quantum Monte Carlo data. Our simulations also provide data for momenta larger than Λ_s. We exploit this scale dependence to determine scales smaller than what we can simulate. (**A**) Representative data close to the weak-coupling fixed point where the on-site Hubbard model (blue data) shows no observable change in Fermi velocity. With long-range interactions, the Fermi velocity increases with decreasing momenta (red data), understood either by using a continuum perturbation theory (PT) (red curve) or lattice perturbation theory with no adjustable parameter (black curve), which diverge logarithmically. (**B**) The Gross-Neveu (GN) critical point is very different. Here, neither the Hubbard model (blue data) nor the data including the long-range Coulomb interaction (red data) show a logarithmic divergence at small Λ_k. A phenomenological fit captures the on-site Hubbard model (blue dashed line). The weak increase in renormalized Fermi velocity as one goes from $\Lambda_k = 2$ to $\Lambda_k = 1$ is seen in both the black curve (lattice perturbation theory) and the quantum Monte Carlo data with finite long-range interactions. (**C**) Logarithmic increase at $\Lambda_k \lesssim 2$. The red curve in (B) is obtained by fitting for this increase using a first-order perturbation theory about the Gross-Neveu critical point.

RPA functions $F_1(\alpha_0)$ and $F_0(\alpha_0)$ have been rederived several times in the literature (*4*, *5*, *16*, *29*). Our nonperturbative numerics verify this functional dependence on α_0. This agreement with the nontrivial functional dependence on α_0 gives us confidence to trust the logarithmic dependence on scale also predicted by the RPA calculation. Our full numerical dataset is consistent with this logarithmic increase in the weak-coupling limit [$P < 10^{-3}$; see figs. S8 and S9 in (*16*) for the full dataset and additional evidence].

Although there is currently no analytical theory for the dependence on U/U_c close to the Gross-Neveu fixed point, we can adequately describe the quantum Monte Carlo data by using a phenomenological model that has three parameters for the pure Hubbard model data and one additional parameter for linear dependence on $\gamma = 3\alpha_0/U$ of the Fermi velocity at the Gross-Neveu critical point $v_{GN}(\gamma)$. Defining the change in interaction strength from the critical value $(\varepsilon) = 1 - U/U_c(\alpha_0)$, we find (*16*) that $(v_F - v_{non})/v_0 = C_0 + C_1\varepsilon + C_2\varepsilon^2 + m_\gamma\gamma$ with $C_0 = -0.384 \pm 0.002$, $C_1 = 1.35 \pm 0.05$, and $C_2 = -1.2 \pm 0.1$ determined just from the Hubbard; the single additional parameter $m_\gamma = 0.333 \pm 0.005$ captures the effects of the long-range Coulomb potential. This fit is shown as the lines plotted on the right-hand side of Fig. 2. In section 5 of (*16*), we describe the suppression of the Fermi velocity at the Gross-Neveu fixed point using an approximate spin-wave theory with short-range interactions, giving the brown star in Fig. 2. Approaching the Gross-Neveu transition, antiferromagnetic fluctuations become progressively slower and couple to the Dirac fermions, thereby reducing the Fermi velocity (*30*).

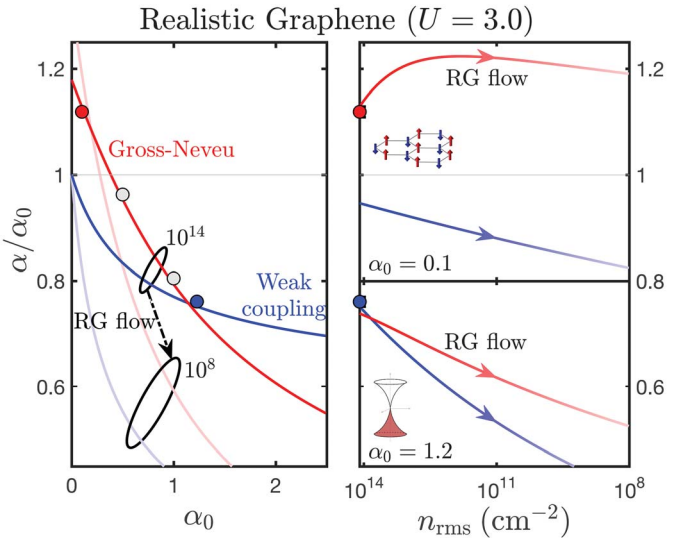


Fig. 4. Theoretical prediction for experimental realizations of graphene. Solid lines in the left panel (red, close to the Gross-Neveu fixed point; blue, close to the weak-coupling fixed point) show our theory for the interaction-induced change in the coupling constant at the energy scale of our numerics for realistic graphene with different α_0, as determined by the choice of substrate. Pale lines show our theory for the experimental energy scale that is numerically inaccessible but can be obtained from our numerical data by using the renormalization group (RG) analysis. Quantum Monte Carlo simulations (data points) have the same values in the left and right panels and can differentiate between these two theories. The dashed arrow shows this RG flow schematically from the numerical scale (small circle) to the experimental scale (larger circle). Right panels show the flow for $\alpha_0 = 0.1$ and 1.2. For the most common realization of graphene on a dielectric substrate (*16*), we predict weak suppression of the coupling constant that changes only slightly under renormalization (the solid and pale lines are not too different for $\alpha_0 \approx 0.6$). Most surprisingly, we predict that, for small α_0 (e.g., graphene on metal substrates), there is an enhancement of the coupling constant (i.e., solid red line is larger than unity) that then further increases with renormalization to the experimental scale (red line increases in the right panel).

Flow beyond numerical scales

The analysis so far has been at the simulation scale $\Lambda_s = 0.48$ that is limited by the largest system size that we can simulate. However, the experimental scale is set mostly by the degree of disorder in the system (currently, energy scales as small as $\Lambda_k = 10^{-3}$ can be measured). How, then, do we extrapolate our numerical findings to the experimental regime? Although we cannot numerically probe scales smaller than Λ_s, we can probe larger energy scales, thereby providing the inputs for a renormalization group flow from the numerical scale to the experimental scale. This procedure is illustrated in Fig. 3, which we now explain.

Figure 3A shows results for $U = 0.1$ in the weak-coupling regime. For the pure Hubbard model ($\alpha_0 = \gamma = 0$), there is no renormalization of the Fermi velocity, even at larger energy scales (blue circles). For small α_0 (shown in Fig. 3A as red data points for quantum Monte Carlo results, with $\gamma = 3\alpha_0/U = 0.1$ as a representative example), the values of the numerical data increase weakly with decreasing Λ_k. To show that this increase is consistent with a logarithm, we show two theoretical analyses. The red curve shows the first-order perturbation theory of a Dirac spectrum with the same single global parameter discussed already in Fig 2. We attribute the disagreement with the quantum Monte Carlo data to the fact that, at such large energy scales, the lattice model has nonlinear terms. To check this, we also solve numerically the first-order perturbation theory in α_0 on a finite lattice [see section 6 of (*16*)]. Because the lattice is specified, there is no adjustable parameter in this calculation. This is shown as the black curve in Fig. 3A, and it agrees with the quantum Monte Carlo at large energy scales (reduced $\chi^2 = 1.96$), giving us confidence in our analysis. In this method, we can go to lattice sizes as large as $L = 1500$, and, in the small Λ_k window, the lattice perturbation theory results look similar to the logarithmic divergence in the continuum perturbation theory. Additional numerical results would be needed to confirm that this observed logarithmic dependence in our quantum Monte Carlo data persists all the way to the Dirac point.

We now turn to the Gross-Neveu critical point. Figure 3B shows quantum Monte Carlo data for the Hubbard model (blue circles) and a representative example $\gamma = 0.5$ (red circles). For only contact interactions $\gamma = 0$, the renormalized Fermi velocity decreases with decreasing Λ_k, with a kink at $\Lambda_k \sim 1$. There is no available theory for this behavior. The blue dashed line shows a phenomenological three-parameter fitting function used to capture this numerical finding (*16*). Including the long-range component, the dominant effect is an overall upward shift in the curve (as already discussed in Fig. 2). However, at large Λ_k, we notice another subtle difference. The Hubbard model always shows the renormalized velocity to be a monotonically decreasing function with decreasing Λ_k, but with long-range interactions, the dependence is nonmonotonic. This observation suggests that the logarithmic increase in Fermi velocity persists into the strong-coupling regime, but the weak increase is masked by the strong decrease induced by the short-range Hubbard interaction. To check this, we obtain the black curve in Fig. 3B by adding the Hubbard results (blue dashed line in Fig. 3B) to the parameter-free lattice perturbation theory results (black line in Fig. 3A) shifted linearly with γ just like we did for the black lines on the right-hand side of Fig. 2, but here, instead of using the lowest-energy quantum Monte Carlo results to do this shift, we use the highest-energy data. Details of the fitting procedure are explained in section 8 of (*16*). The key insight here is that, although the effect of the logarithm is not observable at Λ_s, it can be extracted at larger values of Λ_k.

In Fig. 3B, the increase in the quantum Monte Carlo data (red circles) with decreasing Λ_k—associated with the infrared divergence—is

weaker than that predicted by the lattice perturbation theory (black curve). This weaker divergence is also anticipated by looking at the structure of the terms in a perturbation theory expansion about the Gross-Neveu fixed point [see eq. S32 of (*16*)]. These insights suggest fitting for the logarithm at large Λ_k. Figure 3C shows a blowup of the region around $\Lambda_k = 2$, where a single adjustable parameter $\Lambda_{\mathrm{RHS}} = 70 \pm 20$ captures this increase. The subscript RHS (right-hand side) indicates that the scale dependence of the Fermi velocity close to the Gross-Neveu critical point (or right-hand side of Fig. 2) is different from that at weak coupling, that is, Fig. 3A or the left-hand side of Fig. 2. As a representative example, the red data points and curve in Fig. 3B show our quantum Monte Carlo data and an analytical first-order perturbation theory at the Gross-Neveu critical point with the fit parameter Λ_k for the case of $\gamma = 0.5$ [other values of γ are shown in fig. S10 of (*16*)]. The agreement between the theory and quantum Monte Carlo data is good. Moving away from the Gross-Neveu critical line, we observe that for the Hubbard model data ($\gamma = 0$), the parameter C_1 discussed in association with Fig. 2 also gets a linear shift with energy scale. This introduces one additional parameter in the phenomenological model, that is, $C_1(\Lambda_k) = C_1(0.48) - (0.92 \pm 0.06)(\Lambda_k - 0.48)$, where the slope was determined also by looking at the full quantum Monte Carlo dataset. Therefore, combining this final observation for $C_1(\Lambda_k)$ with the fits already shown in Fig. 2 [i.e., dependence of Fermi velocity on $U/U_c(\alpha_0)$] and Fig. 3 (i.e., dependence on Λ_k), we can now extrapolate [see eq. S37 of (*16*)] our quantum Monte Carlo findings to any value of α_0, U, and Λ_k, thereby making predictions for realistic Dirac fermions at any experimental scale.

Implications for experiments

Our results on the role of electron-electron interactions apply to any Dirac fermion system for which one can define the short-range component of the Coulomb interaction U, its long-range tail α_0, and the experimental probe energy scale Λ_k. However, in Fig. 4, we use our results to make predictions for graphene, because the coupling constant for graphene Dirac fermions $\alpha_0 = e^2/(\kappa\hbar v_0)$ can be tuned by using substrates of different dielectric constants κ, with e and $\hbar$ being the elementary charge and Planck's constant h divided by 2π, respectively. The on-site potential for realistic graphene is about $U = 3.0$ (*16*). Our results apply only at half-filling. However, in real materials, carrier density inhomogeneity makes the Dirac point inaccessible experimentally (*31*). We use this degree of spatial inhomogeneity to define the experimental probe scale $\Lambda_k = \sqrt{\pi n_{\mathrm{rms}}}$ and use the root mean square (rms) of the carrier density n_{rms} in Fig. 4 (instead of Λ_k) for ease of comparison with experimental results. For the most typical situation of graphene on a substrate, typical measured Fermi velocities are 1.1×10^6 to 1.3×10^6 m/s (*12*); however, ab initio calculations predict 0.87×10^6 m/s [see, for example, (*32*)]. This notable discrepancy between theory and experiment has been largely unresolved in the literature—a notable exception is (*32*), where an ab initio density functional theory–GW calculation was used that agrees precisely with our result at our simulation scale, but their spectrum became unphysical for $n_{\mathrm{rms}} \lesssim 3 \times 10^{11}$ cm^{-2}. In our calculations, this interaction-caused enhancement of the Fermi velocity (or, equivalently, suppression in coupling constant) can be seen directly in the left panel of Fig. 4. For example, at $\alpha_0 = 1$ and typical experimental energy scale $n_{\mathrm{rms}} = 1 \times 10^{10}$ cm^{-2}, we predict $\alpha = 0.65 \pm 0.03$, corresponding to Fermi velocity of $(1.34 \pm 0.07) \times 10^6$ m/s [see fig. S11 in (*16*)].

We predict that for topological insulator Bi_2Se_3 ($\alpha_0 \approx 0.05$) or for graphene on metallic substrates, interactions enhance, rather than suppress, the coupling constant in the experimental window. This enhancement originates from the suppression of Fermi velocities in the Hubbard model at the Gross-Neveu fixed point. The phenomenological theory predicts that the coupling constant changes by less than 5% over the entire experimental window. Most surprisingly, for sufficiently small α_0, we predict that flowing Λ_k closer to the Dirac point will result in an increase in α rather than the expected decrease. These predictions challenge the conventional wisdom on the role of electron-electron interactions in 2D Dirac fermions. On the other hand, for suspended graphene, or quantum corral-like graphene, we expect the experiments to be close to the weak-coupling fixed point, and, therefore, the RPA should work well in this regime. The mechanism for this nonuniversal behavior is curious: The lattice-scale physics that depends on experimentally tunable parameters like strain and choice of substrate regularizes the Dirac theory not only in the ultraviolet, as expected, but also at the infrared, which was unexpected. All these predictions can readily be tested with current experimental capabilities.

REFERENCES AND NOTES

1. F. J. Dyson, *Phys. Rev.* **85**, 631–632 (1952).
2. E. Barnes, E. H. Hwang, R. E. Throckmorton, S. Das Sarma, *Phys. Rev. B* **89**, 235431 (2014).
3. J. González, F. Guinea, M. A. H. Vozmediano, *Nucl. Phys. B* **424**, 595–618 (1994).
4. J. González, F. Guinea, M. A. H. Vozmediano, *Phys. Rev. B* **59**, R2474–R2477 (1999).
5. J. Ye, S. Sachdev, *Phys. Rev. Lett.* **80**, 5409–5412 (1998).
6. V. N. Kotov, B. Uchoa, V. M. Pereira, F. Guinea, A. H. Castro Neto, *Rev. Mod. Phys.* **84**, 1067–1125 (2012).
7. I. F. Herbut, *Phys. Rev. Lett.* **97**, 146401 (2006).
8. F. Parisen Toldin, M. Hohenadler, F. F. Assaad, I. F. Herbut, *Phys. Rev. B* **91**, 165108 (2015).
9. Y. Otsuka, S. Yunoki, S. Sorella, *Phys. Rev. X* **6**, 011029 (2016).
10. S. Sorella, paper presented at the Conference on Interactions and Topology in Dirac Systems, Trieste, Italy, 5 Aug 2016; http://indico.ictp.it/event/7605/session/14/contribution/79.
11. A. Giuliani, V. Mastropietro, *Commun. Math. Phys.* **293**, 301–346 (2009).
12. S. Das Sarma, S. Adam, E. H. Hwang, E. Rossi, *Rev. Mod. Phys.* **83**, 407–470 (2011).
13. D. Hsieh *et al.*, *Nature* **452**, 970–974 (2008).
14. D. Kim *et al.*, *Nat. Phys.* **8**, 459–463 (2012).
15. K. K. Gomes, W. Mar, W. Ko, F. Guinea, H. C. Manoharan, *Nature* **483**, 306–310 (2012).
16. Materials and methods are available as supplementary materials.
17. D. C. Elias *et al.*, *Nat. Phys.* **7**, 701–704 (2011).
18. Z. Q. Li *et al.*, *Nat. Phys.* **4**, 532–535 (2008).
19. G. L. Yu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 3282–3286 (2013).
20. C. Hwang *et al.*, *Sci. Rep.* **2**, 590 (2012).
21. J. Chae *et al.*, *Phys. Rev. Lett.* **109**, 116802 (2012).
22. C. Faugeras *et al.*, *Phys. Rev. Lett.* **114**, 126804 (2015).
23. R. Brower, C. Rebbi, D. Schaich, in *Proceedings of Science–XXIX International Symposium on Lattice Field Theory (Lattice 2011)*, paper 056, Squaw Valley, Lake Tahoe, CA, 10 to 16 July 2011.
24. H.-K. Tang *et al.*, *Phys. Rev. Lett.* **115**, 186602 (2015).
25. M. Hohenadler, F. Parisen Toldin, I. F. Herbut, F. F. Assaad, *Phys. Rev. B* **90**, 085146 (2014).
26. C. Jang *et al.*, *Phys. Rev. Lett.* **101**, 146805 (2008).
27. V. Juričić, I. F. Herbut, G. W. Semenoff, *Phys. Rev. B* **80**, 081405 (2009).
28. M. Schüler, M. Rösner, T. O. Wehling, A. I. Lichtenstein, M. I. Katsnelson, *Phys. Rev. Lett.* **111**, 036601 (2013).
29. D. T. Son, *Phys. Rev. B* **75**, 235423 (2007).
30. B. Roy, V. Juričić, I. F. Herbut, *J. High Energy Phys.* **2016**, 18 (2016).
31. J. Xue *et al.*, *Nat. Mater.* **10**, 282–285 (2011).
32. P. E. Trevisanutto, C. Giorgetti, L. Reining, M. Ladisa, V. Olevano, *Phys. Rev. Lett.* **101**, 226405 (2008).
33. H.-K. Tang, Source data of Green's function and figure files. Figshare (2018); doi: 10.6084/m9.figshare.5131840.v2.
34. M. Bercx, F. Goth, J. S. Hofmann, F. F. Assaad, *SciPost Phys.* **3**, 013 (2017).

ACKNOWLEDGMENTS

We acknowledge allocation of computational resources at the CA2DM (Singapore) and the Gauss Centre for Supercomputing (SuperMUC at the Leibniz Supercomputing Center). **Funding:** This project was supported by the Singapore National Research Foundation (NRF-NRFF2012-01), Deutsche Forschungsgemeinschaft (SFB 1170 ToCoTronics, project C01), NSERC of Canada, and Singapore Ministry of Education (MOE2014-T2-1-112 and MOE2017-T2-1-130). **Author contributions:** S.A. and F.F.A. conceived this project. H.-K.T. and F.F.A. developed and optimized the quantum Monte Carlo codes. H.-K.T. and J.N.B.R. analyzed the data under the guidance of P.S. and F.F.A. J.N.L. developed the lattice perturbation theory and analytical results under the guidance of J.N.B.R., S.A., and I.F.H. All authors worked together to interpret and understand the results and to write the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The full collection of data developed in this work is available at https://figshare.com/articles/Source_data_of_Green_s_function_and_figure_files_/5131840 (*33*). For these calculations, we used a projective version of the auxiliary field quantum Monte Carlo approach. All the data presented in this article can be reproduced using the Algorithms for Lattice Fermions (ALF) open-source general implementation of the finite temperature auxiliary field quantum Monte Carlo available at https://alf.physik.uni-wuerzburg.de and documented in (*34*). There, detailed documentation on input-output and error analysis can be found. The ALF implementation allows one to simulate very general lattice models, including the long-range Coulomb repulsion. All the data presented in this paper can be reproduced with ALF by using a nearest-neighbor honeycomb lattice in the low-temperature limit.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/570/suppl/DC1
Materials and Methods
Figs. S1 to S11
Tables S1 and S2
References (*35*–*41*)

4 July 2017; accepted 25 May 2018
10.1126/science.aao2934

REPORT

THERMAL CONDUCTIVITY

Experimental observation of high thermal conductivity in boron arsenide

Joon Sang Kang, Man Li, Huan Wu, Huuduy Nguyen, Yongjie Hu*

Improving the thermal management of small-scale devices requires developing materials with high thermal conductivities. The semiconductor boron arsenide (BAs) is an attractive target because of ab initio calculation indicating that single crystals have an ultrahigh thermal conductivity. We synthesized BAs single crystals without detectable defects and measured a room-temperature thermal conductivity of 1300 watts per meter-kelvin. Our spectroscopy study, in conjunction with atomistic theory, reveals that the distinctive band structure of BAs allows for very long phonon mean free paths and strong high-order anharmonicity through the four-phonon process. The single-crystal BAs has better thermal conductivity than other metals and semiconductors. Our study establishes BAs as a benchmark material for thermal management applications and exemplifies the power of combining experiments and ab initio theory in new materials discovery.

The decreasing size of modern electronics makes heat dissipation one of the most critical technological challenges. The worldwide semiconductor industry, which has powered the information technology revolution since the 1960s, acknowledged in 2016 that Moore's law is nearing its end (*1*). A major issue is the enormous amount of waste heat generated during electronic device operation (*2*, *3*). For example, a U.S. data center devotes about 50% of its total electricity use to cooling (*4*). At the nanoscale, the power density of hot spots in current transistors is approaching that of the Sun's surface (*5*). Low thermal conductivity and heat dissipation rates severely degrade the performance and energy efficiency of electronic and photonic devices. Thermal management is arguably the biggest roadblock for next-generation devices, such as microprocessors and integrated circuits, light-emitting diodes, and high-power radio frequency devices, to name just a few (*1*, *5*).

Discovering high thermal conductivity (HTC) materials is needed to enable efficient heat dissipation from hot spots and improve device performance. So far, much of the research has been focused on carbon-based crystals—diamond, graphene, and carbon nanotubes. Although these materials can have exceptional heat transfer properties, there are several drawbacks for widespread use. Diamond, the most developed material for passive cooling of high-power electronics, suffers from high cost, slow synthesis rates, low quality, and challenging integration with semiconductors. Degradation of thermal conductivity plagues graphene and nanotubes when assembled into practical sizes, owing to ambient interactions and disorder scattering. Their intrinsic anisotropy creates other challenges for applications.

Fundamentally, understanding the origins of HTC remains a challenge. The conventionally accepted criteria for HTC materials are (i) small average atomic mass ($\bar{M}$); (ii) strong interatomic bonding; (iii) simple crystal structure; and (iv) low anharmonicity (*6*–*8*). Criteria (i) and (ii) imply a large Debye temperature (Θ_D) and provide the commonly used rule of thumb that thermal conductivity increases with decreasing $\bar{M}$ and increasing Θ_D. Diamond is the prototypical crystal. Diamond's two-atom primitive unit cell, light carbon mass, and stiff covalent bonding result in an exceptionally high value for thermal conductivity. Recent ab initio calculations show excellent agreement with the measured thermal conductivity of a wide range of materials (*8*–*16*), including silicon, diamond, graphene, and carbon nanotubes. Such calculations provide new physical insights into the nature of phonon thermal transport and the HTC mechanism.

Recent ab initio theoretical work indicates that the conventional criteria for HTC materials are incomplete and points to new ones stemming from fundamental vibrational properties that can lead to HTC (*8*, *16*–*18*). These new criteria applied to binary compounds are (i) a large mass ratio of constituent atoms; (ii) bunching together of the acoustic phonon branches; and (iii) an isotopically pure heavy atom. The large mass ratio provides a large frequency gap between acoustic and optical phonons (a-o gap). According to materials examined thus far, bunching of the acoustic phonon dispersions tends to occur in crystals with light constituent atoms, such as boron and carbon, where it derives from an unusual interatomic bonding that lacks core p electrons (*19*). Criteria (i) and (ii) contribute to unusually weak phonon-phonon scattering (*20*) and a large intrinsic thermal conductivity, whereas criteria (i) and (iii) cause relatively weak scattering of phonons by isotopes (*21*, *22*). The ab initio theory identified the III-V zinc-blende compound, defect-free boron arsenide (BAs), as having an exceptionally high thermal conductivity of more than 1000 W/m·K (*8*, *16*–*18*). This predicted HTC exceeds that of most state-of-the-art HTC materials and more than triples that of the current industrial HTC standard, i.e., silicon carbide. BAs possesses an advantageous combination of properties that incorporates both conventional (light boron mass and stiff, almost pure, covalent bonding) and new criteria [large arsenic-to-boron mass ratio, bunching together of its acoustic phonon branches, and isotopically pure As (heavy) atom] (*8*, *16*–*18*).

Experimental efforts to synthesize and characterize BAs have been scarce (*17*). Although the growth of cubic BAs was reported in the 1950s, its detailed structural characterization and properties were not reported (*23*–*28*). Generally, boron-related materials are notably difficult to obtain in dense bulk form (*29*). The synthesis of BAs is further complicated by the high volatility of arsenic and the introduced vacancy defects, as well as the possible formation of subphases (e.g., $B_{12}As_2$). In a collaborative effort, we reported the earliest thermal measurements on cubic BAs using the time-domain thermoreflectance technique (*30*). Our thermal conductivity value of 190 W/m·K measured in BAs samples with a high density of defects is far below the theoretical expectation. Later study made improvement, but the samples still show defects and grain boundaries that degraded the crystal quality and thermal properties (*31*). On the basis of subsequent analysis and calculation, defect scattering plays a dominant role in those samples, which makes probing the actual intrinsic thermal conductivity of BAs impossible in the absence of high-quality BAs crystals (*32*).

Here, we synthesized high-quality single-crystal BAs and measured an ultrahigh thermal conductivity of 1300 W/m·K in our BAs crystals. This value exceeds that of most HTC materials and is consistent with the ab initio prediction (*8*, *16*–*18*). We characterized our samples with scanning electron microscopy, Raman spectroscopy, powder x-ray diffraction (P-XRD), single-crystal x-ray diffraction (S-XRD), and high-resolution transmission electron microscopy (HRTEM). BAs has a zinc-blende face-centered cubic (fcc) crystal structure in the $F\bar{4}3m$ space group, where boron and arsenic atoms are interpenetrating and covalently bonded to form a tetrahedral geometry (Fig. 1A). The Raman spectroscopy data (Fig. 1C) clearly show two peaks, at 700 and 720 cm^{-1}, corresponding to the separate vibrational behaviors of two boron isotopes (^{10}B and ^{11}B) in their natural abundance, respectively. The P-XRD peaks that we observed (Fig. 1D) are in agreement with the zinc-blende fcc crystal structure. Our S-XRD confirmed the $F\bar{4}3m$ space group and the crystal quality, and was performed to unambiguously verify the single domain, single-crystalline

School of Engineering and Applied Science, University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA.
*Corresponding author. Email: yhu@seas.ucla.edu

nature over entire crystals (*33*). Each of the reflections in the x-ray diffraction pattern (Fig. 1E) appears clearly as a single dot without distortion, indicating that the sample has no grain boundaries. To collect a complete dataset of all reciprocal lattice points through the whole crystal, we rotated BAs samples over 360° under x-ray excitation and collected diffraction data at every 0.3° rotation. The reconstructed reciprocal lattice images from the S-XRD [see [100] plane in Fig. 1F and (*34*)] confirms the single-domain, single-crystalline zinc-blende fcc structure with a cubic lattice constant of 4.78 Å over the entire BAs sample. We used a focused ion beam to thin samples to ~100 nm for HRTEM (Fig. 1G). HRTEM images (Fig. 1H) clearly demonstrate the atomically resolved single-crystal lattice of our BAs sample. The reciprocal lattice peaks obtained from two-dimensional Fourier transforms of the lattice-resolved image (inset to Fig. 1H) were indexed in the zinc-blende structure with the zone axes along the [111] direction. The measured distance between each fringe is 1.69 Å, which is consistent with $(\bar{2}02)$ lattice spacing of BAs crystals (*23*) given the diffraction selection rules (*35*).

We characterized the thermal properties and phonon transport using ultrafast optical pump-probe spectroscopy based on the time-domain thermoreflectance (TDTR) technique (*15*, *30*, *36*). TDTR is well suited for the study as no physical contact is required with the sample and the measurement can provide high spatial resolution at the micrometer scale. We exposed our sample to a train of short pulses from a pump laser that creates a temperature rise at the sample surface (Fig. 2A). The transient temperature decay, caused by the heat impulse, was monitored with another probe pulse that is delayed in time with respect to the pump using a mechanical delay stage. The thermal conductivity was obtained by fitting the full transient decay curve, acquired by varying the time delay, to a thermal model (Fig. 2B). We measured a room-temperature thermal conductivity of our high-quality BAs crystals of ~1300 W/m·K. This value exceeds that of any known metal and semiconductor and is more than three times as high as that of industrial standards (copper and silicon carbide).

We simultaneously performed TDTR measurements on single-crystal BAs, diamond, and cubic boron nitride (BN) samples to ensure accuracy and provide direct comparison between the materials (*34*). We measured the temperature-dependent thermal conductivity of these samples from 300 to 600 K (Fig. 2C). Our thermal conductivity values for diamond (from 2200 to 1050 W/m·K for the temperature range) are consistent with values from the literature (*37*–*39*) and validate our approach. All samples had decreased thermal conductivities with increased temperature, indicating Umklapp scattering due to increasing phonon population. This also indicates a high single-crystal quality as phonon-defect scattering behavior is not observed. The thermal conductivity of BAs is almost twice that of cubic BN, making BAs the second-highest thermally conducting material among all known isotropic materials, exceeded only by diamond. As a semiconductor, BAs has a high potential for manufacturing integration and holds promise for thermal management applications.

Fundamentally, thermal transport in solids can be described by the interactions of phonons, i.e., the quantum-mechanical modes of lattice vibrations (*6*). Thermal conductivity results from phonon scattering processes that are closely related to the structure of materials. Phonon scattering usually includes Umklapp phonon-phonon scattering (τ_U), phonon-electron scattering ($\tau_{ph\text{-}e}$), mass fluctuation scattering (τ_M), and boundary scattering (τ_B) that can be characterized by the following relaxation rate $1/\tau$ (τ is the corresponding relaxation time):

$$\frac{1}{\tau} = \frac{1}{\tau_U} + \frac{1}{\tau_{ph\text{-}e}} + \frac{1}{\tau_M} + \frac{1}{\tau_B} \quad (1)$$

Previous research indicates that the last two terms of this equation—τ_M and τ_B—play important roles in the thermal conductivity of crystals with defects and grain boundaries (*32*). $\tau_{ph\text{-}e}$ is non-negligible for doped semiconductors, or metals with sufficiently high electron density. Notably, the first term of the equation—τ_U—describes the thermal conductivity of perfect single crystals without any defects, and therefore is expected to be the dominant transport mechanism in our high-quality BAs single crystals. However, a persistent fundamental question regarding the high-order phonon anharmonicity, in particular for BAs, remains in the field of atomistic phonon theory. For many decades, thermal transport in solids

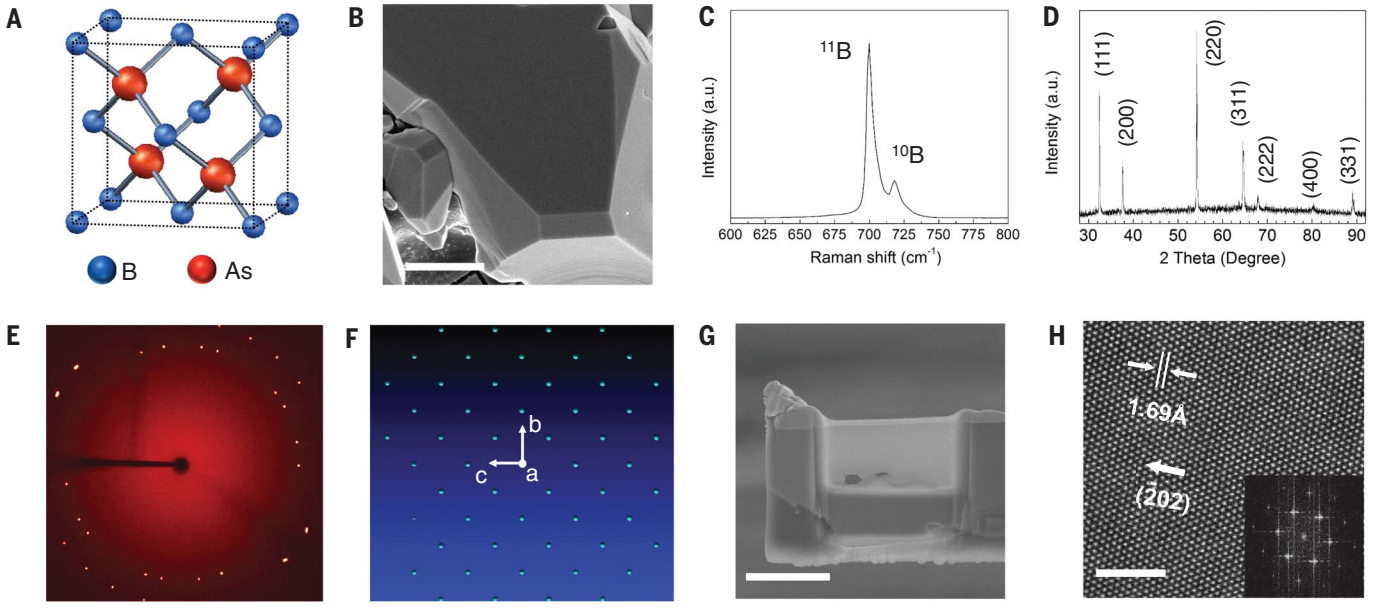


Fig. 1. Structural characterizations of single-crystal BAs. (**A**) Schematic of the zinc-blende crystal structure of cubic BAs, resembling that of diamond. (**B**) Scanning electron microscope (SEM) image of BAs. Scale bar: 5 μm. (**C**) Raman spectra of BAs crystals. (**D**) Powder x-ray diffraction measurements. (**E**) Single-crystal x-ray diffraction image of BAs. (**F**) Constructed reciprocal lattice of BAs from the complete dataset of single-crystal x-ray diffraction measurements, representing a clear single-crystal reciprocal space over the entire crystal. The lattice constant was measured as 4.78 Å for cubic BAs. (**G**) SEM image of a BAs sample thin slice (~100 nm) prepared by focused ion beam for HRTEM studies. Scale bar: 3 μm. (**H**) HRTEM image of BAs showing atomically resolved lattices. Inset: Two-dimensional Fourier transforms of the image depicting the [111] zone axes of BAs; the arrow indicates the crystal direction of $(\bar{2}02)$. Scale bar: 2 nm.

was considered to be governed by the three-phonon scattering process (*6*, *40*), and the effects of four-phonon and higher-order scattering processes were believed to be negligible. The three-phonon scattering was deployed in the initial ab initio calculations of the BAs thermal conductivity (*8*), but according to recent theory (*16*, *41*), the four-phonon scattering is important in certain materials such as BAs. The two density functional theory (DFT)–based calculations predict different thermal conductivity values for BAs. We compared our experimental measurements at different temperatures with two DFT theory predictions (Fig. 2D). In addition, to compare and exclude the effect of point defects, we performed three-phonon DFT calculation with vacancies (*34*) and included it in the figure. We found that our experimental results are in good agreement with the four-phonon DFT calculation, verifying that the Umklapp phonon scattering dominates phonon interactions. Moreover, our experiments indicate that, unlike in most materials, the probability of higher-order phonon scattering in BAs is important and cannot be ignored even around room temperature. The high-order anharmoniciy that we observed for BAs is due to its distinctive band structure (i.e., large a-o gap and large mass ratio) that leads to a relatively weak three-phonon process and provides sufficient numbers of possible four-phonon scattering configurations.

Another important phonon mechanism for HTC in BAs is the unique phonon mean free path (MFP) spectra that come from its extraordinary phonon band structure. MFPs represent the characteristic lengths corresponding to the distance over which heat carriers transmit thermal energy before being scattered; in general, MFPs can span several orders of magnitude—usually from ~1 nm to ~100 μm. According to the ab initio theory, BAs has a large phononic band gap between acoustic and optical phonon branches, which minimizes acoustic-optical phonon scattering and leads to MFPs longer than those of most common materials. In particular, BAs phonons with MFPs longer than 1 μm contribute to more than 90% of its total thermal conductivity (*18*). By comparison, for diamond and cubic BN, phonons with MFPs longer than 1 μm only contribute to about 30% of their total thermal conductivity. To gain further insight and verify the phonon MFP distribution in BAs, we probed the phonon spectral contribution by exploiting the size-dependent ballistic transport (*15*). In essence, thermal conductivity represents the spectral contribution from many different phonon modes and can be quantified using the cumulative thermal conductivity (*42*) expressed as

$$\kappa(\Lambda_m) = -\int_0^{\Lambda_M} \frac{1}{3} C \cdot v \cdot \Lambda \left(\frac{d\Lambda}{d\omega}\right)^{-1} d\Lambda \quad (2)$$

where C, v, and Λ are the phonon mode (ω)–dependent heat capacity, group velocity, and phonon MFP, respectively. The cumulative thermal conductivity—$\kappa(\Lambda_m)$—represents the contribution to the total thermal conductivity from all phonons with MFPs shorter than a certain value, Λ_m. This quantitative spectral information projects the contributions to thermal conductivity into characteristic length scales and is key to understanding thermal properties in connection with atomistic phonon theory (*42*).

Experimentally, this phonon MFP distribution information can be investigated by exploiting the ballistic (or nonequilibrium) heat conduction around a small heating area, with the physical concept described in our recently developed thermal spectral mapping spectroscopy technique (Fig. 3A) (*15*, *36*). In this measurement, hot phonons travel from the heated area into the underlying substrate material. The heat transfer regime is controlled by a characteristic thermal length, which is proportional to the heating size D, in comparison to phonon MFPs. In the diffusive limit, when $D >> \Lambda$, propagating phonons experience enough scattering to reach local thermal equilibrium. In this case, Fourier's law accurately describes the transport and heat flux, and the thermal conductivity of materials is simply the bulk value (κ_{bulk}). From the kinetic theory (*6*), the contribution to the total thermal conductivity from a specific phonon mode is $\kappa_{\omega,bulk} = \frac{1}{3} C_\omega \cdot v_\omega \cdot \Lambda_\omega$. As the heater size D decreases, phonons with $D < \Lambda$ will have fewer opportunities to scatter. In the ballistic limit ($D << \Lambda$), phonons propagate analogously to the thermal radiation over the whole region with a characteristic length of ~D. Therefore, the actual heat flux will deviate from the Fourier's law prediction for the quasi-ballistic, or ballistic, regimes (Fig. 3A). Mathematically, the actual heat flux is measured and fitted to Fourier's law to obtain an effective thermal conductivity (κ_{eff}), which gradually decreases in value for smaller Ds, as a higher portion of phonons evolves from diffusive to ballistic transport. The decrease in κ_{eff} (D) represents MFP spectra and should follow the same trend in $\kappa(\Lambda_m)$. Therefore, the size-dependent thermal conductivity measurement can provide a fundamental understanding of such MFP spectra, although the exact relationship between κ_{eff} and D requires careful atomistic and multiscale simulations and is a function of the heating geometry and materials (*15*, *36*).

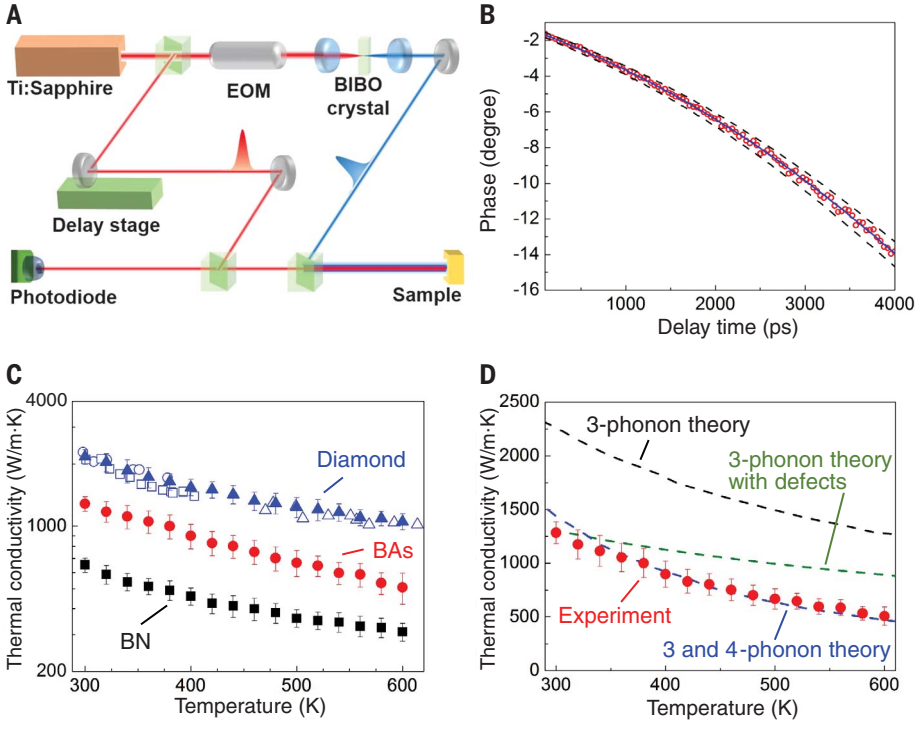


Fig. 2. Temperature-dependent thermal transport measurements. (**A**) Schematic of the setup for ultrafast pump-probe spectroscopy via the time-domain thermoreflectance (TDTR) technique. (**B**) Typical TDTR data: thermal reflectance phase signal versus time (red circles), fitted to the thermal transport model (blue line). Calculated curves (black dashed lines) with the thermal conductivity changed by ±10% of the best values to illustrate measurement accuracy. (**C**) Temperature-dependent (300 to 600 K) thermal conductivity of the three materials with the highest thermal conductivities: diamond, BAs, and boron nitride (BN). Black squares, red circles and blue triangles indicate cubic BN, BAs, and diamond, respectively. Also shown are literature data for diamond [open blue circles (*38*), squares (*37*), and triangles (*39*)]. (**D**) Experimentally measured thermal conductivity of BAs in comparison to ab initio predictions by DFT, considering the three-phonon scattering process (black dashed line) (*8*), the three- and four-phonon scattering process (blue dashed line) (*16*), and the three-phonon scattering process with point defects (green dashed line) (*34*).

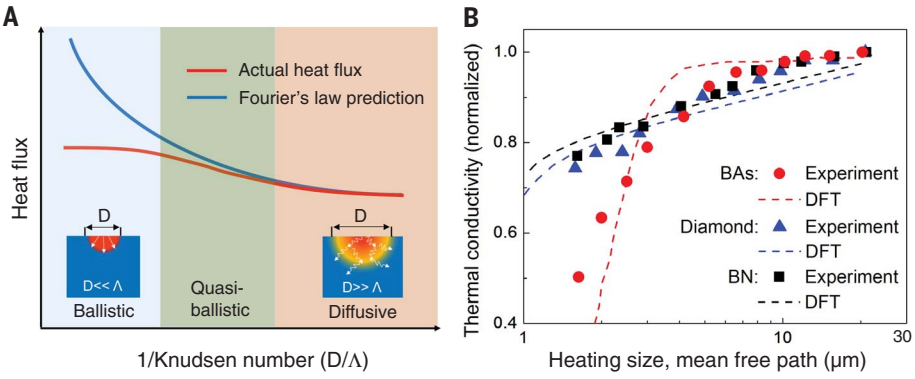


Fig. 3. Probing phonon mean free path spectra of BAs through size-dependent ballistic transport. (**A**) Schematic of heat flux for a fixed temperature difference as a function of the Knudsen number. Red and blue lines indicate the actual heat flux and the flux predicted by Fourier's law, respectively. Insets show that the thermal transport evolves from a diffusive to a ballistic regime when the heating size is gradually reduced. When the actual heat flux was measured and fitted to Fourier's law to obtain an effective thermal conductivity (κ_{eff}), a gradual reduction in κ_{eff} was expected with a decreasing heating size. Essentially, the κ_{eff} decrease is due to the evolution from diffusive to ballistic transport for the phonons with a mean free path comparable to the heating size, and thereby represents the phonon MFP spectra. (**B**) The normalized effective thermal conductivity was measured for the three best thermal conductors (BAs, diamond, and BN), as a function of heating diameters from 21 to 1.6 μm. Experimental results, compared with the MFP spectra calculated with DFT (*8*, *16*, *18*), indicate that in BAs, a large portion of phonons have long mean free paths, owing to the distinctive band structure of BAs.

In our experiments, the laser beam served as the heater, and we measured the effective thermal conductivity of materials by varying the laser heating size from 21 to 1.6 μm. The thermal conductivity of BAs, diamond, and cubic BN were measured simultaneously as a function of the heating sizes (Fig. 3B). The results clearly show that ballistic transport occurs in all three materials, but that their thermal conductivity decreases differ. For diamond and cubic BN, κ_{eff} decreases by ~26 and 23%, respectively, when reducing the laser heating size from 21 to 1.6 μm. By comparison, under the same conditions, a strong thermal conductivity reduction of ~50% is observed for BAs, indicating that long MFP phonons contribute substantially more to heat conduction in BAs crystals. Our measurements (Fig. 3B) are consistent with the DFT-predicted results for these materials (*8*, *16*, *18*). Phonons with long phonon MFPs (1 to 10 μm) contribute to a very high portion of BAs's total thermal conductivity (>50%). By comparison, in most known materials, phonon MFPs are distributed over a wider range (1 nm to 100 μm). Thus, we demonstrated that the ultrahigh thermal conductivity of BAs originates from the enhanced MFPs resulting from its distinctive phonon band structure.

We experimentally observed an ultrahigh thermal conductivity of 1300 W/m·K at room temperature in synthetic high-quality single-crystal BAs. Our study verifies the prediction from ab initio theory and establishes BAs as a benchmark material with the highest isotropic thermal conductivity of bulk metals and semiconductors. Our temperature-dependent data suggest that, unlike most common materials, high-order phonon anharmonicity strongly affects heat conduction in BAs. Furthermore, the phonon spectral contribution to thermal conductivity, investigated by exploiting the ballistic thermal transport, shows that compared to most materials, in BAs long phonon MFPs contribute to a substantially higher portion of the thermal conductivity. The investigation of fundamental transport mechanisms represents an important breakthrough in advancing the experiment-theory synergy for the rational design of new materials. Looking to the future, the ultrahigh thermal conductivity of BAs, together with its semiconducting nature and manufacturing integration, could revolutionize the current technological paradigms of thermal management and possibly extend the roadmap for high-power electronics.

REFERENCES AND NOTES

1. M. M. Waldrop, *Nature* **530**, 144–147 (2016).
2. P. Ball, *Nature* **492**, 174–176 (2012).
3. S. Chu, A. Majumdar, *Nature* **488**, 294–303 (2012).
4. W. L. Arman Shehabi, Sarah Smith, Dale Sartor, Richard Brown, Magnus Herrlin, Jonathan Koomey, Eric Masanet, Nathaniel Horner, Inês Azevedo, United States Data Center Energy Usage Report (2016); https://ses.lbl.gov/publications/united-states-data-center-energy.
5. International Technology Roadmap for Semiconductors, ITRS 2.0 2015 Edition; www.itrs2.net/.
6. J. M. Ziman, *Electrons and Phonons: The Theory of Transport Phenomena in Solids. Oxford Classic Texts in the Physical Sciences* (Oxford Univ. Press, 1960).
7. G. A. Slack, *J. Phys. Chem. Solids* **34**, 321–335 (1973).
8. L. Lindsay, D. A. Broido, T. L. Reinecke, *Phys. Rev. Lett.* **111**, 025901 (2013).
9. N. Mingo, D. A. Broido, *Phys. Rev. Lett.* **93**, 246106 (2004).
10. N. Mingo, D. A. Broido, *Phys. Rev. Lett.* **95**, 096105 (2005).
11. A. Ward, D. A. Broido, D. A. Stewart, G. Deinzer, *Phys. Rev. B* **80**, 125203 (2009).
12. K. Esfarjani, G. Chen, H. T. Stokes, *Phys. Rev. B* **84**, 85204 (2011).
13. L. Lindsay, D. A. Broido, T. L. Reinecke, *Phys. Rev. Lett.* **109**, 095901 (2012).
14. M. N. Luckyanova *et al.*, *Science* **338**, 936–939 (2012).
15. Y. Hu, L. Zeng, A. J. Minnich, M. S. Dresselhaus, G. Chen, *Nat. Nanotechnol.* **10**, 701–706 (2015).
16. T. Feng, L. Lindsay, X. Ruan, *Phys. Rev. B* **96**, 161201 (2017).
17. C. Uher, *Physics (College Park Md.)* **6**, 76 (2013).
18. D. A. Broido, L. Lindsay, T. L. Reinecke, *Phys. Rev. B* **88**, 214303 (2013).
19. R. M. Wentzcovitch, M. L. Cohen, P. K. Lam, *Phys. Rev. B Condens. Matter* **36**, 6058–6068 (1987).
20. M. Lax, P. Hu, V. Narayanamurti, *Phys. Rev. B* **23**, 3095–3097 (1981).
21. L. Lindsay, D. A. Broido, T. L. Reinecke, *Phys. Rev. B* **88**, 144306 (2013).
22. S. Tamura, *Phys. Rev. B* **30**, 849–854 (1984).
23. J. A. Perri, S. Laplaca, B. Post, *Acta Crystallogr.* **11**, 310 (1958).
24. F. V. Williams, R. A. Ruehrwein, *J. Am. Chem. Soc.* **82**, 1330–1332 (1960).
25. J. Osugi, K. Shimizu, Y. Tanaka, *Proc. Jpn. Acad.* **42**, 48 (1966).
26. S. M. Ku, *J. Electrochem. Soc.* **113**, 813 (1966).
27. A. F. Armington, *J. Cryst. Growth* **1**, 47–48 (1967).
28. T. L. Chu, A. E. Hyslop, *J. Appl. Phys.* **43**, 276–279 (1972).
29. P. Rudolph, Ed., *Handbook of Crystal Growth, Vol. II, Bulk Crystal Growth* (Elsevier, 2015); http://store.elsevier.com/Handbook-of-Crystal-Growth/isbn-9780444633033/).
30. B. Lv *et al.*, *Appl. Phys. Lett.* **106**, 074105 (2015).
31. F. Tian *et al.*, *Appl. Phys. Lett.* **112**, 031903 (2018).
32. N. H. Protik, J. Carrete, N. A. Katcho, N. Mingo, D. Broido, *Phys. Rev. B* **94**, 045207 (2016).
33. B. E. Warren, *X-ray Diffraction* (Addison-Wesley, MA, 1969).
34. Materials and methods are available as supplementary materials.
35. D. B. Williams, C. B. Carter, *Transmission Electron Microscopy: A Textbook for Materials Science* (Springer, Boston, MA, 2009).
36. J. S. Kang, H. Wu, Y. Hu, *Nano Lett.* **17**, 7507–7514 (2017).
37. D. G. Onn, A. Witek, Y. Z. Qiu, T. R. Anthony, W. F. Banholzer, *Phys. Rev. Lett.* **68**, 2806–2809 (1992).
38. L. Wei, P. K. Kuo, R. L. Thomas, T. R. Anthony, W. F. Banholzer, *Phys. Rev. Lett.* **70**, 3764–3767 (1993).
39. J. R. Olson *et al.*, *Phys. Rev. B Condens. Matter* **47**, 14850–14856 (1993).
40. A. Maradudin, A. Fein, *Phys. Rev.* **128**, 2589–2608 (1962).
41. L. Lindsay, D. A. Broido, *J. Phys. Condens. Matter* **20**, 165209 (2008).
42. F. Yang, C. Dames, *Phys. Rev. B* **87**, 035437 (2013).

ACKNOWLEDGMENTS

We thank H. Albrecht for editorial help, T. Aoki from UCI Materials Research Institute for help with TEM study, and the support from UCLA CNSI and Nanolab. **Funding:** Y.H. acknowledges support from a CAREER award from the EPM and TTP programs of the National Science Foundation under grant 1753393, a Young Investigator Award from the United States Air Force Office of Scientific Research under grant FA9550-17-1-0149, a PRF Doctoral New Investigator Award from the American Chemical Society under grant 58206-DNI5, the UCLA Sustainable LA Grand Challenge, and the Anthony and Jeanne Pritzker Family Foundation. **Author contributions:** Y.H. proposed and directed the research. Y.H. and J.S.K. designed the experiment; J.S.K., M.L., and H.N. performed the experiment; H.W. performed the ab initio calculations. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/575/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S3
References (*43*–*58*)

12 March 2018; accepted 21 June 2018
Published online 5 July 2018
10.1126/science.aat5522

THERMAL CONDUCTIVITY

High thermal conductivity in cubic boron arsenide crystals

Sheng Li[1]*, Qiye Zheng[2]*, Yinchuan Lv[3], Xiaoyuan Liu[1], Xiqu Wang[4], Pinshane Y. Huang[2], David G. Cahill[2]†, Bing Lv[1]†

The high density of heat generated in power electronics and optoelectronic devices is a critical bottleneck in their application. New materials with high thermal conductivity are needed to effectively dissipate heat and thereby enable enhanced performance of power controls, solid-state lighting, communication, and security systems. We report the experimental discovery of high thermal conductivity at room temperature in cubic boron arsenide (BAs) grown through a modified chemical vapor transport technique. The thermal conductivity of BAs, 1000 ± 90 watts per meter per kelvin meter-kelvin, is higher than that of silicon carbide by a factor of 3 and is surpassed only by diamond and the basal-plane value of graphite. This work shows that BAs represents a class of ultrahigh–thermal conductivity materials predicted by a recent theory, and that it may constitute a useful thermal management material for high–power density electronic devices.

Thermal management is a major challenge for a wide variety of engineering systems (*1*–*3*). In particular, effective thermal management often plays a determining role in the performance and reliability of high-power electronic and optoelectronic devices. A common component in a thermal management system is a high–thermal conductivity (Λ) material that can effectively conduct heat from a small high-power device to a larger heat exchanger that then dissipates heat to a working fluid. Diamond has the highest isotropic Λ of any bulk material (the room-temperature value is ~2200 W/m·K) (*4*) and is sometimes used in demanding heat dissipation applications. Heat spreaders constructed from single-crystal diamond, however, are costly and can be produced in only limited sizes. The thermal conductivity of more cost-effective diamond films produced by chemical vapor deposition (CVD) is often substantially compromised by microstructural defects. Another material with a predicted Λ = 940 W/m·K at room temperature, cubic boron nitride (c-BN) (*5*), requires high-pressure, high-temperature synthesis conditions of 10 GPa and >2000°C.

Cubic boron arsenide (BAs) is predicted to have ultrahigh thermal conductivity comparable to that of diamond (*5*). The main challenge in realizing the potential of this material is synthesizing high-quality crystals, as defects and impurities degrade the thermal properties. However, BAs is an appealing candidate for thermal management for high–power density devices. BAs is chemically inert, with a coefficient of thermal expansion (3.0×10^{-6} K^{-1}) similar to that of Si (2.6×10^{-6} K^{-1}) (*6*). BAs is also growth-compatible with GaN and GaAs (*7*), making it ideal for a variety of applications in the thermal management of high–power density devices.

BAs adopts the cubic zincblende structure and is a semiconductor with band gap of ~1.5 eV. We synthesized single crystals of BAs by a modified chemical vapor transport method using high-purity source materials (B and As powders) and transport agents (*8*). Among the different vapor transport agents we investigated (I_2, H_2, Br_2, and NH_4I), we found that NH_4I yielded the highest–thermal conductivity crystals. We optimized the ratio of starting materials and temperature profiles for our synthesis [see (*9*) and figs. S1 to S4 for the details of the synthetic approach and characterization of microstructure]. Growth runs and individual specimens are labeled with Roman numerals and lower-case letters, respectively (e.g., **Ia**). This helped us track variations between and within growth runs.

We synthesized single crystals of cubic BAs and characterized the crystals with scanning electron microscopy (SEM) (Fig. 1A), Raman spectroscopy (Fig. 1B), and single-crystal x-ray diffraction (Fig. 1C). SEM showed that specimen **Ia** has a lateral dimension of ~0.5 mm and a large crystal facet parallel to the {111} crystal planes. Raman spectroscopy and single-crystal x-ray diffraction confirmed the zincblende structure and high crystalline quality of the crystals. The strong Raman peak we labeled as P1 at 698 cm^{-1} (Fig. 1B, inset), with its overtone at 1398 cm^{-1}, corresponds to the T_2 longitudinal optical phonon. Because isotope disorder typically does not generate a localized phonon mode (*10*), we attributed the P2 peak at 717 cm^{-1} to scattering from the high phonon density of states at the L or W point in the Brillouin zone (*11*) that is induced by isotope disorder. In other words, isotope disorder partially relaxes the crystal momentum conservation rules and allows the peak corresponding to zone boundary phonons to appear in the one-phonon spectra. This assignment is consistent with prior studies of the Raman spectra of Si, Ge, and diamond (*12*–*14*).

We found high-quality single-domain crystals by means of single-crystal x-ray diffraction (XRD) on the selected crystals. We reconstructed a pseudo-precession photograph for the a^*-b^* plane pixel-by-pixel from complete sets of single-crystal XRD images (2178 frames) that we measured with a scan width of Δω = 0.30° (Fig. 1C). We reconstructed precession photographs for other reciprocal lattice planes as well (fig. S2). The pseudo-precession photographs provide an undistorted image of layers of reciprocal space and a direct view of crystal quality (*9*). All of the detected reflections (circled spots) are derived from a single lattice; we did not observe any evidence for crystal twinning in samples grown using the conditions of specimen I. We observed twinning in crystals grown under different conditions (e.g., specimens **IIb** and **IIIb**) (figs. S3 and S4) (*9*). We further used selected-area electron diffraction, diffraction-filtered TEM, and annular dark-field scanning transmission electron microscopy (ADF-STEM) to look for large-scale structural defects (Fig. 1D and fig. S11). We did not detect grain boundaries or twinning in our high-quality single crystals, consistent with single-crystal XRD results. Because of damage caused by a focused ion beam (FIB) during sample preparation and electron beam damage during imaging, we cannot make any conclusions about the potential presence of point defects in our samples.

We measured the thermal conductivity of our crystals by time-domain thermoreflectance (TDTR) (*15*–*17*) on smooth facets of ~100 μm. TDTR is a pump-probe optical metrology with a spatial resolution on the order of 10 μm, allowing for measurements on smooth-faceted crystals of less than 1 mm^3 (*15*, *18*). We determined the thermal conductivity of BAs by comparing the time dependence of the ratio of the in-phase (V_{in}) and out-of-phase (V_{out}) thermoreflectance signals from the lock-in amplifier with a thermal diffusion model (*15*). Determining Λ from TDTR data requires a model fit (*9*), for which we present the corresponding result for our highest-Λ samples (Fig. 2A). We used $1/e^2$ intensity radii of 10.4 μm for both the pump and probe laser beams to collect the data (Fig. 2, A and B). The thermal conductivity of BAs (Λ) and the thermal conductance of the interface between the transducer layer (~80 nm Al) and BAs (G) are the two free parameters in the fit. The best fit of our experimental data for sample **Ia** gives Λ = 1000 ± 90 W W/m·K and G = 130 ± 10 MW m^{-2} K^{-1} (see fig. S7). We show our measurement sensitivity by fixing Λ at either 890 or 1090 W/m·K, and allowing G to vary (Fig. 2A).

We measured the temperature-dependent thermal conductivity of BAs crystal specimens **Ia**, **IIa**, and **IIIa** between 300 and 600 K (Fig. 2B). The 300 K Λ values are ~1000 W/m·K for **Ia**, ~870 W/m·K for **Ig**, ~850 W/m·K for **IIa**, and ~500 W/m·K for **IIIa**, respectively, which are much higher than the previously reported Λ

[1]Department of Physics, University of Texas at Dallas, Richardson, TX 75080, USA. [2]Department of Materials Science and Engineering and Materials Research Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [3]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [4]Department of Chemistry, University of Houston, Houston, TX 77204, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: d-cahill@illinois.edu (D.G.C.); blv@utdallas.edu (B.L.)

values of ~350 and ~200 W/m·K (*8*, *19*, *20*) (see table S1 for other specimens). The high Λ value of the best BAs crystal (**Ia**) also exceeds that of c-BN (~900 W/m·K), copper (~400 W/m·K) and 4H SiC (~320 W/m·K along the *c* axis). The thermal conductivity of BAs specimen **Ia** is lower than that of diamond by a factor of 2 (~2200 W/m·K) and graphite (in-plane direction, ~2000 W/m·K) (fig. S8) (*4*, *21*–*24*). BAs has the second highest measured isotropic thermal conductivity among the known single-phase bulk metal and semiconducting materials.

We need to consider deviations from Fourier's law (*25*, *26*) in TDTR measurements of high–thermal conductivity crystals. Such deviations result from ballistic phonon transport in the sample and from a mismatch in the distribution of phonons that carry heat across the metal transducer–sample interface and the distribution of phonons that carry heat in the sample. A reduction in the characteristic length scales of the temperature gradient by decreasing the laser spot size (w_0) or increasing the pump modulation frequency (f) increases the percentage of low-frequency phonons with long mean free paths that are not in local equilibrium with high-frequency phonons. Such effects cause the apparent thermal conductivity (Λ_A) derived from the thermal model using a small characteristic length scale to be less than the apparent thermal conductivity derived from the thermal model using a larger characteristic length scale (*25*).

We measured Λ_A as a function of spot sizes w_0 in TDTR for specimen **Ia** (Fig. 2C). As a result of ballistic phonon transport, Λ_A drops by approximately 30% as w_0 decreases from 26.5 to 2.7 μm at 300 K. The difference between Λ_A measured with w_0 = 10.4 μm and with w_0 = 26.5 μm is ~7%, comparable to the typical uncertainty of TDTR measurements of thermal conductivity. We did not observe modulation frequency dependence in Λ_A as we varied f from 1.1 to 9.3 MHz. This suggested negligible interfacial non-equilibrium effects (*25*). We used w_0 = 10.4 μm and f = 9.3 MHz in the measurements plotted in Fig. 2B to maintain an improved signal-to-noise ratio. We expect the spot size dependence of Λ_A to be smaller at elevated temperatures where phonon-phonon scattering becomes stronger and phonon mean free paths decrease.

We compared our measurements to the theoretical prediction from first-principles calculations based on three-phonon (3ph) scattering (*5*) and both three- and four-phonon (3ph + 4ph) scattering (Fig. 2B). Previous studies attributed the potential for high thermal conductivity of BAs to a large frequency gap between acoustic and optic phonons, in combination with the bunching of the acoustic phonon dispersions. The large acoustic-optic gap greatly limits the three-phonon scattering rate involving one optical and two acoustic phonons (*5*). However, such a gap does not forbid four-phonon scattering processes between the acoustic and optical phonons, and a calculation found that four-phonon processes substantially reduced the predicted Λ at room temperature from 2200 to 1400 W/m·K (*27*). At elevated temperatures when intrinsic phonon-phonon scattering becomes important (*28*), our measured values of Λ for **Ia** are comparable to the prediction based on 3ph + 4ph scattering. The measured thermal conductivity of **Ia** scales as $\Lambda \sim T^{-1.4}$ in the temperature range 300 to 600 K, close to the calculation including four-phonon scattering (approximately $\Lambda \sim T^{-1.6}$) and much stronger than the original calculation based on only three phonon processes (approximately $\Lambda \sim T^{-0.8}$).

Crystal imperfections and free carriers could both affect phonon thermal transport in semiconductors (*28*). Because the grain size determined by single-crystal XRD is much larger than the TDTR characteristic temperature length scale (*9*), and because we do not observe a high density of stacking faults in TEM, phonon scattering by boundaries should be negligible. On the other hand, we found that specimens with high hole concentrations of 1.9×10^{19} cm^{-3} (**Ie**) and 1.1×10^{20} cm^{-3} (**Id**) according to Hall measurements (fig. S10) have low Λ values of 130 and 110 W/m·K, respectively, much smaller than our best sample.

Relative to the Raman spectra of the best specimen of **Ia**, specimens with high hole concentration such as **Id** (Fig. 3A) show asymmetric one-phonon P1 and P2 phonon peaks and stronger background intensity resulting from Fano interference (i.e., the coherent interaction between a continuum of electronic excitations and a discrete optical phonon) (*29*, *30*). To study the relation between the Raman spectra and Λ, and to avoid fitting for overlapping P1 and P2 peaks, we chose to use the integrated Raman intensity from 1050 to 1150 cm^{-1} (A_{Raman}) where the main contribution is from electronic Raman scattering (*31*). The Λ values of BAs samples grown from several batches show clear correlation with the corresponding A_{Raman} values (Fig. 3B), and samples with high carrier concentration (**Id** and **Ie**) indeed give large integrated Raman intensity and low Λ. In a single specimen with smooth facets, Λ and Raman spectra intensity are relatively uniform, with variations around 10% (see Λ and A_{Raman} map in figs. S9 and S11). Free carriers, presumably resulting from impurities or vacancies, thus may reduce the thermal conductivity in each BAs sample even in our best specimens (*32*). We cannot, however, rule out the contribution from other crystal defects (e.g., point defects and clusters of point defects) in

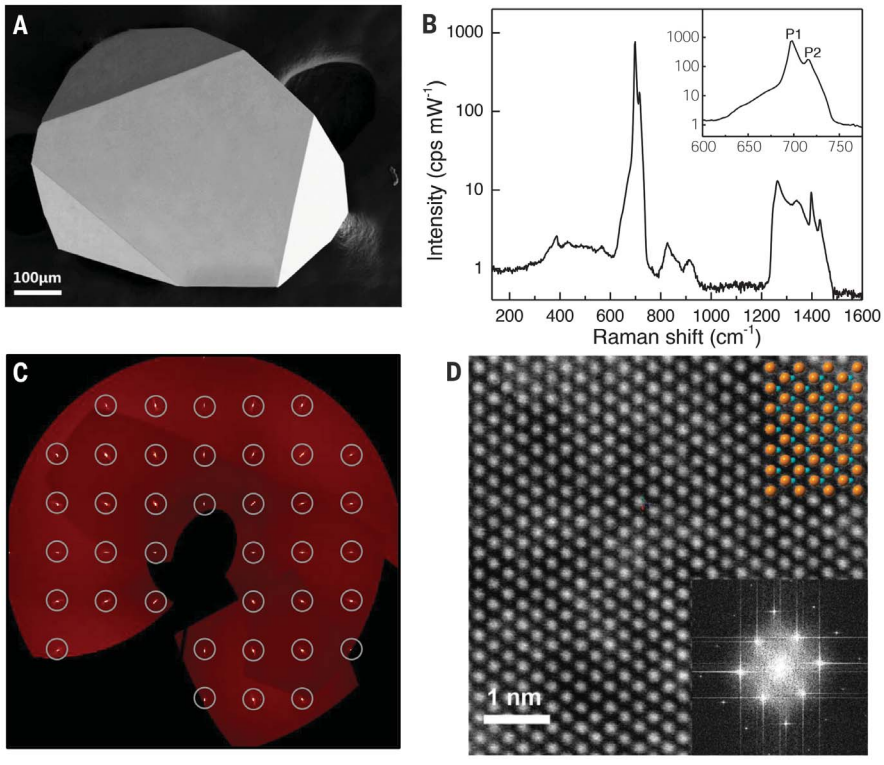


Fig. 1. Structural characterization of BAs crystals. (**A**) SEM image of BAs specimen **Ia**. (**B**) Raman spectrum of BAs specimen **Ia** with the intensity normalized to the excitation laser power (cps, counts per second). The inset shows the details near the T_2 longitudinal optical phonon at the Γ point. Assignment of the second-order peaks is discussed in (*9*). (**C**) Precession image of the *hk0* layer integrated from complete sets of single-crystal x-ray diffraction frames, where all detected reflections are indexed as circled spots, suggests a high-quality single lattice in BAs specimen **Ib**. (**D**) ADF-STEM image of BAs specimen **Ic** along the [110] zone axis, showing the zincblende crystal structure. FIB-induced surface damage and carbon contamination during image acquisition are responsible for the low-frequency contrast modulation across the image. The image is a cross-correlated sum of several image frames; raw data are available in fig. S5. A cartoon of the BAs lattice (orange spheres, As; blue spheres, B) is overlaid on the STEM image (top right). Inset: Fast Fourier transform of the STEM image.

lowering Λ. In a high-Λ specimen such as **Ia**, however, theoretical predictions that relate defect density to Λ indicate that the point defect concentration should be low (*33*). Further improving the controllability of the carrier concentration in the growth of BAs could be an important path for enhancing the yield of BAs with high Λ. Future development of large-scale fabrication of BAs films by CVD using gas precursors may facilitate its widespread application in thermal management of high–power density electronic and optoelectronic devices.

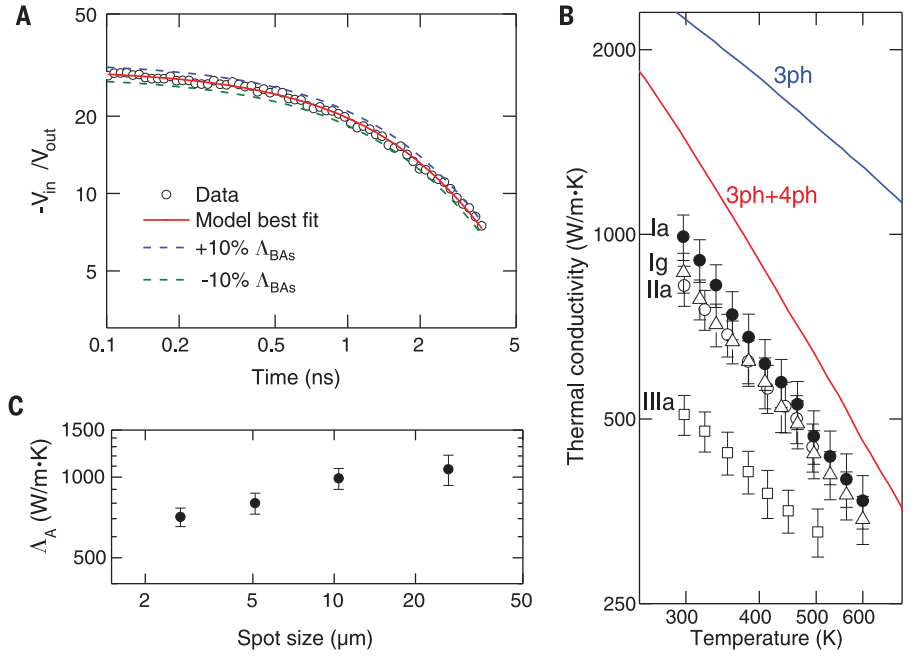


Fig. 2. Thermal conductivity measurement of BAs crystals. (**A**) TDTR ratio data [the ratio of the in-phase (V_{in}) and out-of-phase (V_{out}) signals from the lock-in amplifier] measured using a laser spot size with $1/e^2$ intensity radius of 10.4 μm (circles) and model fitting (solid line) for BAs specimen **Ia**. Model curves using thermal conductivities 10% larger and 10% smaller than the best-fit thermal conductivity (dashed lines) are included to illustrate the measurement sensitivity. (**B**) Temperature-dependent Λ of the BAs specimen **Ia** (solid circles), **Ig** (triangles), **IIa** (open circles), and **IIIa** (open squares) from 300 to 600 K using the same laser spot size as (A) and comparison with the predictions of first-principles calculations based on only three-phonon scattering (blue solid line) (*5*) and both three- and four-phonon process (red solid line) (*27*). (**C**) Spot size radius–dependent (apparent) thermal conductivity from TDTR for specimen **Ia** at 300 K.

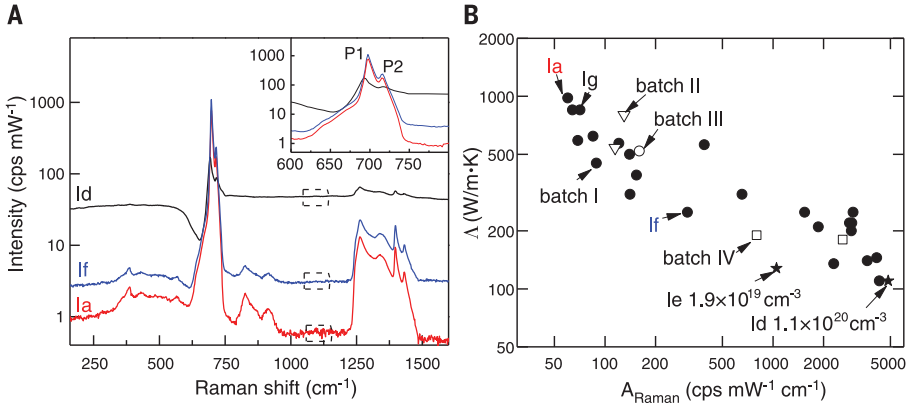


Fig. 3. Raman spectra and thermal conductivity of BAs. (**A**) Raman spectra of three representative BAs specimens, **Ia**, **Id**, and **If**. Specimen **Id** is measured to have a carrier concentration of 1.1 × 10^{20} cm^{-3} and a low thermal conductivity of 110 W/m·K. The black dashed squares indicate the range of the frequency between 1050 and 1150 cm^{-1} where the integration of the electronic Raman scattering intensity in (B), A_{Raman}, is performed. The inset shows the magnified Raman spectra of the samples in (A) near the one-phonon peaks. (**B**) Thermal conductivity of BAs crystals from several growth batches versus integrated Raman spectra intensity A_{Raman}. The data points for specimens in (A), the specimen on which we mapped Λ and A_{Raman} (**Ig**; see figs. S9 and S11), and the specimens with measured hole concentrations (**Id** and **Ie**) are highlighted.

REFERENCES AND NOTES

1. P. Ball, *Nature* **492**, 174–176 (2012).
2. M. M. Waldrop, *Nature* **530**, 144–147 (2016).
3. S. V. Garimella, L. T. Yeh, T. Persoons, *IEEE Trans. Compon. Packaging Manuf. Technol.* **2**, 1307–1316 (2012).
4. J. R. Olson *et al.*, *Phys. Rev. B* **47**, 14850–14856 (1993).
5. L. Lindsay, D. A. Broido, T. L. Reinecke, *Phys. Rev. Lett.* **111**, 025901 (2013).
6. Y. Okada, Y. Tokumaru, *J. Appl. Phys.* **56**, 314–320 (1984).
7. Q. Wang, X. Ren, H. Huang, Y. Huang, S. Cai, *Microelectron. J.* **40**, 87–91 (2009).
8. B. Lv *et al.*, *Appl. Phys. Lett.* **106**, 074105 (2015).
9. See supplementary materials.
10. V. G. Plekhanov, *Isotope-Based Quantum Information* (Springer, 2012).
11. V. G. Hadjiev, M. N. Iliev, B. Lv, Z. F. Ren, C. W. Chu, *Phys. Rev. B* **89**, 024308 (2014).
12. H. D. Fuchs, P. Etchegoin, M. Cardona, K. Itoh, E. E. Haller, *Phys. Rev. Lett.* **70**, 1715–1718 (1993).
13. F. Widulle *et al.*, *Solid State Commun.* **118**, 1–22 (2001).
14. J. Spitzer, P. Etchegoin, M. Cardona, T. R. Anthony, W. F. Banholzer, *Solid State Commun.* **88**, 509–514 (1993).
15. D. G. Cahill, *Rev. Sci. Instrum.* **75**, 5119–5122 (2004).
16. K. Kang, Y. K. Koh, C. Chiritescu, X. Zheng, D. G. Cahill, *Rev. Sci. Instrum.* **79**, 114901 (2008).
17. C. Chiritescu *et al.*, *Science* **315**, 351–353 (2007).
18. C. A. Paddock, G. L. Eesley, *J. Appl. Phys.* **60**, 285–290 (1986).
19. J. Kim *et al.*, *Appl. Phys. Lett.* **108**, 201905 (2016).
20. F. Tian *et al.*, *Appl. Phys. Lett.* **112**, 031903 (2018).
21. W. M. Haynes, *CRC Handbook of Chemistry and Physics* (CRC Press, ed. 96, 2015).
22. X. Qian, P. Jiang, R. Yang, *Mater. Today Phys.* **3**, 70–75 (2017).
23. L. Vel, G. Demazeau, J. Etourneau, *Mater. Sci. Eng. B* **10**, 149–164 (1991).
24. G. A. Slack, *Phys. Rev.* **127**, 694–701 (1962).
25. R. B. Wilson, D. G. Cahill, *Nat. Commun.* **5**, 5075 (2014).
26. R. B. Wilson, D. G. Cahill, *Appl. Phys. Lett.* **107**, 203112 (2015).
27. T. Feng, L. Lindsay, X. Ruan, *Phys. Rev. B* **96**, 161201 (2017).
28. J. M. Ziman, *Electrons and Phonons: The Theory of Transport Phenomena in Solids* (Oxford Univ. Press, 1960), pp. 288–333.
29. G. Abstreiter, M. Cardona, A. Pinczuk, in *Light Scattering in Solids IV*, M. Cardona, G. Güntherodt, Eds. (Springer, 1984), pp. 5–150.
30. F. Cerdeira, T. Fjeldly, M. Cardona, *Phys. Rev. B* **8**, 4734–4745 (1973).
31. M. A. Kanehisa, R. F. Wallis, M. Balkanski, *Phys. Rev. B* **25**, 7619–7625 (1982).
32. J. P. Dismukes, L. Ekstrom, E. F. Steigmeier, I. Kudman, D. S. Beers, *J. Appl. Phys.* **35**, 2899–2907 (1964).
33. N. H. Protik, J. Carrete, N. A. Katcho, N. Mingo, D. Broido, *Phys. Rev. B* **94**, 045207 (2016).

ACKNOWLEDGMENTS

We thank B. Janicek for help with STEM imaging measurement. **Funding:** Supported by Office of Naval Research (ONR) MURI grant N00014-16-1-2436. Thermal conductivity, Raman, and STEM research was carried out in part in Frederick Seitz Materials Research Laboratory (MRL) at the University of Illinois at Urbana-Champaign. S.L., X.L., and B.L. acknowledge support from U.S. Air Force Office of Scientific Research (AFOSR) grant FA9550-15-1-0236, and UT Dallas start-up funds. **Author contributions:** D.C. and B.L. conceived the project; S.L., X.L., and B.L. synthesized the materials; Q. Z, and D.C. performed TDTR and Raman measurements; Y.L. and P.H. performed STEM studies; X.W. and B.L. performed single-crystal analysis; D.C., B.L., S.L., and Q.Z. analyzed the results and co-edited the manuscript with input from all authors. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/579/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
Table S1
References (*34*–*56*)

17 April 2018; accepted 21 June 2018
Published online 5 July 2018
10.1126/science.aat8982

THERMAL CONDUCTIVITY

Unusual high thermal conductivity in boron arsenide bulk crystals

Fei Tian[1]*, Bai Song[2]*, Xi Chen[3]*, Navaneetha K. Ravichandran[4], Yinchuan Lv[5], Ke Chen[2], Sean Sullivan[3], Jaehyun Kim[6], Yuanyuan Zhou[6], Te-Huan Liu[2], Miguel Goni[7], Zhiwei Ding[2], Jingying Sun[1], Geethal Amila Gamage Udalamatta Gamage[1], Haoran Sun[1], Hamidreza Ziyaee[8], Shuyuan Huyan[1], Liangzi Deng[1], Jianshi Zhou[3,6], Aaron J. Schmidt[7], Shuo Chen[1], Ching-Wu Chu[1,9], Pinshane Y. Huang[10], David Broido[4]†, Li Shi[3,6]†, Gang Chen[2]†, Zhifeng Ren[1]†

Conventional theory predicts that ultrahigh lattice thermal conductivity can only occur in crystals composed of strongly bonded light elements, and that it is limited by anharmonic three-phonon processes. We report experimental evidence that departs from these long-held criteria. We measured a local room-temperature thermal conductivity exceeding 1000 watts per meter-kelvin and an average bulk value reaching 900 watts per meter-kelvin in bulk boron arsenide (BAs) crystals, where boron and arsenic are light and heavy elements, respectively. The high values are consistent with a proposal for phonon-band engineering and can only be explained by higher-order phonon processes. These findings yield insight into the physics of heat conduction in solids and show BAs to be the only known semiconductor with ultrahigh thermal conductivity.

Materials with high thermal conductivity (κ) can help to address a range of grand technological challenges, such as keeping nanoelectronic devices cool by removing the high-density heat generated within them. At room temperature (RT), diamond and graphite, the two carbon allotrope bulk crystals, have a record high κ of about 2000 W/m·K (*1*–*4*). However, high-quality natural diamond is scarce and expensive. Although future technological advances may help to alleviate the cost of high-quality synthetic diamond, the large mismatch in the coefficient of thermal expansion between diamond and common semiconductors can introduce large thermal stresses. The κ of graphite, moreover, is highly anisotropic, with the cross-plane value being two orders of magnitude smaller than the corresponding in-plane value (*1*). The thermal anisotropy and the weak interlayer bonding have limited the use of graphite for thermal management. In addition, the large bandgap of diamond and semimetallic behavior of graphite prevent their use as active electronic materials. Common electronic materials such as copper and silicon have a RT κ of about 400 and 150 W/m·K, respectively (*1*), which are well below the diamond value. The highest measured RT κ values for semiconductors are about 490 W/m·K in silicon carbide (*5*) and 460 W/m·K in boron phosphide (*6*). Although these values are comparable to the highest electronic contribution to κ in metals, it is desirable to discover semiconductors with values of κ comparable to the ultrahigh value for diamond.

In semiconductors and nonmagnetic insulators, the thermal conductivity is dominated by the phonon contribution. Thermal conductivity is typically limited by the lowest-order process arising from the anharmonicity of the interatomic potential, three-phonon scattering, at and above RT (*7*). According to the criteria established by Slack about half a century ago (*2*), only crystals composed of strongly bonded light elements should exhibit ultrahigh κ. However, Lindsay, Broido, and Reinecke recently proposed that ultrahigh κ could be achieved in compounds that combine a light and a heavy atom if (i) the frequency gap between heat-carrying acoustic phonons and optic phonons is sufficiently large and (ii) some of the acoustic phonons with different polarizations have regions of similar frequencies away from the Brillouin zone center. First-principles calculations supported this phonon-band engineering concept in predicting that cubic boron arsenide (BAs) should have a RT κ of around 2000 W/m·K when only three-phonon interaction is considered (*8*, *9*). Subsequent theoretical calculations found that four-phonon scattering would lower the calculated RT κ in BAs to about 1400 W/m·K (*10*), which is still exceptionally high, but surprising, because three-phonon scattering accurately describes the measured κ data for many semiconductors and insulators, and higher-order processes are expected to be weak at RT.

Synthesis of high-quality BAs bulk crystals has proved challenging, which has prevented experimental verification of the unusual predicted transport properties. Several measurements have captured RT κ values of only 200 to 350 W/m·K in small BAs particles (*11*–*13*). The inability to measure an ultrahigh κ for BAs has limited adoption of the phonon-band engineering strategy as a viable route for achieving ultrahigh κ, and the possibility of higher-order phonon processes suppressing κ has remained.

We report experimental evidence that validates the phonon-band engineering route. We grew BAs bulk crystals from seed microparticles in a

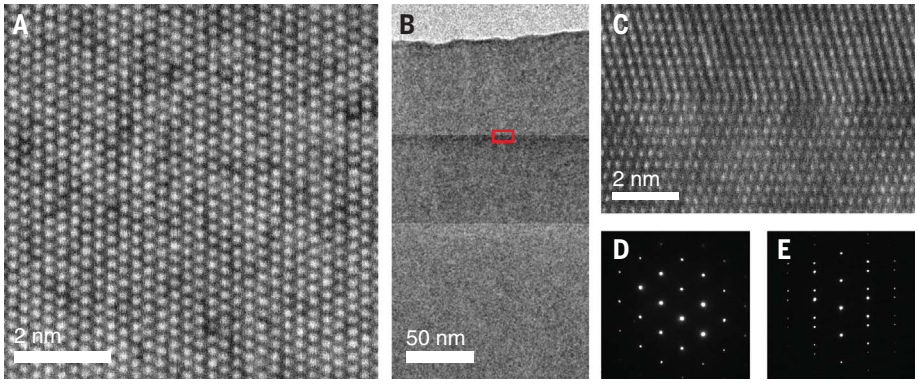


Fig. 1. STEM characterizations of BAs. (**A**) Annular dark-field STEM image within one grain of BAs, looking down the [110] zone axis. (**B**) Low-magnification bright-field TEM image near the surface of the BAs crystal. Horizontal lines indicate the locations of mirror twin boundaries. (**C**) Annular dark-field STEM image showing the atomic structure of the mirror twin boundary from the region highlighted by the red box in (B). (**D**) Electron diffraction pattern of BAs within a single grain. (**E**) Electron diffraction pattern of BAs across the grain boundary, showing the presence of the mirror twin.

[1]Department of Physics and Texas Center for Superconductivity, University of Houston, Houston, TX 77204, USA. [2]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Materials Science and Engineering Program, Texas Materials Institute, The University of Texas at Austin, Austin, TX 78712, USA. [4]Department of Physics, Boston College, Chestnut Hill, MA 02467, USA. [5]Department of Physics, University of Illinois Urbana–Champaign, Urbana, IL 61801, USA. [6]Department of Mechanical Engineering, The University of Texas at Austin, Austin, TX 78712, USA. [7]Department of Mechanical Engineering, Boston University, Boston, MA 02215, USA. [8]Department of Mechanical Engineering, University of Houston, Houston, TX 77204, USA. [9]Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [10]Department of Materials Science and Engineering, University of Illinois Urbana–Champaign, Urbana, IL 61801, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: david.broido@bc.edu (D.B.); lishi@mail.utexas.edu (L.S.); gchen2@mit.edu (G.C.); zren@uh.edu (Z.R.)

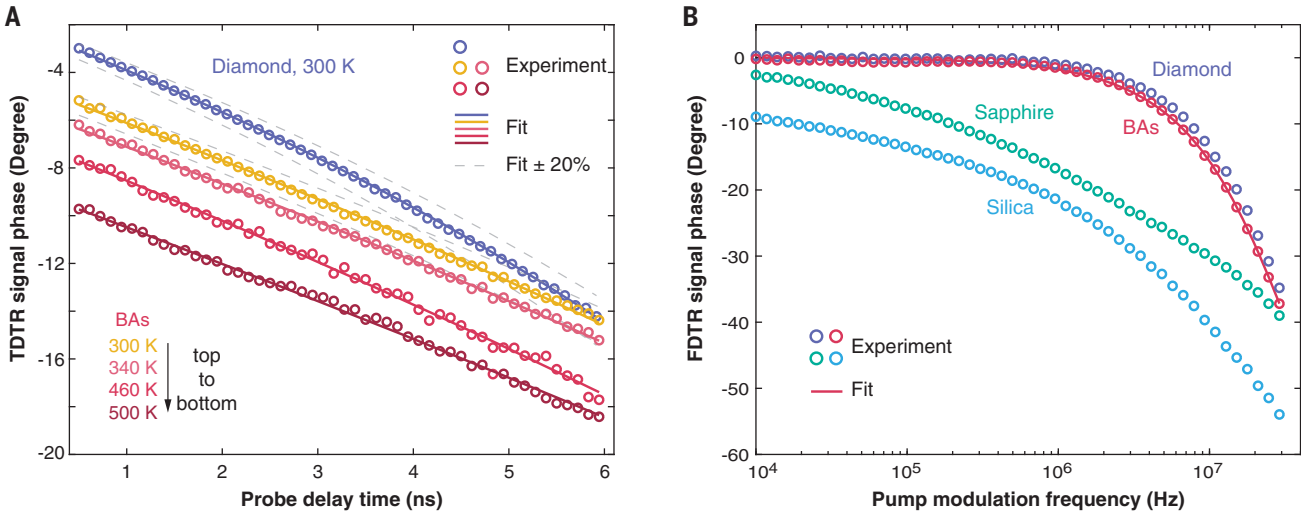


Fig. 2. TDTR and FDTR measurements. (**A**) Representative TDTR phase signals and best-fitted curves for a diamond crystal acquired from Element Six and a BAs crystal at different temperatures. The diamond sample has the natural carbon abundance (1.1% ^{13}C) and a low level of boron [<0.05 parts per million (ppm)] and nitrogen (<1 ppm) impurities. The 300 K data are averaged over 200 and 140 runs at the same location for diamond and BAs, respectively. The data for BAs at higher temperatures are averages of about 10 runs and show slightly increased noise. (**B**) Representative FDTR signal phase as a function of the pump modulation frequency measured on a BAs crystal, diamond, sapphire, and fused silica. The phase lag between the probe and the pump increases with decreasing sample κ.

chemical vapor transport (CVT) process. Local measurements of low-defect regions obtained a RT κ exceeding 1000 W/m·K, whereas multiple local and bulk transport measurement methods yielded average RT κ values of about 800 and 900 W/m·K for two bulk crystal samples. The bulk crystal has a high κ despite twin boundaries and other defects known to decrease κ. Both the peak and average κ values show a rapid decrease with increasing temperature, which is a signature of lattice anharmonicity. This behavior agrees with our detailed first-principles theoretical model that included both three- and four-phonon interactions.

Previously reported efforts to synthesize BAs yielded only particles with maximum dimensions less than about 500 μm (*11*–*13*). Because bulk-size crystals are required for device applications, we investigated a seeded CVT growth mechanism for the synthesis of bulk BAs crystals. In this approach, we used small single BAs crystals with a lateral dimension of a few micrometers as seeds to ensure that the nucleation centers were sparse and under control during the growth process (*13*, *14*). We optimized seed-crystal quality and distribution to obtain BAs crystals as large as about 4 mm by 2 mm by 1 mm within a 14-day period of seed growth followed by another 14-day period of crystal growth from the chosen seed crystals (*15*). This increased crystal size allowed us to use transport measurement techniques established for bulk samples. Increasing the growth time would increase the crystal size. To probe the crystal structure of the BAs, we obtained an aberration-corrected, annular dark-field scanning transmission electron microscopy (STEM) image (Fig. 1A) and a low-magnification bright-field TEM image (Fig. 1B) of representative crystals. We found planar defects (Fig. 1B) that were mirror twin boundaries (Fig. 1, C to E).

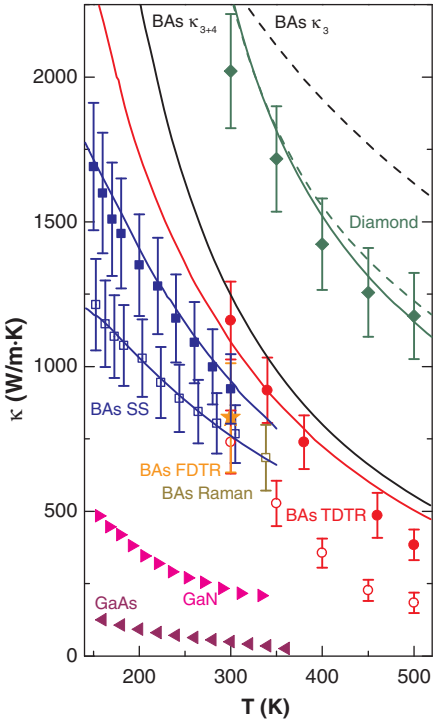


Fig. 3. Measured thermal conductivity of BAs in comparison with values from theoretical calculations and other crystals. Calculated κ versus temperature for BAs (black) and diamond (green) including only three-phonon scattering (dashed lines) and both three- and four-phonon scattering (solid lines); measured κ for diamond by TDTR (green diamonds); measured κ for BAs samples 1 (solid red symbols) and 2 (open red symbols) by TDTR; measured κ for sample 3 by FDTR (solid orange star, mean value), steady-state (SS; open blue squares), and lock-in Raman (open brown square) methods; and measured κ for sample 5 by the steady-state method (solid blue squares). Also shown are the fits to measured steady-state and TDTR κ for BAs (blue and red solid lines, respectively) and reported measured κ for GaN (*21*) and GaAs (*22*) (magenta and purple triangles, respectively). The error bars for the TDTR and FDTR data represent one standard deviation and were obtained via Monte Carlo simulations and derivative matrix-based analysis of uncertainty propagation, respectively (*15*). The error bars for the steady-state and lock-in Raman measurements were calculated by propagating random errors at 95% confidence and combining them with systematic errors (*15*).

We discovered unusually high but nonuniformly distributed κ in these BAs crystals by using time- and frequency-domain thermoreflectance (TDTR and FDTR) techniques with micrometer resolution (*16*–*18*). We used a large 58-μm-diameter pump laser spot and a small 9-μm-diameter probe laser spot in conjunction with a relatively low modulation frequency of 3 MHz to improve the TDTR measurement accuracy (*15*). For both the pump beam and probe beam, the diameter quoted here is the $1/e^2$ diameter of the Gaussian beam. We used the same TDTR platform and parameters to measure the κ of a synthetic diamond crystal (Figs. 2A and 3). The values that we measured for diamond are in good agreement with theoretical calculations and literature values (*1*–*4*). Among the single-spot measurements at five locations on BAs sample 1, the highest and lowest RT κ values were 1160 ± 130 and 640 ± 70 W/m·K, respectively. Among the

10 single-spot TDTR measurements on sample 2, RT κ values ranged from 790 ± 100 to 450 ± 60 W/m·K. We found a sharp decrease in κ as temperature increased to 500 K at the location on sample 1 where we found our maximum κ (Fig. 3). We found the same behavior at a location on sample 2 with a RT κ of 740 ± 110 W/m·K (Fig. 3). This temperature behavior is consistent with dominant anharmonic phonon-phonon scattering. At the same sample 1 spot where our TDTR measurements yielded the highest κ, our single-spot FDTR measurements obtained a κ of 1310 ± 740 W/m·K, where the large uncertainty is due to the use of a small 3.36-μm-diameter pump beam and 2.60-μm-diameter probe spot to measure the high-κ region (*19*). We observed no spot-size dependence of κ when we increased the pump and probe laser spot sizes to 5.60 and 4.80 μm, respectively.

The increased size of the bulk BAs crystals allowed us to make steady-state comparative measurements of the bulk κ (Fig. 4A) (*15*). Without accounting for the contact thermal resistance errors between the thermocouples and the sample, we obtained a κ of 770 ± 100 W/m·K at 305 K on a 0.1 mm by 0.2 mm by 2 mm bar cut from sample 3. The thermal conductivity increased with deceasing temperature. In comparison, we obtained an average value of 820 ± 140 W/m·K with FDTR at 14 locations on another piece cut from sample 3 (Fig. 3). We addressed the uncertainty due to contact resistance by using a lock-in Raman thermometry approach with a sinusoidally modulated heating current at a low modulation frequency (υ) of about 1 mHz (*15*). A fast Fourier transform of the measured Raman peak shift shows clear modulation at the second harmonic frequency corresponding to the Joule heating frequency (Fig. 4B), which we used to measure the temperature drops along the Si and BAs bars (Fig. 4A). The Raman measurements obtained similar temperature gradients as the thermocouple measurements in both silicon and BAs (Fig. 4C) and a κ of 690 ± 120 W/m·K at 338 K (Fig. 3). On samples 4 (fig. S19B) (*15*) and 5 (Fig. 3), we used the steady-state method to measure a bulk κ at 300 K of 570 ± 70 and 920 ± 120 W/m·K, respectively, and found a similar temperature dependence as for samples 1 to 3.

The measurement results agree with first-principles calculations of the κ of BAs, including both three- and four-phonon scattering, scattering of phonons by the natural boron isotope mix, and phonon scattering by point defects and grain boundaries. Although the κ of most high-quality insulating crystals is well described by lowest-order three-phonon scattering, in BAs, the phase space for three-phonon scattering is unusually small (*8*). We show that four-phonon scattering is necessary to accurately capture the intrinsic κ of BAs (*10*). We implemented several changes in our calculation from that of Feng *et al.* (*10*) to improve the accuracy (*15*). In addition, we found hole scattering of phonons to be negligible at the hole concentration of 7.6×10^{18} cm^{-3} that we measured in the p-type BAs semiconductor sample (*15*).

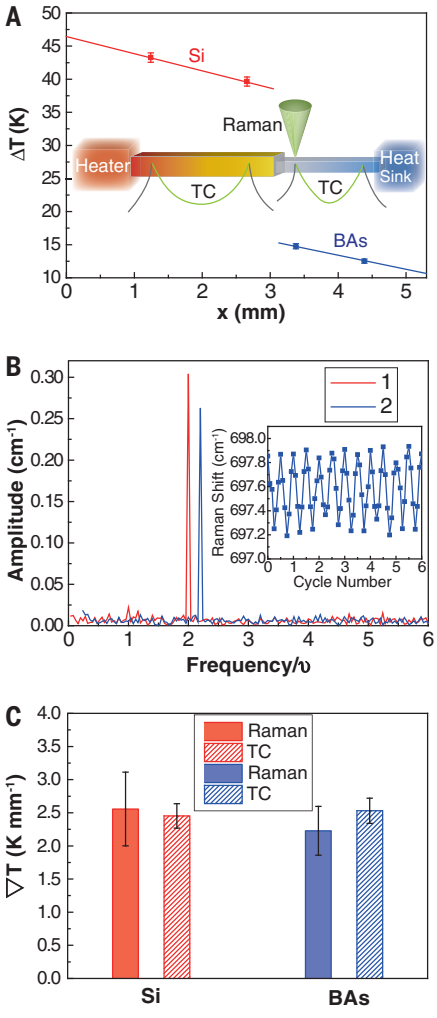


Fig. 4. Steady-state comparative and lock-in Raman thermometry measurements. (**A**) Temperature modulation amplitudes (Δ*T*) measured by Raman thermometry at two locations on the Si bar and two locations on the BAs bar. The lines are linear fittings to the measurement data. The inset is a schematic diagram of the experimental setup for thermocouple (TC) and Raman measurements. (**B**) Amplitude spectrum of the measured Raman peak modulation for BAs at locations x = 3.38 mm (location 1) and 4.39 mm (location 2). The curve for x = 4.39 mm is shifted manually by +0.2 along the x axis so that it can be distinguished from the other curve. The inset shows the modulation of the Raman peak frequency of BAs at location x = 3.38 mm as a function of the cycle number during the first six cycles. (**C**) Temperature gradients (∇*T*) on the Si and BAs bars obtained from TC and Raman measurements. The ambient temperature was 308.9 K, and the heater power amplitude was 0.081 W. The TC measurement error bars include random uncertainties with 95% confidence and systematic uncertainty, and the Raman measurement error bars consist of random uncertainties from the signal-to-noise ratio and systematic error (*15*).

Our calculated BAs κ at 300 K including three-phonon, four-phonon, and phonon-isotope scattering is 1260 W/m·K, about half that obtained without four-phonon scattering (2330 W/m·K) and about 10% smaller than that obtained by Feng *et al.* (*10*). The BAs κ that we calculated including only three-phonon and phonon-isotope scattering, κ_3 (dashed black curve, Fig. 3), lies well above all measured data. The calculation including four-phonon scattering as well, κ_{3+4} (solid black curve), suppresses κ and brings the calculated values close to the measured local high TDTR values (solid red circles). It also provides a strong temperature dependence, as previously found (*10*). Importantly, the TDTR temperature behavior follows the temperature dependence of κ_{3+4}, which is stronger than that of κ_3. We fit the steady-state and TDTR data by including additional scattering from assumed point defects and grain boundaries (Fig. 3) (*15*). Defect scattering mechanisms are typically much less sensitive to temperature change than phonon-phonon scattering, so increasing defect scattering to match the measured RT value weakens the temperature dependence of κ. The large defect concentrations needed to match the magnitudes of the measured data when including only three-phonon scattering cannot produce the steep observed temperature dependence (figs. S30 and S31) (*15*). In contrast, the best fit of κ_{3+4} is excellent for the steady-state data (fig. S19B) (*15*) and reasonably good for the TDTR data.

The comparison between the measurements and theoretical calculations provides strong evidence that BAs is distinct from other known high-κ materials in achieving κ through the phonon-band engineering concept and in having higher-order phonon-phonon interactions play such a large role. By breaking the conventional theoretical criteria, these findings have firmly established a different route to ultrahigh κ and highlighted the rich physics of phonons. Our strategy for the growth of bulk BAs crystal is an important step toward implementation in future applications of BAs, which is now the only known semiconductor with a bandgap (*20*) comparable to silicon and an ultrahigh room-temperature thermal conductivity.

REFERENCES AND NOTES

1. Thermophysical Properties Research Center, Purdue University, *Thermophysical Properties of Matter* (IFI, 1970–1979).
2. G. A. Slack, *J. Phys. Chem. Solids* **34**, 321–335 (1973).
3. L. Wei, P. K. Kuo, R. L. Thomas, T. R. Anthony, W. F. Banholzer, *Phys. Rev. Lett.* **70**, 3764–3767 (1993).
4. J. R. Olson *et al.*, *Phys. Rev. B* **47**, 14850–14856 (1993).
5. G. A. Slack, *J. Appl. Phys.* **35**, 3460–3466 (1964).
6. J. S. Kang, H. Wu, Y. Hu, *Nano Lett.* **17**, 7507–7514 (2017).
7. J. M. Ziman, *Electrons and Phonons: The Theory of Transport Phenomena in Solids* (Oxford Univ. Press, 1960).
8. L. Lindsay, D. A. Broido, T. L. Reinecke, *Phys. Rev. Lett.* **111**, 025901 (2013).
9. D. Broido, L. Lindsay, T. Reinecke, *Phys. Rev. B* **88**, 214303 (2013).
10. T. Feng, L. Lindsay, X. Ruan, *Phys. Rev. B* **96**, 161201 (2017).
11. B. Lv *et al.*, *Appl. Phys. Lett.* **106**, 074105 (2015).
12. J. Kim *et al.*, *Appl. Phys. Lett.* **108**, 201905 (2016).
13. F. Tian *et al.*, *Appl. Phys. Lett.* **112**, 031903 (2018).
14. P. Cubillas, M. W. Anderson, in *Zeolites and Catalysis* (Wiley-VCH, 2010), pp. 1–55.
15. Supplementary materials.
16. D. G. Cahill, *Rev. Sci. Instrum.* **75**, 5119–5122 (2004).

17. A. J. Schmidt, X. Chen, G. Chen, *Rev. Sci. Instrum.* **79**, 114902 (2008).
18. J. Yang, C. Maragliano, A. J. Schmidt, *Rev. Sci. Instrum.* **84**, 104904 (2013).
19. J. Yang, E. Ziade, A. J. Schmidt, *Rev. Sci. Instrum.* **87**, 014901 (2016).
20. T. L. Chu, A. E. Hyslop, *J. Electrochem. Soc.* **121**, 412–415 (1974).
21. G. A. Slack, L. J. Schowalter, D. Morelli, J. A. Freitas Jr., *J. Cryst. Growth* **246**, 287–298 (2002).
22. R. Carlson, G. Slack, S. Silverman, *J. Appl. Phys.* **36**, 505–507 (1965).

ACKNOWLEDGMENTS

We thank S. Huberman, J. Zhou, and Y. Xu for assistance with the experiments. **Funding:** The work was supported by the Office of Naval Research under MURI grant N00014-16-1-2436. **Author contributions:** F.T. and Z.R. designed the crystal growth process. F.T. grew the crystal samples. B.S., K.C., M.G., T.-H.L., Z.D., A.J.S., and G.C. performed the thermal conductivity measurements and analyses by microprobe. X.C., S.S., J.K., Y.Z., J.Z., and L.S. performed the thermal conductivity measurements by steady-state and Raman methods and the electrical transport property measurements. N.K.R. and D.B. performed the thermal conductivity calculations. Y.L. and P.Y.H. performed electron microscopy characterization in the Fredrick-Seitz Material Research Laboratory Central Facilities. J.S., G.A.G.U.G., H.S., H.Z., L.D., S.H., S.C., and C.-W.C. performed structural characterizations. F.T., B.S., X.C., D.B., L.S., G.C., and Z.R. wrote the paper. All authors contributed to the discussion of the results and writing of the manuscript. The project was directed and supervised by Z.R., G.C., L.S., and D.B. **Competing interests:** A provisional patent application has been filed based on this work. D.B. is an inventor on U.S. Patent 9,986,663. **Data and materials availability:** All data are available in the manuscript and supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/582/suppl/DC1
Materials and Methods
Figs. S1 to S33
References (*23*–*30*)

18 April 2018; accepted 21 June 2018
Published online 5 July 2018
10.1126/science.aat7932

RIVER NETWORKS

Global extent of rivers and streams

George H. Allen*† and Tamlin M. Pavelsky

The turbulent surfaces of rivers and streams are natural hotspots of biogeochemical exchange with the atmosphere. At the global scale, the total river-atmosphere flux of trace gasses such as carbon dioxide depends on the proportion of Earth's surface that is covered by the fluvial network, yet the total surface area of rivers and streams is poorly constrained. We used a global database of planform river hydromorphology and a statistical approach to show that global river and stream surface area at mean annual discharge is 773,000 ± 79,000 square kilometers (0.58 ± 0.06%) of Earth's nonglaciated land surface, an area 44 ± 15% larger than previous spatial estimates. We found that rivers and streams likely play a greater role in controlling land-atmosphere fluxes than is currently represented in global carbon budgets.

Water interacts with the atmosphere in a series of complex biogeochemical processes at the water-atmosphere interface as it flows down Earth's rivers and streams (*1*–*5*). At this interface, equilibrium reactions drive mass and energy exchange, amounting to considerable material flux at the global scale (*4*–*6*). For example, estimated outgassing from rivers and streams introduces ~1.8 Pg of carbon per year as carbon dioxide to the atmosphere (*1*), roughly equivalent to one-fifth of combined emissions from fossil fuel combustion and cement production (*7*). Globally, the rates of these processes are partly controlled by the total river and stream surface area (RSSA), which acts as the medium of exchange between the fluvial network and the atmosphere (*1*, *2*, *4*–*6*). Despite the fact that RSSA is one of the principal parameters in large-scale evaluations of river-atmosphere biogeochemical and thermal flux (*1*, *2*, *5*, *6*), the field of large-scale river hydrology has primarily focused on quantifying the volume of water that rivers and streams transport to the ocean, rather than RSSA (*2*, *8*).

Only two studies have attempted to estimate global RSSA to date. In a pioneering effort, Downing *et al.* (*2*) developed stream-order scaling relationships between river width and length under the assumption that all rivers belong to a single branching river network. They made two global RSSA estimates: 485,000 km^2, based on an aggregate estimate of RSSA for rivers wider than 90 m, and 682,000 km^2, based on the length and width of the Amazon trunk river. Raymond *et al.* (*1*) remain the only group to estimate the spatial variability of RSSA globally. They arrived at a total RSSA estimate of 536,000 km^2 (excluding Greenland, Antarctica, and the seasonal effects of freezing rivers on RSSA) by applying a flow-routing algorithm to digital topography and assuming globally constant hydraulic geometry relationships between river width and discharge (*9*). Both of these previous studies are limited by the lack of direct observations of RSSA, quantification of statistical uncertainty, and consideration of regional variability in hydraulic geometry. We used satellite observations of rivers and a statistical approach to produce a direct estimate of river and stream coverage at the global scale.

We built the Global River Widths from Landsat (GRWL) Database to characterize the global coverage of rivers and streams. The GRWL Database is the first global compilation of river planform geometry at a constant-frequency discharge (Fig. 1). We used a global database of 3693 gauge stations (*10*) to determine months that rivers were commonly near mean discharge (fig. S1). Then we acquired 7376 Landsat TM (Thematic Mapper), ETM+ (Enhanced Thematic Mapper Plus), and OLI (Operational Land Imager) scenes captured during these months. We applied previously published image processing techniques (*11*, *12*) to classify rivers and measure their location, width, and braiding index. The GRWL Database contains planform measurements of >2.1 million km (>58 million measurements) of rivers ≥30 m wide at mean annual discharge. It also contains >7.6 million flagged measurements of lakes, reservoirs, and canals connected to the fluvial network. We validated the Landsat-derived width data by using in situ river width measurements from the U.S. Geological Survey and the Water Survey of Canada taken at 1250 gauge stations (Fig. 2) (*13*). We found that GRWL width data are most accurate and complete at widths wider than 90 m (about three Landsat pixels), and thus we considered only rivers wider than this width to assess the statistical distributions of RSSA (*11*, *12*, *14*).

The freely available GRWL vector product and water mask have considerable potential to improve the representation of large-scale fluvial processes and understanding of river resources (*15*). Although other empirical datasets of river width exist, their coverage is not global, or their coarse spatial resolution limits their usefulness for river system models (*11*, *14*, *16*). Subsets of the GRWL data are already being used to improve hydrologic models (*17*), organize remotely sensed surface-water observations (*18*), and improve biogeochemical efflux estimates (*11*). The database will also be used to identify river segments observable by the NASA and Centre National d'Etudes Spatiales SWOT (Surface Water and Ocean Topography) satellite, scheduled to launch in 2021 (*19*). Further, the GRWL Database has applications for fluvial geomorphology (e.g., studies of river sinuosity), determining spatiotemporal variations in river discharge

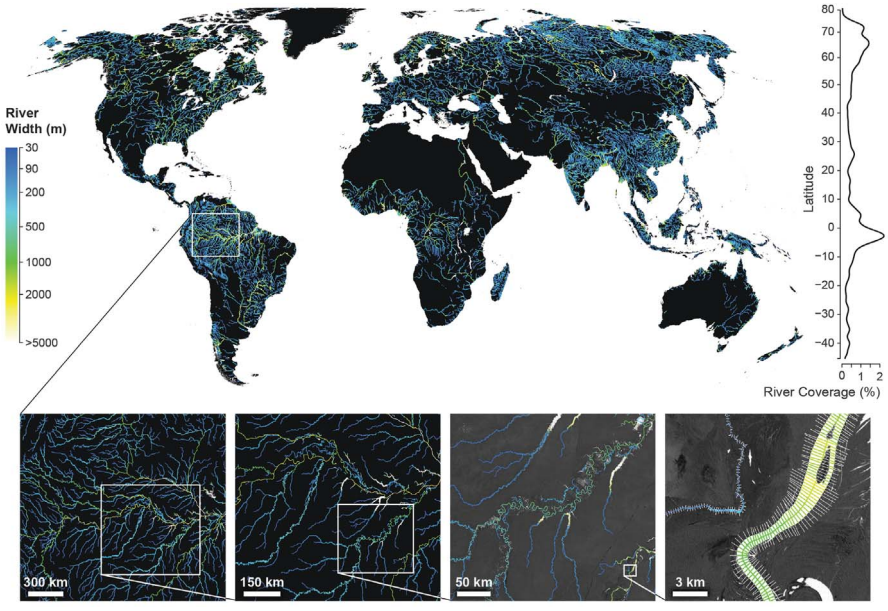


Fig. 1. The Global River Widths from Landsat (GRWL) Database contains more than 58 million measurements of planform river geometry. The line plot on the right shows observed river coverage as a percentage of land area by latitude, and the bottom insets show GRWL at increasing zoom. The rightmost inset shows GRWL orthogonals over which river width was calculated, with only every eighth orthogonal shown for clarity.

Department of Geological Sciences, University of North Carolina, Chapel Hill, NC, USA.
*Present address: Department of Geography, Texas A&M University, College Station, TX, USA.
†Corresponding author. Email: georgehenryallen@gmail.com

at the global scale (*20*), and organizing large multitemporal datasets of surface-water dynamics (*21*).

This newly assembled database of river hydromorphology allows direct quantification of RSSA for large, observable rivers. By summing the product of each river width measurement and its corresponding downstream pixel length (Fig. 3A), the total observed area of rivers measured by the GRWL Database is 468,000 km^2, or 0.35% of Earth's nonglaciated land surface. We excluded reservoirs, lakes, canals, Antarctica, Greenland, and any water bodies measured at mean sea level (*22*) from this analysis to make it comparable to previous studies. The total surface area of rivers wider than 90 m, where GRWL data are most complete and accurate, is 404,000 km^2, which exceeds a previous aggregate estimate of 360,000 km^2 (*2*).

To estimate the surface area of streams and rivers too narrow to accurately observe from Landsat imagery (widths < 90 m), we split GRWL measurements into major global drainage basins (*22*) and grouped these drainage basins into three categories: basins that contain >250,000 measurements (class A), basins that contain 10,000 to 250,000 measurements (class B), and basins that contain ≤10,000 measurements (class C) (Fig. 3B). In class A basins (n = 20), we estimated the total RSSA by extending a fitted Pareto frequency distribution down to the median first-order wetted stream width of 32 ± 8 cm (*23*) (Fig. 3C). Both theoretical (*24*, *25*) and empirical (*2*, *25*, *26*) evidence indicates that RSSA is fractal down to first-order streams, although this assumption should be tested. In class B basins (n = 273), which have insufficient GRWL data to exhibit a well-developed fractal RSSA distribution, we used the average Pareto shape parameter established in class A basins (fig. S2) to extend the RSSA distribution to first-order streams (Fig. 3D). Class C basins have very little GRWL data, so we developed an empirical power-law relationship between climate aridity (*27*), basin area (*22*), and percent basin occupied by RSSA (coefficient of determination R^2 = 0.68; Fig. 3E). This relationship is noteworthy because it indicates a link between variations in climate and the extent of rivers and streams at a global scale (*1*). Adding together the RSSA contained in all basins, the global surface area of rivers and streams at mean annual discharge is 773,000 ± 79,000 km^2, or 0.58 ± 0.06% of Earth's nonglaciated land surface (Fig. 4A). We used a Monte-Carlo simulation to calculate the uncertainty of our RSSA estimates in each global

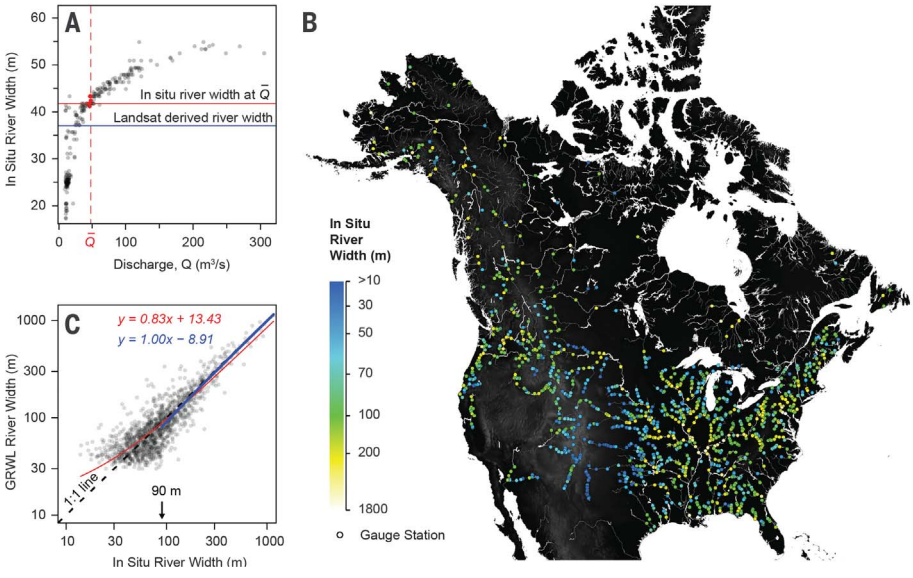


Fig. 2. Validating remote sensing measurements. (**A**) Example of an in situ river discharge-width rating curve used to validate Landsat measurements. (**B**) Gauge stations used in validation, colored by in situ width at mean annual discharge ($\bar{Q}$). (**C**) In situ river widths compared with corresponding Landsat-derived GRWL river widths. Red line, fit to all data; blue line, fit to in situ widths wider than 90 m.

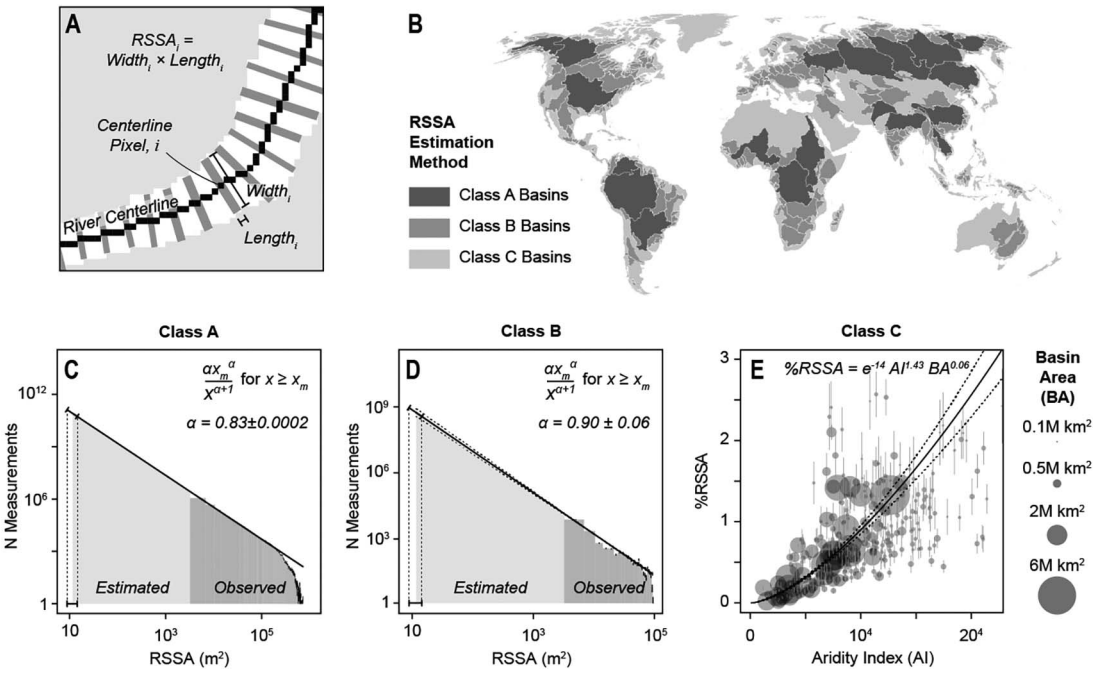


Fig. 3. Estimating the global surface area of rivers and streams. (**A**) Schematic of discretized river and stream surface area measurements ($RSSA_i$, gray bars), with only every fourth measurement shown for clarity. At a given river centerline pixel i, $RSSA_i$ is the product of the river width and the downstream pixel length. (**B**) Map of basins by class; a different RSSA estimation method was used for each class. (**C**) RSSA in class A basins, estimated using a Pareto distribution fit on observed wide rivers and extrapolated to narrower streams unobservable from Landsat (the Amazon basin is shown as an example; see section 2 of the supplementary materials). Throughout the figure, vertical black lines on observed RSSA bins show 1σ uncertainty of GRWL width measurements, and dotted lines are 1σ uncertainty bounds. (**D**) RSSA in class B basins, estimated using a Pareto distribution with a fixed shape parameter, α, fit on observed wide rivers and extrapolated to narrower streams (the Delaware basin is shown as an example). (**E**) RSSA in class C basins, estimated using weighted multiple linear regression of log-transformed percent RSSA (%RSSA) against aridity index and basin area in class A and B basins. Vertical gray lines show 1σ uncertainty bounds.

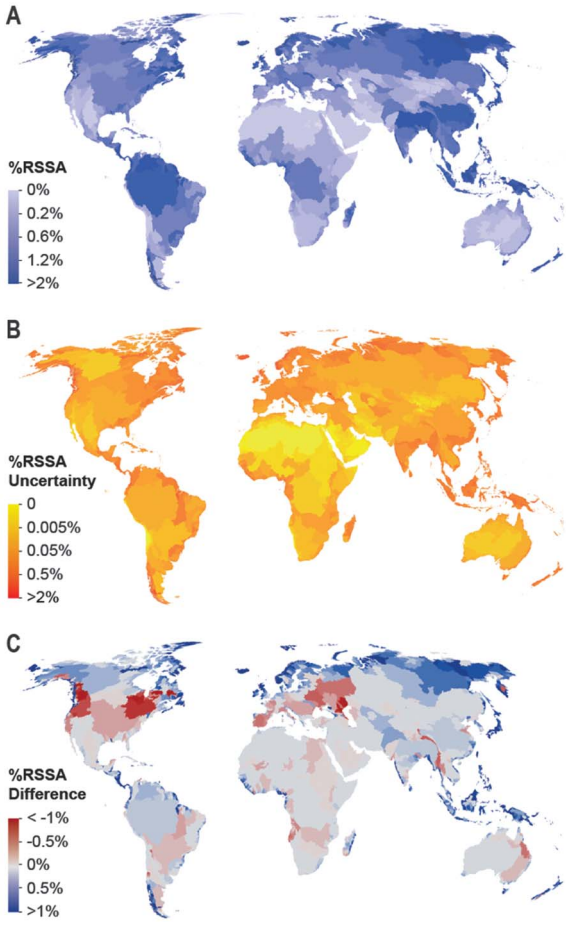


Fig. 4. Global patterns of stream and river coverage. (**A**) Percent of basin covered by river and stream surface area (%RSSA). (**B**) %RSSA uncertainty by basin. (**C**) %RSSA difference between this study and Raymond *et al*. (*1*).

basin, and we found that small, humid basins tend to exhibit the greatest uncertainty (n = 500; Fig. 4B) (*13*).

Our analysis shows that rivers and streams cover a larger portion of Earth's surface than previously estimated (*1*, *2*). We found a RSSA 44 ± 15% greater than that found by Raymond *et al*. (*1*), which is the only other geographically varying global estimate of RSSA (Fig. 4C). Our estimate is also 15 ± 12% greater than the maximum and 59 ± 16% greater than the minimum RSSA estimate from Downing *et al*. (*2*). In the Amazon basin, where a variety of methods have been used to estimate RSSA, we found that rivers and streams occupy 1.33 ± 0.02% of the basin at mean annual discharge, an area 6 to 67% greater than previous estimates (*1*, *28*, *29*). Compared with the current best region-by-region global estimate (*1*), we found more river and stream coverage in the Arctic and less in Europe, the conterminous United States, and some other economically developed regions (Fig. 4C). Previous estimates of global RSSA do not consider extra-climatic influences on RSSA, such as variations in fluvial geomorphology and human modifications to river channels, potentially resulting in an overestimate in some developed regions. For example, RSSA in many developed regions may be less than previously predicted owing to the influence of leveeing and water withdrawal in these areas.

The upward revision of the total global surface area of the fluvial network implies that interactions between rivers and the atmosphere are likely greater than previously thought. The upward revision is particularly pronounced in the Arctic, where the impacts of climate change on carbon fluxes are of major concern (*30*). Our findings also imply that the atmosphere plays a greater role in controlling the thermal dynamics and aquatic chemistry of river and stream water (*5*, *6*). The downward revision of RSSA in economically developed regions may be related to the large-scale impact of human modification on the fluvial network, although this hypothesis requires further testing. The largest sources of unquantified uncertainty in our RSSA estimate likely originate from the distribution of surface area for intermediate-sized rivers and streams and the seasonal variation of RSSA within river networks. As we develop analyses to address these uncertainties, our conclusions provide a robust first-order RSSA estimate that will be useful for improving the accuracy of large-scale fluvial biogeochemical fluxes.

REFERENCES AND NOTES

1. P. A. Raymond *et al*., *Nature* **503**, 355–359 (2013).
2. J. A. Downing *et al*., *Inland Waters* **2**, 229–236 (2012).
3. D. Bastviken, L. J. Tranvik, J. A. Downing, P. M. Crill, A. Enrich-Prast, *Science* **331**, 50 (2011).
4. E. R. Hotchkiss *et al*., *Nat. Geosci.* **8**, 696–699 (2015).
5. D. Butman *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **113**, 58–63 (2016).
6. D. Caissie, *Freshw. Biol.* **51**, 1389–1406 (2006).
7. G. P. Peters *et al*., *Nat. Clim. Chang.* **3**, 4–6 (2012).
8. A. Dai, in *Terrestrial Water Cycle and Climate Change* (John Wiley & Sons, 2016), pp. 17–37.
9. L. B. Leopold, T. Maddock Jr., "The hydraulic geometry of stream channels and physiographic implications" (USGS Professional Paper 252, U.S. Geological Survey, 1953).
10. Global Runoff Data Centre, Long-term mean monthly discharges and annual characteristics of GRDC stations (2017); available at www.bafg.de/.
11. G. H. Allen, T. M. Pavelsky, *Geophys. Res. Lett.* **42**, 395–402 (2015).
12. T. M. Pavelsky, L. C. Smith, *IEEE Geosci. Remote Sens. Lett.* **5**, 70–73 (2008).
13. Materials and methods are available as supplementary materials.
14. Z. F. Miller, T. M. Pavelsky, G. H. Allen, *Hydrol. Earth Syst. Sci.* **18**, 4883–4895 (2014).
15. M. P. Clark *et al*., *Water Resour. Res.* **51**, 5929–5956 (2015).
16. D. Yamazaki *et al*., *Water Resour. Res.* **50**, 3467–3480 (2014).
17. Y. Yoon, E. Beighley, H. Lee, T. Pavelsky, G. Allen, *J. Hydrol. Eng.* **21**, 05015030 (2016).
18. T. M. Pavelsky *et al*., *J. Hydrol. (Amst.)* **519**, 1516–1525 (2014).
19. S. Biancamaria, D. P. Lettenmaier, T. M. Pavelsky, *Surv. Geophys.* **37**, 307–337 (2016).
20. C. J. Gleason, L. C. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4788–4791 (2014).
21. J.-F. Pekel, A. Cottam, N. Gorelick, A. S. Belward, *Nature* **540**, 418–422 (2016).
22. B. Lehner, G. Grill, *Hydrol. Processes* **27**, 2171–2186 (2013).
23. G. H. Allen *et al*., *Nat. Commun.* **9**, 610 (2018).
24. J. W. Kirchner, *Geology* **21**, 591–594 (1993).
25. I. Rodríguez-Iturbe, A. Rinaldo, *Fractal River Basins: Chance and Self-Organization* (Cambridge Univ. Press, 2001).
26. M. E. Morisawa, *Geol. Soc. Am. Bull.* **73**, 1025–1046 (1962).
27. R. J. Zomer, A. Trabucco, D. A. Bossio, L. V. Verchot, *Agric. Ecosyst. Environ.* **126**, 67–80 (2008).
28. R. E. Beighley, V. Gummadi, *Earth Surf. Process. Landf.* **36**, 1059–1071 (2011).
29. L. L. Hess *et al*., *Wetlands* **35**, 745–756 (2015).
30. X. Feng *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14168–14173 (2013).
31. G. H. Allen, T. M. Pavelsky, Global River Widths from Landsat (GRWL) Database. Zenodo (2018); https://doi.org/10.5281/zenodo.1297434.

ACKNOWLEDGMENTS

We thank K. Hinson, C. Destefano, S. Dawson, C. Nelson, M. Eimer, W. Rudisill, S. Steel, J. Witten, N. Burk, C. Peterson, M. Dvorak, and W. Robinson, who assisted with the Landsat image processing. D. Butman and P. Raymond provided the RSSA estimates from (*1*). A. Pietroniro and E. Klyszejko provided the Water Survey of Canada stream gauge data. **Funding:** This work was funded by a NASA NIP grant (NNX12AQ77G) and a NASA THP grant (NNX14AD82G) to T.M.P. G.H.A. was partially supported by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA, including grants from the NASA SERVIR Applied Sciences Team (NNH15ZDA001N-SERVIR) and the NASA SWOT Science team (NNH15ZDA001N-SWOT) to C. H. David. **Author contributions:** T.M.P. conceived of the GRWL Database. G.H.A. produced the GRWL Database, performed the statistical analysis, drafted the figures, and wrote the manuscript with input from T.M.P. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The GRWL vector product and water mask are freely available for download from Zenodo (*31*), and the code used in the RSSA analysis and production of the figures is available at https://github.com/geoallen/RSSA/.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/585/suppl/DC1
Materials and Methods
Figs. S1 to S7
References (*32*–*42*)
Data S1

20 January 2018; accepted 5 June 2018
Published online 28 June 2018
10.1126/science.aat0636

NEUROSCIENCE

Opioid prescribing decreases after learning of a patient's fatal overdose

Jason N. Doctor[1]*, Andy Nguyen[1], Roneet Lev[2], Jonathan Lucas[3], Tara Knight[1], Henu Zhao[1], Michael Menchine[4]

Most opioid prescription deaths occur among people with common conditions for which prescribing risks outweigh benefits. General psychological insights offer an explanation: People may judge risk to be low without available personal experiences, may be less careful than expected when not observed, and may falter without an injunction from authority. To test these hypotheses, we conducted a randomized trial of 861 clinicians prescribing to 170 persons who subsequently suffered fatal overdoses. Clinicians in the intervention group received notification of their patients' deaths and a safe prescribing injunction from their county's medical examiner, whereas physicians in the control group did not. Milligram morphine equivalents in prescriptions filled by patients of letter recipients versus controls decreased by 9.7% (95% confidence interval: 6.2 to 13.2%; $P < 0.001$) over 3 months after intervention. We also observed both fewer opioid initiates and fewer high-dose opioid prescriptions by letter recipients.

The United States is in the grips of its worst drug crisis in history, driven by twin epidemics of illicit and prescription opioid use. Opioid overdoses resulted in 351,630 deaths from 1999 to 2016, with deaths increasing from 2015 to 2016 for both illicit and prescription opioid users (*1*). In addition, more than 1.9 million Americans suffer from opioid addiction (*2*). Before the introduction of extended-release and long-acting opioids in the 1990s, long-term, high-dose opioid therapy was rare. The use of these drugs for cancer pain led to calls to address chronic noncancer pain with the same agents. Opioid proponents discounted historical evidence for opioid harms by frequently citing studies on small convenience samples (*3*, *4*). For example, there were more than 600 citations of a five-sentence letter to the editor published in the *New England Journal of Medicine* in 1980 titled "Addiction rare in patients treated with narcotics" (*3*), most in support of opioid safety (*5*). These and other factors may have encouraged unwarranted opioid prescribing.

In this randomized experiment, we aimed to counter factors that may promote high levels of opioid prescribing. We evaluated the effect of a personal letter from the medical examiner notifying clinicians of a controlled substance overdose drug death in their practice and reiterating guidelines for safe prescribing. We intervened on clinicians and allied health professionals with scheduled drug prescribing privileges in California. These individuals had prescribed a schedule II, III, or IV drug to a person who died as a result of a schedule II, III, or IV accidental overdose between the period of 1 July 2015 and 30 June 2016 in San Diego County. Prescriptions had to have been filled less than 12 months before the decedent's day of death.

Our study was a decedent-cluster randomized trial, in which clusters of individuals, rather than single persons, are randomly allocated to intervention groups. A decedent cluster is a distinct set of clinicians who wrote scheduled drug prescriptions to a person who suffered a fatal scheduled drug overdose. Decedent clusters were randomly assigned to either a control condition or a group receiving a letter signed by the Chief Deputy Medical Examiner of San Diego County (see supplementary text S1) that notified them of a death in their practice. The letter identified the decedent by name, address, and age; outlined the annual number and types of prescription drug deaths seen by the medical examiner; discussed the value of and way to access the state's prescription drug monitoring program; and discussed five U.S. Centers for Disease Control and Prevention (CDC) guideline–recommended safe prescribing strategies.

We abstracted data from the Controlled Substance Utilization Review and Evaluation System (CURES). This database provided a comprehensive record of opioids dispensed at California pharmacies to civilian, non–U.S. Department of Veterans Affairs, and non-institutionalized patients treated by clinicians in our sample. Descriptive and inferential statistics were carried out with the Stata software (*6*). The *cmp* command in Stata was used to compute a difference-in-differences estimator within a mixed-model two-part linear regression analysis (*7*). The difference-in-differences estimator compared the average change over time in milligram morphine equivalents (MMEs) dispensed for prescribers in the intervention group with the average change over time for prescribers in the control group. The natural log transformation of MMEs ensured a normal distribution of data. A two-part regression model has a discrete component and a continuous component: A binary event identifying whether there are opioid fill(s) in a prescriber's name each day represents the discrete part of the model, and the natural log of MMEs dispensed on days when opioids are filled in the name of a prescriber represents the continuous part. We also evaluated high-dosage prescriptions (≥50 or >90 MMEs per day) and the probability that a prescription fill was a "new start" (a new person entering the database).

Sample size calculations, described in the supplementary materials, indicate that we would need 65 decedents per study arm to have an 80% chance to detect an effect. This analysis assumes a 5% difference in daily MME prescribing between groups. Figure S1 illustrates the study timeline. Figure 1 describes the flow of decedent and prescriber identification, intervention allocation, follow-up, and data analysis. The medical examiner investigated 220 deaths in San Diego County between 1 July 2015 and 30 June 2016 for which a schedule II, III, or IV prescription drug was the primary or contributing cause of death. Of this group, 170 decedents (77.3%) had one or more prescriptions found in CURES 12 months before their death. Table S1 presents observed frequencies within randomization strata variables (cause of death and whether decedents received prescriptions from a clinician with multiple deaths), and Table 1 displays decedent characteristics. There were 861 prescribers to the 170 decedents whose deaths were caused by overdoses of schedule II, III, or IV drugs. Of these, 725 had prescribed to only one decedent, and 136 had prescribed to multiple decedents. There were, on average, 5.5 (±5.4) prescribers per decedent and 1.2 (±0.6) decedents per prescriber (numbers in parentheses indicate standard deviations). Sample characteristics of clinicians are presented in Table 2.

We observed 1,279,691 prescriptions filled during the study period. Using the dates of prescriptions dispensed, we computed daily MMEs for prescribers. Table 3 shows the average daily MMEs dispensed per prescriber in each of intervention and control groups 3 months before and 1 to 4 months after letters were sent. In the control group, opioid prescribing increased from 71.6 MMEs daily before intervention to 71.7 MMEs daily after intervention [i.e., +0.1 MME; 95% confidence interval (CI): −2.8 to 3.2 MMEs], whereas intervention group prescribing decreased from 72.5 to 65.7 MMEs (−6.8 MMEs; 95% CI: −9.9 to −3.8 MMEs) per prescriber per day. MMEs prescribed by clinicians in the intervention group decreased by 9.7% (95% CI: 6.2 to 13.2%) compared with control-prescribed MMEs over a period beginning 1 month and ending 4 months after the day the letters were mailed ($P < 0.001$). Letter recipients were 7% (95% CI: 2 to 11%) less likely than control prescribers to start a new patient on opioids ($P < 0.01$). We also

[1]Schaeffer Center for Health Policy and Economics, University of Southern California, Los Angeles, CA, USA. [2]Emergency Department, Scripps Mercy Hospital San Diego, San Diego, CA, USA. [3]Department of Medical Examiner-Coroner, Los Angeles County, Los Angeles, CA, USA. [4]Department of Emergency Medicine, University of Southern California, Los Angeles, CA, USA.
*Corresponding author. Email: jdoctor@usc.edu

observed a significant reduction in high-dose prescribing in the intervention group compared with the control group: a 3% decrease for 50 MME daily doses ($P < 0.05$) and a 4.5% decrease for 90 MME daily doses ($P < 0.05$) dispensed. There was no difference in the proportion of prescribers in the intervention or control group who made substantial (>20%) reductions in opioid prescribing in the postintervention period ($z = 1.279$, $P > 0.05$). Sensitivity analysis that transformed MMEs ensured normality, and a lag-correlation analysis to ensure independence of errors had no effect on results (fig. S2).

A simple letter, supportive in tone, to inform clinicians of a scheduled drug harm to their patient resulted in fewer subsequent opioids dispensed by those clinicians. Thus far, traditional state regulatory approaches to limiting opioid prescribing have not achieved great success. For example, Meara and colleagues found that adoption of controlled substance laws at the state level was not associated with reductions in potentially hazardous use of opioids or overdose among disabled Medicare beneficiaries (*8*). Another effort to send letters to potential high prescribers of controlled substances (individual health care clinicians who were at the 99.7th percentile of prescribing volume among prescribers of schedule II drugs in Medicare Part D) found no effect (*9*). In comparison, the success of the approach used in our study—notifying clinicians of a single fatal overdose—may have a number of explanations. First, people rely upon knowledge that is impactful, recent, and easy to retrieve from memory when judging probabilities and making decisions (*10*, *11*). Decisions to avoid harms could occur more frequently after receipt of the letter, because the effects of opioid harms are available to memory. Second, clinicians are also disproportionately exposed to patients who return to their clinics uneventfully for an opioid refill. The letter may alert clinicians to those patients who do not return, owing to death via an overdose. Third, individual behavior improves when outside persons attend to it (*12*). Clinicians may prescribe with greater care when they perceive that they are being watched, particularly by figures of authority (*13*). A message communicating a patient's overdose death from the medical examiner may have particular weight. Such information appears to encourage more cautious prescribing without restricting clinician freedom to prescribe opioids through mandated prescribing limits. Mandated limits do not account for individual patient circumstances that may arise in the course of care. We observed modest prescribing reductions, which suggest that clinicians exercised greater caution with opioids rather than abandoning use.

Brief exposure to opioids in opioid-naive persons makes long-term opioid use more likely (*14*, *15*), and excess prescription opioids (in medicine cabinets or diverted) are a source of misuse and are linked to transition to heroin (*16*, *17*). Correcting course in prescribing after learning of a patient's death from a prescription overdose may lessen the effect of the aforementioned harms.

Although the generalizability of these findings may be limited, San Diego County is a diverse county with a broad representation of providers and which constitutes ~1% of the U.S. population. The control condition involved no increased awareness of opioids or education; however, our study occurred 5 months after the U.S. Surgeon General issued CDC guideline pocket cards to

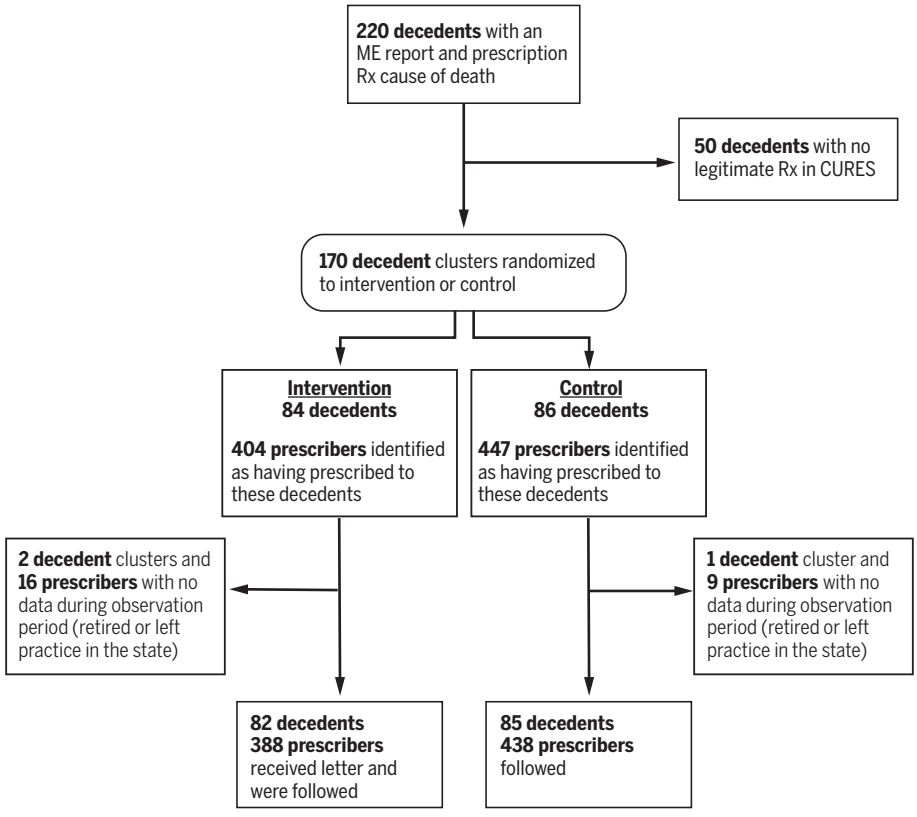


Fig. 1. Consort diagram. ME, medical examiner.

Table 1. Decedent characteristics. OTC, over the counter; *n*, number of clinicians.

Characteristic	Randomization group		Statistic	*P* value
	Letter (*n* = 82)	Control (*n* = 85)		
Age (±SD)	49.75 (11.15)	48.21 (14.85)	$t = 0.059$	0.954
Male	53 (65%)	44 (52%)	$\chi^2(1) = 0.591$	0.442
Race				
Black	6 (7%)	8 (9%)		
Hispanic	8 (10%)	4 (5%)		
Native American	0	1 (1%)	$\chi^2(5) = 4.752$*	0.447*
Asian/Pacific Islander	0	1 (1%)		
Non-Hispanic White	65 (79%)	70 (82%)		
Other	3 (4%)	1 (1%)		
Cause of death				
Prescription	41 (50%)	53 (62%)		
Prescription and illicit	27 (33%)	15 (18%)		
Prescription and alcohol	10 (12%)	13 (15%)	$\chi^2(5) = 5.300$†	0.380†
Prescription, illicit, and alcohol	2 (2%)	2 (2%)		
Prescription and OTC	1 (<1%)	1 (<1%)		
Prescription, alcohol, and OTC	1 (<1%)	1 (<1%)		

*For all races. †For all causes of death.

Table 2. Prescriber characteristics.

Professional practice	Randomization group		Statistic	*P* value
	Letter	Control		
Medical doctor (MD)	277	315		
Doctor of osteopathy (DO)	30	29		
Nurse practitioner (NP)	24	26		
Physician assistant (PA)	44	48	$\chi^2(4) = 1.173$*	0.883*
Dentistry (DDS/DND)	13	20		
Total	388	438		

*For all professional practices.

Table 3. Adjusted daily average milligram morphine equivalents (MMEs) dispensed per prescriber among persons randomized to the intervention or control groups. Values in parentheses are 95% CIs with 5% trimmed means.

Parameter	Randomization group	
	Letter	Control
Prescribers followed	388	438
Preintervention	72.5 (71.3 to 73.7)	71.6 (70.3 to 72.8)
Postintervention	65.7 (63.8 to 67.5)	71.7 (70.0 to 73.5)
Increment (pre- to post-)	−6.8 (−9.9 to −3.8)	0.1 (−2.8 to 3.2)
Difference in increment	−6.9 (−13.1 to −1.0)	
P value	0.001	

all U.S. clinicians, including those in our control arm. We do not address appropriate or inappropriate prescribing at the patient level.

Judicious prescribing represents only one of the components necessary to correct the missteps caused by overly enthusiastic use of opioids to alleviate pain. Access to medication-assisted therapy, counseling, and naloxone for resuscitation after overdose and efforts to address social determinants responsible for increased opioid use all play equally important roles in ending the crisis.

The intervention described here is scalable. Each county in the United States reports prescription opioid deaths to the National Center for Health Statistics, and each state maintains a vital records death file. Each state contains a prescription drug monitoring program that tracks prescriptions to decedents. It is thus feasible to "close the loop" on deaths by encouraging safe prescribing habits through the use of behavioral insights.

REFERENCES AND NOTES

1. P. Seth, L. Scholl, R. A. Rudd, S. Bacon, *MMWR Morb. Mortal. Wkly. Rep.* **67**, 349–358 (2018).
2. B. Han *et al.*, *Ann. Intern. Med.* **167**, 293–301 (2017).
3. J. Porter, H. Jick, *N. Engl. J. Med.* **302**, 123 (1980).
4. R. K. Portenoy, K. M. Foley, *Pain* **25**, 171–186 (1986).
5. P. T. M. Leung, E. M. Macdonald, M. B. Stanbrook, I. A. Dhalla, D. N. Juurlink, *N. Engl. J. Med.* **376**, 2194–2195 (2017).
6. Stata Statistical Software, Release 14 (StataCorp, 2015).
7. J. F. McDonald, R. A. Moffitt, *Rev. Econ. Stat.* **62**, 318–321 (1980).
8. E. Meara *et al.*, *N. Engl. J. Med.* **375**, 44–53 (2016).
9. A. Sacarny, D. Yokum, A. Finkelstein, S. Agrawal, *Health Aff. (Millwood)* **35**, 471–479 (2016).
10. A. Tversky, D. Kahneman, *Cognit. Psychol.* **5**, 207–232 (1973).
11. W. Bruine de Bruin *et al.*, *J. Exp. Psychol. Appl.* **22**, 261–271 (2016).
12. J. McCambridge, J. Witton, D. R. Elbourne, *J. Clin. Epidemiol.* **67**, 267–277 (2014).
13. R. B. Cialdini, *Curr. Dir. Psychol. Sci.* **12**, 105–109 (2003).
14. A. Shah, C. J. Hayes, B. C. Martin, *Morb. Mortal. Wkly. Rep.* **66**, 265–269 (2017).
15. M. L. Barnett, A. R. Olenski, A. B. Jena, *N. Engl. J. Med.* **376**, 663–673 (2017).
16. R. A. Pollini *et al.*, *Subst. Abuse Rehabil.* **2**, 173–180 (2011).
17. T. J. Cicero, M. S. Ellis, H. L. Surratt, S. P. Kurtz, *JAMA Psychiatry* **71**, 821–826 (2014).

ACKNOWLEDGMENTS

We thank K. Pfeifer, D. Meeker, and M. Sullivan for valuable feedback on the manuscript. We thank M. Small and T. Farrales at the U.S. Department of Justice for assistance with CURES. The Institutional Review Board at the University of Southern California approved all study procedures and waived informed consent for participants under HHS regulations at 45 CFR 46.116(c), as the study was evaluating a county public service safe prescribing program. Protected health information was not shared outside of the San Diego County Medical Examiner's Office and its employees and volunteers working in an official capacity. The trial was registered at https://clinicaltrials.gov (NCT02790476) before commencing. **Funding:** This work was supported by the California Health Care Foundation (grant 19413) and the National Institute on Aging at the National Institutes of Health (grant R21-AG057395-01). **Author contributions:** Study concept and design: J.N.D., A.N., R.L., J.L., and M.M. Acquisition, analysis, or interpretation of data: J.N.D., A.N., and H.Z. Drafting of the manuscript: J.N.D. Critical revision of the manuscript for important intellectual content: J.N.D., T.K., R.L., J.L., M.M., A.N., and H.Z. Statistical analysis: A.N. and H.Z. Obtained funding: J.N.D., R.L., and M.M. Administrative, technical, or material support: T.K. Study supervision: J.N.D., R.L., J.L., and M.M. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data in our project were extracted from CURES 2.0 (Controlled Substance Utilization Review and Evaluation System), a database of schedule II, III, and IV controlled substance prescriptions dispensed in California. The CURES database serves the state's public health, regulatory oversight, and law enforcement agencies. Data may be made available through data use agreements among researchers interested in using the data, the County of San Diego Medical Examiner's Office, and the California State Department of Justice. A.N. and J.L. had full access, whereas J.N.D. and H.Z. had deidentified access to all of the data in the study. These authors take responsibility for the integrity of the data and the accuracy of the data analysis.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/588/suppl/DC1
Materials and Methods
Supplementary Text S1
Figs. S1 and S2
Tables S1 to S3
References (*18*–*22*)

1 March 2018; accepted 5 July 2018
10.1126/science.aat4595

EVOLUTION

Ancient convergent losses of *Paraoxonase 1* yield potential risks for modern marine mammals

Wynn K. Meyer[1], Jerrica Jamison[2], Rebecca Richter[3], Stacy E. Woods[4]*, Raghavendran Partha[1], Amanda Kowalczyk[1], Charles Kronk[2], Maria Chikina[1], Robert K. Bonde[5], Daniel E. Crocker[6], Joseph Gaspard[7], Janet M. Lanyon[8], Judit Marsillach[3], Clement E. Furlong[3,9], Nathan L. Clark[1,10]†

Mammals diversified by colonizing drastically different environments, with each transition yielding numerous molecular changes, including losses of protein function. Though not initially deleterious, these losses could subsequently carry deleterious pleiotropic consequences. We have used phylogenetic methods to identify convergent functional losses across independent marine mammal lineages. In one extreme case, *Paraoxonase 1* (*PON1*) accrued lesions in all marine lineages, while remaining intact in all terrestrial mammals. These lesions coincide with PON1 enzymatic activity loss in marine species' blood plasma. This convergent loss is likely explained by parallel shifts in marine ancestors' lipid metabolism and/or bloodstream oxidative environment affecting PON1's role in fatty acid oxidation. PON1 loss also eliminates marine mammals' main defense against neurotoxicity from specific man-made organophosphorus compounds, implying potential risks in modern environments.

As the ancestors of aquatic marine mammals adopted obligate aquatic lifestyles, they evolved many adaptive changes, such as those that improved locomotion and respiration in and perception of their new environment (*1*–*3*). Many of these morphological and physiological changes occurred in parallel in distinct lineages of marine mammals, including cetaceans, pinnipeds, and sirenians. Although convergent trait changes are frequently adaptive, environmental transitions can also result in nonadaptive convergent trait loss due to release from functional constraint. Examples of convergently reduced or lost traits include olfaction in marine mammals (*4*–*6*), bitter taste receptors in carnivorous tetrapods (*7*), and eyes in subterranean species (*8*–*10*). Any convergent evolutionary change in the context of a given environment can carry negative consequences in a different environment as a result of pleiotropy (one genetic locus influencing multiple phenotypes).

To characterize how mammals responded to selective pressures imposed by the marine environment, we identified genes that convergently lost function in marine mammals. We identified candidate pseudogenes with observed early stop codons and/or frameshifts (genetic lesions) in 58 eutherian mammals' genomes in a 100-way vertebrate alignment (*11*). Using our predicted pseudogene calls, we then tested, for each gene, whether its pattern of functional loss was better explained by a model with one loss rate throughout the mammalian phylogeny or by a model in which the loss rate was dependent upon the terrestrial or marine state of a given branch in a likelihood ratio test (LRT) (*12*). To ensure that our results were not strongly influenced by errors in pseudogene calling, we performed manual checks of lesion calls against reference genomes for our top genes, along with comparisons of pseudogene calls at highly conserved genes for marine and terrestrial species (*13*). We used simulations to estimate empirical gene-specific *P* values and study-wide (multiple-test-corrected) false discovery rates (FDR) for all genes (*13*) (Table 1 and table S1). The set of genes with the strongest evidence for a higher loss rate on marine lineages was strongly enriched for functions related to chemosensation, driven by many olfactory and taste receptors (tables S2 to S5). These results are consistent with previous behavioral, anatomical, and genetic studies indicating a reduction of smell and taste in marine mammals (*5*, *14*, *15*).

We also observed a notable pattern of convergent loss in the marine environment at *Paraoxonase 1* (*PON1*) (Table 1) (*13*). *PON1* encodes a bloodstream enzyme that reduces oxidative damage to lipids in low- and high-density lipoprotein (LDL and HDL) particles, potentially preventing atherosclerotic plaque formation (*16*, *17*) (Fig. 1A). PON1 also hydrolyzes the oxon forms of specific organophosphate compounds, such that it is the main line of defense against some man-made pesticide by-products,

[1]Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA, USA. [2]Dietrich School of Arts and Sciences, University of Pittsburgh, Pittsburgh, PA, USA. [3]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA. [4]Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. [5]Wetland and Aquatic Research Center, U.S. Geological Survey, Gainesville, FL, USA. [6]Department of Biology, Sonoma State University, Rohnert Park, CA, USA. [7]Pittsburgh Zoo and PPG Aquarium, Pittsburgh, PA, USA. [8]School of Biological Sciences, The University of Queensland, St. Lucia, QLD 4072, Australia. [9]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [10]Pittsburgh Center for Evolutionary Biology and Medicine, University of Pittsburgh, Pittsburgh, PA, USA.
*Present address: Natural Resources Defense Council, Washington, DC, USA.
†Corresponding author. Email: nclark@pitt.edu

Table 1. Top 10 manually validated genes with evidence for marine-specific loss. Loss rates represent the inferred instantaneous rates of transition from functional gene (1) to pseudogene (0) per unit branch length under the relevant model in BayesTraits (*12*, *13*), restricted to a maximum value of 100 (the default).

Gene	Loss rate (independent)	Marine loss rate (dependent)	Terrestrial loss rate (dependent)	LRT statistic	Empirical *P* value	FDR	Description of gene product
PON1	0.672	49.7	0	22.24	3.08×10^{-6}	0.0154	Paraoxonase 1
OR10Z1	1.15	100	0.467	19.99	7.25×10^{-6}	0.0201	Olfactory receptor
OR8D4	1.25	100	0.510	19.21	1.60×10^{-5}	0.0201	Olfactory receptor
TAS2R1	1.32	100	0.535	19.20	1.60×10^{-5}	0.0201	Taste receptor
OR1F2P	2.03	100	1.18	15.86	5.40×10^{-5}	0.0831	Olfactory receptor
GSTM1	1.48	100	0.762	15.82	3.90×10^{-5}	0.0831	Glutathione *S*-transferase mu 1
OR6K2	2.02	100	1.22	15.79	4.50×10^{-5}	0.0831	Olfactory receptor
OR51D1	1.13	49.3	0.466	15.59	8.60×10^{-5}	0.0831	Olfactory receptor
TAAR5	1.17	48.2	0.484	15.16	9.90×10^{-5}	0.0936	Trace amine–associated receptor 5
OR4C13	1.77	100	0.915	14.88	7.00×10^{-5}	0.0972	Olfactory receptor

including chlorpyrifos oxon and diazoxon (Fig. 1B) (*18*). The *PON1* coding sequence contains genetic lesions in the cetacean, pinniped, and sirenian lineages but is intact in all 53 terrestrial mammal genomes surveyed (Fig. 1C and table S1).

To estimate when PON1 function was lost in the three marine mammal clades, we obtained *PON1* sequences for 14 additional species, including three cetaceans, the dugong, and two pinnipeds, and we estimated evolutionary rates across the mammalian phylogeny (*13*) (Fig. 1C and fig. S1). We observed shared genetic lesions among all sequenced cetaceans and a different shared lesion in sirenians (fig. S2), and the inferred ratio of nonsynonymous to synonymous substitutions (d_N/d_S) was not significantly different from one on the ancestral branches of both clades (cetacean ancestor d_N/d_S = 1.09, P = 0.79; sirenian ancestor d_N/d_S = 1.20, P = 0.57). This suggests that PON1 lost functional constraint in the ancestral cetacean lineage soon after its split with the ancestral hippopotamid lineage, approximately 53 million years (Ma) ago [95% confidence interval (CI) lower bound, 34.5 Ma ago] (*13*, *19*). In sirenians, functional loss occurred soon after the split with the ancestral elephantid lineage, approximately 64 Ma ago (lower bound, 41.7 Ma ago) (*19*).

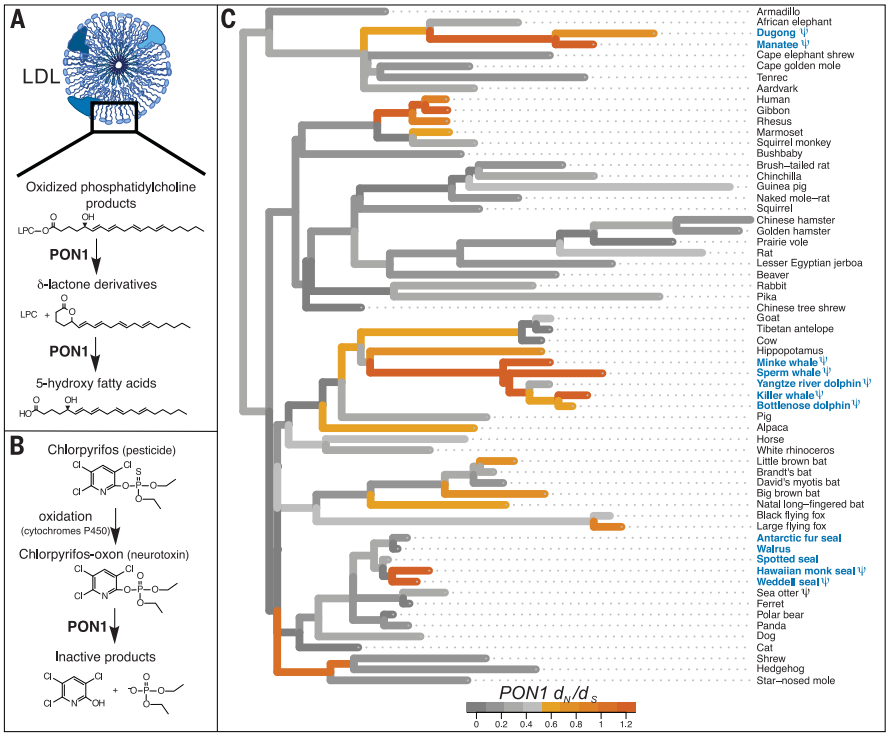


Fig. 1. PON1 functions and evolutionary history. Illustration of PON1's proposed roles in (**A**) preventing oxidative damage to LDL and HDL (*16*, *17*) and (**B**) detoxifying the oxon by-product or metabolite of a common organophosphorus pesticide, chlorpyrifos (*27*). LPC, lysophosphatidylcholine. (**C**) Evolutionary rate of the *PON1* coding sequence across the phylogeny of 62 eutherian mammals. Branch lengths represent d_N, and colors represent d_N/d_S (see the color legend). d_N/d_S values greater than 1.2 were set to 1.2. Blue, marine species; ψ, genetic lesion(s) present.

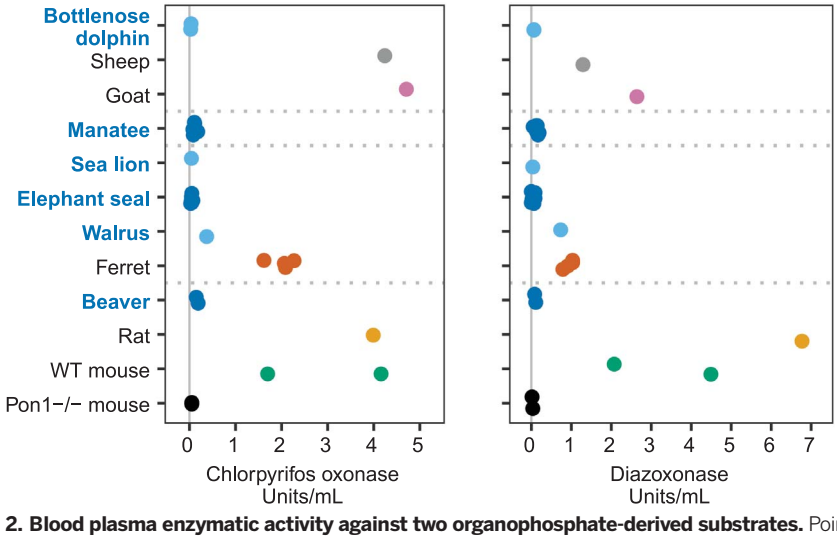


Fig. 2. Blood plasma enzymatic activity against two organophosphate-derived substrates. Points represent rates of hydrolysis of chlorpyrifos oxon (left) or diazoxon (right) in micromoles per minute per milliliter for plasma from marine and semiaquatic species (in blue) and terrestrial out-groups. Values for sheep, goats, and rats are from Furlong *et al*. (*35*), who performed assays under the same experimental conditions used in this study. Vertical solid lines indicate no activity, and horizontal dashed lines separate species from different evolutionary clades. Control assays of alkaline phosphatase activity show that samples were not degraded (fig. S3). WT, wild-type.

In pinnipeds, we observed clear evidence of PON1 functional loss only among a subset of species within the family Phocidae, wherein Weddell seal and Hawaiian monk seal *PON1* sequences contained nonshared genetic lesions (fig. S2). Because these branches are short, it is difficult to estimate precisely when functional loss occurred in pinnipeds; however, there was likely at least one loss since the Phocidae-Otarioidea split approximately 21 Ma ago (95% CI, 0 to 21 Ma ago). This incomplete loss may reflect either a difference between the selective environments experienced by pinnipeds and those experienced by other marine mammals or pinnipeds' more recent colonization of the marine environment (pinnipeds, 24 Ma ago; cetaceans, 44.7 to 37.3 Ma ago; sirenians, 47.1 to 43.9 Ma ago) (*20*).

PON1's functional loss in marine mammals may be related to its role in lipid metabolism via fatty acid beta-oxidation (*21*) (tables S6 and S7). The diets of both herbivorous and carnivorous aquatic mammals contain a higher proportion of ω-3 relative to ω-6 polyunsaturated fatty acids (PUFAs) than those of terrestrial mammals (*22*), and these PUFAs differ in their capacity to sustain oxidative damage (*23*). Marine and terrestrial mammals also have vastly different antioxidant profiles (*24*, *25*), presumably because of the extreme oxidative stress experienced during diving, with repeated cycles of hypoxia and reperfusion. Rewiring of either lipid metabolism or antioxidant networks in ancient marine mammals may have obviated the function of PON1. Supporting the antioxidant hypothesis, the Weddell seal, which carries *PON1* lesions, is one of the longest-diving pinnipeds known, in contrast to the shorter-diving walrus and Antarctic fur seal, which lack lesions but share an aquatic diet (*26*). However, two semiaquatic mammals, the sea otter and the beaver, which are more moderate divers (*26*), also have either lesions or substitutions at sites predicted to be necessary for PON1 function (fig. S2 and table S8).

Whatever the cause, loss of PON1 function may carry negative pleiotropic consequences for the health of marine mammals repeatedly exposed to man-made organophosphate compounds. PON1 alone is protective against the

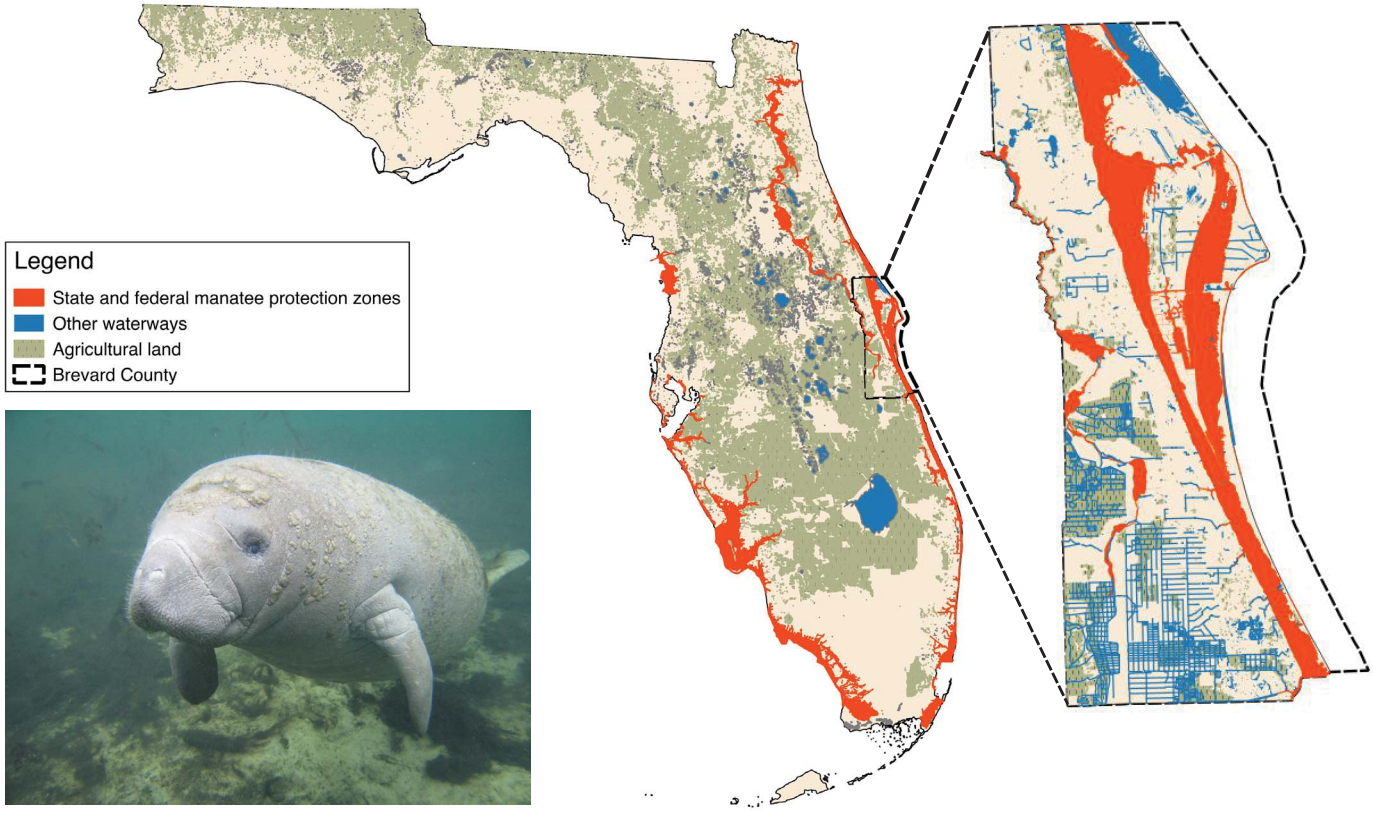


Fig. 3. The manatee and the adjacency of its habitat to agricultural land use. (Left) Florida manatee. (Center) Manatee protection zones and agricultural land in Florida. (Right) Manatee protection zones, waterways, and agricultural land in Brevard County.

highly toxic oxon forms of the heavily used pesticides chlorpyrifos and diazinon; these oxons are formed from the parent compounds in the environment and in vivo by cytochromes P450 (*27*) (Fig. 1B). We tested blood plasma from six marine and semiaquatic species for the capacity to hydrolyze these and other PON1 substrates (Fig. 2 and fig. S3). The plasma from all but one of the assayed marine and semiaquatic species showed activity levels against the PON1 substrates that more closely resembled those of the *Pon1* knockout (*Pon1*$^{-/-}$) mouse than those of terrestrial out-groups. Thus, the genetic deterioration of *PON1* has left these species without a mechanism to break down specific neurotoxic compounds.

Given the sensitivity of *Pon1*$^{-/-}$ mice to organophosphate exposure (*28*), the inability of most marine mammal plasma to detoxify organophosphates suggests the potential for neurotoxicity if sufficient levels of these compounds accumulate in these animals' habitats or food sources. In Florida, agricultural use of organophosphate pesticides is common, and runoff can drain into manatee habitats. In Brevard County, where an estimated 70% of Atlantic coast manatees migrate or seasonally reside (*29, 30*), agricultural lands frequently abut manatee protection zones and waterways (Fig. 3). Limited sampling upstream of Manatee Bay has measured levels of chlorpyrifos as high as 0.023 μg/liter (*31*), and levels could be much higher directly after pesticide applications (*32*). Dugongs may be at risk of exposure to organophosphorus pesticides that are used in the sugarcane industry along the Queensland coast of Australia and have been detected at 5 to 270 pg/liter in coastal river systems (*33*). Carnivorous marine mammals may also ingest these compounds through their diets of invertebrates and fish, which have shown evidence of bioaccumulation of organophosphates in Arctic populations (*34*). In order to improve our understanding of the extent of exposure and attendant risk marine mammals face, we recommend increased monitoring of marine mammal habitats, as well as the testing of tissues from deceased animals for biomarkers of organophosphate exposure.

The presence of these potential risks to many marine mammals due to their loss of PON1 function provides a clear example of the tradeoffs possible in evolution: although PON1 functional loss was not deleterious and may even have been beneficial in ancestral marine environments, it may carry detrimental fitness consequences in modern environments.

REFERENCES AND NOTES

1. F. E. Fish, L. E. Howle, M. M. Murray, *Integr. Comp. Biol.* **48**, 788–800 (2008).
2. D. Wartzok, D. R. Ketten, in *Biology of Marine Mammals*, J. Reynolds, S. Rommel, Eds. (Smithsonian Institution Press, 1999), pp. 117–175.
3. V. Sharma *et al.*, *Nat. Commun.* **9**, 1215 (2018).
4. M. R. McGowen, C. Clark, J. Gatesy, *Syst. Biol.* **57**, 574–590 (2008).
5. M. L. Bills, thesis, University of Florida, Gainesville, FL (2011).
6. M. Chikina, J. D. Robinson, N. L. Clark, *Mol. Biol. Evol.* **33**, 2182–2192 (2016).
7. D. Li, J. Zhang, *Mol. Biol. Evol.* **31**, 303–309 (2014).
8. M. Protas, M. Conrad, J. B. Gross, C. Tabin, R. Borowsky, *Curr. Biol.* **17**, 452–454 (2007).
9. W. R. Jeffery, *Annu. Rev. Genet.* **43**, 25–47 (2009).
10. R. Partha *et al.*, *eLife* **6**, e25884 (2017).
11. UCSC Genome Browser, http://genome.ucsc.edu/.
12. M. Pagel, A. Meade, BayesTraits (2013).
13. Materials and methods are available as supplementary materials.
14. L. Marino, *Anat. Rec.* **290**, 694–700 (2007).
15. M. R. McGowen, J. Gatesy, D. E. Wildman, *Trends Ecol. Evol.* **29**, 336–346 (2014).
16. M. Rosenblat, M. Aviram, *Biofactors* **35**, 98–104 (2009).
17. M. I. Mackness, S. Arrol, P. N. Durrington, *FEBS Lett.* **286**, 152–154 (1991).
18. W. F. Li *et al.*, *Pharmacogenetics* **10**, 767–779 (2000).
19. R. W. Meredith *et al.*, *Science* **334**, 521–524 (2011).
20. J. Thewissen, S. Nummela, in *Sensory Evolution on the Threshold: Adaptations in Secondarily Aquatic Vertebrates*, J. Thewissen, S. Nummela, Eds. (University of California Press, 2008), pp. 1–28.
21. C. E. Furlong, J. Marsillach, G. P. Jarvik, L. G. Costa, *Chem. Biol. Interact.* **259** (Pt B), 51–62 (2016).
22. A.-M. Koussoroplis *et al.*, *Lipids* **43**, 461–466 (2008).
23. K. Miyashita, E. Nara, T. Ota, *Biosci. Biotechnol. Biochem.* **57**, 1638–1640 (1993).
24. J. P. Vázquez-Medina, T. Zenteno-Savín, R. Elsner, R. M. Ortiz, *J. Comp. Physiol. B* **182**, 741–750 (2012).
25. N. Cantú-Medellín, B. Byrd, A. Hohn, J. P. Vázquez-Medina, T. Zenteno-Savín, *Comp. Biochem. Physiol. A* **158**, 438–443 (2011).
26. S. Mirceta *et al.*, *Science* **340**, 1234192–1234192 (2013).
27. C. E. Furlong, *J. Biochem. Mol. Toxicol.* **21**, 197–205 (2007).

PHOTO (LEFT): R. K. BONDE (2006)

28. D. M. Shih *et al*., *Nature* **394**, 284–287 (1998).
29. C. J. Deutsch *et al*., *Wildl. Monogr.* **151**, 1–77 (2003).
30. J. Martin *et al*., *Biol. Conserv.* **186**, 44–51 (2015).
31. J. F. Carriger, G. M. Rand, *Ecotoxicology* **17**, 660–679 (2008).
32. J. R. Aguirre-Rubí *et al*., *Environ. Sci. Pollut. Res. Int.* **25**, 13396–13415 (2018).
33. M. Shaw *et al*., *Mar. Pollut. Bull.* **60**, 113–122 (2010).
34. A. D. Morris *et al*., *Environ. Toxicol. Chem.* **35**, 1695–1707 (2016).
35. C. E. Furlong *et al*., *Neurotoxicology* **21**, 91–100 (2000).

ACKNOWLEDGMENTS

We thank B. Small for assistance with molecular work; K. Goulet and V. Fravel for supplying marine mammal samples; S. Lakdawala and the Lakdawala lab for supplying ferret samples; and A. Lusis, D. Shih, and A. Tward for supplying *Pon1* knockout mice; as well as all members of the Clark and Chikina labs and K. Dolan for feedback. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. government. **Funding:** This study was funded by NIH grants R01HG009299 and U54 HG008540 to N.L.C. and M.C. A.K. was supported by NIH T32 training grant T32 EB009403 as part of the HHMI-NIBIB Interfaces Initiative. C.E.F., J.M., and R.R. were supported by funds from the Biotechnology Research Gift Fund, University of Washington, Division of Medical Genetics. J.M. was supported by grant 16SDG30300009 from the American Heart Association. Dugong samples were collected with funds from the Winifred Violet Scott Foundation and the Sea World Research and Rescue Foundation. The collection of manatee samples was funded by the U.S. Geological Survey. **Author contributions:** N.L.C., M.C., C.E.F., and W.K.M. designed the study. R.K.B., D.E.C., J.G., J.M.L., and C.E.F. provided samples and reagents. J.J., J.M., and R.R. performed laboratory experiments. W.K.M., J.J., R.P., A.K., C.K., and N.L.C. performed analyses. W.K.M., S.E.W., R.P., A.K., and N.L.C. generated figures. W.K.M., C.E.F., and N.L.C. wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data reported in this paper are tabulated in the supplementary materials. Resequencing data for the *PON1* coding sequence in dugongs is available in GenBank (accession number MF197755). Scripts used in analyses are available at https://github.com/nclark-lab/MarineFxLoss.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/591/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S13
References (*36–105*)

25 August 2017; accepted 29 June 2018
10.1126/science.aap7714

KIDNEY CANCER

Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors

Matthew D. Young[1]*, Thomas J. Mitchell[1,2,3]*, Felipe A. Vieira Braga[1]*, Maxine G. B. Tran[4,5], Benjamin J. Stewart[6], John R. Ferdinand[6], Grace Collord[1,2,7], Rachel A. Botting[8], Dorin-Mirel Popescu[8], Kevin W. Loudon[6], Roser Vento-Tormo[1], Emily Stephenson[8], Alex Cagan[1], Sarah J. Farndon[1,9,10], Martin Del Castillo Velasco-Herrera[1], Charlotte Guzzo[1], Nathan Richoz[6], Lira Mamanova[1], Tevita Aho[2], James N. Armitage[3], Antony C. P. Riddick[3], Imran Mushtaq[9], Stephen Farrell[2], Dyanne Rampling[9], James Nicholson[2,7], Andrew Filby[8], Johanna Burge[2], Steven Lisgo[11], Patrick H. Maxwell[12], Susan Lindsay[11], Anne Y. Warren[2], Grant D. Stewart[2,3], Neil Sebire[9,10], Nicholas Coleman[2,13], Muzlifah Haniffa[8,14]†, Sarah A. Teichmann[1]†, Menna Clatworthy[2,6]†, Sam Behjati[1,2,7]†

Messenger RNA encodes cellular function and phenotype. In the context of human cancer, it defines the identities of malignant cells and the diversity of tumor tissue. We studied 72,501 single-cell transcriptomes of human renal tumors and normal tissue from fetal, pediatric, and adult kidneys. We matched childhood Wilms tumor with specific fetal cell types, thus providing evidence for the hypothesis that Wilms tumor cells are aberrant fetal cells. In adult renal cell carcinoma, we identified a canonical cancer transcriptome that matched a little-known subtype of proximal convoluted tubular cell. Analyses of the tumor composition defined cancer-associated normal cells and delineated a complex vascular endothelial growth factor (VEGF) signaling circuit. Our findings reveal the precise cellular identities and compositions of human kidney tumors.

Cancer cell identity is defined by morphological appearance, tissue context, and marker gene expression. Single-cell transcriptomics refines this cellular identity on the basis of a comprehensive and quantitative readout of mRNA. Precise cellular transcriptomes may reveal a tumor's cell of origin and the transcriptional trajectories underpinning malignant transformation (*1*).

We sought to define the identities of normal and cancerous human kidney cells from a catalog of 72,501 single kidney cell transcriptomes, integrated with tumor whole-genome DNA sequences (*2*). We studied Wilms tumor (n = 3 specimens), clear cell renal cell carcinoma (ccRCC) (n = 3), and papillary renal cell carcinoma (pRCC) (n = 1) in relation to healthy fetal (n = 2), pediatric (n = 3), adolescent (n = 2), and adult (n = 5) kidneys, as well as to ureters (n = 4) (table S1).

Normal tissue biopsies were taken from macroscopically normal regions of kidneys resected because of cancer (n = 10 samples) or for transplantation (n = 2 samples). We performed technical replicates of each biopsy and prepared biological replicates, where clinically permissible (table S1). We processed kidneys immediately after resection, generating single-cell solutions enriched for viable cells. We derived counts of mRNA molecules in each cell for further analyses, subject to quality control (*2*).

We split 72,501 fetal, normal, and tumor cells into immune and nonimmune cell compartments (fig. S1). Using a community detection algorithm (*2*), we further segregated transcriptomes into distinct clusters of cells (table S2). We next generated a reference map of normal mature and fetal cells, assigning an identity to each cluster, by cross-referencing cluster-defining transcripts with canonical markers curated from the literature (table S3). Ambiguous clusters were not included in the reference map and are presented in figs. S2 to S8. Highly specific cluster-defining transcripts (potential cell markers) are appended (table S4).

Among 42,809 nonmalignant cells, 37,951 mature kidney cells represented epithelial cells from distinct micro-anatomical regions of the nephron, with a large proportion of proximal tubular cells (Fig. 1, A to C, and fig. S4). Furthermore, there were fibroblasts, myofibroblasts, and vascular endothelial cells (glomerular endothelium and ascending and descending vasa recta) (Fig. 1D and fig. S2). Fetal cells (4858) grouped into developing nephron cells [ureteric bud (UB), cap mesenchyme (CM), and primitive vesicle (PV) cells] and fibroblasts, myofibroblasts, vascular endothelial cells, and ganglion cells (Fig. 2, A to C, and fig. S5).

To determine transcriptional programs underlying nephrogenesis, we identified transcription factors differentially expressed in UB cells versus CM and PV cells (Fig. 2D). Furthermore, we applied pseudotiming methods to identify transcription factors that define the transition from CM to PV (Fig. 2D). Together, these analyses identified both established and previously unknown transcription factors associated with nephron development, included as a reference for subsequent analyses of malignancy (table S5).

Having established the single-cell landscape of healthy kidneys, we characterized the cellular identities of 6333 nonimmune (fig. S7) and 17,821 immune (fig. S8) tumor cells from Wilms tumor (n = 3), ccRCC (n = 3), and pRCC (n = 1) (table S1). Children had received neoadjuvant cytotoxic treatment before nephrectomy, per British practice. Although this pretreatment reduced yield (table S6), recovered cells represent therapeutically relevant surviving cancer cells that determine the degree of adjuvant cytotoxic chemotherapy required (*3*). We used logistic regression to quantify the similarity between tumor and normal cell clusters, validated through intrinsic control populations (*2*). That is, the model found that myofibroblasts from tumors matched myofibroblasts from mature and fetal kidneys (Fig. 3A) and found no match for mast cells, a negative control population inserted into the training data.

This similarity metric may be obfuscated by the phenotypic plasticity of tumor cells. We therefore developed a method to genotype individual cancer cells from mRNA reads by using somatic copy number changes (table S7 and fig. S9) defined by whole-genome sequencing (fig. S10). We validated genotyping calls by phasing single-nucleotide polymorphisms across segments with altered copy numbers, testing for the presence of somatic single-nucleotide variants, and comparing with control populations (figs. S11 to S14).

Integrating genotyping and similarity analyses, we found that Wilms cells resembled normal fetal cells, showing that Wilms cells represent aberrant fetal cells. We found different populations of Wilms tumor that matched UB and PV cells (specific developing nephron populations) (Fig. 3A). One cluster (designated WF), composed of Wilms cancer cells and noncancerous ccRCC fibroblasts, exhibited a fibroblast-myofibroblast transcriptome. In one case, we obtained an anatomically

[1]Wellcome Sanger Institute, Hinxton CB10 1SA, UK. [2]Cambridge University Hospitals NHS Foundation Trust, Cambridge CB2 0QQ, UK. [3]Department of Surgery, University of Cambridge, Cambridge CB2 0QQ, UK. [4]UCL Division of Surgery and Interventional Science, Royal Free Hospital, London NW3 2PS, UK. [5]Specialist Centre for Kidney Cancer, Royal Free Hospital, London NW3 2PS, UK. [6]Molecular Immunity Unit, Department of Medicine, University of Cambridge, MRC Laboratory of Molecular Biology, Cambridge CB2 0QQ, UK. [7]Department of Paediatrics, University of Cambridge, Cambridge CB2 0QQ, UK. [8]Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. [9]Great Ormond Street Hospital for Children NHS Foundation Trust, London WC1N 3JH, UK. [10]UCL Great Ormond Street Hospital Institute of Child Health, London WC1N 1E, UK. [11]Human Developmental Biology Resource, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne NE1 3BZ, UK. [12]Cambridge Institute for Medical Research, University of Cambridge, Cambridge CB2 0XY, UK. [13]Department of Pathology, University of Cambridge, Cambridge CB2 1QP, UK. [14]Department of Dermatology, Royal Victoria Infirmary, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne NE1 4LP, UK.
*These authors contributed equally to this work.
†Corresponding author. Email: sb31@sanger.ac.uk (S.B.); mrc38@medschl.cam.ac.uk (M.C.); st9@sanger.ac.uk (S.A.T.); m.a.haniffa@newcastle.ac.uk (M.H.)

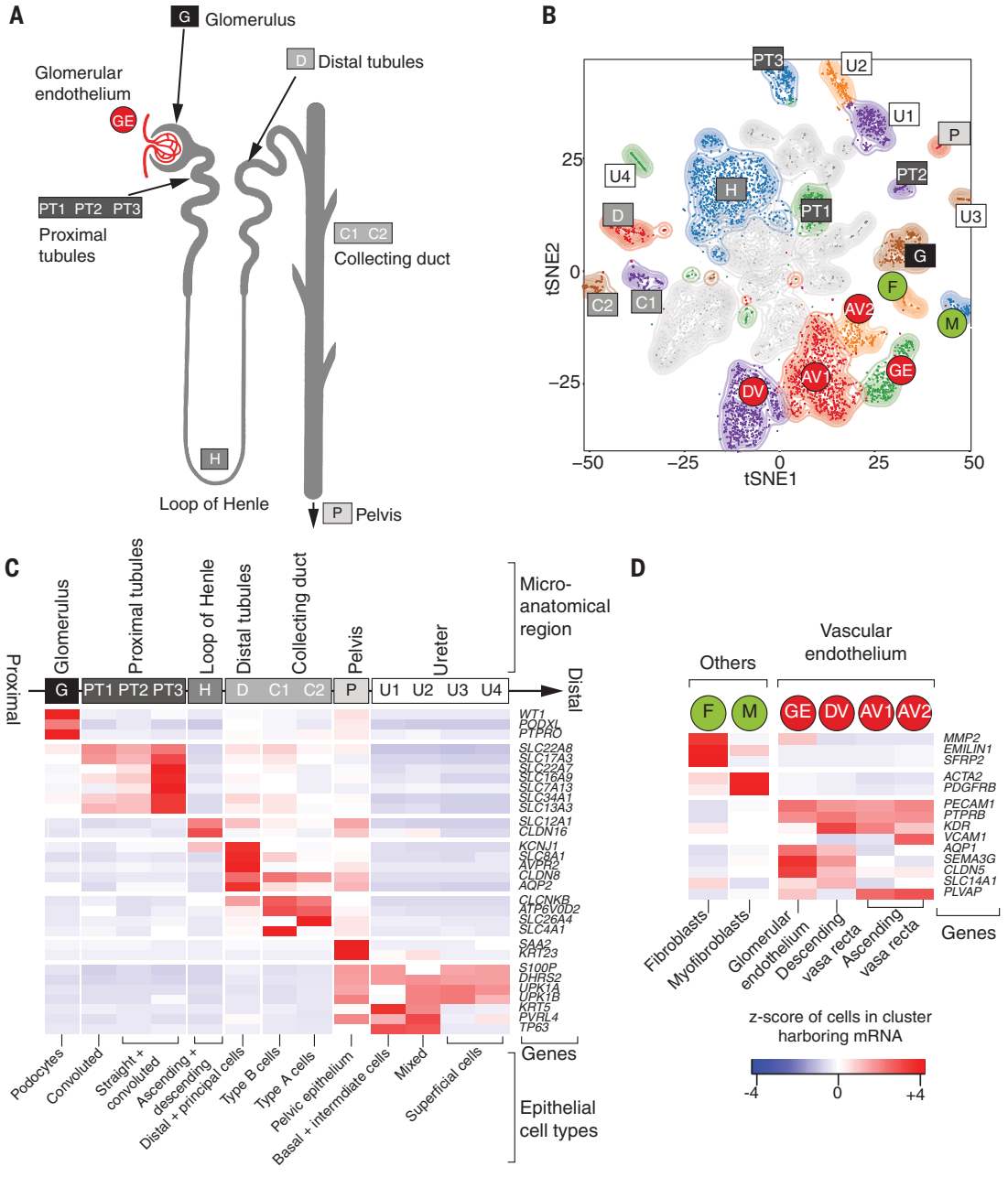


Fig. 1. Canonical cell types in normal human kidneys. (**A**) Illustration of nephron anatomy with cell clusters marked. (**B**) t-SNE (*t*-distributed stochastic neighbor embedding) representation of 8,707 normal epithelial and vascular cells. Clusters are colored, distinctively labeled, and emphasized with density contours. Ambiguous clusters are de-emphasized and fully shown in fig. S2. (**C**) Expression of canonical nephron-specific genes (table S3) in clusters from (A). Colors give the fraction of cells expressing each gene in a cluster, scaled to have a mean of 0 and SD of 1 across all clusters. (**D**) Expression of clusters in (A) not shown in (C) and their canonical genes.

separate perilobar nephrogenic rest, thought to represent a precursor lesion of Wilms tumor. We observed that, like Wilms cancer cells, the nephrogenic rest resembled UB or PV cells. This finding suggests that the potential to generate the different cell states of the fetal nephron is acquired early or was not lost by the developing Wilms cancer, although the basis for this conclusion is only one sample.

To validate the cellular identity of Wilms cells, we interrogated bulk transcriptomes of an independent series of 124 Wilms tumors for cellular signatures of UB and PV (*4*, *5*). We extracted specific markers expressed within UB or PV cells and unexpressed within nontumor cells (table S8) (*2*) and probed bulk transcriptomes for these cluster-defining transcripts. As comparators to Wilms tumor transcriptomes, we included fetal, pediatric, and adult normal tissue bulk transcriptomes (n = 135) and other childhood kidney tumors: 17 congenital mesoblastic nephroma and 65 malignant rhabdoid tumors. Corroborating the presence of these cells in Wilms tumor, signatures of PV and UB cells were seen in, and confined to, Wilms tumor and normal fetal tissues (Fig. 3B).

Placing Wilms tumor cells in pseudotime revealed two transcriptional programs emanating from the UB: one branch describing the development predominantly of nephrogenic rest cells and the other of Wilms cancer cells (Fig. 3C). The transcription factors underpinning these two programs (Fig. 3D and table S9) and normal nephrogenesis overlapped significantly ($P < 10^{-4}$; hypergeometric test). This indicates that developmental relationships exist among Wilms tumor cells that have been adopted from normal nephrogenesis. Our analyses reveal the plasticity and fetal identity of Wilms cells and transcriptionally define developmental cell states and trajectories that may harbor targetable vulnerabilities.

Next we studied ccRCC and pRCC (type 1), including one case of von Hippel–Lindau disease–related ccRCC (table S1). Matching ccRCC and

pRCC with normal mature cells, we found that they retained transcriptional features of cluster PT1, a specific subtype of convoluted proximal tubular cell (Fig. 4A). Most (six of seven) ccRCC clusters and all pRCC cells matched this particular PT1 cell, indicating that it represents an RCC cell state that transcends the diversity of RCC cells within and across tumors. Little is known about the nearest normal cell correlate of RCC, the PT1 cell, which has been found to become more abundant in inflamed renal tissue (*6*).

To validate the identity of the PT1 signature in RCC, we made use of the observation that PT1 cells were defined by SLC17A3 and VCAM1 with the absence of SLC7A13 within our data (Fig. 4B and fig. S2). We measured these transcripts in an independent series of 1019 publicly available bulk kidney tumor and normal tissue transcriptomes. High expression of SLC17A3 mRNA distinguished ccRCC and pRCC (types 1 and 2) from other types of RCC ($P < 10^{-4}$; Mann-Whitney test), whereas SLC7A13 mRNA was significantly depleted in ccRCC and pRCC bulk transcriptomes versus normal transcriptomes ($P < 10^{-4}$; Mann-Whitney test), as were mRNAs representing other regions of the nephron (Fig. 4B). VCAM1 expression, specific to PT1 within proximal tubules, was also significantly elevated across RCC bulk transcriptomes ($P < 10^{-4}$; Mann-Whitney test) (Fig. 4B), with each individual RCC tumor exhibiting PT1 features (fig. S15). Confocal microscopy demonstrated colocalization of VCAM1 and SLC17A3 in CA9^{+} cells, CA9 being

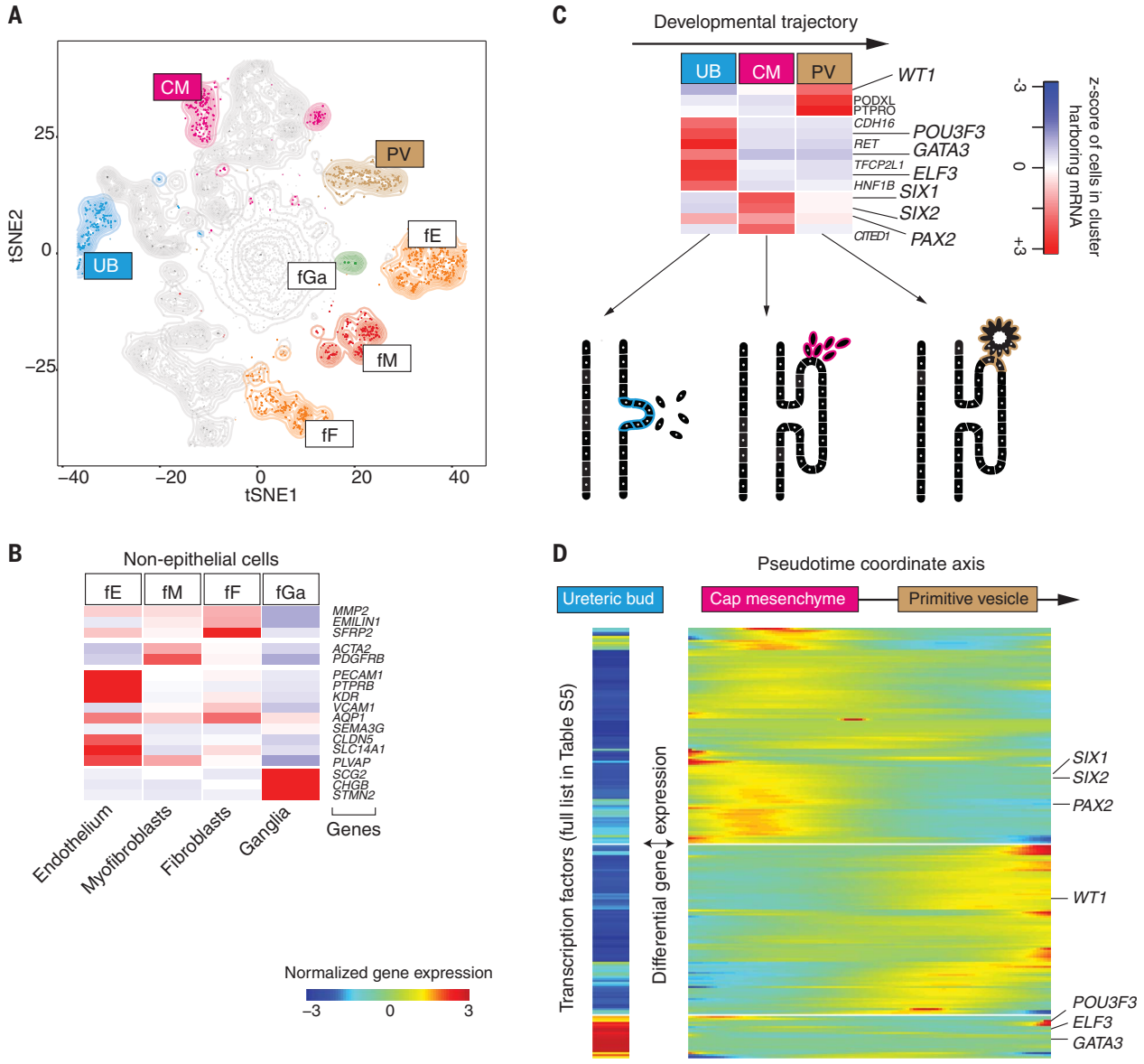


Fig. 2. Fetal cell types and nephrogenesis. (**A**) t-SNE representation of 4,858 fetal epithelial and vascular cells, colored and labeled as in Fig. 1B. fE, fetal endothelium; fGa, fetal ganglia; fM, fetal myofibroblasts; fF, fetal fibroblasts. (**B**) Expression of markers of clusters in (A), colored as in Fig. 1C. (**C**) Expression of nephrogenesis markers from clusters in (A) with illustration of nephron development. Formation of nephrons emanates from the UB, which induces condensation of the overlying mesenchyme into the CM. The CM then forms the PV, the precursor of the glomerulus. The tubular system grows out from both ends of the fetal nephron: the UB and the PV. (**D**) The expression of transcription factors that vary significantly ($P < 0.01$; likelihood ratio test) along the pseudotime trajectory defined by using the CM and PV cells from (C) or that are differentially expressed in UB versus CM and PV cells. UB expression is shown in a separate block on the left. Within the right block, pseudotime increases from left to right and rows are clustered and grouped by hierarchical clustering, with canonical transcription factors of nephrogenesis highlighted (see table S6).

a specific marker of ccRCC cells (Fig. 4C). Furthermore, we studied the earliest precursor lesions of ccRCC: $CA9^+$ proximal tubular cells residing in morphologically normal kidney tissue, predisposed to ccRCC through pathogenic germline mutation of *VHL*. Examining tissue from three individuals, we identified $CA9^+$ $VCAM1^+$ clusters of proximal tubular cells (Fig. 4D). Similarly, tumors arising in these kidneys harbored $CA9^+$ $VCAM1^+$ cells (fig. S16). As expected, VCAM1 was otherwise sparsely expressed on proximal tubular cells. Together, these observations substantiate our proposition that PT1 cells are the nearest normal cell correlate of ccRCC cells. The presence of the PT1 signature in both ccRCC and pRCC may indicate a common origin of these tumors with divergent fates.

Apart from the PT1 signature in pRCC and ccRCC, we found that one ccRCC cell cluster (cR7) matched PT3 cells and that pRCC cells exhibited an additional, weaker match with collecting duct cells (Fig. 4A). Neither signal was enriched in bulk transcriptomes (Fig. 4B). As our study was confined to type 1 pRCC, it is possible that we missed other pRCC cell types.

Finally, we dissected the tumor microenvironment occupied by cancer-associated normal cells, comprised of immune cells, fibroblasts, myofibroblasts, and vascular endothelial cells (predominately ascending vasa recta) (figs. S7, S8, and S17). Within these groups, we studied vascular endothelial growth factor (VEGF) signaling, an established target in RCC treatment (*7*, *8*). The VEGF signaling circuit in renal tumors involves VEGFA secretion from RCC cells, resulting in a response from endothelial cells (*7*, *8*). Measuring expression of the key components of VEGF signaling, we identified tumor-infiltrating macrophages as a further source of VEGFA (fig. S18A), as confirmed by confocal microscopy of ccRCC cells and flow cytometry of an independent ccRCC tumor (fig. S18, B to D). VEGF signaling receptors (KDR, FLT1, and FLT4) were expressed mainly by one population of ascending vasa recta cells (fig. S18A, cluster tE1). The other ascending vasa recta cluster, tE2 (fig. S18A), exhibited lymphangiogenic VEGFC and FLT1. Furthermore, tE2 endothelial cells expressed high levels of ACKR1, a marker of venular endothelium promoting tissue migration of immune cells (*9*). Overall, these findings delina complex VEGF signaling circuit within RCC tissue.

By identifying specific normal cell correlates of renal cancer cells, this study moves our understanding of these malignancies beyond a notion of "fetalness" or an approximate micro-anatomical

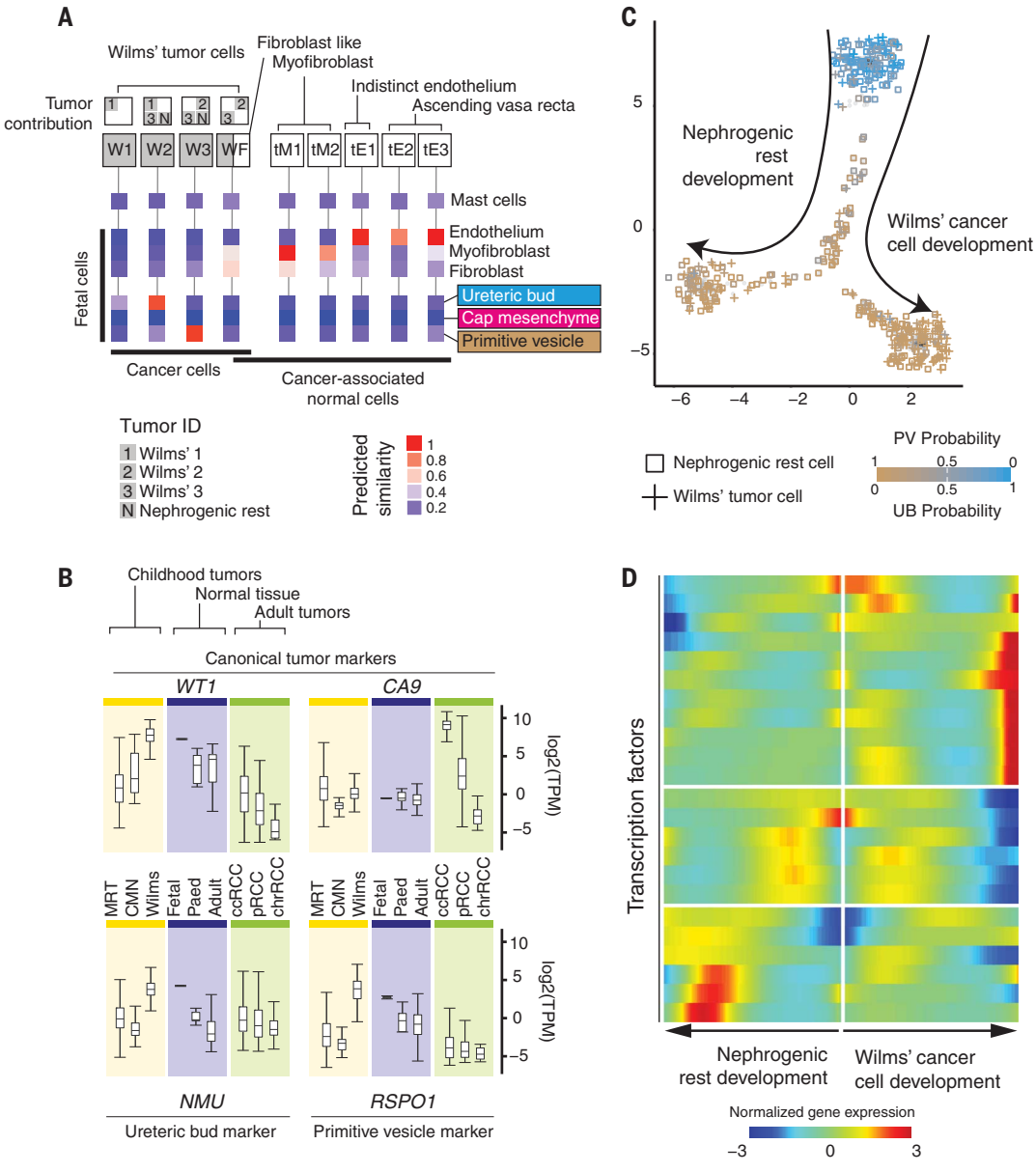


Fig. 3. Matching childhood tumors with normal fetal cells. (**A**) Similarity of Wilms tumor and cancer-associated normal cells to the reference fetal kidney map (Fig. 2A), with mast cells added as a negative control. Square boxes indicate sample contribution. Colors represent the probability that the cluster identified in the column header is "similar" to the fetal cluster identified by the row label (*2*). tM1 and tM2, tumor myofibroblast clusters 1 and 2; tE1 to tE3, tumor endothelial clusters 1 to 3. (**B**) Expression of canonical tumor markers and representative UB- and PV-specific genes (table S8) in bulk transcriptomes of childhood cancers (yellow), normal tissue (blue), or adult cancers (green). MRT, malignant rhabdoid tumor; CMN, congenital mesoblastic nephroma. As positive controls, canonical tumor markers are shown: WT1, Wilms tumor, and CA9, ccRCC. TPM, transcripts per kilobase million. (**C**) Pseudotime trajectory of all Wilms tumor and nephrogenic rest cells. Color indicates the similarity of each cell to the PV or UB fetal population. Jitter has been added to each point's position, with the original position plotted underneath in black (*2*). (**D**) Transcription factors identified as varying significantly along the pseudotime trajectory in (C). The center of the heat map corresponds to the cells at the top of (C), and the map then proceeds left and right along the arrows shown in (C).

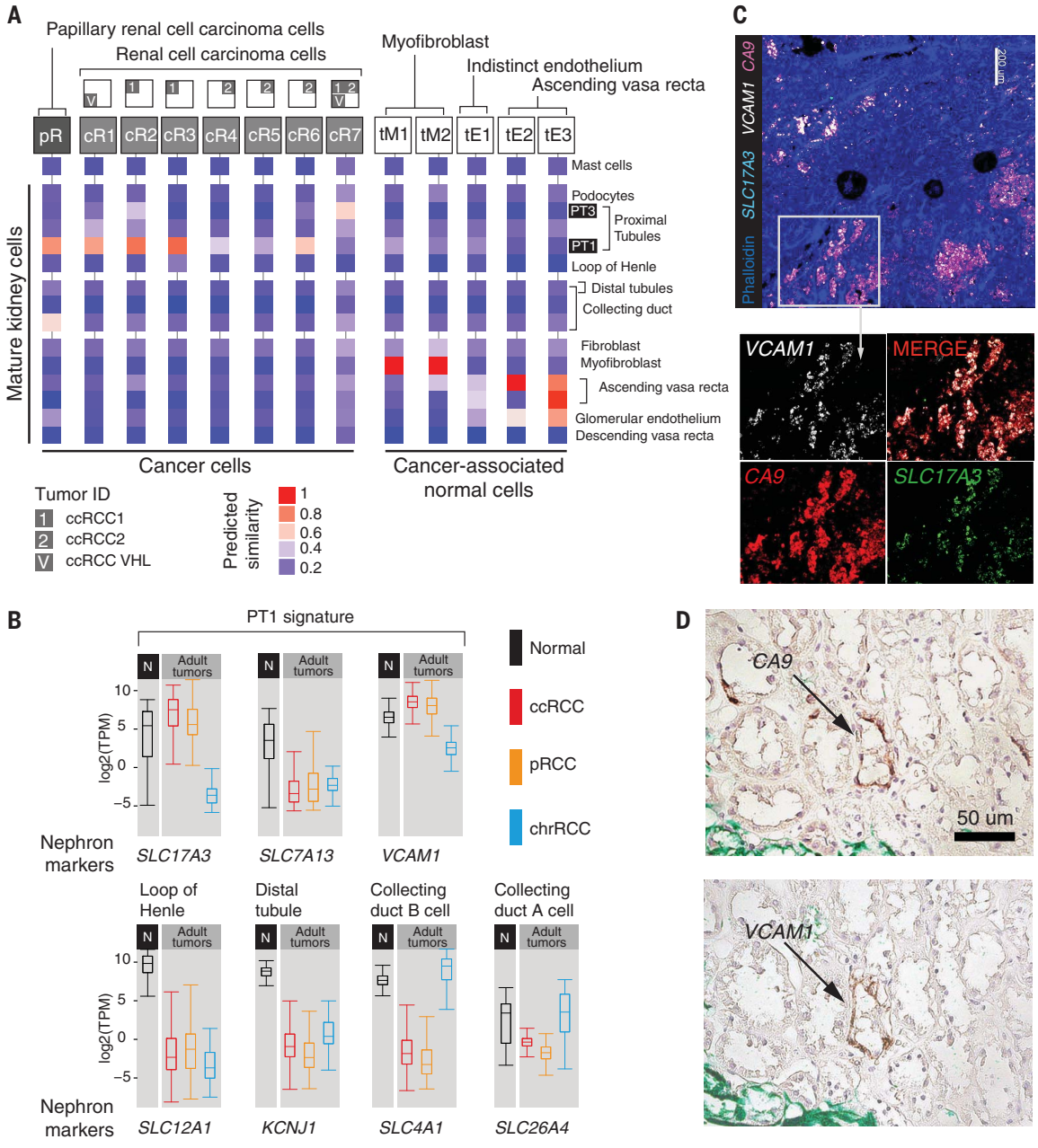


Fig. 4. Matching adult tumors with normal mature kidney cells. (**A**) Similarity of adult cancer and cancer-associated normal cells to the mature kidney reference map (Fig. 1B), with mast cells added as a negative control. Square boxes indicate sample contribution. Colors represent the probability that the cluster identified in the column header is "similar" to the normal cluster identified by the row label (*2*). pR, pRCC cell cluster; cR1 to cR7, ccRRC cell clusters. (**B**) Expression of nephron-specific genes in bulk transcriptomes as in Fig. 3B. pRCC samples are both types 1 and 2. (**C**) Confocal microscopy showing colocalization of PT1 markers (VCAM1 and SLC17A3) in ccRCC cells (CA9). (**D**) Staining of a proximal tubular ccRCC precursor lesion (CA9) for the PT1 marker VCAM1.

region to a precise cellular, molecularly quantitative resolution. Our findings portray the peak incidence of Wilms tumor in early childhood as a corruption of fetal nephrogenesis, in contrast to the lifelong development of RCC in mature kidneys. Our study provides a scalable experimental strategy for determining the identity of human cancer cells.

REFERENCES AND NOTES

1. C. Ziegenhain *et al.*, *Mol. Cell* **65**, 631–643.e4 (2017).
2. Materials and methods are available as supplementary materials.
3. K. Pritchard-Jones *et al.*, *Lancet* **386**, 1156–1164 (2015).
4. E. J. Perlman *et al.*, *Nat. Commun.* **6**, 10013 (2015).
5. S. Gadd *et al.*, *Nat. Genet.* **49**, 1487–1494 (2017).
6. D. Seron, J. S. Cameron, D. O. Haskard, *Nephrol. Dial. Transplant.* **6**, 917–922 (1991).
7. B. Ljungberg *et al.*, *Eur. Urol.* **67**, 913–924 (2015).
8. S. Fernández-Pello *et al.*, *Eur. Urol.* **71**, 426–436 (2017).
9. A. Thiriot *et al.*, *BMC Biol.* **15**, 45 (2017).

ACKNOWLEDGMENTS

We thank M. Stratton, P. Campbell, D. Rowitch, M. Gessler, and M. Ramakrishna for review of the manuscript and M. Gerstung and V. Svensson for advice regarding logistic regression. We are indebted to our patients and their families for participating in this research. **Funding:** This experiment was principally funded by the St. Baldrick's Foundation (Robert J. Arceci International Award to S.B.). Additional funding was received from Wellcome (S.B., M.H., G.C., C.G., P.H.M.), Cambridge Biomedical Research Campus (biobanking infrastructure; M.C.), CRUK Cambridge Centre (biobanking infrastructure), the NIHR Blood and Transplant Research Unit (M.C.), MRC (M.C.), Arthritis Research UK (M.C.), The Lister Institute for Preventative Medicine (M.H.), NIHR and Newcastle Biomedical Research Centre (M.H.), the ISAC SRL-EL program (A.F.), joint Wellcome Trust–MRC funding (S.Lis., S.Lin.), Kidney Cancer UK (M.G.B.T.), Facing up 2 Kidney Cancer (M.G.B.T.), EMBO (R.V.-T.), the Human Frontier Science Program (R.V.-T.), and Children with Cancer UK (S.J.F.).
Author contributions: S.B. conceived of the experiment. M.D.Y., T.J.M., and S.B. analyzed the data, with contributions from F.A.V.B., B.J.S., M.D.C.V.-H., G.C., and M.C. Samples were curated and/or experiments were performed by F.A.V.B., J.R.F., M.G.B.T., P.H.M., R.A.B., D.-M.P., R.V.-T., E.S., K.W.L., S.J.F., C.G., N.R., L.M., T.A., J.N.A., A.C.P.R., I.M., S.F., D.R., J.N., A.F., J.B., S.Lis., S.Lin., and G.D.S. Pathological expertise was provided by A.Y.W., N.S., and N.C. A.C. created illustrations. T.J.M., M.D.Y., and S.B. wrote the manuscript. M.H., S.A.T., M.C., and S.B. codirected the study. **Competing interests:** P.H.M. discloses paid consultancy work for Mission Therapeutics and stock ownership in ReOx plc. A.Y.W. attended a meeting of the Roche UK Scientific Advisory

Board on PDL-1 testing in renal cell carcinoma. G.D.S. discloses paid consultancy work for Pfizer, EUSA Pharma, and Cambridge Medical and an advisory board role for the latter. **Data and materials availability:** Raw sequencing data have been deposited in the European Genome-phenome Archive (EGA) under study IDs EGAS00001002171, EGAS00001002486, EGAS00001002325, and EGAS00001002553. Sample-specific identifiers can be found in tables S6 and S10, a table of mapped unique molecular identifier counts for each cell and gene combination in data S1, and metadata about each cell in table S11. The code necessary to perform the analysis and generate figures can be obtained from https://github.com/constantAmateur/scKidneyTumors.

SUPPLEMENTARY MATERIALS
www.sciencemag.org/content/361/6402/594/suppl/DC1

Materials and Methods
Figs. S1 to S19
Tables S1 to S12
References (*10*–*46*)
Data S1

1 February 2018; accepted 2 July 2018
10.1126/science.aat1699

METABOLISM

Lacteal junction zippering protects against diet-induced obesity

Feng Zhang[1], Georgia Zarkada[1], Jinah Han[1], Jinyu Li[1], Alexandre Dubrac[1], Roxana Ola[1,2], Gael Genet[1], Kevin Boyé[1], Pauline Michon[1,3], Steffen E. Künzel[1], Joao Paulo Camporez[4], Abhishek K. Singh[5], Guo-Hua Fong[6], Michael Simons[1], Patrick Tso[7], Carlos Fernández-Hernando[5], Gerald I. Shulman[4,8], William C. Sessa[9], Anne Eichmann[1,3,8]*

Excess dietary lipid uptake causes obesity, a major global health problem. Enterocyte-absorbed lipids are packaged into chylomicrons, which enter the bloodstream through intestinal lymphatic vessels called lacteals. Here, we show that preventing lacteal chylomicron uptake by inducible endothelial genetic deletion of *Neuropilin1* (*Nrp1*) and *Vascular endothelial growth factor receptor 1* (*Vegfr1*; also known as *Flt1*) renders mice resistant to diet-induced obesity. Absence of NRP1 and FLT1 receptors increased VEGF-A bioavailability and signaling through VEGFR2, inducing lacteal junction zippering and chylomicron malabsorption. Restoring permeable lacteal junctions by VEGFR2 and vascular endothelial (VE)–cadherin signaling inhibition rescued chylomicron transport in the mutant mice. Zippering of lacteal junctions by disassembly of cytoskeletal VE-cadherin anchors prevented chylomicron uptake in wild-type mice. These data suggest that lacteal junctions may be targets for preventing dietary fat uptake.

Dietary fats are absorbed by enterocytes and incorporated into triglyceride-rich lipoproteins, called chylomicrons. Nearly all dietary lipids are transported in chylomicrons from the intestine to tissues via the lymphatic system. Chylomicrons enter the lymphatics through lacteals, specialized lymphatic capillaries located in the center of the intestinal villi. From there, they are transported through the mesenteric lymphatic vessels into the thoracic duct, which drains into the venous circulation, and become metabolized by the liver (*1*–*3*).

Preventing lacteal growth in mice by conditional deletion of the genes encoding VEGF-C or Delta-like 4 (DLL4) renders mice resistant to high-fat diet (HFD)–induced obesity, providing experimental evidence that lacteals could be targeted to prevent obesity (*4*, *5*). The cellular mechanisms controlling chylomicron entry into the lacteals are poorly understood; some studies suggested that chylomicrons enter through junctions between lymphatic endothelial cells (LECs) lining the lacteals (*5*–*7*), but other studies have shown that they pass through LECs by transcytosis (*3*). The molecular mechanisms controlling lacteal chylomicron uptake are currently unknown. Here we show that chylomicron uptake is controlled by VEGF-A signaling, through modulation of lacteal junctions. Lacteals respond to high VEGF-A levels by zippering up their junctions, which impairs chylomicron passage and renders mice resistant to diet-induced obesity.

FLT1 and NRP1 are VEGF-A receptors that were previously implicated in metabolism regulation, but how they function in endothelium in this context remained undefined (fig. S1). We therefore generated pan-endothelial deletions of *Nrp1* and *Flt1* by crossing $Nrp1^{fl/fl}$ and $Flt1^{fl/fl}$ mice to $Cdh5\text{-}Cre^{ERT2}$ mice, where gene deletion can be induced by tamoxifen (TAM) in cells expressing the vascular endothelial cadherin promoter (hereafter referred to as *Nrp1;Flt1iECko* mice). Adult 5-week-old males and Cre-negative littermate controls received TAM to induce efficient gene deletion in the endothelium, as confirmed by Western blot and real-time quantitative polymerase chain reaction on purified endothelial cells (fig. S2). HFD feeding was initiated at week 8 and continued for 16 weeks (wk) (fig. S2). Control mice doubled their body weight after 16wk of HFD feeding, whereas *Nrp1;Flt1iECko* mice gained little weight (Fig. 1, A to C). No weight differences were observed in mice on a normal chow diet (Fig. 1D). The weight of *Nrp1iECko* mice was similar to that of littermate controls (Fig. 1E), whereas weight gain in *Flt1* single mutants was slightly reduced, and attributed to increased adipose tissue angiogenesis (*8*, *9*) (Fig. 1F), as described previously.

HFD-fed *Nrp1;Flt1iECko* mice had reduced fat and increased lean body mass ratios when compared to control littermates, and their body composition resembled that of lean control mice on a normal chow diet (Fig. 1G and fig. S3). In contrast to controls, *Nrp1;Flt1iECko* mice had reduced liver triglyceride content and did not develop hepatic steatosis after 8wk of HFD feeding (Fig. 1H and fig. S4). *Nrp1;Flt1iECko* and control mice showed similar food and water consumption, physical activity, O_2 consumption, CO_2 production, and energy expenditure after 2wk and 8wk of HFD feeding (Fig. 1I and fig. S5). Glucose tolerance in *Nrp1;Flt1iECko* mice was improved compared to controls (Fig. 1J).

Plasma triglycerides, total cholesterol, and high-density lipoprotein cholesterol (HDL) concentrations were reduced in *Nrp1;Flt1iECko* mice when compared to controls after 16wk on a HFD (Fig. 2A), indicating that resistance to diet-induced obesity was either due to defective lipid uptake into the bloodstream, reduction in very-low-density lipoprotein (VLDL) production, or enhanced VLDL and chylomicron catabolism. To determine if lipid uptake was affected, we measured plasma triglyceride uptake after intragastric gavage with olive oil in the presence or absence of the lipoprotein lipase (LPL) inhibitor TritonWR1339. As expected, circulating triglyceride concentrations increased in controls in both conditions, but the effect was abrogated in *Nrp1;Flt1iECko* mice (Fig. 2, B and C). Similar results were obtained when we injected the mice with the LPL inhibitor Poloxamer 407 and fed them via gavage with oil mixed with [^{3}H]-triolein (Fig. 2D), indicating that absence of NRP1 and FLT1 attenuates fat absorption. Bomb calorimetry showed increased calorie content in the feces of *Nrp1;Flt1iECko* mice, suggesting that part of the lipid is not absorbed but rather is cleared in the feces (Fig. 2E). Histological analysis revealed the presence of oil-red-O–labeled lipid particles in the colon of *Nrp1;Flt1iECko* mice, whereas less lipid was present in the colon of control mice (fig. S6). We confirmed reduced intestinal lipid uptake in *Nrp1;Flt1iECko* mice by analyzing postnatal day 7 (P7) mesenteries, after neonatal gene deletion. Chylomicrons were present in the mesenteric lymphatic vessels of control mice and single *Nrp1* or *Flt1* knockout mice, but few lymphatic vessels containing milky chyle could be seen in *Nrp1;Flt1iECko* pups (Fig. 2F and fig. S7). Neonatal *Nrp1;Flt1iECko* mice had decreased weight gain (fig. S8), which is consistent with lipid malabsorption.

Expression of key enzymes and components required for chylomicron assembly, including *Mtp*, *ApoB*, *Sar1b*, *Plagl2*, and fatty acid transporters (*CD36* and *Fabps*), was similar in the jejunum of *Nrp1;Flt1iECko* and control mice, indicating that the enterocytes devoid of both *Nrp1 and Flt1* can assemble and secrete chylomicrons (fig. S9). Transmission electron microscopy (TEM) showed similar morphology of intestinal enterocytes and the presence of chylomicrons within enterocytes and in the interstitial space between capillaries and lacteals regardless of genotype

[1]Cardiovascular Research Center, Yale University School of Medicine, New Haven, CT 06510-3221, USA. [2]Department of Basic, Preventive and Clinical Science, University of Transylvania, 500019 Brasov, Romania. [3]INSERM U970, Paris Cardiovascular Research Center, 75015 Paris, France. [4]Department of Internal Medicine, Yale University School of Medicine, New Haven, CT, USA. [5]Departments of Comparative Medicine and Pathology, Vascular Biology and Therapeutics Program and Integrative Cell Signaling and Neurobiology of Metabolism Program, Yale University School of Medicine, New Haven, CT, USA. [6]Department of Cell Biology, University of Connecticut Health Center, Farmington, CT, 06030-3501, USA. [7]Department of Pathology and Laboratory Medicine, Metabolic Diseases Institute, University of Cincinnati, Cincinnati, OH 45237-0507, USA. [8]Department of Cellular and Molecular Physiology, Yale University School of Medicine, New Haven, CT, USA. [9]Department of Pharmacology, Vascular Biology and Therapeutics Program, Yale University School of Medicine, New Haven, CT, USA.
***Corresponding author. Email: anne.eichmann@yale.edu**

(Fig. 2G and fig. S10). TEM also revealed the presence of lacteal LECs but notably, the lacteal lumen was almost completely lacking chylomicrons in *Nrp1;Flt1iECko* animals, whereas numerous chylomicrons were present within the lacteal lumen in controls (Fig. 2G and fig. S10). High-magnification views showed many open junctions between LECs in controls, but more closed LEC junctions in *Nrp1;Flt1iECko* mice (Fig. 2G and fig. S10). These data are consistent with the concept that altered lacteal junctions result in defective lacteal chylomicron uptake, which limits weight gain in the absence of both NRP1 and FLT1.

Immunostaining of lacteal junctions with VE-cadherin confirmed that *Nrp1;Flt1iECko* mice exhibited fewer button-like LEC junctions and developed more impermeable zipper junctions compared to controls (Fig. 3, A and B). Lacteals devoid of both *Nrp1* and *Flt1* were of similar length, but showed reduced width compared to those lacking either *Nrp1* and *Flt1* alone, or control animals (fig. S11). VE-cadherin staining also revealed enlarged villus capillaries and disrupted blood endothelial cell (BEC) junctions, leading to leakage of intravenously injected fluorescent dextran in *Nrp1;Flt1iECko* mice (Fig. 3, C and D, and fig. S12). Postnatal or adult deletion of both *Nrp1* and *Flt1* caused altered vascular morphology and villus edema, which was surprisingly well tolerated and did not compromise vascular or enterocyte ultrastructure even after prolonged HFD feeding (fig. S10). Staining for pericytes, smooth muscle cells, and macrophages and analysis of inflammatory marker expression revealed no changes between controls and *Nrp1;Flt1iECko* mutants (fig. S13), suggesting a primary effect upon villus vasculature by the absence of *Nrp1* and *Flt1*.

Loss of junctional VE-cadherin staining in BECs is a hallmark of increased VEGF-A signaling (*10*), and FLT1 is a known VEGF-A decoy receptor (*11*). To test if loss of NRP1 and FLT1 increased intestinal VEGF-A signaling, we probed lysates from control and *Nrp1;Flt1iECko* tissues with antibodies against phosphorylated and total VEGFR2. Activation of both tyrosine-1173 (Y1173) and Y949 signaling was significantly increased in the absence of both *Nrp1 and Flt1* when compared to control littermates or single *Flt1iECko* mice (Fig. 3E and fig. S14), indicating that NRP1 enhances FLT1 decoy function. NRP1 also acts as a VEGFR2 co-receptor to activate downstream ERK signaling (*12*), and ERK activation was accordingly deficient in most *Nrp1;Flt1iECko* tissues (fig. S14).

Loss of *Flt1* has been shown to induce adipose tissue browning (*8*, *9*), and we observed increased vascular density in white adipose tissue (WAT) but not in brown adipose tissue (BAT) in *Nrp1;Flt1iECko* mice (fig. S15). Expression levels of the mitochondrial uncoupling protein UCP1 were only slightly increased in *Nrp1;Flt1iECko* epididymal WAT and were unchanged in subcutaneous WAT and BAT (fig. S15). Thus, adipose tissue browning is unlikely to account for resistance to diet-induced obesity in *Nrp1;Flt1iECko* animals.

VEGF-A is known to increase blood-vascular permeability (*13*, *14*), and lacteals express VEGFR2 (fig. S16), but effects of VEGF-A on lacteal junctions have previously not been investigated, to the best of our knowledge. Intravenous VEGF-A injection disrupted BEC junctions after 30 min as expected, but also increased the number of zipper-like junctions in lacteals (Fig. 3, F and G). Intravenous injection of the lymphatic growth factor VEGF-C also increased junction zippering, but to a lesser extent than VEGF-A, whereas a mutant VEGF-C-156S protein that cannot bind VEGFR2 (*15*) had no effect (Fig. 3G and fig. S16). In vitro, treatment of starved confluent LECs with VEGF-A and VEGF-C, but not VEGF-C-156S, promoted the appearance of straighter VE-cadherin–lined junctions and decreased actin stress fiber anchoring to perpendicular arranged VE-cadherin (fig. S16). These data suggested that increased VEGF-A–VEGFR2 signaling in *Nrp1;Flt1iECko* villi might cause defective chylomicron uptake via lacteal junction zippering.

Intestinal villi show high levels of *Vegf-a* expression, lower levels of *Vegf-c*, and little expression of *Vegf-b* (fig. S17). Expression of *Flt1* is increased at birth, concomitantly with the expression of *Mtp* and *ApoB*, a known inducer of *Flt1* (*16*) (fig. S17). RNA sequencing showed that mature intestinal BECs express four times as many copies of *Nrp1* when compared to *Flt1* (fig. S17), consistent with the concept that chylomicrons induce an increase of FLT1, which functions together with NRP1 to prevent excessive VEGF-A signaling, thereby allowing lacteal junction maturation and chylomicron absorption.

Cdh5-Cre^{ERT2} mediates gene deletion in both intestinal BECs and LECs, raising the possibility that NRP1;FLT1 function might be required in both compartments in a cell-autonomous manner. We therefore generated LEC-specific *Nrp1* and *Flt1* deletions using *Prox1-Cre*$^{ERT2(BAC)}$ mice. *Prox1-Cre*$^{ERT2(BAC)}$ specifically recombined

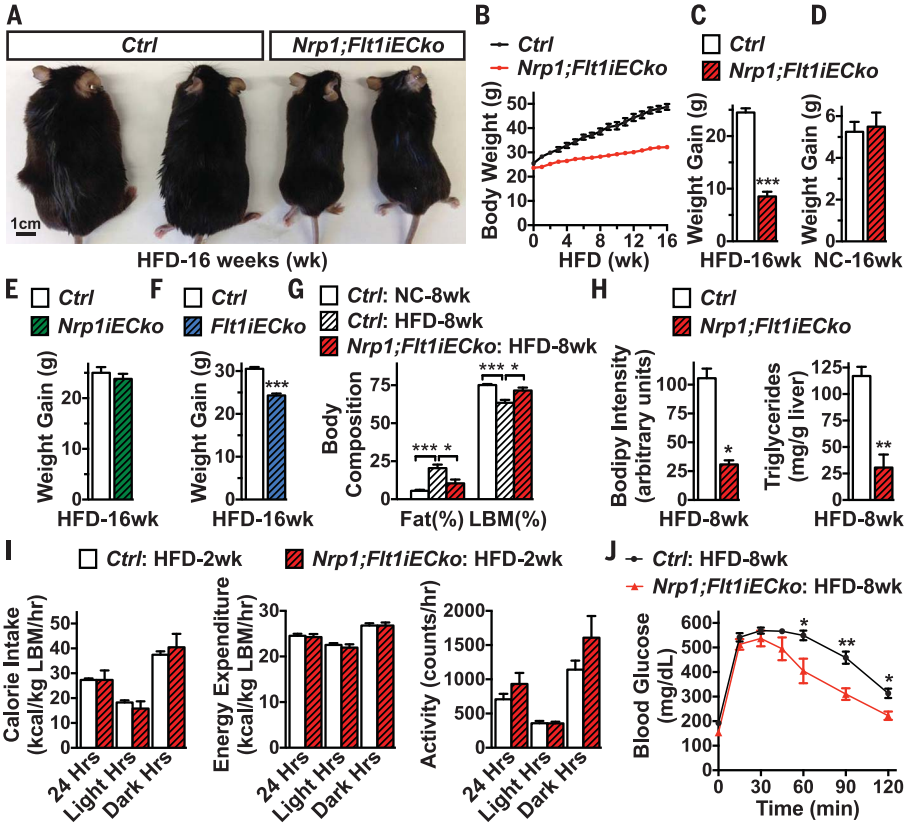


Fig. 1. *Nrp1;Flt1iECko* mice are resistant to diet-induced obesity. (**A**) *Nrp1;Flt1iECko* and *Cre*-negative littermate control (*Ctrl*) mice after 16-week (wk) HFD. (**B** and **C**) Growth curves (B) and weight gain (C) of *Ctrl* (n = 8) and *Nrp1;Flt1iECko* (n = 11) mice after 16wk HFD feeding. Data are mean ± SEM. ***p < 0.001. Mann-Whitney U test. (**D** to **F**) Weight gain of normal chow (NC)- or HFD-fed mice (n = 4 to 11 per group). Data are mean ± SEM. ***p < 0.001. Mann-Whitney U test. (**G**) Body composition of mice (n = 7 to 8 per group) after 8wk NC or HFD. Data are mean ± SEM. *p < 0.05, ***p < 0.001. Mann-Whitney U test. (**H**) Quantification of BODIPY staining intensity of liver cryosections (left) and liver triglyceride content (right) from 8wk HFD-fed mice (n = 4 to 5 per group). Data are mean ± SEM. *p < 0.05, ***p < 0.001. Mann-Whitney U test. (**I**) Calorie intake, energy expenditure, and activity of *Ctrl* (n = 7) and *Nrp1;Flt1iECko* (n = 5) mice after 2wk on HFD. Data represent 24 hours (hrs), light hrs (7:00 a.m. to 7:00 p.m.) and dark hrs (7:00 p.m. to 7:00 a.m.) average, normalized to lean body mass (LBM). Error bars: SEM. (**J**) Intraperitoneal glucose tolerance test in HFD-fed *Ctrl* (n = 10) and *Nrp1;Flt1iECko* (n = 6) mice. Data are mean ± SEM. *p < 0.05; **p<0.01. Mann-Whitney U test.

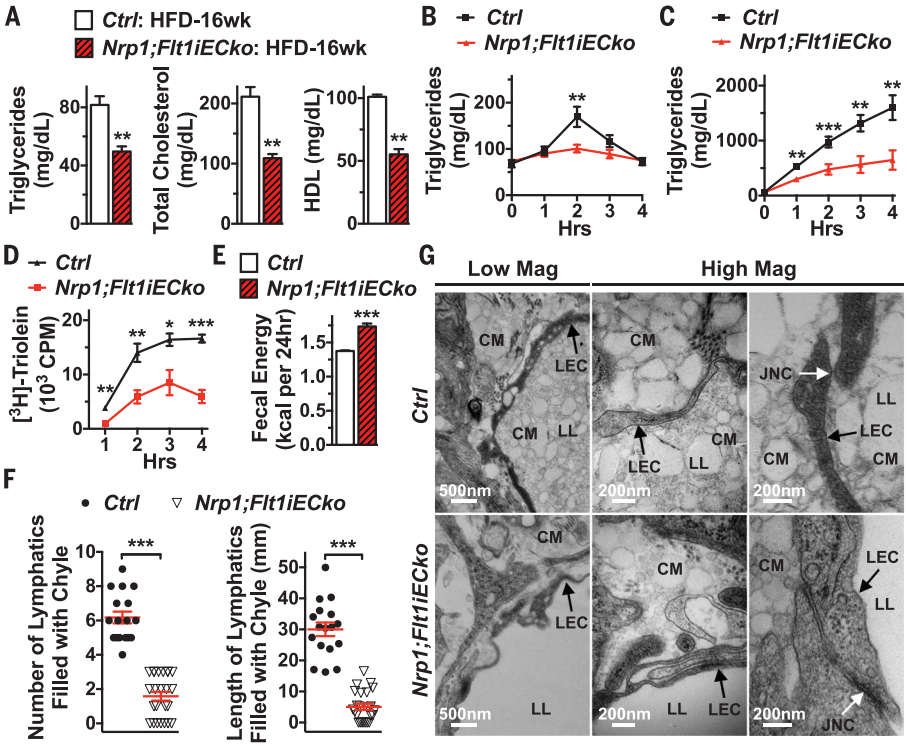


Fig. 2. Endothelial *Nrp1;Flt1* deletion prevents lacteal chylomicron absorption. (**A**) Plasma lipid profile in 6-hours fasted *Ctrl* (n = 6) and *Nrp1;Flt1iECko* mice (n = 6) after 16wk HFD. Data are mean ± SEM. $^{**}p < 0.01$. Mann-Whitney U test. (**B** and **C**) Plasma triglyceride content in NC-fed adult mice after gavage with 200 μl of olive oil. All mice (n = 8 to 10 per group) received 6 hours of fasting, and mice in (C) received Triton WR1339 (0.5 g/kg) intraperitoneally (i.p.) 30 min before gavage. Data are mean ± SEM. $^{**}p < 0.01$, $^{***}p < 0.001$. Mann-Whitney U test. (**D**) Plasma ^{3}H CPM (counts per minute) in NC-fed adult mice after gavage with ^{3}H-triolein–containing lipid. Mice (n = 5 per group) were fasted for 6 hours and injected i.p. with poloxamer 407 (1 g/kg) 30 min before gavage. Data are mean ± SEM. $^{*}p < 0.05$; $^{**}p < 0.01$, $^{***}p < 0.001$. Mann-Whitney U test. (**E**) Fecal bomb calorimetry analysis of 2wk HFD-fed mice (n = 8 per group). Data are mean ± SEM. $^{***}p < 0.001$. Mann-Whitney U test. (**F**) Quantifications of chyle-filled lymphatics in mesenteries of P7 mice after injection with TAM (3 × 100 μg) at P2 to P4. Each symbol represents one mouse (n = 17 to 24 per group). Error bars: SEM. $^{***}p < 0.001$. Mann-Whitney U test. (**G**) Transmission electron microscopy of jejunum central lacteals in P13 mice. LL: lacteal lumen; LEC: lymphatic endothelial cell; CM: chylomicron. JNC: junction. The closed LEC junction and empty lacteal lumen are apparent in *Nrp1;Flt1iECko* mice.

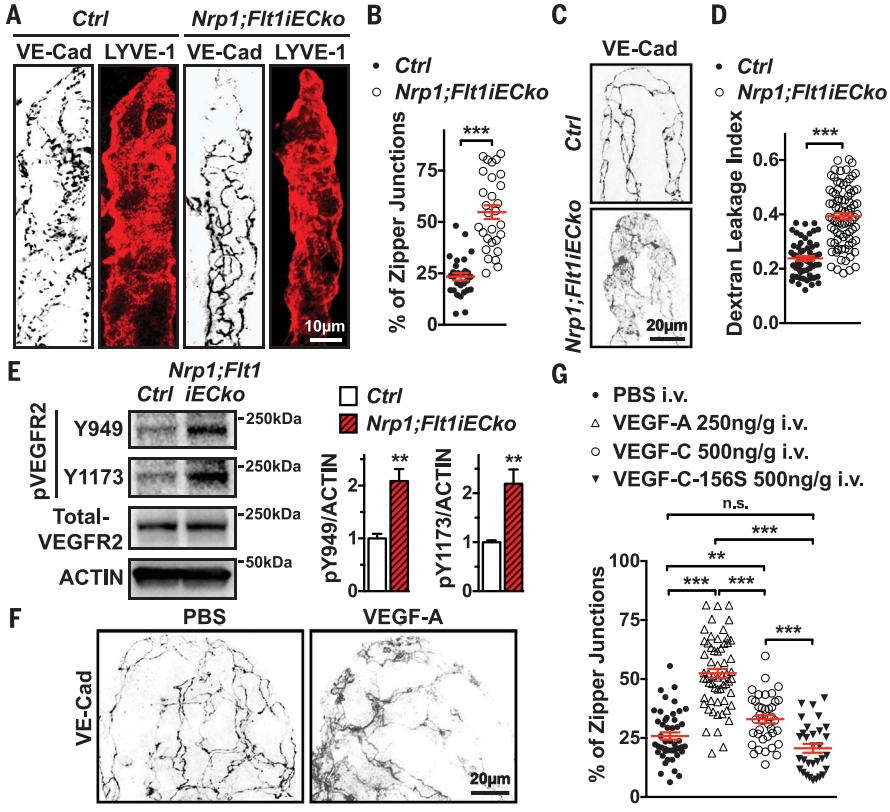


Fig. 3. Increased VEGF-A signaling alters villus endothelial junctions. (**A**) VE-Cad and LYVE-1 staining of whole-mounted jejunum lacteals from P13 to P18 mice after postnatal TAM administration. (**B**) Quantification of zipper-like lacteal junctions in (A). Each symbol represents one lacteal (7 mice per group). Data are mean ± SEM. $^{***}p < 0.001$. Mann-Whitney U test. (**C**) VE-Cad staining of whole-mounted jejunum BEC junctions from P7 mice. (**D**) Quantification of dextran leakage in P11 mice after Rhodamine-dextran and Alexa 647-IsoB4 intravenous (i.v.) injection for 10 min. Each symbol represents one villus (5 to 6 mice per group). Data are mean ± SEM. ***, $p < 0.001$. Mann-Whitney U test. (**E**) Western blots and quantifications of VEGFR2 phosphorylation in P7 jejunum lysates. n = 5 to 6 mice per group. Data are mean ± SEM. $^{**}p < 0.01$. Mann-Whitney U test. (**F** and **G**) VE-Cad staining of villus BEC junctions and quantification of zipper-like lacteal junctions in jejunum from P18 to P21 wild-type mice 30 min after injection of growth factors or phosphate-buffered saline (PBS). Each symbol represents one lacteal (4 to 6 mice per group). Data are mean ± SEM. n.s., not significant; $^{**}p < 0.01$, $^{***}p < 0.001$. Mann-Whitney U test.

mesenteric and lacteal LECs, but loss of *Nrp1;Flt1* in LECs had no detectable effects on postnatal mesenteric chyle uptake or on weight gain in adult HFD-fed mice (fig. S18). Lymphatic button-like junctions were present and responded to VEGF-A by forming zippers (fig. S18). Together with highly enriched expression of *Nrp1* and *Flt1* in intestinal BECs compared to LECs (*17*), these data suggest that NRP1 and FLT1 function in villus BECs to antagonize VEGF-A–VEGFR2 signaling on lacteal LECs.

The effect of *Nrp1;Flt1* deletion on LEC junctions was tissue-specific, as the junction morphology in initial lymphatics in the skin and diaphragm of *Nrp1;Flt1iECko* mice was similar to that of controls (fig. S19). Intravenous VEGF-A, but not VEGF-C or VEGF-C-156S, modestly increased dermal initial lymphatic junction zippering (fig. S19). However, intradermal injection of VEGF-A or VEGF-C robustly increased dermal LEC junction zippering, whereas VEGF-C-156S did not (fig. S19). High VEGFR2 signaling may therefore be a general inducer of lymphatic junction zippering, which is antagonized by NRP1; FLT1 decoys in the small intestine and via other mechanisms in other tissues.

If enhanced VEGF-A signaling accounted for defective chylomicron uptake in *Nrp1;Flt1iECko* mice, blocking VEGFR2 function should rescue this process. Indeed, intraperitoneal injection of the VEGFR2 blocking antibody DC101 (*18*) decreased VEGFR2 phosphorylation, and increased uptake of mesenteric chyle in neonates 6 hours after treatment (Fig. 4, A and B, and fig. S20). Dilation of villus capillaries was reduced, and width of lacteals increased (fig. S20). DC101-treated double mutants showed less dextran leak from intestinal capillaries (fig. S20), had more LEC button-like junctions and fewer zippers, and had higher plasma lipid concentrations in the bloodstream when compared to untreated *Nrp1;Flt1iECko* animals (Fig. 4C-D).

We next determined if changes in lacteal junctions were sufficient to impair chylomicron uptake. Button-like junctions develop postnatally by transformation of zipper-like junctions in a poorly characterized mechanism that involves angiopoietin2 (ANGPT2) and changes in plasma membrane localization of VE-cadherin and LYVE-1 (*19*–*21*). LYVE-1 immunostaining and protein abundances of LYVE-1 and ANGPT2 were comparable between *Nrp1;Flt1iECko* mutants and controls, and inhibition of matrix metalloproteinase (MMP)–mediated LYVE-1 cleavage (*22*) had no effect on mesenteric chylomicron uptake (fig. S21). To test VE-cadherin function, we used the BV13 antibody, which disrupts endothelial adherens junctions in BECs and LECs (*19*). Intraperitoneal injection of BV13 (10 μg/g) into postnatal mice rescued lacteal chylomicron uptake into mesenteric lymphatics in *Nrp1;Flt1iECko* mutants and disrupted LEC junctions (Fig. 4, E and F, and fig. S22). Notably, BV13 treatment disrupted capillary BEC adherens junctions but did not affect chylomicron uptake in control mice (fig. S22). This result indicates that disruption of BEC junctions per se has no effect on

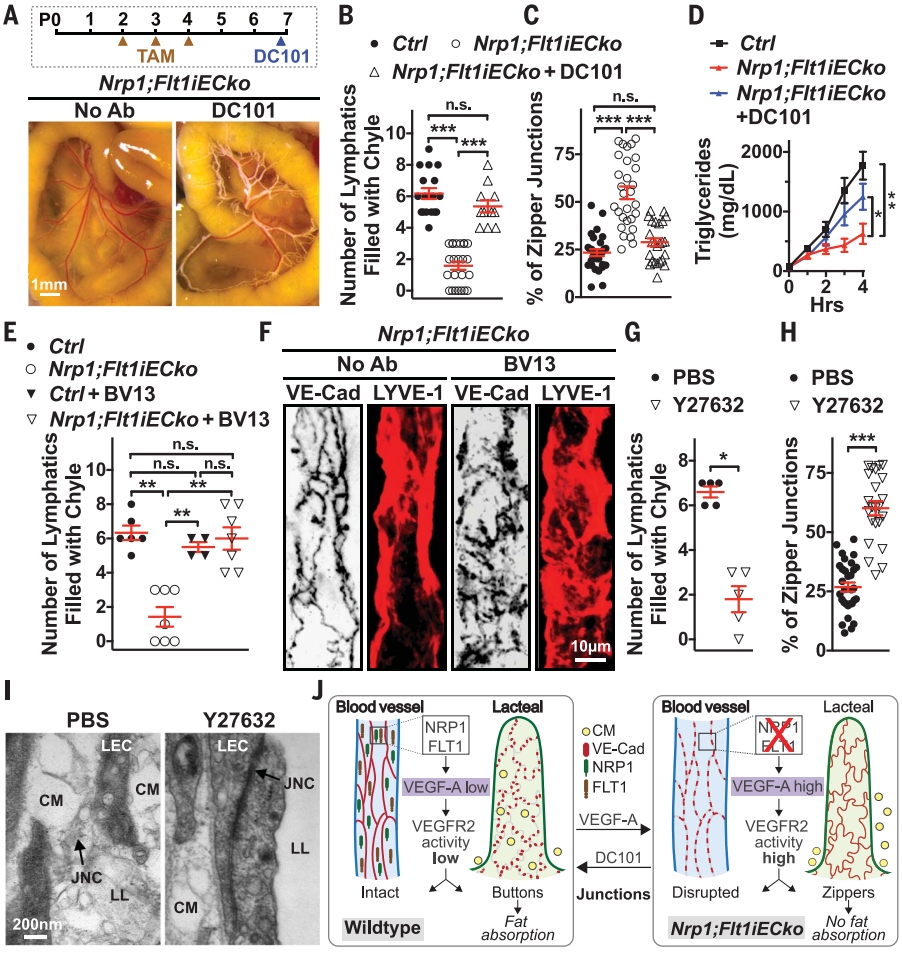


Fig. 4. Lacteal zipper junctions prevent chylomicron absorption. (**A** to **D**) DC101 rescues fat absorption in *Nrp1;Flt1iECko* mice. (**A**) Experimental timeline of TAM injection and DC101 treatment (30 μg/g, i.p., for 6 hours) and mesenteries of P7 *Nrp1;Flt1iECko* mice with or without DC101. (**B**) Quantification of chyle-filled lymphatics in (A). Each symbol represents one mouse (n = 11 to 24 per group). Data are mean ± SEM. n.s., not significant; ***$p < 0.001$. Mann-Whitney U test. (**C**) Quantification of zipper-like lacteal junctions in P13 to P18 *Nrp1;Flt1iECko* jejunum with or without DC101. Each symbol represents one lacteal (4 to 7 mice per group). Data are mean ± SEM. n.s., not significant; ***$p < 0.001$. Mann-Whitney U test. (**D**) Plasma triglyceride content in NC-fed adult mice after 6 hours of fasting and lipid gavage. All mice (n = 5 to 8 per group) received TritonWR1339 (0.5 g/kg, i.p., 30 min before gavage) and some received DC101 (30 μg/g, i.p., 4 hours before gavage). Data are mean ± SEM. *$p < 0.05$, **$p < 0.01$. Mann-Whitney U test. (**E** and **F**) BV13 rescues chylomicron uptake in *Nrp1;Flt1iECko* mice. (**E**) Quantification of chyle-filled lymphatics in P7 mice with or without BV13 treatment (10 μg/g, i.p., for 4 hours). Each symbol represents one mouse (n = 5 per group). Data show mean ± SEM. n.s., not significant; **$p < 0.01$. Mann-Whitney U test. (**F**) VE-Cad and LYVE-1 staining of jejunum lacteals from P13 *Nrp1;Flt1iECko* mice with or without BV13. (**G** to **I**) ROCK inhibition prevents chylomicron uptake in wild-type mice. (**G**) Quantification of chyle-filled lymphatics in mesenteries of P10 mice treated with Y27632 (20 μg/g, i.p., for 4 hours) or PBS. Each symbol represents one mouse (n = 5 per group). Data are mean ± SEM. *$p < 0.05$. Mann-Whitney U test. (**H**) Quantification of zipper-like lacteal junctions in P13 to P15 mice treated with Y27632 or PBS. Each symbol represents one lacteal (4 mice per group). Data are mean ± SEM. ***$p < 0.001$. Mann-Whitney U test. (**I**) TEM analysis of lacteals in P13 mice with or without Y27632 treatment. LL: lacteal lumen; LEC: lymphatic endothelial cell; CM: chylomicron. JNC: junction. (**J**) Model of NRP1;FLT1 effects on BECs and LECs in intestinal villi. VEGF-A binding to NRP1;FLT1 on BECs limits VEGF-A bioavailability for VEGFR2, resulting in continuous and discontinuous cell junctions in BECs and LECs, respectively. Discontinuous button-like LEC junctions allow lacteal chylomicron uptake. NRP1;FLT1 are highly expressed only on BECs. Increased VEGF-A concentrations or *Nrp1;Flt1* deletion in BECs result in excessive VEGFR2 activation, which disrupts BEC junctions while zippering up lacteal LEC junctions, thereby preventing chylomicron uptake. This phenotype can be rescued by inhibition of VEGFR2 signaling with DC101.

chylomicron transport and that lacteal LEC junctions, but not capillary BEC junctions, control chylomicron uptake.

In a complementary approach, we treated *Nrp1;Flt1iECko* mice with dexamethasone, a drug that induces zipper–to–button-like junction transition in tracheal lymphatics (*21*). Short-term dexamethasone treatment decreased the number of zipper-like junctions and increased chyle uptake in mesenteries of a subset of *Nrp1;Flt1iECko* mice (fig. S23). *Nrp1;Flt1iECko* mutant BEC junctions were unaffected by dexamethasone, but treated mutant animals exhibited fewer zipper-like junctions (fig. S23), confirming that LEC but not BEC junctional changes correlate with chylomicron uptake.

We reasoned that junctional changes between continuous VE-cadherin–lined zippers and discontinuous buttons could depend on cytoskeletal VE-cadherin anchoring, which can be inhibited in BECs by antagonizing Rho guanosine triphosphatase (GTPase) signaling (*23*). Indeed, the ROCK inhibitor Y27632 induced straighter junctions of cultured LECs in vitro (fig. S24), and short-term treatment of wild-type mice with Y27632 reduced chylomicron transport to mesenteric lymphatics and increased lacteal junction zippering, without affecting BEC junctions (Fig. 4, G and H, and fig. S24). TEM showed open lacteal junctions and chylomicrons in the lacteal lumen of untreated mice, but closed junctions and almost no chylomicrons in the lacteal lumen in Y27632-treated mice (Fig. 4I).

Together, our findings demonstrate that high VEGF-A signaling has opposing effects on blood and lymphatic vessels: opening capillary cell-cell junctions via well-studied pathways (*10*, *24*), but closing lymphatic junctions via transformation of buttons into zippers. We show that LEC VEGFR2 activation inhibits VE-cadherin cytoskeletal anchoring, which can be mimicked by ROCK inhibition in vitro and in vivo. Cytoskeletal VE-cadherin anchoring might generate pulling forces that help maintain lacteal button junctions open. Notably, we show that short-term effects of VEGF-A gain of function (over hours instead of days) (*25*) is sufficient to functionally switch junction morphology and lipid uptake. This reveals an intestinal-specific lacteal barrier that could be closed on demand to prevent lipid uptake (Fig. 4J). Thus, NRP1 and FLT1 appear to function collaboratively as a double decoy receptor system in intestinal BECs to limit VEGF-A signaling (Fig. 4J). VEGF-A–VEGFR2 signaling might provide a LEC growth signal together with VEGF-C–VEGFR3 (*21*, *26*). In addition, VEGFR2 signaling acts as a physiological inhibitor of lacteal maturation, whose activity must be dampened to allow acquisition of transport function. This mechanism may also apply to other tissues, with implications for edema prevention. Full knowledge of the mechanism responsible for LEC junction zippering may hold promise for identifying molecular targets for the treatment of obesity. Of note, ROCK inhibitors are clinically approved for treatment of cerebral vasospasm in some countries and have been shown to improve metabolism and decrease obesity in rodent models (*27*–*30*). Further studies of the effects of such drugs in obesity prevention, as well as of potential adverse effects of lacteal junction zippering on intestinal health, may be warranted.

REFERENCES AND NOTES

1. J. Bernier-Latmani, T. V. Petrova, *Nat. Rev. Gastroenterol. Hepatol.* **14**, 510–526 (2017).
2. G. J. Randolph, N. E. Miller, *J. Clin. Invest.* **124**, 929–935 (2014).
3. J. B. Dixon, *Ann. N. Y. Acad. Sci.* **1207** (suppl. 1), E52–E57 (2010).
4. H. Nurmi *et al.*, *EMBO Mol. Med.* **7**, 1418–1425 (2015).
5. J. Bernier-Latmani *et al.*, *J. Clin. Invest.* **125**, 4572–4586 (2015).
6. J. R. Casley-Smith, *J. Cell Biol.* **15**, 259–277 (1962).
7. S. M. Sabesin, S. Frase, *J. Lipid Res.* **18**, 496–511 (1977).
8. M. R. Robciuc *et al.*, *Cell Metab.* **23**, 712–724 (2016).
9. T. Seki *et al.*, *J. Exp. Med.* **215**, 611–626 (2018).
10. M. Simons, E. Gordon, L. Claesson-Welsh, *Nat. Rev. Mol. Cell Biol.* **17**, 611–625 (2016).
11. B. K. Ambati *et al.*, *Nature* **443**, 993–997 (2006).
12. A. Lanahan *et al.*, *Dev. Cell* **25**, 156–168 (2013).
13. Z. Sun *et al.*, *J. Exp. Med.* **209**, 1363–1377 (2012).
14. X. Li *et al.*, *Nat. Commun.* **7**, 11017 (2016).
15. V. Joukov *et al.*, *J. Biol. Chem.* **273**, 6599–6602 (1998).
16. I. Avraham-Davidi *et al.*, *Nat. Med.* **18**, 967–973 (2012).
17. G. Jurisic *et al.*, *Circ. Res.* **111**, 426–436 (2012).
18. M. Prewett *et al.*, *Cancer Res.* **59**, 5209–5218 (1999).
19. P. Baluk *et al.*, *J. Exp. Med.* **204**, 2349–2362 (2007).
20. W. Zheng *et al.*, *Genes Dev.* **28**, 1592–1603 (2014).
21. L. C. Yao, P. Baluk, R. S. Srinivasan, G. Oliver, D. M. McDonald, *Am. J. Pathol.* **180**, 2561–2575 (2012).
22. H. L. Wong *et al.*, *Nat. Commun.* **7**, 10824 (2016).
23. S. Huveneers *et al.*, *J. Cell Biol.* **196**, 641–652 (2012).
24. Y. L. Dorland, S. Huveneers, *Cell. Mol. Life Sci.* **74**, 279–292 (2017).
25. C. R. Schlieve *et al.*, *PLOS ONE* **11**, e0151396 (2016).
26. M. Wirzenius *et al.*, *J. Exp. Med.* **204**, 1431–1440 (2007).
27. M. Shibuya *et al.*, *J. Neurosurg.* **76**, 571–577 (1992).
28. Y. Kikuchi *et al.*, *J. Endocrinol.* **192**, 595–603 (2007).
29. H. Tokuyama *et al.*, *Int. J. Obes.* **36**, 1062–1071 (2012).
30. D. Okin, R. Medzhitov, *Cell* **165**, 343–356 (2016).

ACKNOWLEDGMENTS

We thank Yale Mouse Metabolic Phenotyping Center, Yale Electron Microscopy Facility, and Cincinnati Mouse Metabolic Phenotyping Center for assistance; Y. Zhao for ANCOVA analysis; and K. Alitalo, J.-L. Thomas, and S. Lee for discussion. **Funding:** This study was supported by grants from NHLBI (1R01HLI125811), NEI (1R01EY025979-01, P30 EY026878), and Agence nationale de la recherche (ANR Brainwash) to A.E. F.Z. was supported by a T32 grant from NHLBI (2T32HL7950-16A1). G.Z. was supported by an EMBO Long-Term fellowship (ALTF 87-2016) and a T32 grant from NHLBI (2T32HL007974-16). A.D. and R.O. were supported by AHA postdoctoral fellowships 14POST20380207 and 15POST25560114, respectively. M.S. was supported by NIH grants R01HL053793 and P01HL107205. **Author contributions:** F.Z., G.Z., J.H., J.L., A.D., R.O., G.G., K.B., P.M., S.E.K., J.P.C., and A.K.S. performed experiments and analyzed data. A.E. and F.Z. designed experiments, analyzed data, and wrote the manuscript. G.-H.F., M.S., P.T., C.F.H., W.C.S., and G.I.S. provided reagents or expertise. **Competing interests:** A.E. and F.Z. are inventors on a patent application (U.S. Provisional Application no. 62/686,140) submitted by Yale University that covers compositions and methods for preventing dietary lipid uptake and obesity. The authors declare no additional competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/599/suppl/DC1
Materials and Methods
Figs. S1 to S24
Table S1
References (*31*–*51*)

11 September 2017; resubmitted 4 April 2018
Accepted 27 June 2018
10.1126/science.aap9331

CELL BIOLOGY

A liquid phase of synapsin and lipid vesicles

Dragomir Milovanovic, Yumei Wu, Xin Bian, Pietro De Camilli*

Neurotransmitter-containing synaptic vesicles (SVs) form tight clusters at synapses. These clusters act as a reservoir from which SVs are drawn for exocytosis during sustained activity. Several components associated with SVs that are likely to help form such clusters have been reported, including synapsin. Here we found that synapsin can form a distinct liquid phase in an aqueous environment. Other scaffolding proteins could coassemble into this condensate but were not necessary for its formation. Importantly, the synapsin phase could capture small lipid vesicles. The synapsin phase rapidly disassembled upon phosphorylation by calcium/calmodulin-dependent protein kinase II, mimicking the dispersion of synapsin 1 that occurs at presynaptic sites upon stimulation. Thus, principles of liquid-liquid phase separation may apply to the clustering of SVs at synapses.

The presence of synaptic vesicle (SV) clusters is a defining feature of nerve terminals. SVs are tightly packed in these structures, which are well distinguished from the surrounding cytoplasm, although there is no evidence for a restraining boundary (*1*). Vesicles intermix within the clusters (*2*) and can be exchanged between them (*3*). Although SV clusters present at synapses are anchored to active zones of secretion, active-zone proteins are not required for their formation (*4*, *5*) and small clusters also occur in developing axons before synapse formation (*6*). How the motility of SVs within clusters is compatible with their spatial confinement remains unknown. Recently, liquid-liquid phase separation has been shown to be a mechanism through which components of the cytoplasm (proteins and RNAs) can assemble into distinct compartments (biomolecular condensates) not delimited by a membrane (*7*–*9*). A key feature of proteins that can undergo liquid-liquid phase separation is their ability to engage in multivalent, low-affinity interactions, either through intrinsically disordered regions (IDRs) (*10*, *11*) or through association with binding partners (*8*, *12*). A major constituent of the matrix that connects SVs is synapsin (*13*–*15*), whose abundance in nerve terminals is severalfold higher than the abundance of any other protein specifically localized in this matrix (*16*). Synapsin comprises an adenosine triphosphate (ATP)–binding module of unclear physiological function (*15*, *17*), flanked by an N-terminal short region that partially penetrates membranes (*18*) and a C-terminal IDR (*15*) with multiple SRC homology 3 (SH3) domain–binding motifs (*19*, *20*) (fig. S1). This prompted us to hypothesize (*1*) that synapsin may be a key constituent of a biomolecular condensate that includes SVs.

Departments of Neuroscience and Cell Biology, Howard Hughes Medical Institute, Kavli Institute for Neuroscience, Program in Cellular Neuroscience, Neurodegeneration and Repair, Yale University School of Medicine, New Haven, CT, USA.
*Corresponding author. Email: pietro.decamilli@yale.edu

To assess whether synapsin, which forms homo- and heterodimers, can phase-separate through interactions of its IDR, we incubated enhanced green fluorescent protein (EGFP)–tagged synapsin 1 in a buffer of physiological salt concentration and pH on a glass-bottom dish at room temperature (see methods). After a lag time of tens of minutes, synapsin 1 alone formed micrometer-sized droplets (Fig. 1, A and B; fig. S4; and movie S1), with the size of the droplets correlating with its concentration (0.5 to 20 μM concentrations tested) (figs. S2 and S3). The coalescence of synapsin 1 into droplets was confirmed by performing the incubation in suspension and measuring turbidity (fig. S4). These concentrations were not above the physiological range, as synapsin 1 is estimated to reach concentrations above 100 μM in nerve terminals (*16*). Droplets of synapsin 1 had the expected properties of a liquid phase (*9*): They fused with each other (Fig. 1B), and bleaching of several droplets revealed that synapsin 1 molecules swiftly exchanged into and out of synapsin 1 droplets [half-time ($t_{1/2}$) = 65 s; Fig. 1C and movie S2]. Additionally, fluorescence recovery after photobleaching (FRAP) of a small area within the droplet was followed by rapid recovery ($t_{1/2}$ = 40 s) of fluorescence, reflecting local rearrangement of synapsin 1 molecules (Fig. 1D and movie S3). Analysis of two purified fragments of synapsin 1 confirmed that its IDR (amino acids 421 to 706), but not its folded central ATP-binding module (amino acids 113 to 420), which is known to dimerize (*17*), formed droplets (fig. S1B). Droplet formation by the IDR alone was as efficient as droplet formation by the full-length protein (fig. S5). Synapsin 2, a paralog of synapsin 1 that can heterodimerize with synapsin 1 (*15*, *17*), also contains a C-terminal IDR, albeit shorter than the IDR of synapsin 1. Accordingly, synapsin 2 also phase-separated (fig. S6). Increasing the salt concentration above the physiological range impaired droplet formation, implicating charge-dependent interactions in their formation (fig. S7).

Polyvalent interactions between SH3 domain–containing proteins and proteins harboring cognate proline-rich motifs can also generate distinct liquid phases (*8*, *12*). Synapsin 1 interacts with several SH3 domain–containing proteins via its IDR (*19*–*22*). One such protein, intersectin, is a component of a network of protein-protein interactions that facilitates the clustering of SVs in conjunction with synapsin (*22*, *23*). Thus, we examined whether, upon incubation with SH3 domain–containing binding partners such as intersectin (*22*) and growth factor receptor–bound 2 (GRB2) (*19*), synapsin 1 phase-separated together with them. GRB2 and intersectin contain two and five SH3 domains, respectively (*22*).

Synapsin 1 was mixed with either GRB2 (Fig. 2A and movie S4) or a fragment from human intersectin comprising its five SH3 domains [$(SH3)_{5\text{-intersectin}}$] (Fig. 2B and movie S5) (*21*) (all proteins at 10 μM and fused to fluorescent proteins) and incubated at room temperature in physiological salt concentration. After some delay, droplets appeared containing both synapsin and its binding partners (Fig. 2, A and B, and movies S4 and S5). Droplet growth, after the initial nucleation, was faster than with synapsin 1 alone (fig. S4). As before, droplets grew progressively or by fusing with each other, revealing a liquid state. The same concentration (10 μM) was used for synapsin 1 and

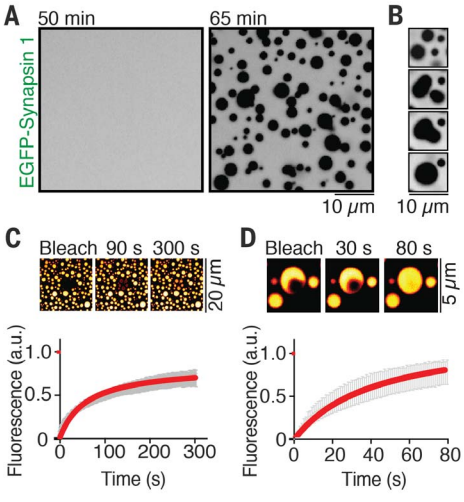


Fig. 1. Synapsin 1 undergoes liquid-liquid phase separation. (**A**) EGFP–synapsin 1 (10 μM) forms droplets when incubated for 1 hour in a buffer of physiological salt concentration at room temperature. (**B**) Droplets of synapsin 1 show liquid behavior by fusing with each other and relaxing into a round-shaped structure, minimizing surface tension. (**C**) Photobleaching of several synapsin 1 droplets with subsequent recovery of fluorescence, as shown by micrograph and by quantification of the fluorescence recovery in the bleached region. (**D**) Fluorescence recovery of synapsin 1 after photobleaching a region within a droplet. Error bars represent SEM, and red shading is the fit with a hyperbolic function. a.u., arbitrary units.

its partners, although synapsin is thought to be the most abundant matrix protein within SV clusters (*16*). This was meant to reflect the presence at synapses of multiple SH3 domain–containing synapsin 1 ligands, and thus a higher collective concentration of these proteins than the concentration of any one of them. The synapsin 1:$(SH3)_{5\text{-intersectin}}$ droplets were larger, which could be explained by the higher valence of the intersectin fragment (five SH3 domains, although with different affinities for synapsin) relative to GRB2 (two SH3 domains) (*22*, *24*).

The cytoplasm of a synaptic bouton is a crowded environment filled with organelles and macromolecules. To mimic this environment in subsequent experiments, we added polyethylene glycol (PEG), a crowding reagent, to the buffer. In the presence of 3% PEG 8000, droplets of synapsin 1 alone (fig. S3) or of synapsin 1 and its binding partners (Fig. 2, C and D) formed immediately, with no lag phase. FRAP confirmed that, even under these conditions, proteins were mobile within the droplets, with a faster recovery time for synapsin 1 in droplets generated with GRB2 ($t_{1/2}$ at 1.5 min) than in those generated with $(SH3)_{5\text{-intersectin}}$ ($t_{1/2}$ at 3.2 min) (Fig. 2, E and F). This, again, possibly reflects the higher valence of this protein relative to GRB2. Recovery of fluorescence was observed both when a region within a droplet (Fig. 2, E and F) and when the entire droplet was bleached (fig. S8), indicating that recovery results from both molecular rearrangements within the droplet and the exchange of molecules with the dilute phase. Presence of a synapsin 1 binding partner, $(SH3)_{5\text{-intersectin}}$, had a biphasic effect on droplet formation. It enhanced this process at low-to-moderate excess stoichiometric ratio but inhibited it when added in large excess (fig. S9). Thus, $(SH3)_{5\text{-intersectin}}$ is not only a "client" of synapsin but also an active player in the formation of the liquid condensate. The negative effect at high stoichiometric ratio may be due to the masking of sites within the synapsin IDR that interact with each other. Recruitment of synapsin interactors into the droplets was specific and did not occur with proteins that do not bind synapsin (fig. S10).

Synapsin binds SVs (*15*, *18*, *25*). If its phase-separating properties are involved in SV cluster formation, synapsin 1 should be capable of capturing vesicles into such a phase. We thus incubated synapsin 1 with small lipid vesicles (~50 to 150 nm in diameter) mimicking SVs in lipid composition supplemented with a fluorescently labeled lipid, 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(cyanine 5) (DOPE-Cy5). Formation of synapsin droplets correlated with the appearance of droplets positive for the labeled lipid, whereas no droplets positive for the lipid were observed in the absence of synapsin 1 (Fig. 3A). Synapsin 1 condensates did not recruit vesicles lacking negatively charged phospholipids (fig. S11), which is expected, given that negatively charged lipids are necessary for synapsin binding to vesicles (*26*). Other well-characterized protein liquid condensates that do not bind lipid membranes, such as droplets composed of cytoplasmic protein noncatalytic region of tyrosine kinase adaptor protein 1 (NCK) and a fragment of neural Wiskott-Aldrich syndrome protein (N-WASP) that does not include the phospholipid-binding region (*8*), did not sequester lipid vesicles (fig. S12). Lipid vesicles were mobile within the synapsin phase, as indicated by FRAP (fig. S13). Furthermore, lipid vesicles and SH3-domain synapsin 1 interactors coassembled with synapsin (fig. S14). Electron microscopy (EM) analysis showed that these droplets were represented by clusters of small vesicles, whereas in the absence of synapsin 1, vesicles remained dispersed (Fig. 3, B and C).

Synapsin 1 is a major presynaptic phosphoprotein that undergoes multisite phosphorylation (*14*). Sustained nerve-terminal stimulation to trigger massive neurotransmitter release also induces the calcium-dependent phosphorylation

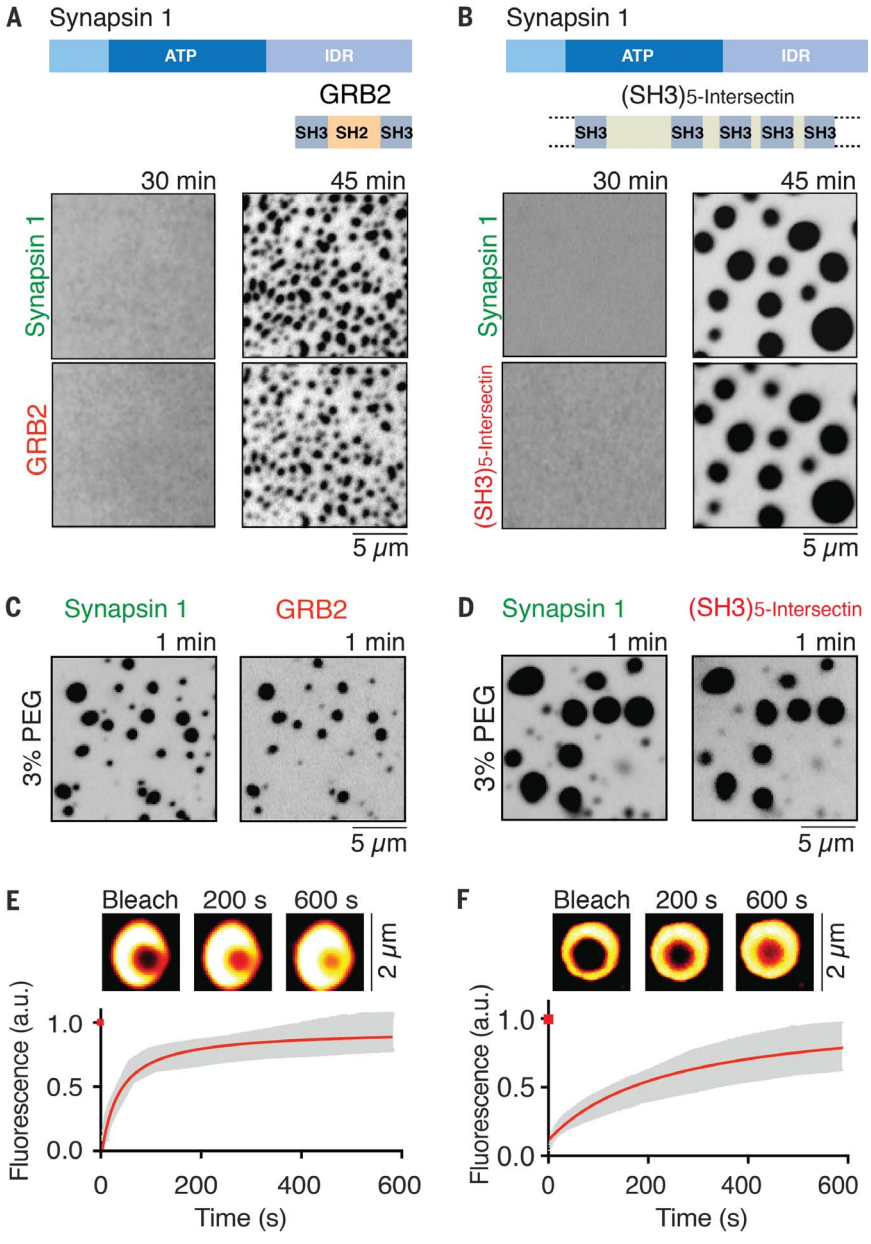


Fig. 2. Synapsin 1 drives phase separation of SH3 domain–containing binding partners. (**A** and **B**) Full-length synapsin 1 (10 μM) and either GRB2 (10 μM) (A) or the SH3 domain–containing region of intersectin (10 μM) (B) form droplets under physiological conditions. The domain organization of the proteins is shown at the top, and fluorescence images of the protein mixtures at 30 and 45 min are shown at the bottom. (**C** and **D**) Fluorescence images of the solution immediately (within 1 min) after mixing of synapsin 1 with GRB2 (C) or $(SH3)_{5\text{-intersectin}}$ (D) in the presence of the crowding reagent (3% PEG 8000). (**E** and **F**) Fluorescence recovery of synapsin 1 after photobleaching a region within a synapsin 1–GRB2 droplet (E) or a synapsin 1–$(SH3)_{5\text{-intersectin}}$ droplet (F). Error bars represent SEM, and red shading is the fit with a hyperbolic function.

of synapsin 1 (*27*). This results in its dissociation from SVs and dispersion within the nerve-terminal cytosol (*28*, *29*), as SVs are consumed by exocytosis. If the formation of a biomolecular condensate by synapsin has a physiological importance in its coassembly with SVs, one would expect synapsin 1 droplets to disassemble upon calcium-dependent phosphorylation. Two prominent phosphorylation sites for calcium/calmodulin-dependent protein kinase II (CaMKII), called sites 2 and 3, are present in its IDR (*14*, *15*). Addition of CaMKII, calcium, and calmodulin to synapsin 1–containing samples did not disperse droplets. Indeed, CaMKII, which binds synapsin 1 (*29*), was recruited into the droplets

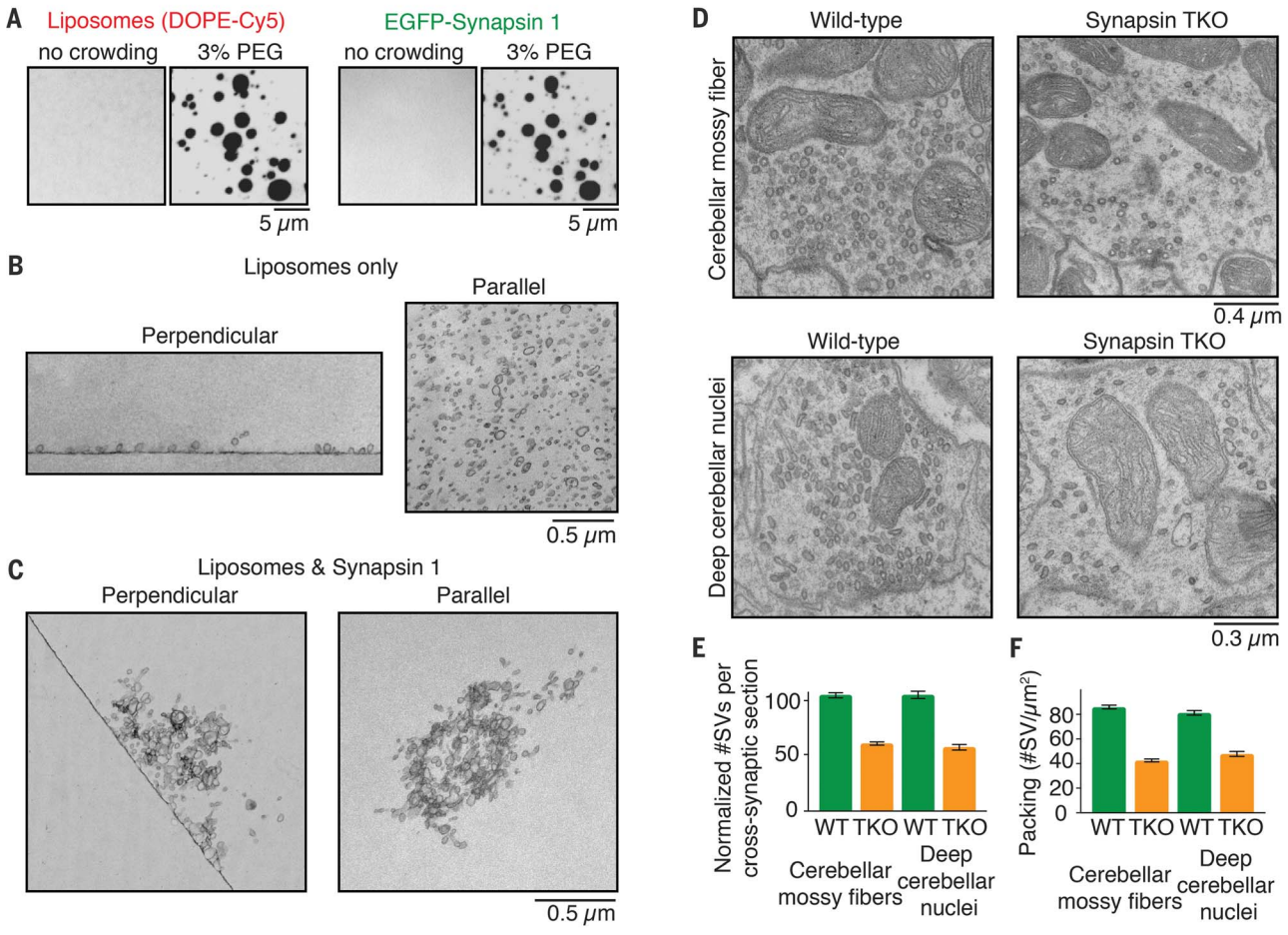


Fig. 3. Synapsin 1 condensates are reaction centers that are able to sequester lipid vesicles. (**A**) Fluorescence images showing the detection of liposomes (left, DOPE-Cy5) and synapsin 1 (right, EGFP–synapsin 1) in a mixture of liposomes and synapsin 1 either without or with crowding agent (3% PEG 8000). (**B**) EM images of liposomes incubated without synapsin 1 in the same buffer conditions used for (A). Left panel shows section perpendicular to the liposome-glass interface. Right panel shows section comprising the layer of liposomes adsorbed to the glass surface. (**C**) Same as in (B), but showing liposomes incubated with synapsin 1. The field shown at right is from a section parallel to the glass surface but above the glass interface. (**D**) EM images of synapses from cerebellar mossy fibers (top) and deep cerebellar nuclei (bottom) obtained from adult wild-type (WT, left) and synapsin triple-knockout (TKO, right) mice. (**E**) Number of SVs in synaptic cross sections of WT and synapsin TKO mice, normalized to WT. (**F**) Number of SVs per unit area of synaptic section in WT and synapsin TKO mice. For each condition, 50 sections from three independent animals were examined. Error bars represent SEM.

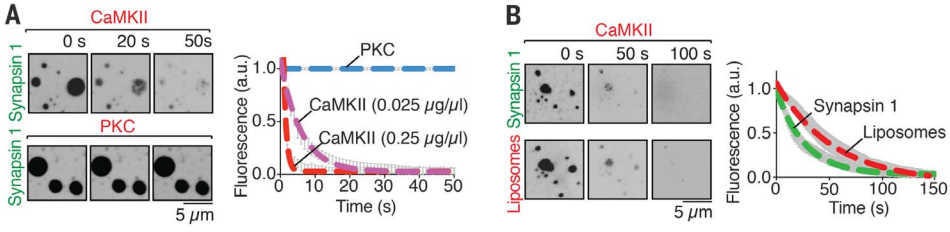


Fig. 4. Phosphorylation of the IDR of synapsin 1 disperses condensates of either synapsin 1 alone or synapsin 1 and liposomes. (**A**) Left panel shows fluorescence images of EGFP–synapsin 1 condensates preincubated with either CaMKII (0.025 μg/μl), calmodulin, and calcium or protein kinase C (PKC), phosphatidylserine, diacylglycerol, and calcium upon addition of ATP (200 μM) at 0 s, demonstrating dispersion of synapsin by CaMKII but not by PKC. Right panel shows time course of the effect of the kinases on the condensates, as assessed by the decrease of fluorescence on regions of interest corresponding to randomly selected droplets. (**B**) Left panel shows fluorescence images of liposome-synapsin condensates preincubated with CaMKII (0.25 μg/μl), calmodulin, and calcium upon addition of ATP (200 μM) at 0 s, demonstrating dispersion of both synapsin and liposomes. Right panel shows time course of the effect of CaMKII on liposome-synapsin droplets dispersion. Error bars represent SEM, and dashed lines represent the fit with a single exponential function. Incubations were carried out at room temperature in a buffer of physiological salt concentration supplemented with 3% PEG 8000.

(Fig. 4A). However, further addition of ATP (200 μM) to induce synapsin 1 phosphorylation caused rapid dispersal of both synapsin 1 and CaMKII [mean lifetime (τ) of 5.9 s] (Fig. 4A, fig. S15, and movie S6). Importantly, CaMKII also disassembled synapsin 1–liposome droplets (Fig. 4B and movie S7). As a control, we added protein kinase C to the droplets, for which synapsin 1 is not a substrate (*30*), but neither addition of the kinase nor the subsequent addition of ATP (200 μM) affected the droplets (Fig. 4A). The lack of effect of ATP in the latter experiment also rules out that droplet dispersion may be explained by a hydrotrope action of this nucleotide (*31*). Such an action, as reported for liquid droplets generated by other proteins, occurs only at much higher ATP concentrations (fig. S16) than the one used in our phosphorylation assay (see also fig. S4).

Strong evidence points to a physiological master role of synapsin in the clustering of SVs at living synapses (*25*, *32*–*34*). In neuronal cultures from mice that lack all three synapsins, the number of SVs at synapses is lower than that in wild-type mice, and this decrease is selective for SVs away from active zones (*33*). We extended these results. Even at synapses in situ—both excitatory (cerebellar mossy fibers) and inhibitory (deep cerebellar nuclei) nerve terminals—not only the total number but also the packing of SVs was notably lower in synapsin triple-knockout mice than in wild-type mice (Fig. 3, D to F).

Collectively, these findings demonstrate that synapsin can form a separate liquid biomolecular condensate either alone or together with binding partners for its IDR, with lipid vesicles, or with both. Interactions occurring at the presynapse in situ are expected to be more complex than the interactions of synapsin with the two SH3 domain–containing proteins and artificial lipid membranes described here. Because SVs are membranous organelles selectively recruited into clusters, there must be additional factors that help provide specificity. However, the minimal systems used here provide some insight into the mechanisms responsible for the properties of SV clusters. Clusters of other membranous organelles may self-organize according to similar principles without the need for a surrounding membrane or protein-based structure to confine them.

REFERENCES AND NOTES

1. D. Milovanovic, P. De Camilli, *Neuron* **93**, 995–1002 (2017).
2. S. O. Rizzoli, W. J. Betz, *Science* **303**, 2037–2039 (2004).
3. K. J. Darcy, K. Staras, L. M. Collinson, Y. Goda, *Nat. Neurosci.* **9**, 315–321 (2006).
4. C. Acuna, X. Liu, T. C. Südhof, *Neuron* **91**, 792–807 (2016).
5. S. S. H. Wang *et al*., *Neuron* **91**, 777–791 (2016).
6. K. Kraszewski *et al*., *J. Neurosci.* **15**, 4328–4342 (1995).
7. C. P. Brangwynne *et al*., *Science* **324**, 1729–1732 (2009).
8. P. Li *et al*., *Nature* **483**, 336–340 (2012).
9. S. F. Banani, H. O. Lee, A. A. Hyman, M. K. Rosen, *Nat. Rev. Mol. Cell Biol.* **18**, 285–298 (2017).
10. A. Patel *et al*., *Cell* **162**, 1066–1077 (2015).
11. J. B. Woodruff *et al*., *Cell* **169**, 1066–1077.e10 (2017).
12. S. F. Banani *et al*., *Cell* **166**, 651–663 (2016).
13. P. De Camilli, S. M. Harris Jr., W. B. Huttner, P. Greengard, *J. Cell Biol.* **96**, 1355–1373 (1983).
14. W. B. Huttner, W. Schiebler, P. Greengard, P. De Camilli, *J. Cell Biol.* **96**, 1374–1388 (1983).
15. T. C. Südhof *et al*., *Science* **245**, 1474–1480 (1989).
16. B. G. Wilhelm *et al*., *Science* **344**, 1023–1028 (2014).
17. M. Hosaka, T. C. Südhof, *J. Biol. Chem.* **274**, 16747–16753 (1999).
18. F. Benfenati, M. Bähler, R. Jahn, P. Greengard, *J. Cell Biol.* **108**, 1863–1872 (1989).
19. P. S. McPherson *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6486–6490 (1994).
20. F. Onofri *et al*., *J. Biol. Chem.* **275**, 29857–29867 (2000).
21. M. Yamabhai *et al*., *J. Biol. Chem.* **273**, 31401–31407 (1998).
22. F. Gerth *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12057–12062 (2017).
23. Å. M. E. Winther *et al*., *J. Neurosci.* **35**, 14756–14770 (2015).
24. A. Pechstein *et al*., *EMBO Rep.* **16**, 232–239 (2015).
25. P. De Camilli, F. Benfenati, F. Valtorta, P. Greengard, *Annu. Rev. Cell Biol.* **6**, 433–460 (1990).
26. F. Benfenati, P. Greengard, J. Brunner, M. Bähler, *J. Cell Biol.* **108**, 1851–1862 (1989).
27. W. B. Huttner, L. J. DeGennaro, P. Greengard, *J. Biol. Chem.* **256**, 1482–1488 (1981).
28. P. Chi, P. Greengard, T. A. Ryan, *Nat. Neurosci.* **4**, 1187–1193 (2001).
29. F. Benfenati *et al*., *Nature* **359**, 417–420 (1992).
30. J. K. Wang, S. I. Walaas, P. Greengard, *J. Neurosci.* **8**, 281–288 (1988).
31. A. Patel *et al*., *Science* **356**, 753–756 (2017).
32. V. A. Pieribone *et al*., *Nature* **375**, 493–497 (1995).
33. D. Gitler *et al*., *J. Neurosci.* **24**, 11368–11380 (2004).
34. L. Siksou *et al*., *J. Neurosci.* **27**, 6868–6877 (2007).

ACKNOWLEDGMENTS

We thank A. C. Nairn for providing CaMKII and for discussion and the HHMI Summer Institute at the Marine Biology Laboratory for providing NCK and N-WASP. **Funding:** P.D.C. was supported by grants from the NIH (R37NS036251 and DA018343), and D.M. and X.B. were supported by a long-term postdoctoral fellowship from the Human Frontier Science Program. **Author contributions:** D.M. and P.D.C. designed the research and wrote the paper. D.M. designed and performed all of the imaging studies, FRAP and phosphorylation experiments, and liposome clustering. Y.W. performed EM analysis. X.B. helped with protein preparation. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/361/6402/604/suppl/DC1
Materials and Methods
Figs. S1 to S16
Table S1
References (*35*–*38*)
Movies S1 to S7

13 March 2018; accepted 25 June 2018
Published online 5 July 2018
10.1126/science.aat5671

SIGNAL TRANSDUCTION

Apoptosis propagates through the cytoplasm as trigger waves

Xianrui Cheng[1] and James E. Ferrell Jr.[1,2]*

Apoptosis is an evolutionarily conserved form of programmed cell death critical for development and tissue homeostasis in animals. The apoptotic control network includes several positive feedback loops that may allow apoptosis to spread through the cytoplasm in self-regenerating trigger waves. We tested this possibility in cell-free *Xenopus laevis* egg extracts and observed apoptotic trigger waves with speeds of ~30 micrometers per minute. Fractionation and inhibitor studies implicated multiple feedback loops in generating the waves. Apoptotic oocytes and eggs exhibited surface waves with speeds of ~30 micrometers per minute, which were tightly correlated with caspase activation. Thus, apoptosis spreads through trigger waves in both extracts and intact cells. Our findings show how apoptosis can spread over large distances within a cell and emphasize the general importance of trigger waves in cell signaling.

Xenopus laevis eggs are large cells, ~1.2 mm in diameter, that are naturally arrested in metaphase of meiosis II. The eggs ultimately adopt one of two fates: Either they become fertilized and enter the embryonic cell cycle, or they remain unfertilized and die, usually through apoptosis. Apoptosis is a relatively nonperturbing form of cell death and may allow the frog to resorb old oocytes and to clean up any eggs retained in the body without provoking an inflammatory response (*1*). The powerful biochemical approaches provided by the *Xenopus* system have made *Xenopus* eggs and extracts useful model systems for the study of apoptosis (*2*, *3*).

The unusual size of the *Xenopus* egg raises the question of how an all-or-none, global process such as apoptosis spreads through the cell. One possibility is that apoptosis spreads through the egg by random walk diffusion, ultimately taking over all of the cytoplasm. A second possibility is suggested by the existence of multiple positive and double-negative feedback loops in the regulatory network that controls apoptosis (Fig. 1A). These loops may allow apoptosis to propagate through self-regenerating trigger waves. Trigger waves are propagating fronts of chemical activity that maintain a constant speed and amplitude over large distances. They can arise when bistable biochemical reactions are subject to diffusion or, more generally, when bistability or something akin to bistability (e.g., excitability or relaxation oscillation) is combined with a spatial coupling mechanism (e.g., diffusion or cell-cell communication) (*4*–*6*). Familiar examples include action potentials; calcium waves; and the spread of a fire through a field, a favorable allele through a population, or a meme through the internet. Trigger waves are an important general mechanism for long-range biological communication, and apoptotic trigger waves may allow death signals to spread rapidly and without diminishing in amplitude, even through a cell as large as a frog egg.

To distinguish between diffusive spread and trigger waves in the propagation of apoptotic signals, we used undiluted cell-free extracts from *Xenopus* eggs (*7*). Such extracts can be placed in long tubes (several millimeters) and imaged by video microscopy (*8*, *9*). Over millimeter distances, the distinction between diffusive spread of apoptosis, which would slow down with increasing distance, and trigger waves, which maintain a constant speed and amplitude, should be readily apparent.

We incubated one portion of a cycloheximide-treated interphase cytoplasmic extract with horse cytochrome *c* (2.4 μM) and verified that caspase-3 and/or caspase-7, executioner caspases (*10*) that have similar peptide sequence specificities, became activated (fig. S1). A 10-kDa Texas Red–conjugated dextran was added to the apoptotic extract as a diffusion marker, and the extract was pipetted into a large (560-μm inner diameter) Teflon reservoir. We then took a second portion of the interphase extract, with added sperm chromatin plus a chimeric protein consisting of glutathione *S*-transferase, green fluorescent protein, and a nuclear localization sequence (GST-GFP-NLS) but no cytochrome *c*, and introduced it into a thin (150-μm inner diameter) Teflon tube. The nuclei act as an easily assessed indicator of apoptosis: In nonapoptotic extracts, GST-GFP-NLS concentrates in the nuclei that form from the sperm chromatin, whereas in an apoptotic extract, caspases attack components of the nuclear pore (*11*) and allow the GST-GFP-NLS to leak out and disperse. The tube containing this naïve extract was gently inserted a short distance into the reservoir containing the apoptotic extract, and the two tubes were placed under mineral oil (Fig. 1B). We used time-lapse fluorescence microscopy to determine whether apoptosis spread up the thin tube in a diffusive fashion, with propagation slowing as apoptosis proceeded, or at a constant velocity as expected for trigger waves.

As shown in movie S1 and Fig. 1, B and C, apoptosis progressed up the thin tube at a constant speed of 27 μm/min over a distance of several millimeters. In five independent experiments, apoptosis always propagated linearly, without showing any signs of slowing down or diminishing, and the average speed was 29 ± 2 μm/min (mean ± SD). In contrast, the 10-kDa dye spread only a few hundred micrometers (Fig. 1B), implying that neither simple diffusion nor any unintended mixing could account for the spread of apoptosis.

If the apoptotic signals are propagated by trigger waves, one prediction is that the waves should be self-sustaining; that is, once the activity is established in the thin tube, continued contact with the reservoir of apoptotic extract would not be required. We tested this possibility by inserting the thin tube into the apoptotic reservoir for 20 min and then removing it and monitoring apoptosis. The apoptotic activity propagated from the induction terminus to the distal terminus at a constant speed of 32 μm/min, consistent with a self-sustaining process (fig. S2). This procedure was used for all subsequent experiments because it provided more reliable focusing for the imaging and allowed more tubes to be imaged per experiment.

A second way of detecting apoptosis is with the fluorogenic caspase substrate carboxybenzyl–Asp-Glu-Val-Asp–rhodamine 110 (Z-DEVD-R110) (*12*). This probe is a rhodamine derivative (R110) with two four–amino acid (DEVD) peptides linked to the fluorophore. It is nonfluorescent when the DEVD-fluorophore bonds are intact but becomes strongly fluorescent once they are hydrolyzed by caspase-3 or -7. We added the probe (2 μM) to an extract and initiated apoptosis as before. Fluorescence spread up the tube at a constant speed (in this experiment, 33 μm/min). In eight experiments, the average speed was 30 ± 3 μm/min (mean ± SD) (movie S2 and Fig. 1D). We also added sperm chromatin and GST-mCherry-NLS as well as Z-DEVD-R110 to compare the propagation of the caspase activity wave with the disappearance of the reconstituted nuclei. In this experiment, the disappearance of the nuclei lagged 40 min behind the front of Z-DEVD-R110 fluorescence (Fig. 1E); in four independent experiments, the lag was 35 ± 5 min (mean ± SD). The speeds of the Z-DEVD-R110 and GST-mCherry-NLS waves—the slopes of the dashed lines in Fig. 1E—were indistinguishable.

We sought to understand the mechanistic basis for the trigger waves. The apoptotic control system includes multiple positive feedback loops (Fig. 1A). One positive feedback circuit (designated "caspase loop" in Fig. 1A) involves only cytosolic proteins, including caspases-3, -7, and -9 and XIAP (X-linked inhibitor of apoptosis protein). Another involves the system that regulates the Bak and Bax proteins (the "BH3 protein loop" in Fig. 1A), which, when activated, bring about mitochondrial outer membrane permeabilization (MOMP). One long loop, in which

[1]Department of Chemical and Systems Biology, Stanford University School of Medicine, Stanford, CA 94305-5174, USA. [2]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305-5307, USA.
*Corresponding author. Email: james.ferrell@stanford.edu

cytosolic cytochrome *c* stimulates caspase-3 and -7 activation and caspase activation stimulates cytochrome *c* release (the "MOMP/caspase/BH3 protein loop"), connects the two shorter loops (Fig. 1A). We tested whether the cytosolic components could support trigger waves in the absence of mitochondria and whether the mitochondrial components could support trigger waves in the absence of the activation of caspase-3 and caspase-7.

We fractionated a crude cytoplasmic extract (*2*, *3*) (fig. S3A) and verified by immunoblotting that the resulting cytosol was largely devoid of mitochondria, as indicated by the absence of a mitochondrial marker protein, the voltage-dependent anion channel (fig. S3B). Horse cytochrome *c* (2.4 μM) was added to this cytosolic fraction, and the activity of caspase-3 and/or -7 was assessed by the chromogenic caspase assay. In agreement with previous reports (*3*, *13*), the caspases were briskly activated (Fig. 2A). Thus, mitochondria are not essential for cytochrome *c*–induced activation of executioner caspases in *Xenopus* extracts.

We tested whether the cytosolic extract could support apoptotic trigger waves. Because nuclei cannot be reconstituted in cytosolic extracts, we used the fluorogenic Z-DEVD-R110 probe for this experiment. The fluorescence spread up the tube in a sublinear fashion (movie S3 and Fig. 2E). To see if this spread was consistent with simple random walk diffusion, we identified points along the propagation front with equal fluorescence intensities (fig. S4, A and B) and plotted the distance squared (x^2) versus time (t) (fig. S4C). The x^2-versus-t relationship was linear for at least an hour (fig. S4C), consistent with random walk diffusion. Thus, over this time scale and this distance scale, we found no evidence for a trigger wave.

To further test the role of the mitochondria, we added back purified mitochondria to the cytosolic extract at a 3% volume-to-volume (v/v) ratio, which is approximately physiological [estimated from the volumes we obtained for the various fractions (see fig. S3) and the concentrations used by others in reconstitution studies (*3*)], and again determined whether apoptosis would propagate diffusively or as a trigger wave. As expected, the reconstituted extract activated caspase-3 and/or -7 in response to horse cytochrome *c* (Fig. 2C). Moreover, the reconstitution restored the trigger waves (movie S3 and Fig. 2D). The propagation distance increased linearly with time (Fig. 2D), as it did in crude cytoplasmic extracts

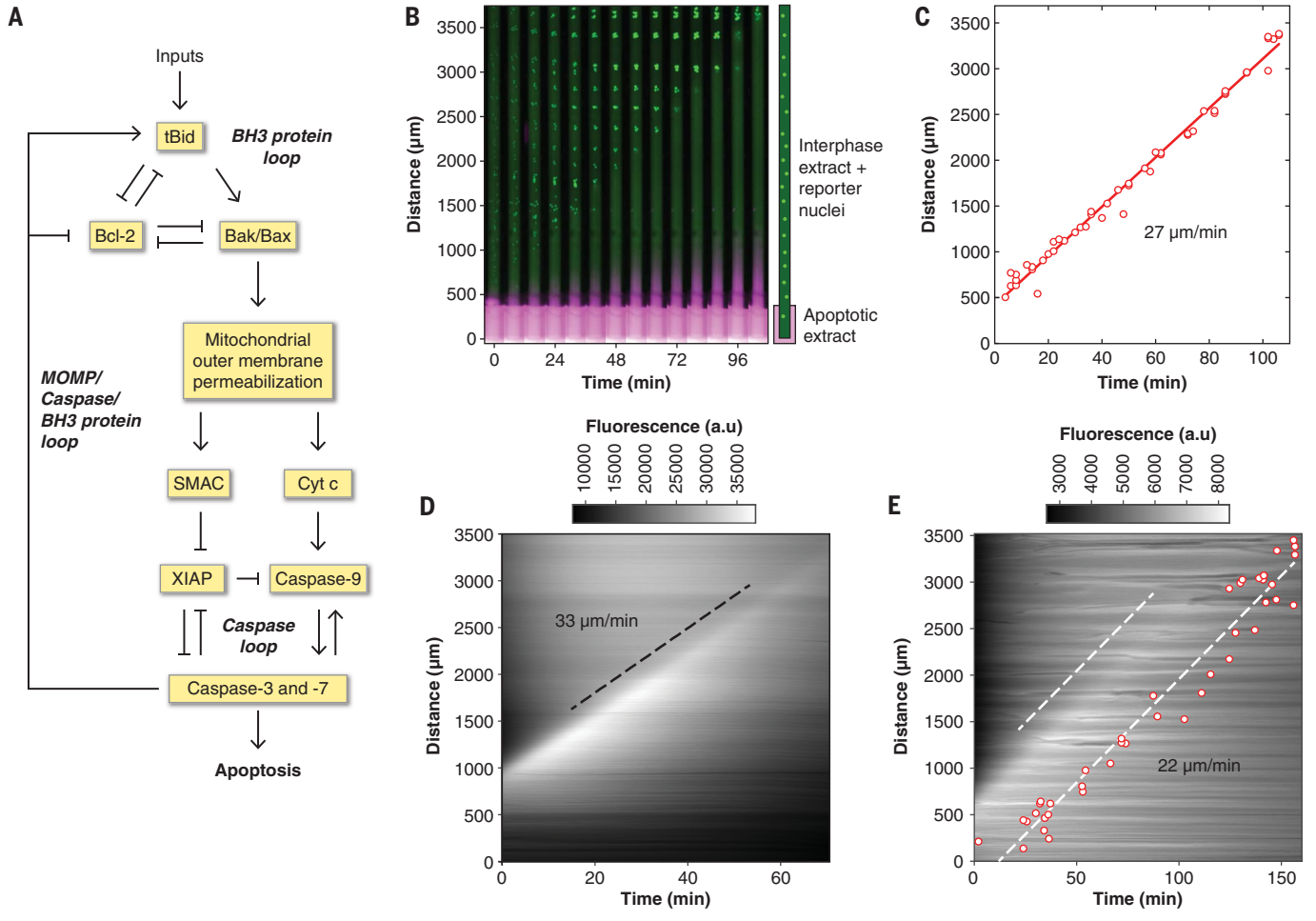


Fig. 1. Apoptosis propagates in interphase-arrested cytoplasmic *X. laevis* egg extracts through trigger waves. (**A**) The control circuit for apoptosis, as conceptualized on the basis of the present study and others' previously published work (*24*–*28*). Cyt c, cytochrome *c*. (**B**) Time-lapse montage of GST-GFP-NLS–filled nuclei (green) in a cytoplasmic extract in a Teflon tube with its lower end in contact with an apoptotic extract reservoir. The extract in the reservoir is marked with 10-kDa dextran–Texas Red dye, shown in magenta. A time-lapse video of this experiment can be found in movie S1. (**C**) Correlation between timing and position of nuclear disappearance for the experiment depicted in (B). The line is a linear least-squares fit to the data. The propagation speed (the slope of the fitted line) is 27 μm/min. (**D**) Kymograph image showing the spatial propagation of caspase-3 and/or caspase-7 activity (indicated by R110 fluorescence) in a crude cytoplasmic extract. The dashed line was manually fitted to the fluorescence front, and it yielded a propagation speed of 33 μm/min. a.u., arbitrary units. (**E**) R110 fluorescence and nuclear disappearance, detected by using GST-mCherry-NLS as a nuclear marker, measured in the same tube. The presence of the nuclei makes the R110 fluorescence less diffuse than it is in (D). One dashed line is manually fitted to the fluorescence front, and the other is a least-squares fit to the nuclear data. The propagation speed was 22 μm/min for both waves.

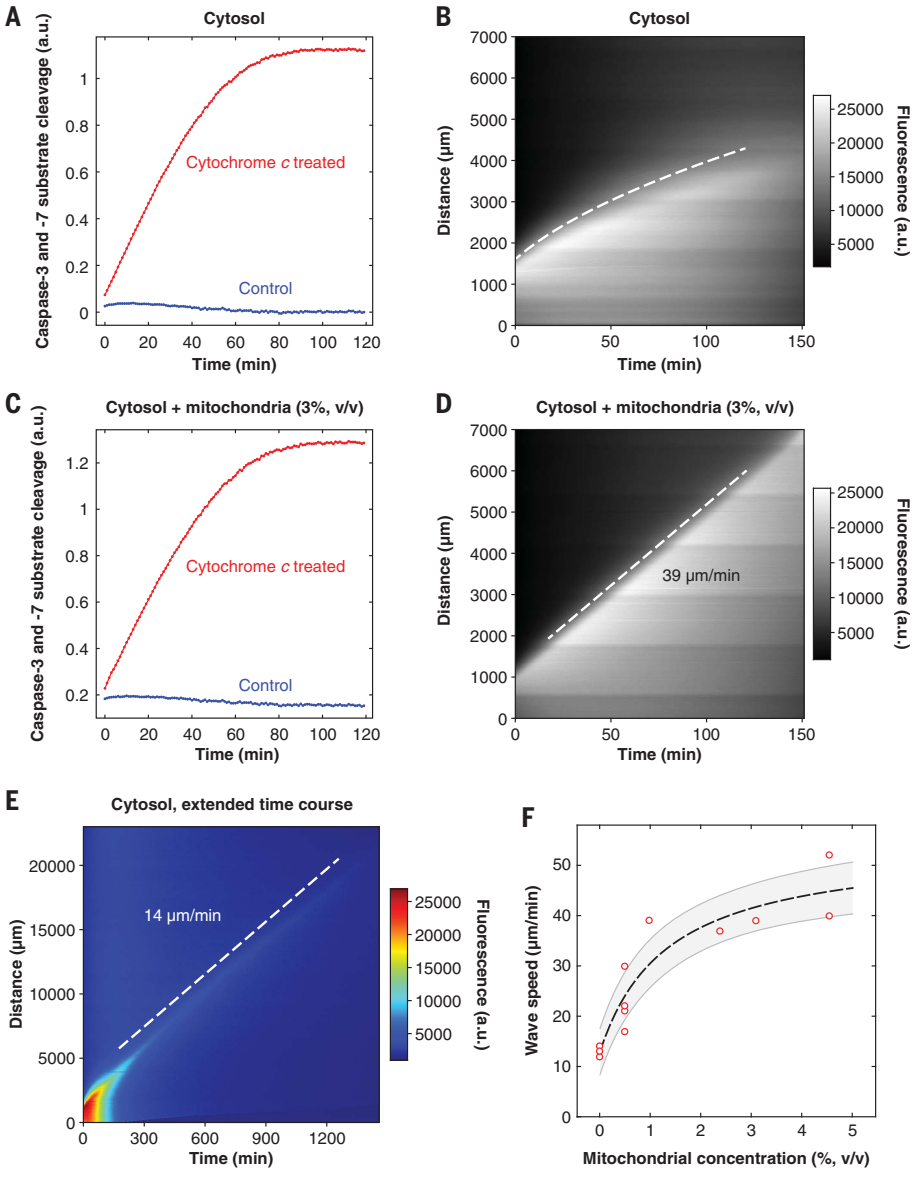


Fig. 2. The speed of the apoptotic trigger wave depends upon the concentration of mitochondria. (**A** and **B**) Cytosolic extract. (A) Activation of caspase-3 and/or caspase-7 as shown by a chromogenic assay. (B) Kymograph image showing diffusive spread of caspase-3 and/or -7 activation, as read out by the Z-DEVD-R110 probe, over the indicated time scale and distance scale. The dashed curve was obtained by defining an equal-fluorescence isocline, replotting the isocline on a distance squared–versus–time plot, carrying out a linear least-squares fit, and then transforming the fitted line for plotting on the original distance-versus-time axes. Further details are provided in fig. S4. (**C** and **D**) Reconstituted extract from the same experiment. (C) Activation of caspase-3 and/or -7. (D) Kymograph image showing trigger wave propagation of caspase activation. (**E**) Slow apoptotic trigger waves detected in cytosolic extracts. This kymograph image represents a cytosolic extract incubated for 24 hours in a 3-cm tube. R110 fluorescence is displayed here on a heat map scale to allow the shape of the wave front to be appreciated both early and late in the time course. (**F**) Wave speed as a function of mitochondrial concentration. Data are from 18 tubes and 8 independent experiments (there are 9 overlapping data points with 0% mitochondria). The dashed line is a Michaelian dose-response curve given by the equation $y = y_0 + y_{max}\frac{x}{K+x}$, where y is the trigger wave speed, x is the mitochondrial concentration, and y_0, y_{max}, and K are parameters determined by fitting the data. The fitted parameters were y_0 = 13.0 ± 1.4 μm/min, y_{max} = 41.3 ± 6.1 μm/min, and K = 1.3 ± 0.6% (means ± SE) and r^2 (coefficient of determination) = 0.98. The region bounded by gray lines represents the SE (68.2% confidence interval) single-prediction confidence band calculated with Mathematica 11.1.1.

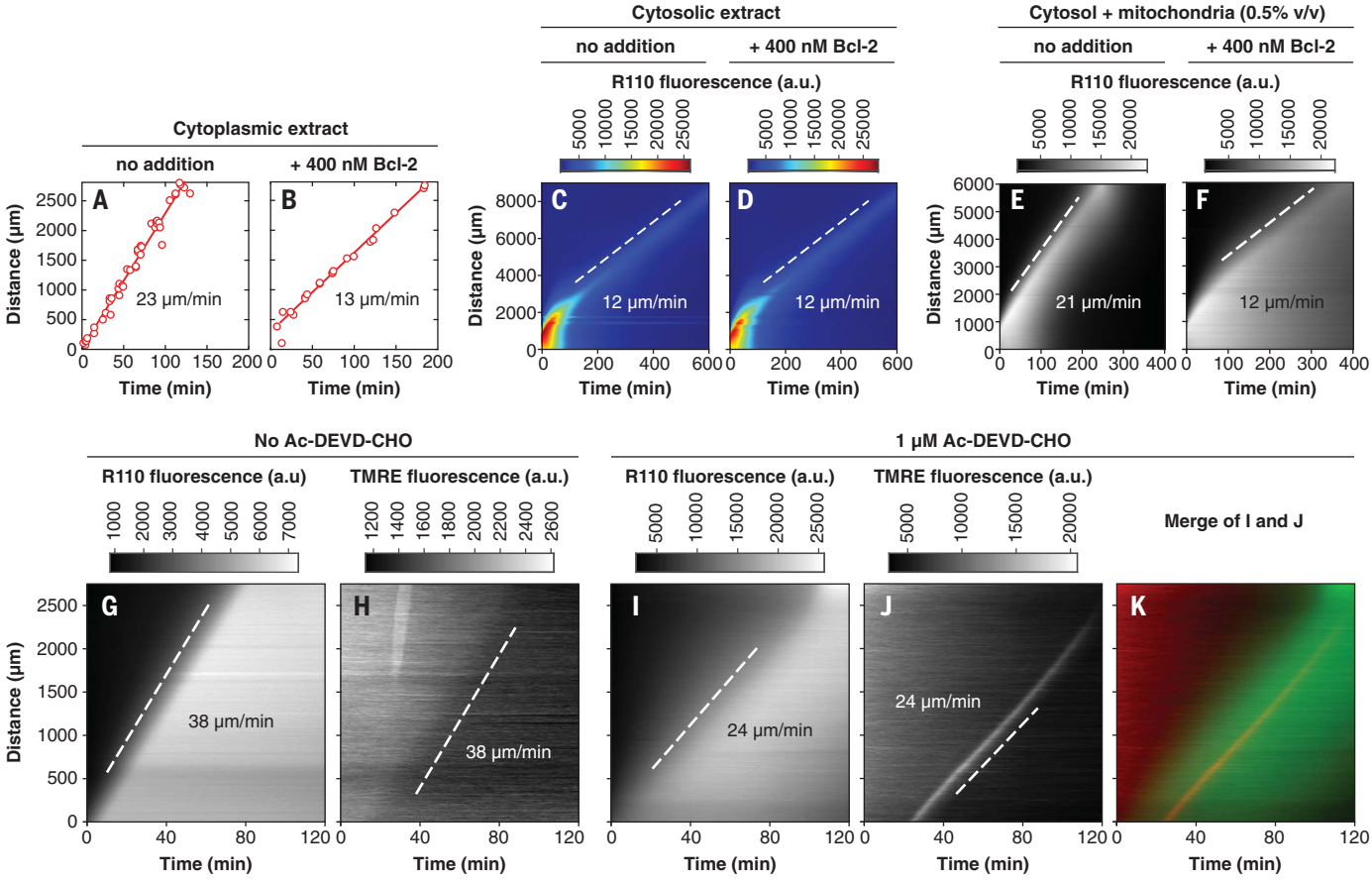
(Fig. 1), and propagation occurred at a constant speed of 39 μm/min, somewhat faster than that observed in cytoplasmic extracts. The signal propagated over a long distance (6000 μm) with little loss of amplitude and no loss of speed (Fig. 2D).

Over a distance scale of a millimeter or so, it is easy to distinguish a ~30-μm/min trigger wave from diffusive spread of even a rapidly diffusing small molecule such as R110. However, by the end of the experiment depicted in Fig. 1B, the speed of the caspase-3 wave had fallen to only ~15 μm/min; a trigger wave any slower than that would be outpaced by diffusion over the same distance. Thus, to determine whether trigger waves were abolished or only slowed in cytosolic extracts, we made use of longer tubes (up to 3 cm) and longer time courses (up to 24 hours). Display of R110 fluorescence on a pseudocolor heat map scale made it easier to distinguish the wave front at both early times and late times (Fig. 2E). As was the case in Fig. 2B, the speed of the wave front fell during the first ~120 min, consistent with diffusive propagation, but once the speed reached ~14 μm/min it remained constant for many hours (Fig. 2E). This suggests that purified cytosol is capable of generating apoptotic trigger waves, albeit with a substantially lower speed than that seen in cytoplasm or in cytosol supplemented with mitochondria.

Trigger wave speeds were measured in eight independent experiments with various concentrations of mitochondria (Fig. 2F). From curve fitting, we determined the trigger wave speed to be half maximal at a mitochondrial concentration of 1.3 ± 0.6% (mean ± SE), which is estimated to be ~40% of the physiological mitochondrial concentration in *Xenopus* eggs. Thus, an average concentration of mitochondria is sufficient to generate apoptotic trigger waves of near-maximal speed, and the wave speed would be expected to drop in mitochondrion-poor regions of the cytoplasm.

Apoptosis almost always initiated first at the end of the tube that was dipped in the apoptotic extract (Figs. 1 and 2 and fig. S2). However, in experiments with either cytoplasmic extracts or reconstituted extracts, more than half of the time (in 12 of 22 or 21 of 37 tubes, respectively) a second spontaneous apoptotic wave emerged elsewhere in the tube (movie S4 and fig. S5A). The velocities of the induced and spontaneous trigger waves were indistinguishable (44 μm/min in the example shown in fig. S5A), indicating that they probably represent the same basic phenomenon. The entire tube of extract usually became apoptotic, as assessed by the Z-DEVD-R110 probe, after 2 hours of incubation (fig. S5) and invariably by 4 hours. This global activation of caspase-3 and -7 did not occur in cytosolic extracts (which allowed the extended-time-course experiment depicted in Fig. 2E).

The dependence of the wave speed upon the concentration of mitochondria implies that the BH3-domain proteins that regulate MOMP may function in generating the trigger waves (Fig. 1A). To further test this possibility, we added recombinant GST–Bcl-2 protein to cytoplasmic



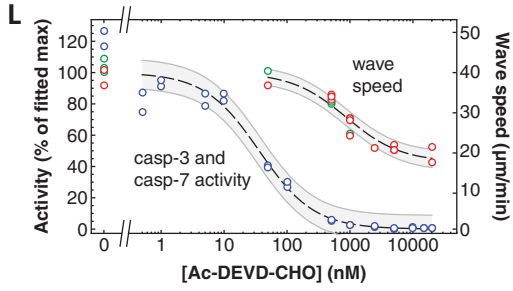


Fig. 3. Both GST–Bcl-2 addition and inhibition of caspase-3 and caspase-7 affect the speed of the trigger waves. (**A** and **B**) GST–Bcl-2 reduces the speed of trigger waves in cytoplasmic extracts. Apoptosis was monitored by the disappearance of reporter nuclei containing GST-mCherry-NLS. (**C** and **D**) GST–Bcl-2 has no effect on the speed of trigger waves in cytosolic extracts. Apoptosis was detected with Z-DEVD-R110, whose fluorescence can be activated by caspase-3 or caspase-7. The pseudocolor heat map scale allows both the initial and final shapes of the wave front to be discerned. (**E** and **F**) GST–Bcl-2 reduces the speed of trigger waves in extracts reconstituted with cytosol and mitochondria (0.5% v/v). (**G** to **K**) Inhibition of caspase-3 and -7 slows trigger waves.
The reporters were Z-DEVD-R110 and the mitochondrial probe TMRE. [(G) and (H)] Cytosolic extract reconstituted with mitochondria (2.4% v/v). [(I) to (K)] A reconstituted extract treated with the caspase inhibitor Ac-DEVD-CHO (1 μM). In experiments with higher concentrations of Ac-DEVD-CHO, a brief (~10 min) period when the TMRE wave appeared to be parabolic rather than linear was observed (fig. S9). (**L**) Inhibition of caspase-3 and -7 activity and slowing of trigger waves as a function of Ac-DEVD-CHO concentration in cytosolic extracts reconstituted with mitochondria (2.4% v/v). Blue data points show caspase activities, measured in extracts diluted 1:20. Green data points indicate wave speeds estimated from Z-DEVD-R110 fluorescence, and red data points indicate wave speeds estimated from TMRE fluorescence. The curves are fits to a Michaelian inhibition function, $y = y_0 + (y_{max} - y_0)\frac{K}{K+x}$, where y is the caspase activity or trigger wave speed, x is the Ac-DEVD-CHO concentration, and y_0, y_{max}, and K are parameters determined by fitting the data. For the caspase activity curve, the fitted parameter values were y_0 = 0.1 ± 2.5, y_{max} = 100 ± 3, and K = 36 ± 8 nM (mean ± SE). For the wave speed curve, the fitted parameter values were y_0 = 17 ± 1 μm/min, y_{max} = 40 ± 1 μm/min, and K = 856 ± 160 nM (mean ± SE). The regions bounded by gray lines represent the SE (68.2% confidence interval) single-prediction confidence bands calculated with Mathematica 11.1.1.

extracts to see whether trigger waves were affected. Bcl-2 is a stoichiometric inhibitor of the pro-apoptotic truncated Bid (tBid) protein and of the pore-forming Bak and Bax proteins, and so the expectation was that GST–Bcl-2 would slow the trigger waves. Adding GST–Bcl-2 decreased the wave speed (Fig. 3, A and B, and movie S5). The maximum effect [at 400 nM added GST–Bcl-2, which compares to the estimated endogenous Bcl-2 concentration of approximately 140 nM (*14*)] was a reduction of the speed to about 13 μm/min, the speed seen in purified cytosol. Added GST–Bcl-2 had no effect on the trigger wave speed in cytosolic extracts (Fig. 3, C and D), which emphasizes that the waves seen in purified cytosol are probably not caused by contaminating mitochondria. GST–Bcl-2 decreased the trigger wave speed in reconstituted (cytosol plus mitochondria) extracts (Fig. 3, E and F), just as it did in cytoplasmic extracts.

If the MOMP–caspase–BH3 protein loop contributed to the generation of trigger waves, inhibition of the executioner caspases would be expected to slow or block the waves (Fig. 1A). To test this, we added the caspase-3 and -7 inhibitor *N*-acetyl–Asp-Glu-Val-Asp–aldehyde (Ac-DEVD-CHO) to the reconstituted extracts. Because high concentrations of the inhibitor make it difficult to monitor trigger waves with the fluorogenic caspase substrate Z-DEVD-R110, we used an additional probe, tetramethylrhodamine ethyl ester (TMRE), a red fluorescent dye that responds to changes in mitochondrial membrane potential.

Nonapoptotic mitochondria accumulate TMRE in their matrices. During apoptosis, the matrix loses TMRE at approximately the time of cytochrome *c* release (*15*, *16*), even in the presence of caspase inhibitors (*17*). We tested TMRE as a reporter of trigger waves by adding 50 nM TMRE to a reconstituted extract, together with Z-DEVD-R110, and looked for evidence of trigger waves in both fluorescence channels. A wave of TMRE loss (and therefore depolarization of mitochondria) could be detected (Fig. 3, G and H). The TMRE wave propagated at the same speed as the caspase trigger wave reported by Z-DEVD-R110 in the same tube (38 μm/min).

We tested whether trigger waves still occurred when caspases-3 and -7 were inhibited. When relatively low (≤ 1 μM) concentrations of the caspase inhibitor Ac-DEVD-CHO were used, there was sufficient residual caspase-3 and/or -7 activity to allow both the Z-DEVD-R110 and TMRE reporters to yield data on trigger wave speed. Both reporters showed that the wave speed decreased as the inhibitor concentration increased (Fig. 3, I to L; fig. S6; and movie S7). In the presence of Ac-DEVD-CHO, a transient increase in TMRE fluorescence occurred at the front of the apoptotic wave, which may result from fluorescence dequenching or transient hyperpolarization and which made the TMRE waves easier to discern (Fig. 3J and movies S6 and S7). At higher concentrations of Ac-DEVD-CHO, the wave speed could be read out only with the TMRE reporter, and it leveled off at ~15 μm/min (Fig. 3L).

These experiments implicate caspase-3 and/or -7 as well as Bcl-2 in the regulation of the apoptotic trigger waves. The experiments also show that the trigger waves are relatively robust; they are still present, though with reduced speeds, in extracts depleted of mitochondria or treated with maximal doses of GST–Bcl-2 or Ac-DEVD-CHO. However, the addition of Ac-DEVD-CHO to cytosolic extracts did completely eliminate trigger waves (fig. S7). Thus, the cytosolic caspase feedback loop (Fig. 1A) appears to mediate the (slow) apoptotic trigger waves seen in purified cytosolic extracts (Figs. 2E and 3C).

To determine whether apoptotic trigger waves occur in intact oocytes and eggs, we imaged oocytes incubated with TMRE or Z-DEVD-R110 and looked for evidence of a wave of fluorescence. However, the opacity of the cells made it difficult to obtain satisfactory imaging data. Nevertheless, we noticed that when oocytes were injected with cytochrome *c*, a wave of changes in the oocyte's natural pigmentation systematically spread from the site of injection to the opposite pole. One particularly notable example of such a wave is shown in montage form in Fig. 4A, as a kymograph image in Fig. 4B, and in video form in movie S8; two more-subtle examples are shown in movie S9. The surface waves propagated at an apparent speed of ~30 μm/min, similar to the wave speeds seen in apoptotic extracts. Control oocytes injected with Texas Red–dextran in water did not exhibit these surface waves. These findings indicate that apoptotic trigger waves can be produced in oocytes.

To test whether eggs might naturally exhibit a similar surface wave at the end of their lifetimes, we incubated eggs with egg-laying buffer at room temperature, conditions that typically lead to apoptosis by 12 to 24 hours. During this same time period, a surface wave often appeared, originating near the vegetal pole and propagating toward the animal pole, with a typical apparent speed of ~30 μm/min (Fig. 4, C and D, and movie S10). This wave was followed by further pigment changes over the next several hours—expansion and then contraction of a white dot at the animal pole (fig. S8).

To see whether the surface wave correlated with caspase activation, we harvested and lysed individual eggs just after they underwent such a wave. Every egg that showed a wave had increased caspase activity (Fig. 4, E and F, red symbols). We also collected eggs during the same time period that had not displayed a wave. None of these eggs had increased caspase activity (Fig. 4, E and F, blue symbols). These findings support the idea that spontaneous apoptosis typically initiates near the vegetal pole of the egg and propagates outward and upward from there as a ~30-μm/min trigger wave.

Our results indicate that apoptosis propagates through cytoplasmic *X. laevis* extracts via trigger waves (Fig. 1). The speed of the apoptotic trigger waves depends upon the concentration of mitochondria, although both cytosolic extracts and GST–Bcl-2–treated cytoplasmic extracts can support slower (~13 μm/min) waves (Figs. 2 and 3). Inhibiting caspase-3 and caspase-7 also slowed but did not abolish the waves. These findings show that the phenomenon is highly robust and must involve multiple interlinked bistable systems. We also found evidence for apoptotic waves in cytochrome *c*–injected oocytes and in spontaneously dying eggs, demonstrating that the trigger waves are not an artifact of the extract system.

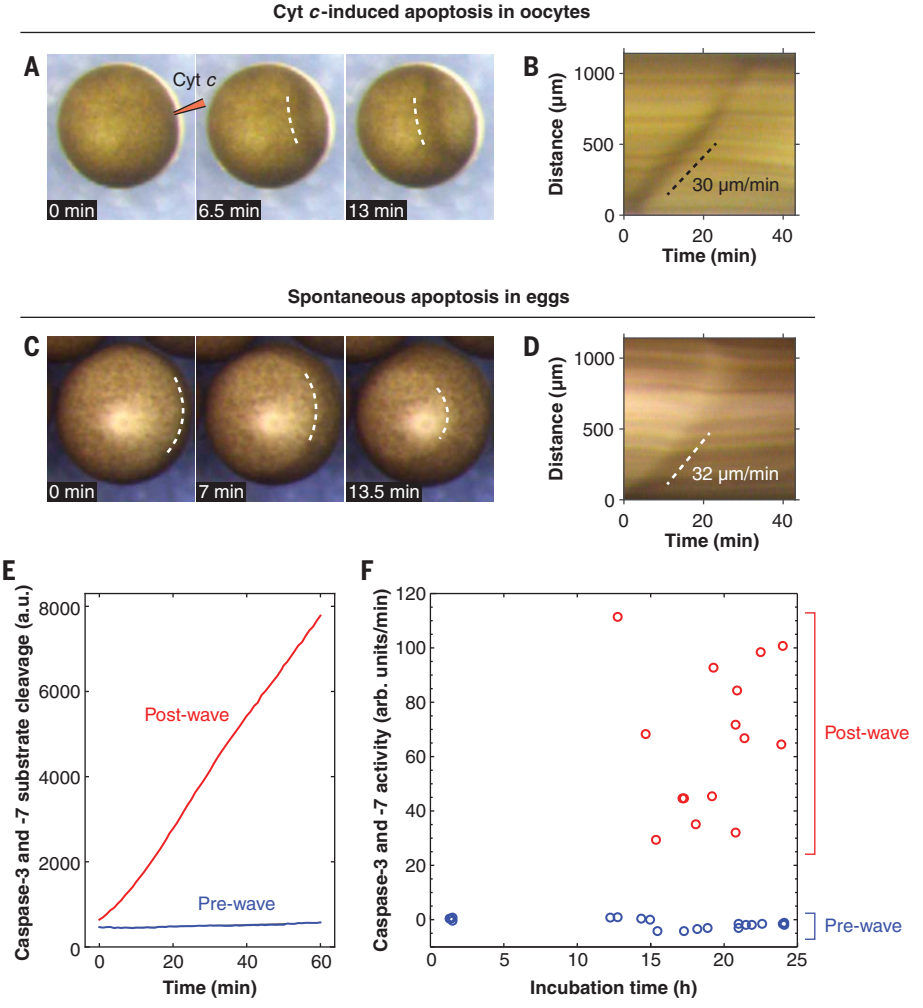


Fig. 4. Apoptotic trigger waves in intact oocytes and eggs. (**A** and **B**) Injection of immature stage VI oocytes with cytochrome *c* (10 nl of 1-mg/ml cytochrome *c*) causes a wave of pigmentation changes to spread from the injection site to the opposite side of the oocyte. (A) Surface wave example shown in montage form. Dashed lines highlight the progress of the wave. (B) Kymograph image. Two other examples of these waves are shown in movie S9. (**C** and **D**) Surface waves occur in spontaneously dying eggs. (C) Surface wave example shown in montage form. (D) Kymograph image. (**E**) Caspase-3 and/or caspase-7 assays for one prewave and one postwave egg. (**F**) Caspase-3 and/or -7 activities for eggs pre- and postwave. The data are from 19 prewave eggs and 15 postwave eggs.

Imaging studies on mammalian cell lines (*18*–*20*), cardiac myotubes (*21*, *22*), and syncytial human trophoblasts (*23*) have shown that apoptosis typically initiates at a single discrete focus or a small number of discrete foci and then spreads rapidly throughout the cell, and in some of these studies the propagation velocities appeared to be constant over distances of ~100 μm (*18*, *19*, *21*, *22*). Although it can be difficult to distinguish between trigger waves and diffusive spread over such short distances, particularly in cells with irregular geometries and inhomogeneous cytoplasm, it seems likely that apoptotic trigger waves occur in many cell types.

There is a close analogy between the mechanisms underpinning the apoptotic trigger waves observed in this study and calcium waves. Calcium waves arise from calcium-induced calcium release from the endoplasmic reticulum; apoptotic waves involve cytochrome *c*–induced release of cytochrome *c* from the mitochondria. The particular proteins, storage organelles, and time scales are different, but the basic logic is the same.

The ingredients needed for generating a trigger wave are simple; they include a spatial coupling mechanism (such as diffusion or intercellular communication) and positive feedback (*6*). Positive feedback is commonplace in signal transduction, from cell fate induction to biological oscillations to prion formation. And whenever positive feedback is present, there is a possibility of trigger waves, allowing signals to propagate quickly over large distances without diminishing in strength or speed. We suspect that many other examples of trigger waves exist in intra- and intercellular communication.

REFERENCES AND NOTES

1. S. Iguchi, T. Iwasaki, Y. Fukami, A. A. Tokmakov, *BMC Cell Biol.* **14**, 11 (2013).
2. D. D. Newmeyer, D. M. Farschon, J. C. Reed, *Cell* **79**, 353–364 (1994).
3. P. Deming, S. Kornbluth, *Methods Mol. Biol.* **322**, 379–393 (2006).
4. J. J. Tyson, J. P. Keener, *Physica D* **32**, 327–361 (1988).
5. A. T. Winfree, *Faraday Symp. Chem. Soc.* **9**, 38–46 (1974).
6. L. Gelens, G. A. Anderson, J. E. Ferrell Jr., *Mol. Biol. Cell* **25**, 3486–3493 (2014).
7. A. W. Murray, *Methods Cell Biol.* **36**, 581–605 (1991).
8. J. B. Chang, J. E. Ferrell Jr., *Nature* **500**, 603–607 (2013).
9. J. B. Chang, J. E. Ferrell Jr., *Cold Spring Harb. Protoc.* 10.1101/pdb.prot097212 (2018).
10. S. W. Tait, D. R. Green, *Nat. Rev. Mol. Cell Biol.* **11**, 621–632 (2010).
11. E. Ferrando-May, *Cell Death Differ.* **12**, 1263–1276 (2005).
12. J. Liu *et al.*, *Bioorg. Med. Chem. Lett.* **9**, 3231–3236 (1999).
13. H. Zou, Y. Li, X. Liu, X. Wang, *J. Biol. Chem.* **274**, 11549–11556 (1999).
14. M. Wühr *et al.*, *Curr. Biol.* **24**, 1467–1475 (2014).
15. J. C. Goldstein, N. J. Waterhouse, P. Juin, G. I. Evan, D. R. Green, *Nat. Cell Biol.* **2**, 156–162 (2000).
16. M. Rehm, H. Düssmann, J. H. Prehn, *J. Cell Biol.* **162**, 1031–1043 (2003).
17. N. J. Waterhouse *et al.*, *J. Cell Biol.* **153**, 319–328 (2001).
18. L. Lartigue *et al.*, *J. Cell Sci.* **121**, 3515–3523 (2008).
19. P. D. Bhola, A. L. Mattheyses, S. M. Simon, *Biophys. J.* **97**, 2222–2231 (2009).
20. M. Rehm *et al.*, *Cell Death Differ.* **16**, 613–623 (2009).
21. P. Pacher, G. Hajnóczky, *EMBO J.* **20**, 4107–4121 (2001).
22. C. Garcia-Perez *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 4497–4502 (2012).
23. M. S. Longtine, A. Barton, B. Chen, D. M. Nelson, *Placenta* **33**, 971–976 (2012).
24. E. H. Cheng *et al.*, *Science* **278**, 1966–1968 (1997).
25. E. A. Slee, S. A. Keogh, S. J. Martin, *Cell Death Differ.* **7**, 556–565 (2000).
26. J. G. Albeck, J. M. Burke, S. L. Spencer, D. A. Lauffenburger, P. K. Sorger, *PLOS Biol.* **6**, e299 (2008).
27. S. Legewie, N. Blüthgen, H. Herzel, *PLOS Comput. Biol.* **2**, e120 (2006).
28. T. Zhang, P. Brazhnik, J. J. Tyson, *Biophys. J.* **97**, 415–434 (2009).

ACKNOWLEDGMENTS

We thank the members of J. Chen's and D. Jarosz's labs for sharing their microscopes, H. Funabiki and M. Dasso for providing the GST-GFP-NLS construct, and J. Kamenz and the rest of the Ferrell lab for helpful discussions and comments on the manuscript. **Funding:** This work was supported by grants from the National Institutes of Health (R01 GM110564 and P50 GM107615). **Author contributions:** X.C. and J.E.F. jointly designed the studies, carried out the computations, made the figures, and wrote the paper. X.C. carried out the experiments. **Competing interests:** None declared. **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/361/6402/607/suppl/DC1
Materials and Methods
Figs. S1 to S9
Movies S1 to S10

21 June 2016; resubmitted 15 December 2017
Accepted 3 July 2018
10.1126/science.aah4065

2018 WORLD LIFE SCIENCE CONFERENCE

http://wlsc2018.csi.org.cn/

Oct. 27th-29th, 2018 Beijing, China

Organized by

China Association for Science and Technology
Ministry of Science and Technology of the People's Republic of China

THEME SCIENCE FOR BETTER LIFE

Conference Chairs

Zhu Chen
Honorary President of the Shanghai Institute of Hematology, Academician of the Chinese Academy of Sciences

David Baltimore
1975 Nobel laureate in Physiology or Medicine, California Institute of Technology

Organizing Committee Chairs

Xuetao Cao
President of China Union of Life Science Societies, Nankai University

Victor J. Dzau
President of the United States National Academy of Medicine, Harvard Medical School

Scientific & Program Committee Chairs

Zihe Rao
The First President of China Union of Life Science Societies, Tsinghua University

Vaida Vinson
Deputy Editor, Science

Confirmed Invited Plenary Speakers

David Baltimore
1975 Nobel laureate in Physiology or Medicine, California Institute of Technology

Aaron Ciechanover
2004 Nobel Laureate in Chemistry, Israel Institute of Technology

Jules A Hoffmann
2011 Nobel Laureate in Physiology or Medicine, University of Strasbourg

John E. Walker
1997 Nobel Laureate in Chemistry, University of Cambridge

Kurt Wüthrich
2002 Nobel Laureate in Chemistry, The Scripps Research Institute and ETH Zurich

Longping Yuan
2001 Wolf Prize in Agriculture, 2004 World Food Prize, Hunan Academy of Agricultural Sciences

Yunde Hou
2017 Highest Science and Technology Award, Vice President of China Academy of Engineering

Youyou Tu
2015 Nobel Laureate in Physiology of Medicine, China Academy of Traditional Chinese Medicine

Symposia

01 Biological Consequences of Global Change
02 Insect Chemical Communication
03 Virtual Human _ Imaging Across Scales
04 Insect Vectors for Animal and Human Diseases
05 Genome-guided application in insect studies
06 New advances in Agriculture
07 Plant genomes & evolution
08 Cell Fate Determination
09 Genome Editing
10 Interventional and Minimally Invasive Medicine
11 Molecules: Make Life more Precise
12 Nanobiotech
13 Medical Imaging
14 Nervous System Development
15 Multiscale Brainnetome Atlas
16 Emerging Infectious Diseases
17 Innate Immune Response and Inflammation
18 Drug-induced Damage and Clinical Rational Drug Use
19 Natural Products and Traditional Medicine
20 Wildlife Gut Microbiomes and Evolution
21 Adaptive evolution of animals
22 Chemical Ecology from Gene to Behavior
23 Prospect and application of Insect Microbiome
24 Plant and Environment
25 Plant Biotic Interactions
26 Structural Biology
27 Cell Death and Disease
28 Stem Cell and Regenerative Medicine
29 Diagnosis and Therapy of Rare Diseases
30 Frontier of Photosynthesis Research
31 Molecules: Make Life more Precise
32 International Standards and New Progress in Medical Robots
33 Chronic Obstructive Pulmonary Disease
34 Microbiome and Human Health
35 Global Health and Infection Diseases
36 Infection and Immunity
37 Infant Nutrition and Scientific Feeding
38 Research Progress in Agents Against Drug-resistant Infection
39 Chronic Infectious Disease
40 Visible Human and Digital Medicine
41 The Symposium on Cancer Frontier
42 Evolution and Ecology of Infectious diseases
43 Transplantation
44 Plant Metabolome and Human Health
45 Plant Genetics and Genomics
46 Cell Metabolism and Metabolic Diseases
47 Epigenetics
48 Genetic Diversity
49 Sustainable Agriculture
50 Cellular Imaging
51 Protein Research Frontier
52 Nanomedicine
53 Frontiers in Sleep Studies
54 Aging and Disease
55 Neurodegenerative Diseases: From Biology to Clinical Applications
56 New Vaccines Development and Application
57 Nuclear Receptors and Metabolic Disease
58 New Drug Discover
59 Immunotherapy and Biotherapy
60 Synthetic Biology

Special Forums Young Scientists Forum, Biomedical Industrial Zone Forum, Bioeconomy Forum, Biotechnology Development Forum, Scientific Publishing Forum

Young Scientist Travel Awards | 2018 WLSC offers travel awards to help young scientists offset the costs of attending the conference. Please find the details in the official website.

Hosted by China Union of Life Science Societies, China National Center for Biotechnology Development

CHINA ZOOLOGICAL SOCIETY · BOTANICAL SOCIETY OF CHINA · THE ENTOMOLOGICAL SOCIETY OF CHINA · CHINESE SOCIETY FOR MICROBIOLOGY
THE CHINESE SOCIETY OF BIOCHEMISTRY AND MOLECULAR BIOLOGY · CHINESE SOCIETY FOR CELL BIOLOGY · CHINESE SOCIETY FOR PLANT BIOLOGY
BIOPHYSICAL SOCIETY OF CHINA · GENETICS SOCIETY OF CHINA · CHINESE ASSOCIATION FOR LABORATORY ANIMAL SCIENCES · THE CHINESE NEUROSCIENCE SOCIETY
CHINESE SOCIETY OF BIOTECHNOLOGY · CHINESE ASSOCIATION OF INTEGRATIVE MEDICINE · CHINESE ASSOCIATION FOR PHYSIOLOGICAL SCIENCES
CHINESE SOCIETY FOR ANATOMICAL SCIENCES · CHINESE SOCIETY OF BIOMEDICAL ENGINEERING · CHINESE NUTRITION SOCIETY · CHINESE PHARMACOLOGICAL SOCIETY
CHINESE ANTI-CANCER ASSOCIATION · CHINESE SOCIETY FOR IMMUNOLOGY · CHINESE PREVENTIVE MEDICINE ASSOCIATION · CHINESE SOCIETY FOR COGNITIVE SCIENCE

SCIENCE TRANSCENDING BOUNDARIES

AAAS | ANNUAL MEETING
Washington, DC | Feb. 14–17, 2019

visit aaas.org/meetings

BE A PART OF THE CONVERSATION

Artificial Intelligence | Astrophysics | Genomics | Sustainability
Science Diplomacy | Food Security | Digital Privacy | Infections Control
Public Engagement | Climate Change | Diversity | Big Data

and more...

Advance registration now available online

Scientific program released in October

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Present your work to the world

Are you a representative of the upcoming generation of thought leaders in your field? Together we look forward to your application for the new Sartorius & *Science* Prize for Regenerative Medicine & Cell Therapy.

Apply now!

www.passionforscience.com/prize

The Sartorius & *Science* Prize for Regenerative Medicine & Cell Therapy

Awarded by sartorius Science

2019 MRS® SPRING MEETING & EXHIBIT

April 22–26, 2019 | Phoenix, Arizona

CALL FOR PAPERS

Abstract Submission Opens
September 28, 2018

Abstract Submission Closes
October 31, 2018

Spring Meeting registrations include MRS Membership July 1, 2019 – June 30, 2020

BROADER IMPACT

BI01 High Impact Practice—Increasing Ethnic and Gender Diversification in Engineering Education

CHARACTERIZATION, PROCESSING AND THEORY

CP01 Advances in *In Situ* Experimentation Techniques Enabling Novel and Extreme Materials/Nanocomposite Design
CP02 Design and *In Situ* TEM Characterization of Self-Assembling Colloidal Nanosystems
CP03 Advances in *In Situ* Techniques for Diagnostics and Synthetic Design of Energy Materials
CP04 Interfacial Science and Engineering—Mechanics, Thermodynamics, Kinetics and Chemistry
CP05 Materials Evolution in Dry Friction—Microstructural, Chemical and Environmental Effects
CP06 Smart Materials for Multifunctional Devices and Interfaces
CP07 From Mechanical Metamaterials to Programmable Materials
CP08 Additive Manufacturing of Metals
CP09 Mathematical Aspects of Materials Science—Modeling, Analysis and Computations

ELECTRONICS AND PHOTONICS

Soft Organic and Bimolecular Electronics

EP01 Liquid Crystalline Properties, Self-Assembly and Molecular Order in Organic Semiconductors
EP02 Photonic Materials and Devices for Biointerfaces
EP03 Materials Strategies and Device Fabrication for Biofriendly Electronics
EP04 Soft and Stretchable Electronics—From Fundamentals to Applications
EP05 Engineered Functional Multicellular Circuits, Devices and Systems
EP06 Organic Electronics—Materials and Devices

Semiconductor Devices, Interconnects, Plasmonic and Thermoelectric Materials

EP07 Next-Generation Interconnects—Materials, Processes and Integration
EP08 Phase-Change Materials for Memories, Photonics, Neuromorphic and Emerging Application
EP09 Devices and Materials to Extend the CMOS Roadmap for Logic and Memory Applications
EP10 Heterovalent Integration of Semiconductors and Applications to Optical Devices
EP11 Hybrid Materials and Devices for Enhanced Light-Matter Interactions
EP12 Emerging Materials for Plasmonics, Metamaterials and Metasurfaces
EP13 Thermoelectrics—Materials, Methods and Devices

ENERGY AND SUSTAINABILITY

Energy Storage

ES01 Organic Materials in Electrochemical Energy Storage
ES02 Next-Generation Intercalation Batteries
ES03 Electrochemical Energy Materials Under Extreme Conditions
ES04 Solid-State Electrochemical Energy Storage

Catalysis, Alternative Energy and Fuels

ES05 Cooperative Catalysis for Energy and Environmental Applications
ES06 Atomic-Level Understanding of Materials in Fuel Cells and Electrolyzers
ES07 New Carbon for Energy—Materials, Chemistry and Applications
ES08 Materials Challenges in Surfaces and Coatings for Solar Thermal Technologies
ES10 Rational Designed Hierarchical Nanostructures for Photocatalytic System
ES11 Advanced Low Temperature Water-Splitting for Renewable Hydrogen Production via Electrochemical and Photoelectrochemical Processes
ES12 Redox-Active Oxides for Creating Renewable and Sustainable Energy Carriers

Water-Energy Materials and Sustainability

ES09 Advanced Materials for the Water-Energy Nexus
ES13 Materials Selection and Design—A Tool to Enable Sustainable Materials Development and a Reduced Materials Footprint
ES14 Materials Circular Economy for Urban Sustainability

Photovoltaics and Energy Harvesting

ES15 Fundamental Understanding of the Multifaceted Optoelectronic Properties of Halide Perovskites
ES16 Perovskite Photovoltaics and Optoelectronics
ES17 Perovskite-Based Light-Emission and Frontier Phenomena—Single Crystals, Thin Films and Nanocrystals
ES18 Frontiers in Organic Photovoltaics
ES19 Excitonic Materials and Quantum Dots for Energy Conversion
ES20 Thin-Film Chalcogenide Semiconductor Photovoltaics
ES21 Nanogenerators and Piezotronics

QUANTUM AND NANOMATERIALS

QN01 2D Layered Materials Beyond Graphene—Theory, Discovery and Design
QN02 Defects, Electronic and Magnetic Properties in Advanced 2D Materials Beyond Graphene
QN03 2D Materials—Tunable Physical Properties, Heterostructures and Device Applications
QN04 Nanoscale Heat Transport—Fundamentals
QN05 Emerging Thermal Materials—From Nanoscale to Multiscale Thermal Transport, Energy Conversion, Storage and Thermal Management
QN06 Emerging Materials for Quantum Information
QN07 Emergent Phenomena in Oxide Quantum Materials
QN08 Colloidal Nanoparticles—From Synthesis to Applications

SOFT MATERIALS AND BIOMATERIALS

SM01 Materials for Biological and Medical Applications
SM02 Progress in Supramolecular Nanotheranostics
SM03 Growing Next-Generation Materials with Synthetic Biology
SM04 Translational Materials in Medicine—Prosthetics, Sensors and Smart Scaffolds
SM05 Supramolecular Biomaterials for Regenerative Medicine and Drug Delivery
SM06 Nano- and Microgels
SM07 Bioinspired Materials—From Basic Discovery to Biomimicry

www.mrs.org/spring2019

Meeting Chairs

Yuping Bao The University of Alabama
Bruce Dunn University of California, Los Angeles
Subodh Mhaisalkar Nanyang Technological University
Ruth Schwaiger Karlsruhe Institute of Technology—Institute for Applied Materials
Subhash L. Shinde University of Notre Dame

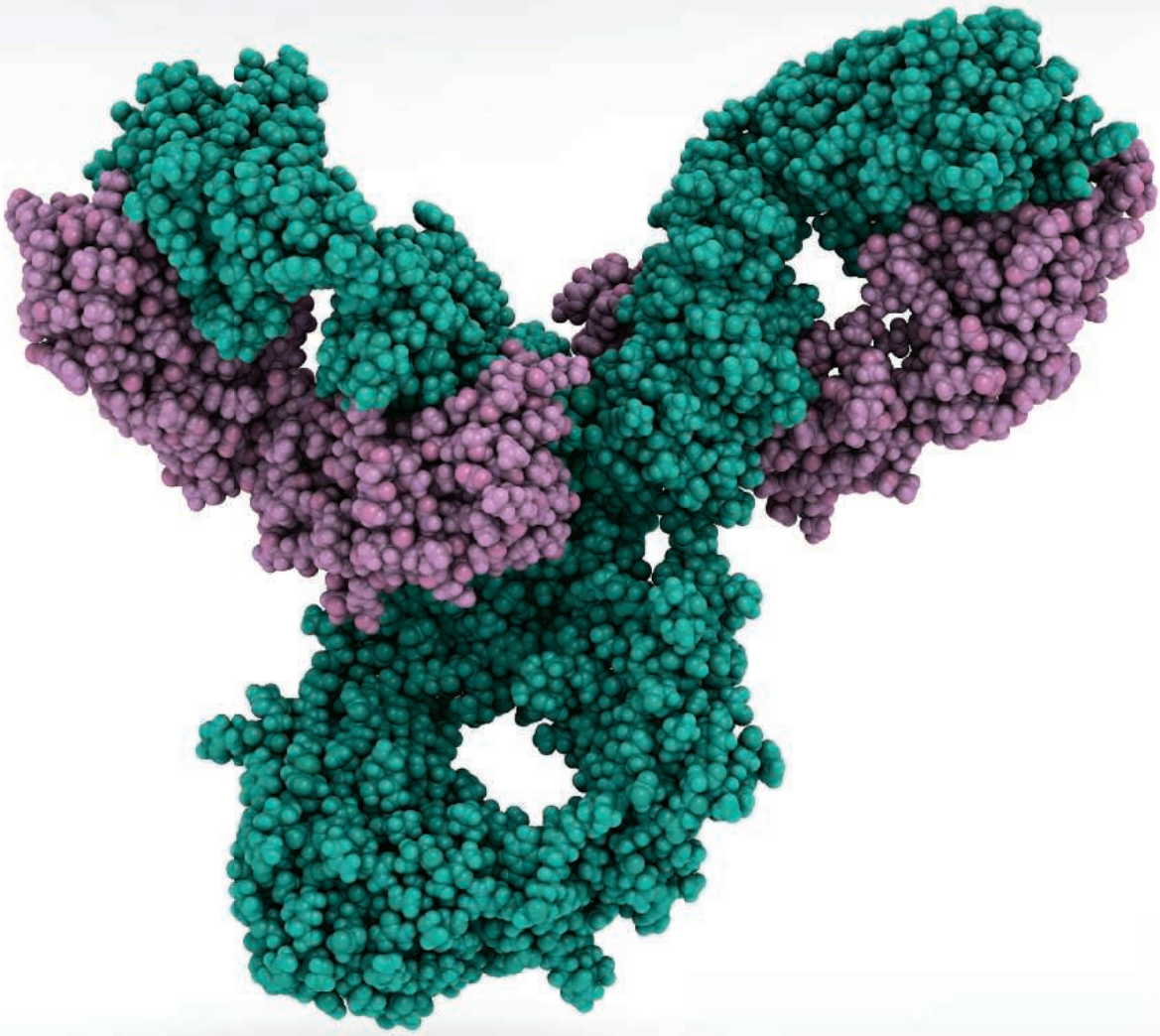
Don't Miss These Future MRS Meetings!

2019 MRS Fall Meeting & Exhibit
December 1–6, 2019, Boston, Massachusetts

2020 MRS Spring Meeting & Exhibit
April 13–17, 2020, Phoenix, Arizona

MRS MATERIALS RESEARCH SOCIETY®
Advancing materials. Improving the quality of life.
506 Keystone Drive • Warrendale, PA 15086-7573
Tel 724.779.3003 • Fax 724.779.8313 • info@mrs.org • www.mrs.org

Publish your research in *Science Immunology*

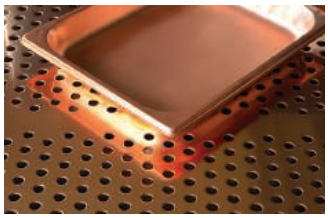


Science Immunology publishes original, peer-reviewed, science-based research articles that report critical advances in all areas of immunological research, including important new tools and techniques.

For more information: **ScienceImmunology.org**

Science Immunology

AAAS



Copper Incubator Accessories

An accessory set for the BINDER Series C CO_2 incubators includes three copper shelves and a copper water pan. As the pan always contains water, and the shelves directly contact incubating cell culture vessels, they can pose contamination risks. Combined with BINDER's 180°C hot-air sterilization cycle and humidity management, these accessories provide an added level of contamination control. Copper ions have a demonstrable biocidal effect and resist the growth of most contaminants, eliminating a potential source of cross-contamination should any culture medium spill on the shelves. Using copper for the water pan prevents contaminating growth in the humidification water. The copper equipment reduces the time and effort required for cleaning the incubator while further ensuring a high level of contamination resistance.

BINDER
For info: +49-(0)-7462-2005-0
www.binder-world.com/us

Neuronal Culture System

The Gibco B-27 Plus Neuronal Culture System, consisting of B-27 Plus Supplement and Neurobasal Plus Medium, allows neurobiology researchers to replace current neuronal cell culture supplements and basal media, so they can maximize survival of neurons in short- and long-term cultures without having to change workflows. For researchers who are either maintaining primary rodent and human pluripotent stem cell (hPSC)-derived neurons or differentiating hPSC-derived neural stem cells to neurons, the combination of Gibco B-27 Plus Supplement, Neurobasal Plus Medium, and CultureOne Supplement removes the challenges of culturing neurons and enables an improved downstream experience.

Thermo Fisher Scientific
For info: 866-356-0354
www.thermofisher.com/b27plus

Plate Washer

The DA-Cell system from Curiox BioSystems features a unique 96-well plate and washer that uses gravity and laminar flow to wash your cells in 2–4 min. The system is fully automated and operator independent, helping you attain more consistent results. It lets you achieve more than 95% cell retention, even after multiple washes of permeabilized cells. The DA-Cell technology enables you to work with a 96-droplet–based format instead of a conventional microwell plate. It also helps you to prepare samples for multicolor flow cytometry and bypasses the need for centrifugation and washing of suspension cells.

Curiox BioSystems
For info: 650-226-8420
www.curiox.com

Virus Production Media

Serum and di- and tripeptide plant-derived hydrolysates can cause variability in cell expansion, hindering viral replication, and ultimately reducing the productivity of the cell culture system. Therefore, media with undefined components can increase the time and resources required to manufacture a vaccine. OptiVERO is a chemically defined, blood- and plant hydrolysate–free cell culture media for vaccine manufacturing that outperforms popular VERO media. Unrefined components are replaced with pure and biologically active recombinant human transferrin and albumin to formulate a chemically defined media optimized specifically for VERO-based virus production in both 2D and 3D cell cultures.

InVitria
For info: 800-916-8311
invitria.com

Automated Cell Counter

During manual cell counting, the subjective evaluation of the definition of a cell introduces bias to the result. The automatic cell counter NucleoCounter NC-200 is designed to limit human interference in counting. With the Via1-Cassette, all errors introduced during pipetting and staining are avoided. The cassette includes an in-built pipette, and the immobilized fluorescent dyes acridine orange and DAPI automatically stain the total and dead cell populations, respectively. The Via1-Cassette is volume-calibrated, ensuring high precision in determining cell count and viability. This patented one-step cell count and viability procedure takes less than 50 s.

ChemoMetec
For info: +45-(0)-48131020
chemometec.com/cell-counters/cell-counter-nc-200-nucleocounter

Recombinant Wnt3a for Organoid Culture

The recombinant Wnt3a from AMS Biotechnology has a Wnt stabilizer that significantly extends the activity of this protein in serum-free medium from a half-life of just 2 h to around 30 h. Wnt3a is a protein encoded by the *Wnt3a* gene. The *Wnt* gene family consists of structurally related genes that encode secreted signaling proteins. These proteins have been implicated in oncogenesis, adipogenesis, and in several other developmental processes including regulation of cell fate and patterning during embryogenesis. Wnt3a has been a key reagent in the organoid culture techniques developed by Hans Clevers and coworkers, based on the LGR5 stem cell marker/Wnt signaling system. In serum-free culture conditions, Wnt proteins are very unstable, losing their activity completely in 16 h. In the presence of Wnt protein stabilizer, purified recombinant Wnt3a can support even colon organoid cultures that need strong Wnt activity.

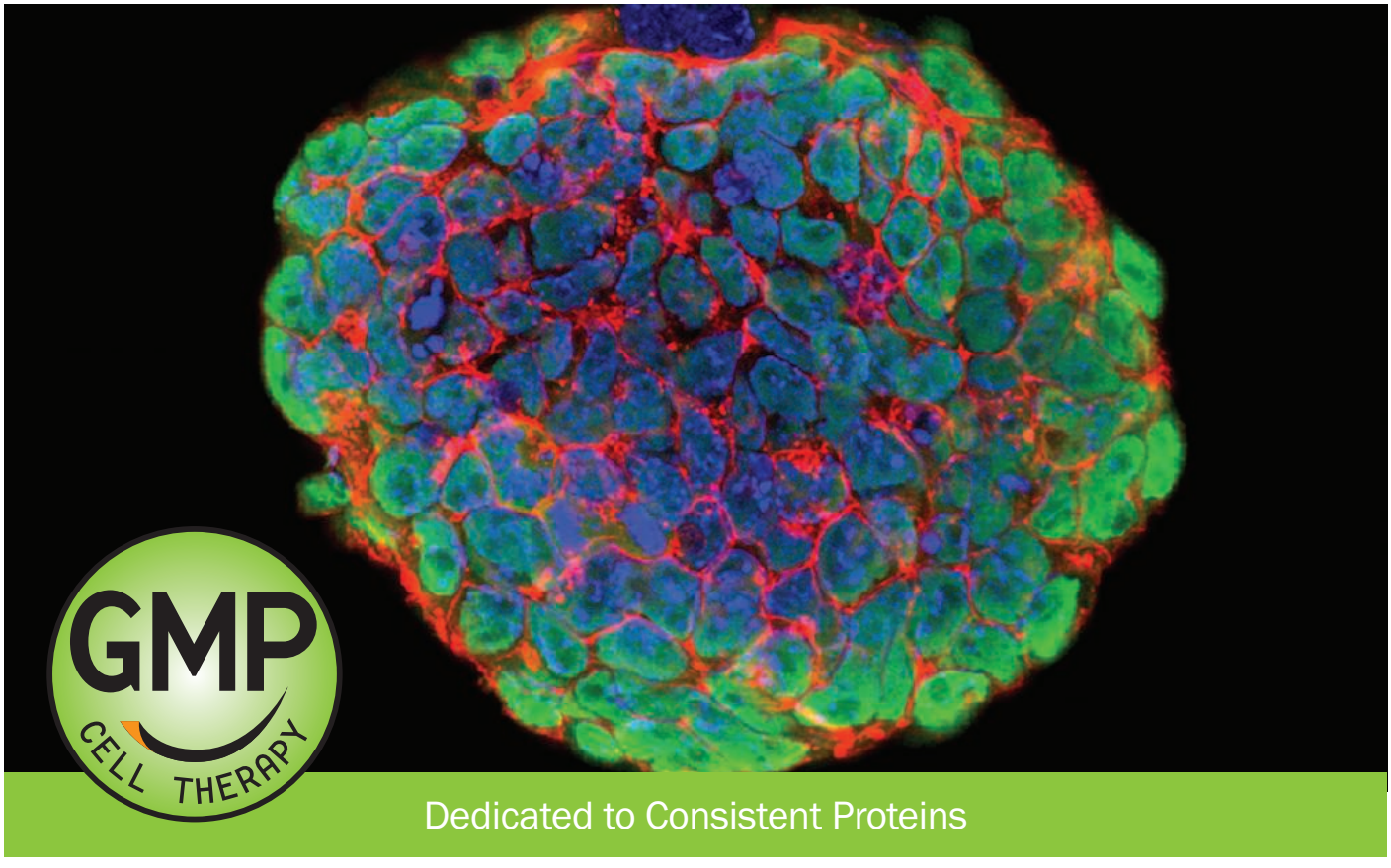
AMS Biotechnology
For info: +44-(0)-1235-828200
www.amsbio.com/wnt-signaling-cancer-proteins.aspx

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.

R&D SYSTEMS
a bio-techne brand

GMP-Grade Growth Factors



Dedicated to Consistent Proteins

The Widest Selection for

- Cell Manufacturing
- Immunotherapy
- Regenerative Medicine
- Stem Cell Therapy

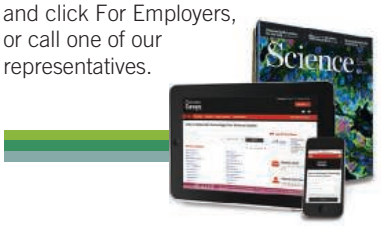
Learn more | rndsystems.com/gmp

bio-techne®

Global bio-techne.com info@bio-techne.com TEL +1 612 379 2956 **North America** TEL 800 343 7475
Europe | Middle East | Africa TEL +44 (0)1235 529449 **China** info.cn@bio-techne.com TEL +86 (21) 52380373

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

ScienceCcareers.org

INDIANA UNIVERSITY
BLOOMINGTON

Tenure-Track Faculty Position in Functional Genomics Genome, Cell, and Developmental Biology (GCDB) Program Department of Biology, College of Arts and Sciences https://biology.indiana.edu/index.html

The GCDB Program within the Biology Department at Indiana University invites applications for a tenure-track position at the **assistant professor** level with a research emphasis in functional genomics. This position is part of the Indiana University Precision Health Initiative Grand Challenge https://grandchallenges.iu.edu/precision-health/index.html. We are interested in candidates whose research uses high-throughput genomic methods to elucidate the molecular mechanisms underlying either cancer or neurodegenerative disorders. We are seeking individuals that will complement and/or extend our existing strengths in chromatin structure, DNA replication and repair, cell cycle regulation, cancer biology, and neural development. Candidates working on model organisms as well as human cell lines and/or primary samples are encouraged to apply.

Successful candidates will have full access to state-of-the-art facilities for genomics and bioinformatics, light and electron microscopy, flow cytometry, structural biology, protein analysis, and molecular biology. IU researchers have many opportunities for collaboration, including the IU Simon Cancer Center (http://cancer.iu.edu) and the IU Program in Neuroscience (https://neuroscience.indiana.edu/index.html). Our faculty members are expected to establish and maintain a vigorous well-funded research program and to participate in both undergraduate and graduate education. More information on our program can be found at https://biology.indiana.edu/graduate/genome-cell-development/index.html

Bloomington is a family-friendly city with urban amenities, including theatre productions, live music, art exhibits, and a wide-range of culinary options, while still being located in the hilly, forested landscape of south-central Indiana (http://www.visitbloomington.com/).

Applicants should hold a Ph.D. or equivalent degree and have relevant postdoctoral experience with a strong record of research accomplishments. Applicants should submit a cover letter, CV, research statement, and a list of three references with contact information using the submission link at **http://indiana.peopleadmin.com/postings/6285**. Please address inquires to Jennifer Tarter, **jenjones@indiana.edu**; 812-856-3984; 1001 E. Third Street, Bloomington, IN 47405-3700. Applications received by **October 15, 2018** will be assured of full consideration.

The College of Arts and Sciences is committed to building and supporting a diverse, inclusive, and equitable community of students and scholars.

Indiana University is an Equal Employment and Affirmative Action Employer and a provider of ADA services. All qualified applicants will receive consideration for employment without regard to age, ethnicity, color, race, religion, sex, sexual orientation or identity, national origin, disability status or protected veteran status.

POSITIONS OPEN

Yale

YALE UNIVERSITY SCHOOL OF MEDICINE POSTDOCTORAL ASSOCIATE
Infectious Disease Pathogenesis/Immunology

Positions available to study the interactions between ticks, pathogens and the vertebrate host. The goal is to develop new strategies to prevent diverse tick-borne infections, such as Lyme disease, anaplasmosis, babesiosis and Powassan virus. An M.D. or Ph.D. in microbial pathogenesis, immunobiology, entomology, cell biology or molecular biology is necessary.

Send curriculum vitae and recent publications to: **Erol Fikrig M.D., Investigator, Howard Hughes Medical Institute, Yale University School of Medicine, Section of Infectious Diseases, P.O. Box 208022, New Haven, CT 06520-8022** or email: lynn.gambardella@yale.edu. *Yale University is an Affirmative Action, Equal Opportunity Employer. Applications from women and minorities are encouraged.*

Tenure-Track Professor in Marine Biology

The Department of Organismic and Evolutionary Biology (OEB), in partnership with the Museum of Comparative Zoology (MCZ), seeks to appoint a tenure-track professor in Marine Biology. Research areas of interest include, but are not limited to, human impacts on marine life and ecosystems, conservation biology, marine ecology, systematics, physiology and evolutionary biology, with an emphasis on animals (invertebrates or vertebrates). We seek an outstanding individual who will establish an innovative research program and teach both undergraduate and graduate students. The position will likely be associated with a curatorial appointment in the MCZ with shared oversight responsibilities for the museum's research collections. The department and museum have strong linkages to a number of allied institutions, including the Harvard Forest, Arnold Arboretum, Harvard University Herbaria, and Harvard Center for the Environment. Please submit applications online at **http://academicpositions.harvard.edu/postings/8358**. Required materials include a curriculum vitae; a statement of research and teaching interests; four representative publications; and the names, institutional affiliations, and e-mail addresses of three to five references. This appointment will be made initially at the untenured rank (Assistant or Associate Professor). Review of applications and nominations will begin **September 15, 2018** and conclude when the position is filled.

Basic Qualifications: Doctorate in Marine Biology or related discipline is required by the time the appointment begins.

Additional Qualifications: Demonstrated strong commitment to teaching is desired.

Special Instructions: Please submit the following materials through the ARIeS portal (http://academicpositions.harvard.edu/postings/8358). Candidates are encouraged to apply by September 15, 2018; applications will be reviewed until the position is filled.

1. Cover letter
2. *Curriculum Vitae*
3. Teaching statement (describing teaching approach and philosophy)
4. Research statement
5. Names and contact information of 3-5 referees, who will be asked by a system-generated email to upload a letter of recommendation once the candidate's application has been submitted. Three letters of recommendation are required, and the application is considered complete only when at least three letters have been received.
6. Publications or copies of creative works, if applicable

Further information about OEB and MCZ is available at http://www.oeb.harvard.edu/ and https://mcz.harvard.edu/. Address questions about the position to fac_search@oeb.harvard.edu.

We are 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.

The University of Texas at Austin
College of Pharmacy

Assistant or Associate Professor
Division of Pharmacology and Toxicology

The Division of Pharmacology and Toxicology at The University of Texas at Austin College of Pharmacy invites applications for a full-time tenure-track or tenured faculty position, with appointment as Assistant or Associate Professor, with an anticipated start date of August 2019. Division faculty members (http://sites.utexas.edu/pharmtox/faculty/) are a highly interdisciplinary and collaborative group with cutting-edge basic and translational research programs. The University of Texas at Austin is home to the Waggoner Center on Alcohol and Addiction Research which provides strong support for collaborative research. The successful applicant will be expected to engage in productive research, high-quality teaching, and actively participate in college and university service. The successful applicant will be expected to establish and maintain a vigorous and externally funded research program.

We are seeking applicants with research interests in the pharmacology and neurobiology of addiction, including alcohol use disorders. Candidates with expertise in genomics, electrophysiology, or signal transduction are particularly encouraged to apply. Teaching in the professional (Pharm.D.) program, and Ph.D. graduate program are required. Qualified candidates must have a Ph.D., M.D., or equivalent degree in a relevant field. Only candidates with a sustained record of extramural funding and outstanding scholarship will be considered for appointment as Associate Professor.

Applications received before **October 1, 2018** will receive first consideration, but applications will be accepted until the position is filled. The position carries a highly competitive salary, benefits, and start-up package, commensurate with experience.

This institution is using Interfolio's ByCommittee to conduct this search. Applicants may access this position at **http://apply.interfolio.com/52701** and will be prompted to submit a cover letter, curriculum vitae, research plan, and a list of 3 references.

The University of Texas at Austin is an Equal Opportunity Employer with a commitment to diversity at all levels. All qualified applicants will receive consideration for employment without regard to race, color, religion, gender, national origin, age, disability, or veteran status. (Compliant with the new VEVRAA and Section 503 Rules). Security sensitive; conviction verification conducted on applicant selected. If hired, you will be required to complete the federal Employment Eligibility Verification form, I-9. You will be required to present acceptable and original documents to prove your identity and authorization to work in the United States. UT Austin is a tobacco-free campus.

WAYNE STATE UNIVERSITY

Faculty Positions in Integrative Biosciences

IBio

Wayne State University (WSU) is recruiting up to15 faculty (open rank) for research and development programs as part of the continued expansion of a broad institutional initiative in Integrative Biosciences. This initiative leverages a new 200,000 sq. ft. Integrative Biosciences Center (IBio) that houses coordinated inter- and trans-disciplinary research teams, and a Clinical Research Center. Programmatic themes include a focus on pathophysiologies and accumulated stressors affecting health in evolving urban environments with a strong emphasis on basic disease mechanisms, clinical translation and community health impact.

Ten faculty have been recruited to date as part of the IBio initiative and we are now entering the next phase of thematic-based, programmatic growth. Faculty recruitment (tenured, tenure-track, or research-track) will focus on six primary thematic areas: Behavioral Health (#043728); Bio & Systems Engineering (#043729); Environmental Sciences and Health (#043725); Health Disparities (#043726); Metabolic Diseases (Cardiovascular, Diabetes and Obesity) (#043727); and Translational Neurosciences (#043730). Each theme includes basic discovery-driven research as well as translational, community and implementation sciences cutting across departments, programs, centers, and colleges.

Faculty recruits (tenured or tenure-track) will integrate with departments and colleges or schools consistent with their areas of expertise and shared interests and engage in all aspects of our academic mission including research, education and service. Faculty are expected to either already have established extramural research funding and/or are on a clear path to secure and sustain extramural funding in support of their research programs.

Candidates must have a Ph.D., M.D., Pharm.D. and/or related degree(s) in disciplines aligning with the focus areas and possess a demonstrated track record of exceptional science, creative discovery and/or knowledge translation and application. We would be pleased to receive applications from groups of faculty from one or several institutions who may wish to work together. Qualified candidates should submit applications to the specific thematic position posting # identified above through the Wayne State University Online Hiring System **https://jobs.wayne.edu/applicants/jsp/shared/Welcome_css.jsp**. Applications should include a *curriculum vitae* and a brief narrative describing their research and how it relates to the Integrated Biosciences initiative (http://www.IBio.wayne.edu) with a cover letter addressed to the IBio Steering Committee Chair, Stephen M. Lanier, Ph.D. Vice President for Research. Review of applications for the next phase of recruitments will begin immediately with applications accepted through October 30, 2018. Competitive recruitment packages are available with salary and rank based on qualifications.

Wayne State University, which holds the highest Carnegie Foundation for the Advancement of Teaching designations in both research and community engagement, is a premier, public, urban, comprehensive research university located in the heart of Detroit where students from all backgrounds are offered a rich, high quality education. Our deep-rooted commitment to excellence, collaboration, integrity, diversity and inclusion creates exceptional opportunities for students and faculty in a diverse, global society. WSU encourages applications from women, people of color, and other underrepresented people. Wayne State is an Affirmative Action/Equal Opportunity Employer.

Founded in 1868, Wayne State University offers more than 370 academic programs through 13 schools and colleges to nearly 28,000 students. The campus in Midtown Detroit comprises 100 buildings over 200 acres including the School of Medicine, the Eugene Applebaum College of Pharmacy and Health Sciences and the College of Nursing. The university is home to the Perinatology Research Branch of the National Institutes of Health, the Karmanos Cancer Center, a National Cancer Institute-designated comprehensive cancer center, and a National Institute of Environmental Health Sciences Core Center - *Center for Urban Responses to Environmental Stressors (CURES)*.

Career Feature:

Postdoc Careers

Issue date: August 31
Book ad by August 16

Ads accepted until August 24 if space allows

To book your ad:
advertise@sciencecareers.org

The Americas
+ 202 326 6577
Europe
+44 (0) 1223 326527
Japan
+81 3 6459 4174
China/Korea/Singapore/ Taiwan
+86 131 4114 0012

Produced by the *Science*/AAAS Custom Publishing Office.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

Why choose this Postdoc Feature for your advertisement?

- Relevant ads lead off the career section with a special "Postdoc" banner
- Link on the job board homepage directly to postdoc positions
- Dedicated landing page for postdoc positions.

THERE'S ONLY ONE *SCIENCE*.

EGL
Charitable Foundation

The EGL Charitable Foundation invites you to apply to the

Gruss Lipper Post-Doctoral Fellowship Program

Eligibility

- Israeli citizenship
- Candidates must have completed PhD and/or MD/PhD degrees in the Biomedical Sciences at an accredited Israeli University/Medical School or be in their final year of study
- Candidates must have been awarded a postdoctoral position in the U.S. host research institution.

Details regarding the fellowship are available at www.eglcf.org

Application Deadline is October 12, 2018

WORKING LIFE

By **Alona Fyshe**

How to start a research lab

As September approaches, a new cohort of junior faculty members are taking up their first positions as research group leaders. I was there 3 years ago, making career-shaping decisions—sometimes without much mentoring or support. I learned a lot in my first years—how to write a grant, manage rejection, and supervise students, to name just a few—and it was all trial by fire. Though I made it through and had some successes along the way, I certainly could have used advice about how to set up and run my lab. I've learned that my experience is the norm, which inspired me and a group of other early-career principal investigators to interview leaders in our fields about how they built successful research groups. Here are some of the lessons they shared.

FIND YOUR NICHE. Before you even begin to interview for a faculty job, you need to decide what your lab's focus will be. You must be a pioneer, "carving out a niche for yourself that is unique, and where you'll be at the top of the heap," says Margaret McFall-Ngai, a professor at the University of Hawaii at Manoa who was among the first to study squid-bacteria interactions. Identify how your skills intersect with the science that excites you, in the most promising uncharted territory. Plan big while identifying key publication checkpoints along the way. Your tenure case depends on it.

"I certainly could have used advice about how to set up and run my lab."

FIND YOUR PEOPLE. Good science is done by talented people. "If you have an excellent person who wants to work with you, try to hire them at all costs, even if you have to spend the last of your money," says Gregor Weihs, a professor of photonics at the University of Innsbruck in Austria. On the flip side, hiring the wrong people can be a real drain on the group. "Never hire just because you can," Weihs says. Get to know prospective lab members by teaching graduate classes and taking on undergraduates for smaller projects, and use your network to find promising graduate students at other institutions.

IT PAYS TO WORK TOGETHER. To secure major funding, "it's better to try and see if you can chase it together rather than all competing for the same buck," says Melvyn Goodale, a professor of neuroscience at Western University in London, Canada, who helped form an 11-institution research consortium, the Canadian Action and Perception Network. If your goals are aligned with those of other labs, then it makes sense to work together on a joint application rather than against each other. Writing grant applications as a group can help spark new ideas, and many minds working together can increase your chance of success. Even if you don't get funded, writing a group grant can deepen your collaborative relationships for years to come.

BUILD A NETWORK. To make connections when you're just starting out, you need to be your own marketing department. "It's not just doing the research; it's making it known," says Yoshua Bengio, a professor at the University of Montreal in Canada who works on artificial intelligence. These days, a lot of networking is done online. Get on Twitter and follow your 10 favorite research labs. Look at who they follow to find more connections. Tweet about the work you publish and interesting papers you read to help people identify your niche and get to know your research brand.

Finally, don't be afraid to reach out to senior faculty members to seek out mentoring, share your work, and ask for input on grant applications. Their feedback will be invaluable, and some day you will pay it forward to a new cohort of junior faculty members. There is knowledge and experience all around. You may be surprised at people's willingness to share it. ■

Alona Fyshe is an assistant professor in the Department of Computing Science and Department of Psychology at the University of Alberta in Edmonton, Canada, and a Canadian Institute for Advanced Research (CIFAR) Azrieli Global Scholar. *She thanks CIFAR Azrieli Global Scholars Gerhard Kirchmair, Graham Taylor, and Katherine Amato, and the senior faculty members who were interviewed. Full interview transcripts are available at www.cifar.ca/labs.*

ILLUSTRATION: ROBERT NEUBECKER