Can "any idiot" get CRISPR to work? *p. 541*

Mutational signatures of tobacco smoke *pp. 549 & 618*

Single-cell technique reveals diversity in brain genome *p. 557*

# Science

$15
4 NOVEMBER 2016
sciencemag.org

AAAS

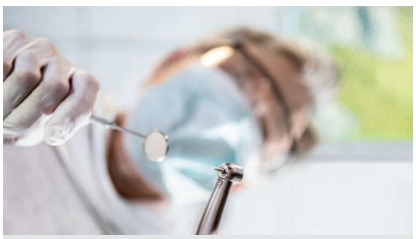
SPECIAL ISSUE

# PAIN

# CONTENTS

547
Birds in trouble

4 NOVEMBER 2016 • VOLUME 354 • ISSUE 6312



SPECIAL SECTION

## Pain research



ON THE COVER


A woman holds her hand in ice water, one of several techniques used to measure pain tolerance. Typically, healthy people can submerge their hands for up to 5 minutes. However, for some of those taking opioids, 10 to 15 seconds is the limit, suggesting opioid-induced hypersensitivity to pain. For more on recent developments in pain research, see the special section beginning on page 564. *Photo: Matthew Rakola*

# NEWS




# INSIGHTS

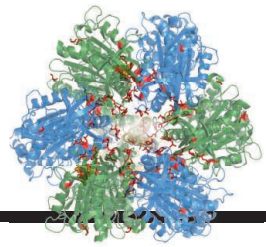






Published by AAAS

# CONTENTS



599
Protein mutations that rescue defects

4 NOVEMBER 2016 • VOLUME 354 • ISSUE 6312



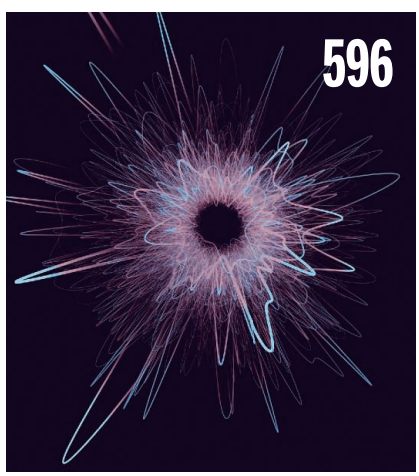







PHOTO: SINATRA *ET AL.*

**SCIENCE (ISSN 0036-8075) is published weekly on Friday, except the 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 and subscription (51 issues): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1522; 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 #1254 88122. 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:** $15.00 current issue, $20.00 back issue prepaid includes surface postage; bulk rates on request. 

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 Services** 202-326-6417 or http://membercentral.aaas.org/discounts

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

**INFORMATION FOR AUTHORS** See pages 624 and 625 of the 5 February 2016 issue or access www.sciencemag.org/authors/science-information-authors

EDITORIAL

# Forecasting the opioid epidemic

Since 2000, almost half a million Americans have died from drug overdoses. This modern plague—largely driven by opioid addiction—degrades health, saps productivity, spawns crime, and devastates families, all at enormous societal cost. How did we get here, and what do we do now?

About 20 years ago, compassionate advocacy for better treatment of chronic pain, combined with aggressive marketing of high-dose opioid formulations, led to a sharp increase in the prescribing of legal opioids by physicians in the United States. An unintended consequence of this well-meaning movement was that millions of Americans became dependent on opioids. Drug cartels seized the opportunity to sell heroin as a cheaper alternative to this ready-made consumer base. As users switched from high-quality pharmaceuticals to street drugs of unreliable composition and quality, deaths mounted. Mortality rates surged further as potent illicit synthetic drugs—such as fentanyl—were mixed in with heroin.

***"...a new, public health–oriented approach is needed."***

Law enforcement officials have realized that "we can't arrest our way out of the epidemic" and that a new, public health–oriented approach is needed. How can a new approach be designed? Better understanding of the underlying system dynamics may prove critical. Computational models are routinely used by engineers to understand the dynamics of physical systems, and models and simulations are increasingly being used in public health to understand the dynamics of contagious disease epidemics in human populations. Today, every major public health agency has access to a modeling unit that provides policy decision support for epidemics such as HIV/AIDS, influenza, Ebola, and Zika.

The opioid epidemic can be approached as a dynamical system, composed of networks of interacting individuals including nonusers, users, legal prescribers, illicit suppliers, treatment providers, supporters (family and friends), law enforcement officials, and others. Data on the frequency of opioid drug prescribing, illicit drug availability, the natural history of substance use disorder, treatment effectiveness, availability of support services, and other key factors can then be used to construct and parameterize models. Borrowing from methods used to model contagious diseases, these dynamical models can be employed to forecast the likely future trajectory of the opioid epidemic in the population (not a single path, but a range of possible paths as bounded by uncertainty) and then to evaluate—in silico—interventions designed to alter the trajectory. Policy-makers could estimate the likely effectiveness—and cost-effectiveness—of possible public health interventions such as increasing the availability of naloxone (medication used to block the effects of opioids), decreasing prescriptions, expanding treatment, and widening education, as well as concurrent deployment of two or more of these interventions. Computational modeling can also be used to explore synergies between public health and law enforcement.

Increased data openness will be needed. Each country will have its own data access issues. In the United States, overdose deaths are available from public death records, but these are compiled slowly, and are not readily available at detailed geographic specificity. Prescription data can be obtained from commercial vendors and from state monitoring programs, but these are either not publicly available or are available only at exorbitant cost. Illicit drug seizure data are collected, but these are not readily shared outside law enforcement. Self-reported drug use patterns are available from national surveys, but these do not provide regional or local data. Urine drug testing data are collected, but are held as proprietary. Finding creative ways to open up and analyze these and other data sources can lead to valuable insights into the dynamics of the opioid epidemic.

The opioid epidemic is a complex, dynamical process, and it should be approached as such in the development and evaluation of policy. A coordinated national opioid epidemic modeling program could help solve this difficult problem.

**–Donald S. Burke**

*Donald S. Burke is Dean of the Graduate School of Public Health at the University of Pittsburgh, Pittsburgh, Pennsylvania. Email: donburke@pitt.edu*

10.1126/science.aal2943

PHOTOS: (INSET) PHANIE/ALAMY STOCK PHOTO; (TOP RIGHT) ALAN ADAMS

# NEWS

> “I never should have shown my 4yo #NASApumpkin before we carved our own. Each ... has involved a motor, dry ice, and fireworks.”
>
> **@SamDillenback**, tweeting about the elaborate entries in the annual NASA–Jet Propulsion Laboratory pumpkin carving competition.

## IN BRIEF

Cape Hallett, in Antarctica's Ross Sea, is a breeding ground for Adélie penguins, considered an important bellwether species for the impact of climate change.

# Nations agree to create world's largest marine reserve in Antarctica

A key polar habitat will be getting some extra protection. After years of negotiations, nations agreed on 28 October to establish the world's largest marine reserve in the Ross Sea off Antarctica. The deal bans commercial fishing in the 1.55-million-square-kilometer swath of international waters, but allows research in 28% of the preserve. It calls for the reserve to open on 1 December 2017 and exist for at least 35 years. The pact came after Russia, China, and Ukraine dropped their objections at a meeting of the Commission for the Conservation of Antarctic Marine Living Resources—which includes 24 countries and the European Union—in Hobart, Australia. Researchers say the reserve, which borders a major ice shelf, includes two biologically rich seamounts and is home to an estimated one-half of the world's South Pacific Weddell seals, one-third of Adélie penguins and Atlantic petrels, and 25% of emperor penguins. The Wildlife Conservation Society in New York City welcomed the move to protect “this unique seascape,” adding that “the world is coming to grips with the need to conserve large swaths of our oceans.”

AROUND THE WORLD

## Cargo ships to report emissions

LONDON | In a step toward quantifying and reducing the climate impact of the shipping industry, the International Maritime Organization (IMO) announced last week it will require large cargo ships to record and report their fuel consumption. Countries will provide aggregated annual data on their fleets' fuel use to a database starting in 2018. IMO's Marine Environment Protection Committee will analyze the data to develop a strategy to cut greenhouse gas emissions from ships. One option is tightening energy efficiency standards for new ships, which IMO set in 2011. The committee also cut the amount of sulfur allowed in fuel oil, which contributes to air pollution, by 85% starting in 2020. It remains much higher than the limit for diesel fuel used by cars and trucks.

## U.S. science test shows gains

WASHINGTON, D.C. | Students in grades four and eight did slightly better in science than their cohorts in 2009 on a test

dubbed the Nation's Report Card. Results released last week from the 2015 National Assessment of Educational Progress in science also showed that the large achievement gap between white and minority students (a difference of roughly 35 points for blacks and 25 points for Hispanics on a scale of 1 to 300) narrowed slightly. In contrast, scores for the nation's 12th graders didn't budge, nor did the yawning racial disparity shrink. The results for younger students are also broken down by state, with Tennessee showing the largest increase. At the same time, only one-third of students earned a proficient score, and one-third scored below the basic level.

## A backup site for the TMT

LA PALMA, IN SPAIN | The location of the Thirty Meter Telescope (TMT) remains in the balance as hearings continue this month in Hawaii over its disputed building permit. This week, the board of the TMT International Observatory announced its backup site: the Roque de los Muchachos on La Palma, one of Spain's Canary Islands. Roque de los Muchachos is at an elevation of only 2400 meters, compared with Mauna Kea's 4050 meters, which would make infrared observations more challenging: Atmospheric moisture absorbs infrared, and the light will travel through more atmosphere. But La Palma is also an established observatory with existing infrastructure. Native Hawaiian activists object to siting the TMT on Mauna Kea, which is considered sacred by the Hawaiian religion, and have had the TMT's building permit overturned by Hawaii's supreme court on a technical issue. The TMT will be one of the world's largest telescopes when it begins operating next decade.

## Turkish academics get ax

ANKARA | In a surprise decree over the weekend, Turkey's government announced the firing of more than 1000 academics at public universities in the wake of a failed coup attempt in July. Many of those fired are learning of their fate by finding themselves on a long list of names that the government released. The government also announced that it will now directly appoint the leaders of Turkey's universities. Insiders say that the government has long held indirect control over the election of university rectors, and that the new decree simply makes the government's control of academic leadership official.

NEWSMAKERS

## Three Qs

As a graduate student at the University of Cambridge in the United Kingdom studying heart valve bioengineering, **Jacob Brubert** spent last summer cavorting in public: dancing with colleagues as polymer molecules, simulating a beating heart valve, even juggling. For his creative moves and painful honesty—Brubert ends his dance in despair as his experiment goes haywire—he earned top prize in the annual *Science*/AAAS Dance Your Ph.D.

BY THE NUMBERS

**6.5**

Magnitude of an earthquake that struck 100 kilometers northeast of Rome on 30 October—Italy's strongest quake since 1980.

**$18 million**

Amount fronted by the Bill & Melinda Gates Foundation and the Wellcome Trust to expand the release of bacteria-infected mosquitoes in South America to fight the spread of Zika.

**58%**

Fraction of wild animal populations lost between 1970 and 2012, the Zoological Society of London and the World Wildlife Fund report.

Researchers found a striping gene in this African striped mouse.

PHOTO: J. F. BROEKHUIS

## Striped mouse reveals clue to mammalian color patterns

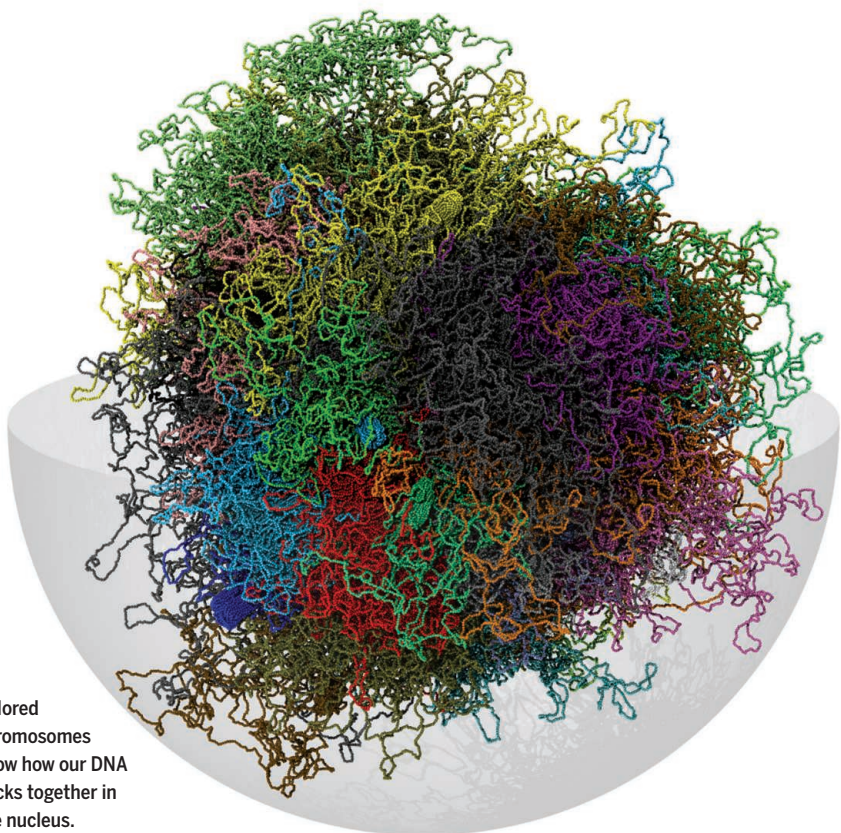
Everyone wants to know how the zebra got its stripes, but zebras are hard to study in the lab. So Harvard University evolutionary biologist Hopi Hoekstra and her colleagues turned to the African striped mouse (*Rhabdomys pumilio*), which lives in southern Africa and has alternating dark and light stripes running down its back. When they studied skin development in the mouse embryos, they discovered that a gene called *Alx3* was very active where light stripes form—its protein suppresses the formation of dark pigment, Hoekstra and her colleagues report this week in *Nature*. The same proved true where eastern chipmunks develop stripes, and because chipmunks and these mice are separated by 70 million years of evolution, Hoekstra and her colleagues think this gene may lead to stripes and other distinctive color patterns across the mammals—including zebras. A preliminary study found the *Alx3* gene was more active in the white parts of zebra skin than in the black parts. "But we will still need to do additional work to make any real case," she cautions. **http://scim.ag/Rodentstripes**

contest: $1000 and a trip to Boston for the 2017 AAAS Annual Meeting. *Science* caught up with Brubert, now in medical school at the University of Oxford in the United Kingdom.

**Q: What motivated your research?**
**A:** About 2% of people will have a heart valve problem in their lifetime, and about 300 million artificial valves get implanted each year. The mechanical ones last, but … you have to take anticlotting drugs that have painful and dangerous side effects. Animal tissue valves don't have this problem, but don't last as long. We're trying to make polymer valves that last long and are gentle on the blood.

**Q: Did your Ph.D. experiments really fail?**
**A:** I had a great Ph.D. but … not everything worked. The crucial thing we haven't figured out is how to make the valve durable. It can beat for 100 million cycles before failing—but that's only 2 years of human life. The natural valve lasts for 3 billion beats.

**Q: Has the contest changed your life?**
**A:** Explaining my Ph.D. to friends, family, and strangers on the bus is far more amusing now. [But] I now get asked to translate every presentation I do into interpretive dance!



Colored chromosomes show how our DNA packs together in the nucleus.

## Your entire genome in 3D

We used to think of our 1.8-meter-long genomes as linear molecules of DNA somewhat randomly squished into the nucleus. But over this past decade, researchers have come to realize very deliberate loops bring specific parts of each chromosome into contact to help control what genes are active. Now, biophysicist Marco Di Stefano at the National Centre for Genomic Analysis in Barcelona, Spain, and his colleagues have used sophisticated statistical approaches to convert experimental data that provided indirect information about individual connections into a comprehensive, biologically correct 3D model. It shows how all our chromosomes (different colors) pack together in the nucleus, the team reported in *Scientific Reports*. The model is "summarizing a large portion of the knowledge we have on the DNA organization in the nucleus," says Di Stefano, who did the work while a graduate student at the International School for Advanced Studies in Trieste, Italy. "It is indeed beautiful," says Job Dekker, a biologist at the University of Massachusetts Medical School in Worcester who pioneered methods to determine DNA contact points.

FINDINGS

## Ebola mutation boosts infection

Two separate prominent international groups of researchers have found evidence that the Ebola virus may have adapted to humans during the West African epidemic that occurred from 2013 to 2016. As researchers describe in two papers published in the 3 November issue of *Cell*, a mutation in one amino acid of the Ebola virus made it easier for the virus to infect human cells. The papers, which looked at more than 1000 genetic sequences from West African isolates of the virus, are careful not to overstate the meaning of the finding: One group writes that they were "left with a strong suspicion" that a human adaption occurred, and they further noted that "it remains an open question whether this change led not only to increased replication in human hosts, but also to increased transmission between humans." More than 28,000 people became infected with Ebola virus during the epidemic—by far the most human-to-human spread that has ever occurred. Although a key driver of the explosive thread likely was the fact that the virus for the first time made it to densely populated cities that had poor health infrastructures, these findings do raise the possibility that evolutionary forces were at play, too.

## First dinosaur brain fossil found

A decade ago, a fossil hunter was combing the beach in southeastern United Kingdom when he found a strange, brown pebble. The surface of it caught his eye: It was smooth and strangely undulating, and also slightly crinkly in some places. That oddly textured pebble, scientists reported last week at the Society for Vertebrate Paleontology meeting, is actually an endocast—an impression preserved in the rock—that represents the first known evidence of fossilized brain tissue of a dinosaur (likely a close relative of *Iguanodon*, a large, herbivorous type of dinosaur that lived about 133 million years ago). The structure of the brain, studied with scanning electron microscopes, reveal similarities to both birds and crocodiles. It's the first known evidence of such a dinosaur brain—but now that they know to look for it, the researchers say, they might go back and look at other endocranial casts of dinosaurs to see whether they might contain traces of other such structures. **http://bit.ly/dinobrains**

IMAGE: SISSA

One metric ton of $CO_2$ costs the Arctic 3 square meters of summer sea ice.

CLIMATE CHANGE

# *Sea ice shrinks in step with carbon emissions*

## Models may underestimate the pace of ice loss because they are missing Arctic warming

*By* **Warren Cornwall**

The jet fuel you burned on that flight from New York City to London? Say goodbye to 1 square meter of Arctic sea ice.

Since at least the 1960s, the shrinkage of the ice cap over the Arctic Ocean has advanced in lockstep with the amount of greenhouse gases humans have sent into the atmosphere, according to a study published this week in *Science*. Every additional metric ton of carbon dioxide ($CO_2$) puffed into the atmosphere appears to cost the Arctic another 3 square meters of summer sea ice—a simple and direct observational link that has been sitting in data beneath scientists' noses. "It's really basic," says co-author Dirk Notz, a sea ice expert at the Max Planck Institute for Meteorology in Hamburg, Germany. "In retrospect, it sounds like something someone should have done 20 years ago."

If both the linear relationship and current emission trends hold into the future, the study suggests the Arctic will be ice free by 2045—far sooner than some climate models predict. The study suggests that those models are underestimating how warm the Arctic has already become and how fast that melting will proceed. And it gives the public and policymakers a concrete illustration of the consequences of burning fossil fuels, says Edward Maibach, director of the Center for Climate Change Communication at George Mason University in Fairfax, Virginia. "Concrete information is always more engaging than abstract information," he says.

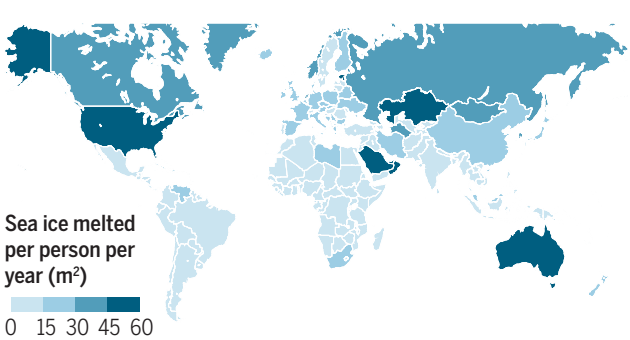
According to the new calculations, for instance, the average annual carbon emissions from a U.S. family of four would claim nearly 200 square meters of sea ice. Over 3 decades, that family would be responsible for destroying more than an American football field's worth of ice—a tangible threat to ice-dependent creatures such as polar bears. The study also makes for vivid comparisons between nations: Each person in the United States, for instance, is responsible for the destruction of 10 times as much ice each year as someone in India (see graphic, below).

Sea ice retreat is already a poster child for global warming, and some earlier studies had shown that the retreat closely tracks atmospheric $CO_2$ levels. But Notz and co-author Julienne Stroeve, an expert in sea ice satellite measurements at the National Snow and Ice Data Center in Boulder, Colorado, identified a tighter link to human activity by compiling annual human-caused greenhouse gas emissions and comparing those numbers with historic observations of Arctic sea ice coverage during September—when sea ice is at its smallest. To avoid problems with year-to-year ice fluctuations, they used a 30-year moving average of ice coverage, allowing them to study the years from 1968 to 2000.

To see whether the linear relationship they found between emissions and sea ice also emerged in computer simulations, they checked 36 of the world's major climate models. In simulations in which $CO_2$ levels rose 1% every year, they found the same telltale pattern every time, Stroeve says. Yet the models' sensitivity was off: They tended to underestimate the amount of ice loss. "Models are not perfect," Stroeve says. "And if you can use observations by themselves to forecast when Arctic ice will go away maybe that's in some ways better."

### On thin ice

Based on $CO_2$ emissions in 2013, each U.S. resident led to the melting of 49 square meters of Arctic sea ice—nearly 10 times as much as someone in India.




CREDITS: (PHOTO) PAUL NICKLEN, NATIONAL GEOGRAPHIC CREATIVE; (GRAPHIC) J. YOU/*SCIENCE*; (DATA) WORLD BANK

By tracing the trajectory of their observations into the future, Stroeve and Notz estimate that another 1000 gigatons of $CO_2$ would push summer sea ice coverage to below 1 million square kilometers, an area the size of Texas and New Mexico combined. That's essentially ice free, because the remaining ice would be tucked into pockets around places like the northern edge of Greenland, where winds concentrate ice. If today's carbon emissions of 35 gigatons per year persist, that translates to an ice-free ocean, at least in summer, by about 2045. Scientists say the same 1000 gigatons that would erase sea ice would also warm the world by 2°C—the threshold that the Paris climate agreement intends to stay below.

Stroeve and Notz aren't certain why the shrinking of the sea ice tracks emissions so neatly. But they point to a simple potential explanation: As rising emissions warm the Arctic air, the ice retreats to more northerly latitudes, where there is less heat from incoming direct sunlight. Notz thinks that climate models tend to downplay the amount of sea ice lost to each ton of $CO_2$ because they're underestimating how warm the Arctic is getting. Although that might seem easy to check, data for air temperature over the Arctic sea ice is quite sparse, says Gregory Flato, a sea ice scientist at the Canadian Centre for Climate Modelling and Analysis in Victoria.

Alexandra Jahn, a sea ice modeler at the University of Colorado in Boulder, says Stroeve and Notz's mechanism "finally explains this linear relationship that people have been seeing, but didn't have a physical explanation for." But she thinks further studies are needed to rule out other factors that might influence sea ice decline, such as changes in cloud cover. Stroeve and Notz say the impact of clouds and ocean temperatures aren't significant.

They do acknowledge, however, that sea ice half a world away in the Antarctic doesn't track emissions as faithfully. There, forces including wind and ocean temperatures have a stronger impact on sea ice behavior, they say.

François Massonnet, a sea ice modeler at the Barcelona Supercomputing Center in Spain, cautions against assuming that what happened in the recent past will last into the future. A curveball, such as a shift in the amount of $CO_2$ the oceans can absorb, could upset the pattern by breaking the correlation between emissions and Arctic warming. "It's tempting, but it's very dangerous to be too confident that we can extrapolate linear relationships," he says. "We know that the Arctic climate is very nonlinear." ■

FUNDING

# Ambitious web fundraising startup fails to meet big goals

## Hundreds of scientists paid fees to have their work featured on Benefunder, but very few received donations

*By* **Mark Harris**

It seemed like a promising deal. Pay up to $2000 to have your research project featured on a slick website aimed at affluent donors looking to support a good cause and get a tax break, and then watch the money roll in. But it hasn't paid off for most of the 650 researchers who signed up with Benefunder, a 3-year-old firm based in San Diego, California, that once envisioned raising more than a billion dollars annually for research. Just six researchers have received donations totaling some $250,000 through Benefunder, *Science* has learned. The startup has had to lay off most of its employees and is struggling to revamp its business model.

Benefunder's stumble represents a cautionary tale from the promising but sometimes perilous frontier of internet fundraising. Over the past decade, so-called crowdfunding websites, including Kickstarter, Indiegogo, and GoFundMe, have enabled moviemakers, entrepreneurs, and others to raise sometimes vast sums from small donors around the world. A few sites have specialized in raising money for scientists, with varying success. Experiment has raised nearly $7 million for more than 550 researchers since it began 5 years ago, with nearly half of proposed projects meeting funding goals. But another site, Petridish, quietly closed its doors in 2014 after funding just a few dozen experiments over 3 years.

Benefunder, the brainchild of entrepreneur Christian Braemer and Gert Lanckriet, a machine learning professor at the University of California, San Diego, aimed to become a science funding star. Instead of seeking donations from rank-and-file web users, it targeted affluent contributors and the organizations they use to make charitable donations. Benefunder even set up a financial mechanism, known as a donor advised fund, to ease a donor's ability to claim U.S. tax breaks. By taking a 10% cut of the donations, Benefunder expected to earn some $1 million in 2014, $13 million in 2015, and a staggering $165 million in 2016—representing more than $1 billion in donations—according to forms Braemer filed with the state of California.

> ***"The idea of targeting small investors and exploiting a tax break strategy felt very innovative."***
>
> **François Baneyx**, University of Washington

Benefunder's vision appealed to hundreds of scientists, many from top universities. Each paid about $500 to have their work profiled on the website, and some spent up to $1500 more for a short video produced by the firm. "The idea of targeting small investors and exploiting a tax break strategy felt very innovative," says François Baneyx, a chemical engineer at the University of Washington in Seattle, who signed up in 2014. "I had reasonable expectations, perhaps getting $10,000 to $50,000."

But even as Benefunder bulged with projects, donors remained scarce. "We were never able to get off the ground," Braemer says. Donor funds "were not willing to take the reputational risk [on] an unknown entity," he says. And the firm received just a few "small transactions ... a bit out of the blue."

To stay afloat, Benefunder ramped up sales of the profiles and videos. In 2014 and 2015, it earned more than $660,000 this way but attracted just $62,000 in gifts, tax forms show. In late 2015, as the firm ran out of cash, it abruptly stopped recruiting researchers, left some videos unfinished, and laid off all but three of the 12 employees who worked for it and an allied firm.

Benefunder is now "licking its wounds," Braemer says, and with the help of new investors has "pivoted" to advising large wealth management firms on how to match their donors with its researchers. It's not clear how that strategy will fare. But the firm, Braemer says, no longer expects "anyone to go to the website and actually make a donation." ■

*Mark Harris is a writer in Seattle, Washington.*

South Africa's high HIV infection rate makes it an ideal place to stage a vaccine trial. Many people becoming infected are young, like this woman hospitalized with AIDS earlier this year in KwaZulu-Natal.

PHOTO: JASON KANE

AIDS EPIDEMIC

# *Controversial HIV vaccine strategy gets a second chance*

## Modest success in Thailand inspires South Africa trial

*By* **Jon Cohen**

A two-pronged HIV vaccine strategy that delivered lackluster results in a trial in Thailand 7 years ago will get another chance in South Africa. Last week, researchers injected the first of what they hope will be 5400 healthy men and women between 18 and 35; each will receive either a combination of two vaccines or placebo shots. The $130 million study should show once and for all whether the combination—the only HIV vaccine so far to show any hints of benefit in an efficacy study—actually works.

Some researchers say the results in Thailand were so unimpressive that the massive study, funded primarily by the National Institute of Allergy and Infectious Diseases (NIAID) and the Bill & Melinda Gates Foundation, amounts to a waste of research funds. "That money could be much better spent developing promising concepts that have a real chance of working," says Ronald Desrosiers, an HIV vaccine researcher at the University of Miami in Florida. But Anthony Fauci, head of the NIAID in Bethesda, Maryland, says there was enough evidence of efficacy in the Thai trial to warrant another test. "We're just going to have to let it play out," Fauci says.

Trial leaders argue that even a modestly efficacious vaccine would help South Africa, which has more than 6 million HIV-infected people—19% of the world's total. "There have been all these naysayers," says Glenda Gray, who heads the South African Medical Research Council in Cape Town and will lead the new study. "Do they not care? It means they're completely out of touch with the epidemic and that HIV continues to ravage Africa."

The Thai trial, which involved more than 16,000 people, found that the vaccine reduced the risk of HIV infection by a modest 31.2%. Volunteers first received a canarypox virus engineered to deliver several HIV genes. This was followed by a shot that only contained HIV's surface protein, gp120, plus alum, an immune stimulant called an adjuvant. Both vaccines were designed to protect against strains of HIV in circulation in Thailand; the dual approach was meant to stimulate different immune responses.

For the new study, researchers built similar vaccines with strains common in South Africa but chose a stronger adjuvant called MF59. Participants will also receive an extra injection of the gp120 vaccine in an attempt to make the protection more durable.

Critics say the problem isn't just the low level of protection seen in Thailand. The study also failed to show a mechanism that could explain the modest benefit, they argue. Researchers studying stored blood samples from the Thai study found that vaccinated people who did not become infected had higher levels of antibodies that bind to gp120. But in test tube experiments, the same binding antibodies do not prevent the virus from infecting cells, which some researchers insist is critical for protection. "Fundamentally, there's no logical pathway from what we know to this vaccine trial," says virologist Paul Bieniasz of the Aaron Diamond AIDS Research Center in New York City. "There should be more intelligent debate about what moves forward in the pipeline."

Bieniasz and others think vaccine researchers should focus on other approaches, like the rare antibodies that have unusual power to neutralize HIV in the test tube and work against a huge range of mutant strains—crucial for a vaccine against the constantly mutating virus. Many groups are attempting to develop vaccines that can teach the body how to make these broadly neutralizing antibodies, but progress has been slow.

Genoveffa Franchini, a vaccine researcher at the U.S. National Cancer Institute in Bethesda, has another concern about the trial: the new adjuvant. In May, Franchini and co-workers reported in *Nature Medicine* that the vaccine duo works in rhesus macaques when gp120 is combined with alum, but not with MF59. "If MF59 is doing the same thing that it's doing in the macaques, then you're going to end up having no protection and you won't know why," Franchini said at a recent HIV/AIDS conference at Cold Spring Harbor Laboratory in New York. She suggested MF59 might even overstimulate the immune system and create more target cells for the virus to infect, leaving vaccinated people more vulnerable to infection. "There's this oath that we take when we finish up medical school: Don't harm," she cautioned.

But MF59 did work in another monkey study, said Lawrence Corey of the Fred Hutchinson Cancer Research Center in Seattle, Washington, who heads the NIAID-funded network overseeing the study. "People of good faith will have different answers," he said.

Gray, the study leader, thinks Franchini makes a "compelling" case to add an alum arm to the trial—if the researchers can find the money. But she says the South African study is well designed as is, with many checkpoints at which it can be halted if the vaccine is harmful or even if it becomes clear that vaccinated people are getting infected at the same rate as those in the placebo group. "Someone has to put their stake in the ground and have the courage to move forward, knowing we might fail," Gray says. ■

One side of a slab holds the fossil's body and a mold of its skull (top); the "counterpart" has the skull and a mold of the body.

PHOTO: H. TISCHINGER

PALEONTOLOGY

# 'Four-legged snake' may be ancient lizard instead

## Controversy swirls around fossil, which is no longer accessible

*By* **Carolyn Gramling**, *in Salt Lake City*

It is a tiny, fragile thing: a squashed skull barely a centimeter in length; a sinuous curving body about two fingers long; four delicate limbs with grasping hands. In a major paper last year, researchers called this rare fossil from more than 100 million years ago the first known four-legged snake. But at a meeting of the Society of Vertebrate Paleontology (SVP) here last week, another team suggested that it's a marine lizard instead. Even as scientists debate the identity of this controversial specimen, the only one of its kind, it appears to be inaccessible for further study. And paleontologists are mad as hell.

"It's horrifying," says Jacques Gauthier, a paleontologist at Yale University. As far as he's concerned, if the fossil can't be studied, it doesn't exist. "For me, the take-home message is that I don't want to mention the name *Tetrapodophis* ever again."

A year ago, researchers led by David Martill of the University of Portsmouth in the United Kingdom reported in *Science* that the fossil, which they named *Tetrapodophis amplectus* (for four-footed snake), was a missing link in the snake evolutionary tree (24 July 2015, p. 416). Researchers knew snakes had evolved from four-limbed reptiles, but few transitional forms had been discovered, and researchers continue to wrangle over whether the first lizards to lose their limbs and become snakes were terrestrial burrowers or aquatic swimmers.

Martill and colleagues reported that the fossil, which they described as a specimen in a German museum, originated from a Brazilian outcrop of the Crato Formation, a 108-million-year-old limestone layer rich in both marine and terrestrial species. They identified snakelike features in the fossil, including a long body consisting of more than 150 vertebrae, a relatively short tail of 112 vertebrae, hooked teeth, and scales on its belly. Those features, they say, support the hypothesis that snakes evolved from burrowing ancestors.

> ***"Tetrapodophis is no longer science. ... It's not repeatable, it's not testable."***
>
> **Jason Head**, University of Cambridge

But many paleontologists weren't convinced. Last week at the meeting, paleontologist Michael Caldwell of the University of Alberta in Edmonton, Canada, and colleagues presented their own observations of the specimen, rebutting Martill's paper point by point to a standing-room-only crowd.

The new analysis hinges on the "counterpart" to the original fossil, which was also housed in the Bürgermeister-Müller Museum in Solnhofen, Germany. When the slab of rock containing the fossil was cracked open, the body of the organism stayed mostly in one half of the slab, whereas the skull was mostly in the other half, paired with a mold or impression of the body. This counterpart slab, Caldwell says, preserved clearer details of the skull in particular. In his group's analysis of the counterpart, he says, "every single character that was identified in the original manuscript as being diagnostic of a snake was either not the case or not observable."

For example, in snake skulls, a bone called the quadrate is elongated, which allows snakes to open their jaws very wide. This fossil's quadrate bone is more C-shaped, and it surrounds the animal's hearing apparatus—a "characteristic feature" of a group of lizards called squamates, says co-author Robert Reisz, a paleontologist at the University of Toronto in Mississauga, Canada. He and Caldwell add that although the fossil has more vertebrae in its body than in its tail, the tail isn't short, but longer than that of many living lizards. They are working on a paper arguing that the fossil is probably a dolichosaur, an extinct genus of marine lizard.

Martill and co-author Nicholas Longrich of the University of Bath in the United Kingdom, neither of whom was at the meeting, stand firmly behind their original analysis. Longrich cites all the snakelike features discussed in the original paper. "In virtually every single respect [it] looks like a snake, except for one little detail—it has arms and legs," he told *Science* by email.

Many researchers who attended the talk, including Gauthier and paleontologist Jason Head of the University of Cambridge in the United Kingdom, are persuaded that

*Tetrapodophis* is not a snake. But as for what it is, there may be as many opinions as there are paleontologists. Hong-yu Yi of the Institute of Vertebrate Paleontology and Paleoanthropology in Beijing, China, says she's reserving judgment on the specimen's identity until further analysis. "I was always waiting for a longer description of the specimen. I'm still waiting," she says.

That analysis may never happen. Caldwell says he went back to the Bürgermeister-Müller Museum several months ago to study the specimen again; separately, Head says he also attempted to study the fossil. Neither could get access to it. Caldwell says that the fossil wasn't actually part of the museum's collection, but was on loan from a private owner. Researchers who declined to be named because of ongoing discussions around the fossil say that it may have been damaged during study, prompting the collector to restrict access to it. "I don't even know if a publication at this moment is appropriate because no one else will be able to access this specimen," Yi says.

In fact, some researchers question whether the original paper should have been published, because the fossil was not officially deposited in a museum or other repository, so the authors couldn't guarantee that future researchers could access it. "I have nothing against" private fossil collecting, Gauthier says. But when a fossil enters the scientific literature, he says, "then it has to be available. Science requires repeatability." In response, *Science* Deputy Editor Andrew Sugden in Washington, D.C., says, "Our understanding at the time of publication and in subsequent correspondence was that the specimen was accessible at the museum, as stated at the end of the paper."

Researchers had raised other questions about the fossil's transport out of Brazil. Brazil passed laws in the 1940s making all fossils property of the state rather than private owners. "Most of the exploration of the limestone quarries in that region of the country began in the second half of the 20th century," says Tiago Simões, a paleontologist at the University of Alberta, who was also an author on the SVP talk. "So the vast majority" of fossils from those areas were collected after the law had passed. "That really touches on some very sensitive ethical boundaries."

Head agrees. "The best way to move forward is to literally erase the specimen from our research program. *Tetrapodophis* is no longer science. … It's not repeatable, it's not testable. If any good can come out of *Tetrapodophis*, it's the recognition that we have got to maintain scientific standards when it comes to fossils … they have to be accessible." ■

DATA CHECK

BEHIND THE NUMBERS

# California rules U.S. corporate research

*By* **Jeffrey Mervis**

Every state wants to attract the next Alphabet (Google's parent company). Such high-tech firms not only provide high-paying jobs, but also help boost the research capacity of its universities. But the latest data from the National Science Foundation (NSF) show that most state economic development officials have their work cut out for them: Three of every 10 corporate research dollars are now spent in California, and over the past 5 years the Golden State has widened its lead over the rest of the country.

The data come from NSF's annual Business R&D and Innovation Survey (BRDIS). The survey documents a $265 billion enterprise in which some 500 companies account for 80% of the total spending. The survey finds that California was home to $77 billion in corporate research expenditures in 2013, thanks to companies such as Google, Facebook, and Lockheed Martin. That total equals the combined spending in the next five states: Michigan, Massachusetts, Washington, Texas, and Illinois. Together, the top six states account for nearly 57% of the $255 billion in corporate research that NSF was able to link to a particular location. Yet those states comprise just 37% of the U.S. population and generate 35% of the country's gross domestic product.

The report also shows that California's corporate research economy weathered the Great Recession better than did the economies of other states. California's research output grew 42% between 2008 and 2013, according to the survey, compared with a 7% increase nationwide. That healthy performance boosted California's share of the U.S. total from 25% in 2008 to 30% in 2013, a figure that is well above California's roughly 13% share of the country's population and the national GDP.

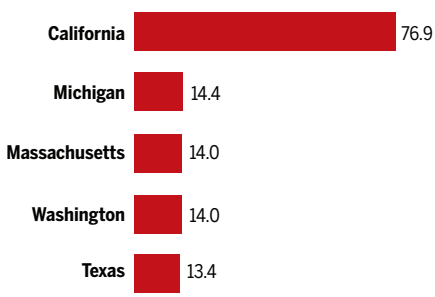
The Golden State's increasing dominance of corporate research was fueled by a boom in information and communications technology services, which grew faster than any other industrial research sector in the past 5 years. California also has benefited from the fact that software research is no longer done exclusively by information technology (IT) companies like Microsoft, based in Seattle, Washington. Overall, software research, for the first time, accounts for the majority of the $121 billion spent on all IT research, which includes research done by semiconductor manufacturers.

The 2008 financial crisis was hard on corporate research in states dominated by a single sector. For example, downsizing by pharmaceutical companies dropped New Jersey from second to seventh place in industrial research activity. A rebound in car sales allowed Michigan, where the automotive industry accounts for 74% of the state's overall research activity, to remain a distant second behind California.

The BRDIS also zooms in on research activity at the level of metropolitan areas. California is home to the top two corridors: the San Jose–San Francisco–Oakland and the Los Angeles–Long Beach regions.

## A big lead

California is ahead of the pack when it comes to corporate research spending ($ billions, 2013).




Together, research-intensive companies (those spending more than $3 million annually) reported performing 35% of their research in those two urban corridors. Washington state's Seattle-Tacoma-Olympia corridor placed third, slightly ahead of Michigan's Detroit–Warren–Ann Arbor complex and the Boston-Worcester-Providence axis stretching from Massachusetts to Rhode Island.

The report is a sobering reminder that the ongoing competition among states to land new research facilities can be a winner-take-all struggle. Some 52% of the companies with the largest research portfolios report doing all of that work at only one site (typically their headquarters), and 69% perform all of their research at just two locations. That reality suggests state officials who are California dreamin' might be better off nurturing an in-state startup—perhaps a university spinoff—rather than trying to lure a high-tech giant from another state. ■

CREDITS: (DIAGRAM) D. MALAKOFF/*SCIENCE*; (DATA) NATIONAL SCIENCE FOUNDATION

Rudolf Jaenisch made the first transgenic mouse in 1974 and has pioneered the engineering of CRISPR mice.

# MICE MADE EASY

## The genome-editing tool CRISPR upends the vital business of creating mutant mice

*By* **Jon Cohen**

In the beginning of 2013, Michael Wiles sat down with high-level managers of the Jackson Laboratory in Bar Harbor, Maine, and told them about a novel way to cut DNA that had amazing power. The lab, called JAX for short, genetically engineers mice that it sells to researchers under a trademarked brand: JAX® Mice, it likes to boast, "are the highest quality and most-published mouse models in the world." Wiles evaluates and develops technologies for the lab, and he was convinced that this new tool, ingeniously adapted from an immune strategy that bacteria and archaea use to protect themselves from viruses, would revolutionize the way JAX engineered mice. "Of about a dozen people, nine were asleep," Wiles says. "No one had heard of CRISPR."

Now, most every mouse developer has. JAX and other labs making new mouse strains have long relied on a laborious multistep process that involves genetically altering mouse embryonic stem (ES) cells, injecting them into an embryo, and breeding multiple generations of animals. Even JAX's crack team took up to 2 years to engineer a mouse. CRISPR replaces all that with a molecular complex that can do targeted genetic surgery on a fertilized egg. It can produce a strain of transformed mice in 6 months. "It's night and day," Wiles says. "We had five or six people working with ES cells. They were close friends of mine and I said, 'You better look for a job.'"

Mice genetically modified to cripple or "knock out" genes or to add or "knock in" genetic information have become key research models for a wide array of human diseases, from cancer and atherosclerosis to Alzheimer's, osteoarthritis, muscular dystrophy, and Parkinson's. Knockout and knockin mice also offer a powerful tool for probing the functions of specific genes.

Most investigators get their engineered mice from colleagues or by purchasing them from commercial outfits like JAX or academic-based repositories. Popular engineered mice, such as JAX's immunodeficient NOD scid gamma strains, sell for as little as a few hundred dollars, but a custommade mutant could cost as much as $20,000. By making the engineering of mice far simpler and cheaper, CRISPR opens the way for more labs to do it themselves. "When you made knockout mice before, you needed some skills," says Rudolf Jaenisch at the Massachusetts Institute of Technology (MIT) in Cambridge. "Now, you don't need them anymore. Any idiot can do it."

If CRISPR's talent with rodents is shaking up individual labs, it is causing an earthquake in an international consortium to knock out all 21,000 mouse genes, one by one, in order to reveal their functions. The consortium, which includes JAX, has spent $350 million to date and is about a quarter of the way to its goal. Many investigators hope the speed and cost savings of CRISPR will accelerate progress. The National Institutes of Health (NIH), for one, is so impressed with CRISPR's ease and power that it no longer funds consortium investigators to use ES cells.

> ***"It's really changed the time and efficiency of getting these engineered animals."***
> **Tak Mak**, University of Toronto

But that's where some mouse engineers have second thoughts about the rush to CRISPR. Few doubt its potential, but the technique is still a work in progress, and its ability to alter genomes has one big gap. Although CRISPR knocks out genes with ease, it is less efficient at inserting, or knocking in, new DNA. That's important not just for giving an animal a novel function, but also for creating a knockin known as a "conditional" knockout, an animal model in which researchers can turn off a target gene at specific times of life or in specific tissues.

Because CRISPR is less adept at making conditional knockouts, William Skarnes, who led a team making mutant mouse ES cells at the Wellcome Trust Sanger Institute in Hinxton, U.K., worries that NIH is overemphasizing the new approach. "The decision to abandon the ES resource in favor of making simple knockouts is a mistake," Skarnes says. "You still want to make conditionals through the ES route."

CRISPR researchers are now refining the technique to do knockins with greater efficiency. But that entails tinkering with the mechanisms that cells use to repair broken DNA, which are critical to their health. "I'm cautious about overmanipulation of biology to increase efficiency," says Steve Murray, who helps run JAX's contribution to the Knockout Mouse Phenotyping Program (KOMP2), which is part of the international consortium. "We're waiting in the wings for the wizards in the field to help us with this."

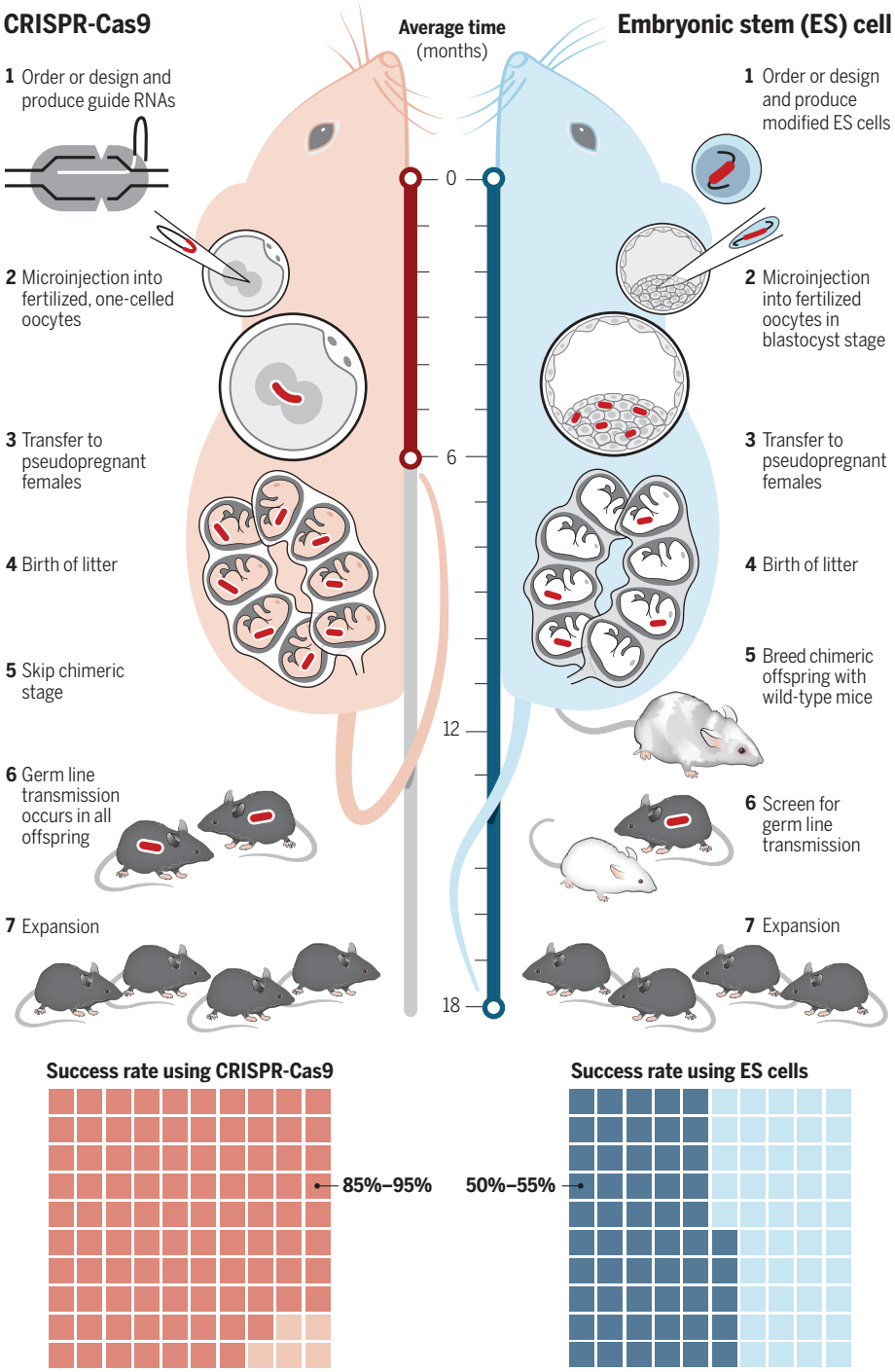
**CRISPR STANDS FOR** "clustered regularly interspaced short palindromic repeats," which is a description of the prokaryotic genetic material from which it was derived. It uses what's called a guide RNA to send biological scissors—usually the CRISPR-associated protein, Cas9—to a precise spot in a genome. Once Cas9 enzymatically makes the cut, the cell tries to heal the wounded DNA. One repair mechanism leads to knockouts, whereas the second leads to knockins. "All CRISPR does is cut the DNA," Wiles says. "Everything else is the cell repair system, and that's what we're hitching on to."

The cell's standard response is to try to paste the double-stranded DNA back together at the break points. This often requires eating away or adding a few bases—the As, Cs, Ts, and Gs that make up DNA—which leads to insertions or dele-

PHOTO: (OPPOSITE PAGE) KEN RICHARDSON

## See how CRISPR runs

Using the genome-engineering tool to alter a fertilized egg is a quicker and more efficient way to engineer a mouse than the traditional route, which starts by modifying an embryonic stem cell.




GRAPHIC: K. SUTLIFF/*SCIENCE*

tions. In effect, the repair effort introduces typos into the DNA text, disabling the gene.

MIT's Jaenisch was the first to show the power of CRISPR for producing mouse knockouts. In a 2 May 2013 paper in *Cell* that appeared 5 months after researchers first showed CRISPR could work in mammalian cells, he and co-workers reported that the technique successfully disrupted five genes in a single set of mouse ES cells, something that was not possible before. More important, they showed that they could bypass ES cells altogether and simultaneously knock out two genes in single-celled mouse zygotes, or fertilized eggs. No longer would researchers have to modify ES cells and painstakingly breed several generations of mice to produce an animal that carried the mutant gene in its egg or sperm cells. And researchers who wanted mice with two mutations would no longer have to interbreed single mutants and go through a similarly time-consuming, cumbersome process to arrive at progeny with the altered germ line. As the title of Jaenisch's paper declared triumphantly, "One-step generation of mice carrying mutations in multiple genes."

Since then, more than 500 papers have detailed how CRISPR can both knock out and knock in genes in mice. "The impact it's had is enormous," says Jaenisch, who in 1974 created the first transgenic mouse. "It's really changed the time and efficiency of getting these engineered animals," adds biochemist Tak Mak of the University of Toronto in Canada, who was also a pioneer in the mouse-mutating business. Mak estimates it's about 30% cheaper to engineer a mouse with CRISPR than with ES cells, bringing his average cost down to about $100,000.

CRISPR's impact is measured in more than savings. The ease and speed of the technique makes it possible to engineer mice on the fly, to solve specific puzzles like one that C. C. Hui of the Hospital for Sick Children in Toronto recently confronted: a knockout in which the missing gene didn't have any observable effect. Hui realized that the knocked-out gene was linked to another gene that might be compensating for it. He took the problem to Lauryl Nutter, who oversees mouse making at the Centre for Phenogenomics in Toronto. She used CRISPR to mutate the offending gene in a zygote from the original knockout. "We got the zygote, injected it with CRISPR-Cas9, and 8 weeks later he had a double mutant on the ground," Nutter says. "That would have taken years with ES cells."

The revolution is not limited to making mice with germline mutations. CRISPR has allowed investigators to mutate several suspected cancer genes simultaneously in the somatic cells of adult mice, for example. CRISPR knockins have also corrected disease-causing gene defects in adult mice, such as the mutations that cause hemophilia and sickle cell anemia. And several groups plan to inject CRISPR into a developing mouse; the goal is to create mutations that act as barcodes and allow scientists to track cell lineages as they differentiate.

Mousemaking outfits like the Centre for Phenogenomics and JAX expect that CRISPR will vastly expand the range of mutants they produce. "Now I can take a really exotic mouse that has three genetic modifications and modify it again," JAX's Wiles says. "We couldn't do sequential modification with ES cells. We could breed a mouse

# A reporter does CRISPR

*By* **Jon Cohen**

I speak biology fluently, but the molecular complexities of the novel genome-editing tool called CRISPR left me as befuddled as when I peruse descriptions of the inflationary universe. So I decided to test what one investigator told me: CRISPR (for "clustered regularly interspaced short palindromic repeats") may sound intimidating, but it is so simple to use that "any idiot" could do it.

I would give it a try.

CRISPR works best at crippling, or knocking out, genes, so that's how I choose to use it. (For a video, see bit.ly/vid-6312.) But I aim high: I target an immune gene that, I theorize, could lead to insights into reducing the harm done by Zika virus. (My admittedly wild hypothesis is that the gene, *CD32*, may help drive Zika virus to copy itself to higher levels if a person was previously infected with dengue and has antibodies to that virus.)

Roland Wagner, a postdoc in the lab of Sumit Chanda at the Sanford Burnham Prebys Medical Discovery Institute in San Diego, California, agreed to serve as my CRISPR sensei. An experienced rock climber originally from Austria, Wagner approaches everything methodically. He pulls up the sequence of the *CD32* gene, which has five distinct protein-coding regions. If we cut the DNA in one region, the gene most likely would be knocked out: It would no longer make its protein.

Biologist Roland Wagner (left) watches as Jon Cohen attempts a key pipetting step in creating a CRISPR construct.

CRISPR uses a guide made of RNA to direct molecular scissors—part of the CRISPR-associated protein, or Cas9—to exact spots in a genome. We could buy the guide RNA (gRNA), but the idea appalls Wagner. "I would assume it's probably $500 to buy the gRNA, but I wouldn't know," he says. "We're making our own and we're spending about $5."

The gRNA sequence must complement a stretch of 20 nucleotides on the segment of the *CD32* gene we want to cut. But the same DNA sequence could recur elsewhere in the genome, leading the molecular scissors to cut in the wrong place. Such "off-target" effects can cause mayhem, and eliminating them is a key goal of those honing their CRISPR skills. To make the match more specific, Cas9 requires an additional sequence flanking the targeted 20 nucleotides: N-G-G, in which "N" can be any nucleotide. Where Cas9 finds N-G-G immediately following the 20 nucleotides, it attaches to and opens the double helix, allowing the gRNA to bind. Cas9 then cuts each strand of the DNA.

To homebrew our gRNA, Wagner copies the sequence of the *CD32* segment we've identified and pastes it into a freely available database, Optimized CRISPR Design, that looks for a matching set of 20 nucleotides followed by N-G-G. There are 41 options within *CD32*. The database scans the entire human genome to see whether there are identical matches elsewhere—potential sites of off-target cuts. We select a sequence that appears unique, and then he goes to another website and orders a stretch of DNA—an oligonucleotide—with that sequence.

The oligo arrives, and I lose my modern pipetting virginity. I have not worked in a lab since I was an undergraduate more than 30 years ago. Back then, I learned a pipetting technique that probably was invented by Louis Pasteur: I put a finger in my mouth and then sucked up a chemical into a thin glass tube, capping it with my fingertip when I had drawn up enough.

Now, at Wagner's lab bench, I face a rack of fancy plastic gizmos that look like squirt guns but enable users to suck up precise microliters of liquid with a push of a button. My task is to pipette the oligo from one tiny test tube into another. The second tube holds a plasmid, which is a circular piece of DNA that will act as a Trojan horse. This plasmid, customized for CRISPR experiments, already holds the gene for Cas9. It also contains a 60-nucleotide "hairpin" sequence that ultimately will attach to the 20 nucleotides I add to make the full gRNA.

I use one of the fancy pipettes to move the oligo into the plasmid tube, and I also add buffer, water, and an enzyme. If all goes well, the enzyme will cut open the plasmid, removing a piece of its DNA and allowing the oligo to take its place.

All does not go well.

"Oops!" Wagner says as I pipette the enzyme. "You failed a little bit." I apparently hit the pipette button before submerging the tip into the liquid.

In the end, I manage the procedure. After waiting for the chemical reactions to take place, we take my CRISPR plasmid to an electrophoresis machine, a tray that has wires hooked to it. We add a liquid that quickly turns to gelatin, and then I pipette a few drops of my CRISPR plasmid into different lanes on the device. I flip a switch to apply an electric current, which should separate the DNA into bands based on weight. The small piece of DNA I cut out with the enzyme should form a distinct band.

My gel electrophoresis only has one band, from the plasmid.

"It doesn't look like it worked," Wagner says gently. "I didn't want to be all picky, but it could be that you messed up the enzyme with your pipetting." He allows that when he was starting out, his experiments often failed. "I'd go home and I'd say, 'I hate my life,'" he confided. "There are a lot of setbacks in science."

I've already learned that any idiot cannot do CRISPR: It takes, at least, basic laboratory skills.

Wagner conducts the experiment in parallel with me, and his plasmid properly incorporates the oligo that will guide Cas9 to its target. We then coax a cell line made from an embryonic kidney into taking up the Trojan horse plasmid. After a few days, we isolate the DNA from the cells, amplify it with the polymerase chain reaction, and use electrophoresis to show that the *CD32* gene has been cut into pieces. Voilà, our knockout worked.

"You did great," Wagner tells me. "Way to go!"

I did not do great. But CRISPR did its job. ■

PHOTO: BRAD OLANDER

with two modifications with another that had two modifications and the alleles scattered like the wind. It would take years to get all four modifications."

Wiles says this likely won't affect JAX's bottom line. "Instead of shipping thousands of boxes with one variety, we will have hundreds of boxes with tens of varieties."

**ON ONE FRONT**, however, the CRISPR revolution is faltering. Three months after his lab's first CRISPR report, Jaenisch and co-workers published a second paper in *Cell* that suggested CRISPR could easily perform more complex genetic surgery, knocking in chunks of DNA rather than simply disabling genes. As a demonstration, they used CRISPR to knock fluorescent tags into mouse zygotes, which lighted up whenever a specific gene was turned on. They also created conditional mutants, which are key to many research efforts, including the knockout consortium.

CRISPR works in every strain of mouse, whereas ES-cell technology is mostly limited to one inbred strain.

Conditionals get around a barrier to making knockouts. About one-third of mouse genes are essential for embryonic growth; the mouse is never born if they are disabled from the start. So researchers working with ES cells cleverly designed a system called Cre-Lox recombination that knocks out genes only after the mouse has developed enough to survive their loss. It requires adding extra DNA: *Lox* sequences flanking the targeted gene plus a *Cre* gene, which can be turned on to produce an enzyme that modifies the DNA between the *Lox* sites. Using CRISPR to insert this same system into zygotes, Jaenisch's team reported making conditional mice with relatively "high efficiency"—about 16% of the zygotes led to mouse pups with the correct mutations.

Skarnes is one of many researchers bowled over by Jaenisch's initial reports, but he was disappointed when he tried to take the technique into his own lab. "It looked from his papers that this was going to be straightforward and I was quite confident this would make ES obsolete," Skarnes says. "What was disappointing is none of us could reproduce at the efficiencies reported by Jaenisch. … It works at 1% or 2% at JAX and a lot of projects are failing. It's really not proven to be a robust method."

There are several reasons why a CRISPR cut more readily leads to a knockout than a knockin. To create knockins with CRISPR, researchers introduce stretches of "donor" DNA—anything from a few bases to an entire gene—designed to integrate at the break points created by Cas9. Splicing in the donor DNA requires that its ends match, or are "homologous" with, the damaged DNA. A process of homologous-directed repair (HDR) then stitches the ends together.

The knockout repair mechanism, which is called nonhomologous end joining, can happen at any stage in the cell division cycle and occurs quickly. HDR, in contrast, mainly happens in one phase of the cell cycle and is far slower. Some genes, such as those Jaenisch selected for his initial knockin experiments, are also more conducive to HDR than others. "The paper reported what we found," Jaenisch says. "Now, we see there are issues."

To others, CRISPR's limitations raise questions about the knockout mouse consortium's decision to abandon ES cell technology. Launched in 2003, the project has created a repository of mutant ES cells, most of them conditionals, for nearly 18,000 genes. Any researcher can order a cell line and spend a year or more making a needed knockout mouse. It has also bred 5011 mutant mouse strains that have germ line transmission of the knockout. This summer, as part of KOMP2, NIH decided to extend the tally of live knockouts to 8000, funding JAX, the University of California, Davis, and Baylor College of Medicine in Houston, Texas, to do the work. But it specified that the knockouts should be made using CRISPR alone.

Colin Fletcher, a mouse geneticist at NIH's National Human Genome Research Institute in Rockville, Maryland, who oversees KOMP2, says advisers endorsed the switch to CRISPR. "You can't cling to the old technology," Fletcher says. "A lot of people have abandoned the ES cell repository and, on the other hand, a lot of people have come in to the field because of the new technology. People are voting with their feet. People are putting much more effort into making conditional alleles with CRISPR rather than making ES cells."

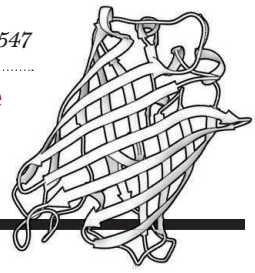
Skarnes, who has just moved from Wellcome to the JAX genomic medicine branch in Farmington, Connecticut, calls the shift premature. But he concedes that researchers will "eventually" figure out how to tweak CRISPR so that it makes conditional mutant mice with high efficiency. One route is to block an enzyme crucial to nonhomologous end joining. Another is to enhance a protein critical to the HDR process that makes knockins possible. Still other investigators have toyed with lengthening the cell-cycle phase that is most favorable to that repair process, zapping zygotes with electric pulses to aid the entry of the CRISPR-Cas9 construct, and creating mutant Cas9s called "nickases" that only break a single DNA strand and preferentially induce HDR.

Whatever CRISPR's shortcomings appear to be at this point, Wiles emphasizes that its potential for engineering mice should not be underestimated. "There is a massive number of things CRISPR can do that people are just beginning to grasp," he says. "We're really at the very, very early phases of development and the tool has infinite possibilities." ■

PHOTO: JENNIFER L. TORRANCE COURTESY OF THE JACKSON LABORATORY

Save the migrants p. 547

Two ways to new side chains p. 553 ▸



PERSPECTIVES

NEUROSCIENCE

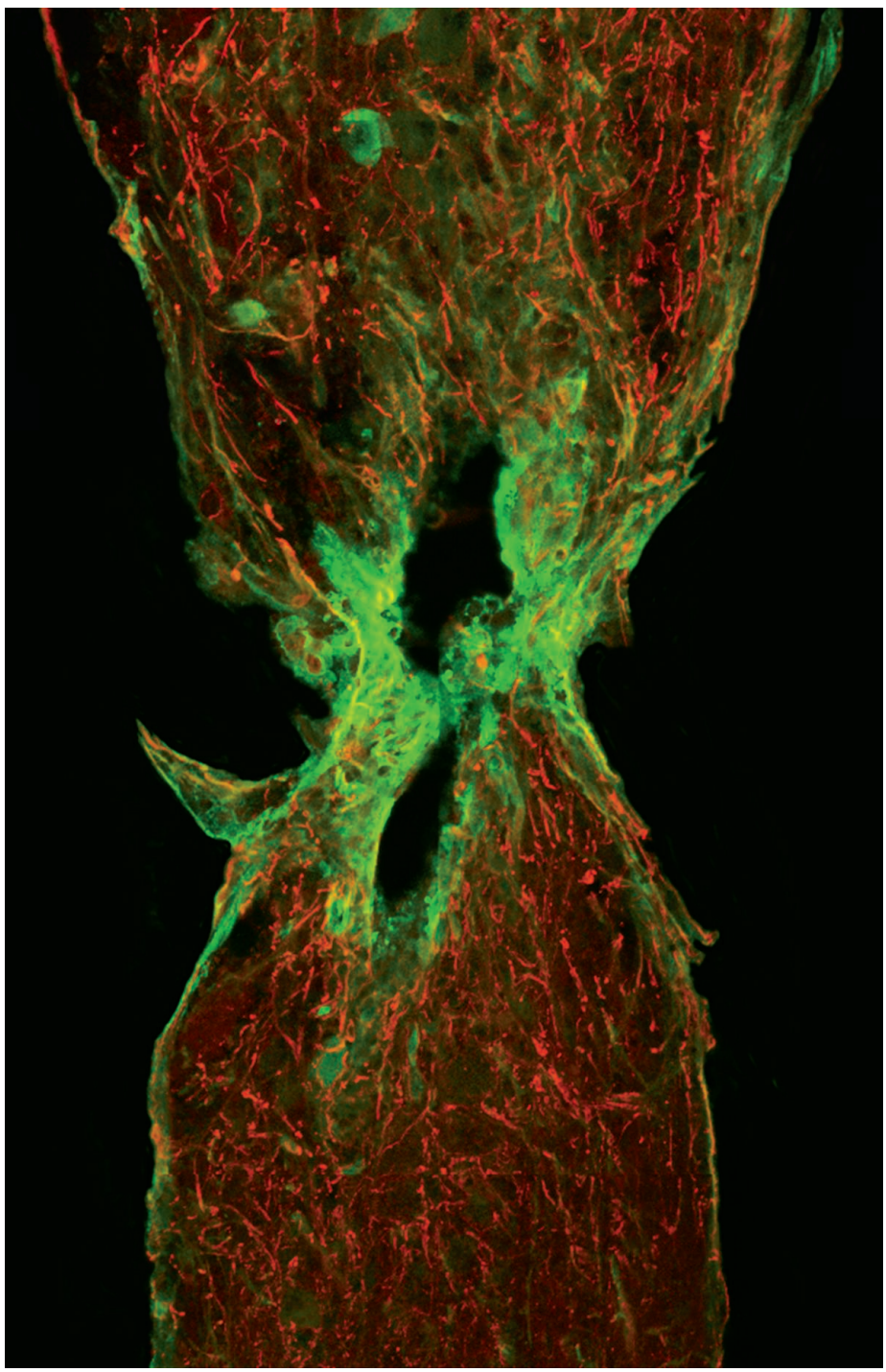
# Building bridges to regenerate axons

## A secreted factor stimulates glia to mend spinal cord injury

*By* **Philip R. Williams** *and* **Zhigang He**

After traumatic spinal cord injury, local damage responses have an important effect on regenerating axons. In lower organisms (i.e., zebrafish and newt), glial cells and other non-neuronal cell types proliferate, migrate, and differentiate to form a bridge between the two ends of a transected spinal cord. This glial bridge supports axon regeneration across the lesion site, enabling functional recovery. However, in mammals, a glial scar composed of mixed cell types and extracellular matrix forms to seal such a wound. Although early studies emphasized the inhibitory nature of this scar, recent studies have revealed that in mammals, as in lower organisms, it can also serve as a bridge to facilitate axon regeneration (*1*–*3*). Yet, how this glial reaction is regulated remains largely unknown. On page 630 of this issue, Mokalled *et al.* (*4*) report that connective tissue growth factor a (CTGFa) is crucial for directing glial bridging and subsequent axon regeneration in a zebrafish model of spinal cord injury. Although it remains to be determined whether this is

*F.M. Kirby Neurobiology Center, Children's Hospital, and Department of Neurology, Harvard Medical School, 300 Longwood Avenue, Boston, MA 02115, USA. Email: zhigang.he@childrens.harvard.edu*



**Tissue section of a zebrafish spinal cord reveals regeneration after a transection injury. Enhanced green fluorescent protein (green) reports expression of CTGFa in tissue that is bridging the two severed ends. Glia are indicated in red using an antibody against glial fibrillary acidic protein. Images were acquired using a Zeiss LSM 700 confocal microscope. Shown is a maximum projection of images that were tiled and processed using ZEN imaging software.**

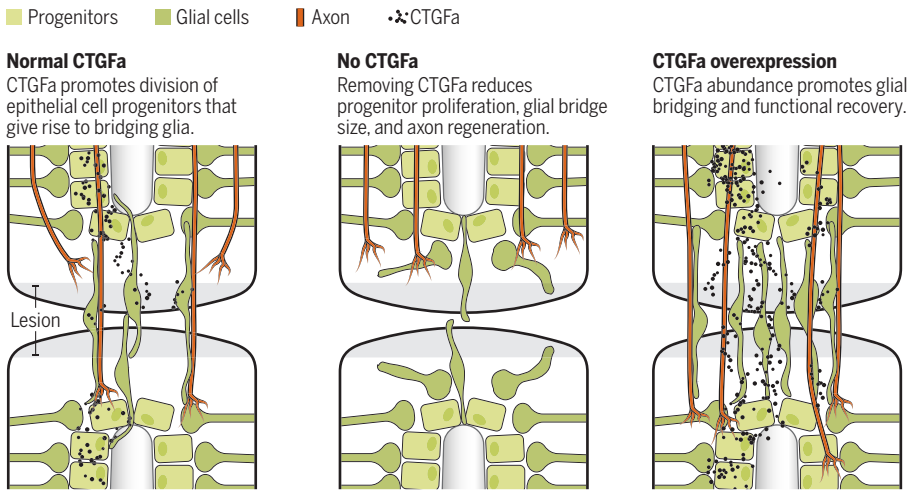
PHOTO: M. H. MOKALLED AND K. D. POSS

the case in mammals, this finding could inform the design of neural repair strategies.

Mokalled *et al.* examined whether the injured spinal cord releases factors that coordinate the glial bridging response. By analyzing messenger RNA expressed in the injured spinal cord of zebrafish, the authors identified a few secreted proteins whose expression transform into bridging cells with a typical elongated morphology (see the photo). In CTGFa mutants, the most obvious defect was the reduced proliferation of glial cells. Consistently, the proliferation-promoting carboxyl-terminal region of CTGFa, but not the profibrotic amino-terminal of the protein, could mimic the effects of full-length CTGFa.

## Forging the glial bridge

In a zebrafish model of spinal cord injury, regeneration and functional recovery requires a secreted factor that stimulates glial cells to form a bridge across the lesion.




in bridging glia and in ependymal cells (the glial cells from which bridging glia develop) was induced at 1 or 2 weeks after injury. One such factor, CTGF, is a cytokine that is conserved across species from zebrafish to humans (*5*). Zebrafish express two isoforms, CTGFa and CTGFb, whereas mammals express only CTGF. This matricellular protein is implicated in multiple processes such as the formation and remodeling of blood vessels, bone, and connective tissue. Notably, in injury conditions, CTGF is a profibrotic factor that stimulates connective tissue formation and wound healing. Zebrafish carrying a *ctgfa* mutation exhibited defects in glial cell proliferation and bridge formation after spinal cord injury. Consequently, axon regeneration and functional recovery were impaired in these mutants. Thus, CTGFa secretion from glia is required for glial bridging and self-repair. In wild-type zebrafish with spinal cord injury, forced expression of CTGFa increased glial bridging, axon regeneration, and functional recovery (see the figure). A soluble form of human CTGF protein applied to the lesion site in zebrafish mimicked the effect of zebrafish CTGFa in promoting glial bridging and axon regeneration.

Several steps are needed to form a glial bridge after spinal cord transection in fish (*6*). A subset of glial cells must proliferate locally, migrate into the lesion, and then Thus, it is possible that the proliferation of pioneering glial cells is the limiting step and CTGFa primarily acts at this point. It is also possible that CTGFa has a direct role in other steps such as the migration of glial cells and their morphological transformation. Because human CTGF interacts with various growth factors and cytokines and modifies other signaling pathways, it is possible that zebrafish CTGFa enhances the function of other glial bridge-promoting factors such as fibroblast growth factor (*6*).

These results also spur speculation as to how observations in regeneration-competent zebrafish could be relevant to mammals. As with zebrafish, expression of CTGF is also induced in reactive astrocytes, invading fibroblasts, and endothelial cells in rats after spinal cord lesion (*7*). Reactive astrocytes in the spinal cord lesion site in mammals, particularly elongated astrocytes (*8–10*), could be permissive for axon regeneration.

With such similarities, why do injured axons fail to regenerate in mammals? An obvious reason is the injured axons' poor intrinsic regenerative ability in the adult mammalian central nervous system. With recent advances in achieving axon regeneration by boosting the intrinsic regenerative ability of injured neurons (*11*), there is a better opportunity to assess which cell types around the lesion are permissive and which are inhibitory. For example, corticospinal axons in the mouse can be induced to grow across a spinal cord injury site when the corticospinal neurons no longer express a gene called phosphatase and tensin homolog (*Pten*). In this scenario, corticospinal axons associate with reactive astrocytes while avoiding fibroblasts and macrophages (*10*). These findings are consistent with the notion that reactive astrocytes are permissive for axon outgrowth, and also indicate that fibroblasts and macrophages are likely inhibitory.

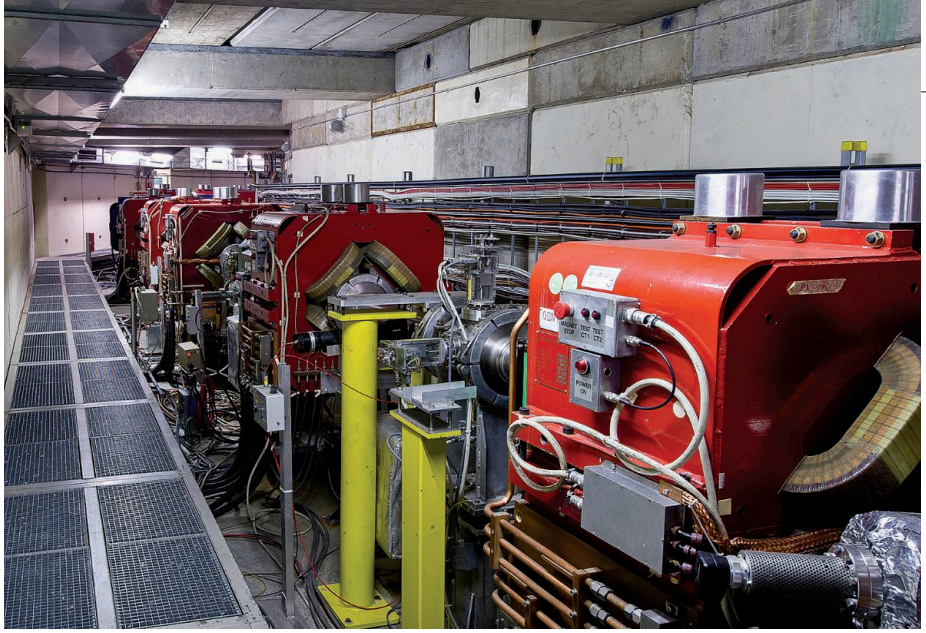
Also unclear is whether all reactive astrocytes are permissive for axon regeneration. It appears that the morphology of astrocytes affects their ability to support regenerating axons. In mammals, regenerating axons often associate with elongated bipolar astrocytes instead of stellate ones (*8–10*). It is unknown whether different astrocytes reflect different lineages or different functional states (immature or mature) (*12–15*). In this regard, a notable difference between zebrafish and mammals is the duration of CTGF expression. CTGFa is transiently expressed in the zebrafish after spinal cord injury (*4*), whereas CTGF is persistently expressed in spinal cord–injured rats (*7*). In fibrotic lesions, persistent CTGF expression can drive the initial wound healing process into a sustained scarring response with a mixed cell composition and extracellular matrix accumulation (*5*). An analogous scar in spinal cord injury could result in a physical and biochemical barrier that blocks axon regeneration. It would be interesting to assess the effects of manipulating the timing of CTGF expression on glial bridge formation and axon regeneration in both zebrafish and mammals. Furthermore, in human patients, inflammatory reactions in the spinal cord lesion site lead to cavity formation surrounded by a dense glial scar. In such a condition, a strategy for promoting glial bridges, such as manipulating CTGF expression and other pathways that control scar formation (*6, 8*), could be useful for promoting axon regeneration and functional recovery in human patients. ■

### REFERENCES


1. J. Silver *et al.*, *Cold Spring Harb. Perspect. Biol.* **7**, a020602 (2014).
2. J. E. Burda, M. V. Sofroniew, *Neuron* **81**, 229 (2014).
3. M. A. Anderson *et al.*, *Nature* **532**, 195 (2016).
4. M. H. Mokalled *et al.*, *Science* **354**, 630 (2016).
5. A. Leask, D. J. Abraham, *Biochem. Cell Biol.* **81**, 355 (2003).
6. Y. Goldshmit *et al.*, *J. Neurosci.* **32**, 7477 (2012).
7. S. Conrad *et al.*, *J. Neurosurg. Spine* **2**, 319 (2005).
8. R. E. White *et al.*, *Exp. Neurol.* **214**, 10 (2008).
9. A. R. Filous *et al.*, *Dev. Neurobiol.* **70**, 826 (2010).
10. K. Zukor *et al.*, *J. Neurosci.* **33**, 15350 (2013).
11. Z. He, Y. Jin, *Neuron* **90**, 437 (2016).
12. J. Silver, *Exp. Neurol.* 10.1016/j.expneurol.2016.06.018 (2016).
13. F. Barnabé-Heider *et al.*, *Cell Stem Cell* **7**, 470 (2010).
14. J. L. Zamanian *et al.*, *J. Neurosci.* **32**, 6391 (2012).
15. O. A. Bayraktar *et al.*, *Cold Spring Harb. Perspect. Biol.* **7**, a020362 (2014).


10.1126/science.aal2112

GRAPHIC: V. ALTOUNIAN/*SCIENCE*



The Antiproton Decelerator at CERN used in the preparation of the antiprotonic helium atoms.

PHYSICS

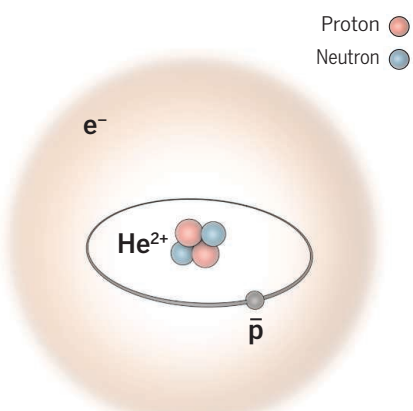
# A testing time for antimatter

## Precision measurement of antiprotonic helium provides a test of physical laws

*By* **Wim Ubachs**

Spectroscopy is the most accurate branch of science. Optical transition frequencies in isolated atoms and molecules can nowadays be measured to many-digit accuracies by applying the tools developed in the atomic physics community: ultrastable lasers, locked via frequency-comb lasers to atomic clocks, and the techniques to cool and control the motion of atoms. Precision measurements on small quantum systems can be compared with theoretical descriptions of these systems at the most fundamental level, allowing physics theories to be tested and enabling the search for physics beyond the standard model (*1*). On page 610 of this issue, Hori *et al.* (*2*) apply these tricks of the trade to a small atomic quantum system with a built-in antiparticle to perform precise spectroscopic measurement in antiprotonic helium (see the figure). The technique of buffer-gas cooling is demonstrated for the first time on a composite matter–antimatter particle.

The laws of physics conspire to create the conditions for producing relatively long-lived states in the exotic antiprotonic helium atom ($e\bar{p}He$). In all other cases of antiprotons stopping in matter, the characteristic annihilation process by which a matter particle and an antimatter particle end their lives in a flash of energy takes place on a time scale of less than a nanosecond. The Heisenberg uncertainty principle then prohibits a precision measurement. The capture of antiprotons in a helium atom, under replacement of an electron, is the exception (*3*). The inclusion of the antiproton produces a $\bar{p}He^+$ ionic core in a heavy-Rydberg two-particle state. Energy conservation dictates that $\bar{p}He^+$ emerges in a quantum state $n$ of typically $n = 38$; in the experiment by Hori *et al.*, states with $n = 31$ to $n = 40$ are probed. The overlap of the wave functions of $\bar{p}$ and $He^{2+}$ is minimal if the high-$n$ Rydberg states exhibit a large angular momentum—for example, in a state such as $n = 38$ with $l = 36$ or 37 (where $l$ denotes the quantized angular momentum). These circular states have typical lifetimes of microseconds, long enough to perform a laser experiment. Narrow spectral lines can be measured to determine level energies to high accuracy. The Rydberg constant of the heavy-Rydberg $\bar{p}He^+$ ion pair is so large that the transitions by which the $n$-quantum number changes by unity lie in the wavelength range of the near-infrared, visible, and ultraviolet, where Hori *et al.* have narrowband and stable lasers at hand.




**Structure of the antiprotonic helium system ($e\bar{p}He$).** The positive $He^{2+}$ and negative $\bar{p}$ ions can be considered to be bound as a heavy-Rydberg ion (*11*). For $n = 38$, the typical state in which the antiproton is captured, this yields a characteristic distance of 0.16 Å. The electron is then bound in the field of the two-particle ionic core of charge 1, similar to the way it is bound in the hydrogen atom at a typical distance of the Bohr radius $a_0 = 0.5$ Å. The quantum-level structure of antiprotonic helium can be calculated to 10-digit precision (*6*, *7*).

The second electron of the helium atom remains in its 1s orbital, and it acts as a spectator during the capture process. It favorably shields the composite newly built electrically neutral $e\bar{p}He$ system during collisions with the outside world. This property makes it possible for the antiprotonic helium atom to survive while it undergoes multiple collisions in the surrounding bath of cold helium atoms at a temperature of 1.5 K above absolute zero. This buffer-gas cooling technique, originally developed for the preparation of cold atoms and molecules (*4*), is applied to the exotic atom bearing an antiparticle in its structure. After a number of collisions, without annihilation, the exotic atoms are equilibrated and take up the temperature of the bath. At 1.5 K, the velocity of the particles is very low, and hence the usual spectral line broadening due to the Doppler effect is suppressed, allowing for the measurement of narrow spectral lines.

The precision frequency measurements performed by Hori *et al.* on 13 energy-level transitions in antiprotonic helium, and obtained with nine-digit precision, can be interpreted in three ways. First, it is a test of charge-parity-time (CPT) invariance, arguably the most fundamental theorem on symmetries in physics. All calculations in particle physics would lose validity if CPT symmetry were not valid (*3*). The exchange of particles by their antiparticles (C), reflection in a mirror (P), and letting time run backwards (T) should not cause a change in the measurable properties of a physical system. The measurements verify that CPT is a valid symmetry, at least at the level of accuracy now obtained.

Second, antiprotonic helium is a three-particle Coulomb system, the stability of which was already investigated by Poincaré in the framework of classical mechanics. In the fully quantized version of electromagnetism, quantum electro-dynamics (QED), the stability of the helium atom (*5*) and the $H_2^+$

*Department of Physics, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, Netherlands. Email: w.m.g.ubachs@vu.nl*

molecular ion (*6*) is well established, and the bound-state level energies can be calculated to 10-digit precision. The same is true for the $e\bar{p}He$ system (*6, 7*) that is found to withstand a test of QED.

Third, the transition frequencies sensitively depend on the mass ratios of the constituent particles. Similarly, as in a spectroscopic determination of the proton-electron mass ratio from $HD^+$ spectroscopy (*8*), evaluation of the data of Hori *et al.* yields a value for the antiproton–electron mass ratio. We learn that antiprotons weigh the same as protons, up to the 10th digit.

The frontier of particle physics is commonly approached in the realm of high-energy physics, using particle accelerators like the Large Hadron Collider (LHC). The detection of the Higgs boson at the LHC marks the culmination of the standard model of physics, but the quest is on to explore new physics beyond that. Alternative approaches exist in the low-energy domain by performing extreme precision measurements on small atomic and molecular systems for which the quantum-level structure is calculable (*1*). This can be done, for example, through the search for an electric dipole moment of the electron in molecules that might reveal signatures of supersymmetry (*9*). Laser precision measurements of optical transitions in molecules constrain the existence of higher dimensions beyond the 3 + 1 (spatial and time) that we regularly observe, or the occurrence of a fifth force beyond the three forces known in the standard model plus gravity (*1*). Previous measurements on antiprotonic helium have already set limits on the strength of such a hypothetical fifth force at sub-ångström length scale (*10*); the present data constrain these phenomena even further.

This exotic atom involving antimatter seems a fortunate accident of nature. It exhibits long-lived (metastable) quantum states that can be probed with lasers, and it survives the collisional conditions needed to cool its kinetic motion, as Hori *et al.* have demonstrated. It is likely that these properties may be further exploited to reveal new physics in future experiments on this extraordinary atom-like particle. ■

## REFERENCES


1. W. Ubachs *et al.*, *J. Mol. Spectrosc.* **320**, 1 (2016).
2. M. Hori *et al.*, *Science* **354**, 610 (2016).
3. R. S. Hayano *et al.*, *Rep. Prog. Phys.* **70**, 1995 (2007).
4. J. D. Weinstein *et al.*, *Nature* **395**, 148 (1998).
5. K. Pachucki, *Phys. Rev. A* **74**, 062510 (2006).
6. V. I. Korobov *et al.*, *Phys. Rev. Lett.* **112**, 103003 (2014).
7. V. I. Korobov *et al.*, *Hyperfine Int.* **233**, 75 (2015).
8. J. Biesheuvel *et al.*, *Nat. Commun.* **7**, 10385 (2016).
9. J. Baron *et al.*, *Science* **343**, 269 (2014).
10. E. J. Salumbides *et al.*, *J. Mol. Spectrosc.* **297**, 41 (2014).
11. E. Reinhold, W. Ubachs, *Mol. Phys.* **103**, 1329 (2005).


10.1126/science.aah6215

CONSERVATION

# Migratory birds under threat

## Habitat degradation and loss, illegal killings, and climate change threaten European migratory bird populations

*By* **Franz Bairlein**

The populations of migratory bird species that breed in Europe and overwinter in sub-Saharan Africa are declining considerably faster than those of non-migratory resident species or of migratory species that overwinter in Europe (*1*). Likely factors are habitat changes due to changes in land use, illegal killing and taking along the northern African coasts, and climate-induced changes in timing of migration and breeding. However, not only European trans-Saharan migrants are declining fast. This holds also for North American long-distance migrants wintering in Central and South America. To halt these declines, preservation of remaining habitats and restoration of habitats both at breeding and nonbreeding grounds is essential, as well as stopping illegal killing and taking of birds along their migration routes.

### ILLEGAL KILLING AND TAKING

Every year, between 11 million and 36 million birds are killed or taken illegally in the Mediterranean region (*2*). The areas of greatest concern are in the eastern and central Mediterranean, with more than 5 million birds taken in both Egypt and Italy and an estimated 1 million each in Cyprus and Lebanon. Common migratory species such as Eurasian chaffinch, blackcap, and song thrush are most affected, but many less common migratory species are also taken in substantial numbers, including species of global conservation concern such as red kite and Eurasian curlew. On average, the annual illegal killings and takings of threatened or near-threatened migratory bird species amount to 1.0 to 3.5% of their populations (*2*)—percentages that are very likely to have considerable impacts on the fate of these species.

Illegal trapping can cause a collapse in population numbers within a short period of time. For example, the yellow-breasted bunting was abundant in its Eurasian breeding range until illegal takings in China caused an 84% population decline between 1980 and 2013 (*3*). Similarly, the passenger pigeon was once the most abundant migratory bird in North America, numbering around 3 billion to 5 billion birds in the early 19th century. Massive-scale hunting as cheap meat resulted in its extinction at the turn of the 20th century (*4*).

*Institute of Avian Research, Wilhelmshaven, Germany.*
*Email: franz.bairlein@ifv-vogelwarte.de*

### HABITAT DEGRADATION AND LOSS

Many European migratory birds that overwinter in sub-Saharan Africa do not travel across those areas of the eastern Mediterranean most affected by illegal killings. Rather, they use a flyway across the western Mediterranean where illegal taking is much less intense (see the figure). Other factors must play a role in their decline.

A large number of European migratory species overwinter in the dry savannas of sub-Saharan Africa. Annual survival of many of these species correlates with rainfall in the Sahel zone (*1*). However, despite an increase in rainfall in the Sahel in recent decades, bird populations have continued to decline. Thus, factors other than rainfall must contribute to the population declines. Land-use and land-cover changes are the most important (*5*). Between 1975 and 2000, agriculture increased by 57% in sub-Saharan Africa at the expense of natural vegetation, with nearly 5 million hectares of forest and nonforest vegetation lost per year (*6*). Most affected are the Sahel and Guinea Savanna zones where the majority of the Eurasian migratory species overwinter. An analysis of breeding-bird survey trends of 26 long-distance migratory species in the United Kingdom shows that wintering habitat is the most important determinant of population trends, with specialist species that occupy either open or woodland habitats in Africa showing declines (*7*).

### CLIMATE CHANGE

Climate change is another major driver for biodiversity changes, including responses of bird populations (*8*). Many migratory species, including those that overwinter in sub-Saharan Africa, now arrive earlier at their spring breeding grounds (*9*). However, different organisms do not respond to climate change at the same pace, which has led to an ecological mismatch between some consumers and their prey (*10*).

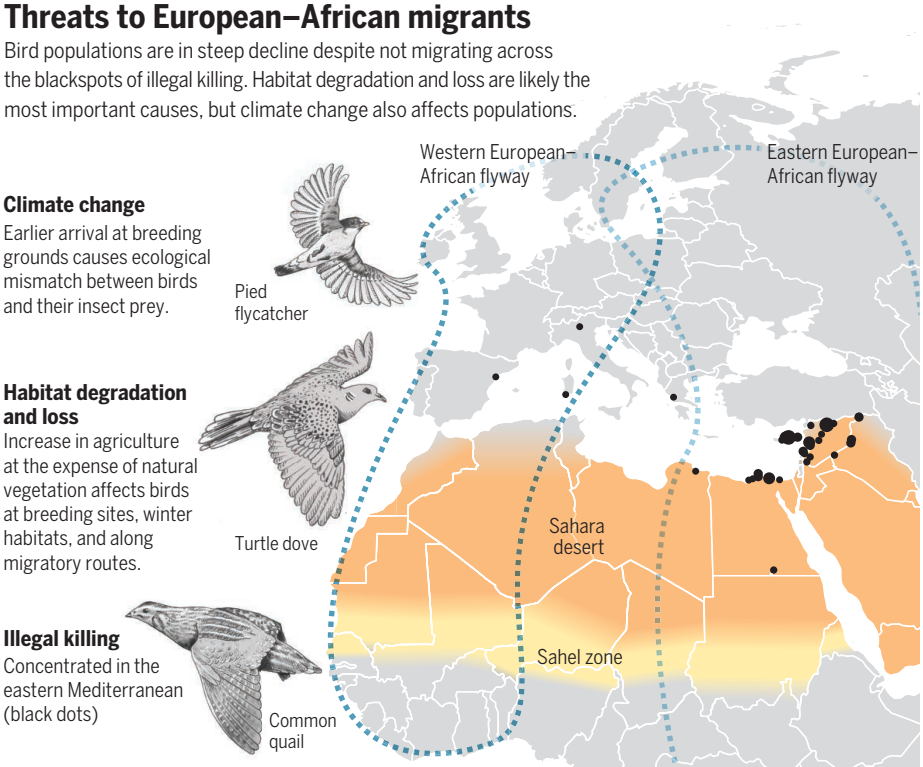
Dutch pied flycatchers, which overwinter in sub-Saharan Africa, do not arrive earlier at breeding grounds, but the populations of their insect food peak earlier as a result of warmer spring temperatures.

This mismatch between breeding and food availability has caused a decline of up to 90% in Dutch pied flycatcher population sizes (*11*).

Similarly, an analysis of 242 time series over the past five decades for 117 European migratory bird species revealed a larger population decline for species with a larger mismatch between food availability and consumer requirement (*12*). Climate change–induced changes in migratory bird populations are thus evident, but their relative contribution to population changes in long-distance migrants wintering requires further clarification.

## Threats to European–African migrants

Bird populations are in steep decline despite not migrating across the blackspots of illegal killing. Habitat degradation and loss are likely the most important causes, but climate change also affects populations.




## A GLOBAL ISSUE

Although I have focused on the European–African migration system, many migratory bird species are declining around the world. In North America, the Breeding Bird Survey—a roadside census throughout the United States and parts of Canada that has been running since 1966—shows that half of the migratory bird species are declining (*13*); declines in long-distance Neotropical migrants are more pronounced than those of birds that migrate short distances. As for the European–African migrants, climate mismatch is a factor, but the long-distance migrants are particularly sensitive to habitat changes (*14*). Similar patterns might be expected for migratory birds in East Asia, but large-scale and long-term breeding bird population surveys, population trends, and nonbreeding habitat assessments are largely missing for that region.

## OUTLOOK

Any attempt to understand and ameliorate migratory bird losses must consider threats far away from their breeding sites. These threats could include killing and taking, human disturbance at staging sites, pesticide exposure, or collisions with human obstacles such as wind turbines and traffic. The most important drivers of population declines in migratory bird species are likely to be land-use changes and connected habitat degradation and loss, but few studies have investigated the specific impacts of land-use and land-cover changes on migratory bird populations in detail (*7*). Such studies are urgently needed to disentangle the various factors acting on populations of migratory species.

These studies must include carry-over effects, because conditions at nonbreeding grounds can affect reproductive success in the breeding season (*15*). Conversely, conditions at breeding grounds can affect nonbreeding fitness. In the red knot, climate-induced malnutrition at their Arctic breeding grounds resulted in shorter bills and reduced survival rates at their African wintering grounds (*16*).

Future studies must also consider the role of stopover sites. Most long-distance migratory species rely on considerable fueling at such stopover sites before they embark on migratory flight, particularly if they need to cross oceans or deserts with no or limited feeding opportunities. The main fuel for migratory flights is fat. Consequently, migrants must accumulate enormous amounts of fat before migratory flights, some doubling their body mass within just a few weeks. To achieve that timed fueling, suitable habitats and food must be available at stopover sites. The effect of habitat loss and degradation at stopover sites on population trends in trans-Saharan migrants remains uninvestigated.

Consequently, future studies also need to clarify where species of conservation concern stop over and overwinter, and which migration routes they take. For example, British common cuckoos migrate along two routes to the same winter destination in sub-Saharan Africa. They face a much higher en route mortality when migrating along a western route than along an easterly route (*17*). Emerging technologies for tracking individual migratory birds throughout their annual cycle (*18*) will reveal migratory routes and destinations in more detail than past bird marking, allowing more detailed and frequent assessment of the drivers of migratory species declines.

Existing data can, however, already be translated into immediate conservation actions to halt the decline of these migratory species. Wetlands can be protected from drainage; woody vegetation can be protected from grazing or even be replanted. Such efforts would not only support migratory birds but also the local biodiversity and livelihoods of local farmers and pastoralists. In addition, illegal taking and killing can be stopped. The required political instruments, such as the African–Eurasian Waterbird Agreement and the African–Eurasian Migratory Landbird Action Plan, are already in place. We just need to act, and we can if we wish. ■

### REFERENCES AND NOTES


1. J. A. Vickery *et al.*, *Ibis* **156**, 1 (2014).
2. A.-L. Brochet *et al.*, *Bird Conserv. Int.* **26**, 1 (2016).
3. J. Kamp *et al.*, *Conserv. Biol.* **29**, 1684 (2015).
4. C.-M. Hung *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **29**, 10636 (2014).
5. B. A. Walther, *Afr. J. Ecol.* **54**, 268 (2016).
6. A. B. Brink, H. D. Eva, *Appl. Geogr.* **29**, 501 (2009).
7. N. Ockendon *et al.*, *Bird Study* **59**, 111 (2012).
8. P. A. Stephens *et al.*, *Science* **352**, 84 (2016).
9. M. Zalakevicius *et al.*, *J. Ornithol.* **147**, 326 (2006).
10. S. J. Thackeray *et al.*, *Global Change Biol.* **16**, 3304 (2010).
11. C. Both *et al.*, *Nature* **441**, 81 (2006).
12. N. Saino *et al.*, *Proc. R. Soc. B* **278**, 835 (2012).
13. USGS Patuxent Wildlife Research Center; www.mbr-pwrc.usgs.gov/cgi-bin/guild12.pl.
14. J. Faaborg *et al.*, *Ecol. Appl.* **20**, 398 (2010).
15. D. R. Norris *et al.*, *Proc. R. Soc. B* **271**, 59 (2003).
16. J. A. van Gils *et al.*, *Science* **352**, 819 (2016).
17. C. M. Hewson *et al.*, *Nat. Commun.* **7**, 12296 (2016).
18. M. Wikelski, G. Tertitski, *Science* **352**, 775 (2016).



### ACKNOWLEDGMENTS

I thank H. Schmaljohann for suggestions.

10.1126/science.aah6647


GRAPHIC: K. SUTLIFF/*SCIENCE*

CANCER

# How tobacco smoke changes the (epi)genome

## Distinct mutation types are found in diverse cancers associated with smoking

*By* **Gerd P. Pfeifer**

Worldwide, more than 1 billion people are tobacco smokers. Cigarette smoking drastically increases the risk of lung cancer. However, many other cancer types also occur more frequently in smokers than in nonsmokers, including cancers of the oral cavity, larynx, pharynx, esophagus, liver, cervix, pancreas, bladder, and kidney. Genome sequencing efforts are beginning to provide more sophisticated clues as to the processes at work that are shaping the mutational landscape of tumors. On page 618 of this issue, Alexandrov *et al.* (*1*) focused specifically on smoking-associated cancers and dissected genetic and epigenetic differences in tumors between smokers and lifetime nonsmokers.

Smoking increases the risk of developing several types of cancer.

The specific molecular origins of most human cancers are unknown. By examining the variety of genetic and epigenetic alterations found in tumors, a process described as molecular archaeology can help to reconstruct the events that have taken place in the past and have led to the evolution of tumor genomes. This field has its beginnings more than two decades ago when mutations in the *TP53* tumor suppressor gene were found to be common but to differ quite substantially in their nature between different tumor types (*2*). Today, whole-genome or -exome sequencing data have generally confirmed tumor type–specific mutations similar to those found earlier in *TP53* (*3*).

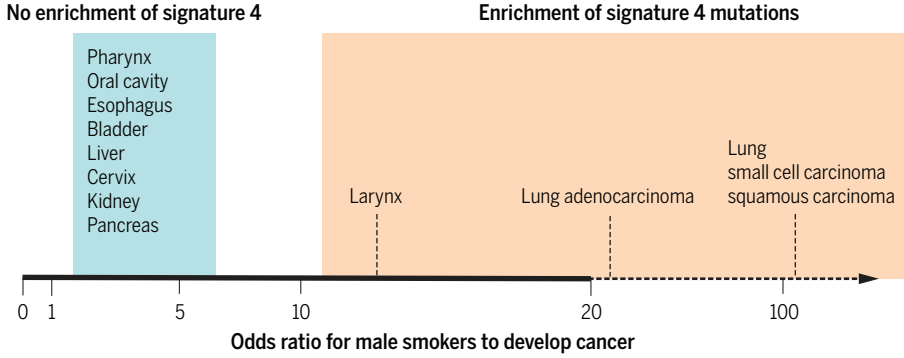
Alexandrov *et al.* examined 13 different cancer types that have been linked to increased risk in smokers and analyzed more than 5000 tumor genomes or exomes. They had previously developed an algorithm based on the six different types of possible base substitutions (C>A, C>G, C>T, T>A, T>C, T>G) and the 16 types of nearest neighboring bases, resulting in 96 specific trinucleotide substitutions (*4*). These substitutions occur in characteristic combinations in different tumor types and were assigned to ~30 so-called mutational signatures. Each type of tumor has one or several such signatures common among patients. Beginning with lung cancer, the authors provided comprehensive comparisons for lung adenocarcinomas in smokers and nonsmokers. This type of lung tumor is also seen in nonsmokers, unlike small cell and squamous cell lung cancers, which almost always develop only in smokers. Smokers' adenocarcinomas stood out by a dramatic increase of "signature 4" mutations, defined as C>A/G>T base substitutions, which were almost absent in nonsmokers. The same signature 4 mutations have previously been implicated in an in vitro experiment with benzo[*a*]pyrene (*5*), a prominent member of the class of polycyclic aromatic hydrocarbons found in the tar fraction of cigarette smoke. In addition, benzo[*a*]pyrene forms strong DNA damage at tobacco-specific mutation hotspots, mechanistically linking polycyclic aromatic hydrocarbons to lung cancer mutations (*6*). Because signature 4 mutations were most prominent in tissues directly exposed to tobacco smoke (the lung and larynx), it is plausible that the signature is derived from DNA damage at guanine, which is then incorrectly bypassed by DNA polymerases, leading to the observed mutations. Signatures 4 and 5 were enriched in clonal mutations (they were found in all cells) of smokers, suggesting that they are due to cigarette smoke exposure before clonal expansion and tumor progression.

Importantly, when the odds ratios for cancer in smokers versus nonsmokers are highest (ranging from over 100 in small cell and squamous cell lung cancers to ~22 in lung adenocarcinomas and ~13 in laryngeal cancers), the contribution of signature 4 to mutations in smokers is the greatest (see the figure). However, as the odds ratio becomes less than 5, such as for oral cavity or esophageal cancers, signature 4 is no longer enriched in smokers. Thus, signature 4 is chiefly responsible for an increased mutational load and for the most strongly enhanced cancer risk in smokers.

Another mutational signature found enriched in several tobacco-associated cancers is signature 5, which is less distinct than signature 4 and contains several types of base substitutions (*1*, *4*). The molecular origin of this signature remains unknown. Although not experimentally proven, signature 4 might also be produced by aromatic amines, another class of chemicals found in tobacco smoke that has been linked to bladder cancer.

### More reasons to quit

Enrichment of a specific mutational signature in cancers of smokers is strongly linked to the odds ratio of developing cancer.




*Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI, USA. Email: gerd.pfeifer@vai.org*

However, the absence of signature 4 in bladder cancer might suggest that aromatic amines are not directly involved in this process or that the mutations caused were too few to be detected. Instead, Alexandrov *et al.* see a strong enrichment of signatures 2 and 13 in bladder tumors, which are characterized by C to T or C to G mutations at TpC dinucleotides. These signatures have been linked to overactivity of the APOBEC (apolipoprotein B editing catalytic polypeptide) DNA editing enzymes (*7*). Signatures 2 and 13 are also present in lung adenocarcinomas, where they seem to be enhanced by smoking. In fact, in 4 of 50 lung adenocarcinomas of both smokers and nonsmokers, signatures 2 and 13 accounted for more than 70% of the mutations (*1*), raising questions about a possible unique etiology or biology of these individual tumors.

Besides mutations, tumor DNA also retains information regarding the modified base 5-methylcytosine at CpG sequences. Alexandrov *et al.* compared CpG methylation data for smokers and nonsmokers but did not find substantial differences in methylation (*1*). This finding is inconsistent with the common belief that environmental exposures, including smoking, have a major effect on the DNA methylation machinery. However, for lung adenocarcinoma, they did find that a small fraction (<0.1%) of the queried CpG sites were differentially methylated in smokers with a tendency for hypomethylation. It will be interesting to see whether these hypomethylated CpG sites coincide with sites of intense DNA damage by tobacco carcinogens.

The new study presents the most comprehensive view to date on smoking-related genetic and epigenetic differences in human tumors. It provides strong support for a prominent role of tobacco carcinogens in cancer of tissues directly exposed to smoke but raises intriguing questions regarding why more "remote" tissues show an enhanced cancer risk in smokers in the absence of a well-defined mutational specificity. Although smaller-scale mutational events are still possible, more indirect effects of smoking should be considered. These might include other epigenetic events, immune system-related events such as inflammation, and/or metabolic disturbances caused by tobacco smoking. ■


**REFERENCES**

1. L. B. Alexandrov *et al.*, *Science* **354**, 618 (2016).
2. M. Hollstein, D. Sidransky, B. Vogelstein, C. C. Harris, *Science* **253**, 49 (1991).
3. P. Hainaut, G. P. Pfeifer, *Cold Spring Harb. Perspect. Med.* 10.1101/cshperspect.a026179 (2016).
4. L. B. Alexandrov *et al.*, *Cell Rep.* **3**, 246 (2013).
5. S. Nik-Zainal *et al.*, *Mutagenesis* **30**, 763 (2015).
6. M. F. Denissenko, A. Pao, M. Tang, G. P. Pfeifer, *Science* **274**, 430 (1996).
7. S. Nik-Zainal *et al.*, *Cell* **149**, 979 (2012).


10.1126/science.aal2114

PHOTO: COURTESY OF LOPHELIA II 2012 EXPEDITION, NOAA-OER/BOEM

GEOLOGY

# *Deep-sea corals feel the flow*

## Uranium isotope data from deep-water corals help to resolve conflicting results from shallow-water corals

*By* **Yusuke Yokoyama**[1] *and* **Tezer M. Esat**[2]

Shallow-water corals provide the only direct way of determining the absolute timing and magnitude of sea levels over the past 600,000 years. Their uranium and thorium abundances and uranium isotope ratios, combined with coral-reef elevations, help to determine past sea levels (*1*). The veracity of the dating results depends on the state of preservation of fossil corals over extended time periods. Models have been developed to correct for specific complications. Deep-sea corals (see the photo) do not provide sea-level information. However, as Chen *et al.* (*2*) show on page 626 of this issue, they appear to be responsive to ocean circulation changes and continental riverine and ice-sheet meltwater inputs to the oceans. In addition, the data help to resolve one of the most contentious issues in uranium series dating, namely whether systematic variations in $^{234}U$ are due to physical ocean processes or are artifacts of poor preservation.

Corals take up uranium from the oceans, where it exists as a trace element, and concentrate it by a factor of 1000. The residence time of uranium in the oceans is over half a million years. In its source rocks, $^{238}U$ decays via the short-lived $^{234}Th$ to $^{234}U$. This radioactive decay process damages the mineral locally, enabling eroding water to preferentially leach $^{234}U$. Furthermore, $^{234}U$ can recoil out into pore spaces, further assisting its preferential leaching. Thus, rivers carry large excesses of $^{234}U$, and the oceans are enriched by 15%. Uranium is highly soluble whereas thorium tends to amalgamate with insoluble particles; therefore, corals take up uranium but almost no thorium. The further decay of $^{234}U$ to $^{230}Th$ provides an age for the coral, and the ocean

[1]*Atmosphere and Ocean Research Institute and Department of Earth and Planetary Sciences, School of Science, The University of Tokyo, Chiba, Japan.* [2]*Research School of Earth Sciences, Research School of Physics and Engineering, The Australian National University, Canberra, ACT 0200, Australia. Email: yokoyama@aori.u-tokyo.ac.jp; tezer.esat@anu.edu.au*

This deep-sea reef is formed by the coral species *Lophelia pertusa* at 450 m below the surface of the Gulf of Mexico, with a brisingid starfish in the center. As Chen *et al.* show, data from deep-sea corals can help to explain past variations in ocean uranium levels, thus putting uranium dating of shallow-water corals on firmer footing.

$^{234}$U levels can be back-calculated to when the coral was alive. For this dating method to work, however, a key requirement is that the sample remain closed to external addition or removal of uranium or thorium; it does not work if the system is "open" due to exchange of U and Th between samples and their surrounding environments.

Historically, efforts to calculate coral ages with the uranium-thorium method required the back-calculated $^{234}$U/$^{238}$U ratio to equal the present oceanic value; deviations from this value were interpreted as an indication of external influences and taken to mean that the age was unreliable (*3*). As data accumulated, scientists focused on the last interglacial, ~130,000 years ago, to determine the reliability of the dating method. Coral samples from the last interglacial are easy to collect as they are above the present sea level. The data showed that the last interglacial extended from ~130,000 to 120,000 years ago. However, a substantial number of corals from the last interglacial had higher $^{234}$U/$^{238}$U values than present-day corals (*3*–*5*); the $^{234}$U/$^{238}$U values appeared to increase almost linearly with age. To explain these observations, Thomson and Goldstein proposed that circulating fluids containing $^{234}$Th and $^{230}$Th had contaminated some of the samples (*6*). The age-$^{234}$U trend was used to correct for high $^{234}$U/$^{238}$U by projecting it to the current $^{234}$U value and to younger ages and was called an open-system age.

There are good reasons for opposing this model. However, most scientists now accept that during the last ice age, the $^{234}$U/$^{238}$U ratio of the ocean was lower than it is today and that it gradually increased to the present value over the past 20,000 years (*7, 8*). Applying the model to the most precise single data set yielded a start date for the last interglacial at 125,000 years ago (*6*), rather than 130,000 years ago as previously estimated. This discrepancy put the utility of uranium-series dating in doubt. Is the timeline of climate and sea-level variations over the past 600,000 years derived from this dating reliable?

An alternative explanation for high $^{234}$U is that there were transient inputs of additional $^{234}$U to the oceans at times of rapid climate and sea-level changes, but this explanation is not generally accepted (*9*). One of the main reasons for this lack of acceptance is the difficulty of intuitively understanding the trends implied by the uranium-thorium equations. For example, it is generally held that older coral reefs have high current $^{234}$U contents, yet exactly the opposite is true: They have low $^{234}$U/$^{238}$U ratios. The high ratios are the back-calculated ocean values at the time when the coral was growing.

This is why Chen *et al.*'s study of deep-sea corals is important. The authors use the solitary corals as time markers of past ocean $^{234}$U levels at depths from 400 to 2000 m. First, they confirm the low $^{234}$U values during the last ice age. Second, their data show a transient peak in $^{234}$U at the end of one of the periodic stop/start episodes of the Atlantic overturning circulation at 15,000 years ago. One could argue that the excess $^{234}$U is due to open-system behavior, entailing random external addition of $^{234}$Th (which rapidly decays to $^{234}$U) and $^{230}$Th. However, the systematics of the data and the quality of the samples do not support this conclusion.

In addition, the samples used by Chen *et al.* are from solitary corals that grow individually and are not embedded in a reef. It is highly unlikely that they could be affected by circulating fluids enriched with Th isotopes. There is simply not enough Th in the water column to cause the observed effects, even over the course of 100,000 years. Therefore, the transient high-$^{234}$U levels are intrinsic to hydrospheric processes and not due to postmortem alteration of coral U/Th trace-element abundances. The excess $^{234}$U was circulated from the surface to the deeper ocean following the resumption of the Atlantic current. It probably included nutrients, promoting a spurt in deep-sea coral growth.

The present demonstration that transient ocean $^{234}$U excesses could have caused the observed high coral $^{234}$U values casts doubt on the veracity of the open-system model. If the ocean sustained transient high-$^{234}$U episodes during the last deglaciation, the same is likely to have happened during the penultimate deglaciation. A recent analysis of diverse evidence for the timing of the last interglacial favors the 130,000 year start date over the open-system estimate of 125,000 years (*10*).

Chen *et al.* postulate that the source of the excess $^{234}$U was the subglacial basal melt, which was in contact with frozen sediments laden with excess $^{234}$U (*11*). This is a likely scenario, but other evidence points to episodic sea-level rise as a probable mechanism. A large fraction (>80%) of uranium supplied by rivers is sequestered in nearshore sediments, salt marshes and mangroves, estuaries, and continental margins during periods of warm climate and high sea levels. These nearshore areas are exposed during periods of low sea level, resulting in rapid oxidation of uranium into highly soluble phases. The subsequent release of $^{234}$U-enriched uranium into the oceans occurs over a sustained period as sea levels rise and can be taken up by corals growing close to coastal areas before the excess uranium is dispersed into the wider ocean.

There is abundant evidence for a connection between $^{234}$U and sea level at locations far from polar ice sheets, including at Huon Peninsula, Papua New Guinea (*9*). In contrast, it is surprising that the low-latitude Pacific deep-sea corals do not seem to have been as affected given their proximity to Antarctica (*12*). Future studies will better delineate the sources and scope of external transient inputs of $^{234}$U into the oceans as indicators of meltwater pulses and rapid sea-level changes, as opposed to stochastic open-system artifacts in coral. ■

## REFERENCES


1. R. L. Edwards, J. H. Chen, G. J. Wasserburg, *Earth. Planet. Sci. Lett.* **81**, 175 (1987).
2. T. Chen *et al.*, *Science* **354**, 626 (2016).
3. C. H. Stirling *et al.*, *Earth. Planet. Sci. Lett.* **160**, 745 (1998).
4. C. D. Gallup *et al.*, *Science* **295**, 310 (2002).
5. H. Cheng *et al.*, *Science* **326**, 248 (2009).
6. W. G. Thompson, S. L. Goldstein, *Science* **308**, 401 (2005).
7. T. M. Esat, Y. Yokoyama, *Geochim. Cosmochim. Acta* **70**, 4140 (2006).
8. E. Bard *et al.*, *Geochim. Cosmochim. Acta* **55**, 2285 (1991).
9. T. M. Esat, Y. Yokoyama, *Geochim. Cosmochim. Acta* **74**, 7008 (2010).
10. G. Marino *et al.*, *Nature* **522**, 197 (2015).
11. G. Henderson *et al.*, *Chem. Geol.* **226**, 298 (2006).
12. Y. Yokoyama *et al.*, *Proc. Nat. Acad. Sci. U.S.A.* **113**, 2354 (2016).


10.1126/science.aak9817

BIOTECHNOLOGY

# Tips for battling billion-dollar beetles

## Tiny insecticidal proteins (Tips) kill pests that evolved resistance to engineered corn

*By* **Bruce E. Tabashnik**

In the never-ending war on insect pests, the widespread soil bacterium *Bacillus thuringiensis* (*Bt*) is one of the greatest heroes. Insecticidal crystalline (Cry) proteins and vegetative insecticidal proteins (Vips) from *Bt* are treasured for their effectiveness against some devastating pests and their safety for beneficial insects, wildlife, and people (*1*). Sprays containing *Bt* proteins have been used for more than 70 years and remain valuable in organic and conventional agriculture, forestry, and vector control (*1*). Crops genetically engineered to produce *Bt* proteins were introduced 20 years ago and quickly became a cornerstone of pest management. They have suppressed pest populations, reduced reliance on insecticide sprays, enhanced control by natural enemies, and increased farmer profits (*2*–*4*). In 2015, *Bt* crops were planted on 84 million hectares globally (*5*), including 81% of the corn and 84% of the cotton in the United States (*6*). But with increasingly rapid evolution of pest resistance to *Bt* crops (*7*, *8*) (see the figure), alternatives are urgently needed. On page 634 of this issue, Schellenberger *et al.* (*9*) report the discovery of tiny insecticidal proteins (Tips) from other soil bacteria that could be part of the solution.

The Tips found by Schellenberger *et al.* kill one of the most destructive and adaptable crop pests in North America and Europe, a beetle called the western corn rootworm (*Diabrotica virgifera virgifera*). The larvae of this pernicious pest and closely related species devour corn roots, costing growers more than $1 billion each year (*10*). The advance by Schellenberger *et al.* is timely because evolution of resistance has undermined control of rootworms by *Bt* corn (*11*–*15*).

The first transgenic corn targeting rootworms produces *Bt* toxin Cry3Bb and was commercialized in the United States in 2003. Field-evolved rootworm resistance to Cry3Bb was initially detected in Iowa in 2009 and subsequently in several other midwestern states (*11*, *12*). Four *Bt* toxins active against rootworms are available now in *Bt* corn: Cry3Bb, mCry3A, eCry3.1Ab, and Cry34/35Ab. However, resistance to Cry3Bb causes cross-resistance to its close relatives mCry3A and eCry3.1Ab (*13*, *15*), and resistance to Cry34/35Ab was detected in Iowa in 2013 (*14*). Factors hastening evolution of rootworm resistance to *Bt* corn include the moderate efficacy of the toxins and growing corn that produces the same *Bt* toxin year after year (*11*, *12*).

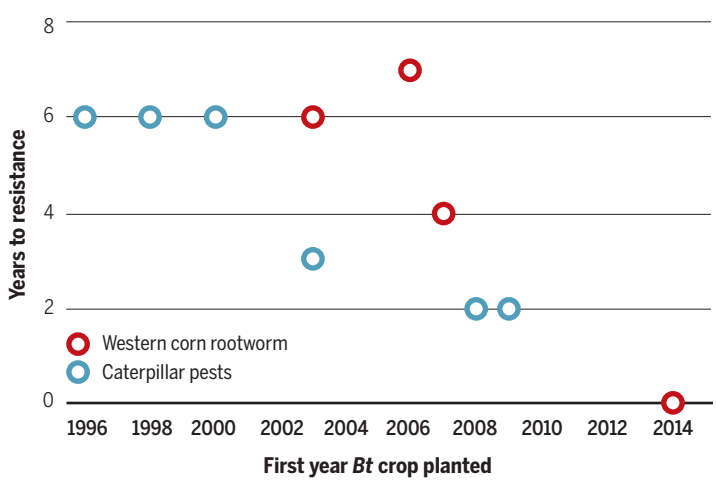
The U.S. Environmental Protection Agency requires farmers to plant corn that does not produce *Bt* toxins near *Bt* corn to promote survival of susceptible insects and thereby delay evolution of resistance (*11*, *12*). However, the current regulations have been challenged as insufficient to achieve this goal because they allow up to 95% *Bt* corn on each farm (*11*). Moreover, adaptation to *Bt* corn plants making either a Cry3 toxin or Cry34/35Ab will accelerate evolution of resistance to recently introduced *Bt* corn plants that produce a Cry3 toxin in combination with Cry34/35Ab (*8*). Therefore, controls effective against rootworms resistant to the currently available *Bt* toxins are especially significant.

The proteins identified by Schellenberger *et al.* may fit the bill. By systematically screening soil for microbes other than *Bt*, they discovered that the bacterium *Pseudomonas chlororaphis* has a potent inhibitory effect on rootworms. They isolated an insecticidal protein from *P. chlororaphis* and determined its amino acid sequence. They named it IPD072Aa and matched it to a protein of 86 amino acids predicted from a gene in the *P. chlororaphis* genome. From various databases, they identified 18 proteins with 38 to 83% sequence identity to IPD072Aa: four from *P. chlororaphis*, eight from six other species of *Pseudomonas*, three from the switchgrass rhizosphere microbial community metagenome, and one each from the bacteria *Burkholderia*, *Xenorhabdus*, and *Photorhabdus*. Four of these homologs were toxic when fed to corn rootworms in artificial diet. IPD072Aa and its homologs are dubbed here as Tips because they are much smaller than previously described insecticidal proteins from bacteria.

Schellenberger *et al.* genetically engineered corn plants to make IPD072Aa. In greenhouse and field trials, the protection from rootworm injury provided by these transgenic plants was similar to that from *Bt* corn producing Cry34/35Ab. In a field trial in Iowa during 2014, rootworm injury to corn was much higher for plants producing Cry3Bb than for those making IPD072Aa. The extensive damage to corn producing Cry3Bb implies rootworm resistance to that toxin. The efficacy of IPD072Aa in this field trial suggests its utility against Cry3Bb-resistant rootworms. Laboratory bioassays show that IPD072Aa killed laboratory-selected strains of rootworm resistant to mCry3A or Cry34/35Ab.

The absence of strong cross-resistance to IPD072Aa caused by resistance to Cry34/35Ab or Cry3 toxins is not surprising because Tips are strikingly different from *Bt* toxins. Indeed, Schellenberger *et al.* report that none of the Tips examined had any motifs, domains, or signatures known from other proteins. It will be important to determine how these

**Pests evolving resistance faster to engineered *Bt* crops.** In the 10 cases of field-evolved practical resistance to *Bt* crops in five pest species, the time from the first commercial planting of the *Bt* crop to the first evidence of resistance decreased over the past two decades ($r^2$ = 0.60, df = 8, $P < 0.01$) (*7*, *8*, *11*–*15*). Western corn rootworm resistance to the four rootworm-active *Bt* proteins made by transgenic corn includes cross-resistance to eCry3.1Ab detected in the field before plants producing this toxin were commercialized.




GRAPHIC: N. CARY/*SCIENCE*

*Department of Entomology, University of Arizona, Tucson, AZ 85721, USA. Email: brucet@cals.arizona.edu*

intriguingly distinct proteins kill rootworms.

The billion-dollar question remains: Will Tips be a valuable alternative to *Bt* toxins for deployment in transgenic plants? Factors that might impede application of bacterial insecticidal proteins other than *Bt* toxins include their large size, lack of oral toxicity to pests, and broad toxicity to nontarget organisms. Tips are small and orally toxic to two species of corn rootworm (*9*). Bioassays with five caterpillar pests and one piercing-sucking pest indicate that they are not broadly toxic to insects (*9*). To better understand their toxicity

Rootworm larvae can devour corn roots.

spectrum, testing of more species is needed, including vertebrates and beneficial beetles such as the "ladybirds" famed for their role in biological pest control.

If Tips or other new control tactics such as RNA interference are widely adopted, rootworms will surely evolve resistance, just as they have to *Bt* corn and the time-honored tactic of crop rotation (*10*–*15*). In light of the remarkable adaptability of this and other insect pests, the increased efforts to find new natural insecticides spurred by the discovery of Tips could pay great dividends. ■

## REFERENCES AND NOTES


1. V. Sanchis, *Agron. Sustain. Dev.* **31**, 271 (2011).
2. W. D. Hutchison *et al.*, *Science* **329**, 222 (2010).
3. B. E. Tabashnik *et al.*, *Nat. Biotechnol.* **28**, 1304 (2010).
4. National Academies of Sciences, Engineering, and Medicine, *Genetically Engineered Crops: Experiences and Prospects* (National Academies Press, 2016).
5. C. James, *20th Anniversary (1996 to 2015) of the Global Commercialization of Biotech Crops and Biotech Crop Highlights in 2015* (ISAAA Brief No. 51, International Service for the Acquisition of Agri-Biotech Applications, 2015).
6. USDA Economic Research Service, *Adoption of Genetically Engineered Crops in the U.S.; www.ers.usda.gov/data-products/adoption-of-genetically-engineered-crops-in-the-us.aspx*.
7. B. E. Tabashnik *et al.*, *Nat. Biotechnol.* **31**, 510 (2013).
8. Y. Carrière *et al.*, *Trends Biotechnol.* **34**, 291 (2016).
9. U. Schellenberger *et al.*, *Science* **354**, 634 (2016).
10. M. Gray *et al.*, *Annu. Rev. Entomol.* **54**, 303 (2009).
11. B. E. Tabashnik, F. Gould, *J. Econ. Entomol.* **105**, 767 (2012).
12. D. Andow *et al.*, *J. Econ. Entomol.* **109**, 1 (2016).
13. S. N. Zukoff *et al.*, *J. Econ. Entomol.* **109**, 1387 (2016).
14. A. J. Gassmann *et al.*, *J. Econ. Entomol.* **109**, 1872 (2016).
15. S. R. K. Jakka *et al.*, *Sci. Rep.* **6**, 27860 (2016).


## ACKNOWLEDGMENTS


I thank R. L. Smith for comments and the U.S. Department of Agriculture National Institute of Food and Agriculture for support.

10.1126/science.aag101


PHOTO: MARLIN E. RICE/PLANT HEALTH PROGRESS

CHEMICAL BIOLOGY

# *A radical approach to posttranslational mutagenesis*

## Introducing diverse protein side chains via carbon-carbon bond-forming reactions

*By* **Raphael Hofmann** *and* **Jeffrey W. Bode**

The structure and function of proteins is extensively modulated and expanded by posttranslational modifications (PTMs) on many of the canonical amino acids. As a means of unraveling the role and interplay of PTMs, methods to produce proteins with site-specific modifications have attracted considerable attention. In addition to PTMs, incorporation of unnatural side chains is of interest for protein engineering, and systematic studies using analogs of amino acids offer insights into the molecular mechanisms by which proteins and enzymes work. On pages 597 and 623 of this issue, Wright *et al.* (*1*) and Yang *et al.* (*2*) report potentially general strategies to chemically introduce a wide variety of natural, unnatural, posttranslationally modified, and labeled side chains via an unprecedented carbon-carbon bond-forming reaction on intact proteins. This approach will be of particular interest to chemical biologists aiming to introduce authentic protein PTMs, as well as to protein chemists interested in introducing unnatural side chains of their choice.

Genetic code expansion approaches such as amber-codon suppression (*3*) have lifted the narrow constraints imposed by the genetic code and routinely enable the recombinant production of proteins containing certain PTMs or unnatural amino acids. However, recombinant incorporation of unnatural amino acids relies on the availability of, and efficient processing by, an orthogonal transfer RNA (tRNA) and tRNA synthetase pair. Protein synthesis (*4*, *5*) or semisynthesis (*6*) allows for full flexibility with regard to unnatural residues but can pose a challenging task for large protein targets. Alternatively, site-specific chemical modifications of expressed proteins at cysteine (Cys) or dehydroalanine (Dha) residues—which serve as potent nucleophiles and electrophiles, respectively—represent versatile and relatively simple strategies and have enabled the generation of bioconjugates or PTM mimics (*7*, *8*). However, such "chemical mutagenesis" strategies have so far resulted in products containing an unnatural thioether bond in the side chain.

Wright *et al.* and Yang *et al.* now describe a three-step chemical mutagenesis strategy (see the figure). First, the site of modification in the expressed protein undergoes "genetic tagging." This tagged residue is converted to the common Dha reactive precursor. The Dha residue is allowed to react with alkyl carbon free radicals, generated from alkyl halides, to introduce the desired side chain via a carbon-carbon single bond. In this regard, the authors have leveraged the analogous alkyl halides like those used for Cys alkylation strategies but obtain products containing an all-carbon backbone instead of the thioether linkage.

For tagging and activation of the modification site, Wright *et al.* rely on Cys mutations and their posttranslational conversion to Dha using bis-alkylation with 2,5-dibromohexane diacetamide, followed by elimination in mildly basic buffer (pH 8.0) (*8*). Yang *et al.* use their previously established system for recombinant incorporation of *O*-phosphoserine (Sep) (*9*) and a base-promoted elimination reaction (40 mM barium hydroxide) to convert Sep into Dha. The existence of several other methods to obtain Dha-containing proteins (*10*–*12*) increases the chance of finding a procedure compatible with a given protein target.

Wright *et al.* report, in detailed model reactions with small molecules, peptides, and proteins, the successful and chemoselective zinc ($Zn^0$)–mediated radical reaction between Dha and alkyl halides in aqueous buffer. The occurrence of undesired side reactions—including substitution of the Cα capto-dative radical with a second equivalent of the alkyl radical, as well as oxidative protein cleavage with molecular triplet oxygen—prompted further optimization. The best results were obtained by using sodium borohydride ($NaBH_4$) instead of $Zn^0$, along with oxygen-free conditions (glovebox), which resulted in selective bond formation in many common buffers (pH 4 to 8) and suppressed side reac-


*Laboratory of Organic Chemistry, Department of Chemistry and Applied Biosciences, ETH Zürich, Zürich 8093, Switzerland. Email: bode@org.chem.ethz.ch*

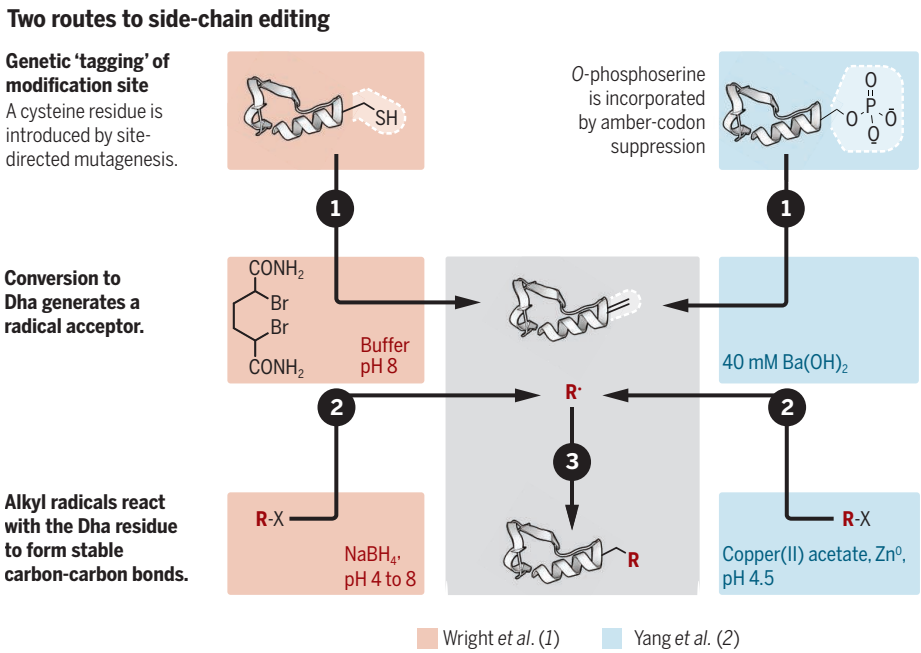
tions. Yang *et al.* found that a radical reaction promoted by $Zn^0$ and copper(II) acetate in the presence of surfactants and metal ligands (tetramethylethylenediamine) in aqueous acidic conditions (sodium acetate, pH 4.5) gave the best conversion. Both procedures call for up to 3000 equivalents of alkyl halide for good conversion, which in some cases is limited by their solubility in water.

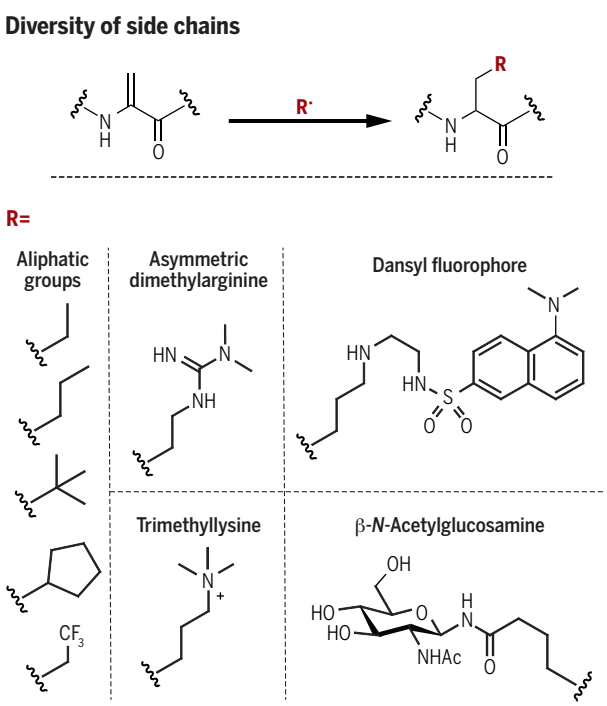
The reaction tolerates primary, secondary, tertiary, and cyclic aliphatic halides, as well as alkyl halides bearing unprotected amines, phosphonates, guanidines, amides, or hydroxyl groups. The choice of the appropriate organic halide, typically readily available either commercially or synthetically, allows for systematic variation of side chains. The value of the system is demonstrated by the introduction of several side chains previously unattainable with codon suppression techniques. Yang *et al.* report the incorporation of mono-, di-, and trimethylated lysine at position 79 of histone H3 and its assembly into chromatin templates. This feat allowed the authors to study the effects of the various methylation states on in vitro transcription. Incorporation of asymmetric dimethylarginine at position 26 of histone H3 and its assembly into mononucleosomes by Wright *et al.* allowed the identification of partner proteins through affinity enrichment proteomics in human cell extracts.

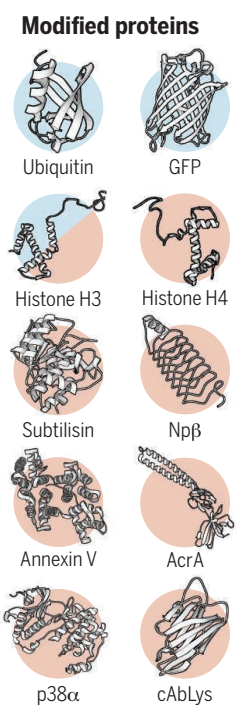
Combined, the authors achieve site-specific modification of 10 structurally diverse proteins with a multitude of different side chains. Wright *et al.* also demonstrate that their conditions are compatible with a disulfide-containing single-domain antibody fragment. It remains to be seen what other complex protein targets are compatible with these reaction conditions. For example, many research groups will be curious about the modification of antibodies, which contain easily reducible interchain disulfide bonds (*13*). The generation of diastereomeric products at the α-center of the modified residues might limit its use for the production of pharmaceutical bioconjugates. For the moment, the method is restricted to a single site of modification within a protein. Regardless, this novel carbon-carbon bond-forming reaction could be used readily in chemical biology, given its versatility with regard to both protein substrates and side chains while retaining the flexibility offered by other chemical mutagenesis approaches. By demonstrating the chemoselectivity of protein radical reactions, this work may further provide fruitful ground for the development of other radical-based protein modifications. ■

## Introducing unnatural side chains

Cysteine or *O*-phosphoserine residues at a particular protein site are chemically converted to dehydroalanine (Dha), which can react with carbon-centered radicals generated from the corresponding halides R-X to introduce new side chains.




### REFERENCES


1. T. H. Wright *et al.*, *Science* **354**, aag1465 (2016).
2. A. Yang *et al.*, *Science* **354**, 623 (2016).
3. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **244**, 182 (1989).
4. P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **266**, 776 (1994).
5. T. J. Harmand, C. E. Murar, J. W. Bode, *Nat. Protoc.* **11**, 1130 (2016).
6. T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6705 (1998).
7. M. D. Simon *et al.*, *Cell* **128**, 1003 (2007).
8. J. M. Chalker *et al.*, *Chem. Sci.* **2**, 1666 (2011).
9. S. Lee *et al.*, *Angew. Chem. Int. Ed.* **52**, 5771 (2013).
10. J. Wang, S. M. Schiller, P. G. Schultz, *Angew. Chem. Int. Ed.* **46**, 6849 (2007).
11. F. P. Seebeck, J. W. Szostak, *J. Am. Chem. Soc.* **128**, 7150 (2006).
12. Y. O. You, M. R. Levengood, L. A. F. Ihnken, A. K. Knowlton, W. A. van der Donk, *ACS Chem. Biol.* **4**, 379 (2009).
13. J. R. Junutula *et al.*, *Nat. Biotechnol.* **26**, 925 (2008).


10.1126/science.aai8788

POLICY FORUM

WATER

# Water strategies for the next administration

## Water policy offers opportunity for nonpartisan agreement

*By* **Peter H. Gleick**

Issues around fresh water are not particularly high on the U.S. political agenda. They should be. Water problems directly threaten food production, fisheries, energy generation, foreign policy, public health, and international security. Access to safe, sufficient, and affordable water is vital to well-being and to the economy. Yet U.S. water systems, once the envy of the world, are falling into disrepair, and new threats loom. Drinking water disasters in Flint, Michigan, droughts and floods increasingly attributable to anthropogenic climate change (*1*), and growing violence worldwide over water offer a glimpse of what we face unless new efforts are made to address failing infrastructure, worsening climate conditions, and ineffective policies and regulations (*2*). Yet, if there is any issue that offers the opportunity for nonpartisan agreement, it is to create and implement a 21st-century national water policy. In that vein, I detail national and international water challenges and recommendations for the next U.S. president, administration, and Congress.

*Federal agencies, authorities, and policies are often inconsistent, overlapping, and inefficient.* Addressing water challenges requires consistent, effective, and efficient management and institutions. Yet ~30 different federal agencies or departments have overlapping and conflicting responsibilities for fresh water. For example, the Bureau of Reclamation (BoR), Army Corps of Engineers, and agencies like the Tennessee Valley Authority each build and manage dams. The Environmental Protection Agency oversees tap water quality, but the Food and Drug Administration oversees bottled water quality. The National Park Service, BoR, Forest Service, and others manage water resources on lands under their jurisdiction, often within the same watershed. A mix of federal and state agencies and commissions manage international agreements over the shared waters with Canada and Mexico.

Production of food by U.S. farmers is at risk because federal water, energy, and agricultural policies often have conflicting and contradictory priorities and objectives (*3*). National policies designed to boost biofuels production had unanticipated impacts on food production and regional water demands. For example, an average of 780 liters of water are required to produce a liter of ethanol from irrigated corn, much of this from overdrafted aquifers in the Great Plains (*4*).

The next U.S. president should create a bipartisan water commission to evaluate and recommend changes to national water policies. The commission would provide guidance to reorganize and streamline the diverse and uncoordinated federal water responsibilities and laws, including better coordination among energy, water, and food policies. We have had no such guidance since the final report of the last U.S. National Water Commission in 1973, which first called for—and helped drive acceptance of—environmental water policies, improved water-quality regulation, and better economic tools for utilities (*5*). Such commissions offer the opportunity to generate nonpartisan recommendations that can overcome political barriers to action.

> ***"...if there is any issue that offers the opportunity for nonpartisan agreement, it is to create and implement a 21st-century national water policy."***

*Basic water science and data collection remain undone.* Vital water data are not collected or analyzed, and fundamental hydrologic science remains incomplete (*6*). There is massive groundwater overdraft in California and the Great Plains–Ogallala aquifer but little accurate information about withdrawals or recharge rates. The U.S. Geological Survey collects and publishes water-use data only every 5 years (*7*), and data are not collected in a comprehensive or consistent manner. Links between clean and adequate water and healthy aquatic ecosystems are strong, but little information is available on sustainable watersheds and freshwater management.

A national program to expand collection, management, and release of water supply and use data is key to developing sustainable policies and improving water sciences. This includes federal support for remote sensing platforms, such as replacing the SMAP (Soil Moisture Active Passive) satellite sensors and fully funding the National Oceanic and Atmospheric Administration's (NOAA's) Joint Polar Satellite System. Funding and expanding the new National Water Center, coordinated by NOAA, is a step in the right direction.

*Critical water infrastructure is often obsolete and decaying.* The United States pioneered and built water treatment and delivery systems that provide nearly all Americans with safe water and sanitation and eliminated cholera, dysentery, and other water-related diseases prevalent in other parts of the world. But hundreds of thousands—if not millions—of Americans still lack access to safe water. Recent failures—such as in Flint, Michigan, where bad technical, financial, and management decisions led to high levels of lead in the water—highlight underinvestment in system maintenance and replacement. Water in rivers, streams, and lakes is inadequately protected from contamination by weak or unenforced regulations, especially nonpoint sources of pollution from agriculture and urban development. Public and private water agencies are not adequately monitoring and enforcing existing laws and regulations.

The next president and Congress must work together to modernize water-quality laws—in particular the Clean Water Act and the Safe Drinking Water Act (SDWA)—and give federal agencies resources to oversee and enforce these laws. Challenges include improving our ability to understand and mitigate uncontrolled sources of pollution in streams, rivers, and lakes; adding regulations to address long-ignored risks to groundwater; and accelerating rules for new contaminants in drinking water. Hundreds of unregulated chemicals and microbes may pose health risks (*8*) but no new contaminant has been added under the SDWA for decades. Other priorities should be the complete elimination of lead fixtures in cities, the testing of water in every school, and remediation of any contamination problems, and investment in new water treatment and reuse technologies.

In regions where water availability is a

*Pacific Institute, Oakland, CA 94612, USA.*
*Email: pgleick@pipeline.com*

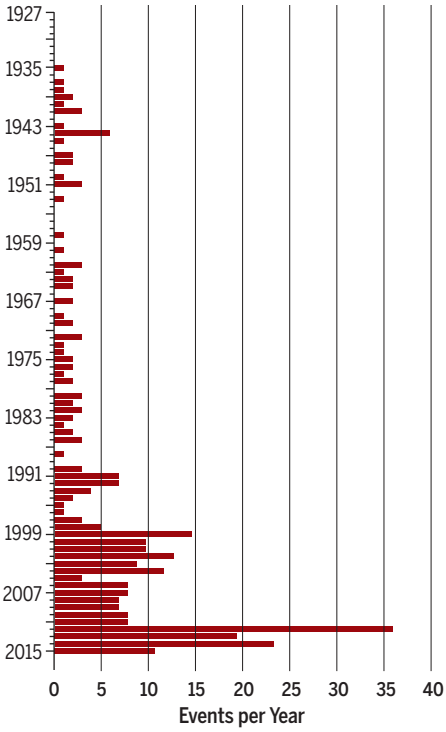
growing problem, especially the western United States, urban and farm water use can be made far more efficient with technologies such as precision irrigation, soil-moisture monitoring, and modern appliances and by using policy tools such as the Environmental Quality Incentives Program in the Farm Bill, trade laws, national efficiency standards, and tax-code revisions that promote water-efficiency investments for growers, industry, and communities. State revolving-loan programs, tax incentives, and direct support, especially for economically disadvantaged communities, can expand funds available for upgrading infrastructure. A 2013 report concluded that these strategies could produce hundreds of thousands of new jobs in urban and agricultural conservation and efficiency, stormwater management, alternative water supplies, and ecological restoration (*9*).

*Links between water conflicts and national security are clear and growing.* There is a long history of political insecurity and instability in regions where access to fresh water is a problem (see the figure). Recent experience in the Middle East shows that water problems contribute to food shortages, create environmental refugees, weaken governments, and worsen civil conflict (*10*). In Syria, severe drought contributed to economic and political destabilization, and attacks on water systems have led to a worsening humanitarian disaster and new tensions between the United States and Russia. There are ongoing examples of violence related to water scarcity and control in eastern Africa and central and southern Asia. An updated list of such conflicts can be found at the Water Conflict Chronology database (http://worldwater.org/water-conflict/). In 2012, the National Intelligence Council released an assessment of national security threats to the United States from global water challenges (*11*). Similar assessments show how climate change may contribute to state collapse and threaten peace and security (*12*). To understand and reduce these risks, the intelligence community must monitor water-related threats to U.S. security and interests. The Department of State should expand diplomatic efforts to prevent water-related conflicts worldwide, and the United States should ratify the United Nations Convention on the Law of the Non-Navigational Uses of International Watercourses—the major international agreement that provides guidelines for peacefully managing shared watersheds (*13*).

*Many people still lack basic safe water and sanitation.* More than 2 billion people—nearly a third of the global population—lack reliable, affordable access to basic water and wastewater services. Water-related diseases are prevalent in many developing countries, leading to nearly 250 million illnesses and

## Water Conflict Events per Year, 1927–2015

Data from the Water Conflict Chronology List 2015.




millions of preventable deaths a year, mostly among children (*14*). Even in the United States, many communities continue to suffer from contaminated drinking water or lack the financial resources to install modern treatment and distribution systems.

As part of the international effort to reach the new Sustainable Development Goal of achieving 100% coverage of safe water and sanitation by 2030, Congress should boost the modest resources currently available for domestic and international programs to meet basic human needs for water and to monitor water-related diseases. The next U.S. president will be responsible for developing and issuing a Global Water Strategy in fall 2017, as required by the 2014 Paul Simon Water for the World Act (*15*). This law, passed with bipartisan support in Congress, aims to redirect and expand U.S. foreign aid to increase access to safe water, sanitation, and hygiene in high-priority countries; work to improve watershed management in such countries; and help reduce water-related conflicts.

*Climate change impacts on water resources and systems are worsening.* The most recent national scientific assessment of climate risks for the United States identifies a wide range of growing risks to water resources (*16*). Rising temperatures are increasing demands for water and rapidly melting snow and ice. Rising sea levels are threatening coastal aquifers and wetlands. Some floods and droughts, already the nation's most destructive natural disasters, are now more extreme because of human-caused climate change (*17*).

We need to integrate climate change into water management and planning to help adapt to unavoidable and increasingly harsh impacts of climate change. All federal agencies that manage land and water must integrate scientific findings around climate impacts and adaptation into long-term plans and current operations. The next administration will have to implement the provisions of the Paris Agreement that call for reducing emissions of greenhouse gases and work to prepare the country for the impacts of climate change that can no longer be avoided.

We have neglected the nation's fresh water far too long. The next administration and Congress have the opportunity and responsibility to ensure that federal agencies, money, and regulations work to protect our waters, citizens, communities, and national interests. ■

### REFERENCES AND NOTES


1. J. E. Hay, D. Easterling, K. L. Ebi, A. Kitoh, M. Parry, *Weather Clim. Extremes* **11**, 103 (2016).
2. P. H. Gleick, *Clim. Change* **100**, 125 (2010).
3. J. Christian-Smith, P. H. Gleick, Eds., *A 21st Century U.S. Water Policy* (Oxford Univ. Press, New York, 2011).
4. C. W. King, A. S. Stillwell, K. M. Twomey, M. E. Webber, *Nat. Resour. J.* **53**, 117 (2013).
5. U.S. National Water Commission, "New directions in U.S. water policy: Summary, conclusions, and recommendations from the final report of the National Water Commission" (U.S. Government Printing Office, Washington, DC, 1973).
6. P. H. Gleick *et al.*, in *Climate Science for Serving Society: Research, Modeling and Prediction Priorities*, G. R. Asrar and J. W. Hurrell, Eds. (Springer, Dordrecht, 2013), pp. 151–184.
7. M. A. Maupin *et al.*, *Estimated Use of Water in the United States in 2010* (USGS circular 1405, USGS, Reston, VA, 2014).
8. K. E. Murray, S. M. Thomas, A. A. Bodour, *Environ. Pollut.* **158**, 3462 (2010).
9. E. Moore *et al.*, *Sustainable Water Jobs: A National Assessment of Water-Related Green Job Opportunities* (Pacific Institute, Oakland, CA, 2013).
10. P. H. Gleick, *Weather Clim. Soc.* **6**, 331 (2014).
11. Defense Intelligence Agency, *Global Water Security: Intelligence Community Assessment* (ICA 2012-08, DIA, Washington, DC, 2012).
12. S. Hsiang, M. Burke, E. Miguel, *Science* **341**, 1235367 (2013).
13. S. McCaffrey, Introduction, United Nations: Convention on the Law of the Non-Navigational Uses of International Watercourses, New York, 1997 (2016); http://legal.un.org/avl/ha/clnuiw/clnuiw.html.
14. UNICEF and World Health Organization, "Progress on sanitation and drinking water–2015 update and MDG assessment" (UNICEF and WHO, Geneva, 2015).
15. Senator Paul Simon Water for the World Act of 2014, Pub. L. No. 113-289 (2014).
16. J. M. Melillo, T. C. Richmond, G. W. Yohe, Eds., *Climate Change Impacts in the United States: The Third National Climate Assessment* (U.S. Global Change Research Program, Washington, DC, 2014).
17. P. Stott, *Science* **352**, 1517 (2016).


### ACKNOWLEDGMENTS

Thank you to two reviewers for thoughtful comments and suggestions and to M. Heberger for helping to format and maintain the online Water Conflict Chronology database. Partial funding for this work was provided by the Flora Family Foundation and the Horace W. Goldsmith Foundation.

10.1126/science.aaj2221

ESSAY

NEUROSCIENCE

# One brain, many genomes

## Single-cell genomic techniques shed light on somatic mutations in brain development and neurologic disorders

*By* **Gilad D. Evrony**

We each begin life as a single cell harboring a single genome, which—over the course of development—gives rise to the trillions of cells that make up the body. From skin cells to heart cells to neurons of the brain, each bears a copy of the original cell's genome. But as anyone who has used a copy machine or played the childhood game of "telephone" knows, copies are never perfect. Every cell in an individual actually has a unique genome, an imperfect copy of its cellular ancestor differentiated by inevitable somatic mutations arising from errors in DNA replication and other mutagenic forces (*1*). Somatic mutation is the fundamental process leading to all genetic diseases, including cancer; every inherited genetic disease also has its origins in such mutation events that occurred in an ancestor's germline cells. Yet how many and what kinds of somatic mutations accumulate in our cells as we develop and age has long been unknown and a blind spot in our understanding of the origins of genetic disease.

*Harvard Medical School, Boston, MA 02115, USA, and Mount Sinai Hospital, New York, NY 10029, USA. Email: g.evrony@gmail.com*

While in the laboratory of Christopher Walsh at Boston Children's Hospital and Harvard Medical School, I became intrigued by reports of neurologic diseases caused by somatic mutations, including rare cases of epilepsy, neurodegeneration, intellectual disability, brain malformations, and autism spectrum disorder (*2*–*8*). There were also long-standing hypotheses in the field that somatic genetic diversity may be prevalent in the human brain (*9*). Still, the common view was that the brain operates from a unitary genome.

I wondered how much of a genomic patchwork is our brain? What kinds and how many somatic mutations are present, and do they affect brain function? Could somatic mutations underlie some of the neuropsychiatric diseases whose causes remain unknown?

Answering these questions, we recognized, would require development of a technology to sequence the vanishingly small amount of DNA (6 picograms) present in single brain cells. Any individual somatic mutation may be present in only a very small fraction of cells or even just one cell, making it undetectable by standard DNA sequencing, which mixes together DNA from thousands or millions of cells. Together with colleagues, I developed methods to sequence the genomes of single brain cells, allowing detection of even the rarest somatic mutations. Along with novel bioinformatic approaches developed with Eunjung Alice Lee in the lab of Peter Park, this provided the first systematic, genome-wide measurements of somatic mutation in the brain (*10*, *11*). Together with single-cell studies of cancer and sperm (*12*–*14*), this heralded the emergence of the field of single-cell genome sequencing, spurred by the confluence of whole-genome amplification technologies (*15*) and decreasing DNA-sequencing costs.

Our single-neuron genomics studies have identified remarkably diverse somatic mutations that reveal a wide gamut of mutation processes impacting the brain, from small point mutations and microsatellite polymorphisms to larger retrotransposon insertions, copy-number variants, and aneuploidy (*10*, *11*, *16*, *17*) (see the figure). Notably, we are finding that each type of mutation occurs at distinct rates and patterns (*10*, *11*, *16*–*18*). These and single-neuron studies by others provide a proof of principle for the systematic quantification of somatic mutations in any human tissue (*19*, *20*).

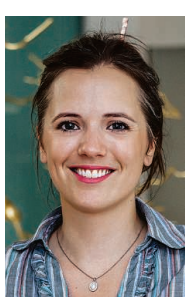
During my doctoral work, I was part of a team led by Ann Poduri that identified the first brain-specific somatic mutations causing neurologic disease (*21*). The disease in question was hemimegalencephaly, a rare congenital brain malformation in which one brain hemisphere is severely dysplastic

eppendorf & Science PRIZE FOR NEURO BIOLOGY

GRAND PRIZE WINNER:

**Gilad Evrony**

Gilad Evrony received his undergraduate degree from the Massachusetts Institute of Technology. He served in the Intelligence Division of the Israel Defense Forces and completed an M.D. and Ph.D. at Harvard Medical School, with graduate research in the laboratory of Dr. Christopher Walsh at Boston Children's Hospital. Dr. Evrony is currently pursuing clinical training in pediatrics at Mount Sinai Hospital and continuing his research developing novel technologies for studying the brain and neuropsychiatric diseases.



FINALIST:

**Anna Beyeler**

Anna Beyeler received her undergraduate degree from the University of Bordeaux, in southern France, where she then completed her Ph.D. degree requirements. As a postdoctoral fellow at the Massachusetts Institute of Technology, she has been exploring the neural circuit mechanisms underlying rewarding and aversive memories. Dr. Beyeler is in the midst of establishing an independent research program aimed at identifying neural substrates of anxiety disorders at the University of Lausanne, Switzerland.

www.sciencemag.org/content/354/6312/558.1

FINALIST:

**Arjun Krishnaswamy**

Arjun Krishnaswamy received undergraduate and Ph.D. degrees from McGill University. As a postdoctoral fellow at Harvard University, he has been using molecular, electrophysiological, and genetic approaches to learn how developing neurons in the mouse retina choose synaptic targets and establish wiring patterns important for retinal function. Dr. Krishnaswamy will continue this line of research at McGill University.

www.sciencemag.org/content/354/6312/558.2

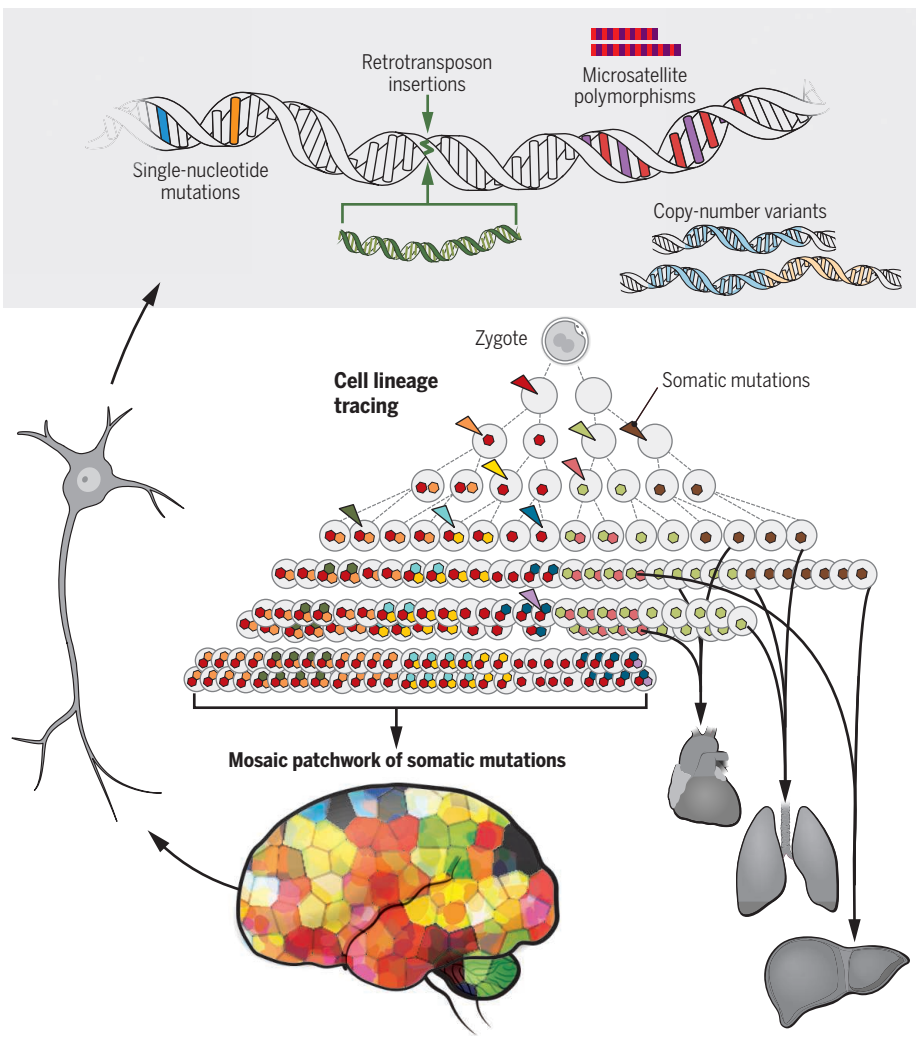
PHOTOS: (LEFT TO RIGHT) JAMES BAREHAM; ELIZABETH CLARK; SIRI POWELL

and grows too large, leading to intractable epilepsy. Using single-cell sequencing, we found that ~20 to 30% of both glial and neuronal cells carry the disease-causing mutations, allowing us to pinpoint neuroglial progenitors of the cortex as the cell type where the mutations occurred (*8*, *10*, *16*). This finding provides insight into the source of hemimegalencephaly and related focal neurologic diseases and may potentially be relevant to other epilepsies of unknown origin. Whereas hemimegalencephaly is visible by imaging, our findings suggest that radiographically invisible somatic mutations— for example, in ion channels or synaptic proteins—may be an occult cause of some neurologic diseases. As a result of these studies, numerous groups have begun systematic investigations of somatic mutations in neuropsychiatric disease.

A key tool in developmental biology is a technique known as lineage tracing, in which the fate of all the offspring of a particular cell (or group of cells) are tracked as the body forms. In animal models, this is achieved by tagging cells with fluorescent proteins or other invasive markers that cannot be used in humans. Somatic mutations, however, occur naturally and, as we were able to demonstrate, can serve as noninvasive lineage markers in humans (*1*, *11*, *17*).

Using somatic mutations identified by our single-neuron sequencing, we have been able to trace cell lineages in the human brain and reconstruct their proliferation and migration during development (*11*, *17*). Clear spatial patterns were revealed, including cell lineages distributed over an entire hemisphere at surprisingly low mosaicism indicating spatial mixing among early brain progenitors, as well as somatic mutations marking focally distributed lineages present in only one small (<1 $cm^2$) area of the frontal cortex (*11*, *17*). The latter pattern would seem to indicate that every brain is a mosaic patchwork of focal somatic mutations. It is therefore possible that rare, as-yet unrecognized brain disorders may exist in which a focal somatic mutation affects only one small region responsible for a particular cognitive function, while sparing the rest of the brain.

Our studies have generated a number of captivating new questions. Might the brain be particularly susceptible to harmful somatic mutations because of the unique interconnectedness of its cells? Would neurogenetic diseases—for example, genetic cases of autism—manifest differently if the inciting mutation were present in only one hemisphere, one lobe, or just one gyrus of the brain? Single-neuron sequencing may also be useful for measuring the extent to which neurons, whose genomes must function for decades, accumulate mutations with age and determining whether these mutations eventually impair function. We believe that single-cell genomics combined with single-cell epigenomics, transcriptomics, and proteomics will ultimately revolutionize our understanding of brain development and function. ■




**Somatic mutation events.** Single-cell genomic techniques enable the systematic measurement of somatic mutations that occur throughout normal body and brain development. Each somatic mutation event (triangles) is inherited by all offspring of the cell in which it occurred. Somatic mutations can therefore also be used as spontaneously occurring, endogenous markers for lineage tracing in human tissues, enabling reconstruction of patterns of progenitor proliferation and migration in the brain.

GRAPHIC: ADAPTED FROM G. D. EVRONY BY K. SUTLIFF/*SCIENCE*

## REFERENCES AND NOTES


1. E. Shapiro, T. Biezuner, S. Linnarsson, *Nat. Rev. Genet.* **14**, 618 (2013).
2. J. G. Gleeson *et al.*, *Am. J. Hum. Genet.* **67**, 574 (2000).
3. C. Depienne *et al.*, *J. Med. Genet.* **47**, 404 (2010).
4. M. Topçu *et al.*, *Eur. J. Hum. Genet.* **10**, 77 (2002).
5. A. Alzualde *et al.*, *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **153B**, 1283 (2010).
6. P. Castronovo *et al.*, *Clin. Genet.* **78**, 560 (2010).
7. J. A. Beck *et al.*, *Hum. Mol. Genet.* **13**, 1219 (2004).
8. A. Poduri, G. D. Evrony, X. Cai, C. A. Walsh, *Science* **341**, 1237758 (2013).
9. A. R. Muotri, F. H. Gage, *Nature* **441**, 1087 (2006).
10. G. D. Evrony *et al.*, *Cell* **151**, 483 (2012).
11. G. D. Evrony *et al.*, *Neuron* **85**, 49 (2015).
12. N. Navin *et al.*, *Nature* **472**, 90 (2011).
13. Y. Hou *et al.*, *Cell* **148**, 873 (2012).
14. J. Wang, H. C. Fan, B. Behr, S. R. Quake, *Cell* **150**, 402 (2012).
15. F. B. Dean *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5261 (2002).
16. X. Cai *et al.*, *Cell Rep.* **8**, 1280 (2014).
17. M. A. Lodato *et al.*, *Science* **350**, 94 (2015).
18. G. D. Evrony, E. Lee, P. J. Park, C. A. Walsh, *Elife* **5**, e12966 (2016).
19. M. J. McConnell *et al.*, *Science* **342**, 632 (2013).
20. Y. Wang, N. E. Navin, *Mol. Cell* **58**, 598 (2015).
21. A. Poduri *et al.*, *Neuron* **74**, 41 (2012).


## ACKNOWLEDGMENTS

I am grateful to C. A. Walsh for his mentorship, P. J. Park and E. A. Lee for rewarding collaborations, and X. Cai and all members of the Walsh laboratory. This work was supported by NIH Medical Scientist Training Program grant T32GM007753, the Louis Lange III Scholarship in Translational Research, and the Howard Hughes Medical Institute.

10.1126/science.aak9761

ESSAY

NEUROSCIENCE

# Parsing reward from aversion

## Investigating the neural circuits encoding emotional valence

*By* **Anna Beyeler**

Starting from the moment we hear our alarms in the morning, our emotions guide the thousands of decisions we make every day. More specifically, it is the valence of our emotions that determines our subsequent behavior. Valence is a concept that was originally defined in psychology and corresponds to the value we assign to the perceptions of our external and internal environments (*1*). Valence varies from negative, when we are afraid or anxious, to positive, when we are happy or peaceful. In the case of the morning alarm, if your emotional state has a positive valence you might jump out of bed, eager to engage with whatever is motivating you. Conversely, if your emotional state has a negative valence, you might choose to stay in bed to avoid the causes of your negative emotions.

Beyond allowing or preventing your timely arrival to work, the neural circuits supporting valence encoding are critical for your survival. Imagine the consequences if you were to assign a positive valence to, and therefore seek, dangerous situations, for example. In fact, misassignment of valence is not rare, and dysfunctions of the circuits encoding valence are thought to underlie many psychiatric illnesses including anxiety, depression, addiction, and compulsive disorders (*2*, *3*). From a fundamental and clinical point of view, understanding how the brain attributes valence to contexts and salient elements of our environment is one of the main challenges of modern neuroscience.



It has been known for decades that the basolateral amygdala (BLA) is necessary for associative learning of both positive and negative valence (*4*, *5*). However, understanding how a single brain region can be responsible for encoding such a diverse range of information has been elusive.

One hypothesis to explain this paradox is that specific subpopulations of the BLA encode opposite valence. As a postdoctoral fellow in the Laboratory of Professor Kay Tye at the Massachusetts Institute of Technology (MIT), I have dedicated the past 4 years to testing this hypothesis, combining cutting-edge approaches in functional neuroanatomy, electrophysiology, and optogenetics. During this time, my collaborators and I assessed different subpopulations of the BLA that relay integrated information to distinct regions. Over multiple studies, I identified circuit, synaptic, and molecular mechanisms by which three distinct neural populations in the BLA encode innate and learned behaviors (*6*–*8*).

Consistent with recent studies in other brain regions (*9*), we found that different BLA neuronal populations (as defined by their projection targets) differentially encode valence. To do so, we used optogenetic-mediated photo-tagging in combination with large-scale electrophysiological recordings during the retrieval of positive or negative associative memories in mouse models and examined the activity of BLA neurons projecting to the nucleus accumbens, the centromedial amygdala, or the ventral hippocampus (see the figure).

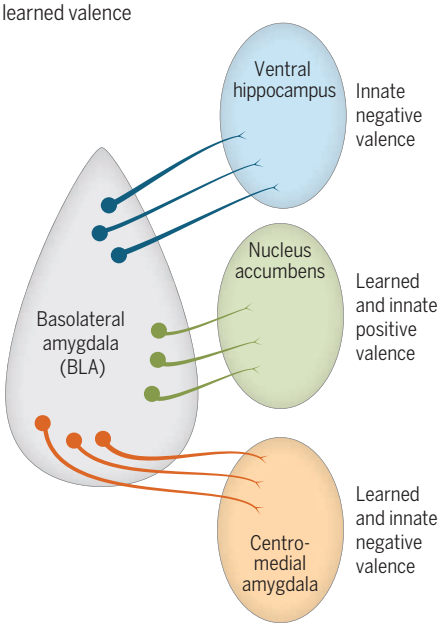
Although heterogeneity of response types was known to exist in the entire BLA, we were surprised to find that when mice encounter cues predicting a rewarding or aversive outcome (*10*, *11*), heterogeneity of response types was also noted in BLA subpopulations. However, two BLA populations differentially encode valence during the retrieval of cues of opposite valence (*7*). Specifically, BLA neurons projecting to the nucleus accumbens were preferentially excited by cues predicting a reward, and those projecting to the central amygdala were preferentially excited by cues predicting an aversive outcome. The response-type heterogeneity we observed in BLA subpopulations could play a role in behavioral flexibility in cases in which the valence of environmental stimuli changes.

Consistent with our recordings in vivo, the synaptic inputs of BLA neurons projecting to the nucleus accumbens are potentiated after mice learn to associate a cue with a reward and depressed after mice learn to fear a cue associated with a foot shock (*8*). Conversely, the synaptic inputs of BLA neurons projecting to the centromedial amygdala were potentiated after fear learning and depressed after reward learning (*8*). Optogenetic activation of these projections induced approach and avoidance behavior, respectively (*8*), showing that these divergent pathways causally control valence encoding.

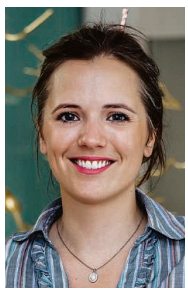
Although we had previously found that optogenetically simulating the BLA neurons that project to the ventral hippocampus

### Valence circuitry

A proposed model for the role of basolateral amygdala projector populations in innate and learned valence




*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA. Email: annabey@gmail.com*



FINALIST:

**Anna Beyeler**

Anna Beyeler received her undergraduate degree from the University of Bordeaux, in southern France, where she then completed her Ph.D. degree requirements. As a postdoctoral fellow at the Massachusetts Institute of Technology, she has been exploring the neural circuit mechanisms underlying rewarding and aversive memories. Dr. Beyeler is in the midst of establishing an independent research program aimed at identifying neural substrates of anxiety disorders at the University of Lausanne, Switzerland.

**www.sciencemag.org/content/354/6312/558.1**

drives negative valence in innate anxiety-related behaviors (*6*), we determined that these neurons do not preferentially respond to cues predicting unpleasant outcomes (*7*). These results add a level of complexity to our understanding of valence circuits and support a model of distinct routing for the attribution of valence for innate and learned information.

We have begun to unravel how BLA populations encode valence in physiological conditions (*6*–*8*). We can now start to investigate how valence circuits are dysfunctional in pathologies such as anxiety and depression. Ultimately, we hope this research will lead to the development of novel strategies to restore the function of those circuits in human patients. ■

**REFERENCES**


1. J. A. Russell, *J. Pers. Soc. Psychol.* **39**, 1161 (1980).
2. K. M. Tye, K. Deisseroth, *Nat. Rev. Neurosci.* **13**, 251 (2012).
3. A. Lüthi, C. Lüscher, *Nat. Neurosci.* **17**, 1635 (2014).
4. J. LeDoux, *Cell. Mol. Neurobiol.* **23**, 727 (2003).
5. P. H. Janak, K. M. Tye, *Nature* **517**, 284 (2015).
6. A. C. Felix-Ortiz *et al.*, *Neuron* **79**, 658 (2013).
7. A. Beyeler *et al.*, *Neuron* **90**, 348 (2016).
8. P. Namburi *et al.*, *Nature* **520**, 675 (2015).
9. S. Ciocchi, J. Passecker, H. Malagon-Vina, N. Mikus, T. Klausberger, *Science* **348**, 560 (2015).
10. S. J. Shabel, P. H. Janak, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15031 (2009).
11. J. J. Paton, M. A. Belova, S. E. Morrison, C. D. Salzman, *Nature* **439**, 865 (2006).


10.1126/science.aak9762

ESSAY

NEUROSCIENCE

# Building connections

## Studies in the developing retina provide insights into neural circuit assembly

*By* **Arjun Krishnaswamy**

Much of a child's first year is spent asleep, punctuated by the occasional feeding, fit, or flatulence. The serenity is only skin deep, as a flurry of action goes on beneath. The brain is building itself at a frantic pace during this year, forming millions of synapses per second. Remarkably, speed does not incur a cost in precision—children reach their first birthday with specific, largely mature patterns of connectivity already formed, ready to become terrible at 2. How the brain manages this remarkable feat is not altogether clear.

We do know part of the story. It begins with the generation of neurons, which migrate to appropriate locations and extend axons long distances to target areas (*1*). The puzzle is what happens next, when axons and dendrites form their remarkably specific patterns of connectivity. Forming synapses is an astonishingly promiscuous process. Decades of study indicate that neurons readily synapse with each other if they share matched transmitters and receptors, as well as generic organizing molecules. Yet, axons in fact synapse with only small subsets of potential targets and often only on specific portions of a target—for example, a distal dendrite, a soma, or an axon hillock (*1*, *2*).

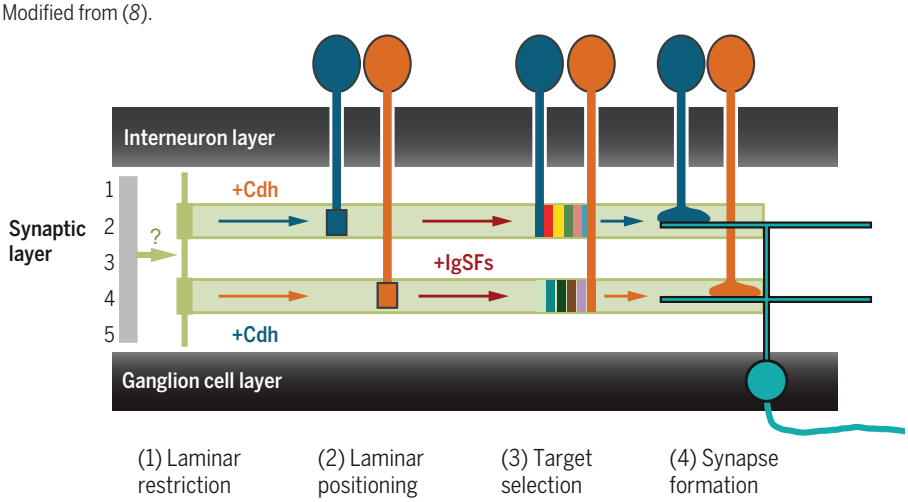


How do these selective connections arise? One idea finds its roots in the classical work by Langley and Sperry: Selectivity between potential synaptic partners arises because they have a specific chemical affinity for one another (*3*–*5*). Endowing neurons with this affinity might be the role for the large families of recognition molecules that take up ~3% of the genome (*1*). There is ample evidence that such proteins regulate cell-cell contact but little evidence that their combinatorial expression patterns are related to specific connectivity. Could such expression patterns encode some or all of the nervous system's blueprint?

*Department of Molecular and Cellular Biology and Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA. Email: akrish@mcb.harvard.edu*

### Retinal circuit assembly

Model for selective synapse formation and circuit assembly in the retina. Progressively narrowing the options available to developing neurons leads to specific connections among interneurons and retinal ganglion cells. Modified from (*8*).




### VISUALIZING THE WIRING OF THE RETINA

Conceptually, deciphering the code requires us to draw links among genes, circuitry, and function. Technically, this requires a list of the neural components to be wired, tools to mark and manipulate them, and methods to map circuitry. For much of the central nervous system, this is a tall order, but the retina is a notable exception.

The retina is a sophisticated neural computer whose parallel circuits preprocess the visual scene to highlight salient features before this information is transmitted to the brain (*6*). Each of the 30 or more retinal circuits begins with photoreceptors, which work as photon detectors. Each ends with a retinal ganglion cell (RGC), which acts as a feature detector that conveys visual information to the brain. There are ~30 types of RGCs, each uniquely attuned to features such as motion, color, contrast, and so on. In between are 70 or more types of interneurons. RGCs are endowed with their preference for different features via the synapses that they receive from specific subsets of interneurons, called amacrine cells (ACs) and bipolar cells (BCs) (*6*, *7*). How this precision is established and its contribution to retinal circuitry are unclear.

Initially, methods to map retinal circuitry were lacking. An ever-growing compendium of mouse lines that let one mark and manipulate many of the ~100 cell types in the retina, however, made this problem tractable.

We designed and built a custom two-photon microscope–physiology rig to provide spatiotemporally restricted optogenetic excitation to hundreds of genetically defined interneurons in the mouse retina while recording intracellularly from defined RGC subtypes. Armed with this approach, we could map functional connectivity among the retinal types and study how these maps arise.

### CADHERINS DIRECT LAYER-SPECIFIC INNERVATION

Using the technique described above, we first discovered a pair of related BCs that provide visual input to directionally selective RGCs (DSRGCs) (*8*). Anatomically, these BC axons target layers occupied by DSRGC dendrites, which suggests that BCs recognize these locations in the retinal neuropil. Using genetic screens, we discovered that a pair of cadherins were expressed by these BCs and later determined that these cadherins position the BC axons in layers containing DSRGCs dendrites. This organization is critical for DSRGC visual responses—perturbing it by deleting cadherins disconnects BC input and effectively renders DSRGCs blind (*8*). Conceptually, choosing a layer mitigates the wiring chal-

GRAPHIC: ADAPTED FROM A. KRISHNASWAMY BY K. SUTLIFF/*SCIENCE*

FINALIST:

**Arjun Krishnaswamy**

Arjun Krishnaswamy received undergraduate and Ph.D. degrees from McGill University. As a postdoctoral fellow at Harvard University, he has been using molecular, electrophysiological, and genetic approaches to learn how developing neurons in the mouse retina choose synaptic targets and establish wiring patterns important for retinal function. Dr. Krishna-swamy will continue this line of research at McGill University.
**www.sciencemag.org/content/354/6312/558.2**

lenge by limiting the total number of potential partners that must be connected.

## A "SIDEKICK" ENHANCES PAIRING WITH POTENTIAL SYNAPTIC PARTNERS

Layering alone cannot explain the circuitry of the retina. Simply put, there are ~30 kinds of retinal circuits but only 5 to 10 layers in which to contain their wiring—how do neurons find each other within this tangled thicket? We reasoned that such selectivity should result in enriched connectivity between a subset of colaminar neurons and set out to determine whether such patterns existed by functionally mapping more than a dozen colaminar interneuron-RGC pairs. We discovered that ACs defined by expression of the vesicular glutamate transporter 3 (VG3-ACs) connect well to one type of RGC, W3B RGCs, but poorly to other RGCs in the same layer (*9*). This selectivity is remarkable because VG3-W3B connections are interdigitated with those of several others.

Such complexity poses a serious wiring challenge for these two neurons and suggests the presence of organizers within layers that operate at an incredibly fine scale. Parallel experiments confirmed this suspicion—VG3 and W3B connect strongly because they both express a homophilic immunoglobulin superfamily (IgSF) adhesion molecule called Sidekick 2 (*9*). Connections of VG3 and W3B with equally proximate partners that are Sidekick-negative are substantially weaker. This indicates that recognition biases local connectivity in favor of particular pairings.

## A HIERARCHY OF WIRING CHOICES BUILDS RETINAL CIRCUITS

This work leads to a speculative model for retinal circuit assembly that revolves around progressively narrowing the options available to developing neurons (see the figure) (*8*, *9*). This model is attractive because it unburdens the genome of having to provide unique recognition for every synapse. For example, the division of labor with cadherins for layers and IgSFs for targets might allow the same IgSFs to be used in several layers at once. Reality might be even simpler, since ~10 other kinds of wiring choices have been documented, each enacted by largely distinct genetic programs (*1*). Combinations of such genetic programs might direct neurons through a hierarchical sequence of choices that result in particular connectivity patterns.

Do neurons use all of these strategies? Or only a subset? How are they combined? Could the path of a given neuron influence the wiring of another? And perhaps most important, how do activity and experience act on this initial scaffold to refine patterns of connectivity? Understanding these systems-level interactions among circuit patterning rules is critical for developmental neurobiology's ultimate goal—to know how to build a brain. This goal, if realized, might offer major clues to brain function and new avenues of diagnosis and treatment of dysfunctions caused by mental illness. ■

### REFERENCES


1. J. R. Sanes, M. Yamagata, *Annu. Rev. Cell Dev. Biol.* **25**, 161 (2009).
2. J. de Wit, A. Ghosh, *Nat. Rev. Neurosci.* **17**, 22 (2016).
3. R. W. Sperry, *Proc. Natl. Acad. Sci. U.S.A.* **50**, 703 (1963).
4. R. W. Sperry, *J. Comp. Neurol.* **79**, 33 (1943).
5. J. N. Langley, *Philos. Trans. R. Soc. London B* **183**, 85 (1892).
6. T. Gollisch, M. Meister, *Neuron* **65**, 150 (2010).
7. J. R. Sanes, R. H. Masland, *Annu. Rev. Neurosci.* **38**, 221 (2015).
8. X. Duan, A. Krishnaswamy, I. De la Huerta, J. R. Sanes, *Cell* **158**, 793 (2014).
9. A. Krishnaswamy, M. Yamagata, X. Duan, Y. K. Hong, J. R. Sanes, *Nature* **524**, 466 (2015).


10.1126/science.aak9763

PHOTO: SIRI POWELL

PHOTO: NATIONAL GEOGRAPHIC CREATIVE/ALAMY STOCK PHOTO

Fossils found in the Burgess Shale have yielded insights into our planet's extraordinary biodiversity.

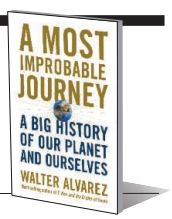
GEOLOGY

# *Against all odds*

## A geologist revels in the unlikely reality of life on Earth

*By* **Jill S. Schneiderman**

**A Most Improbable Journey**
**A Big History of Our Planet and Ourselves**
*Walter Alvarez*
Norton, 2016. 256 pp.




When esteemed geologist Walter Alvarez and his colleagues searched the lowlands of Eastern Mexico for ancient debris that had been ejected from the Chicxulub impact crater, they were following in the footsteps of renowned early-20th-century paleontologist Charles Doolittle Walcott. In his own time, Walcott wandered the Canadian Rockies seeking Cambrian-aged fossils that could shed light on this crucial time period when multicellular life began to flourish. As recounted in his obituary, "One of the most striking of Walcott's faunal discoveries came at the end of the field season of 1909, when Mrs. Walcott's horse slid in going down the trail and turned up a slab that at once attracted her husband's attention. Here was a great treasure—wholly strange Crustacea of Middle Cambrian time.... Snow was even then falling, and the solving of the riddle had to be left to another season." Serendipity combined with persistence eventually led Walcott to one of the most important discoveries in the history of geology—the soft-bodied fauna of the Burgess Shale.

In the case of the Alvarez group, two broken jeeps on the last day of the field season of 1991 nearly prevented the researchers from reaching the strange sand bed that Alvarez had chanced to read about in a book about the geology of the region published in 1936. As they "bounced along the rough road following the Arroyo el Mimbral, worrying as the Sun got lower in the sky," they came upon a layer of glass spherules exposed in a steep bluff along the dry river bed. They suspected that the spherules were droplets of impact melt—"the most wonderful outcrop I have seen in five decades as a geologist," writes Alvarez in his new book, *A Most Improbable Journey: A Big History of Our Planet and Ourselves*.

Alvarez is best known for helping to establish that a meteor struck the Yucatán, causing the mass extinction of half the genera of animals on Earth. In *A Most Improbable Journey*, he tells the story of the cosmos, Earth, life, and humanity using the interdisciplinary approach of Big History, which combines traditional historical scholarship with scientific insights. Alvarez aims to instill in his readers a sense of wonder that, despite enormous odds, there exists a planet (Earth) supremely suited for life. At the same time, he seeks to cultivate an expanded view of the nature of history, replete with contingency and consequent improbability, and to foster appreciation for the enormous stretches of time and space across which history has unfolded.

In each section of *A Most Improbable Journey*—"Cosmos," "Earth," "Life," and "Humanity"—Alvarez uses both broad questions (have there been recognizable patterns, regularities, cycles, and contingencies in the history of continental motions?) and "little Big History" (how did the Spanish language come to dominate the Iberian Peninsula and then much of Latin America?) to help us comprehend "our whole situation."

Throughout the book, Alvarez uses evocative phrases and images: History is "violent and chancy," rocks "remember [their] history," and mountains "are not wrecks—they are sculptures." Compelling images of rock and architectural engravings, relief and sketch maps, and historical photographs and drawings enrich the discussion but, unfortunately, are not referred to directly in the text.

Alvarez invites his audience to read the book chapters in any order. Though I chose to read the book cover-to-cover, each chapter does indeed stand alone and therefore lends itself to this type of engagement. The book contains an appendix of resources that Alvarez annotates thoroughly, and it acts as an additional chapter that readers will likely enjoy perusing.

The paleontologist and historian of science Stephen Jay Gould wrote in an essay about the discovery of the Burgess Shale fauna, "So much of science proceeds by telling stories.... Even the most distant and abstract subjects, like the formation of the universe or the principles of evolution, fall within the bounds of necessary narrative" (*1*). In *A Most Improbable Journey*, Alvarez harnesses such narrative, enabling readers to experience the power of Big History. ■

*The reviewer is at the Department of Earth Science and Geography, Vassar College, Poughkeepsie, NY 12604, USA. Email: schneiderman@vassar.edu*

### REFERENCES

1. S. J. Gould, "Literary Bias on the Slippery Slope," in *Bully for Brontosaurus* (Norton, 1992).

10.1126/science.aah5116

ECONOMICS

# Priceless

## A cautionary tome probes the cost of black box algorithms

*By* **Barry Nalebuff**

In *Virtual Competition*, law professors Ariel Ezrachi and Maurice Stucke begin with a scary premise: What if computers could collude? How would we even know it was happening, and what could we do about it?

For starters, most pricing is already done by computer algorithms. A single airline flight may have more than 40 different fares. Uber's decision to surge is based on dynamic estimates of capacity utilization. Amazon has more than 200 million products for sale, too many for even an army of wizards behind the curtain to set prices.

In the good old days, firms that wanted to collude would get together in bars or industry conferences and share pricing sheets. Today, they share pricing algorithms. In 2015, in the first Internet commerce antitrust case, David Topkins pleaded guilty to sharing pricing formulas for collectible posters sold on Amazon's Marketplace. If Topkins and his competitors used the same pricing scheme, then no seller would undercut another.

Self-learning algorithms are increasingly being used for retail pricing and logistics.

But that's old wine in new bottles. Far more frightening are three new ways that computers can collude. There's hub-and-spoke collusion, in which a firm sets prices for all its independent contractors. Uber, for example, sets the price for a trip; individual drivers can't undercut each other. At least this coordination has the advantage of being visible and easy to detect.

More problematic is when a nefarious programmer writes code that encourages collusion, creating, for example, a system that immediately matches any price cut and follows any price hike. Such behavior trains rivals (humans or machines) to refrain from price cutting and to raise prices without fear of losing share. Although frightening, the collusive intent is hardwired into the code and thus discoverable.

Scariest of all is a scenario in which a computer figures out both the advantages of collusion and how to make it happen. Here, the situation might resemble what happened with AlphaGo, the computer program developed to play the board game Go. The program's success was mostly due to machine learning. The computer played countless games against itself and figured out what worked best. The end result is a black box: We don't really know how the computer is making decisions, only that it works. Because successful collusion leads to higher profits, it would make sense that computers—left to their own devices—would figure this out. Antitrust authorities would have no way to punish this type of collusion under existing laws.

That's not all that worries our digital Cassandras. The use of big data may allow monopolists to be much more entrenched and much more effective at raking in profits, often at consumer expense. For example, the cookies on my computer can tell that I never go to price-comparison websites. I fall into the "asleep-at-the-wheel" category, which makes me a prime target for price increases. Amazon may have sworn off charging different consumers different prices for the same book, but other retailers, credit cards, and insurance companies provide personalized pricing.

Will big data lead to "the end of competition as we know it"? Not yet. Ezrachi and Stucke recognize that computers can also provide solutions. I can give my proxy to a program that specializes in ferreting out the best deals, for example, leading to a battle of buyer bots versus seller bots.

The authors often do a better job describing the problems than offering solutions. One proposed solution to a potential Uber monopoly has a government regulator setting prices. The fact that San Francisco is able to dynamically monitor and set prices for parking meters doesn't give me faith that a regulator could do a better job dynamically setting prices for Uber—even if they had all of the company's data. An additional concern is that such regulation could choke off innovations such as Uber pool or driverless cars.

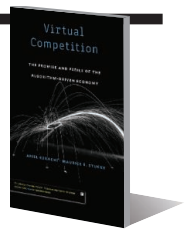
Another option the authors consider is for Uber to give up central pricing and have drivers bid on each trip. If that sounds complicated, it is. A ride-sharing app called Sidecar tried it and shut down in 2015. Even if it worked—perhaps with driver computer bots placing bids and passenger bots deciding which to accept—this could be a recipe for drivers to engage in local price discrimination. Today, the nearest driver usually takes the job at the Uber set price. With driver bidding, a nearby driver, knowing he or she is closest, could charge extra for the quicker pickup.

Most important is the question of whether this is a problem that needs fixing. Today, Uber is in a price war with Lyft; in the future, it will face competition from Google and perhaps Apple. New technologies displace what once looked like entrenched monopolies. Remember Blackberry? Blockbuster?

We owe the authors our deep gratitude for anticipating and explaining the consequences of living in a world in which black boxes collude and leave no trails behind. They make it clear that in a world of big data and algorithmic pricing, consumers are outgunned and antitrust laws are outdated, especially in the United States. They want their readers to agitate for new antitrust laws. Sign me up. When the masses get mad enough, perhaps they'll elect a new trust-busting Teddy Roosevelt for the digital era. In the meantime, we can hope the ideas discussed in *Virtual Competition* get on the political agenda. ■

*The reviewer is at the Yale School of Management, Yale University, New Haven, CT 06511, USA. Email: barry.nalebuff@yale.edu*

**Virtual Competition**
**The Promise and Perils of the Algorithm-Driven Economy**
*Ariel Ezrachi and Maurice E. Stucke*
Harvard University Press, 2016. 364 pp.




10.1126/science.aaj2011

Children play on fishing boats at Lake Tanganyika, one of the East African Great Lakes threatened by oil exploitation.

PHOTO: BLICKWINKEL/ALAMY STOCK PHOTO

# *Oil extraction imperils Africa's Great Lakes*

AS THE WORLD'S demands for hydrocarbons increase (*1*), remote areas previously made inaccessible by technological limitations are now being prospected for oil and gas deposits. Virtually unnoticed by the public, such activities are ongoing in the East African Great Lakes region, threatening these ecosystems famed for their hyper-diverse biota, including the unique adaptive radiations of cichlid fishes (*2*). Countries in the region see exploitation of hydrocarbon reserves as a vital economic opportunity. In the Lake Albert region of Uganda, for example, the government foresees a $3.6 billion oil profit per year starting in 2018—a sum almost as high as the country's current annual budget (*3*). However, oil extraction in the East African Great Lakes region poses grave risks to the environment and local communities.

The thousands of oil spills reported in Nigeria (*4*) demonstrate that the extraction and transport of oil are prone to accidents. This is especially bad news for the African Great Lakes because they are virtually closed ecosystems. For example, for Lake Tanganyika, which contains about one-fifth of the world's surface freshwater (*5*), the flushing time is ~7000 years (*5*). This time frame implies that the recovery from an oil spill could take millennia. To make matters worse, the lakes' location in a remote part of the world would impede a quick and effective reaction to an oil spill. Appropriate infrastructures are currently unavailable at the lakes, and bringing in heavy equipment at the time of a spill would be cumbersome, logistically impossible, or prohibitively expensive.

An oil spill would markedly affect the health, water supply, and food security of local communities (*6*). More than 10 million people depend on Lake Tanganyika alone for fisheries and water resources, and many more along the Congo River, into which the lake drains, are highly dependent on the lakes' ecosystem (*7*). In addition to the toll on humans, an oil spill in these lakes would be a global catastrophe for biodiversity. Combined, these lakes are home to thousands of species, almost all of them endemic (*2*). An accident might deal a final blow to these ecosystems, which have already been rendered fragile by anthropogenic stressors such as overfishing, deforestation, and global warming (*8*).

Finally, large parts of the East African region still lack political stability and security (*9*). In addition to the possibility of accidents, competition for hydrocarbon resources could lead to sabotage, as has unfortunately been frequently observed in the Niger delta (*4*).

We are concerned that the risks associated with the intended exploitation of fossil hydrocarbons in the East African Great Lakes region are seriously underestimated. We urge the countries involved in these undertakings to engage with the scientific and lake management communities to identify strong mitigation and control measures that could be put in place before hydrocarbon production begins. Local governments should foster alternative, sustainable plans to develop the region in accordance with the United Nations Sustainable Development Goals (*10*). To this end, the local population, regional stakeholders, governments, nongovernmental organizations, and scientists must cooperate to develop economically and ecologically viable strategies for the region, as is currently being attempted for the Virunga National Park in the DR Congo (*11*).

***Erik Verheyen,* on behalf of the Cichlid Science 2015 Meeting and concerned scientists***

*Corresponding author.
Email: erik.verheyen@naturalsciences.be

The full list of authors is available online.

## REFERENCES

1. International Energy Agency, Oil Market Report (2016); www.iea.org/oilmarketreport/omrpublic/.
2. W. Salzburger, B. van Bocxlaer, A.S. Cohen, *Annu. Rev. Ecol. Evol. Syst.* **45**, 519 (2014).
3. M. L. Oketch, "Uganda to earn Shs9 trillion from oil annually—Tullow," *Daily Monitor* (2014); www.monitor.co.ug/News/National/Uganda-to-earn-Shs9-trillion-from-oil-annually---Tullow/688334-2451478-rag0a9/index.html.
4. Nigerian Oil Spill Monitor (https://oilspillmonitor.ng/).
5. R. H. Spigel, G. W. Coulter, 1996, in *The Limnology, Climatology and Paleoclimatology of the East African Lakes*, T. C. Johnson, E. O. Odada, Eds. (Gordon and Breach Science Publishers, Amsterdam, 1966), pp. 103–140.
6. NCF/WWF/IUCN, "Niger Delta Natural resources damage assessment and restoration project scoping report, May 2006" [Nigerian Conservation Foundation, UK World Wildlife Federation, and International Union for

Conservation of Nature, Commission on Environmental, Economic, and Social Policy from the Federal Ministry of Environment (Abuja), 2006].
7. H. Mölsä, J. E. Reynolds, E. J. Coenen, O. V. Lindqvist, *Hydrobiologia* **407**, 1 (1999).
8. A. S. Cohen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 9563 (2016)
9. A. Adusei, *African Security Rev.* **24**, 332 (2015).
10. United Nations, "17 goals to transform our world" (2015); www.un.org/sustainabledevelopment/sustainable-development-goals/.
11. Global Witness, "Protecting Virunga National Park from oil companies" (2016); www.globalwitness.org/en/campaigns/oil-gas-and-mining/protecting-virunga-national-park-oil-companies/.

**SUPPLEMENTARY MATERIALS**

www.sciencemag.org/content/354/6312/561.1/suppl/DC1
Full author list

10.1126/science.aal1722

# Time for responsible peatland agriculture

THE 15TH INTERNATIONAL Peat Congress, held in Asia for the first time, brought together industry, policy-makers, and academia to discuss responsible peatland management. In Southeast Asia, peatland management is largely driven by the palm oil industry. After the Congress, misleading reports were published by leading Asian newspapers. They claimed that oil palm plantations on peatland can be viewed as sustainable [e.g., (*1*)] and supported the continuation of business-as-usual peatland agriculture. This is contrary to the opinion of an overwhelming number of tropical peatland scientists (*2*) and the vast majority of science published in the past two decades.

Deep, carbon-rich peat deposits are maintained by a combination of steady organic matter inputs and high water tables, which inhibit microbial decomposition (*3*). Conversion of peat swamp forest (the natural vegetation of Southeast Asian peatlands) to agricultural land requires removing vegetation and lowering groundwater tables. The combination of slash and burn techniques and drainage used to prepare peat for agriculture promotes smoldering fires and rapid peat oxidation. Peat fires are globally significant for their greenhouse gas emissions and threats to human health and regional economies (*4*). Peat oxidation leads to high $CO_2$ emissions and land subsidence. As the land surface falls toward river and sea levels, it will be subject to periodic and eventually permanent flooding, limiting future agricultural production (*5*). Agricultural use of peatlands cannot, therefore, be considered sustainable from either environmental or socioeconomic perspectives.

Industry and academia are working together to develop peatland agricultural systems (*6*) that minimize negative environmental and commercial impacts. In the interim, steps should be taken to improve hydrological management of peatlands under agriculture and to implement landscape-scale management planning. Denial of known issues slows progress toward responsible solutions, which are urgently needed to prevent avoidable losses of Southeast Asia's peatlands, as well as global consequences.

***Lahiru S. Wijedasa,[1,2,3]* Susan E. Page,[4] Christopher D. Evans,[5] Mitsuru Osaki[6]***

[1]Department of Biological Sciences, National University of Singapore, 117543, Singapore. [2]ConservationLinks, 120433, Singapore. [3]Rimba, Malaysia, Selangor, MY 43650, Malaysia. [4]Centre for Landscape & Climate Research, Department of Geography, University of Leicester, Leicester, LE1 7RH, UK. [5]Centre for Ecology and Hydrology, Bangor LL57 2UW, UK. [6]Research Faculty of Agriculture, Hokkaido University, Hokkaido 060-0808, Japan.

*Corresponding author. Email: lahirux@gmail.com

**REFERENCES**

1. B. Nurbianto, "Malaysia challenges the world over palm oil on peatland," *The Jakarta Post* (2016); www.thejakartapost.com/news/2016/08/24/malaysia-challenges-the-world-over-palm-oil-on-peatland.html.
2. L. S. Wijedasa *et al.*, *Glob. Chang. Biol.* 10.1111/gcb.13516 (2016).
3. T. Hirano *et al.*, *Glob. Chang. Biol.* **18**, 3410 (2012).
4. R. A. Chisholm, L. S. Wijedasa, T. Swinfield, *Conserv. Biol.* **30**, 5 (2016).
5. E. Sumarga, L. Hein, A. Hooijer, R. Vernimmen, *Ecol. Soc.* **21**, 52 (2016).
6. International Peat Society, "Statement regarding the Jakarta Post article of 18th August" (2016); www.peatlands.org/news/ips-statement-congress-may-change-views-cultivation-peatland-jakarta-post.

10.1126/science.aal1794

# Protecting China's soil by law

AFTER SEVERE PROBLEMS with air and water pollution, China is getting serious about its soil (*1*, *2*). On 31 May, China's State Council released an action plan for soil pollution prevention and remediation, aiming to make 90% of polluted, arable land safe for human use by 2020 and increase this to 95% by 2030 (*3*). This ambitious action plan calls for the support of strong environmental laws to monitor, prevent, and remediate serious levels of soil contamination. However, national legislation protecting soil quality has lagged behind that of air and water for more than a decade, while Chinese lawmakers debate the focus and purpose of a soil protection law.

On 3 September, China released a draft of its first environmental tax law, which designates four taxable types of pollution: airborne and water pollutants, solid waste, and noise (*4*). Soil is conspicuously absent. A soil protection law could close the current environmental legislation system's loopholes, make China's new environmental tax system more comprehensive, and protect China's soil.

To make a pragmatic soil protection law, the central government must clearly identify local government's liability and responsibility for soil pollution, as ambiguous responsibility has been one of the major problems in soil management in the past. The law must stipulate the division of duties between government agencies, establish a surveying and monitoring system, and introduce funding mechanisms. Remediation of contaminated soil is extremely costly, and China needs to create a national soil fund by allocating an adequate percentage of its land revenues and environmental tax revenues. Some members of the soil pollution plan panel initially suggested that 10% of land revenues be designated to the fund (*5*). Because China lacks comprehensive risk assessment systems for contaminated land management (*6*), the law should stipulate risk management and control approaches for contaminated sites.

China's soil pollution has become a critical issue that affects public health and creates social unrest and instability (*1*, *7*, *8*). China should not repeat its past mistakes of focusing on economic growth at the expense of the environment.

***Jinnan Wang,[1] Qing Hu,[2] Xiahui Wang,[1] Xiaoliang Li,[1,3] Xiao Jin Yang[3]****

[1]Chinese Academy of Environmental Planning, Beijing, 100012, China. [2]South University of Science and Technology of China, Shenzhen, Guangdong, 518055, China. [3]Beijing University of Chemical Technology, Beijing, 100029, China.

*Corresponding author. Email: yangxj@mail.buct.edu.cn

**REFERENCES**

1. C. Larson, *Science* **343**, 1415 (2014).
2. Ministries of Environmental Protection and Land and Resources of the People's Republic of China, National Soil Pollution Report 2013 (2014); www.mlr.gov.cn/xwdt/jrxw/201404/t20140417_1312998.htm [in Chinese].
3. Central Government of the People's Republic of China, "The State Council issues an action plan for soil pollution prevention and control" (2016); www.gov.cn/zhengce/content/2016-05/31/content_5078377.htm [in Chinese].
4. The National Congress of the People's Republic of China, Environmental Protection Tax Law (Draft), (2016); www.npc.gov.cn/npc/flcazqyj/2016-09/02/content_1996531.htm [in Chinese].
5. E. Wang, "A suggestion of allocating 10% of land revenues for soil pollution remediation fund," *21st Century Business Herald* (2016); http://epaper.21jingji.com/html/2016-03/08/content_33885.htm [in Chinese].
6. F. Coulon *et al.*, *Environ. Int.* **91**, 196 (2016).
7. G. He, "Special report: The victims of China's soil pollution crisis," *China Dialogue* (2014); www.chinadialogue.net/article/show/single/en/7073-Special-report-The-victims-of-China-s-soil-pollution-crisis.
8. Xinhua News, "No sure cure for China's soil pollution," *China Daily* (2014); www.chinadaily.com.cn/china/2014-04/29/content_17474094.htm.

10.1126/science.aal1847

# THE FUTURE OF PAIN RESEARCH

*By* **Peter Stern** *and* **Leslie Roberts**

It can hit as a piercing jolt, or sometimes as a dull ache that radiates through the body. Pain can be exquisite, or debilitating and chronic, which can destroy one's quality of life.

Pain starts with a stimulus at the periphery of the body, which then travels through neurons to the brain. But the process is far more complex. A number of neuronal circuits are involved in pain transmission. Most of these circuits are plastic and can change when something goes awry. Once we understand these circuits, we can interfere and block them at the right level.

However, pain is not only a matter of neurons. The tissue around them plays an important role, too. Other cells—such as skin, immune, or glia cells, to name just a few—participate in the pathogenesis and also the resolution of pain. Nor is there a simple one-way street from the periphery to the brain. Powerful cognitive processes shape the way that we perceive pain. This perception is determined by our expectations and the situation in which we find ourselves. Clinicians need this knowledge to develop techniques for personalized treatment of chronic pain and to prevent pain from spiraling out of control.

Failed attempts to manage pain have contributed to the opioid epidemic in the United States. Understanding how opioids work, and how in some circumstances they might actually amplify pain, should lead to alternatives. As many states legalize the medical use of marijuana, hints are emerging that cannabis could be one, but research is complicated by strict federal regulations.

**INSIDE**



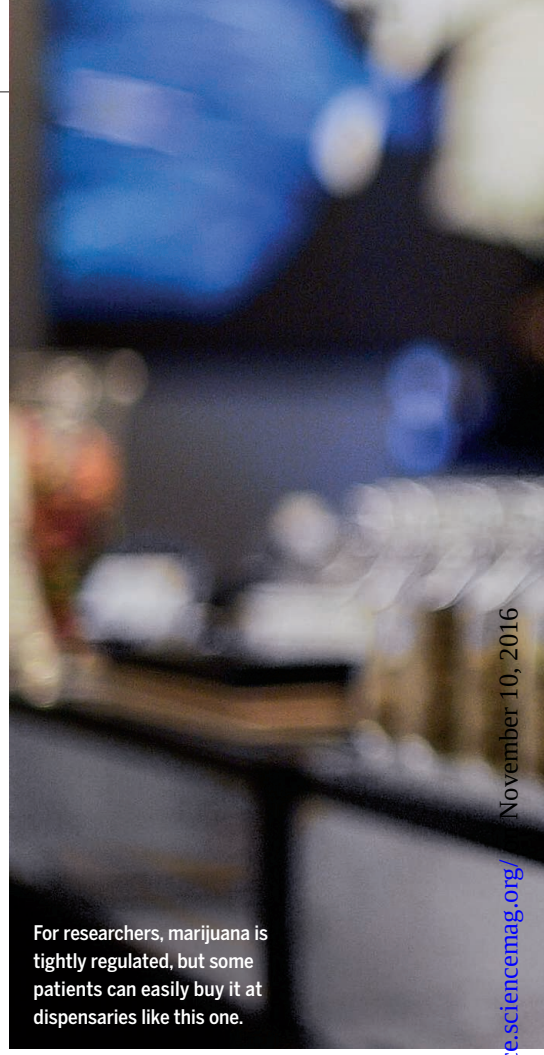
Pain is a subjective experience. People distracted by a suspense-packed movie experience less pain at the dentist's office.

PHOTO: MATTHEW RAKOLA

# POT AND PAIN

## Hints are emerging that cannabis could be an alternative to opioid painkillers

*By* **Greg Miller**



For researchers, marijuana is tightly regulated, but some patients can easily buy it at dispensaries like this one.

PHOTO: KEVORK DJANSEZIAN/GETTY IMAGES

In the mid-19th century, some European doctors became fascinated with a plant-derived drug recently imported from India. Cannabis had been used as medicine for millennia in Asia, and physicians were keen to try it with their patients. No less an authority than Sir John Russell Reynolds, the house physician to Queen Victoria and later president of the Royal College of Physicians in London, extolled the medical virtues of cannabis in *The Lancet* in 1890. "In almost all painful maladies I have found Indian hemp by far the most useful of drugs," Reynolds wrote.

Like other doctors of his day, Reynolds thought cannabis might help reduce the need for opium-based painkillers, with their potential for abuse and overdose. "The bane of many opiates and sedatives is this, that the relief of the moment, the hour, or the day, is purchased at the expense of tomorrow's misery," he wrote. "In no one case to which I have administered Indian hemp, have I witnessed any such results."

More than 125 years later, the misery caused by opioids is clearer than ever, and there are new hints that cannabis could be a viable alternative. Some clinical studies suggest that the plant may have medical value, especially for difficult-to-treat pain conditions. The liberalization of marijuana laws in the United States has also allowed researchers to compare overdoses from painkiller prescriptions and opioids in states that permit medical marijuana versus those that don't. Yet following up on those hints isn't easy. Clinical studies face additional hurdles because the plant is listed on Schedule I, the U.S. Drug Enforcement Administration's (DEA's) list of the most dangerous drugs.

Some researchers worry that rigorous research is being outpaced by informal experimentation, as millions of people with access to medical marijuana treat themselves. "It's clear that the policy has gone way out in front of the science in terms of allowing access to products that haven't been through the standard clinical trials process," says Mark Ware, a pain specialist at McGill University in Montreal, Canada.

**NEARLY 2 MILLION AMERICANS** were addicted to or abusing prescription opioid drugs in 2014, according to the Centers for Disease Control and Prevention, and the Kaiser Family Foundation estimates that more than 21,000 died from overdoses. That same year, a study published in *JAMA Internal Medicine* hinted that medical marijuana could make a dent in that alarming toll. The researchers, led by Marcus Bachhuber, then at the Philadelphia Veterans Affairs Medical Center in Pennsylvania, examined death certificates in all 50 states between 1999 and 2010. They found that the annual rate of deaths due to overdose on an opioid painkiller was nearly 25% lower in states that permitted medical marijuana. In 2010, that translated into 1729 fewer deaths in those states. The researchers also found that the effect grew stronger in the 5 to 6 years after the states approved medical marijuana.

More recently, David Bradford, a health economist at the University of Georgia in Athens, and his daughter Ashley, a master's student there, sought to investigate whether marijuana was supplanting conventional drugs in states where it's legal. Analyzing Medicare drug prescription data from 2010 to 2013, they found a significant difference in the number of prescriptions for several conditions, including anxiety and nausea, in states with medical marijuana. But one condition stood out from the rest: "The effect for pain was three to four times larger than all of the others," David Bradford says. In medical marijuana states, each physician prescribed an average of 1826 fewer doses of conventional pain medication each year, they reported in the July issue of *Health Affairs*. That translates into many millions of doses per year in those states.

The Bradfords haven't yet analyzed how many of those doses were opioid drugs versus other painkillers, but David Bradford suspects it's a large chunk. "It's suggestive evidence that medical marijuana might help divert people away from the path where they would start using [an opioid drug], and of course if they don't start, they're not on that path to misuse and abuse and potentially death."

In a follow-up study, the Bradfords analyzed prescription data from Medicaid recipients, a younger population than the Medicare enrollees in their previous study. So far, the reduction in pain prescriptions appears to be even more dramatic in this group, David Bradford says.

Additional evidence about whether cannabis can reduce opioid use could come from Canada, which legalized medical marijuana in 2001 and might legalize recreational use as soon as next year. In Quebec, researchers established a patient registry in 2015 to collect demographic data on patients who use medical marijuana, the type and dose they take, and the conditions they're seeking treatment for, along with self-reports on benefits and adverse outcomes. McGill's Ware, who is leading the ef-

fort, says the registry is also collecting data on opioid use. "We'll certainly be looking at whether patients who manage their pain with cannabis can reduce their opioid doses over time or even wean themselves off opioids entirely," he says.

Yet in the United States, where 25 states and Washington, D.C., have legalized medical marijuana, there are no state-wide efforts to collect data on how patients are using cannabis or on whether they have been affected for good or ill, in part because marijuana is still illegal at the federal level. That's a huge missed opportunity, says Ryan Vandrey, a behavioral pharmacologist at Johns Hopkins University in Baltimore, Maryland. "It's mind-boggling that we have millions of people in the U.S. using cannabis for medicine and we not only don't have the proper data to help them take it appropriately, we're not doing a good job of collecting it."

Researchers have good reason to think marijuana might relieve pain. Tetrahydrocannabinol, or THC, the plant's main psychoactive ingredient, binds to a class of receptors on neurons that are involved in mediating pain, appetite, and mood, among other things. "It's working directly on pain pathways in the brain, spinal cord, and periphery," says Ethan Russo, a neurologist and medical director of Phytecs, a Los Angeles, California–based company developing therapies based on compounds isolated from marijuana. Previously, Russo oversaw international clinical trials for Sativex, an oral spray made by GW Pharmaceuticals in Salisbury, U.K., that has been approved in 27 countries for treating spasticity caused by multiple sclerosis and in Canada for certain types of pain. Sativex combines THC with cannabidiol, another compound in marijuana that may counteract the anxiety and cognitive side effects associated with THC and that appears to have anti-inflammatory effects.

But few marijuana-based therapies have gone through clinical trials. A meta-analysis published last year in the *Journal of the American Medical Association* found just 28 randomized clinical trials investigating cannabis for chronic pain. (Sativex accounted for nearly half of them.) The authors concluded there was "moderate quality evidence" to support its use. Part of the reason for the scarcity of cannabis trials is that whole plants and natural extracts aren't patentable, giving pharmaceutical companies little incentive to pursue them.

Recently, however, some states that have legalized medical marijuana have begun to fund clinical studies. California, which in 1996 became the first U.S. state to legalize medical marijuana, led the way with its Center for Medicinal Cannabis Research, which has done several placebo-controlled studies on pain. Barth Wilsey, a pain management physician at the University of California in San Diego, led two of them. The first, published in 2008, found that smoking marijuana reduced pain caused by nerve damage in 38 patients, with minimal side effects. The second, published in 2013, found that vaporized cannabis, even in low doses, relieved pain in a similar group of patients who hadn't responded to traditional medications, including opioid analgesics.

A trial just getting underway at the University of Colorado (CU) Anschutz Medical Campus in Aurora will be the first to directly compare cannabis and opioid painkillers in patients with back and neck pain. "There's definitely emerging evidence in the literature for [using cannabis to treat] neuropathic pain, but there's hardly anything for chronic back and neck pain, which is one of the most common reasons people go see their doctor," says neurobiologist Emily Lindley, who will run the CU study.

Hers is one of nine medical marijuana re-

search grants funded so far by the state of Colorado with a total of $9 million from tax collected on marijuana sales. The impetus for the study was a survey done a few years ago at CU Hospital's Spine Center. Nearly one-fifth of the 184 patients with chronic back and neck pain who responded to the survey reported using marijuana to treat their pain. Of those, 86% reported that it "moderately" or "very much" relieved their pain, and 77% said marijuana provided as much or more relief than their opioid prescription pain-killers. "We expected to see some positive effects regarding pain control but not quite to that extent and not with that many patients," says Vikas Patel, chief of orthopedic spine surgery at CU.

Now, Lindley's study will enroll 50 patients with back and neck pain, who will visit the university three times and receive either vaporized cannabis, the opioid drug oxycodone, or a placebo. (In the case of cannabis, the placebo is marijuana with the THC chemically extracted; for oxycodone, the placebo is a pill.) At each visit, the patients will be given a battery of tests to assess their pain levels and look for side effects like impairments of memory, attention, and concentration.

**BUT SUCH RESEARCH** faces regulatory obstacles, because DEA still classes marijuana as a Schedule I drug: the most dangerous drugs with no known medical benefits. It has taken Lindley nearly 2 years from the time she received her grant to start her study. Getting the required Schedule I license from DEA took about 6 months. Prior to that, the university spent tens of thousands of dollars to install secure narcotics cabinets to meet DEA's requirements and a new ventilation system to comply with its own no smoking policy.

In August, DEA rejected two petitions to remove marijuana from Schedule I. The decision was made after a scientific review by the Food and Drug Administration (FDA) concluded that the evidence for the medical benefits of marijuana did not meet their standards for new drug approval.

FDA noted that most cannabis studies to date have been fairly small—with a few dozen participants, not hundreds—and they've followed patients for a few hours, not the 12 weeks or more that's typical in the clinical trials pharmaceutical companies conduct. Another complication is the variation in how cannabis is delivered. Patients in many early studies smoked it, and people ingest varying amounts of THC per puff. Newer delivery systems, such as vaporizers and edible products, add still more uncertainty about the doses patients actually receive. Then there's the natural variation in the concentration of THC and other cannabinoids in different strains of marijuana.

Even scientists who are bullish on the potential for medical marijuana acknowledge that consistent dosing is an issue. Yet many researchers see the situation as a Catch-22: The Schedule I listing and other restrictions on marijuana research hinder the type of studies that are needed to convince regulators to loosen those restrictions.

Two bills introduced in Congress this year aim to lower some of these hurdles. The bills would limit the time that DEA spends reviewing proposed research studies (just as FDA has 30 days to review drug studies). They would also restrict DEA's role in making sure listed drugs are stored securely. Now, DEA also has to weigh in on changes of scientific protocol, and that can really slow things down, says Vandrey of Johns Hopkins, who is collaborating on a study there to compare the analgesic effects of cannabis and the opioid drug hydrocodone in healthy subjects.

A third bill, introduced in July, aims to ease research with cannabidiol and other chemical components of marijuana. "The current interpretation [of the Controlled Substances Act] is that anything in the plant is Schedule I," Vandrey says. Even though there is no evidence that cannabidiol is prone to abuse, researchers interested in studying it have to jump through the same hoops as if their study involved whole-leaf marijuana. "That, in my mind, is just silly," Vandrey says. Research on terpenes, still another group of cannabis compounds that may have analgesic effects, faces the same hurdles.

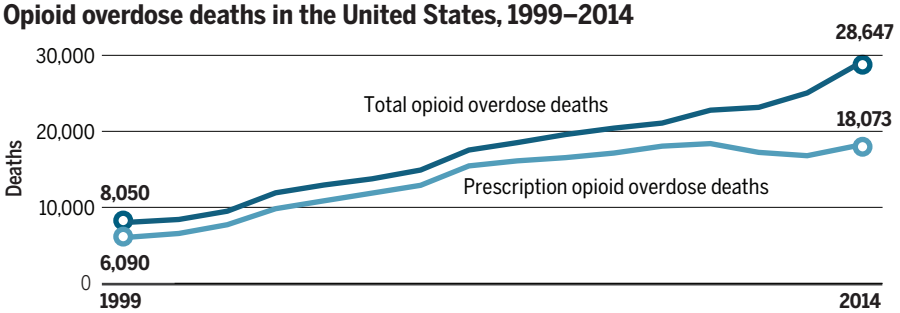
"With changing attitudes and changing policy, I'm hopeful that research can proceed with fewer barriers," McGill's Ware says. He and others hope they'll soon be able to firm up the case for marijuana as an effective pain treatment. "I'd hate to think we're still asking the same questions 10 years from now," Ware says. ■

*Greg Miller is a science and technology journalist in Portland, Oregon.*

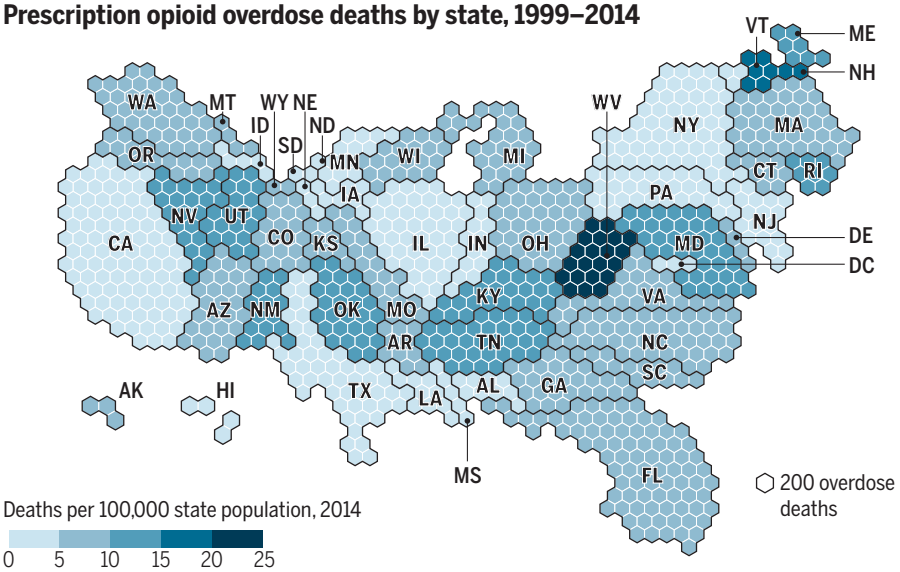
## A rising toll

Statistics from the Kaiser Family Foundation show a growing epidemic of opioid abuse in the United States over the last 15 years. In the cartogram below, the size of each state reflects the total number of prescription opioid overdose deaths from 1999–2014. Darker colors indicate more overdose deaths relative to the state's population in 2014.

### Opioid overdose deaths in the United States, 1999–2014




### Prescription opioid overdose deaths by state, 1999–2014




CREDITS: (GRAPHIC) J. YOU/*SCIENCE*; (DATA) KAISER FAMILY FOUNDATION/U.S. CENSUS BUREAU

Cold pressor tests estimate a person's pain threshold based on how long they can tolerate ice-cold water.



# PRIMED FOR PAIN

## Do opioids make chronic users more sensitive to pain?

*By* **Kelly Servick**; *Photography by* **Matthew Rakola**

Mark Hutchinson could read the anguish on the participants' faces in seconds. As a graduate student at the University of Adelaide in Australia in the late 1990s, he helped with studies in which people taking methadone to treat opioid addiction tested their pain tolerance by dunking a forearm in ice water. Healthy controls typically managed to stand the cold for roughly a minute. Hutchinson himself, "the young, cocky, Aussie bloke chucking my arm in the water," lasted more than 2 minutes. But the methadone patients averaged only about 15 seconds.

"These aren't wimps. These people are injecting all sorts of crazy crap into their arms. … But they were finding this excruciating," Hutchinson says. "It just fascinated me." The participants were taking enormous doses of narcotics. How could they experience such exaggerated pain?

The experiment was Hutchinson's first encounter with a perplexing phenomenon called opioid-induced hyperalgesia (OIH). At high doses, opioid painkillers actually seem to amplify pain by changing signaling in the central nervous system, making the body generally more sensitive to painful stimuli. "Just imagine if all the diabetic medications, instead of decreasing blood sugar, increased blood sugar," says Jianren Mao, a physician and pain researcher at Massachusetts General Hospital in Boston who has studied hyperalgesia in rodents and people for more than 20 years.

But how prevalent hyperalgesia is, and whether it plays a role in the U.S. epidemic of opioid abuse and overdose, is unclear. A lack of reliable testing methods and a series of contradictory papers have created believers and skeptics. A few researchers, like Mao, think hyperalgesia is an underappreciated puzzle piece in the opioid epidemic—a force that can pile on pain, drive up doses, and make it harder for chronic users to come off their drugs. Some of those

researchers are looking for ways to turn down hyperalgesia, to help patients function on lower doses of their oxycodone, for example, or make it easier to taper off it altogether. Others see OIH as an oddity in the literature—real, and a powerful clue to the workings of pain pathways, but unlikely to tighten the grip of opioids on most patients. Hutchinson thinks the majority of physicians are either unaware of hyperalgesia or unconvinced of its importance. "I think if you surveyed prescribers of opioids, they would be divided probably 60–40."

**PARADOXICAL AS IT MAY SEEM**, OIH makes evolutionary sense. "Nature didn't come up with pain just to torture mankind," says Martin Angst, an anesthesiologist and clinical pharmacologist at Stanford University in Palo Alto, California. Pain causes us to recoil from a hot stove and to stay off an injured leg while it heals. And when it's crucial that we temporarily ignore pain—say, when we run on that injured leg to evade a charging lion—the body has a way of numbing it, in part by releasing its own opioids. These natural molecules bind to receptors on neurons to block pain signals and activate reward centers in the brain.

But doses of prescription opioids are orders of magnitude higher than our endogenous levels, Angst says. Confronted by these, "your biology fights back and says, 'I'm blindfolded to pain by all these chemicals. I need to be able to sense pain again.'"

Some studies of hyperalgesia rely on gradually heated probes applied to the skin.

Mao was among the first to delve into potential mechanisms of OIH in an animal model. In 1994, while at Virginia Commonwealth University in Richmond, he and his colleagues showed that after 8 days of spinal morphine injections, rats were quicker to pull their paws away from a gradually heated glass surface. The animals' baseline pain threshold had changed, and the effect was something more than tolerance, in which the body requires increasing doses of a drug to get the same effect. In this case, a higher dose could actually increase sensitivity to pain.

The researchers found they could reverse the hyperalgesic effect by blocking certain receptors on neurons in the animals' spinal cord. These N-methyl-D-aspartate (NMDA) receptors pick up chemical signals—notably an excitatory molecule called glutamate—released by sensory neurons projecting from the skin and organs, and transmit pain signals up to the brain. Researchers already knew that even without opioids, some people with chronic pain from nerve damage or fibromyalgia, for example, experience hyperalgesia when normal pain signaling gets reinforced and amplified over time. It appeared that, at least in animals, opioids had a similar effect.

By 2000, Mao was turning his attention to patients, and the population of opioid users was expanding. Doctors had begun to consider the drugs relatively safe options for managing chronic pain. With the release and aggressive marketing of the long-acting narcotic OxyContin in 1996, a class of drugs that had largely been reserved for cancer patients was becoming a go-to treatment for conditions such as lower back pain.

As prescribing skyrocketed, so did overdoses. U.S. deaths from prescription opioids have roughly quadrupled in the last 2 decades, reaching 21,000 in 2014. Making things worse, abundant prescription opioids have been diverted for recreational use, which has driven up rates of heroin addiction as users have sought cheaper or more accessible alternatives. Both prescription and illegal opioids kill when high doses slow breathing, especially when combined with alcohol or antianxiety drugs called benzodiazepines.

"I'm not sure you could find an example of physicians doing more harm to human beings than we have achieved in our liberal opiate prescribing," says David Clark, an anesthesiologist at Stanford.

Mao and others wondered whether hyperalgesia was another important opioid side effect. People might be seeking a higher dose as drug-induced pain compounded the original pain, he thought. If so, doctors who ignore hyperalgesia might bump up the dose when the right decision was to reduce it. And when a patient tried to taper off a drug, a temporarily lowered pain threshold might make it harder for them to manage without it. "If they're hyperalgesic, they can just go back to the drug again to feel okay," says Jose Moron-Concepcion, a neuroscientist at the Washington University School of Medicine in St. Louis in Missouri.

The evidence for hyperalgesia is clearest in people taking extreme doses—for instance, in opioid abusers or terminal cancer patients managing severe pain. Surgical patients given large amounts of the opioid remifentanil have shown signs of hyperalgesia; they have larger areas of soreness around their wounds and seem predisposed to chronic pain following surgery. But what about patients who take lower doses of opioids daily over months or years to manage chronic pain? As a pain specialist at a large teaching hospital, Mao frequently encounters patients who can't find relief from increasing opioid doses and who tell him that their pain has become worse—diffuse, nagging, and harder to pinpoint.

But just how many people experience OIH, and at what opioid dose, is hard to say. The phenomenon can be very hard to distinguish from tolerance, when pain increases as the drug loses its effectiveness over time. (It's also possible that a patient's underlying condition has changed, or that the chronic pain itself has kicked their pain signaling into high gear.)

Because diagnosing hyperalgesia can be a guessing game in the clinic, some researchers have turned to the lab. They have tried to document changing pain thresholds with quantitative sensory tests, like the so-called cold pressor test Hutchinson witnessed in the methadone patients in Australia, or contraptions that apply heat or pressure to the skin. But the studies have been small and the results inconsistent. "Nobody has actually shown that that particular stimulus in a human being is a valid way to say, 'Yes, this person has become hyperalgesic,'" Angst says.

PHOTO: MATTHEW RAKOLA

In 2006, for instance, a team that included Angst and Clark gave the cold pressor test to six people with chronic lower back pain before and after a monthlong course of morphine pills. After the drug treatment, the team found signs of hyperalgesia: On average, the subjects registered pain from the ice water about 2 seconds earlier, and removed their hands about 8 seconds earlier, than they had beforehand. But those results didn't hold up in a larger group of 139 patients randomized to take opioids or placebo, nor did they appear in a different pain test that applied a gradually heated probe to the forearm. Then in 2013, a study with a different methodology seemed to confirm the effect. A research team in Israel reported evidence of hyperalgesia in 17 of 30 patients with radiating spinal nerve pain by asking them to rate the intensity of heat pain on a numerical scale before and after a 4-week course of hydromorphone.

If you can't reliably diagnose hyperalgesia, it's hard to predict its long-term effects, says Michael Hooten, an anesthesiologist at the Mayo Medical School in Rochester, Minnesota. His group found evidence in 91 patients tapering off opioids that those whose doses were higher at the start, forcing them to make greater reductions over the 3-week program, had worse measures of heat pain hyperalgesia. But the team wasn't able to track these patients long-term to ask the bigger questions: How long until their pain thresholds bounced back to normal? Do hyperalgesic patients who manage to quit taking opioids ultimately see improvements in pain? Are hyperalgesic patients more or less prone to addiction or relapse?

For some, this lack of evidence makes research into hyperalgesia look like a dead end. "When I go to work every day, I don't think about opioid-induced hyperalgesia," says Gary Bennett, a pain researcher at the University of California in San Diego. "We know that it's real. We don't know how important it is, and it's really, really hard to answer that question, so let's move on."

**MAO ISN'T READY** to move on. He believes the risk of hyperalgesia should motivate doctors to try tapering patients off their opioids when their pain worsens without an obvious cause. But in his experience, only about a third of chronic pain patients are willing to try that. So he's hoping for a different solution: a drug that targets the mechanisms behind hyperalgesia and that might be given alongside an opioid, either when it's first prescribed or when a doctor suspects OIH.

Mao is recruiting patients for clinical trials to test two candidate drugs. One is ketamine, an anesthetic that blocks NMDA receptors. The other, guanfacine, is currently used to treat high blood pressure and is thought to keep sensory neurons from releasing glutamate into the spinal cord. A team led by Peggy Compton of Georgetown University in Washington, D.C., meanwhile, is investigating a pain and antiseizure drug called gabapentin that may block neural transmission to reduce excessive pain signals.

Other groups are attacking opioid side effects, including hyperalgesia, from a very different angle. In the early 2000s, researchers began exploring the role of glia, star-shaped immune cells in the brain and spinal cord, which were traditionally thought to function as mere "housekeepers," offering structural support for neurons and removing debris. But when the immune system becomes activated in response to an illness or injury, glia in regions associated with pain processing seem to take on another role: They release inflammatory molecules that interact with nearby neurons to amplify pain signals.

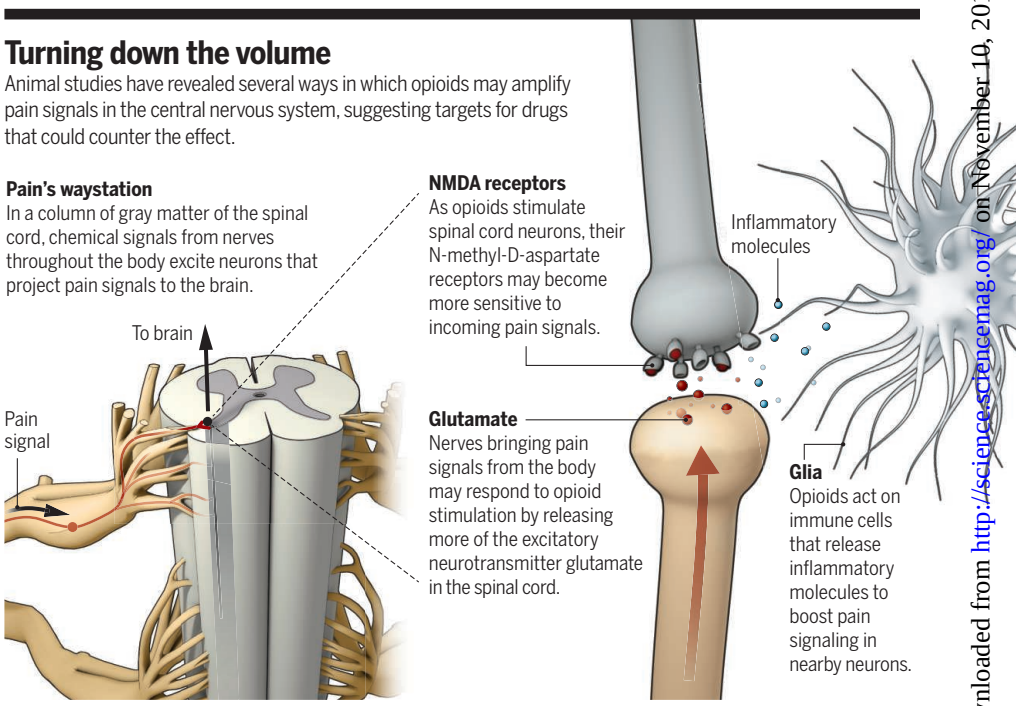
In 2001, researchers at the Chinese Academy of Sciences in Shanghai reported that chronic morphine administration in rats activated glial cells called astrocytes in the spinal cord. Subsequent studies showed that inhibiting the inflammatory molecules released by glia could reverse hyperalgesia and tolerance in the rats. The results suggested that opioids may trigger glia to set off system-wide pain signaling that both counteracts the pain relief from the drug and makes the body generally more sensitive to pain.

Many see dampening this inflammatory response as a promising way to fight hyperalgesia, because it would not interfere with opioids' pain-relieving activity on neural receptors. Several efforts are underway. The San Diego, California–based biotech company MediciNova recently completed a phase II trial of a glia-inhibiting drug called ibudilast, already approved as an asthma treatment in Japan, to relieve pain and treat withdrawal in opioid abusers. A study led by researchers at Yale University is testing the antibiotic acne medication minocycline, which is also thought to block glial activation in the brain. And research spun out of neuroscientist Linda Watkins's group at the University of Colorado in Boulder is testing a new pain drug that may tame glia in the spinal cord by blocking a signaling protein on their surface.

If inflammation turns out to be a key driver of OIH, it might also point the way to a better test for the effect, says Lesley Colvin, a pain researcher at the University of Edinburgh. Markers of inflammation in the blood might correlate with clinical signs of hyperalgesia or declining pain thresholds on sensory tests. Colvin says she already sees strong evidence of hyperalgesia in high-dose opioid users at the clinic where she works. With so much at stake, she is eager to understand the phenomenon and how it might affect them long term. "Although it's complicated," she says, "that doesn't mean we shouldn't try and work out the details." ■

## Turning down the volume

Animal studies have revealed several ways in which opioids may amplify pain signals in the central nervous system, suggesting targets for drugs that could counter the effect.




GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REVIEW

# Pain regulation by non-neuronal cells and inflammation

Ru-Rong Ji,[1,2]* Alexander Chamessian,[1] Yu-Qiu Zhang[3]

**Acute pain is protective and a cardinal feature of inflammation. Chronic pain after arthritis, nerve injury, cancer, and chemotherapy is associated with chronic neuroinflammation, a local inflammation in the peripheral or central nervous system. Accumulating evidence suggests that non-neuronal cells such as immune cells, glial cells, keratinocytes, cancer cells, and stem cells play active roles in the pathogenesis and resolution of pain. We review how non-neuronal cells interact with nociceptive neurons by secreting neuroactive signaling molecules that modulate pain. Recent studies also suggest that bacterial infections regulate pain through direct actions on sensory neurons, and specific receptors are present in nociceptors to detect danger signals from infections. We also discuss new therapeutic strategies to control neuroinflammation for the prevention and treatment of chronic pain.**

Clinically, inflammation is characterized by five cardinal signs: rubor (redness), calor (increased heat), tumor (swelling), dolor (pain), and functio laesa (loss of function). Acute inflammation is a protective response involving immune cells, blood vessels, and molecular mediators (inflammatory mediators). The function of inflammation is to eliminate the initial cause of cell injury and initiate tissue repair. Acute pain, also known as nociceptive pain, is a cardinal feature of inflammation. The majority of known inflammatory mediators cause pain by binding to their receptors on nociceptive primary sensory neurons in the peripheral nervous system (PNS) (nociceptors) that innervate injured skin, muscle, and joint tissues (Fig. 1) (*1*–*3*). Once thought to be a passive process, the resolution of acute inflammation is now recognized as a distinct, active process involving specialized pro-resolution mediators (SPMs) such as resolvins, protectins, and maresins, derived from omega-3 unsaturated fatty acids (*2*, *4*), as well as other pro-resolution mechanisms (*5*). Resolvins not only regulate the resolution of acute inflammation but also inhibit inflammatory pain through direct actions on nociceptors through specific receptors; activation of ChemR23 by resolvin E1 potently inhibits transient receptor potential ion channel V1 (TRPV1), a key ion channel for pain transduction (*4*).

In contrast to acute inflammation, chronic inflammation is often detrimental, leading to a host of diseases such as periodontitis, atherosclerosis, rheumatoid arthritis, and even cancer (*2*). It is unclear whether chronic inflammation is also as critical for driving chronic pain as acute inflammation is for acute pain. Pain research in the past several decades has established that neuronal plasticity is a key mechanism for the development and maintenance of chronic pain (*1*, *6*). Peripheral sensitization in nociceptors is essential for the development of chronic pain (*3*) and transition from acute pain to chronic pain (*7*). Central sensitization (enhanced responses of pain circuits in the spinal cord and brain) regulates the chronicity of pain, causes the spread of pain beyond the site of injury, and influences the emotional and affective aspects of pain (*8*).

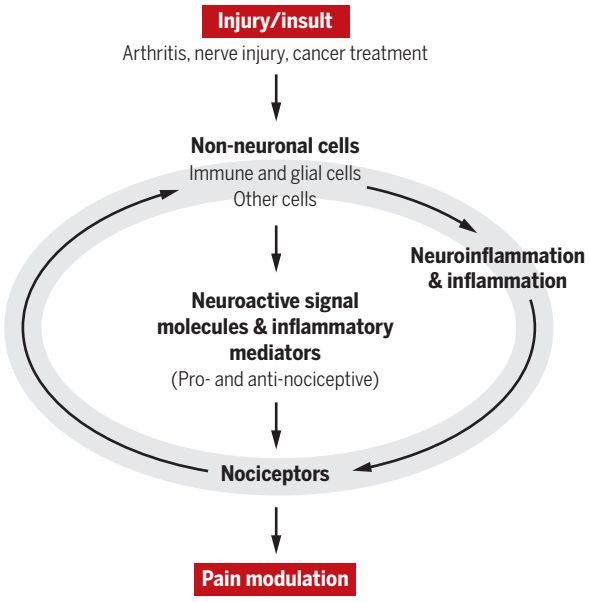



**Fig. 1. Interactions between non-neuronal cells, neurons, and inflammation/neuroinflammation in different pain conditions after injury and insult.** Non-neuronal cells can modulate pain in different directions by producing either pro- or anti-nociceptive mediators.

Neuroinflammation is a localized inflammation occurring in the PNS and central nervous system (CNS) in response to trauma, neurodegeneration, bacterial or viral infection, autoimmunity, and toxins (*2*, *9*). The hallmarks of neuroinflammation are activation and infiltration of leukocytes, activation of glial cells, and increased production of inflammatory mediators. Neuroinflammation is also associated with changes of vascular cells that facilitate leukocyte infiltration (*2*, *9*). Compared with inflammation, neuroinflammation is more persistent in chronic pain conditions and therefore plays a more important role in chronic pain maintenance (*2*). For example, fibromyalgia, a widespread chronic pain syndrome, is associated with small fiber neuropathy and neuroinflammation, although its correlation with systemic inflammation is unclear (*10*).

The interactions between inflammation and pain are bidirectional (Fig. 1). Nociceptive sensory neurons not only respond to immune signals but also directly modulate inflammation. For example, nociceptors express receptors for and respond to cytokines and chemokines and also produce these inflammatory mediators (*11*, *12*). In a process called neurogenic inflammation, noxious stimulation causes nociceptors to release neuropeptides such as Substance P (SP) and calcitonin gene–related peptide (CGRP), leading to the extravasation of fluid and cells from the blood. Consistently, silencing nociceptors reduces allergic airway inflammation (*13*). Nociceptors also serve to dampen and constrain the immune response: Ablation of nociceptors abrogated pain during bacterial infection but concurrently worsened inflammation via CGRP (*14*). Activation of pain circuits also regulates neuroinflammation in the CNS, referred to as neurogenic neuroinflammation in chronic pain and neurodegenerative diseases (*9*).

Numerous non-neuronal cell types influence pain sensation, including immune, glial, epithelial, mesenchymal, cancer, and bacterial cells. In this Review, we focus on non-neuronal cells that interact with nociceptors in distinct anatomical compartments in the PNS and CNS (glial cells) under normal and pathological conditions (Fig. 2). Despite the diversity of these cells, the ways in which they modulate pain are surprisingly consistent. In response to an injury or insult, non-neuronal cells release neuromodulatory substances in close proximity to nociceptors, which either promote or dampen pain depending on the specific identities of the mediators involved (Figs. 1 and 2). Although the details will differ in each case, this general model will serve as a useful lens through which to examine specific non-neuronal cell types in pain.

## Pain modulation by non-neuronal cells

### *Monocytes and macrophages*

Monocytes and macrophages serve three main functions in the immune system: phagocytosis,

[1]Department of Anesthesiology, Duke University Medical Center, Durham, NC 27710, USA. [2]Department of Neurobiology, Duke University Medical Center, Durham, NC 27710, USA. [3]Institute of Neurobiology, Institutes of Brain Science and State Key Laboratory of Medical Neurobiology, Collaborative Innovation Center for Brain Science, Fudan University, Shanghai 200032, China.
*Corresponding author: Email: ru-rong.ji@duke.edu

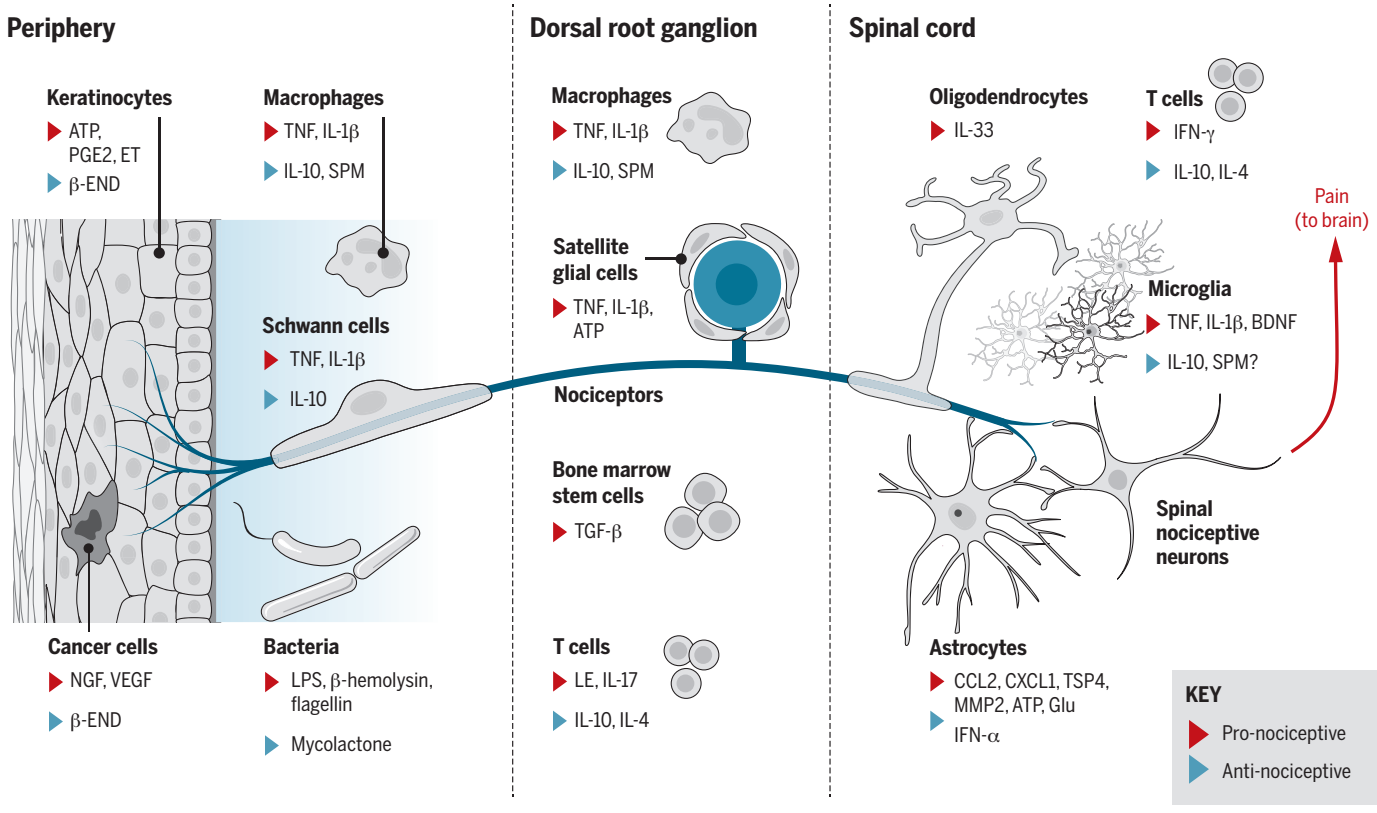
antigen presentation, and cytokine production. Monocytes and macrophages in the periphery play an active role in pain, exhibiting diverse mechanisms that are shaped by the causes and context of pain. In most cases, these cells produce pain through the release of proinflammatory mediators such as tumor necrosis factor (TNF) and interleukin-1β (IL-1β) (*11*), resulting in enhanced pain transduction and conduction via modulation of ion channels such as transient receptor potential ankyrin 1 (TRPA1), TRPV1, and Nav1.7-1.9 (*1*, *2*). Cell-specific depletion of proliferating monocytes and macrophages impairs the development of mechanical and thermal hypersensitivity caused by sterile incision and pathogens, in parallel with a decrement in IL-1β and other pro-algesic mediators at the site of inflammation (*15*). However, in a nerve injury model, deletion of peripheral monocytes does not affect neuropathic pain development (*16*). In a model of chemotherapy-induced neuropathic pain, $CX3CR1^{+}$ monocytes migrate into peripheral nerves and produce reactive oxygen species, which then elicit pain via the activation of TRPA1 (*17*). After peripheral nerve injury, infiltration of monocytes and macrophages to the spinal cord is limited, arguing against a central role of these cells in neuropathic pain (*16*, *18*). Conversely, monocytes and macrophages can also effect analgesia by releasing anti-inflammatory mediators such as IL-10 and SPMs, together promoting the resolution of the initial insult (*2*, *4*). In support of this view, depletion of monocytes and macrophages delayed the resolution of inflammatory pain (*19*). Macrophages have different phenotypes related to their functional states, including proinflammatory M1-like and anti-inflammatory M2-like phenotypes, which may play distinct roles in the induction and resolution of pain.

### *T lymphocytes*

T cells are critical elements of adaptive immunity that have also been implicated in pain, with the most evidence for a role in neuropathic pain. After nerve injury, T cells infiltrate the dorsal root ganglia (DRG) and release the pro-algesic mediator leukocyte elastase (LE), resulting in mechanical allodynia. Consistently, inhibitor of LE reduces nerve injury–induced allodynia (*20*). In the spinal cord, T cell infiltration occurs after nerve injury and is required for the development of mechanical hypersensitivity, as indicated by a reduction of pain behaviors in T cell–deficient *Rag1*-null mice (*21*). A recent report supported the role of spinal T cells in nerve injury–induced neuropathic pain but limited this function only to female mice, with male mice instead depending on microglial signaling for pain. This intriguing finding was attributed to the sexually dimorphic expression of peroxisome proliferator–activated receptors (PPARs) α and γ in T cells (*22*). Different types of T cells play different roles in chronic pain. For example, adoptive transfer of toxic T cells ($CD8^{+}$ T cells) via intrathecal injection enhances neuropathic pain, whereas injection of regulatory T cells ($T_{reg}$ cells) decreases neuropathic pain after chemotherapy (*23*).

### *Keratinocytes*

Keratinocytes are the primary cells of the epidermis. They reside near the peripheral terminals of nociceptors and produce various neuroactive mediators such as adenosine 5′-triphosphate (ATP), IL-1β, prostaglandin E2, endothelin, and nerve growth factor (NGF) that are known to elicit pain, suggesting that these cells can directly activate nociceptors. Photostimulation of channelrhodopsin-expressing keratinocytes is sufficient to generate nocifensive behaviors and evoke action potentials in specific subsets




GRAPHIC: K. SUTLIFF/*SCIENCE*

**Fig. 2. Interactions between distinct parts of a nociceptor with different types of non-neuron cells.** Included are keratinocytes, Schwann cells, satellite glial cells, oligodendrocytes, and astrocytes, as well as immune cells (for example, macrophages and T cells), microglia, cancer cells, and stem cells. These non-neuronal cells produce both pro-nociceptive (highlighted in red) and anti-nociceptive (highlighted in blue) mediators, which can bind their respective receptors on the nociceptor to modulate its sensitivity and excitability. The central terminal of the nociceptor forms a nociceptive synapse with a postsynaptic neuron in the spinal cord dorsal horn to mediate pain transmission in the CNS. END, endorphin; ET, endothelin; Glu, glutamate; IFN-α, interferon-α; LE, leukocyte elastase.

of cutaneous nociceptive afferents (*24*). Keratinocytes are multifunctional, promoting pain or analgesia depending on the condition and context. The familiar experience of sunburn illustrates the dual nature of these cells. Before the pain of sunburn sets in, the feeling of sun on the skin is a pleasurable experience. Here, keratinocytes participate by releasing the endogenous opioid peptide β-endorphin (β-END), which produces both analgesia and reward (*25*). After light overexposure, when inflammation and injury occur, keratinocytes also elicit pain via the activation of TRPV4 and the subsequent release of endothelin onto cutaneous nociceptors (*26*). Although keratinocytes are not immune cells, they can release inflammatory mediators including cytokines (*27*), illustrating how nonimmune cells can also regulate inflammation.

ATP, colony-stimulating factor–1 (CSF1), chemokines (CCL2 and CX3CL1), and proteases, which can originate from injured or activated sensory neurons (Fig. 3). In parallel, expression of the receptors for ATP and CX3CL1 (P2X4, P2X7, P2Y12, and CX3CR1) is increased selectively on spinal microglia in response to nerve injury (*28*, *29*). Activation of these receptors typically converges on an intracellular signaling cascade involving the phosphorylation of p38 mitogen-activated protein (MAP) kinase, which leads to the increased production and release of TNF-α, IL-1β, IL-18, and brain-derived growth factor (BDNF), as well as increased expression of cyclooxygenase (COX) and subsequent synthesis of prostaglandin $E_2$ (*6*). These neuromodulators can then fine-tune both excitatory and inhibitory synaptic transmission, which ultimately enhances pain signal transmission to the brain.

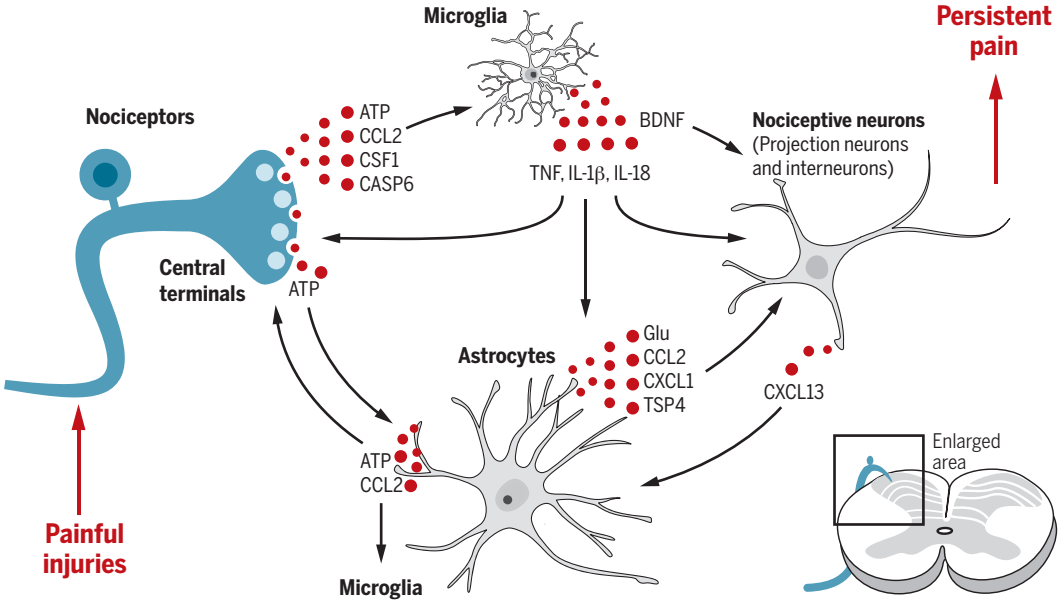



**Fig. 3. Neuron-glial interactions in the spinal cord for the amplification of chronic pain.** Painful injuries such as nerve injury, arthritis, cancer, and treatment (chemotherapy) cause hyperactivity of nociceptors and secretion of glial modulators from their central terminals, leading to the activation of microglia and astrocytes in the spinal cord dorsal horn. Upon activation, microglia and astrocytes secrete neuromodulators to drive chronic pain by inducing synaptic and neuronal plasticity. Pre- and postsynaptic neurons can both "listen" and "talk" to microglia and astrocytes. CASP6, caspase-6.

### *Central glial cells: Microglia, astrocytes, and oligodendrocytes*

Microglia serve as the resident macrophages of the spinal cord and brain, and abundant evidence supports their role in pathological pain (*28*, *29*). A hallmark of microglia is their rapid activation in response to even minor pathological changes in the CNS. Peripheral nerve injury induces a marked proliferation and activation of microglia in parallel with up-regulation of microglial markers such as IBA1 and CD11b in the spinal cord. Indeed, spinal microglia have been strongly implicated in the pathogenesis of neuropathic pain after nerve injury (*18*, *29*–*32*). The signals that activate these microglia include

For example, TNF and IL-1β enhance excitatory synaptic transmission and suppress inhibitory synaptic transmission in spinal cord lamina II neurons (*33*). BDNF disrupts chloride homeostasis and suppresses γ-aminobutyric acid (GABA)–mediated inhibitory synaptic transmission (disinhibition) in spinal lamina I projection neurons (*6*, *34*). Selective activation of spinal microglia by fractalkine (CX3CL1) is sufficient to rapidly facilitate synaptic strength between primary afferent C-fibers and lamina I projection neurons (*35*). Caspase-6 is only expressed in axonal terminals in the spinal cord and mediates neuron-microglial interaction. Recombinant caspase-6 triggers marked TNF release from microglia and drives robust acute inflammatory pain behaviors before the morphological activation of microglia (*36*). Spinal microglia signaling is sex-dependent, and microglial signaling inhibitors such as minocycline and p38 inhibitor reduce neuropathic pain primarily in male mice, with little to no effect on female mice (*22*). Although microglia are critical to the development of chronic pain, they may function in the maintenance of chronic pain as well. For example, examination of microglial enhancers by using chromatin immunoprecipitation sequencing (ChIP-seq) revealed persistent alterations in close proximity to transcriptionally regulated genes, leading to the hypothesis that these changes may contribute to the "memory" recorded at a molecular level (*18*). Microglia are also involved in opioid-induced hyperalgesia through microglia-mediated disruption of neuronal chloride homeostasis (*34*). Morphine paradoxically prolongs neuropathic pain in rats, and selective inhibition of spinal cord microglia in vivo through DREADD (Designer Receptor Exclusively Activated by Designer Drugs) reverses morphine-induced pain sensitization for weeks (*37*). In the future, it will also be important to determine whether microglia are involved in the resolution of chronic pain, which would be expected on the basis of the dual role of their peripheral macrophage counterparts (*19*).

Astrocytes perform numerous critical functions such as neurotransmitter recycling, formation of the blood-brain barrier, regulation of extracellular ion concentration, and modulation of synaptic transmission, among many others. Nerve injury induces myriad changes in astrocytes that lead to enhanced pain. For example, after nerve injury, astrocytes lose their ability to maintain the homeostatic concentrations of extracellular potassium ($K^+$) and glutamate, leading to neuronal hyperexcitability (*28*). Astrocytes can also signal directly to neurons through physically coupled networks mediated by gap junctions to facilitate intercellular transmission. Gap junction communication is mediated by connexin-43 (Cx43), the predominant connexin expressed in astrocytes. Nerve injury induces persistent up-regulation of Cx43 in astrocytes and switches the function of Cx43 from gap junction communication to paracrine modulation (*38*). This paracrine regulation leads to the increased release of glutamate, ATP, and chemokines through a paracrine mechanism. The astrocyte-derived chemokines act as neuromodulators and can potentiate excitatory synaptic transmission in the spinal cord pain circuitry (*38*). Furthermore, nerve injury

GRAPHIC: K. SUTLIFF/*SCIENCE*

up-regulates CXCL13 in spinal cord neurons, which can activate astrocytes via CCR5 to maintain neuropathic pain (*39*). Thus, chemokines facilitate neuropathic pain via bidirectional neuron-astrocyte interactions (Fig. 3). Nerve injury also induces spinal cord and cortical astrocytes to up-regulate thrombospondin-4 (TSP4), which promotes neuropathic pain through the formation of new synapses and rewiring of somatosensory cortical circuits (*40*, *41*). A single human astrocyte may contact more than 1 million synapses, and such complexity points to a more important role of astrocytes in humans that bears further investigation. As compared with microglial activation, astrocyte activation in chronic pain conditions is more persistent, indicating their contribution to the chronicity of pain (*28*).

Oligodendrocytes create the myelin sheath that provides support and insulation to axons in the CNS (Fig. 2). In spite of their ubiquity and importance, only recently has a role for oligodendrocytes been revealed in pain. In the chronic constriction injury model of nerve injury–induced neuropathic pain, oligodendrocyte-derived IL-33 contributes to pain hypersensitivity via MAP kinases and nuclear factor κB (NF-κB) (*42*). Similarly, in post mortem spinal cord samples from HIV patients, expression of oligodendrocyte markers such as NG2, Oligo2, and platelet-derived growth factor receptor α (PDGFRα) increases, reflecting a persistent activation of oligodendrocytes in chronic pain (*43*). Conversely, toxin-mediated ablation of oligodendrocytes induces neuropathic pain symptoms, suggesting a potential protective role of these cells (*44*). These divergent findings suggest that as with other non-neuronal cells, oligodendrocytes play active and context-specific roles in pain.

### *Peripheral glia: Schwann cells and satellite glial cells*

Like their central counterparts, the major glial cells of the PNS, the Schwann cells and the satellite glial cells (SGCs), contribute to pain. In response to painful stimuli, these peripheral glial cells are activated before central glia and release various inflammatory mediators, sensitizing nociceptors at axons (Schwann cells) and cell bodies (SGCs). After nerve injury, activated Schwann cells mediate the breakdown of the blood-nerve barrier via the secretion of matrix metalloproteinase 9 (MMP-9), which promotes the recruitment of immune cells from the vasculature and their subsequent release of more pro-nociceptive mediators (*27*). SGCs surround the somata of DRG neurons and are directly coupled to each other via gap junctions. After nerve injury, SGCs become activated and proliferate (*27*). SGCs contribute to chronic pain sensitization by producing cytokines and MMPs that regulate the cleavage and activation of cytokines (*28*). Nociceptive activity also causes ATP release from neuronal soma to activate P2X7 in SGCs, leading to TNF release from SGCs and subsequent increase in neuronal excitability (*45*). Activation of P2X7 receptors in SGCs also reduces pain through down-regulation of P2X3 receptors in nociceptive neurons (*46*).

### *Stem cells*

Bone marrow stromal cells or bone marrow stem cells (BMSCs) produce many beneficial effects to promote tissue regeneration and tissue repair by secreting growth factors. These cells can also effectively control inflammation and neuroinflammation by secreting anti-inflammatory mediators. The duration of pain relief by BMSCs is remarkable after systemic or local injection (*47*, *48*). A single intrathecal injection of BMSCs inhibits nerve injury–induced neuropathic pain for many weeks via secretion of transforming growth factor–β1 (TGF-β1), a potent anti-inflammatory cytokine (*48*). The analgesic effects of intrathecal BMSCs were abolished by a neutralizing antibody against TGF-β1. Intrathecal BMSCs effectively suppress nerve injury–induced glial activation and neuroinflammation in DRGs and spinal cord. After intrathecal injection, BMSCs migrate to DRG via a chemotaxic signal (CXCL12) triggered by nerve injury (*48*). This paracrine modulation of pain by BMSCs is very different from other stem cell strategies for chronic pain management. For example, implantation of forebrain GABAergic precursor cells into the spinal cord reduced nerve injury–induced neuropathic pain because these precursor cells can differentiate into functional GABAergic neurons in the spinal cord (*49*).

***"...non-neuronal cells can communicate with nociceptive neurons by 'listening' and 'talking' to neurons..."***

### *Cancer cells*

Cancer pain is a complex pain state, involving inflammatory, neuropathic, compressive, and ischemic mechanisms. Cancer pain is commonly caused by tumors that metastasize from distant sites to the bone and by tumors infiltrating the nerve. Cancers generate and secrete algogenic mediators (such as protons, bradykinin, endothelins, prostaglandins, and proteases) that sensitize and activate nociceptors in the cancer microenvironment (*50*). In particular, cancer cells secrete NGF and vascular endothelial growth factor (VEGF), which induce hyperinnervation of pain-mediating nerve fibers in cancer tissues (*50*, *51*). VEGF promotes cancer pain through a nonvascular regulation. VEGF receptor 1 (VEGFR1) is expressed in nociceptors, and tumor-derived VEGF family members increase nociceptor excitability and produce pain hypersensitivity through selective activation of VEGFR1 (*51*). Certain cancer cells, such as Walker-256 rat mammary gland carcinoma cells, also secrete TGF-β. In early stages of cancer, TGF-β acts as a tumor suppressor through antiproliferative and proapoptotic actions. During tumor progression, TGF-β becomes an oncogenic factor and causes angiogenesis, immunosuppression, and proliferation, invasion, and metastasis of cancers. TGF-β1 also promotes bone cancer pain (*52*). Thus, this anti-inflammatory cytokine might have different roles in different pain conditions via different signaling mechanisms. Microglia also play an active and distinct role in bone cancer pain. Spinal microglia contribute to the maintenance of bone cancer pain in female rats (*53*). Bone cancer causes increased ATP secretion in cerebrospinal fluid and up-regulation of P2X7 receptor in spinal microglia. Activation of P2X7 results in increased synthesis and release of IL-18 via p38 MAP kinase. Spinal inhibition of the P2X7/p-38/IL-18 pathway reduces the advanced phase of bone cancer pain and suppresses hyperactivity of spinal-wide dynamic-range neurons (*53*). This result suggests that spinal cord microglia have an active role in late-phase bone cancer pain in female rats. Further investigation is warranted to test sex-dependent regulation of microglial signaling in different animal models and also in different phases of chronic pain. Some cancers such as melanoma are not painful in the majority of patients. It is possible that melanoma may produce some analgesic molecules such as β-endorphin to suppress pain (Fig. 2), which can promote the optimal growth of cancers, dampening the warning function of pain.

### Pain modulation by bacterial infections through direct interactions with sensory neurons

Infections caused by bacterial, fungal, and viral pathogens are commonly associated with pain. Until recently, it was generally believed that infections cause pain in an indirect manner through the intermediating effects of immune cells and the inflammatory substances they secrete (Fig. 4). However, emerging evidence suggests that pathogens can also directly activate nociceptors and elicit pain using a suite of previously unknown and intriguing mechanisms (*14*). For example, the bacterium *Staphylococcus aureus* possesses a pore-forming toxin called α-hemolysin, which enables it to destroy host cells and also to cause pain by directly forming pores in nociceptors, permitting cation influx and action potential firing through metallopeptidase A disintegrin and metalloproteinase domain–containing protein 10 (ADAM10) (*14*). Lipopolysaccharide (LPS) is the best-studied element derived from bacterial cell walls and is also known to produce pain hypersensitivity by sensitizing TRPV1 in nociceptors (*54*). Bacteria also produce pain by secreting substances that are ligands for receptors on nociceptors, as is the case with N-formyl peptides, which are metabolic by-products of bacteria that activate nociceptors and elicit pain by binding to the formyl peptide receptor 1 (FPR1), a G protein–coupled receptor. The pain produced

through these mechanisms is independent of the immune response (*14*). Bacterial infection is not always painful, however. *Mycobacterium ulcerans*, the etiological agent of Buruli ulcer, causes extensive skin lesions but not pain. *M. ulcerans* produces remarkable analgesia, through a specific activation of sensory neurons triggered by the secreted mycobacterial polyketide mycolactone. Mycolactone elicits analgesia through activation of type 2 angiotensin II receptors (AT2Rs), leading to the activation of K2P potassium channels TRAKK (KCNK4) and hyperpolarization of nociceptors (*55*).

Although viruses are not cells, viral infections are frequently associated with inflammation and pain, such as in sore throat. Several types of viruses—such as herpes simplex virus 1 (HSV-1), HSV-2, and VZV (varicella zoster virus)—can infect sensory neurons to produce painful syndromes. (pathogen-activated molecular patterns) but also by DAMP (danger-activated molecular patterns), the endogenous ligands released after tissue injury. Activation of TLR4 in spinal microglia and astrocytes critically contributes to the development and maintenance of inflammatory and neuropathic pain, as well as opioid-induced hyperalgesia, by eliciting the synthesis of TNF-α and IL-1β (*37*, *56*, *57*). Secretion of these cytokines interacts with nociceptive neurons in the spinal cord pain circuitry (Fig. 4).

Primary sensory neurons in the DRG and trigeminal ganglia also express TLRs (*54*, *58*–*60*). TLRs are differentially expressed in primary sensory neurons and modulate different sensory functions. TLR4 appears to be functionally coupled with TRPV1 in nociceptors, so that LPS binding to TLR4 increased $Ca^{2+}$ influx in sensory neurons (*54*, *60*). In contrast to its intracellular localization in immune cells, TLR7 expresses on the cell surface as well as in axons of DRG nociceptors. Activation of TLR7 by synthetic ligands causes immediate activation and excitation of nociceptors (*61*). MicroRNA (miRNA) let-7b is an endogenous ligand of TLR7 and activates TLR7 through the GUUGUGU motif. Let-7b causes rapid activation and excitation of nociceptors through the coupling of TLR7 and TRPA1. Consistently, intraplantar injection of let-7b elicits rapid spontaneous pain through TLR7 and TRPA1 (*58*).

TLR5 is expressed by large-sized A-fibers, especially Aβ fibers that normally mediate touch sensation. The canonical ligand for TLR5 is flagellin, a globular protein that forms the filament in bacterial flagella. In large A-fibers, activation of TLR5 by flagellin permits the rapid transport of certain small molecules across the neuronal membrane through a to-be-determined mechanism. This Aβ-fiber–specific transport process was exploited to deliver the membrane-impermeable anesthetic QX-314, resulting in selective blockade of A-fibers. Silencing A-fibers in this manner revealed that these neurons inhibit pain under normal conditions but give rise to mechanical allodynia in chemotherapy-induced neuropathy (*59*), in support of the "Gate Control Theory." Further studies are warranted to investigate the ion channels and signaling molecules that give rise to this previously unknown function of TLR5.

Whereas coupling of TLRs with ion channels in nociceptors is important for the induction of acute pain, activation of conventional TLR signaling in nociceptors through myeloid differentiation factor 88 (MyD88), a key downstream signaling molecule for TLRs, is required for the development of chronic pain after painful insults (*23*, *58*). Deletion of MyD88 in nociceptors results in deficits in neuropathic pain as well as in innate and adaptive immunity in the PNS after chemotherapy-induced neuropathy (*23*). Thus, nociceptors share similar features with immune cells by expressing TLRs, cytokines, chemokines, and inflammatory cascades (*62*). Through the coupling with ion channels, nociceptor TLRs act as a distinct class of "danger receptors" for the rapid detection of PAMP and DAMP. Thus, primary sensory neurons not only sense pain, itch, temperature, and touch (*63*) but also sense "danger" (DAMP and PAMP, which are partially overlapped with algogenic and pruritogenic agents), and this danger sensation will trigger protective responses such as a rapid

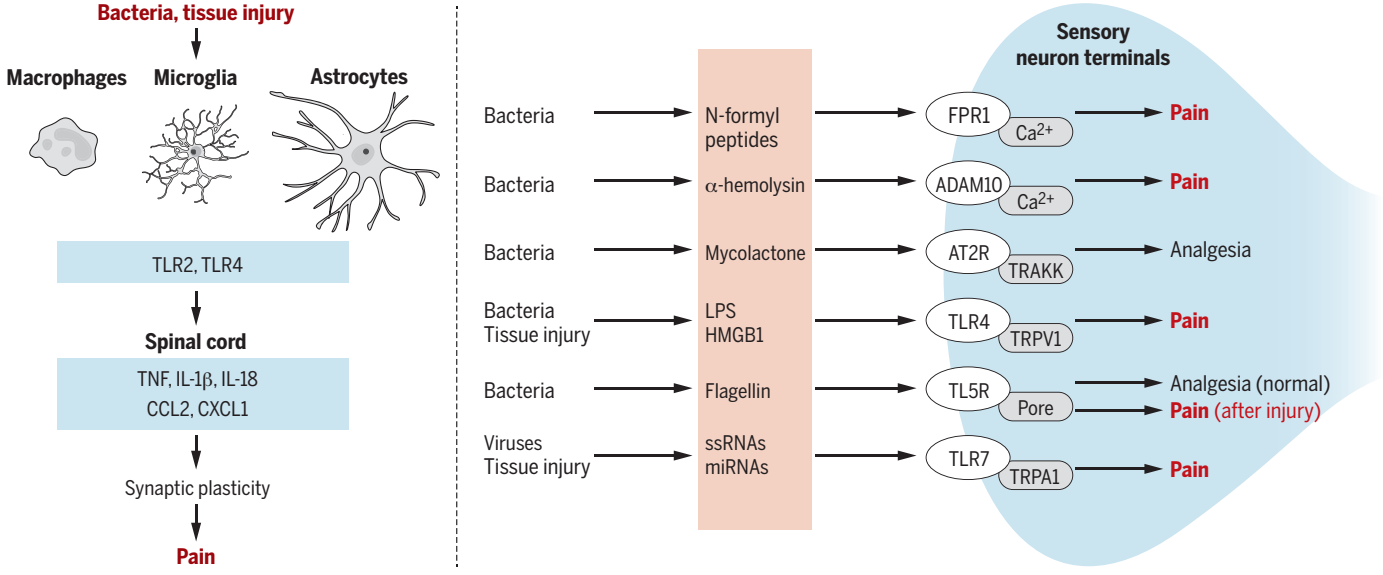



**Fig. 4. Infections and tissue injury regulate pain via both neuronal and non-neuronal mechanisms through TLRs.** Bacterial infections can modulate pain through direct interactions with specific receptors such as FPR and TLR on primary sensory neurons, including nociceptors and mechanoceptors, leading to increased and decreased pain sensitivity. TLRs are expressed by neurons, glial cells, and immune cells and are activated by PAMP (infections), as well as DAMP, endogenous TLR ligands (such as HMGB1 or miRNAs) released after tissue injury. HMGB1, high mobility group box 1 protein; ssRNA, single-stranded RNA.

Shingles, also called herpes zoster, is a painful skin rash and usually appears in a defined region of face or body that corresponds to dermatomes. It is most common in older adults and people who have weak immune systems. Viral infections may activate Toll-like receptors (TLRs) in nociceptors to modulate pain and neuroinflammation.

## Toll-like receptors mediate bilateral neuron-immune and neuron-glial interactions

Mounting evidence suggests that TLRs are instrumental in bilateral interactions between neurons and non-neuronal cells (Fig. 4). TLRs are typically expressed by immune and glial cells to regulate innate immunity through the activation of the NF-κB and MAP kinase signaling pathways. TLRs are activated not only by PAMP

withdrawal or scratching response (*61*) as well as delayed innate and adaptive immunity responses through MyD88 signaling in sensory neurons (*23*). This evolutionarily conserved signaling pathway persists because of its critical biological function, but excessive activity of the same pathway can also lead to autoimmune diseases if left unchecked.

## Concluding remarks and future directions

Chronic pain is a rising health concern worldwide, affecting up to 30% of adults. Despite this enormous burden, the current treatments for chronic pain are inadequate. Chronic pain typically includes inflammatory pain after tissue injury, cancer pain, and neuropathic pain after nerve injury as a result of diabetic neuropathy, viral infection, or major surgeries. Common treatments, such as chemotherapy and chronic opioid treatment, also cause chronic pain. In the past decade, great progress has been made to demonstrate the critical roles of glial cells and immune cells—as well as keratinocytes, cancer cells, and stem cells—in the pathogenesis and resolution of chronic pain. These non-neuronal cells can communicate with nociceptive neurons by "listening" and "talking" to neurons, modifying their activities by secreting neuroactive signaling molecules that can be both pro- and anti-nociceptive (Figs. 1 to 3). There is increasing appreciation that many inflammatory mediators such as cytokines and chemokines are indeed neuromodulators in the PNS and CNS (*2*). Thus, chronic pain could be caused by both "neuropathy" and "gliopathy" (*28*), as well as dysregulations of other non-neuronal cells (Fig. 2). We extend the boundary of non-neuronal cells to include immune cells, cancer cells, and stem cells because they can infiltrate damaged or normal tissues to interact with nociceptors and also produce neuroactive modulators to alter pain states. Given the close link between inflammation and pain, we discussed non-neuronal cells in the context of inflammation and especially neuroinflammation (*2*) that is tightly associated with chronic pain. Inflammation and neuroinflammation can be modulated not only by immune cells but also by other non-neuronal cells such as keratinocytes, stem cells, and cancer cells and even by nociceptive neurons in the PNS and CNS. Infections by bacteria and other pathogens not only affect immune and glial cells but also have a direct impact on sensory neurons, eliciting rapid nociceptive responses through activation of neuronal receptors such as FRP1 and TLRs. In particular, TLRs in immune or glial cells and sensory neurons not only sense the danger signals from tissue injury and infections but also regulate bilateral interactions between neurons and neuron-neuronal cells, and these interactions are important for the induction and chronicity of inflammation and pain.

Although several types of non-neuronal cells contribute to the pathogenesis of pain, deletion of a cell type entirely could be detrimental. We should bear in mind that these non-neuronal cells, except for cancer cells, are protective in the normal conditions but dysregulated in pathological conditions. Although blocking their function may help pain relief for a short period, restoring their normal function through pro-resolution approaches is probably the best way. On the basis of this principle, we propose the following therapeutic strategies as future directions. First, different pharmaceutical strategies should be considered at different development stages of pain and inflammation. Anti-inflammatory drugs such as inhibitors of TNF are more effective in the early stage of pain and inflammation. Resolution of inflammation is a new therapeutic frontier (*5*), and SPMs such as resolvins and protectins may promote the resolution of inflammation and pain, which can be partially achieved via dietary control (*2*). In particular, neuroprotectin D1 is effective in controlling neuroinflammation and neuropathic pain (*64*). Second, cell therapies such as implantation of BMSCs are promising, demonstrating long-term pain relief and even final resolution of pain (*47*, *48*). However, only a very small percentage of cells in bone marrow are stem cells for autologous transplantation. Instead, macrophages are an abundant cell source in blood, and implantation of resident and M2-like macrophages could be an alternative to promote the resolution of inflammation and pain. Alternatively, inducible pluripotent stem cells (iPSCs) from abundant cells (such as skin or blood) could be transformed into the desired cell type and then transplanted. Moreover, transcription activator–like effector nuclease (TALEN)– or clustered regularly interspaced short palindromic repeats (CRISPR)–mediated gene editing ex vivo could first be used to enhance the production of pro-resolution mediators and the migration capability of the cells to be transplanted so that they can target the damaged tissue more accurately. Last, neuromodulation such as vagus nerve stimulation and acupuncture has been used for inflammation control and pain treatment (*65*, *66*). In addition to local acute analgesia via adenosine from non-neuronal cells (*67*), acupuncture and sciatic nerve activation also regulate inflammation through vagus nerve activation and dopamine release (*68*). It will be of great interest to test whether neuromodulation can also help to restore the normal function of non-neuronal cells in chronic pain as a non-pharmaceutical approach.

## REFERENCES AND NOTES


1. A. I. Basbaum, D. M. Bautista, G. Scherrer, D. Julius, *Cell* **139**, 267–284 (2009).
2. R. R. Ji, Z. Z. Xu, Y. J. Gao, *Nat. Rev. Drug Discov.* **13**, 533–548 (2014).
3. M. S. Gold, G. F. Gebhart, *Nat. Med.* **16**, 1248–1257 (2010).
4. Z. Z. Xu *et al.*, *Nat. Med.* **16**, 592–597, 1p, 597 (2010).
5. J. N. Fullerton, D. W. Gilroy, *Nat. Rev. Drug Discov.* **15**, 551–567 (2016).
6. J. A. Coull *et al.*, *Nature* **438**, 1017–1021 (2005).
7. D. B. Reichling, J. D. Levine, *Trends Neurosci.* **32**, 611–618 (2009).
8. C. J. Woolf, M. W. Salter, *Science* **288**, 1765–1768 (2000).
9. D. N. Xanthos, J. Sandkühler, *Nat. Rev. Neurosci.* **15**, 43–53 (2014).
10. N. Üçeyler *et al.*, *Brain* **136**, 1857–1867 (2013).
11. M. Zelenka, M. Schäfers, C. Sommer, *Pain* **116**, 257–263 (2005).
12. F. A. White *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14092–14097 (2005).
13. S. Talbot *et al.*, *Neuron* **87**, 341–354 (2015).
14. I. M. Chiu *et al.*, *Nature* **501**, 52–57 (2013).
15. N. Ghasemlou, I. M. Chiu, J. P. Julien, C. J. Woolf, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E6808–E6817 (2015).
16. J. Peng *et al.*, *Nat. Commun.* **7**, 12029 (2016).
17. E. A. Old *et al.*, *J. Clin. Invest.* **124**, 2023–2036 (2014).
18. F. Denk, M. Crow, A. Didangelos, D. M. Lopes, S. B. McMahon, *Cell Reports* **15**, 1771–1781 (2016).
19. H. L. Willemen *et al.*, *J. Pain* **15**, 496–506 (2014).
20. L. Vicuña *et al.*, *Nat. Med.* **21**, 518–523 (2015).
21. M. Costigan *et al.*, *J. Neurosci.* **29**, 14415–14422 (2009).
22. R. E. Sorge *et al.*, *Nat. Neurosci.* **18**, 1081–1083 (2015).
23. X. J. Liu *et al.*, *Cell Res.* **24**, 1374–1377 (2014).
24. K. M. Baumbauer *et al.*, *eLife* **4**, e09674 (2015).
25. G. L. Fell, K. C. Robinson, J. Mao, C. J. Woolf, D. E. Fisher, *Cell* **157**, 1527–1534 (2014).
26. C. Moore *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, E3225–E3234 (2013).
27. M. Calvo, J. M. Dawes, D. L. Bennett, *Lancet Neurol.* **11**, 629–642 (2012).
28. R. R. Ji, T. Berta, M. Nedergaard, *Pain* **154** (suppl. 1), S10–S28 (2013).
29. P. M. Grace, M. R. Hutchinson, S. F. Maier, L. R. Watkins, *Nat. Rev. Immunol.* **14**, 217–231 (2014).
30. M. Tsuda *et al.*, *Nature* **424**, 778–783 (2003).
31. A. K. Clark *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10655–10660 (2007).
32. Z. Guan *et al.*, *Nat. Neurosci.* **19**, 94–101 (2016).
33. Y. Kawasaki, L. Zhang, J. K. Cheng, R. R. Ji, *J. Neurosci.* **28**, 5189–5194 (2008).
34. F. Ferrini *et al.*, *Nat. Neurosci.* **16**, 183–192 (2013).
35. A. K. Clark *et al.*, *J. Neurosci.* **35**, 4552–4570 (2015).
36. T. Berta *et al.*, *J. Clin. Invest.* **124**, 1173–1186 (2014).
37. P. M. Grace *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, E3441–E3450 (2016).
38. G. Chen *et al.*, *Brain* **137**, 2193–2209 (2014).
39. B. C. Jiang *et al.*, *J. Clin. Invest.* **126**, 745–761 (2016).
40. D. S. Kim *et al.*, *J. Neurosci.* **32**, 8977–8987 (2012).
41. S. K. Kim *et al.*, *J. Clin. Invest.* **126**, 1983–1997 (2016).
42. A. C. Zarpelon *et al.*, *FASEB J.* **30**, 54–65 (2016).
43. Y. Shi, J. Shu, Z. Liang, S. Yuan, S. J. Tang, *Mol. Pain* **12**, 1 (2016).
44. S. Gritsch *et al.*, *Nat. Commun.* **5**, 5472 (2014).
45. X. Zhang, Y. Chen, C. Wang, L. Y. Huang, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 9864–9869 (2007).
46. Y. Chen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16773–16778 (2008).
47. W. Guo *et al.*, *Stem Cells* **29**, 1294–1303 (2011).
48. G. Chen, C. K. Park, R. G. Xie, R. R. Ji, *J. Clin. Invest.* **125**, 3226–3240 (2015).
49. J. M. Bráz *et al.*, *Neuron* **74**, 663–675 (2012).
50. P. Mantyh, *Pain* **154** (suppl. 1), S54–S62 (2013).
51. D. Selvaraj *et al.*, *Cancer Cell* **27**, 780–796 (2015).
52. Q. Xu *et al.*, *J. Neurosci.* **33**, 19099–19111 (2013).
53. Y. Yang *et al.*, *J. Neurosci.* **35**, 7950–7963 (2015).
54. A. Diogenes, C. C. Ferraz, A. N. Akopian, M. A. Henry, K. M. Hargreaves, *J. Dent. Res.* **90**, 759–764 (2011).
55. E. Marion *et al.*, *Cell* **157**, 1565–1576 (2014).
56. F. Y. Tanga, N. Nutile-McMenemy, J. A. DeLeo, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5856–5861 (2005).
57. C. A. Christianson *et al.*, *Pain* **152**, 2881–2891 (2011).
58. C. K. Park *et al.*, *Neuron* **82**, 47–54 (2014).
59. Z. Z. Xu *et al.*, *Nat. Med.* **21**, 1326–1331 (2015).
60. Y. Li *et al.*, *J. Neurosci.* **35**, 13487–13500 (2015).
61. T. Liu, Z. Z. Xu, C. K. Park, T. Berta, R. R. Ji, *Nat. Neurosci.* **13**, 1460–1462 (2010).
62. I. M. Chiu, C. A. von Hehn, C. J. Woolf, *Nat. Neurosci.* **15**, 1063–1067 (2012).
63. A. Zimmerman, L. Bai, D. D. Ginty, *Science* **346**, 950–954 (2014).
64. Z. Z. Xu *et al.*, *Ann. Neurol.* **74**, 490–495 (2013).
65. L. V. Borovikova *et al.*, *Nature* **405**, 458–462 (2000).
66. M. D. da Silva *et al.*, *Mol. Neurobiol.* **51**, 19–31 (2015).
67. N. Goldman *et al.*, *Nat. Neurosci.* **13**, 883–888 (2010).
68. R. Torres-Rosas *et al.*, *Nat. Med.* **20**, 291–295 (2014).


## ACKNOWLEDGMENTS


Our research is supported by NIH (R01 grants DE17794, DE22743, and NS87988 to R.-R.J.) and National Natural Science Foundation of China (grant 31420103903 to Y.-Q.Z and R.-R.J.). All the figures have not been published before.

10.1126/science.aaf8924

REVIEW

# Neural circuits for pain: Recent advances and current views

Cedric Peirs[1,2] and Rebecca P. Seal[1,2]*

**The mammalian nervous system encodes many different forms of pain, from those that arise as a result of short-term low-grade interactions with noxious thermal, chemical, or mechanical sources to more serious forms of pain induced by trauma and disease. In this Review, we highlight recent advances in our understanding of the neural circuits that encode these types of pain. Promising therapeutic strategies based on recent advances are also highlighted.**

The ability to sense pain protects us from harm and is thus an essential aspect of our well-being. Patients suffering from channelopathies that eliminate the ability to feel pain have very high rates of early mortality, largely due to self-mutilation and repetitive fractures (*1*). Pain, of course, is also a source of substantial discomfort, and humans have long sought ways to ameliorate pain, as exemplified by the Ebers Papyrus from ancient Egypt. As a concept, pain has evolved from archaic notions of demonic punishment to more contemporary views of biological circuit–based origins for the sensation (*2*). The most recent definition of pain proposed by the International Association for the Study of Pain, unchanged since its first publication in 1979, is "an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage." The full biological complexity of pain and its underlying circuitry is not wholly conveyed by this definition, however. Pain stems from a varied array of peripheral sensors that detect nociceptive stimuli within internal tissues and those produced by the external world. The information is subsequently distributed to a series of complex neural circuits in the spinal cord dorsal horn and then to numerous brain regions, producing a diverse set of emotions, actions, and sensations (Fig. 1). Pain also manifests with different qualities, including stabbing, pricking, burning, or aching, further highlighting the heterogeneity of the underlying neural circuitry. Longer-lasting pain states produced by nerve and tissue damage, which are also heterogeneous in nature, provide ongoing awareness of the injured area. Pain can become chronic and debilitating when the tissue damage persists and, in some conditions, even after the wound has healed. Because of the high prevalence and lack of adequate treatment options, unraveling the biological basis of persistent pain continues to be an area of intense study. The objective of this Review is to provide an overview of pain circuitry, with an emphasis on recent important advances in our understanding of pathological pain. Insights into neural circuits for pain that have therapeutic potential are also discussed.

[1]Departments of Neurobiology and Otolaryngology, University of Pittsburgh School of Medicine, 3501 Fifth Avenue, BST3, Pittsburgh, PA 15213, USA. [2]Pittsburgh Center for Pain Research, University of Pittsburgh School of Medicine, 200 Lothrop Street, Pittsburgh, PA 15213, USA.
*Corresponding author: Email: rpseal@pitt.edu

## How is pain detected in the periphery?

Pain is produced through the activation of nociceptive primary sensory neurons categorized into the classes C, Aδ, and to a lesser extent Aβ, depending on their axon caliber, degree of myelination, and conductivity properties. Nociceptors innervate the skin, deep tissues, and internal organs and are tuned to detect a wide variety of noxious mechanical, thermal, and chemical stimuli through the activation of modality-specific sensory transduction molecules. A number of these key molecules have been identified (*3*). Most if not all modalities are conveyed through the activation of more than one transducer, as exemplified by heat transduction. Genetic deletion of the heat-sensing transient receptor potential channel vanilloid 1 (TRPV1) in mice impaired the response to heat only at high noxious temperatures (*4*), pointing to the existence of one or more additional heat sensors. Recent efforts have now identified TRPM3 (*5*) and the calcium-activated chloride channel ANO1 (*6*) as important contributors to the sensation of heat. As was observed with TRPV1, deletion of either channel alone strongly reduced, but did not completely eliminate, noxious heat sensitivity. In contrast, ablation of TRPV1$^+$ fibers by intrathecal injection of the TRPV1 ligand capsaicin (*7*) or silencing the cells by selective uptake of the voltage-gated sodium channel blocker QX-314 (*8*) completely abolished noxious heat sensitivity, demonstrating that the fibers are sufficient to account for heat pain, despite a more limited contribution from each channel. Given that TRPV1, TRPM3, and ANO1 show considerable overlap in their distribution, analyses of double and triple genetic deletion of the channels in mice may be required to fully understand at the cellular level how noxious heat is transduced. Cold is also transduced through multiple channels. Mice with a genetic deletion of the menthol-sensitive TRPM8, a critical transducer of innocuous cooling, showed only partial avoidance of noxious cold temperatures, whereas selective ablation of TRPM8-expressing cells completely abolished noxious cold sensitivity (*9*). The molecules that transduce mechanical nociception remain stubbornly elusive. Deletion of the leading candidate molecules in mice unexpectedly produced little to no change in noxious mechanical pain sensitivity (*10*, *11*).

*"Pain stems from a varied array of peripheral sensors that detect nociceptive stimuli within internal tissues and those produced by the external world."*

## Transduction in the periphery: More than just neurons

Epithelial cells such as keratinocytes and Merkel cells directly interact with peripheral axon terminals and have been implicated in the modulation of sensory transduction. Recently, several elegant studies have demonstrated an active role for epithelial cells in tuning the response of sensory neurons. Merkel cells are the epidermal end organ of slowly adapting type 1 (SA1) mechanoreceptors and are involved in the detection of features such as edges and textures. Depolarization of the Merkel cells by optogenetic stimulation or by touch, which depends on the innocuous mechanotransduction channel Piezo2, directly excites SA1 fibers through still-unidentified neurotransmitter signaling mechanisms (*12*). Epithelial cells have also been shown to actively contribute to nociceptive transmission. Activation of channelrhodopsin or TRPV1 channels ectopically expressed by keratinocytes in mice induced action potential firing in sensory neurons, neuronal activity in the spinal cord, and nocifensive behaviors (*13*, *14*). Conversely, inhibition of the cells blocked the response of nociceptors and other primary sensory neurons to cutaneous stimulation (*14*). Studies that identify the salient signaling molecules or establish the extent to which epithelial cells contribute behaviorally to pain will be of interest.

## Is pain a labeled line?

Most nociceptors express various combinations of sensory transducers that are specific for different modalities (often heat and mechanical) and therefore are classified as "polymodal." Large-scale efforts using single-cell RNA transcriptome analyses have sought to better understand the molecular and functional logic of primary sensory neurons by parsing the transcriptome profiles of hundreds of individual cells into groups defined by convergent gene expression patterns (*15*–*17*). These cluster analyses offer a richer picture of the diversity of nociceptors than previous molecular marker–based

classifications, as well as a deeper understanding of the molecular constituents that give rise to the functional properties of sensory neurons (Fig. 2). Although the precise relationships between the molecular constituents and the response properties of the neurons will require further study, the transcriptome data do show some predictive capacity: For example, the tyrosine hydroxylase ($TH^+$) cluster expresses high levels of Piezo2, but not the heat sensors TRPV1, TRPM3, or ANO1, consistent with the response of the cells to light touch, but not to heat (*18*–*20*).

However, limits on the ability to predict the functional output of nociceptors on the basis of their molecular and biophysical properties were also recently demonstrated in studies in which select polymodal populations were ablated in mice. Ablation of nociceptors that express the Mas-related G protein–coupled receptor subtype D (MRGPRD) markedly reduced acute and persistent mechanical pain but had no effect on thermal sensitivity (*7*). Similarly, ablation of the calcitonin gene–related peptide (CGRP) neurons, a population that partially overlaps with the heat sensors TRPV1, TRPM3, and to a lesser extent ANO1, predictably produced a profound loss of heat but did not alter innocuous or noxious mechanical sensation (*21*). Loss of CGRP-expressing sensory neurons, which are not responsive to cooling agents (*22*), greatly increased tonic and evoked activity in spinal neurons associated with cold. This latter finding, mechanistically resembling the activation of dorsal horn itch circuits after silencing pain afferents (pain normally inhibits itch at the level of the spinal cord), further highlights the complexity of primary afferent

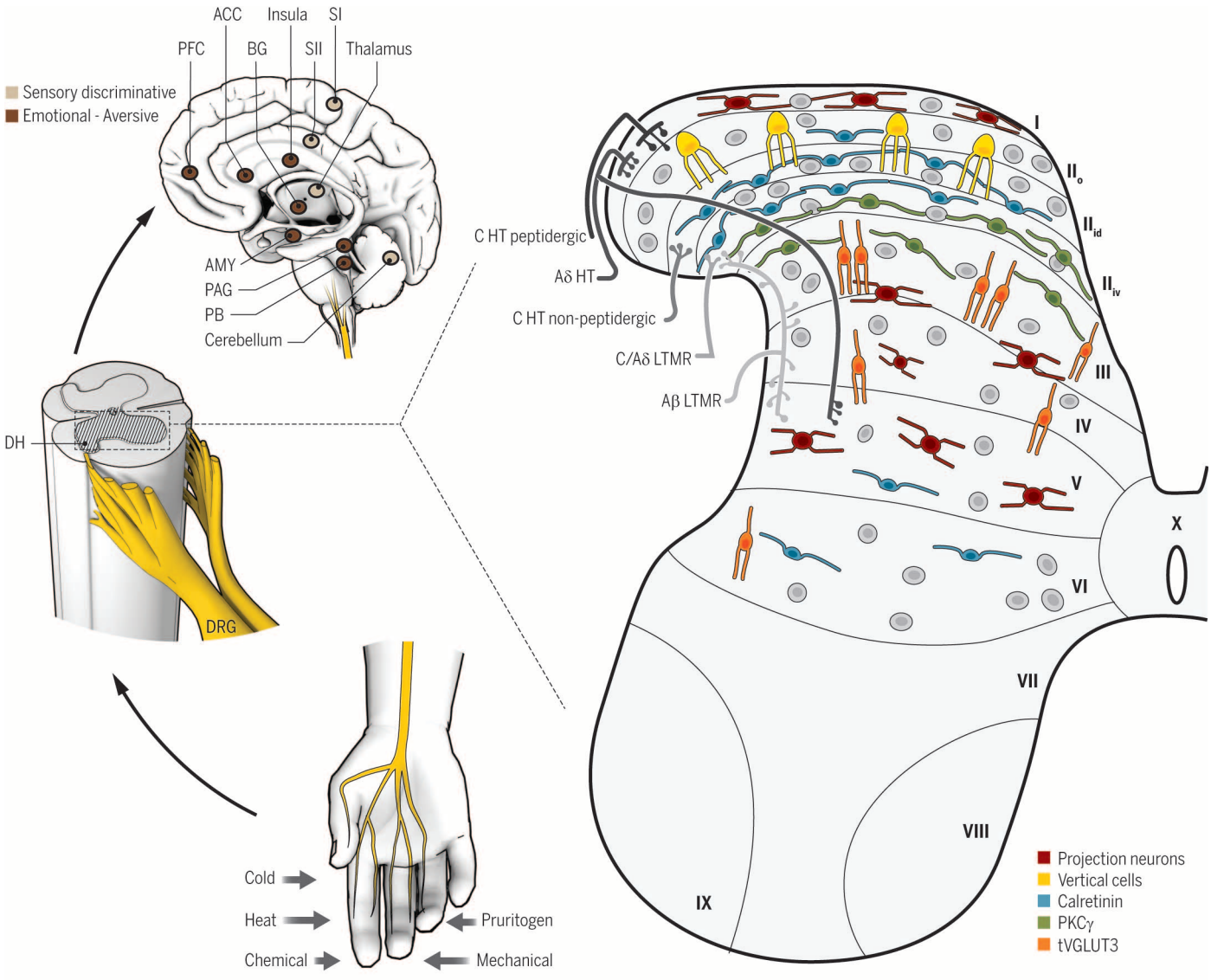



**Fig. 1. Overall organization of somatosensory circuits.** Cutaneous sensory neurons (DRG) are activated by a variety of stimuli (bottom left) and project to the spinal cord dorsal horn (DH, middle left). In the DH (right), the central terminals of high-threshold nociceptors (HT) are located in the most superficial laminae [lamina I to the dorsal part of inner lamina II ($II_{id}$)] and lamina V. Low-threshold mechanoreceptors (LTMR) preferentially end in the deep dorsal horn [ventral part of inner lamina II ($II_{iv}$) to lamina V]. The spinal cord is divided into 10 laminae (the DH is I to VI) and is composed of numerous neuronal populations. Some identified populations are organized in longitudinal layers (only excitatory neurons are represented): neurons transiently expressing VGLUT3 (tVGLUT3, orange) in laminae III and IV, PKCγ (green) in lamina $II_{iv}$, calretinin (blue) in outer lamina II ($II_o$) and lamina $II_{id}$, vertical cells (yellow) in lamina $II_o$, and projection neurons (red) in laminae I, IV, and V. Projection neurons send information to the brainstem and thalamus and then on to several brain regions implicated in sensory-discriminative (upper left, light brown) and emotional (upper left, dark brown) sensory perception. ACC, anterior cingular cortex; SI (II), primary (secondary) somatosensory cortex; PAG, periaqueductal gray area; PB, parabrachial nucleus; AMY, amygdala; PFC, prefrontal cortex; BG, basal ganglia.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

coding, as well as an important role for central processing in shaping the sensory percept.

### Integration in the spinal cord

Primary sensory neurons innervate the six Rexed laminae of the spinal cord dorsal horn, the major locus for the integration of peripheral sensory input and descending supraspinal modulation. The central terminals of sensory neurons are somatotopically organized, forming ventrodorsal-oriented columns similar to the arrangement of the primary somatosensory cortex. Most C- and Aδ-nociceptive afferents form synaptic contacts in the superficial laminae (I and II), whereas low-threshold Aδ- and Aβ-afferents generally project to deeper laminae (III to V) (Fig. 1). Accordingly, in vivo electrophysiological recordings in mammals have shown that laminae I and II respond mainly to noxious peripheral stimulation, whereas neurons in deep layers are more sensitive to touch. The dorsal horn is composed of a large number of excitatory (75%) and inhibitory (25%) interneuron populations, as well as a smaller number of projection populations located in laminae I, III, IV, and V. The projection neurons relay information to numerous supraspinal sites to give rise to both qualitative and affective aspects of the pain sensation.

Recent studies of dorsal horn nociceptive circuits have focused largely on persistent pain because of its high prevalence and the need for better treatment options (*23*, *24*).

### Mechanical allodynia: Peripheral contributions

Nerve and tissue damage produces dramatic changes in peripheral and central somatosensory circuits. Although acute pain typically follows the activation of nociceptors, light touch–activated neurons can be recruited into the nociceptive network to cause pain after injury. This pathological condition, termed mechanical allodynia, occurs in numerous peripheral neuropathies and central pain disorders, presenting in up to 50% of patients with neuropathic pain (*25*). Still unknown is the identity of the light touch–sensitive primary sensory population(s) that conveys mechanical allodynia. A large number of primary afferents respond to light mechanical touch and are thus potential candidates (*26*).

Although the identity of the mechanosensory neurons engaged during mechanical allodynia remains unclear, progress has been made in our understanding of innocuous sensory mechanotransduction (*27*). As mentioned previously, innocuous tactile sensitivity is primarily dependent on the cation channel Piezo2 (*28*). In primary sensory neurons, the channel is almost exclusively expressed by low-threshold mechanoreceptors. Conditional deletion of Piezo2 in all sensory neurons in mice resulted in markedly reduced mechanical activation of low-threshold mechanoreceptors, as well as reduced behavioral touch sensation, with no change in acute mechanical pain (*29*). Because low-threshold mechanoreceptors are thought to convey some forms of mechanical allodynia, Piezo2 could have a role in this type of pain—although after an inflammatory insult, mice with a conditional deletion of Piezo2 in primary sensory neurons did not show evidence of a defect in their mechanical pain threshold under the conditions tested. In contrast, a separate study, which demonstrated sensitization of Piezo2 by the cyclic adenosine monophosphate sensor EPAC1 after neuropathic injury, pointed to a potential role for the mechanotranducer in mechanical allodynia (*30*).

### Mechanical allodynia: The silent dorsal horn circuit

Changes in the dorsal horn circuitry contribute to the expression of mechanical allodynia. One

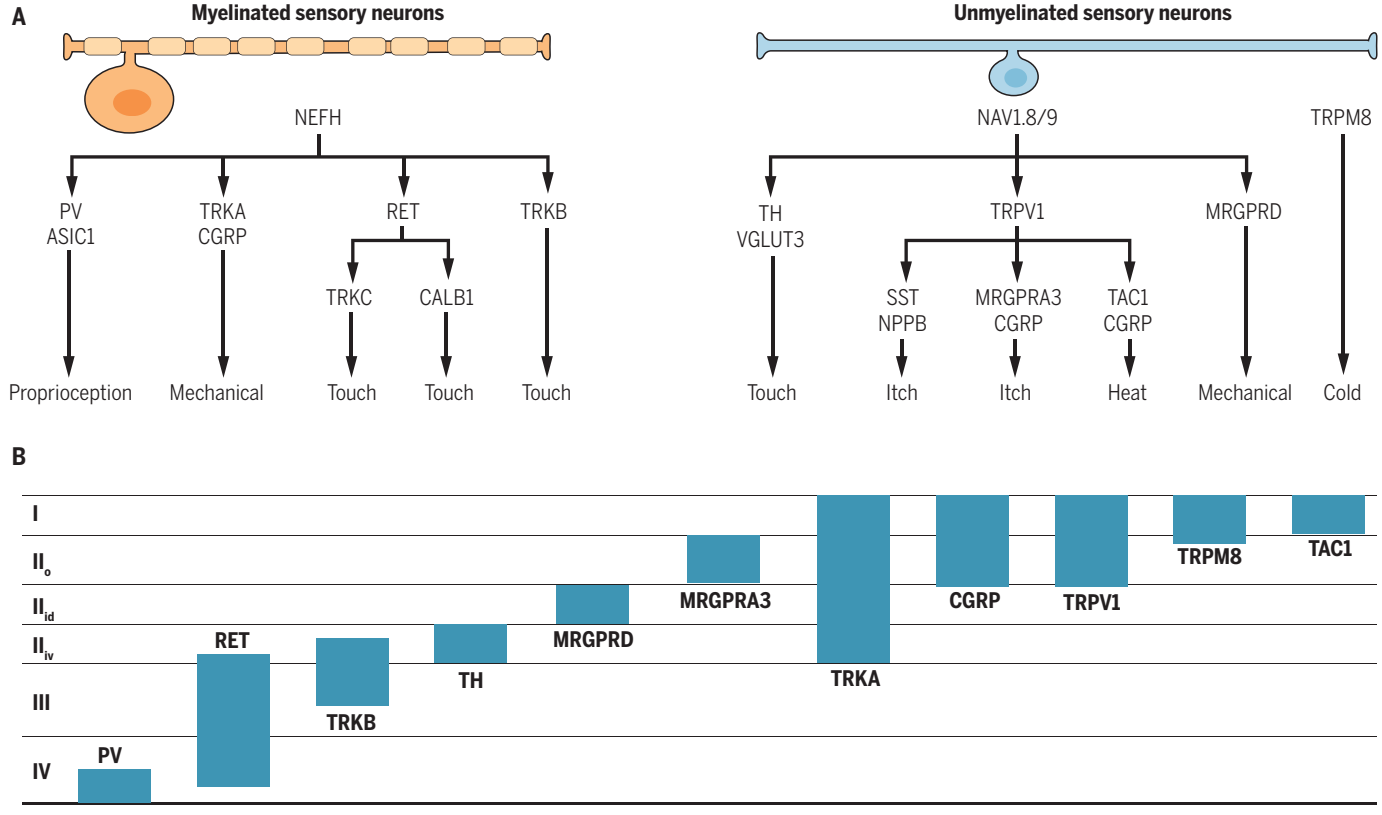



**Fig. 2. Organization of primary sensory neurons.** (**A**) The categories of myelinated and unmyelinated neurons and their respective functional roles are derived from large-scale transcriptional analyses, behavioral analyses, and the literature. (**B**) The locations of the central terminations of the primary sensory neuron categories are shown. The schematic is based on analyses of gene reporter mouse lines (PV, RET, TRKB, TH, MRGPRD, MRGPRA3, CGRP, and TRPM8) and immunohistological analyses (TRKA, TRPV1, TAC1). Myelinated neurons preferentially express neurofilament heavy chain (NEFH), and unmyelinated neurons preferentially express the sodium channels Nav1.8 and -1.9. Laminae are indicated on the left. RET, ret proto-oncogene; CALB1, calbindin 1; TAC1, tachykinin 1; SST, somatostatin; NPPB; natriuretic peptide type B.

GRAPHIC: ADAPTED BY V. ALTOUNIAN/*SCIENCE*

of the principal concepts to explain these changes is based on the gate control theory in which touch inhibits pain through a feed-forward inhibitory circuit within the superficial layers of the dorsal horn (*31*). In the case of mechanical allodynia, injury impairs the feed-forward inhibitory circuit, allowing light touch to engage the nociceptive network. A seminal study, performed in rodent spinal cord slices, showed that low-threshold mechanosensory A-fiber input is able to activate lamina I pain projection neurons through a polysynaptic network when inhibitory receptor antagonists are present (mimicking the injury-induced decrease in inhibition) (*32*). Importantly, the study established that the mechanical allodynia circuit, which is normally silent, is already in place under physiological conditions.

### Dorsal horn mechanical allodynia circuits: Excitatory neurons

The dorsal horn circuit for mechanical allodynia was originally modeled as a dorsally directed pathway that begins with excitatory interneurons at the border between laminae II and III that express the γ isoform of protein kinase C (PKCγ) and ends with lamina I pain projection neurons (*33*) (Fig. 3). Two intermediary populations in this pathway were recently identified through paired recordings, the transient central cells in inner lamina II and vertical cells in outer lamina II (*34*). The dendrites of vertical cells extend deep into lamina III and receive input from most classes of primary afferents, suggesting a potential role as integrators of the network (*35*). An expansion of the circuit into deeper laminae was shown by studies of the vesicular glutamate transporter 3 (VGLUT3). A population of excitatory interneurons was identified in lamina III that transiently expresses VGLUT3 and receives almost exclusively low-threshold input. Selective activation of the cells by using a designer receptor strategy (DREADD) further demonstrated their important role in the transmission of touch as painful (*36*). Calretinin-expressing excitatory interneurons in inner lamina II were also identified as an essential element of the circuit. A functional study of the calretinin and PKCγ populations that used c-Fos analysis revealed the existence of distinct excitatory microcircuits for mechanical allodynia, which are differentially engaged depending on whether the injury is inflammatory or neuropathic in nature (*36*).

### Mechanical allodynia circuits: The "gates"

Inhibitory neurons are a fundamental element of the mechanical allodynia network in the spinal cord, forming "gates" that prevent the recruitment of low-threshold fibers into the nociceptive network under physiological conditions (Fig. 3). Using paired recordings in spinal cord slices, a recent study showed that the PKCγ interneurons, which receive Aβ-fiber input, are normally under feed-forward glycinergic disynaptic inhibition (*34*). After neuropathic injury,

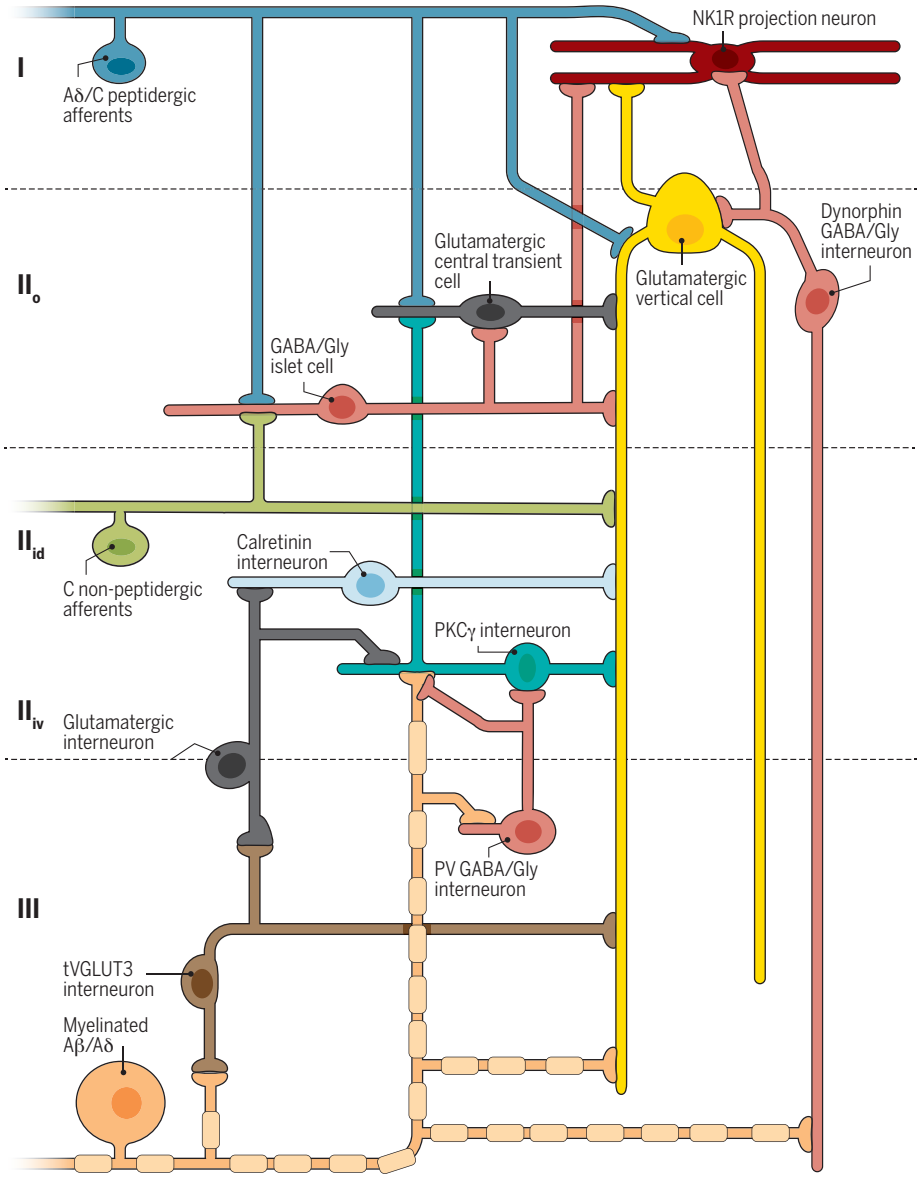



**Fig. 3. Organization of the dorsal horn circuit for pain.** Peripheral nociceptors (blue) project onto excitatory interneurons in lamina II$_o$ (central cell, dark gray; vertical cells, yellow) and onto neurokinin 1 receptor (NK1R) projection neurons (red) in lamina I. Nonpeptidergic afferents expressing MRGPRD (green) project to lamina II$_{id}$, including to excitatory vertical cells with ventrally directed elongated dendrites. Both primary afferents also contact inhibitory islet cells (horizontally elongated, pink). Stimulation of nociceptive afferents activates excitatory central cells, vertical cells, and NK1R projection neurons to mediate noxious pain. Inhibitory islet cells modulate this activity. Innocuous afferents (orange) project onto excitatory interneurons expressing tVGLUT3 in lamina III (brown), PKCγ in lamina II$_{iv}$ (teal), and vertical cells in lamina II$_o$. Myelinated afferents also contact PV inhibitory interneurons in lamina III (radial, pink) and dynorphin inhibitory vertical cells in lamina II$_o$ (vertical, pink). tVGLUT3 interneurons project onto excitatory vertical cell dendrites and intermediate excitatory interneurons in lamina III. Intermediate interneurons project onto PKCγ and calretinin excitatory interneurons in lamina II$_{iv}$. Inhibitory interneurons prevent A-fiber–mediated activation of the nociceptive network through feed-forward circuits that act on PKCγ interneurons, vertical cells, and NK1R projection neurons. After nerve injury, inhibition by PV and dynorphin interneurons is reduced, allowing A-fiber–mediated activation of a dorsally directed circuit that includes tVGLUT3 and PKCγ neurons. After inflammatory injury, reduction in a still-unknown mechanism of disinhibition allows A-fiber–mediated activation of a dorsally directed circuit that includes tVGLUT3 and lamina II$_{iv}$ calretinin neurons.

ADAPTED BY VAL ALTOUNIAN/*SCIENCE*

which is known to impair inhibitory transmission in the dorsal horn, Aβ-fibers were able to drive action potentials in PKCγ neurons and presumably the rest of the nociceptive network. Consistent with this finding, selective ablation of lamina III parvalbumin (PV)–expressing interneurons, which form contacts with the PKCγ neurons, induced mechanical allodynia, indicating a role for PV interneurons in gating the allodynic pathway (*37*). Conversely, selective activation of the PV interneurons by the excitatory DREADD strategy reversed mechanical allodynia induced by nerve injury. In addition to PKCγ neurons, PV interneurons were also proposed to act directly on the Aβ-afferents innervating these neurons, suggesting alternative mechanisms for inhibitory control of the circuit (*38*). The involvement of a second population of inhibitory interneurons in the feed-forward gating of mechanical allodynia was also recently demonstrated. Ablation of dynorphin-expressing interneurons in outer lamina II and at the border of laminae II and III induced mechanical allodynia. The dynorphin gate was further demonstrated to control somatostatin-expressing interneurons, a major population of excitatory neurons that partially overlaps with PKCγ neurons and potentially also vertical and transient central cells (*39*). Studies of the mechanical allodynia network are beginning to not only reveal individual elements but also produce conceptual advances, including the understanding that there are distinct excitatory microcircuits and multiple distinct gates related to specific types of injuries.

### Mechanisms of disinhibition underlying mechanical allodynia

The central sensitization responsible for mechanical allodynia produces various forms of disinhibition in the dorsal horn (*40*). One mechanism involves the injury-induced release of brain-derived neurotrophic factor (BDNF), which triggers down-regulation of the potassium chloride exchanger 2 (KCC2) in tropomyosin receptor kinase B (TRKB)–expressing pain-transmitting neurons, reducing the driving force for chloride and thus the strength of inhibitory transmission (*41*, *42*). The mechanism of injury-induced BDNF release was recently shown to differ between males and females. Males require the activation of microglia, whereas females require adaptive immune cells, possibly T lymphocytes (*43*). Sex differences are frequently observed in animal models of chronic pain and may provide insight into the many clinically observed sex differences in pain in humans (*44*).

Investigations into the mechanisms underlying BDNF release in the dorsal horn of male rodents after injury have focused on adenosine triphosphate signaling through purinergic P2X4 receptors, which requires microglia activation, and on the involvement of chemokine signaling cascades that promote and maintain microglia activation (*45*). De novo production and release of colony-stimulating factor 1 (CSF1) by injured primary sensory neurons is required for the initial activation of the microglia, as well as for cell proliferation and self-renewal. Deletion of CSF1 specifically from primary sensory neurons or deletion of its downstream effector, the microglial membrane adaptor protein DAP12, was sufficient to markedly reduce microglial activation, as well as the behavioral expression of mechanical allodynia induced by nerve injury (*46*).

### Pain processing in the brain

Functional magnetic resonance imaging (fMRI) studies in humans have demonstrated the coordinated activation of several brain areas in response to noxious somatic and visceral stimuli, including the thalamus, anterior cingulate cortex (ACC), insular cortex, primary and secondary sensory cortices, prefrontal cortex, basal ganglia, cerebellum, and amygdala. This network of brain regions involved in both sensory-discriminative and emotional-affective aspects of pain is termed the "pain matrix" (*47*).

Several recent studies have questioned whether activation of the pain matrix specifically represents pain or is a more generalized system to detect salient events and the state of the body. The posterior insular cortex encodes nociceptive intensity, and lesions in this brain area disrupt pain, whereas direct stimulation of the area induces pain. However, the insular cortex also has a role in affective, cognitive, and homeostatic functions for nonnociceptive sensory modalities. Intracerebral recordings of the anterior and posterior insular cortex showed activity levels that responded similarly to different sensory modalities (visual, auditory, and somatosensory), as well as to noxious and innocuous somatosensory stimuli, consistent with the idea that the insula is not specific for pain-related information (*2*). Additionally, patients with loss-of-function mutations in the sodium channel Nav1.7 (SCN9) show congenital insensitivity to pain due to a loss of noxious peripheral sensory drive (*1*). In a recent study, the response of patients and controls to normally noxious peripheral stimuli was assessed by fMRI. Both groups reported experiencing similar levels of sensation; although controls reported the stimuli as painful and patients reported no pain, the two groups nevertheless showed similar patterns and levels of activity within the pain matrix (*48*). Gross activity measured in supraspinal structures thus may not reliably predict the pain experience. Rather, pain-related information is likely encoded by specific subregional patterns of activity throughout the pain matrix, which would require analysis by other methods and by experimental designs that allow for causal inference. Machine-learning algorithms to assess fMRI patterns of activity across the pain matrix are showing a high predictive capacity for distinct forms of physical pain, as well as for social pain such as vicarious pain (*49*).

> ***"...pain-related information is likely encoded by specific subregional patterns of activity throughout the pain matrix..."***

Complementing these recent findings from brain-imaging studies in humans, work in rodents has focused on synaptic-level mechanisms of pain processing and plasticity in particular regions of the pain matrix. These studies provide further support for the idea that individual brain areas have multiple functional roles. Emotional aspects of pain such as anxiety and fear have been linked to the ACC. A study in mice identified a form of presynaptic long-term potentiation (pre-LTP) at thalamocortical synapses in the ACC that was induced by either anxiety-evoking experiences or chronic pain models (*50*). It was concluded that the pre-LTP represents anxiety, which can be triggered by chronic pain. The coexistence of this form of pre-LTP with a previously identified form of postsynaptic LTP required for the behavioral expression of chronic pain (*51*) suggests an important role for these ACC synapses in the mutually enhancing interactions of anxiety and chronic pain.

Mechanisms of plasticity affecting motivation in models of persistent pain were also investigated recently. The nucleus accumbens has been implicated in subjective pain processing in humans and in motivation in rodents. The induction of persistent pain by either nerve injury or an inflammatory mediator impaired motivational behavior in mice, as measured using a progressive ratio operant task (*52*). The decrease in motivation was linked to enhanced release of the neuropeptide galanin in the nucleus accumbens core, which induced a form of long-term depression (LTD) in dopamine receptor 2–expressing medium spiny neurons (D2-MSNs), thereby reducing their activity. The injury-induced effects on D2-MSNs and motivated behavior were reversed by knockdown of the galanin receptor GALR1 and by interfering with intracellular signaling pathways required for the LTD. Synaptic-level investigations of pain-related mechanisms are thus providing important insights into the role of individual brain areas in the pain experience, including interactions with other brain functions. Such studies are also revealing potential new therapeutic avenues.

### Blocking pain circuits at the periphery

Nerve growth factor (NGF) and its major receptor TRKA have essential roles in nociceptor function both during development and in adults. Specific mutations in the genes encoding NGF or TRKA in humans produce phenotypes that include insensitivity to pain, largely resulting from defects in the development of nociceptors (*53*). In adults, NGF signaling through TRKA receptors at the periphery contributes

considerably to the pain associated with many types of injuries and diseases' including bone cancer, lower back injury, diabetic neuropathy, and osteoarthritis. As a new class of analgesic treatment, humanized antibodies raised against NGF, such as tanezumab (Pfizer), now in phase III clinical trials for osteoarthritis pain, are showing promise. Inhibitors of the TRKA receptor are also of therapeutic interest.

Recent studies have focused also on the therapeutic potential of targeting voltage-gated sodium channels in sensory neurons to relieve pain. In addition to the Nav1.7 loss-of-function mutations that cause congenital insensitivity to pain in humans, studies of Nav1.7 and Nav1.8 in mice indicate a required role for the channel in many forms of acute and pathological pain (*54*, *55*). A highly specific Nav1.7 inhibitor isolated from centipede venom, μ-SLPTX-Ssm6a, when injected intraperitoneally in mice, showed potent analgesic effects on chemical-induced pain with an efficacy equivalent to or exceeding that of morphine (*56*). Similarly, monoclonal antibodies against Nav1.7, when injected into mice, effectively reduced mechanical allodynia and thermal hyperalgesia induced by inflammation and neuropathic injury (*57*). Given the promising therapeutic benefit of targeting Nav1.7 channels to relieve pain, several selective inhibitors are now in phase II clinical trials.

Animal venoms that elicit pain have been particularly useful for identifying and characterizing therapeutically relevant nociceptive sensory molecules. MitTx, from the venom of the Texas coral snake *Micrurus tener tener*, is a highly specific and strong potentiator of the acid-sensing channel ASIC1. Injection of this peptide into the hind paw of mice induced pronounced nocifensive behaviors (*58*). Conversely, central and peripheral delivery of the isopeptides mambalgin 1 and 2, from the black mamba snake, reversed mechanical allodynia and heat hyperalgesia induced by inflammatory agents by inhibiting ASIC channels, highlighting the therapeutic potential of this family of ion channels (*59*).

Recent advances in the in vivo application of optogenetic methods also provide attractive therapeutic opportunities for pain. One of the first studies to report the use of optogenetics to evaluate pain circuits in vivo targeted channelrhodopsin to Nav1.8-expressing primary sensory neurons in mice. Cutaneous light-mediated activation of the neurons elicited robust nocifensive behaviors and conditioned place aversion (*60*). In testing of opsins for therapeutic purposes, the inhibitory opsin NpHR was targeted to small-diameter primary sensory neurons by injection of AAV6-NpHR into the sciatic nerve. Cutaneous light-mediated inhibition of the infected primary sensory neurons alleviated both inflammatory and neuropathic pain (*61*). The low penetration of light through the skin may be technically limiting; however, implantable miniaturized optoelectronic systems for wireless control of primary sensory and spinal cord neurons are now being developed (*62*).

Another innovative approach to inhibit nociceptors targets the voltage-gated sodium channel blocker QX-314 to select populations, thereby blocking nerve conduction. Coapplication of capsaicin, which permits entry of QX-314 through the cation channel TRPV1, to the mouse plantar hind paw markedly reduced the response to heat and mechanical pressure (*8*). QX-314 has also been used to ameliorate persistent pain. The Toll-like receptor TLR5 was identified as a receptor specific to A-fibers (*63*). Coapplication of the TLR5 ligand flagellin and QX-314 to the mouse plantar hind paw blocked A-fiber activity, as measured electrophysiologically in vitro and in vivo. Although the precise mechanism by which QX-314 enters the cells is still unclear, the approach successfully reversed mechanical allodynia and ongoing pain in a number of chronic neuropathic pain models in mice without producing obvious deficits in touch sensation.

Lastly, the use of gene therapy to restore the normal expression of dysregulated proteins in sensory neurons has become another attractive strategy to treat pain. The expression of the microRNA miR-7a is dramatically reduced after neuropathic injuries. Increasing its expression by intraganglionic injection of AAV6-miR-7a abolished mechanical allodynia and thermal hyperalgesia induced by nerve injury, although not by inflammation (*64*). In contrast, nerve injury substantially up-regulates the sodium channel Nav1.3. Intraganglionic injection of AAV5 expressing a small hairpin RNA against Nav1.3 reduced its expression and partially reversed nerve injury–induced mechanical allodynia (*65*).

***"...gene therapy to restore the normal expression of dysregulated proteins in sensory neurons has become another attractive strategy to treat pain."***

## Blocking pain circuits centrally

Because mechanisms of disinhibition in the dorsal horn have a critical role in the expression of mechanical allodynia and spontaneous pain, therapeutic efforts have been directed toward the restoration of inhibitory tone. Administration of a small-molecule enhancer of KCC2 activity (CLP290) to rodents restored inhibitory drive in dorsal horn neurons, as measured electrophysiologically in spinal cord slices, and reversed behavioral evidence of mechanical allodynia induced by nerve injury. Similarly, overexpression of the cotransporter KCC2 in the spinal dorsal horn and primary sensory neurons of rodents reversed behavioral evidence of mechanical allodynia induced by peripheral nerve injury (*66*). Humans with a mutation in the *HSN2*-containing splice variant of the kinase WNK1 suffer from congenital insensitivity to pain. WNK1/HSN2 plays a role in nerve injury–induced mechanical allodynia by reducing KCC2 activity in lamina II neurons of the dorsal horn (*67*). Antagonizing WNK activity in rodents restored GABA (γ-aminobutyric acid)–mediated inhibition of spinal cord lamina II neurons and reversed the behavioral evidence of mechanical allodynia induced by peripheral nerve injury, but not by inflammatory injury.

Using an entirely different strategy, based on the observation that GABAergic precursors from the medial ganglion eminence grafted into the forebrain correct neurological disorders of hyperexcitability, such as epilepsy, Bráz *et al.* tested whether injecting the telencephalic GABAergic precursors directly into the dorsal horn of nerve-injured mice could restore inhibitory tone and reverse mechanical allodynia (*68*). Although the precursor cells maintained their cortical identity, they integrated successfully into the dorsal horn, forming local synaptic connections with primary sensory and spinal cord neurons. By 3 weeks after transplantation, mechanical thresholds were elevated to pre–nerve injury levels. In contrast, inflammation-induced hypersensitivity was not altered by cell transplantation. This work demonstrates that strategies to reverse or compensate for injury-induced disinhibition in the dorsal horn show promise for the treatment of pain.

## Conclusions and the road ahead

Recent advances at the molecular, cellular, and systems level are transforming our understanding of how the complex sensation of pain is encoded in the nervous system and providing exciting new avenues for therapeutic intervention. Several key aspects of the nociceptive network and the sensation of pain, however, remain to be addressed. With respect to physiological sensory processing, major outstanding questions include: What is the identity of the noxious mechanotransducer(s)? What are the specific contributions of molecularly distinct primary sensory neurons to modality coding? What is the logic of nociceptor coding of pain? How do spinal cord circuits contribute to modality coding? What is the logic of the projection neuron populations? Which neuronal populations in the brain process the different aspects of pain? In terms of pathological pain, major outstanding questions include: What are the neural networks for spontaneous pain? What are the peripheral and central components of the neural networks that underlie mechanical allodynia induced by different types of injuries, such as inflammatory and neuropathic? How do the neural circuits for pain differ as a function of chronicity? How does chronic pain manifest in the brain? Understanding the precise contributions of the neurons and molecules that underlie each form of pain will greatly facilitate therapeutic drug development.

## REFERENCES AND NOTES

1. D. L. Bennett, C. G. Woods, *Lancet Neurol.* **13**, 587–599 (2014).
2. R. Sabatowski, D. Schäfer, S. M. Kasper, H. Brunsch, L. Radbruch, *Curr. Pharm. Des.* **10**, 701–716 (2004).
3. D. Julius, J. R. Carlson, *Curr. Opin. Neurobiol.* **34**, v–vi (2015).
4. M. J. Caterina *et al.*, *Science* **288**, 306–313 (2000).
5. J. Vriens *et al.*, *Neuron* **70**, 482–494 (2011).
6. H. Cho *et al.*, *Nat. Neurosci.* **15**, 1015–1021 (2012).
7. D. J. Cavanaugh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 9075–9080 (2009).
8. C. Brenneis *et al.*, *J. Neurosci.* **33**, 315–326 (2013).
9. W. M. Knowlton *et al.*, *J. Neurosci.* **33**, 2837–2848 (2013).
10. Y. Roudaut *et al.*, *Channels* **6**, 234–245 (2012).
11. R. Sharif-Naeini, *Prog. Mol. Biol. Transl. Sci.* **131**, 53–71 (2015).
12. S. Maksimovic *et al.*, *Nature* **509**, 617–621 (2014).
13. Z. Pang *et al.*, *Pain* **156**, 656–665 (2015).
14. K. M. Baumbauer *et al.*, *eLife* **4**, e09674 (2015).
15. D. Usoskin *et al.*, *Nat. Neurosci.* **18**, 145–153 (2015).
16. I. M. Chiu *et al.*, *eLife* **3**, e04660 (2014).
17. C. L. Li *et al.*, *Cell Res.* **26**, 83–102 (2016).
18. M. C. Delfini *et al.*, *Cell Rep.* **5**, 378–388 (2013).
19. L. Li *et al.*, *Cell* **147**, 1615–1627 (2011).
20. R. P. Seal *et al.*, *Nature* **462**, 651–655 (2009).
21. E. S. McCoy *et al.*, *Neuron* **78**, 138–151 (2013).
22. E. S. McCoy, B. Taylor-Blake, M. J. Zylka, *PLOS ONE* **7**, e36355 (2012).
23. A. J. Todd, *Nat. Rev. Neurosci.* **11**, 823–836 (2010).
24. J. Braz, C. Solorzano, X. Wang, A. I. Basbaum, *Neuron* **82**, 522–536 (2014).
25. T. S. Jensen, N. B. Finnerup, *Lancet Neurol.* **13**, 924–935 (2014).
26. V. E. Abraira, D. D. Ginty, *Neuron* **79**, 618–639 (2013).
27. A. Zimmerman, L. Bai, D. D. Ginty, *Science* **346**, 950–954 (2014).
28. B. Coste *et al.*, *Science* **330**, 55–60 (2010).
29. S. S. Ranade *et al.*, *Nature* **516**, 121–125 (2014).
30. N. Eijkelkamp *et al.*, *Nat. Commun.* **4**, 1682 (2013).
31. R. Melzack, P. D. Wall, *Science* **150**, 971–979 (1965).
32. C. Torsney, A. B. MacDermott, *J. Neurosci.* **26**, 1833–1843 (2006).
33. L. S. Miraucourt, R. Dallel, D. L. Voisin, *PLOS ONE* **2**, e1116 (2007).
34. Y. Lu *et al.*, *J. Clin. Invest.* **123**, 4050–4062 (2013).
35. T. Yasaka *et al.*, *Mol. Pain* **10**, 3 (2014).
36. C. Peirs *et al.*, *Neuron* **87**, 797–812 (2015).
37. H. Petitjean *et al.*, *Cell Rep.* **13**, 1246–1257 (2015).
38. D. I. Hughes *et al.*, *J. Physiol.* **590**, 3927–3951 (2012).
39. B. Duan *et al.*, *Cell* **159**, 1417–1432 (2014).
40. H. U. Zeilhofer, H. Wildner, G. E. Yévenes, *Physiol. Rev.* **92**, 193–235 (2012).
41. J. A. Coull *et al.*, *Nature* **438**, 1017–1021 (2005).
42. J. A. M. Coull *et al.*, *Nature* **424**, 938–942 (2003).
43. R. E. Sorge *et al.*, *Nat. Neurosci.* **18**, 1081–1083 (2015).
44. J. S. Mogil, *Nat. Rev. Neurosci.* **13**, 859–866 (2012).
45. S. Beggs, T. Trang, M. W. Salter, *Nat. Neurosci.* **15**, 1068–1073 (2012).
46. Z. Guan *et al.*, *Nat. Neurosci.* **19**, 94–101 (2016).
47. P. Schweinhardt, M. C. Bushnell, *J. Clin. Invest.* **120**, 3788–3797 (2010).
48. T. V. Salomons, G. D. Iannetti, M. Liang, J. N. Wood, *JAMA Neurol.* **73**, 755–756 (2016).
49. A. Krishnan *et al.*, *eLife* **5**, e15166 (2016).
50. K. Koga *et al.*, *Neuron* **85**, 377–389 (2015).
51. H. Xu *et al.*, *J. Neurosci.* **28**, 7445–7453 (2008).
52. N. Schwartz *et al.*, *Science* **345**, 535–542 (2014).
53. O. P. Carvalho *et al.*, *J. Med. Genet.* **48**, 131–135 (2011).
54. M. S. Minett *et al.*, *Cell Rep.* **6**, 301–312 (2014).
55. M. A. Nassar *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12706–12711 (2004).
56. S. Yang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17534–17539 (2013).
57. J. H. Lee *et al.*, *Cell* **157**, 1393–1404 (2014).
58. C. J. Bohlen *et al.*, *Nature* **479**, 410–414 (2011).
59. S. Diochot *et al.*, *Nature* **490**, 552–555 (2012).
60. I. Daou *et al.*, *eNeuro* 10.1523/ENEURO.0140-15.2016 (2016).
61. S. M. Iyer *et al.*, *Nat. Biotechnol.* **32**, 274–278 (2014).
62. S. I. Park *et al.*, *Nat. Biotechnol.* **33**, 1280–1286 (2015).
63. Z. Z. Xu *et al.*, *Nat. Med.* **21**, 1326–1331 (2015).
64. A. Sakai *et al.*, *Brain* **136**, 2738–2750 (2013).
65. O. A. Samad *et al.*, *Mol. Ther.* **21**, 49–56 (2013).
66. L. Li *et al.*, *Cell Rep.* **15**, 1376–1383 (2016).
67. K. T. Kahle *et al.*, *Sci. Signal.* **9**, ra32 (2016).
68. J. M. Bráz *et al.*, *Neuron* **74**, 663–675 (2012).



## ACKNOWLEDGMENTS

Funding was provided by the Rita Allen Foundation and the American Pain Society to R.P.S.

10.1126/science.aaf8933


REVIEW

# Deconstructing the sensation of pain: The influence of cognitive processes on pain perception


**Katja Wiech**[1,2]*



**Phenomena such as placebo analgesia or pain relief through distraction highlight the powerful influence cognitive processes and learning mechanisms have on the way we perceive pain. Although contemporary models of pain acknowledge that pain is not a direct readout of nociceptive input, the neuronal processes underlying cognitive modulation are not yet fully understood. Modern concepts of perception—which include computational modeling to quantify the influence of cognitive processes—suggest that perception is critically determined by expectations and their modification through learning. Research on pain has just begun to embrace this view. Insights into these processes promise to open up new avenues to pain prevention and treatment by harnessing the power of the mind.**


Watching a captivating film while a dentist is fixing your tooth can help you endure the much-dreaded visit with surprisingly little pain. Cognitive processes, such as distraction, have the potential to change the way we perceive pain—for better or for worse, as I will show below. Based on a rich psychological literature, brain imaging studies in humans have sought to describe and characterize the influence of cognitive factors on the neural processing and perception of pain since the 1980s. I will first give an overview of our current understanding of mechanisms and neural pathways to cognitive pain modulation and highlight the most important recent strands of research in this field, with an emphasis on experimental studies in healthy individuals. In the second part, I will outline how these findings may be integrated with modern concepts of perception by using computational models to explore the influence of cognition on pain at a more fundamental level.

## The search for a "signature of pain in the brain"

A modulatory influence of cognitive factors on the perception of pain has been documented for a number of processes including attention, anticipation, catastrophizing, (re-)appraisal, and perceived control over pain (*1*). Undoubtedly, the most impressive and most extensively studied example is a placebo analgesic response. Patients with agonizing levels of pain can report complete pain relief after administration of a sugar pill they think is a powerful painkiller. But do such changes in pain reports indeed reflect a change in pain perception and neural pain processing—or just a report bias? Answering this question is notoriously difficult given the highly subjective nature of pain. Attempts to identify an objective readout of pain based on brain imaging data are controversially discussed (*2*). Brain activation induced by noxious input might in fact not be pain-specific but also reflect processes that are (inherently) linked to pain, such as the detection of salient events (*3*). Historically, the network of brain regions involved in pain processing ("pain matrix") has been divided into sensory-discriminative and cognitive-affective systems. The sensory-discriminative system, which includes the lateral thalamus and primary and secondary somatosensory cortex (SI and SII, respectively), was thought to process nociceptive input, including its intensity, localization, and quality. In contrast, the cognitive-affective system, comprising regions such as the anterior insula and anterior cingulate cortex, was implicated in psychological aspects of pain. However, this strict dichotomy turned out to be an oversimplification, as sensory-discriminative brain regions, for instance, can also be sensitive to cognitive processes. Moreover, studies using a decoding approach (i.e., the prediction of pain perception based on activation patterns in the brain) demonstrated that the prediction is significantly improved when the different brain regions are considered together. In the so-far most rigorous attempt to characterize the "neurological pain signature" (NPS), Wager and colleagues (*4*) used a machine-learning algorithm to predict the perceived intensity of experimentally induced heat pain in healthy volunteers. The identified network—comprising brain regions such as thalamus, SI, SII, anterior insula, and anterior cingulate cortex—afforded a specificity of about 90% in discriminating physical pain from related phenomena (e.g., "pain" due to social exclusion). Although these first findings are encouraging, further


[1]Oxford Centre for Functional Magnetic Resonance Imaging of the Brain, Nuffield Department of Clinical Neurosciences, University of Oxford, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK. [2]Nuffield Department of Clinical Neurosciences, Nuffield Division Anaesthetics, University of Oxford, John Radcliffe Hospital, Headley Way, Oxford OX3 9DU, UK.
***Corresponding author. Email: katja.wiech@ndcn.ox.ac.uk**

validation of this "objective marker" is needed, including its translation to clinical pain and generalizability to different types of pain and pain modulation.

### The descending pain control system: Top-down modulation of pain

Neuroimaging studies have not only been concerned with target regions of cognitive pain modulation but also with areas that implement the modulation. The so-called descending pain control network comprises regions such as the dorsolateral prefrontal cortex (DLPFC), rostral anterior cingulate cortex, and periaqueductal gray (PAG) (*5*). Activation and functional connectivity between these regions are positively correlated with the level of pain relief reported. The engagement of this modulatory control network has been linked to reduced activation in other pain-related brain regions, albeit with varying consistency. Furthermore, the top-down influence has been shown to alter responses in the spinal dorsal horn, which suggests that it can modulate nociceptive processing at an early stage (*6*). The DLPFC is thought to play a pivotal role in top-down control of pain, because its transient lesioning by transcranial magnetic stimulation (TMS) abolish the placebo analgesia (*7*). Most of our knowledge of the pain-control system stems from neuroimaging studies on placebo analgesia, but this system has also been found to be engaged during other types of cognitive operations leading to pain reduction [e.g., distraction (*8*)]. Descending pain inhibition is largely mediated through endogenous opioids (*9*). There is, however, evidence for the contribution of other neurotransmitters, including cannabinoids (*10*) and dopamine (*11*). Taken together, research on the descending pain control system has described a network that is sensitive to cognitive manipulations and can interact with other brain regions involved in pain processing.

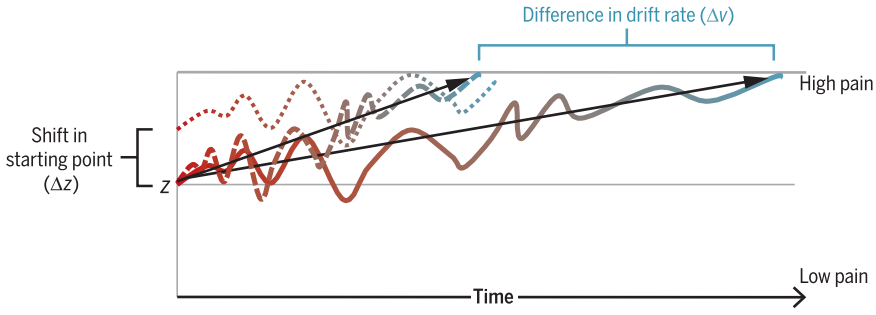
### The frontostriatal system: Valuation of nociceptive input and higher-order integration of different aspects of pain

Do changes in perception based on cognitive modulation differ from those induced by changes in noxious input? Woo and colleagues (*12*) directly compared the modulation of pain through different intensity levels of heat and through cognitive self-regulation of pain in the same individuals. Although the former was indeed reflected in changes in the NPS, self-regulation had no effect on the NPS but was associated with changes in functional connectivity (i.e., the cross-talk between brain regions) of mesolimbic brain structures, including the ventromedial prefrontal cortex (vmPFC) and nucleus accumbens (NAc). This finding is remarkable for two reasons. First, it challenges the concept of the NPS as a universal signature of pain in the brain. If the NPS is to be established as an objective readout of pain, it is expected to reflect changes in pain irrespective of the type of modulation that led to the change in perception. Second, it highlights the contribution of the mesolimbic network that has been

**Fig. 1. Influence of expectations on pain.**

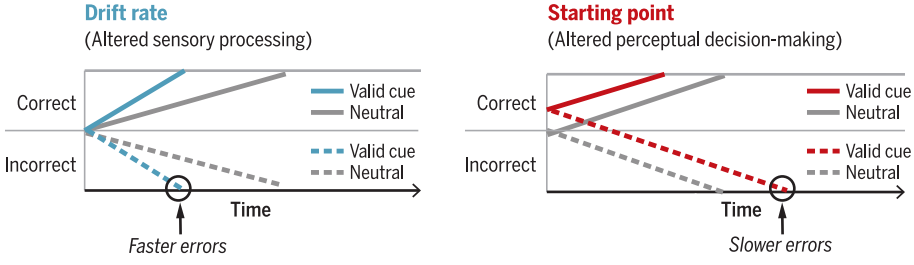
**Diffusion decision model**

Accumulator model within the framework of perception as inference to describe the decision in a binary decision-making task. The decision can be understood as an accumulation of evidence that initiates a response as soon as the decision threshold for one of the options is reached. Without prior information, the probability of both alternatives is identical (thin center line). If high-intensity pain is expected, this prior information could lead to altered sensory processing (dashed line) or a shift in prior toward the high-pain boundary (dotted line).



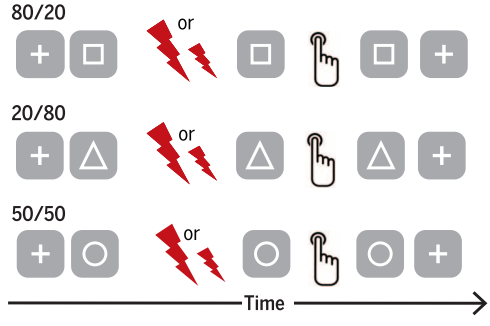

**Two types of bias**

The two types of bias can be distinguished because they result in different response times for incorrect choices.



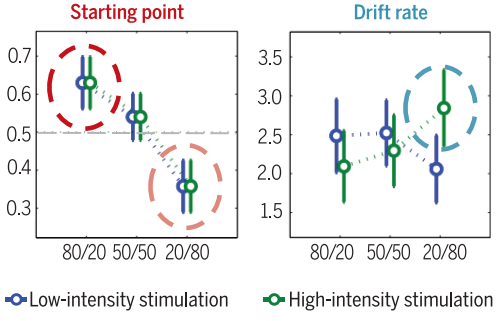

**Experimental design**

Influence of prior information on perceptual decision-making in the context of pain (*30*). Participants saw one of three visual cues on each trial, and these signaled different probabilities for low-intensity and high-intensity pain: (i) 80% high, 20% low; (ii) 20% high, 80% low; or (iii) 50% high, 50% low. Participants had to indicate as quickly as possible whether they received a high- or low-intensity stimulus. Response time and accuracies were recorded.




**Results: Prior information affected perceptual decision-making**

A direct comparison of both types of bias revealed that this influence of prior information was based on a decision-making bias. Expectation shifted the prior toward the expected outcome. This was true for both the expectation of high-intensity pain (red circle) and that of low-intensity pain (pink circle). Only when participants expected a high-pain stimulation but received a low-intensity stimulation was there an increased drift rate (blue circle).




GRAPHIC: ADAPTED BY N. CARY/*SCIENCE*

implicated in learning and valuation rather than in pain processing as such. It could therefore be speculated that this network translates sensory, cognitive, and affective aspects of pain into a "common currency" to integrate them and give rise to one unified pain experience. In a longitudinal study involving brain imaging (*13*), functional and structural characteristics of vmPFC and NAc have been shown to predict the development of chronic pain. On the basis of these findings, it has been postulated that the frontostriatal system is key not only for the conversion of nociception into the perception of pain but also for the transition from acute to chronic pain (*14*, *15*).

## Attention and the influence of spontaneous brain activity on pain perception

Changes in pain perception have not only been observed following deliberate cognitive operations but can also occur spontaneously (*16*). Recent work linked this finding to spontaneous fluctuations in attention to pain that depend on dynamic changes in resting state activity in three distinct brain circuitries (*17*): (i) the "salience network," which is involved in the detection of biologically relevant stimuli and events and comprises brain regions such as mid-cingulate cortex, anterior insula, temporoparietal junction, and DLPFC (*18*); (ii) the default mode network (DMN), which shows a reduced signal level during activity compared with a relaxed nontask state and includes regions such as the medial prefrontal cortex, the posterior cingulate cortex and precuneus, the lateral posterior lobe, and the medial temporal lobe (*19*); and (iii) the descending pain control system described above. Using an experience sampling approach in which participants indicate to which extent they had paid attention to pain, activation in the salience network was found when attention spontaneously focused on pain (*20*). In contrast, the DMN was engaged when attention was focused away from pain (*20*). "Individuals' intrinsic attention to pain" (defined by the test-retest reproducibility of an individual's tendency to attend away from pain) was related to their structural and functional connectivity between DMN and the descending pain control system [and the PAG in particular (*18*)]. Also, alterations in the interplay between the salience network, DMN, and descending pain control network have been related to heightened attention to pain in chronic pain patients (*17*). Although speculative at this time point, research into the neural basis of altered spontaneous focusing on pain might therefore also be relevant for understanding the "interruptive function of pain" (*21*) on concomitant cognitive processes in clinical pain populations.

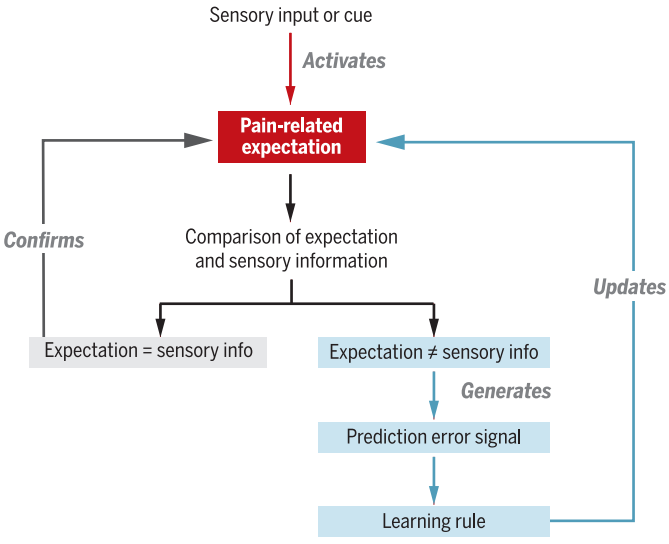



**Fig. 2. Prediction error processing and learning in the context of pain: A schematic overview.** Sensory input or pain-related cues trigger a pain-related expectation. Subsequently, nociceptive input is compared with the expectation that reflects prior information. If incoming information is in line with prior assumptions, the expectation is confirmed. If they diverge, a prediction error signal is generated, and the expectation is updated through a learning rule. Note that the generation of a prediction error might not necessarily lead to a revision of the expectation; following up on prediction errors might selectively be impaired in a pathological state and might contribute to aberrant learning in the context of pain.

## The need for a comprehensive unified framework of cognitive pain modulation

The studies portrayed above have provided important insights into the implementation of cognitive pain modulation. They leave unanswered, however, critical questions regarding its initiation, maintenance, and integration. The processes described in the previous sections need to be carefully orchestrated to integrate momentary demands with long-term goals (*22*) to ensure survival of the organism. What triggers cognitive pain modulation and what prevents or stops it? Furthermore, the actual interface between cognitive processes and pain as a perceptual experience has only insufficiently been described so far. How are cognitive influences woven into the perceptual process that gives rise to the experience of pain?

## The construction of a pain experience: "Perception as inference"

Modern concepts of perception outside the pain field have begun to address these questions by using computational models. In computational modeling, measurable indices of behavior that result from the inferential process (e.g., categorization of stimuli as painful, versus nonpainful, response times) are used to inform a theoretical model, which maps sensory input (e.g., noxious stimulation) onto behavior. Indicators of this mapping are subsequently used in the analysis of functional neuroimaging data to characterize the inferential process. In the most prominent theoretical framework of this lineage, termed "predictive coding," perception is conceptualized as an inferential process in which prior information is used to generate expectations about future perception and to interpret sensory input (*23*). During the perceptual process, incoming sensory information is compared against a "template" or expectation that reflects prior information. The concept of predictive coding acknowledges that we are more likely to perceive sensory information in accordance with our template than with competing interpretations. Perception is thereby understood as a process that favors expected outcomes and weighs down information that is incongruent with the expectation. Evidence for such biased perception in the context of pain comes from studies using explicit cues to signal the intensity or onset of an upcoming noxious stimulus (*24*, *25*) (Fig. 1) or, in more complex paradigms, the predictability or controllability of the stimulation [e.g., (*26*)]. Moreover, the generation of expectations is a shared feature of most cognitive processes that have been related to pain modulation, despite their many differences (*27*). Biased perceptual inference has recently been postulated to contribute to various diseases, including functional motor and sensory syndromes and psychiatric disorders (*28*).

Studies in animals and humans have begun to unravel neural mechanisms underlying the inferential process and to characterize the influence of prior information and expectations (*29*). Collectively, they show that expectations can bias perception by introducing changes not only in sensory brain regions but also in those involved in interpreting the incoming information. This concept extends the traditional view that a cognitive influence has to be implemented in a top-down fashion (as, for instance, reflected in the concept of the descending pain control system) by emphasizing the relevance of higher-order cortical processes that translate incoming information into perception. In line with this notion, it was recently shown that biased decision-making can explain the influence of prior expectations on the perception of pain (*30*). In sum, the concept of "perception as inference" allows for the integration of cognitive factors into the perceptual process itself and highlights the relevance of expectations for perception formation.

## Learning and updating internal models about pain

If the influence of expectations on perception is so profound, why do we not simply foster the most extreme expectations of radical pain relief as part of any pain treatment? Although the influence of expectations is undoubtedly strong, there are clearly limits to the extent expectations can influence the perception of sensory signals

(*31*). Our representations of reality should first and foremost enable us to successfully navigate the environment with minimal costs, which renders delusional ideations impractical. A substantial deviation of our expectations from reality should therefore lead to course correction—an updating or revision of our expectations. Predictive coding and learning models rooted in this approach assume that when expected and observed sensory information diverge, a "prediction error message" is generated in the brain that serves as a teaching signal for model updating (Fig. 2). So far, prediction error (PE) signaling has been studied extensively in the context of reward learning and perceptual decision-making in the visual and auditory systems. However, a limited number of studies have explored aversive PE processing by using noxious stimuli as an aversive experience during learning in humans (*32*–*35*).

But what constitutes an important enough deviation to challenge our current model and trigger updating? Both premature and delayed model updating can be risky. Premature or overly frequent updating might lead to a highly volatile and unreliable model and also absorb large amounts of attentional resources for constant monitoring. Conversely, delayed model updating might bear the risk that true changes are not considered early enough to prevent costly erroneous decisions. Individuals may optimally integrate newly available information into their existing model when short-lasting noxious stimuli are applied under controlled experimental conditions. There is, however, ample evidence that information integration and learning are suboptimal in patients suffering from chronic pain (*36*), and lead to change-resistant mental representations of pain and delayed updating.

***"The concept of pain as an actively constructed experience...has far-reaching implications for pain treatment and prevention."***

Premature updating of expectations, on the other hand, might be particularly detrimental when the maintenance of positive expectations could aid treatment success and their (premature) downward corrections might compromise the outcome. As with placebo analgesia, the experience of insufficient pain relief during treatment might be irreconcilable with the expectation of a successful treatment—and, therefore, lead to model updating (*37*). Treatment expectations are—if at all—often only assessed once before treatment onset, but it seems reasonable to assume that expectations are modified if the treatment effect falls short of pretreatment expectations. This downward adjustment may not only cause patients to drop out of treatment programs but might directly prejudice treatment success, because the inferential processes of pain perception and treatment judgment are no longer supported by a positive expectation of pain relief. Critically, negative treatment experiences have a prolonged effect that can also hamper subsequent unrelated treatment (*38*). In sum, learning models characterize the modification of expectations when new information becomes available and could be applied to explore aberrant learning that is frequently found in the context of chronic pain.

## Conclusions and outlook

With the departure from a rather rigid concept of pain as a direct readout of sensory input, neuroimaging of cognitive pain modulation has provided valuable insights into the complex nature of pain and its neural basis. However, further research is needed to integrate the various efforts into a coherent model that addresses all aspects of modulation, including its initiation, implementation, and monitoring, and also includes the existing behavioral and neuroimaging literature (*39*). The concept of predictive coding could provide the theoretical framework for this endeavor. For example, Büchel and colleagues (*37*) pointed out that although the descending pain-control system is commonly interpreted as a top-down influence, its constituting brain regions have reciprocal connections, which allow for up- and downward projections of information. They therefore suggest that—in line with the concept of predictive coding—the descending system could be part of a larger recurrent network exchanging PE signals at all levels of the neuraxis. Several findings point at a critical role of the DLPFC in orchestrating this network. As described above, the DLPFC is part of various networks that are involved in cognitive pain modulation. It plays a key role in evidence accumulation during perceptual decision-making, as described in the context of perception as inference (*40*) and learning. The integration of these different strands of research with respect to DLPFC functioning and its governance of learning networks can be expected to provide the much-needed unifying model of neural mechanisms underlying cognitive pain modulation.

The concept of pain as an actively constructed experience that is determined by expectations and their modification through learning has far-reaching implications for pain treatment and prevention. Treatment success is known to be critically depending on patients' expectations, not only in the context of placebos but also with active interventions such as analgesic drugs (*41*). Expectations—in turn—are shaped by the information that is provided by health care practitioners. How could information be designed to optimally guide expectations for maximum treatment outcome? How could aberrant information processing be addressed using the framework of predictive coding? Future research should explore the translation of research on the inferential process underlying the perception of pain into clinical practice to optimally inform pain prevention and treatment strategies.

Patients' complaints about pain that persists despite numerous treatment attempts are often dismissed as being "all in their head." Modern pain research has shown that this notion is in fact true for any kind of pain, acute or chronic, easy to treat or resistant to all treatments currently available. We are only beginning to understand that the head (or the brain, for that matter) also holds the key to new ways to help patients conquer their pain.

### REFERENCES AND NOTES


1. K. Wiech, I. Tracey, *Front. Neurosci.* **7**, 46 (2013).
2. L. Hu, G. D. Iannetti, *Trends Neurosci.* **39**, 212–220 (2016).
3. A. Mouraux, A. Diukova, M. C. Lee, R. G. Wise, G. D. Iannetti, *Neuroimage* **54**, 2237–2249 (2011).
4. T. D. Wager *et al*., *N. Engl. J. Med.* **368**, 1388–1397 (2013).
5. I. Tracey, P. W. Mantyh, *Neuron* **55**, 377–391 (2007).
6. F. Eippert, J. Finsterbusch, U. Bingel, C. Büchel, *Science* **326**, 404 (2009).
7. P. Krummenacher, V. Candia, G. Folkers, M. Schedlowski, G. Schönbächler, *Pain* **148**, 368–374 (2010).
8. F. Eippert, C. Buechel, in *Placebo and Pain: From Bench to Bedside*, L. Colloca, M. Flaten, K. Meissner, Eds. (Academic Press, Waltham, MA, ed. 1, 2013), pp. 53–71.
9. F. Eippert *et al*., *Neuron* **63**, 533–543 (2009).
10. F. Benedetti, M. Amanzio, R. Rosato, C. Blanchard, *Nat. Med.* **17**, 1228–1230 (2011).
11. M. Peciña, J. K. Zubieta, *Mol. Psychiatry* **20**, 416–423 (2015).
12. C.-W. Woo, M. Roy, J. T. Buhle, T. D. Wager, *PLOS Biol.* **13**, e1002036 (2015).
13. M. N. Baliki *et al*., *Nat. Neurosci.* **15**, 1117–1119 (2012).
14. M. N. Baliki, A. V. Apkarian, *Neuron* **87**, 474–491 (2015).
15. J. P. Rauschecker, E. S. May, A. Maudoux, M. Ploner, *Trends Cogn. Sci.* **19**, 567–578 (2015).
16. M. Ploner, M. C. Lee, K. Wiech, U. Bingel, I. Tracey, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 355–360 (2010).
17. A. Kucyi, K. D. Davis, *Trends Neurosci.* **38**, 86–95 (2015).
18. W. W. Seeley *et al*., *J. Neurosci.* **27**, 2349–2356 (2007).
19. M. E. Raichle *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 676–682 (2001).
20. A. Kucyi, T. V. Salomons, K. D. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18692–18697 (2013).
21. C. Eccleston, G. Crombez, *Psychol. Bull.* **125**, 356–366 (1999).
22. G. Crombez, C. Eccleston, S. Van Damme, J. W. Vlaeyen, P. Karoly, *Clin. J. Pain* **28**, 475–483 (2012).
23. K. Friston, *Neural Netw.* **16**, 1325–1352 (2003).
24. T. Koyama, J. G. McHaffie, P. J. Laurienti, R. C. Coghill, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12950–12955 (2005).
25. J. R. Keltner *et al*., *J. Neurosci.* **26**, 4437–4443 (2006).
26. T. V. Salomons, T. Johnstone, M. M. Backonja, R. J. Davidson, *J. Neurosci.* **24**, 7199–7203 (2004).
27. L. Y. Atlas, T. D. Wager, *Neurosci. Lett.* **520**, 140–148 (2012).
28. M. J. Edwards, R. A. Adams, H. Brown, I. Pareés, K. J. Friston, *Brain* **135**, 3495–3512 (2012).
29. C. Summerfield, F. P. de Lange, *Nat. Rev. Neurosci.* **15**, 745–756 (2014).
30. K. Wiech *et al*., *Curr. Biol.* **24**, R679–R681 (2014).
31. G. Crombez, K. Wiech, *Pain* **152**, 1449–1450 (2011).
32. A. Ploghaus *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **97**, 9281–9286 (2000).
33. B. Seymour *et al*., *Nat. Neurosci.* **8**, 1234–1240 (2005).
34. M. Roy *et al*., *Nat. Neurosci.* **17**, 1607–1612 (2014).
35. S. Zhang, H. Mano, G. Ganesh, T. Robbins, B. Seymour, *Curr. Biol.* **26**, 52–58 (2016).
36. J. W. S. Vlaeyen, *Pain* **156** (suppl. 1), S86–S93 (2015).
37. C. Büchel, S. Geuter, C. Sprenger, F. Eippert, *Neuron* **81**, 1223–1239 (2014).
38. S. Kessner *et al*., *PLOS ONE* **9**, e109014 (2014).
39. M. Ploner, U. Bingel, K. Wiech, *Trends Cogn. Sci.* **19**, 180–182 (2015).
40. M. J. Mulder, L. van Maanen, B. U. Forstmann, *Neuroscience* **277**, 872–884 (2014).
41. U. Bingel *et al*., *Sci. Transl. Med.* **3**, 70ra14 (2011).


### ACKNOWLEDGMENTS


K.W. would like to thank U. Bingel, F. Eippert, T. Makin, M. Ploner, J. O'Reilly, I. Tracey, and N. Weiskopf for valuable comments and the U.K. Medical Research Council for research support.

10.1126/science.aaf8934

REVIEW

# Exploring pain pathophysiology in patients

**Claudia Sommer***

**Although animal models of pain have brought invaluable information on basic processes underlying pain pathophysiology, translation to humans is a problem. This Review will summarize what information has been gained by the direct study of patients with chronic pain. The techniques discussed range from patient phenotyping using quantitative sensory testing to specialized nociceptor neurophysiology, imaging methods of peripheral nociceptors, analyses of body fluids, genetics and epigenetics, and the generation of sensory neurons from patients via inducible pluripotent stem cells.**

When we speak of pain, we understand it as a deeply human experience, because it involves not only nociception and the immediate physiological reactions to it, but also emotional, cognitive, and social consequences. Therefore, the ideal subject for the study of pain is the human being, and the best animal experiments or in vitro studies can only give information about partial aspects of the phenomenon of pain. However, pain research in humans, and even more so in patients with pain, is greatly restricted, for a number of reasons. Despite ever-improving imaging and neurophysiological methods, our access to the human organism is limited, and ethical restrictions apply to all manipulations involving volunteers or patients. Human beings are exposed to a myriad of influences, both over a lifetime and at any given moment, which makes standardization much more difficult than in an inbred rodent colony. With all these caveats, the study of pain pathophysiology in humans has made great progress in recent years. Some of these findings, including work by the author's group, will be highlighted in this Review with a focus on neuropathic and other chronic pain and on such findings that have already led to a change in patient management or are likely to do so.

## Sensory phenotyping—not all pain is alike

Neuropathic pain is defined as pain due to a lesion or disease of the somatosensory system (*1*). It has been hypothesized that the pattern of sensory loss and gain caused by such a lesion might reflect the underlying pathophysiology and give guidance toward individualized therapy. Quantitative sensory testing (QST) is a psychophysical method that uses a battery of sensory stimuli with predetermined physical properties following specific protocols (*2*, *3*). QST is thus able to capture and quantify stimulus-evoked negative and positive sensory phenomena (Fig. 1). Physical stimuli are transduced by activating specific receptors, and resulting physiological signals are then conducted in specialized peripheral nerve fibers and central pathways. The nerve fibers involved have unique physiologic properties, such that a sensory deficit or hyperphenomenon detected by QST should allow conclusions as to which type of nerve fibers or their central connections are affected.

QST has uncovered involvement of small nerve fibers (i.e., nociceptors) in a number of disorders where it was unexpected, such as Parkinson's disease, neuropathies that so far have been considered to be of the large-fiber type only, and fibromyalgia syndrome (*4*, *5*). It has been inferred from such findings that primary or secondary peripheral nociceptor involvement may be part of the pathophysiology of pain in these disorders. In diseases where nociceptor malfunction was already known to play a role, QST has helped to identify patterns—i.e., sensory profiles—that may give clues toward the individual pathophysiology of pain and the potential treatment response. For example, the antiepileptic oxcarbazepine was more effective in reducing neuropathic pain in patients with preserved (and potentially hyperactive) nociceptor function, according to QST, than in patients with loss of nociceptor function (*6*). By contrast, a topical lidocaine patch, expected to work well if it finds active nociceptors to work on, was equally efficient in nociceptor-deprived skin (*7*, *8*), whereas one recent study came to the opposite conclusion that lidocaine was more efficacious in skin with the QST features of overactive nociceptors (*9*). Another topical treatment, the 8% capsaicin patch, seems to work better in patients with QST signs of hyperalgesia, whereas indicators for small–nerve fiber loss had no predictive value (*10*). In a recent analysis of 361 patient records from trials with antidepressants and anticonvulsants, the usefulness of sensory phenotyping by QST for decisions on individualized treatment was found to be very limited (*11*). Many researchers and clinicians have adopted the QST protocol described by the German Research Network on Neuropathic Pain (DFNS), and this standardization has been of great value for multinational studies. Others have argued that by focusing on thresholds, the dynamic stimulus–effect relationship and suprathreshold-nociceptive processing are neglected in this protocol. Therefore, the issue of whether QST sensory threshold phenotyping helps guide individual treatment is still open and will be reanalyzed with more selective treatments and more comprehensive patient phenotyping. To understand pain pathophysiology, QST data are particularly valuable when put into relation with neurophysiological and molecular phenotyping.

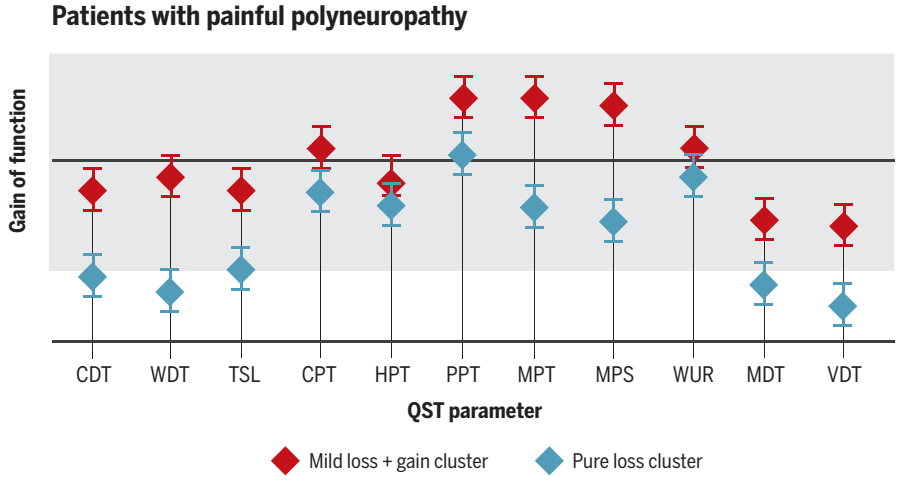



**Fig. 1. A hypothetical example of two types of sensory profiles (clusters) typically found when analyzing QST data from large groups of patients ($n \approx 150$ in each group) with painful polyneuropathies.** Positive $z$ scores indicate positive sensory signs (hyperalgesia), negative $z$ values indicate negative sensory signs (hypoesthesia, hypoalgesia). The shaded area indicates the 95% confidence interval for healthy subjects ($-1.96 < z < +1.96$). Red symbols indicate a cluster with mild sensory loss and gain in the mechanical parameters; blue symbols indicate a cluster with sensory loss. WDT, warm detection threshold; CDT, cold detection threshold; TSL, detection of temperature changes; CPT, cold pain threshold; HPT, heat pain threshold; PPT, pressure pain threshold; MPT, mechanical pain threshold to pinprick; MPS, mechanical pain sensitivity; WUR, wind-up ratio; MDT, mechanical detection threshold to von Frey hairs; VDT, vibration detection threshold.

GRAPHIC: K. SUTLIFF/*SCIENCE*

Department of Neurology, University of Würzburg, Josef-Schneider-Straße 11, D-97080 Würzburg, Germany.
***Corresponding author. Email: sommer@uni-wuerzburg.de**

### Neurophysiology–nociceptor function can be captured, if properly done

Classical sensory neurophysiology measures large nerve fibers and their afferent connections. This is true for standard nerve conduction studies and somatosensory evoked potentials. Special techniques are available to stimulate and record the activity of small nerve fibers and their afferent connections. In general, a pathological finding concerning evoked potentials does not provide information about where in the course from periphery to central nervous system (CNS) the lesion is located (Fig. 2). Among the neurophysiological stimulation techniques selectively activating Aδ and C fibers, the recording of laser-evoked potentials (LEPs) has the longest history and is best established (*12*, *13*). A contact-free 20-ms radiant-heat pulse is applied, which is absorbed by the upper 20 to 50 μm of the skin. Some of the recent contributions of LEP studies to the pathophysiology of pain include support for spinothalamic pathway damage as a predictor for pain in patients with multiple sclerosis (*14*) and after stroke (*15*).

Contact heat–evoked potentials (CHEPs) are a method for which normative data have recently been published (*16*, *17*). CHEPs reflect Aδ activity, but can be used to analyze C-fiber function using special techniques (*18*). It appears that CHEP amplitudes may reflect the reported magnitude of evoked pain (*17*). The "pain-related evoked potentials" (PREPs) are electrically evoked Aδ fiber potentials (*19*, *20*) that correlate with skin innervation in homogenous populations (*19*).

In even more sophisticated setups, neurophysiology combined with psychophysics has been used to uncover pain mechanisms—for example, using the concept of conditioned pain modulation (*21*). Here, a protocol testing temporal summation and conditioned pain modulation was used. For temporal summation testing, a train of stimuli of constant intensity is administered, and a typical increase in pain ratings by the patient along the series is expected. For conditioned pain modulation, heat pain ratings were taken from one arm, while the other arm was immersed in a hot water bath. Patients exhibiting increased ratings in the temporal summation protocol and deficient conditioned pain modulation are supposed to be lacking endogenous pain inhibition, thus having a "pro-nociceptive" pattern of pain modulation (*21*). Interestingly, those patients who had deficient conditioned pain modulation responded best to duloxetine, a serotonin noradrenalin reuptake inhibitor, assumed to work by strengthening descending pain inhibition. Not surprisingly, patients with fibromyalgia syndrome had greatly impaired conditioned pain modulation (*22*). Thus, such complex neurophysiological tests may be valuable tools for assessing pain pathophysiology and potential response to treatment.

Microneurography uses a very different approach. Here, a microelectrode is directly inserted into a nerve fascicle, in close contact to the axons, and spontaneous or evoked nerve discharges are recorded, such that the individual firing behavior of C fibers can be assessed. Thus, whereas an abnormality found in any of the evoked-potential techniques may arise from anywhere in the neuraxis, an abnormality found in microneurography arises from the studied axon on site. Using microneurography and advanced techniques such as threshold tracking in humans, researchers have recently shown spontaneous activity in nociceptors in diabetic and other painful neuropathies (*23*, *24*). Spontaneous activity of C fibers was also shown in fibromyalgia syndrome (*25*), a disorder that was generally thought to be a CNS disorder.

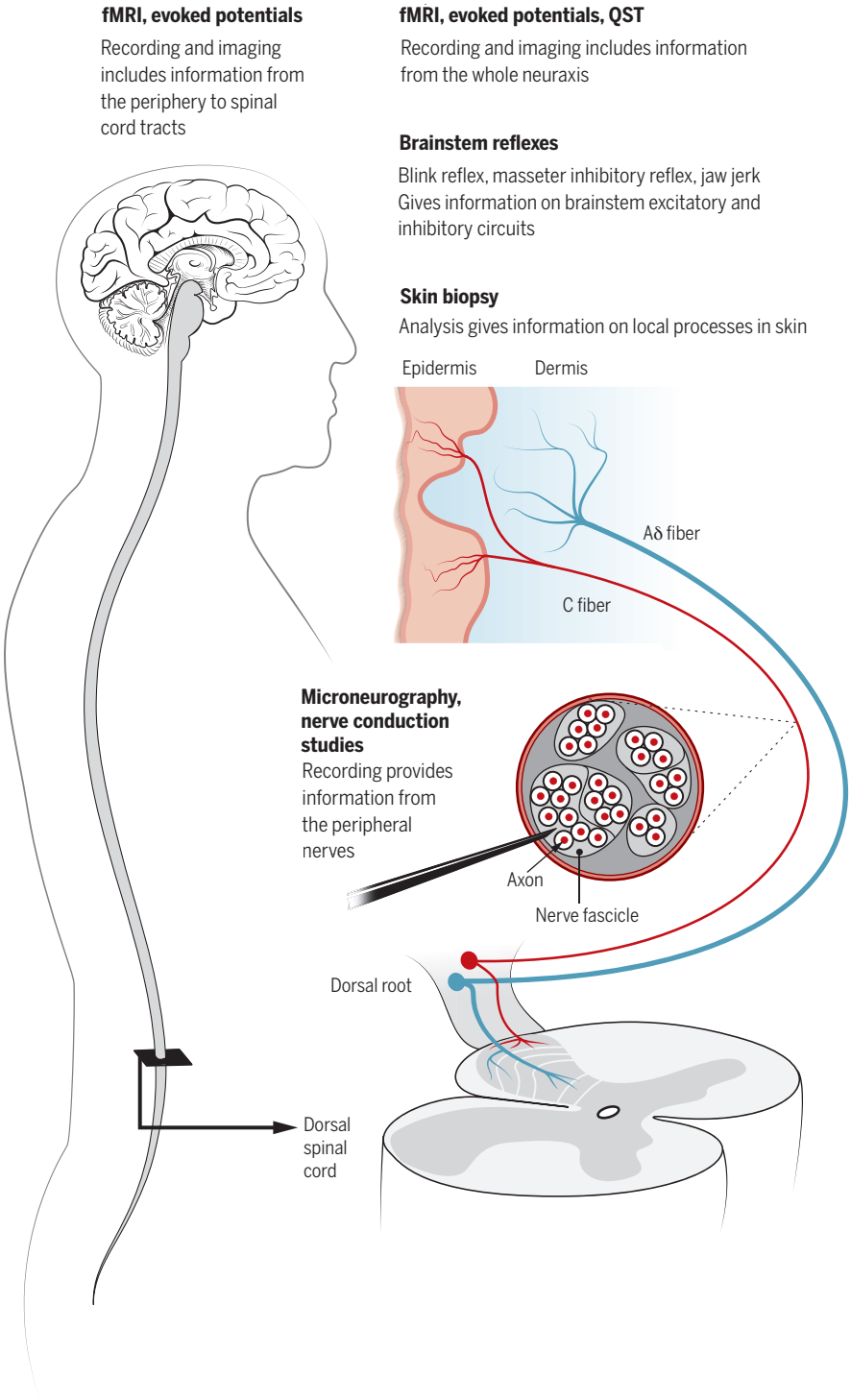



**Fig. 2. Schematic illustrating the information potentially gained from neurophysiological and skin biopsy studies in humans.** When stimulating in the periphery and recording from the CNS, there is information about the entire course of the somatosensory system. Any abnormality detected cannot be localized to a particular neuroanatomic level.

GRAPHIC: K. SUTLIFF/*SCIENCE*

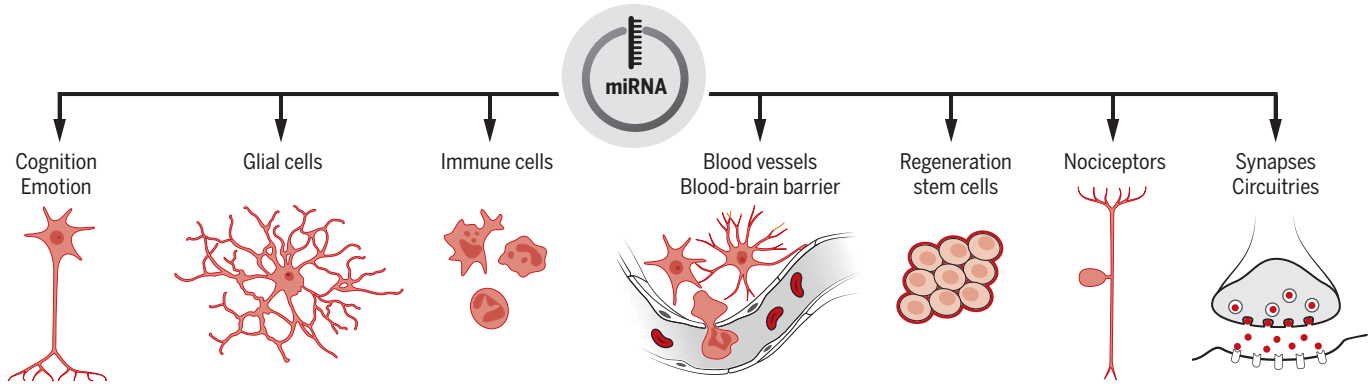



**Fig. 3. microRNAs (miRNAs) as an example of a "master switch" that could regulate numerous pain-related pathways at once.**

Moreover, the pathological C-fiber patterns ("slowing") were different between fibromyalgia and small-fiber neuropathy (SFN), indicating distinct pathophysiological processes behind these disorders, both of which are associated with reduced skin innervation by C fibers (see below).

### Nociceptor histology and noninvasive nociceptor imaging—imaging pain in the periphery

The experience of pain needs a brain, and thus brain imaging with its different modern techniques, such as brain and spinal cord functional magnetic resonance imaging, voxel-based morphometry, quantitative arterial spin labeling, near-infrared spectroscopy (NIRS), and advanced positron emission topography, provide valuable information. These techniques are beyond the scope of this article and will in part be covered in an accompanying article (*26*). Here, I will summarize the methods that allow the very distal parts of the nociceptors to be visualized.

The analysis of nociceptors in skin punch biopsies has moved the field forward considerably. Like functional aspects of skin innervation—as, for example, assessed by QST—morphological quantification of skin innervation has been used to find predictors for drug responses. Local injections of botulinum toxin were more effective in patients with neuropathic pain with preserved skin innervation than in those with marked denervation (*27*).

The whole complex of SFN, an often painful neuropathy with reduced temperature sensitivity and burning feet but lack of classical neuropathy signs, could only be defined after these patients had been investigated with intraepidermal nerve fiber (IENF) quantification in skin biopsies (*28*). IENF quantification has helped define SFN (*29*), to identify length-dependent and non–length-dependent subgroups and point toward a cytokine-related pathogenesis of the pain in the latter (*30*). Small-fiber pathology in the skin has also been detected in disorders where it was not expected, thus extending the ideas on their pathophysiology. Examples include complex regional pain syndrome type I (CRPS I) (*31*), peripheral arterial disease (*32*), and fibromyalgia syndrome (*5*). The finding of decreased skin innervation in a subset of patients with fibromyalgia is of particular interest, for several reasons. First, for decades, this syndrome had been regarded as a chronic pain disorder based on faulty pain processing in the CNS (*33*), or as a somatoform disorder (i.e., not based on organic disease at all). Second, the initial finding was rapidly replicated by other groups (*34*, *35*) and supported by evidence for functional changes using QST and evoked potentials (*5*, *35*), and in particular microneurography (*25*). Third, the finding led to a vivid discussion in the scientific community over whether all disorders showing a reduced IENF density should be called "small-fiber neuropathy."

> ***"Chemosensitive or sensitized nociceptors increase their activity upon stimulation with the appropriate chemical or combination of chemicals..."***

Although this notion sounds attractive, there are some important counterarguments. The neurophysiological characteristics of remaining, viable nociceptors have been shown to be different between the different disorders (*25*). The abnormally high slowing of C nociceptors in conduction velocity when stimulated at 0.25 Hz was only present in fibromyalgia, not in SFN, indicating different mechanisms underlying the loss of nerve fibers in both conditions. This phenomenon might be due to a thinner diameter of these axons (*36*). Showing a peripheral nervous system pathology in fibromyalgia patients, however, does not preclude a CNS component of the disorder, as postulated by others (*33*). Indeed, our own findings using NIRS to assess brain activity in fibromyalgia patients upon painful stimuli point toward an additional central amplification (*37*).

Although skin biopsy as a method to image nociceptors is minimally invasive, other, even less invasive methods have been developed. One of them is corneal confocal microcopy (CCM), which images nociceptors in the cornea with high resolution in real time and allows their quantification. The density of corneal innervation appears to correlate with skin innervation at the leg (where it is usually assessed in patients with painful neuropathies) (*38*). In diabetic neuropathy, where CCM has been most extensively used, CCM is considered valuable for detecting and assessing early nerve damage (*39*). How far this can be generalized and what else can be learned from CCM in particular for the study of pain remain to be seen. In fibromyalgia, as in skin biopsy studies, small-fiber pathology was detected in about 50% of patients by CCM (*40*). Different subgroups of fibromyalgia patients could be identified with a combination of CCM parameters and sensory testing, which may lead to more personally tailored and efficient treatment.

### Pain mediators in body fluids and tissue—reviving the "inflammatory soup"?

Chemosensitive or sensitized nociceptors increase their activity upon stimulation with the appropriate chemical or combination of chemicals, once termed "inflammatory soup" (*41*). The increase of neuronal activity in in-vitro preparations and in experimental animals after application of algesic mediators such as prostaglandins and proinflammatory cytokines is well established. In humans, one of the questions is, which body fluid or tissue should be investigated to correlate levels of such mediators with measures of pain? Then, if such a correlation is found, there remains the question of whether this is a causal relationship or a chance finding. Longitudinal studies and, in particular, studies with interventions directed at the mediator in question may provide evidence for one or the other. Having pursued the subject of inflammatory mediators in patients with neuropathic pain over many years, our group could show such correlations in some but not in all situations tested. A cohort of patients with chronic widespread pain had a systemic proinflammatory cytokine profile with reduced blood concentrations of anti-inflammatory cytokines (*42*). Local cytokine expression in the skin was not different in fibromyalgia compared

to controls (*43*). Unexpectedly, in patients with postherpetic neuralgia, systemic cytokine profiles were not different from those of controls, and local cytokine concentrations in affected skin were not different from those on the unaffected side (*44*). Systemic elevations of proinflammatory cytokines were found in patients with complex regional pain syndromes (CRPS) (*45*) and with painful (but not in painless) peripheral neuropathy (*46*). In length-dependent SFN, a biopsy specimen from the distal, painful skin harbored an increased expression of proinflammatory cytokines compared to proximal, nonpainful skin (*30*). Analyzing cytokine profiles in local blister fluid from patients with CRPS, we found bilaterally increased proinflammatory tumor necrosis factor–α and macrophage inflammatory protein– 1β and decreased anti-inflammatory interleukin-1 (IL-1) receptor antagonist protein concentrations compared to non-CRPS patients (*47*). After 6 months of analgesic treatment, the increased cytokine protein levels in CRPS patients reverted to the level of non-CRPS patients. Such findings, similar to those by Wang *et al.* (*48*), who showed an increase in systemic IL-8 in patients with fibromyalgia that was normalized after a course of multidisciplinary therapy, support an interpretation of a pathophysiological importance. However, comparison between studies is hampered by the wide range of methods used. Furthermore, in contrast to data from animal studies, these correlations are far from providing proof of a causal relationship. Intervention studies directed at the respective mediators would be needed to show a causal link. Unbiased modern methods like proteomics may provide a further solution to this problem (*49, 50*). A factor that may have a major impact on the balance of pro- and anti-inflammatory processes is the gut microbiome. Bidirectional influences between gut microorganisms, the cytokine balance, stress and the hypothalamic-pituitary-adrenal axis, and pain have been postulated (*51*).

## Pain is in the genes—but what manipulates them?

Large-scale genomics analyses of patient samples have thus far been disappointing. By contrast, monogenetic pain disorders, such as the diseases associated with mutations in voltage-gated sodium channels ($Na_v$), erythromelalgia, paroxysmal extreme pain disorder, congenital insensitivity to pain, and painful hereditary SFN, have been immensely instructive as models of human pain disorders (*52*). Gain-of-function mutations of $Na_v1.7$ ion channels underlie erythromelalgia and paroxysmal extreme pain disorder, and a loss-of-function mutation in the same gene induces congenital analgesia. Although these disorders are rare, up to 30% of a cohort of patients with idiopathic SFN have mutations in the genes for the sodium channels $Na_v1.7$, $Na_v1.8$, and $Na_v1.9$ (*52*). Thus, the important role of voltage-gated sodium channel activity in nociceptors for pain perception has been convincingly shown in humans, and detailed information on the malfunction of the channel in different mutations has been gained (*53*).

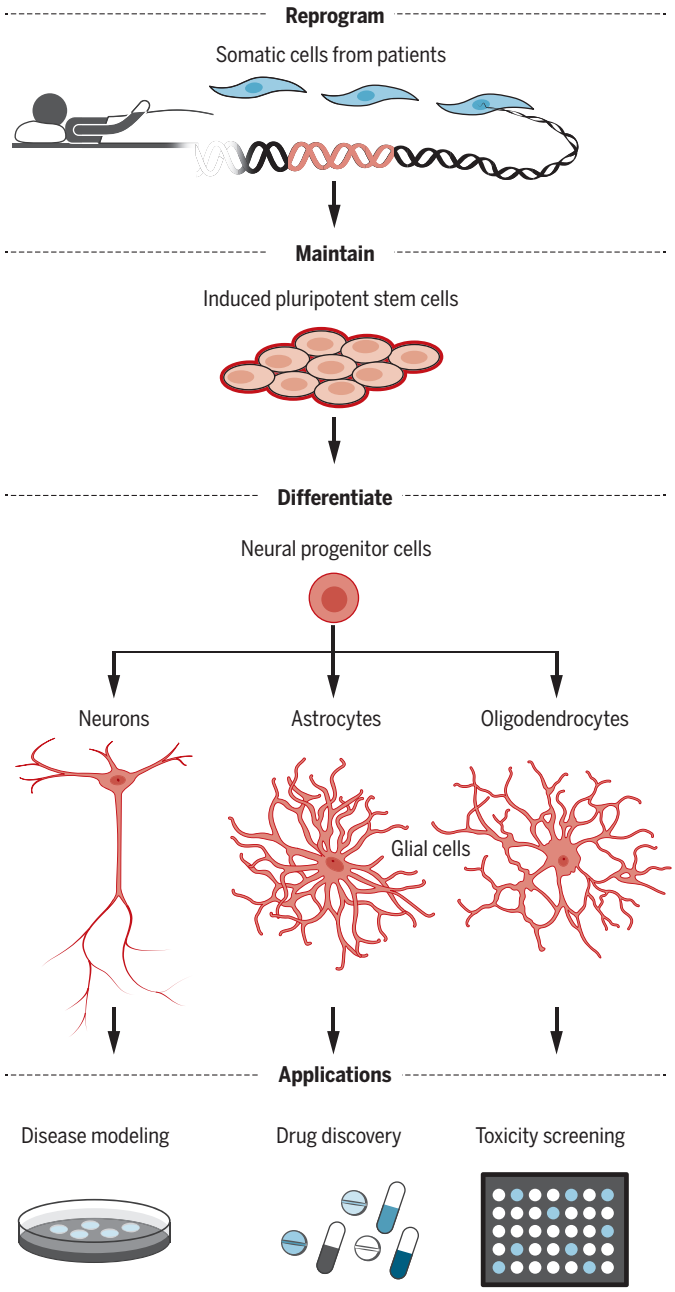



**Fig. 4. Somatic cells from patients can be reprogrammed to neuronal or glial cells, which can be used to study pathophysiology or screen for drug treatment.**

GRAPHIC: K. SUTLIFF/*SCIENCE*

A single-nucleotide polymorphism in the *SCN9A* gene [rs6746030, substitution of arginine-1150 with tryptophan (R1150W)] that is present in ~18% of Caucasians (*54*, *55*) may predispose to increased pain sensation. Further studies have demonstrated an association of this polymorphism with pain in Parkinson's disease, with multiple regional pain syndromes, and with pain in interstitial cystitis. Because not every person with this polymorphism develops a pain disease, additional factors influencing the phenotype have been postulated. We described a family with exercise-induced pain and the R1150W polymorphism that in addition had a distal demyelinating sensory neuropathy (*56*). In these patients, as in patients with erythromelalgia, a sodium channel–blocking drug was an effective analgesic.

Human diseases caused by mutations in the genes for nerve growth factor (NGF) or its tyrosine kinase receptor TrkA illustrate the importance of NGF for pain perception. These rare autosomal recessive mutations lead to congenital insensitivity to pain or decreased pain perception (*57*). The disorders are termed hereditary sensory and autonomic neuropathy (HSAN) type IV and type V. HSAN-IV, also called congenital insensitivity to pain with anhidrosis (CIPA), is caused by mutations in the gene coding for TrkA. In this severe congenital disease, the affected children do not respond to painful stimuli, which may lead to mutilations, Charcot arthropathy, osteomyelitis, and autoamputation. HSAN-V is caused by mutations in the *NGF* gene. Accordingly, drugs blocking NGF have strong analgesic actions in clinical trials on osteoarthritic and neuropathic pain (*58*).

Sometimes a single mutation in a pain-relevant gene within one family may lead to quite different pain phenotypes (*59*). Epigenetic modification of genes is one factor that explains such variability. In animal experiments, DNA methylation correlated between brain and peripheral blood cells (*60*), which encourages the use and interpretation of methylation in human blood cells as surrogates. Indeed, a twin study showed a relation between methylation of the pain-related ligand-gated ion channel *TRPA1* gene, where higher methylation and lower expression of the gene was correlated with higher pain sensitivity (*61*). Methylation of the promoter domain of the endothelin 1B receptor gene was shown in patients with oral squamous cell carcinoma, but only if these were painful (*62*). Many more examples

of methylation of pain-related genes exist in the animal literature, but human studies are lagging behind.

With the exception of rare monogenetic diseases like the ones mentioned above, changes in one gene or its product often do not have a major effect on pain perception. Thus, researchers have been looking for "master switches" that would regulate multiple gene products and orchestrate multiple pathways, acting on circuitries rather than on individual receptors or targets. MicroRNAs have the potential to function as such master switches (Fig. 3) and have been suggested to link pain and inflammation (*63*). MicroRNA "signatures" have been proposed for several pain diseases such as neuropathic pain (*64*) and fibromyalgia (*65*, *66*). Again, such human microRNA profiles can provide hypotheses, but causal connections can only be tested in animal models, until specific drugs for human use become available.

## Re-creating patient cells and tissue

The possibility of creating patient neurons from pluripotent stem cells (PSCs) out of patient blood cells or tissue cells (for example, skin fibroblasts) has opened entirely new horizons for patient-related research on pain pathophysiology (*67*) (Fig. 4). Because the induced PSCs (iPSCs) and the neurons created from them retain the genetic identity of the donor patient, these cells can be used to study various aspects of disease pathophysiology. A striking example is the recent report on iPSC-derived sensory neurons from patients with inherited erythromelalgia caused by a *SCN9A* mutation (*68*). These neurons, like the patients, had hyperexcitability and aberrant responses to heat stimuli. Furthermore, the heat hypersensitivity could be reversed by a blocker of the mutated ion channel, $Na_v1.7$. The same drug had a clinical effect on the patients' pain symptoms.

iPSCs have also been created from patients with Fabry's disease (*69*), a monogenetic disorder with a characteristic pain phenotype, and the potential to mature them into sensory neurons will help to elucidate the pain mechanism. The same option is certainly available for other monogenetic pain disorders, and such neurons will lend themselves to high-throughput screening methods for modulators of neuropathic pain. The technique is still new, and major challenges need to be overcome. Apart from the technical issues and the question of standardization between laboratories, iPSC-derived neurons will have to display a measurable disease-specific phenotype with a stable expression across passages in order to be useful as a disease model and potential drug-screening platform.

Organotypic cultures, thus far realized using human skin cells and rodent nerves to innervate them (*70*), may provide a further step toward human disease models, preferably when constructed entirely out of patient cells. Such organotypic cultures can be studied with high-throughput microscopy, proteomics, and electrophysiology, and even optogenetic methods may be applied (*71*).

## Conclusions—What we know and what we would like to know

Ideally, we would like to understand the underlying pathophysiology of pain in each patient on an individual basis. With the exception of the rare monogenetic diseases with overactive sodium channels, we are still far from this goal.

Other monogenetic disorders associated with typical pain states like hereditary migraines, Fabry disease, sickle cell disease, and others, provide opportunities to uncover more of these mechanisms in model diseases. For the major pain diseases like osteoarthritis and low back pain, twin studies and population studies have shown that genetic risk factors can explain some of the individual differences in pain perception, but that environmental factors are equally important. These studies will continue to provide valuable information on pain risk factors. The yield from genome-wide association studies may improve when more uniform cohorts are included, and ethnicity and other variables are more stringently considered. The field of gene–gene and gene–environment interactions is just beginning to be explored regarding human pain disorders. Epigenetic factors include and are influenced by diet, exercise, and the social environment. Prospective cohort studies may give insight into these interactions.

We have tools for assessing sensory phenotypes, including pain thresholds and hypersensitivity, most of which have been well standardized with international agreement. Consequently, using these in clinical trials may help to identify responder subgroups in trials that would otherwise have a negative outcome. In the research setting, more information may be obtained from individual patients by combining the threshold techniques with methods assessing suprathreshold function and central pain processing. Combining such results with those from genomics and proteomics may push the field forward. Imaging methods of the brain and spinal cord will complement these findings.

Methods for gaining information out of biomaterials from patients have improved considerably, including the generation of neurons out of iPSCs, organotypic cultures, and high-throughput methods. It can be expected that combined with technologies like neurophysiology, proteomics, and optogenetics, human in vitro models will lead to rapid advances in pain pathophysiology and drug response.

### REFERENCES AND NOTES


1. R. D. Treede *et al.*, *Neurology* **70**, 1630–1635 (2008).
2. M. M. Backonja *et al.*, *Clin. J. Pain* **25**, 641–647 (2009).
3. R. Rolke *et al.*, *Pain* **123**, 231–243 (2006).
4. M. Nolano *et al.*, *Brain* **131**, 1903–1911 (2008).
5. N. Üçeyler *et al.*, *Brain* **136**, 1857–1867 (2013).
6. D. T. Demant *et al.*, *Pain* **155**, 2263–2273 (2014).
7. G. Wasner, A. Kleinert, A. Binder, J. Schattschneider, R. Baron, *J. Neurol.* **252**, 677–686 (2005).
8. D. N. Herrmann, V. Pannoni, R. L. Barbano, J. Pennella-Vaughan, R. H. Dworkin, *Muscle Nerve* **33**, 42–48 (2006).
9. D. T. Demant *et al.*, *Pain* **156**, 2234–2244 (2015).
10. T. Mainka *et al.*, *Eur. J. Pain* **20**, 116–129 (2016).
11. J. V. Holbech, F. W. Bach, N. B. Finnerup, T. S. Jensen, S. H. Sindrup, *Pain* **157**, 1305–1313 (2016).
12. R. M. Dotson, *J. Clin. Neurophysiol.* **14**, 32–45 (1997).
13. R. D. Treede, J. Lorenz, U. Baumgärtner, *Neurophysiol. Clin.* **33**, 303–314 (2003).
14. A. Truini *et al.*, *Pain* **153**, 2048–2054 (2012).
15. N. Vartiainen *et al.*, *Brain* **139**, 708–722 (2016).
16. V. Lagerburg *et al.*, *Muscle Nerve* **51**, 743–749 (2015).
17. Y. Granovsky *et al.*, *Pain* **157**, 1156–1163 (2016).
18. C. S. Madsen, B. Johnsen, A. Fuglsang-Frederiksen, T. S. Jensen, N. B. Finnerup, *Neuroscience* **223**, 92–101 (2012).
19. Z. Katsarava *et al.*, *Headache* **46**, 1511–1517 (2006).
20. J. P. Lefaucheur *et al.*, *Neurophysiol. Clin.* **42**, 199–206 (2012).
21. D. Yarnitsky, M. Granot, H. Nahman-Averbuch, M. Khamaisi, Y. Granovsky, *Pain* **153**, 1193–1198 (2012).
22. S. Potvin, S. Marchand, *Pain* **157**, 1704–1710 (2016).
23. K. Ørstavik *et al.*, *J. Neurosci.* **26**, 11287–11294 (2006).
24. I. P. Kleggetveit *et al.*, *Pain* **153**, 2040–2047 (2012).
25. J. Serra *et al.*, *Ann. Neurol.* **75**, 196–208 (2014).
26. K. Wiech, *Science* **354**, 584–587 (2016).
27. N. Attal *et al.*, *Lancet Neurol.* **15**, 555–565 (2016).
28. A. C. Chan, E. P. Wilder-Smith, *Muscle Nerve* **53**, 671–682 (2016).
29. G. Devigili *et al.*, *Brain* **131**, 1912–1925 (2008).
30. N. Üçeyler *et al.*, *Neurology* **74**, 1806–1813 (2010).
31. A. L. Oaklander, H. L. Fields, *Ann. Neurol.* **65**, 629–638 (2009).
32. E. Gröne *et al.*, *Pain* **155**, 1784–1792 (2014).
33. D. J. Clauw, *JAMA* **311**, 1547–1555 (2014).
34. A. L. Oaklander, Z. D. Herzog, H. M. Downs, M. M. Klein, *Pain* **154**, 2310–2316 (2013).
35. M. de Tommaso *et al.*, *J. Neurol.* **261**, 461–472 (2014).
36. K. Doppler, H. L. Rittner, M. Deckart, C. Sommer, *Pain* **156**, 2319–2325 (2015).
37. N. Üçeyler *et al.*, *BMC Neurol.* **15**, 210 (2015).
38. M. Tavakoli *et al.*, *Exp. Neurol.* **223**, 245–250 (2010).
39. M. S. Jiang, Y. Yuan, Z. X. Gu, S. L. Zhuang, *Br. J. Ophthalmol.* **100**, 9–14 (2016).
40. L. Oudejans *et al.*, *Sci. Rep.* **6**, 23573 (2016).
41. W. Kessler, C. Kirchhoff, P. W. Reeh, H. O. Handwerker, *Exp. Brain Res.* **91**, 467–476 (1992).
42. N. Üçeyler *et al.*, *Arthritis Rheum.* **54**, 2656–2664 (2006).
43. N. Üçeyler, S. Kewenig, W. Kafke, S. Kittel-Schneider, C. Sommer, *BMC Neurol.* **14**, 185 (2014).
44. N. Üçeyler, M. Valet, W. Kafke, T. R. Tölle, C. Sommer, *PLOS ONE* **9**, e105269 (2014).
45. N. Üçeyler, T. Eberle, R. Rolke, F. Birklein, C. Sommer, *Pain* **132**, 195–205 (2007).
46. N. Üçeyler, J. P. Rogausch, K. V. Toyka, C. Sommer, *Neurology* **69**, 42–49 (2007).
47. M. Lenz *et al.*, *Pain* **154**, 2142–2149 (2013).
48. H. Wang, M. Buchner, M. T. Moser, V. Daniel, M. Schiltenwolf, *Clin. J. Pain* **25**, 1–4 (2009).
49. E. Bäckryd, B. Ghafouri, A. K. Carlsson, P. Olausson, B. Gerdle, *J. Pain Res.* **8**, 321–333 (2015).
50. G. Oki *et al.*, *Pain* **153**, 532–539 (2012).
51. R. D. Moloney *et al.*, *CNS Neurosci. Ther.* **22**, 102–117 (2016).
52. S. D. Dib-Hajj, S. G. Waxman, *Sci. Transl. Med.* **6**, 249sr4 (2014).
53. S. G. Waxman *et al.*, *Lancet Neurol.* **13**, 1152–1160 (2014).
54. F. Reimann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5148–5153 (2010).
55. M. Estacion *et al.*, *Ann. Neurol.* **66**, 862–866 (2009).
56. J. U. Harrer *et al.*, *Pain* **155**, 2199–2203 (2014).
57. D. L. Bennett, C. G. Woods, *Lancet Neurol.* **13**, 587–599 (2014).
58. M. F. Seidel, B. L. Wise, N. E. Lane, *Osteoarthr. Cartil.* **21**, 1223–1228 (2013).
59. M. Estacion *et al.*, *Mol. Pain* **7**, 92 (2011).
60. R. Massart *et al.*, *Sci. Rep.* **6**, 19615 (2016).
61. J. T. Bell *et al.*, *Nat. Commun.* **5**, 2978 (2014).
62. C. T. Viet *et al.*, *Pain* **152**, 2323–2332 (2011).
63. M. Kress *et al.*, *Front. Mol. Neurosci.* **6**, 33 (2013).
64. K. K. Bali, M. Hackenberg, A. Lubin, R. Kuner, M. Devor, *Mol. Pain* **10**, 22 (2014).
65. J. L. Bjersing, C. Lundborg, M. I. Bokarewa, K. Mannerkorpi, *PLOS ONE* **8**, e78762 (2013).
66. G. Cerdá-Olmedo, A. V. Mena-Durán, V. Monsalve, E. Oltra, *PLOS ONE* **10**, e0121903 (2015).
67. S. M. Chambers *et al.*, *Nat. Biotechnol.* **30**, 715–720 (2012).
68. L. Cao *et al.*, *Sci. Transl. Med.* **8**, 335ra56 (2016).
69. S. Kawagoe *et al.*, *Mol. Genet. Metab.* **109**, 386–389 (2013).
70. N. Lebonvallet *et al.*, *Exp. Dermatol.* **23**, 73–75 (2014).
71. M. Andersson *et al.*, *Sci. Rep.* **6**, 24818 (2016).


### ACKNOWLEDGMENTS

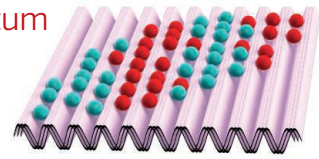
C.S. is a consultant for the following companies, which provide medication for neuropathic pain or neuropathies: Baxalta, Genzyme, Grifols, Air Liquide, CSL-Behring, and Astellas. The author has given educational talks on neuropathic pain or neuropathies at events sponsored by the following companies: Baxalta, Genzyme, Grifols, Kedrion, Novartis, Astellas, and Pfizer.

10.1126/science.aaf8935

# RESEARCH

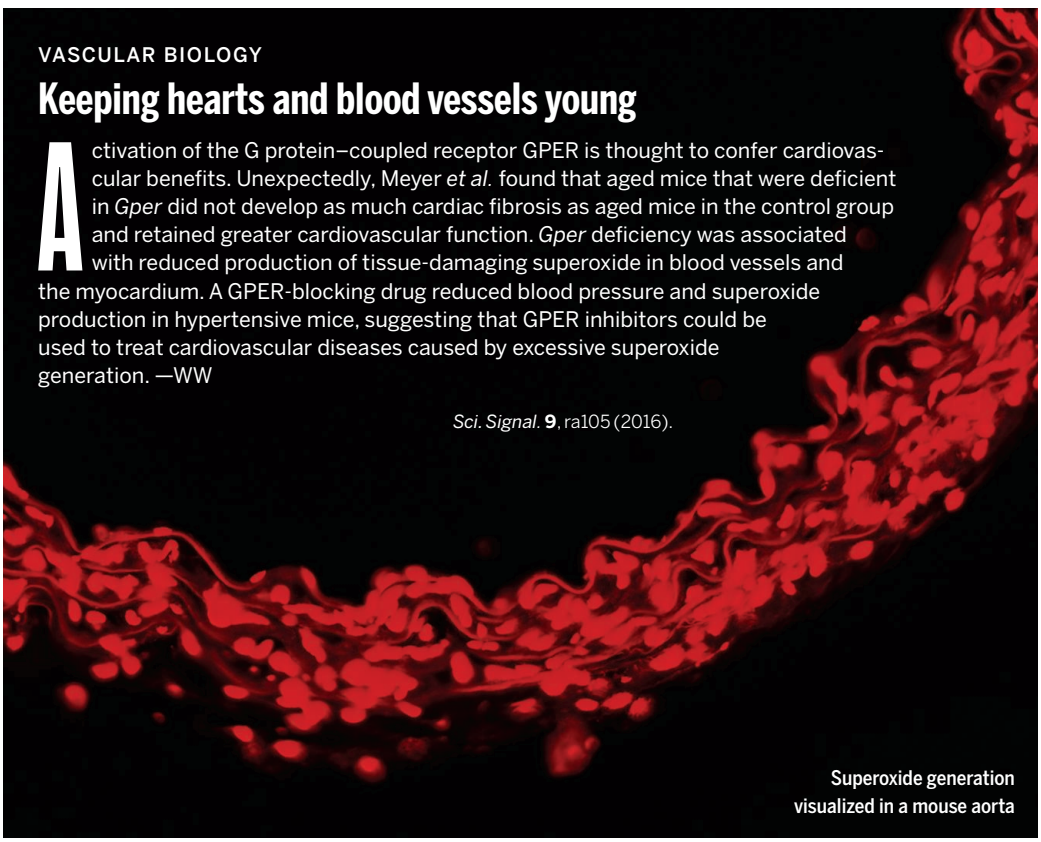
Capturing a quantum phase transition

*Clark* et al., *p. 606*



## IN *SCIENCE* JOURNALS

*Edited by* **Stella Hurtley**



VASCULAR BIOLOGY

### Keeping hearts and blood vessels young

Activation of the G protein–coupled receptor GPER is thought to confer cardiovascular benefits. Unexpectedly, Meyer *et al.* found that aged mice that were deficient in *Gper* did not develop as much cardiac fibrosis as aged mice in the control group and retained greater cardiovascular function. *Gper* deficiency was associated with reduced production of tissue-damaging superoxide in blood vessels and the myocardium. A GPER-blocking drug reduced blood pressure and superoxide production in hypertensive mice, suggesting that GPER inhibitors could be used to treat cardiovascular diseases caused by excessive superoxide generation. —WW

*Sci. Signal.* **9**, ra105 (2016).

Superoxide generation visualized in a mouse aorta

CHEMICAL BIOLOGY

### Radicals push proteins beyond genes

Chemically modifying proteins after their translation can expand their structural and functional roles (see the Perspective by Hofmann and Bode). Two related methods describe how to exploit free radical chemistry to form carbon-carbon bonds between amino acid residues and a selected functional group. Wright *et al*. added a wide range of functional groups to proteins containing dehydro-alanine precursors, with borohydride mediating the radical chemistry. Yang *et al*. employed a similar approach, using zinc in combination with copper ions. Together, these results will be useful for introducing functionalities and labels to a wide range of proteins. —NW

*Science*, this issue pp. 597 and 623; see also p. 553

PHYSICS

### Exotic molecule tests fundamental symmetry

Spectroscopy of exotic molecules can offer insight into fundamental physics. Hori *et al*. studied the transition frequencies of an unusual helium atom in which one of the two electrons was substituted by an antiproton, the negatively charged antiparticle partner of the proton (see the Perspective by Ubachs). The antiprotonic helium was cooled down to low temperatures to allow the frequencies to be measured with high precision. The extracted mass of the antiproton (relative to the electron mass) was in good agreement with previous measurements of the proton mass. This finding is in keeping with the implications of the combined charge, parity, and time-reversal symmetry of physical laws. —JS

*Science*, this issue p. 610; see also p. 546

CANCER ETIOLOGY

### Assessing smoke damage in cancer genomes

We have known for over 60 years that smoking tobacco is one of the most avoidable risk factors for cancer. Yet the detailed mechanisms by which tobacco smoke damages the genome and creates the mutations that ultimately cause cancer are still not fully understood. Alexandrov *et al*. examined mutational signatures and DNA methylation changes in over 5000 genome sequences from 17 different cancer types linked to smoking (see the Perspective by Pfeifer). They found a complex pattern of mutational signatures. Only cancers originating in tissues directly exposed to smoke showed a signature characteristic of the known tobacco carcinogen benzo[*a*]pyrene. One mysterious signature was shared by all smoking-associated cancers but is of unknown origin. Smoking had only a modest effect on DNA methylation. —PAK

*Science*, this issue p. 618; see also p. 549


**Cigarette smoking produces a variety of mutational signatures.**

CREDITS (FROM TOP): CLARK *ET AL.*; MEYER *ET AL.*; LUCKY STRIKE CIGARETTES PACK (B/W PHOTO)/UNDERWOOD ARCHIVES/UIG/BRIDGEMAN IMAGE

VACCINES

## Rejuvenating viral vectors

Adenoviral (Ad5) vaccine vectors elicit mixed responses: They induce protective $CD8^+$ T cells, but these cells may be partially exhausted. Now Larocca *et al.* demonstrate that this exhausted phenotype may result from Ad5 vector–induced antigen-specific $CD4^+$ T cells that express interleukin-10 (IL-10) and PD-1 in both mice and macaques. These IL-$10^+$ $CD4^+$ T cells suppress the vaccine-induced $CD8^+$ T cell response, and their inhibitory function may depend in part on IL-27. Targeting this inhibitory pathway may thus enhance protection of viral vector–based vaccines. —AC

*Sci. Immunol.* **1**, eaaf7643 (2016).

OPTICS

## Shrinking spectrometers

Dual-comb spectroscopy is a powerful technique that uses the interference of two closely related combs to map spectroscopic features directly into a frequency domain that can be read by electronics. Suh *et al*. developed a dual-comb spectroscopy approach using combs produced by silica microresonators fabricated on a silicon chip. Perhaps high-resolution spectroscopy will soon be shrunk to the chip scale, doing away with the need for bulky spectrometers. —ISO

*Science*, this issue p. 600

OCEAN CHEMISTRY

## Uranium in the deep sea

The ratio of $^{234}U$ to $^{238}U$ in seawater underlies modern marine uranium-thorium geochronology, but it is difficult to establish the ratio precisely. Chen *et al*. report two $^{234}U/^{238}U$ records derived from deep-sea corals (see the Perspective by Yokoyama and Esat). The records reveal a number of important similarities to and differences from existing records of the past 30,000 years. Higher values during the most recent 10,000 years than during earlier glaciated conditions may reflect enhanced subglacial melting during deglaciation. —HJS

*Science*, this issue p. 626; see also p. 550

NEUROREGENERATION

## Spinal cord regeneration in zebrafish

Unlike humans, zebrafish can regenerate their spinal cord. Mokalled *et al*. identified a growth factor in zebrafish that helps this process (see the Perspective by Williams and He). The protein encoded by *ctgfa* (*connective tissue growth factor a*) is secreted after injury and encourages glial cells to form a bridge across the spinal lesion. Addition of this protein improved spinal cord repair in injured zebrafish. —PJH

*Science*, this issue p. 630; see also p. 544

PLANT SCIENCE

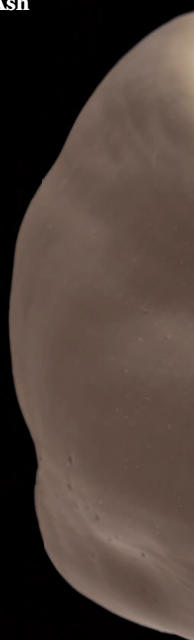
## Soil microbes yield insecticidal peptide

The microbial peptide BT, derived from the bacterium *Bacillus thuringiensis*, is widely used to protect crops from insect pests. Schellenberger *et al*. identified another insecticidal peptide from a different soil-dwelling bacterium, *Pseudomonas chlororaphis* (see the Perspective by Tabashnik). Corn plants expressing the *Pseudomonas* peptide were protected from attack by western corn rootworm. Rootworms that were resistant to BT were susceptible to the *Pseudomonas* peptide. Addition of another insecticidal peptide diversifies the arsenal against insect pests, which may slow down the development of resistance. —PJH

*Science*, this issue p. 634; see also p. 552

# IN OTHER JOURNALS

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



PUBLIC HEALTH

## Worms remodel immune responsiveness

Rural populations in less developed countries commonly show poor immunogenicity in vaccination programs. Helminth infestations remain common in some rural areas, and cellular immune hyporesponsiveness is a hallmark of chronic helminth infections. Community deworming programs are in general believed to be a good thing to reverse the morbidity that a large worm burden can impose on children. Wammes *et al.* set up a 2-year clinical trial to systematically test the immunological consequences of deworming in >1000 villagers in Indonesia. After treatment, subjects showed significant immune remodeling, with reduced expression of CTLA-4 (cytotoxin T lymphocyte–associated antigen 4) and elevated proinflammatory cytokine responses to malaria parasite antigens. The challenge in the longer term could be that restored immune responsiveness might increase the prevalence of inflammatory diseases. —CA

*Proc. Nat. Acad. Sci.* 10.1073/pnas.1604570113 (2016).

POROUS MATERIALS

## Printing emulsions and foams in 3D

The properties of porous materials can depend on the overall volume, size, and size distribution of the pores, but these can be difficult features to simultaneously control during fabrication. Emulsion or foam techniques can produce pores of only one narrow size range, but in combination with a direct ink-writing process, Minas *et al*. developed an approach to create porous materials with highly tunable properties. Shear thinning inks made from modified alumina particles are progressively deposited to form an overall structure, where large pores are made by the writing process and small ones from the emulsion or foam, with pores either open or closed. A percolating network of load-bearing struts results in structures with very high strength-to-weight ratios. —MSL

**A printed cubic lattice scaffold for the creation of porous ceramics**

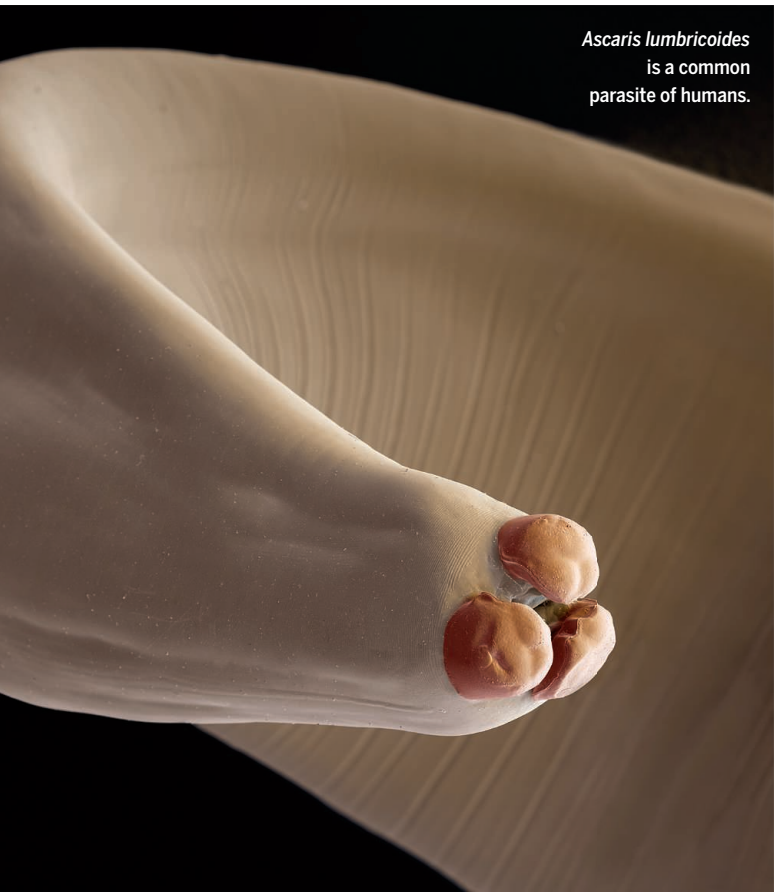
*Adv. Mater.* 10.1002/adma.201603390 (2016).

NEUROSCIENCE

## Cognition, behavior, and the globus pallidus

Deep inside the brain, the external segment of the globus pallidus receives many inputs from the neighboring basal ganglia. Its role in basic cognitive functions has rarely been measured directly. While monkeys were learning a complex task, Schechtman *et al.* recorded discharges from prototypical neurons in the structure. The

PHOTOS (FROM TOP): EYE OF SCIENCE/SCIENCE SOURCE; MINAS *ET AL.*, MATERIAL VIEWS (10 AUGUST 2016) ©2016



*Ascaris lumbricoides* is a common parasite of humans.

PHOTO: SUSUMU NISHINAGA/SCIENCE SOURCE

animals had to switch to a new stimulus-reward association once a predefined learning criterion had been reached. The activity of the neurons predicted whether the monkeys would change or maintain their previous stimulus choice. Neuronal spike activity also encoded whether such choices were successful. When the animals' performance became more automated and less demanding, discharge rates in the context of choice selection decreased. —PRS

*Proc. Nat. Acad. Sci.* 10.1073/pnas.1612392113 (2016).

CELL BIOLOGY

## Microtubule repair and rescue

The microtubule network that forms the cytoskeleton is essential for cell polarization and migration. Microtubules tend to grow slowly, and they can shrink rapidly in a process known as catastrophe. Aumeier *et al.* report that where the network is dense, microtubules bend and cross over, leading to increased mechanical stress and filament damage. Damaged filaments can self-repair in a process that also prevents complete depolymerization. Most of the tubulin incorporates into the growing ends of microtubules, but some incorporates along the shaft of preexisting microtubules at damaged regions. This repair-and-rescue mechanism provides insight into how the network can remain stable in this highly dynamic system. —SH and SMH

*Nat. Cell Biol.* **18**, 1054 (2016).

CHEMISTRY

## Sourcing hydrogen directly from wax

Hydrogen is an appealing fuel because of the energy released in its reaction with oxygen and the nearly ideal safety profile of the sole product, water. The challenge lies in transporting and storing the hydrogen before its use. Gonzalez-Cortes *et al.* now show that simple paraffin wax can be an efficient source of hydrogen under the right circumstances. Specifically, they embedded carbon-supported ruthenium nanoparticles in the wax and then irradiated it with microwaves, thereby releasing up to half of the available gravimetric hydrogen content. The authors attribute the catalytic efficiency to a possible combination of local superheating and field-induced plasma formation. —JSY

*Sci. Rep.* 10.1038/srep35315 (2016).

EDUCATION

## What a (scientific) argument is not

Engaging in arguments based on evidence is a practice found in STEM (science, technology, engineering, and mathematics) education frameworks. Are students being taught to argue evidence in the same way that professional scientists do? MacPherson interviewed 10 ecologists about arguments both current and ongoing in their field. These data were compared to ecological arguments portrayed in assessment tasks written for middle and high school science students. Ecologists discussed causal claims, whereas school science tasks instead focused on descriptive and prescriptive claims. This mismatch results in students not receiving an accurate picture of how professional ecologists argue. How can the STEM education community move forward with designing tasks that ask students to critique evidence in a truly scientific way? —MM

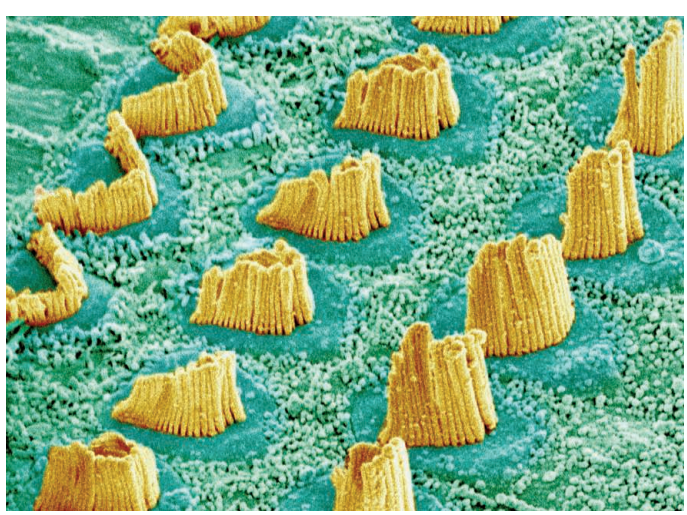
*Sci. Ed.* 10.1002/sce.21246 (2016).

NEURODEVELOPMENT

## Layered haircut underlies hearing

Hair cells deep within the ear transduce sound into hearing. On any single hair cell, a pack of stereocilia is neatly arranged from tallest to shortest. When the stereocilia do not develop adequately, deafness ensues. Studying mice, Tarchini *et al.* discovered some of the key signaling components that organize stereocilia during development. Two regulators of G proteins, the leucine-glycine-asparagine repeat protein and an inhibitory α-subunit of heterotrimeric G protein, coordinate to define the tallest row of stereocilia. Both regulators are expressed in the bare zone of the hair cell, a surface domain that will not produce stereocilia, and also in the very tips of the row of stereocilia adjacent to the bare zone. These stereocilia will emerge as the tallest of the crowd. If this signaling pathway is disrupted, stereocilia develop to more even and modest heights, and the animal is deaf. —PJH

*Development* 10.1242/dev.139089 (2016).



Inner-ear sensory hair cells are specifically organized.

## ALSO IN *SCIENCE* JOURNALS

*Edited by* **Stella Hurtley**

SCIENCE COMMUNITY

### Scientific impact—that is the *Q*

Are there quantifiable patterns behind a successful scientific career? Sinatra *et al*. analyzed the publications of 2887 physicists, as well as data on scientists publishing in a variety of fields. When productivity (which is usually greatest early in the scientist's professional life) is accounted for, the paper with the greatest impact occurs randomly in a scientist's career. However, the process of generating a high-impact paper is not an entirely random one. The authors developed a quantitative model of impact, based on an element of randomness, productivity, and a factor *Q* that is particular to each scientist and remains constant during the scientist's career. —BJ

*Science*, this issue p. 596

SYSTEMS BIOLOGY

### Complex transcription factor interactions

To respond to environmental changes, such as drought, plants must regulate numerous cellular processes. Working in the model plant *Arabidopsis*, Song *et al*. profiled the binding of 21 transcription factors to chromatin and mapped the complex gene regulatory networks involved in the response to the plant hormone abscisic acid. The work provides a framework for understanding and modulating plant responses to stress. —LMZ

*Science*, this issue p. 598

YEAST GENETICS

### A global genetic suppression network

The genetic background of an organism can influence the overall effects of new genetic variants. Some mutations can amplify a deleterious phenotype, whereas others can suppress it. Starting with a literature survey and expanding into a genome-wide assay, van Leeuwen *et al*. generated a large-scale suppression network in yeast. The data set reveals a set of general properties that can be used to predict suppression interactions. Furthermore, the study provides a template for extending suppression studies to other genes or to more complex organisms. —LMZ

*Science*, this issue p. 599

OPTICAL PROCESSING

### Taking the pulse of optimization

Finding the optimum solution of multiparameter or multifunctional problems is important across many disciplines, but it can be computationally intensive. Many such problems defined as computationally difficult can be mathematically mapped onto the so-called Ising problem, which looks at finding the minimum energy configuration for an array of coupled spins. Inagaki *et al*. and McMahon *et al*. show that an optical processing approach based on a network of coupled optical pulses in a ring fiber can be used to model and optimize large-scale Ising systems. Such a scalable architecture could help to optimize solutions to a wide range of complex problems. —ISO

*Science*, this issue pp. 603 and 614

AUTOIMMUNITY

### TYK2's balancing act

Determining the biological consequences of the thousands of genetic variants that contribute to common diseases is challenging. Genetic variants that influence autoimmune diseases have been identified in the gene encoding TYK2 (tyrosine kinase 2), but conflicting evidence regarding their biological impact obscures the therapeutic potential of TYK2. In resolving this conflict, Dendrou *et al*. have revealed a genetic effect that drives an optimal degree of immune signaling: low enough to be protective against autoimmunity but high enough to prevent immunodeficiency. These findings indicate that TYK2 may be a potential drug target in a number of autoimmune conditions. —OMS

*Sci. Transl. Med.* **8**, 363ra149 (2016).

CONSERVATION

### Can Europe's migratory birds be saved?

Many migratory birds breed in Europe and overwinter in sub-Saharan Africa. In a Perspective, Bairlein outlines several factors that are causing their populations to decline. Illegal takings and killings reduce populations, particularly during migration. Climate change causes ecological mismatches between birds and their insect prey. But, perhaps most importantly, human land-use changes are causing habitat degradation and loss in the birds' breeding and winter habitats, as well as along their migration routes. Political agreements and action plans are already in place but must be implemented to halt the population declines and avoid species extinctions. —JFU

*Science*, this issue p. 547

QUANTUM GASES

### Shaking the lattice uncovers universality

Most of our knowledge of quantum phase transitions (QPTs)—which occur as a result of quantum, rather than thermal, fluctuations—comes from experiments performed in equilibrium conditions. Less is known about the dynamics of a system going through a QPT, which have been hypothesized to depend on a single time and length scale. Clark *et al*. confirmed this hypothesis in a gas of cesium atoms in an optical lattice, which was shaken progressively faster to drive the gas through a QPT. —JS

*Science*, this issue p. 606

RESEARCH ARTICLE SUMMARY

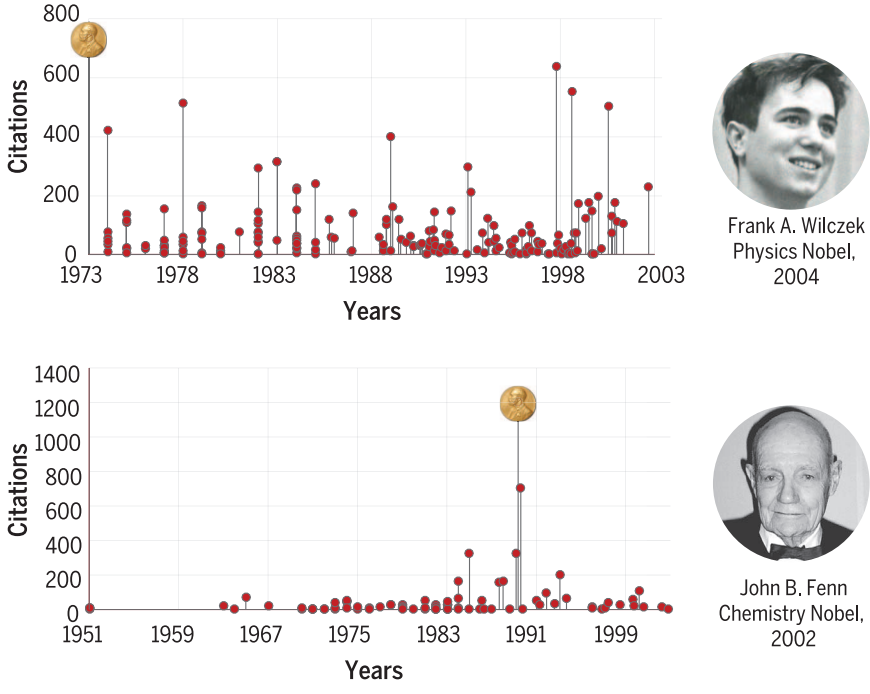
SCIENCE COMMUNITY

# Quantifying the evolution of individual scientific impact

**Roberta Sinatra, Dashun Wang, Pierre Deville, Chaoming Song, Albert-László Barabási***

**INTRODUCTION:** In most areas of human performance, from sport to engineering, the path to a major accomplishment requires a steep learning curve and long practice. Science is not that different: Outstanding discoveries are often preceded by publications of less memorable impact. However, despite the increasing desire to identify early promising scientists, the temporal career patterns that characterize the emergence of scientific excellence remain unknown.

**RATIONALE:** How do impact and productivity change over a scientific career? Does impact, arguably the most relevant performance measure, follow predictable patterns? Can we predict the timing of a scientist's outstanding achievement? Can we model, in quantitative and predictive terms, scientific careers? Driven by these questions, here we quantify the evolution of impact and productivity throughout thousands of scientific careers. We do so by reconstructing the publication record of scientists from seven disciplines, associating to each paper its long-term impact on the scientific community, as quantified by citation metrics.






Frank A. Wilczek
Physics Nobel,
2004



John B. Fenn
Chemistry Nobel,
2002

**Random-impact rule.** The publication history of two Nobel laureates, Frank A. Wilczek (Nobel Prize in Physics, 2004) and John B. Fenn (Nobel Prize in Chemistry, 2002), illustrating that the highest-impact work can be, with the same probability, anywhere in the sequence of papers published by a scientist. Each vertical line corresponds to a research paper. The height of each line corresponds to paper impact, quantified with the number of citations the paper received after 10 years. Wilczek won the Nobel Prize for the very first paper he published, whereas Fenn published his Nobel-awarded work late in his career, after he was forcefully retired by Yale. [Image of Frank A. Wilczek is reprinted with permission of STS/Society for Science & the Public. Image of John B. Fenn is available for public domain use on Wikipedia.org.]

**RESULTS:** We find that the highest-impact work in a scientist's career is randomly distributed within her body of work. That is, the highest-impact work can be, with the same probability, anywhere in the sequence of papers published by a scientist—it could be the first publication, could appear mid-career, or could be a scientist's last publication. This random-impact rule holds for scientists in different disciplines, with different career lengths, working in different decades, and publishing solo or with teams and whether credit is assigned uniformly or unevenly among collaborators.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aaf5239

The random-impact rule allows us to develop a quantitative model, which systematically untangles the role of productivity and luck in each scientific career. The model assumes that each scientist selects a project with a random potential $p$ and improves on it with a factor $Q_i$, resulting in a publication of impact $Q_i p$. The parameter $Q_i$ captures the ability of scientist $i$ to take advantage of the available knowledge in a way that enhances ($Q_i > 1$) or diminishes ($Q_i < 1$) the potential impact $p$ of a paper. The model predicts that truly high-impact discoveries require a combination of high $Q$ and luck ($p$) and that increased productivity alone cannot substantially enhance the chance of a very high impact work. We also show that a scientist's $Q$, capturing her sustained ability to publish high-impact papers, is independent of her career stage. This is in contrast with all current metrics of excellence, from the total number of citations to the $h$-index, which increase with time. The $Q$ model provides an analytical expression of these traditional impact metrics and allows us to predict their future time evolution for each individual scientist, being also predictive of independent recognitions, like Nobel prizes.

**CONCLUSION:** The random-impact rule and the $Q$ parameter, representing two fundamental characteristics of a scientific career, offer a rigorous quantitative framework to explore the evolution of individual careers and understand the emergence of scientific excellence. Such understanding could help us better gauge scientific performance and offers a path toward nurturing high-impact scientists, potentially informing future policy decisions. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: alb@neu.edu
Cite this article as R. Sinatra *et al.*, *Science* **354**, aaf5239 (2016).
DOI: 10.1126/science.aaf5239

RESEARCH ARTICLE

SCIENCE COMMUNITY

# Quantifying the evolution of individual scientific impact

**Roberta Sinatra,[1,2] Dashun Wang,[3,4] Pierre Deville,[1,5] Chaoming Song,[6] Albert-László Barabási[1,7,8,9]***

**Despite the frequent use of numerous quantitative indicators to gauge the professional impact of a scientist, little is known about how scientific impact emerges and evolves in time. Here, we quantify the changes in impact and productivity throughout a career in science, finding that impact, as measured by influential publications, is distributed randomly within a scientist's sequence of publications. This random-impact rule allows us to formulate a stochastic model that uncouples the effects of productivity, individual ability, and luck and unveils the existence of universal patterns governing the emergence of scientific success. The model assigns a unique individual parameter $Q$ to each scientist, which is stable during a career, and it accurately predicts the evolution of a scientist's impact, from the $h$-index to cumulative citations, and independent recognitions, such as prizes.**

Productivity, representing the number of publications authored by a scientist over time, and impact, often approximated by the number of citations a publication receives (*1*–*4*), are frequently used metrics to gauge a scientist's performance. Despite their widespread use, we lack a quantitative understanding of the patterns these metrics follow during a scientist's career (*5*). This is particularly alarming (*6*–*11*), given that they are increasingly adopted for academic assessment (*4*, *11*) and serve as the input for numerous indicators, like the $h$-index and its variants, which are frequently used to compare individual performance (*12*–*14*). Given the increasing interest in predicting the value of these indicators (*5*, *15*), here we ask: How do impact and productivity change over a typical scientific career? Does impact, arguably the most relevant performance measure, follow predictable patterns? Can we predict the timing of a scientist's outstanding achievement? Can we untangle the role of impact, productivity, and luck within a scientific career?

To address these questions, we reconstruct the publication profile of scientists from multiple disciplines and associate each of their publications with an impact, as captured by $c_{10}$, the number of citations 10 years after publication (Fig. 1A; see Methods and section S1).

Motivated partly by the exceptional awareness of a scientist's highest-impact work (*16*, *17*), like radioactivity for Marie Curie and the double helix for Watson and Crick, we identified for each researcher her most-cited paper, $c_{10}^{*}$, that is, the paper with the highest number of citations 10 years after its publication. The distribution $P(c_{10}^{*})$ for the studied scientists indicates that only 5% have $c_{10}^{*} \geq 200$; hence, most scientific careers have limited maximal impact. To systematically distinguish the careers on the basis of their peak impact, we group each scientist into high maximum impact (top 5%, $c_{10}^{*} \geq 200$), low maximum impact (bottom 20%, $c_{10}^{*} \leq 20$), and medium maximum impact (middle 75%, $20 < c_{10}^{*} < 200$) categories (Fig. 1B and section S2).

## Productivity and impact patterns in scientific careers

The total number of papers scientist $i$ publishes up to time $t$ after her first publication, $N_i(t)$, asymptotically follows $N_i(t) \sim t_i^{\gamma}$ (Fig. 1C) (*18*). Hence, yearly productivity, $n_i(t)$, follows the same scaling with exponent $(\gamma_i - 1)$ (fig. S5). Yet, the scaling exponent is different for low-, medium-, and high-impact scientists (Fig. 1C). We find that for low-impact scientists, $\langle\gamma\rangle = 1.55$, indicating on average a steady increase in their productivity. The increase is much faster for high-impact researchers, for whom $\langle\gamma\rangle = 2.05$ (Fig. 1D). These trends are also confirmed by changes in the yearly productivity $\langle n(t)\rangle$: For high-impact scientists, productivity increases almost threefold during their career, whereas the increase is modest for low-impact scientists (Fig. 1E). Together, Fig. 1 (D and E) indicates that productivity changes throughout a scientific career. We find, however, that this trend is modulated by impact: Productivity growth is more pronounced for high-impact scientists and is modest for low-impact scientists (Fig. 1, C to E).

As Fig. 2A indicates, impact appears to follow similar patterns to productivity (Fig. 1E): Although $c_{10}$ increases during a high-impact scientist's career, an increase is hardly noticeable for average and low-impact individuals. Yet, we observe a markedly different pattern if we examine the impact in the vicinity of $t^*$, the publication time of the most-cited work $c_{10}^{*}$. Plotting $\langle c_{10}\rangle$ for the sequence of papers before and after an individual's most-cited paper, $c_{10}^{*}$ (Fig. 2B), we do not see a gradual increase in impact as a scientist approaches $t^*$, nor do we observe elevated citations after this breakthrough. Instead, the observed pattern exhibits a singular behavior. This singularity could be a simple result of averaging random-impact fluctuations present in careers. We find, however, that the result is robust if we use a moving average or consider only the publication with maximum impact in a rolling window (section S2.1 and fig. S6) and is validated using the fitting hypothesis $c_{10}^{i}(t) \sim \alpha_i t$, lacking differences in $\alpha_i$ before or after $t^*$ (section S2.2 and fig. S7) (*19*). Also, the papers published before and after $t^*$ show no discernible differences in their average number of citations (Fig. 2C). Finally, we randomize each career by leaving all productivity measures [total number of papers, $N_i$, and $n_i(t)$] unchanged but shuffling the impact of each paper within each career (Fig. 2C). The lack of differences between the original and the randomized careers supports our overall conclusion: There are no detectable changes in impact leading up to or following a scientist's highest-impact work. We tested the robustness of this measure for different samples of scientists (figs. S8 and S9), for different definitions of impact (section S1.6 and fig. S10), and in data sets where we attribute different impact shares to each author of a paper (section S6 and fig. S11), arriving at the same conclusion. Yet, we cannot exclude that there are other choices of impact variables or data-set selection that can detect patterns before or after the highest-impact paper.

To understand when a scientist publishes her most important work, we measured the probability $P(t^*)$ that the highest-impact paper is published at time $t^*$ after a scientist's first publication (Fig. 2D). The high $P(t^*)$ between 0 and 20 years indicates that most scientists publish their highest-impact paper early or midcareer. The drop in $P(t^*)$ after 20 years suggests that it is unlikely that a scientist's most-cited work will come late in her career, a result well documented by the literature about creativity (see section S3.1) (*20*, *21*). To understand the origin of this pattern, we shuffled $c_{10}$ among all papers published by the same scientist, preserving the scientist's time-dependent productivity and paper-by-paper impact and randomizing only the order of her publications. The fact that $P(t^*)$ for these synthetic careers is indistinguishable from the original data (Fig. 2D) indicates that variations in $P(t^*)$ are not due to specific impact sequences or other features

[1]Center for Complex Network Research and Physics Department, Northeastern University, Boston, MA 02115, USA. [2]Center for Network Science and Math Department, Central European University, Budapest, Hungary. [3]Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA. [4]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208, USA. [5]Department of Applied Mathematics, Université catholique de Louvain, Louvain-la-Neuve, Belgium. [6]Department of Physics, University of Miami, Coral Gables, FL 33124, USA. [7]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. [8]Center for Network Science, Central European University, Budapest, Hungary. [9]Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Boston, MA 02115, USA.
*Corresponding author. Email: alb@neu.edu

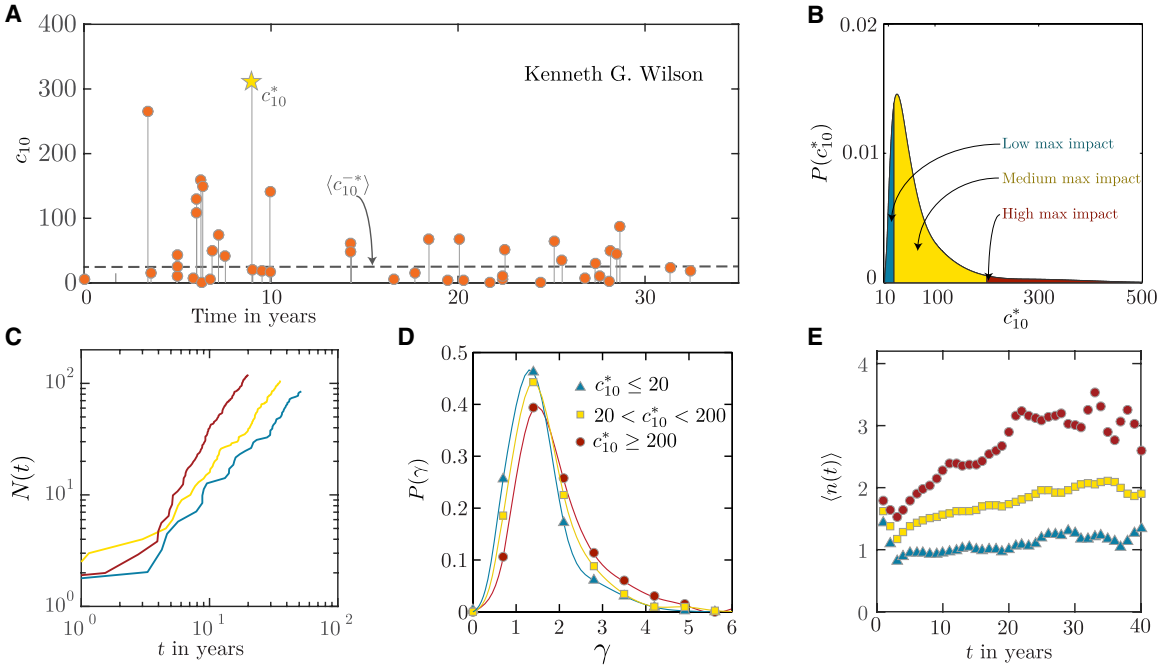



**Fig. 1. Patterns of productivity during a scientific career.** (**A**) Publication history of Kenneth G. Wilson (Nobel Prize in Physics, 1982). Horizontal axis indicates the number of years after the scientist's first publication, and each vertical line corresponds to a research paper. The height of each line corresponds to $c_{10}$, that is, the number of citations the paper received after 10 years (sections S1.3 and S1.6). The highest-impact paper of Wilson was published in 1974, 9 years after his first publication, and it is the 17th of his 48 papers; hence, $t^* = 9$, $N^* = 17$, and $N = 48$. (**B**) Distribution of the highest-impact paper $P(c^*_{10})$ across all scientists. We highlight in blue the bottom 20% of the area, corresponding to low maximum impact scientists ($c^*_{10} \le 20$); red area indicates the high maximum impact scientists (top 5%, $c^*_{10} \ge 200$); yellow corresponds to the remaining 75% medium maximum impact scientists ($20 < c^*_{10} < 200$). These cutoffs do not change if we exclude review papers from our analysis (see figs. S4 and S36). (**C**) Number of papers $N(t)$ published up to time $t$ for three scientists with low, medium, and high impact but with comparable final number of papers throughout their career. (**D**) Distribution of the productivity exponents $\gamma$ (*18*). The productivity of high-impact scientists grows faster than does that of low-impact scientists. (**E**) Dynamics of productivity, as captured by the average number of papers $\langle n(t)\rangle$ published each year for high-, average-, and low-impact scientists. $t = 0$ corresponds to the year of a scientist's first publication.

but are entirely explained by year-by-year variations in productivity throughout a career (fig. S12) (*20*, *21*).

These results prompted us to explore the position $N^*$ of the highest-impact paper in the sequence of $N$ publications of a scientist by measuring $P(N^*/N)$, that is, the probability that the most-cited work is early ($N^*/N$ small) or late ($N^*/N \simeq 1$) within the sequence of papers published by a scientist. We find that $P(N^*/N)$ is flat (Fig. 2E, inset), a finding supported by the cumulative $P(\ge N^*/N)$ (Fig. 2E), which decreases independently of impact as $(N^*/N)^{-1}$, in line with a uniform $P(N^*/N)$. Together, we arrive at a rather unexpected conclusion, representing our main empirical finding: Impact is randomly distributed within a scientist's body of work, regardless of publication time or order in the sequence of publications. We call this the random-impact rule because it indicates that the highest-impact work can be, with the same probability, anywhere in the sequence of $N$ papers published by a scientist. We find that the random-impact rule holds for scientists in different disciplines, with different career lengths, working in different decades, and publishing solo or with teams and whether credit is assigned uniformly or unevenly among collaborators (sections S1.4 and S6.1) (*22*).

The random-impact rule can explain the growing impact during a scientist's career (Fig. 2A). To see this, we again randomly shuffle the impact of the papers within each career, leaving the individual productivity unchanged. The variations of impact of the randomized careers are indistinguishable from the original data for both high- and low-impact individuals (Fig. 2A). Hence, the growing average impact documented in Fig. 2A is the result of combining the increasing average productivity (Fig. 1E) with the heavy-tailed nature of the citation distribution (*6*, *23*–*25*). $\langle c_{10}\rangle$ is not stable but increases with the number of publications, resulting in the observed growing impact (Fig. 2A). Hence, growing productivity, rather than increasing ability or excellence, can account for the growth in average impact during a career in science.

The defining role of productivity in the timing of the highest-impact work persists if we extend the analysis to different samples of scientists, not only those with at least 20 years of publication record. We considered different selections of scientists, such as (i) grouping them by different career lengths (figs. S13 and S14), (ii) grouping them by decade of active career (figs. S15 and S16), (iii) removing multi-authored papers (fig. S17), (iv) including only papers published in one subarea of physics (fig. S38), (v) creating no filter and including all scientists (figs. S18 and S19), (vi) using different definitions of impact (figs. S20, S21, and S37), or (vii) considering the six different disciplines in data set (ii) (figs. S22 and S23). In all these cases, the location of the peak of the highest-impact work probability changes, but we never observe a difference with the randomized careers. Hence, the specific shape of $P(t^*)$ is only a function of the selection of scientists and of their temporal productivity patterns, whereas impact is always randomly distributed within a scientist's sequence of publications.

The documented random-impact rule raises an important question: What is the role of a researcher's own ability, if any, in scientific excellence? We propose two quantitative models to answer this question.

## Random-impact model (R-model)

We can rely on the random impact rule to build a null model of scientific careers: We assume that each scientist publishes a sequence of papers whose impact is randomly chosen from the same impact distribution $P(c_{10})$. Consequently, the only difference between two scientists is their overall productivity $N$. With the observed $P(c_{10})$ and $P(N)$ distributions (Fig. 3, A and B) as input, the obtained $R$-model (section S4.2) accurately reproduces the randomness of the impact sequence $P(N^*/N)$ (Fig. 2E), but it also makes two predictions that are at odds with the data.

(a) Productivity alone begets success: If each paper's impact is randomly drawn from the same $P(c_{10})$, a productive scientist (high $N$) will more likely score a high $c^*_{10}$ (see eqs. S7 and S18) (*26*, *27*). However, the $R$-model does not correctly reproduce the observed increase of $\langle c^*_{10}\rangle$ as a function of $N$ (Fig. 3C).

(b) Divergent impact: The higher the average impact of a scientist's publications without the most-cited publication $\langle c^{-*}_{10}\rangle$ (Fig. 1A), the higher the impact of the most-cited paper, $c^*_{10}$ (Fig. 3D). Hence, papers with truly high impact

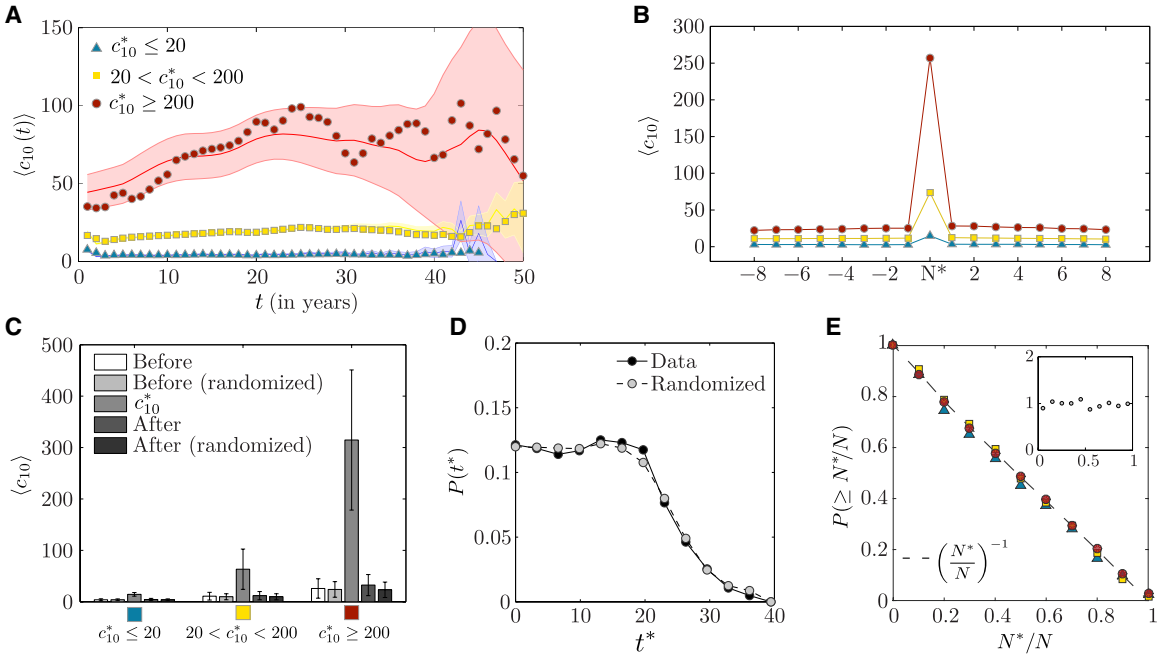



**Fig. 2. Patterns of impact during a scientific career.** (**A**) Dynamics of impact captured by the yearly average impact of papers $\langle c_{10}(t) \rangle$ for high, medium, and low maximum impact scientists, where $t = 0$ corresponds to the year of a scientist's first publications. The symbols correspond to the data, whereas the shaded area indicates the 95% confidence limit of careers where the impact of the publications is randomly permuted within each career. (**B**) Average impact $\langle c_{10} \rangle$ of papers published before and after the highest-impact paper $c_{10}^*$ of high-, middle-, and low-impact scientists. The plot indicates that there are no discernible changes in impact before or after a scientist's highest-impact work. (**C**) $\langle c_{10}^* \rangle$ and $\langle c_{10} \rangle$ before and after a scientist's most-cited paper. For each group, we calculate the average impact of the most-cited paper, $\langle c_{10}^* \rangle$, as well as the average impact of all papers before and after the most-cited paper. We also report the same measures obtained in publication sequences for which the impact $c_{10}^*$ is fixed, whereas the impact of all other papers is randomly permuted. (**D**) Distribution of the publication time $t^*$ of the highest-impact paper for scientists' careers (black circles) and for randomized-impact careers (gray circles). The lack of differences between the two curves ($P = 0.70$ for the Mann-Whitney $U$ test between the two distributions) supports the random-impact rule; that is, impact is random within a scientist's sequence of publication. Note that the drop after 20 years is partly because we focus on careers that span at least 20 years (see fig. S22). (**E**) Cumulative distribution $P(\ge N^*/N)$ for scientists with $N \simeq 50$, where $N^*/N$ denotes the order $N$ of the highest-impact paper in a scientist's career, varying between $1/N$ and 1. The cumulative distribution of $N^*/N$ is a straight line with slope 1, indicating that $N$ has the same probability to occur anywhere in the sequence of papers published by a scientist. The flatness of $P(N^*/N)$ (all scientists, inset) supports the conclusion that the timing of the highest-impact paper is uniform. The small differences between the three curves are due to different number of publications $N$ in the three groups of scientists [see fig. S24 for the plot of $P(\ge N^*/N)$ for other values of $N$ and figs. S25 and S26 for the impact autocorrelation throughout a scientific career].

are published by scientists with a consistent record of high impact. The $R$-model cannot account for this behavior, predicting that $\langle c_{10}^* \rangle$ diverges when $\langle \log (c_{10}^{-*}) \rangle \to 1.97$ (Fig. 3D), a consequence of the log-normal nature of $P(c_{10})$ (section S4.1 and fig. S27).

Failures (a) and (b) prompt us to abandon our hypothesis that research papers are all drawn from the same impact distribution and hence researchers have no distinguishable individual impact, forcing us to explore more closely the relationship between productivity, impact, and chance.

## $Q$-model

Crucially, in the $R$-model, scientists with similar productivity have indistinguishable impact. In reality, impact varies greatly between scientists (Fig. 3E), suggesting the existence of a hidden parameter $Q_i$ that modulates impact, which has a unique value for each scientist $i$.

The log-normal nature of $P(c_{10})$ (Fig. 3A) (*24*) indicates the presence of multiplicative processes, prompting us to write the impact $c_{10,i\alpha}$ of paper $\alpha$ published by scientist $i$ as

$$c_{10,i\alpha} = Q_i p_\alpha \qquad (1)$$

where $p_\alpha$ is the potential impact of paper $\alpha$ in the sequence of papers published by scientist $i$. The parameter $Q_i$ captures the ability of scientist $i$ to take advantage of the available knowledge in a way that enhances ($Q_i > 1$) or diminishes ($Q_i < 1$) the potential impact of paper $\alpha$. We take the value of this parameter $Q_i$ to be constant throughout a scientist's career, a hypothesis we validate later (Fig. 5 and section S4.9). The obtained model assumes that each scientist randomly selects a project with potential $p_\alpha$ and improves on it with a factor $Q_i$ that is unique to the scientist, resulting in a paper of impact (Eq. 1). Truly high-impact publications are therefore the result of a high $Q_i$ scientist selecting by chance a high $p_\alpha$ project; any scientist, independently of her parameter $Q_i$, can publish low-impact papers by selecting a low $p_\alpha$.

The stochastic process behind the model (Eq. 1) is determined by the joint probability $P(p, Q, N)$, with unknown correlations between $p$, $Q$, and $N$. The log-normal nature of $P(c_{10})$ (Fig. 3A) allows us to measure $P(p)$, finding that it can also be fitted with a log-normal function (Fig. 3F). Assuming that $Q$ is also a log-normal (confirmed later), we denote with $\hat{p} = \log p$, $\hat{Q} = \log Q, \hat{N} = \log N$, obtaining the trivariate normal distribution $P(\hat{p}, \hat{Q}, \hat{N}) = \mathcal{N}(\mu, \Sigma)$. Using a maximum-likelihood approach (see section S4.4), we estimate from the data the mean $\mu \equiv (\mu_p, \mu_Q, \mu_N) = (0.92, 0.93, 3.34)$ and the covariance matrix

$$\sum \equiv \begin{pmatrix} \sigma_p^2 & \sigma_{p,Q} & \sigma_{p,N} \\ \sigma_{p,Q} & \sigma_Q^2 & \sigma_{Q,N} \\ \sigma_{p,N} & \sigma_{Q,N} & \sigma_N^2 \end{pmatrix} = \begin{pmatrix} 0.93 & 0.00 & 0.00 \\ 0.00 & 0.21 & 0.09 \\ 0.00 & 0.09 & 0.33 \end{pmatrix} \qquad (2)$$

The matrix (Eq. 2) leads to two key predictions:

(i) $\sigma_{p,N} = \sigma_{p,Q} \simeq 0$ indicates that the paper potential impact $p_\alpha$ is independent of a scientist's productivity $N_i$ and her hidden parameter $Q_i$. Therefore, scientists select the potential impact of each paper randomly from a $P(p)$ distribution that is the same for all individuals, being independent of $Q$ and $N$, capturing a universal—that is, scientist-independent—luck component behind impact.

(ii) The nonzero $\sigma_{Q,N}$ indicates that the hidden parameter $Q$ and productivity $N$ do depend on each other (section S4.4), but its small value also shows that high $Q$ is only slightly associated with higher productivity.

The lack of correlations between $p$ and $(Q,N)$ allows us to analytically calculate the dependence of the highest-impact paper $\langle \log c_{10}^* \rangle$ on productivity

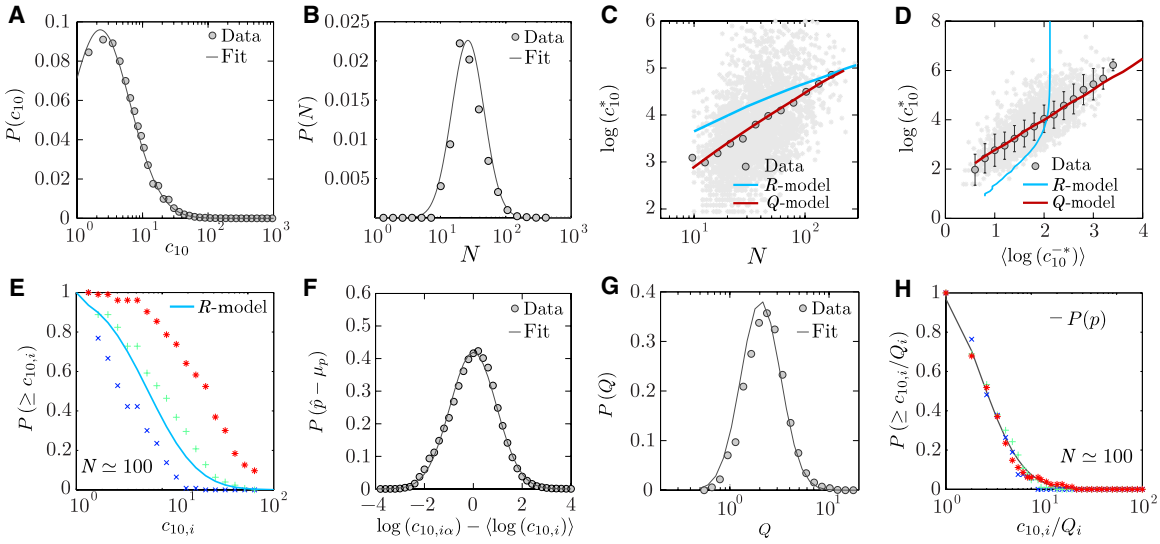



**Fig. 3. The *Q*-model.** (**A**) Distribution of the paper impact $c_{10}$ across all publications in the data set. The gray line corresponds to a log-normal function with average $\mu = 1.93$ and SD $\sigma^2 = 1.05$ ($R^2 = 0.98$). (**B**) Distribution of the total number of papers published by a scientist (productivity). The gray line is a log-normal with $\mu = 3.6$ and $\sigma^2 = 0.57$ [weighted Kolmogorov-Smirnov (KS) test, $P = 0.70$]. (**C**) Citations of the highest-impact paper, $c^*_{10}$, versus the number of publications $N$ during a scientist's career. Each gray point of the scatterplot corresponds to a scientist. The circles are the logarithmic binning of the scattered data. The cyan curve represents the prediction of the *R*-model, assuming that the impact of each paper is extracted randomly from the distribution $P(c_{10})$ of Fig. 2A. The red curve corresponds to the analytical prediction (see eq. S35) of the *Q*-model ($R^2 = 0.97$; see section S4.6 and fig. S29 for goodness of the fit). (**D**) $\log c^*_{10}$ versus $\langle \log c^*_{10} \rangle$. Each gray point in the scatterplot corresponds to a scientist, where $\langle \log c^*_{10} \rangle$ is the average logarithm of her paper impact, excluding the most-cited paper $c^*_{10}$. We report in cyan the *R*-model prediction and in red the analytical prediction (see eq. S36) of the *Q*-model ($R^2 = 0.99$; see section S4.6 and fig. S29 for goodness of the fit). (**E**) Cumulative impact distribution of all papers published by three scientists with the same productivity, $N \simeq 100$, but different $Q$. (**F**) Distribution $P(\hat{p})$ across all publications. For each paper $\alpha$ of scientist $i$, we have $\log p_\alpha = \log c_{10,i\alpha} - \log Q_i$, where $\log Q_i = \langle c_{10,i} \rangle - \langle \hat{p} \rangle$. Therefore, the distribution of $\hat{p} = \log p$, except for a common translational factor $\mu_p$, corresponds to the distribution of $\log c_{10,i\alpha} - \langle \log c_{10,i} \rangle$, which is a normal with $\mu = 0$ and $\sigma^2 = 0.95$ (KS test, $p = 0.48$). (**G**) Distribution of parameter $Q$, $P(Q)$, for all scientists. The gray line corresponds to a log-normal function with $\mu = 0.93$ and $\sigma^2 = 0.46$ (weighted KS test, $p = 0.59$). (**H**) Cumulative distribution of the rescaled impact $c_{10,i\alpha}/Q_i$ for the three scientists in (E). The black line corresponds to the universal distribution $P(p)$. The collapse is predicted by Eq. 1.

$N$ (section S4.10) and on the average impact of the other papers published by the same scientist $\langle \bar{c}^*_{10} \rangle$ (see S4.10). The model prediction for $\langle \log c^*_{10} \rangle(N)$ and $\langle \log c^*_{10} \rangle(\langle \bar{c}^*_{10} \rangle)$ is in excellent agreement with the data (Fig. 3, C and D, and fig. S30), indicating that the hidden parameter $Q$ and variations in the productivity $N$ can explain the empirically observed impact differences between scientists, correcting the shortcomings of the *R*-model.

In summary, the *Q*-model allows us to generate synthetic sequences of publications, by assigning to each scientist an individual parameter $Q$ and a productivity $N$, extracted from the distribution $P(Q, N)$. Each paper in the sequence is assigned an impact calculated as $p \times Q$, where $p$ is randomly drawn from the distribution $P(p)$, identical for all scientists.

## The measurement and accuracy of the hidden parameter *Q*

The model allows us to calculate the parameter $Q_i$ from the sequence of publications $c_{10,i\alpha}$ of each scientist (section S4.5), obtaining for large $N_i$ (see eq. S28 for finite $N_i$ and fig. S28 for the relation between the two estimations of the $Q$ parameter)

$$Q_i = e^{\langle \log c_{10,i} \rangle - \mu_p} \qquad (3)$$

Given its dependence on $\log c_{10,i\alpha}$, $Q$ is not dominated by a single high-impact (or low-impact) discovery but captures instead a scientist's sustained ability to systematically turn her projects into high-impact (or low-impact) publications. For example, although the three scientists in Fig. 3E have the same productivity $N \simeq 100$, Eq. 3 predicts widely different $Q$ values for them, namely, $Q = 9.99$, 3.31, and 1.49. These values accurately reflect persistent differences in their sequence of publications: The $Q = 9.99$ researcher consistently publishes high-impact papers, whereas the publications of the $Q = 1.49$ researcher are consistently of limited impact. Hence, the parameter $Q$ captures a scientist's differentiating ability to take random projects $p$ and systematically turn them into high-impact (or low-impact) publications.

The *Q*-model makes the unexpected prediction that despite the obvious differences in individual career paths, differences in the impact of individual papers should disappear if we use the reduced variable $p_\alpha = c_{10,i\alpha}/Q_i$, a rescaling standard in statistical physics (*28*, *29*). Although the individual $P(c_{10,i\alpha})$ distributions differ greatly, the $P(c_{10,i\alpha}/Q_i)$ distributions for all scientists collapse into a single universal curve $P(p)$ (Fig. 4B), confirming the universal nature of impact across all careers (*30*). Finally, the log-normal $P(Q)$ (Fig. 3G) confirms the model's mathematical self-consistency.

A fundamental limitation of all metrics used in science is their nonstationarity: Productivity, the cumulative number of citations, and the *h*-index all grow in time, making it difficult to compare individuals at different stages of their career. In contrast, we find that the $Q$ parameter is independent of the career stage. To show this, we used a $\Delta N = 30$ paper window to measure changes in $Q$ during the career of a scientist, observing that the $Q$ parameter fluctuates narrowly throughout each career, without systematic changes (Fig. 5A). The magnitude of these fluctuations is explained for 75% of scientists by the stochastic nature of $Q$ (section S4.9), because the estimated $Q$ parameter lies within the uncertainty envelope provided by the model. In the remaining 25% of the cases, the variation in $Q$ is slightly higher than the variation predicted by the stochastic nature of the model (Fig. 5B). However, the magnitude of this surplus variation never exceeds 15%, and the average relative error is always below 10% (section S4.9 and figs. S31 and S32).

Finally, to test the stability of the $Q$ parameter throughout the overall career, and not as a function of productivity, $N$, we consider careers with at least 50 papers and calculate their early and late $Q$ parameters ($Q_{\text{early}}$ and $Q_{\text{late}}$, respectively) using Eq. 3 on the first and second half of their papers, respectively. In this case, the stochastic uncertainty explains the differences between $Q_{\text{early}}$ and $Q_{\text{late}}$ for the large majority of scientists (95.1%, Fig. 5C). Together, these measurements indicate that the $Q$ parameter is generally stable throughout a career, allowing

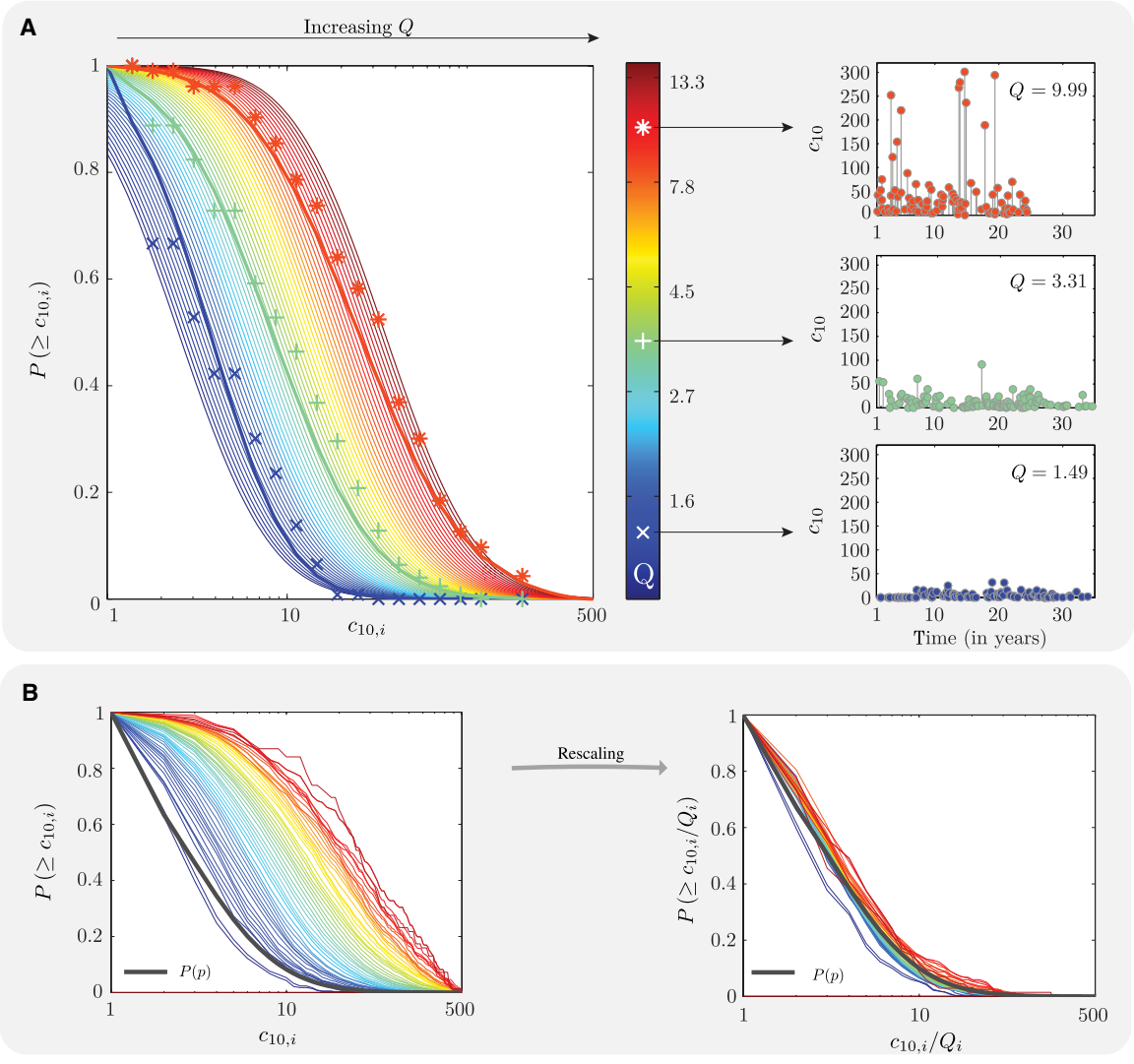



**Fig. 4. Careers and their *Q* parameter.** (**A**) Left: Analytically predicted cumulative impact distributions for different $Q$. The plot also highlights the impact distribution of the three scientists shown in Fig. 2E. The detailed publication record of each scientist is reported on the right, documenting the notable differences between them, given their different $Q$. (**B**) Left: Individual cumulative impact distributions $P(c_{10,i})$. Given the modest number of publications $N$ characterizing most scientists and the impossibility to compute statistically meaningful distributions for many of them, each distribution is computed across all publications of all scientists with the same $Q_i$. The color code captures their $Q$ parameter, as shown in (A). Right: Cumulative distributions of the rescaled impact $c_{10,i}/Q_i$ for the scientists, indicating that the individual distributions collapse on the universal distribution $P(p)$.

us to offer quantitative predictions on the evolution of a scientific career.

## The predictive power of the hidden parameter *Q*

The true value of the $Q$ parameter comes in its predictive power:

(i) The $Q$ parameter allows us to estimate the number of papers a scientist needs to write so that her highest-impact paper gathers $c_{10}^*$ citations (Fig. 6B). We find that scientists with low $Q$ (≃1.2) must write at least 100 papers so that one of them gathers on average 30 citations. Yet, a scientist with the same productivity but $Q = 10$ is expected to author a $c_{10}^* = 250$ paper. Doubling productivity will enhance only with seven citations the highest-impact paper of the low-$Q$ scientist ($Q = 1.2$), whereas it will boost with more than 50 citations $c_{10}^*$ for the high-$Q$ scientist. Overall, Fig. 6B documents that for low-$Q$ scientists, increased productivity cannot boost substantially the chance of publishing a high-impact work; hence, it is very unlikely that they "get lucky."

(ii) A scientist's $h$-index, indicating that her $h$ most-cited papers gather at least $h$ citations (*12*, *15*), is jointly determined by the $Q$ parameter and the productivity $N$ (section S4.11). This analytical prediction reproduces not only the observed $h$-index of all scientists (fig. S33B) but also the evolution of the $h$-index during a scientist's career (Fig. 6, C and D, and fig. S34A). Similar equations describe the cumulative number of citations (Fig. 6D and figs. S33, D to F, and S34B) and the $g$-index (section S4.11), indicating that the traditional performance measures are uniquely determined by $Q$. Given that $Q$ is constant in time, we conclude that productivity only can account for career-wide changes in these measures (Fig. 6, C and D).

(iii) By determining the value of $Q$ during the early stages of a scientific career, we can use it to predict future career impact. The estimation error $\Delta Q$ of $Q$ decreases with the number of published papers $N$ and drops below 10% already after $N = 20$ publications (section S4.12). We can therefore estimate $Q$ based on a scientist's first $N_0$ published papers in Eq. 3 and then use the analytical expression of the $h$-index and of the total number of citations to predict the future impact of a scientist (section S4.12 and fig. S35). Given the stochastic nature of the $Q$-model, an uncertainty envelope accompanies the most likely value of each impact metric. In Fig. 6E, for two scientists, we show the $h$-index prediction up to $N = 150$ after we estimated $Q$ from the first $N_0 = 20$ (top) and $N_0 = 50$ (bottom) papers. Although the initial $h$-index overlaps for the two scientists, their long-term impact diverges, a difference accurately predicted by the $Q$-model. Generalizing for a larger sample of scientists, we find a strong correlation between the predicted and observed $h$-index (Fig. 6F). To quantify the $Q$ model's overall predictive accuracy, we measured the fraction of times that the $h$-index falls within the envelope for scientists with at least 100 papers. The $z_N$

score for each scientist captures the number of SDs the real $h$-index deviates from the most likely $h$-index after $N$ publications. We find that 71% of scientists have $z_{40} \leq 2$ based on $N_0 = 20$, which improves to 81% for $N_0 = 50$ and $z_{70}$ (Fig. 6G). Together, we conclude that the estimation of the $Q$ parameter at early stages has the capability to unveil the long-term career impact.

(iv) To test whether $Q_i$ correlates with outstanding impact, we ranked scientists on the basis of $Q$, $N$, $C_{tot}$, $c^*_{10}$, and their $h$-index. To validate these rankings, we use a receiver operating characteristic (ROC) plot that measures the fraction of Nobel laureates at the top of the ranked list (Fig. 6A). We find that the $Q$-based ranking predicts Nobel-winning careers most accurately, offering the highest area of all ranking measures (Fig. 6A) and the highest precision and recall (section S7 and fig. S45). Equally notable is the finding that the predictive powers of $C_{tot}$, $c^*_{10}$, and the $h$-index are indistinguishable from each other and that the productivity $N$ is the least predictive. Similar results are obtained if we use $Q_i$ to detect Dirac and Boltzmann medalists (figs. S46 and S47). The early-career $Q$ has also the best accuracy in predicting Nobel laureates (section S7.1 and fig. S48).

High-impact discoveries often result from collaborative work (*31*–*33*), mixing scientists with different $Q_i$. To explore the influence of collaborators (*34*, *35*), we used a credit allocation algorithm (*22*, *36*) to attribute different impact shares to each author. We then repeated our entire analysis, finding that the $Q$-model, with slightly revised parameters, can explain the results (section S6.1 and figs. S40 to S43). Further, we find that $Q_i$ is robust to the omission of individual collaborators (section S6.2 and fig. S44). Hence, although collaborative and team effects modulate the success of a particular publication, individual collaborators have only limited influence on $Q_i$.

## Summary and discussion

In summary, we offer empirical evidence that impact is randomly distributed within the sequence of papers published by a scientist, implying that temporal changes in impact during a scientific career can be explained by temporal changes in productivity, luck, and the heavy-tailed nature of a scientist's individual impact distribution. This finding allowed us to systematically untangle the role of productivity, luck, and a scientist's $Q$, predicting that truly high-impact papers require a combination of high $Q$ and luck (high $p$) and that high productivity alone has only a limited effect on the likelihood of high-impact work in a scientific career, if it is not associated with high $Q$. The measurable $Q$ parameter represents a scientist's sustained ability to publish high-impact (or low-impact) papers.

Virtually, all currently used metrics of performance change during the career of a scientist, capturing progression, not sustained ability. In contrast, $Q$ is constant throughout a scientist's career for most scientists (76%), and it is not dominated by a single paper or collaborator, being a measure intrinsically linked to an individual. The fundamental nature of the $Q$ parameter is supported by the fact that the currently used metrics of success, from the $h$-index to cumulative citations, can be calculated from it. $Q$ predicts not only the value but also the time evolution of the traditional impact metrics (Fig. 6, C to F).

All findings presented above are based on a subset of 2887 physicists with a career spanning at least 20 years and a persistent publication record. These scientists have reached a mid- or late-career stage and hence can be considered successful as they survived many selection processes in academia. Although our findings hold in at least six more different disciplines (see section S1.2) and are robust to relaxing the selection criteria (see section S1.4), the studied data sets do not feature young scientists who have left academia early and hence have published only a few papers.

Throughout this work, we have treated long-term impact, as captured by $c^*_{10}$, as an exogenous variable. It seems reasonable, however, that productivity and impact could influence each other. From a mechanistic perspective, for example, some early promising publications might help attract the resources leading to further productivity growth. Early-career impact, quantified with the average $\langle c_{10} \rangle$ for the first 10 papers of a scientist, is associated with career longevity, indicating that the probability to stay in academia is slightly influenced by the impact of a scientist's early publications (fig. S49). The $Q$-model also indicates that the overall number of papers in a career weakly correlates with high $Q$ (Eq. 2). Although the $Q$-model and the predictions provided here are immune from a possible coupling between early impact and overall productivity (section S5), these preliminary findings call for more measurements and models that can accurately capture the coevolution of short-term early impact and productivity (*37*).

Although $Q$ can accurately predict a career impact, the dependence of $Q$ on exogenous factors, such as the quality of the education and current institution (*38*, *39*), size of the research community (*24*, *40*), gender (*41*, *42*), dynamics of subfields (*43*, *44*), or publication habits, remains unknown. Mathematically speaking, the model remains the same if the $Q$ parameter reflects other factors that characterize a scientist. The various robustness checks we performed to discover possible confounding factors, such as career length, decade, team effects, and the analysis

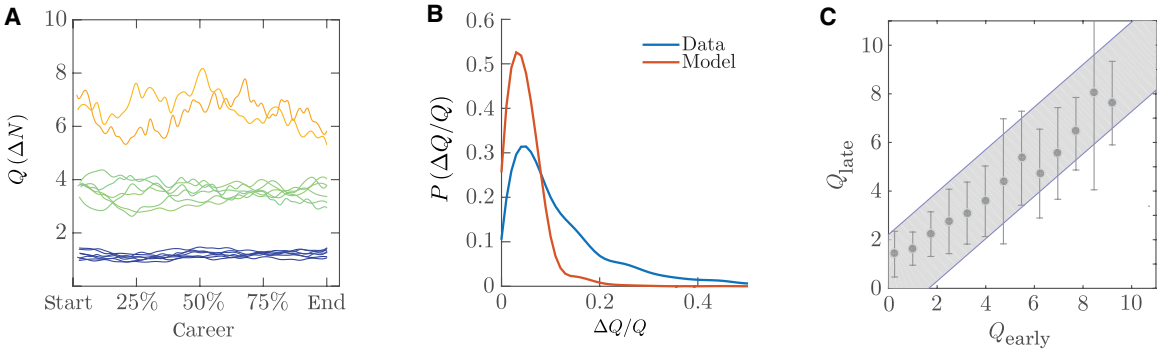



**Fig. 5. Stability of the $Q$ parameter.** (**A**) Time variation of the $Q$ parameter during individual careers. For scientists with at least 100 papers and $Q \simeq 1.2$, $Q \simeq 3.8$, and $Q \simeq 6.5$, we report $Q(\Delta N)$, measured in a moving window of $\Delta N = 30$ papers. For 75% of the scientists, the fluctuations are because we have a finite number of papers in the moving window, the magnitude of the changes being comparable to that predicted by the model with a constant $Q$ (section S4.9). (**B**) Fluctuations of the $Q$ parameter in model and data. We study the distribution of the uncertainty, $\frac{Q(\Delta N)}{Q}$, in both data and synthetic careers with constant $Q$ ($\Delta N = 5$). For 74.7% of the scientists, the fluctuations are comparable to those of the model. For the remaining 25.3%, the SD is slightly higher than the one predicted by the model. (**C**) Comparison between early and late $Q$ parameter. We compare the $Q$ parameter at early-career ($Q_{early}$) and late-career ($Q_{late}$) stage of 823 scientists with at least 50 papers. We measured the two values of the parameters using only the first and second half of published papers, respectively. We perform these measurements on the real data (circles) and on randomized careers, where the order of papers is shuffled (gray shaded areas). For most of the careers, 95.1%, the changes between early- and late-career stages fall within the fluctuations predicted by the null model with randomized paper order, indicating that the $Q$ parameter is stable throughout a career. The observed fluctuations are explained by the finite number of papers in a scientist's career.

of different disciplines and data sets, have failed to offer a simple, straightforward explanation for the origin of the different $Q$ values scientists have. Most likely, the $Q$ parameter is affected by multiple factors, rather than a single one, and more information about its nature might be unveiled once other detailed career information, such as grants and awards, will be available and included in the analysis. Nevertheless, the key differentiating factor of $Q$ from luck is that it has to be sustained. $Q$ is not determined by a single paper, a lucky draw, but by a sustained high performance, throughout the scientist's career. This is reflected in the $\langle \log c_{10} \rangle$ term in Eq. 3, indicating that a single very high impact paper has only a small impact on $Q$. A scientist needs multiple high $c_{10}$ papers to ensure a high $Q$. Uncovering the origin of the $Q$ parameter is a promising future goal, which not only could offer a better understanding of the emergence and evolution of scientific excellence but also might improve our ability to train and nurture high-impact scientists.

## Methods

### *Data sets*

We explore two types of data sets: (i) the publication record of 236,884 physicists publishing in the journal family *Physical Review* from 1893 to 2010 [American Physical Society (APS) data set, see section S1.1 and figs. S1 and S2] and (ii) the combination of 24,630 Google Scholar career profiles with Web of Science (WoS) data, covering 514,896 publications in biology, chemistry, cognitive sciences, ecology, economics, and neuroscience (WoS data set, described in section S1.2 and fig. S3). The results shown in this article refer to 2887 scientists, whose publication record spans at least 20 years, who have at least 10 publications and have authored at least one paper every 5 years, derived from the APS data set (see section S1.3).

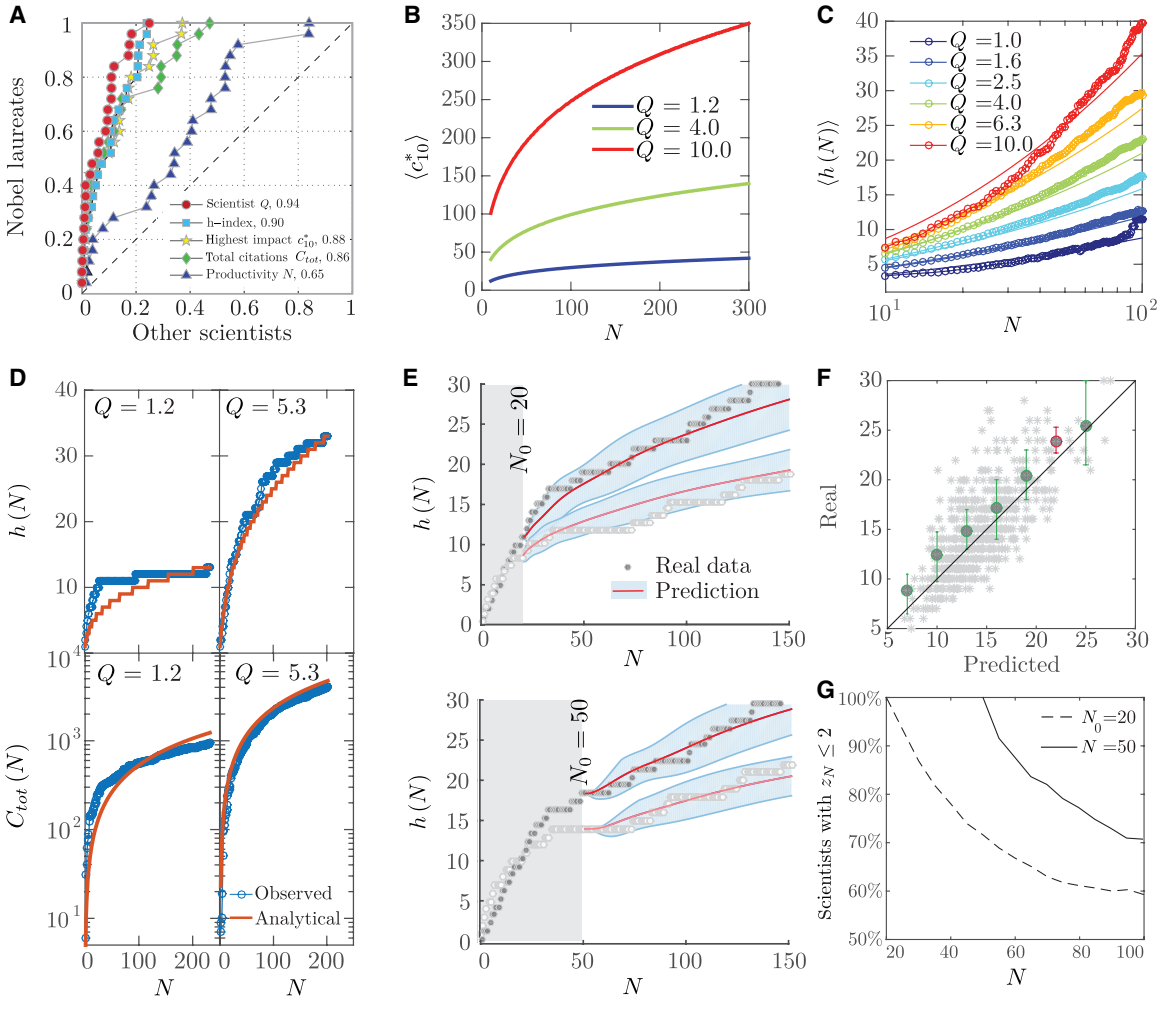



**Fig. 6. Relation between $Q$ and other impact indicators.** (**A**) ROC plot capturing the ranking of scientists based on $Q$, $C_{tot}$, $h$-index, $c^*_{10}$, and $N$. Each curve represents the fraction of Nobel laureates versus the fraction of other scientists for a given rank threshold. The diagonal (no-discrimination line) corresponds to random ranking; the area under each curve provides our accuracy to rank high Nobel laureates. The ranking accuracy is reported in the legend, 1 being the maximum. Precision and recall as a function of rank are discussed in section S7. (**B**) Expected citations to the highest-impact paper, $c^*_{10}$, for a scientist with parameter $Q$ and $N$ publications. The plot illustrates the very low chance of a low $Q$ researcher to publish a high-impact paper. (**C**) Observed versus predicted growth of the $h$-index for scientists with different $Q$. The plot documents the agreement between the analytically predicted $h$-index (eq. S38, continuous line) and the observed value $\langle h(N) \rangle$, obtained by averaging the $h$-index for scientists with the same $Q$ (circles). (**D**) Top: Growth of the $h$-index for two scientists with at least 200 papers and different $Q$ as a function of the productivity $N$ (blue circles), compared with the prediction of eq. S38 (red line). Bottom: For the two scientists in the top panels, we measure the cumulative number of citations as a function of N, Ctot (N), and compare with the prediction of eq. S39. The close agreement between observation and prediction in (C) and (D) shows that the time-independent $Q$ captures an intrinsic property of a scientist and that other indicators, like the $h$-index or cumulative citations, are uniquely determined by $Q$ and productivity. (**E**) For two scientists, we show the $h$-index prediction as a function of $N$ using only early-career information, namely, $N_0 = 20$ (top) and $N_0 = 50$ (bottom), to estimate the $Q$ parameter. Although the initial $h$-index up to $N_0 = 20$ highly overlaps for the two scientists, their long-term impact diverges, a difference accurately predicted by the $Q$-model. (**F**) Scatterplots of predicted and real $h$-index at $N = 60$ based on $Q$ estimated at $N_0 = 20$. The error bars indicate prediction quartiles (25 and 75%) in each bin and are colored green if $y = x$ lies between the two quartiles in that bin and red otherwise. The circles correspond to the average $h$-index in that bin. (**G**) The $z_N$ score for each scientist captures the number of SDs the real $h$-index deviates from the most likely $h$-index after $N$ publications. $z_N \leq 2$ indicates that the real data are within the prediction envelope.

Note that the APS data set contains only citations within the *Physical Review* corpus (see section S1.1); for this reason, the specific number citation numbers are systematically smaller compared to the citations reported by the WoS database.

Our findings are also supported by the analysis of different samples of scientists in the APS data set, selected using a number of different criteria (see section S1.4), and by the analysis of all other disciplines in the WoS data set, which are reported in the Supplementary Materials and referenced throughout the article.

### Citation measures

Citation-based measures of impact are affected by three major problems: (i) citations follow different dynamics for different papers (*6*, *45*), (ii) the average number of citations changes over time (*24*) and (iii) citation count is subfield-dependent (*24*). To overcome (i) for each paper, we use the cumulative number of citations the paper received 10 years after its publication, $c_{10}$, as a measure of its scientific impact (*6*, *45*). We can correct for (ii) and (iii) by normalizing $c_{10}$ by the average $\langle c_{10} \rangle$ of papers published in the same year. Because these corrections do not alter our conclusions for the APS data set, we report results without normalization. For the WoS data set, we instead used normalized citation counts.

### Q-model

The stochastic process behind the $Q$-model is determined by the joint probability $P(p,Q,N)$. The model assumes that a scientist $i$ has a productivity $N_i$ and a parameter $Q_i$ sampled from the marginal distribution $P(Q,N)$, and then extracts $N_i$ values of $p$ from the conditional distribution $P(p|Q,N)$. By assuming that $P(p,Q,N)$ follows a trivariate log-normal distribution with parameters $\mu$ and $\Sigma$, we can write the likelihood function $\mathcal{L}_i$ that a scientist $i$ with $Q_i$ and $N_i$ has a sequence of papers $\{\alpha\}$ with impact $\{Q_i p_\alpha\}$ (see Eq. 1). Finally, with numerical optimization methods, we identify the maximum of the overall log-likelihood function $\log \mathcal{L} = \Sigma_i \mathcal{L}_i$, which provides the numerical estimate of $\mu$ and $\Sigma$ reported in Eq. 2 (see also sections S4.3 and S4.4). This approach also estimates $Q_i$, obtained by maximizing the likelihood function $\mathcal{L}_i$ for each scientist. The maximization provides an analytical expression for $Q_i$, which, for large productivity $N_i$, converges to Eq. 3 (see section S4.5)

This procedure and the measured parameters allow us to generate synthetic sequences of publications: We first extract an individual parameter $Q$ and a productivity $N$ from the distribution $P(Q,N)$. Then, each paper in the synthetic sequence is assigned an impact $pQ$, where $p$ is randomly drawn from the distribution $P(p)$, identical for all scientists.

### REFERENCES AND NOTES


1. I. Fuyuno, D. Cyranoski, Cash for papers: Putting a premium on publication. *Nature* **441**, 792 (2006). doi: 10.1038/441792b; pmid: 16778850
2. J. A. Evans, J. Reimer, Open access and global participation in science. *Science* **323**, 1025 (2009). doi: 10.1126/science.1154562; pmid: 19229029
3. P. Azoulay, Research efficiency: Turn the scientific method on ourselves. *Nature* **484**, 31–32 (2012). doi: 10.1038/484031a; pmid: 22481340
4. B. Owens, Research assessments: Judgement day. *Nature* **502**, 288–290 (2013). doi: 10.1038/502288a; pmid: 24132272
5. O. Penner, R. K. Pan, A. M. Petersen, K. Kaski, S. Fortunato, On the predictability of future impact in science. *Sci. Rep.* **3**, 3052 (2013). doi: 10.1038/srep03052; pmid: 24165898
6. D. Wang, C. Song, A.-L. Barabási, Quantifying long-term scientific impact. *Science* **342**, 127–132 (2013). doi: 10.1126/science.1237825; pmid: 24092745
7. J. A. Evans, Future science. *Science* **342**, 44–45 (2013). doi: 10.1126/science.1245218; pmid: 24092715
8. S. Lehmann, A. D. Jackson, B. E. Lautrup, Measures of measures. *Nature* **444**, 1003–1004 (2006). doi: 10.1038/4441003a; pmid: 17183295
9. S. Lehmann, A. D. Jackson, B. E. Lautrup, A quantitative analysis of indicators of scientific performance. *Scientometrics* **76**, 369–390 (2008). doi: 10.1007/s11192-007-1868-8
10. F. Radicchi, S. Fortunato, B. Markines, A. Vespignani, Diffusion of scientific credits and the ranking of scientists. *Phys. Rev. E* **80**, 056103 (2009). doi: 10.1103/PhysRevE.80.056103; pmid: 20365039
11. D. Hicks, P. Wouters, L. Waltman, S. de Rijcke, I. Rafols, Bibliometrics: The Leiden Manifesto for research metrics. *Nature* **520**, 429–431 (2015). doi: 10.1038/520429a; pmid: 25903611
12. J. E. Hirsch, An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16569–16572 (2005). doi: 10.1073/pnas.0507655102; pmid: 16275915
13. S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, F. Herrera, h-Index: A review focused in its variants, computation and standardization for different scientific fields. *J. Informet.* **3**, 273–289 (2009). doi: 10.1016/j.joi.2009.04.001
14. L. Bornmann, R. Mutz, S. E. Hug, H.-D. Daniel, A multilevel meta-analysis of studies reporting correlations between the *h* index and 37 different *h* index variants. *J. Informet.* **5**, 346–359 (2011). doi: 10.1016/j.joi.2011.01.006
15. D. E. Acuna, S. Allesina, K. P. Kording, Future impact: Predicting scientific success. *Nature* **489**, 201–202 (2012). doi: 10.1038/489201a; pmid: 22972278
16. B. F. Jones, B. A. Weinberg, Age dynamics in scientific creativity. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 18910–18914 (2011). doi: 10.1073/pnas.1102895108; pmid: 22065777
17. P. Azoulay, J. S. Graff Zivin, G. Manso, Incentives and creativity: Evidence from the academic life sciences. *RAND J. Econ.* **42**, 527–554 (2011). doi: 10.1111/rand.2011.42.issue-3
18. A. M. Petersen, M. Riccaboni, H. E. Stanley, F. Pammolli, Persistence and uncertainty in the academic career. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5213–5218 (2012). doi: 10.1073/pnas.1121429109
19. C.-K. Peng *et al.*, Long-range correlations in nucleotide sequences. *Nature* **356**, 168–170 (1992). doi: 10.1038/356168a0; pmid: 1301010
20. D. K. Simonton, Career landmarks in science: Individual differences and interdisciplinary contrasts. *Dev. Psychol.* **27**, 119–130 (1991). doi: 10.1037/0012-1649.27.1.119
21. D. K. Simonton, Creative productivity: A predictive and explanatory model of career trajectories and landmarks. *Psychol. Rev.* **104**, 66–89 (1997). doi: 10.1037/0033-295X.104.1.66
22. H.-W. Shen, A.-L. Barabási, Collective credit allocation in science. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12325–12330 (2014). doi: 10.1073/pnas.1401992111; pmid: 25114238
23. D. de Solla Price, *Little Science, Big Science... and Beyond* (Columbia University, 1963).
24. F. Radicchi, S. Fortunato, C. Castellano, Universality of citation distributions: Toward an objective measure of scientific impact. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17268–17272 (2008). doi: 10.1073/pnas.0806977105; pmid: 18978030
25. M. J. Stringer, M. Sales-Pardo, L. A. Nunes Amaral, Statistical validation of a global model for the distribution of the ultimate number of citations accrued by papers published in a scientific journal. *J. Am. Soc. Inf. Sci. Technol.* **61**, 1377–1385 (2010). doi: 10.1002/asi.v61:7; pmid: 21858251
26. E. J. Gumbel, *Statistics of Extremes* (Dover Publications, 1958).
27. S. Coles, *An Introduction to Statistical Modeling of Extreme Values* (Springer-Verlag, 2001).
28. D. Rybski, S. V. Buldyrev, S. Havlin, F. Liljeros, H. A. Makse, Scaling laws of human interaction activity. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12640–12645 (2009). doi: 10.1073/pnas.0902667106; pmid: 19617555
29. D. Garlaschelli, G. Caldarelli, L. Pietronero, Universal scaling relations in food webs. *Nature* **423**, 165–168 (2003). doi: 10.1038/nature01604; pmid: 12736684
30. T. Vicsek, A question of scale. *Nature* **411**, 421 (2001). doi: 10.1038/35078161; pmid: 11373653
31. S. Wuchty, B. F. Jones, B. Uzzi, The increasing dominance of teams in production of knowledge. *Science* **316**, 1036–1039 (2007). doi: 10.1126/science.1136099; pmid: 17431139
32. B. F. Jones, S. Wuchty, B. Uzzi, Multi-university research teams: Shifting impact, geography, and stratification in science. *Science* **322**, 1259–1262 (2008). doi: 10.1126/science.1158357; pmid: 18845711
33. J. A. Evans, J. G. Foster, Metaknowledge. *Science* **331**, 721–725 (2011). doi: 10.1126/science.1201765; pmid: 21311014
34. J. Moody, D. R. White, Structural cohesion and embeddedness: A hierarchical concept of social groups. *Am. Soc. Rev.* **68**, 103–127 (2003).
35. R. D. Malmgren, J. M. Ottino, L. A. N. Amaral, The role of mentorship in protégé performance. *Nature* **465**, 622–626 (2010). doi: 10.1038/nature09040; pmid: 20520715
36. C.-T. Zhang, A proposal for calculating weighted citations based on author rank. *EMBO Rep.* **10**, 416–417 (2009). doi: 10.1038/embor.2009.74
37. K. Börner, J. T. Maru, R. L. Goldstone, The simultaneous evolution of author and paper networks. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5266 (2004). doi: 10.1073/pnas.0307625100; pmid: 14976254
38. P. Deville *et al.*, Career on the move: Geography, stratification, and scientific impact. *Sci. Rep.* **4**, 4770 (2014). doi: 10.1038/srep04770
39. A. Clauset, S. Arbesman, D. B. Larremore, Systematic inequality and hierarchy in faculty hiring networks. *Sci. Adv.* **1**,e1400005 (2015). doi: 10.1126/sciadv.1400005; pmid: 26601125
40. J. Kaur, F. Radicchi, F. Menczer, Universality of scholarly impact metrics. *J. Informet.* **7**, 924–932 (2013). doi: 10.1016/j.joi.2013.09.002
41. V. Larivière, C. Ni, Y. Gingras, B. Cronin, C. R. Sugimoto, Bibliometrics: Global gender disparities in science. *Nature* **504**, 211–213 (2013). doi: 10.1038/504211a; pmid: 24350369
42. S. F. Way, D. B. Larremore, A. Clauset, *Proceedings of the 25th International Conference on World Wide Web*, Geneva, Switzerland, 11 to 15 April 2016.
43. K. K. Mane, K. Börner, Mapping topics and topic bursts in PNAS. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 5287–5290 (2004). doi: 10.1073/pnas.0307626100; pmid: 14978278
44. R. Sinatra, P. Deville, M. Szell, D. Wang, A.-L. Barabási, A century of physics. *Nat. Phys.* **11**, 791–796 (2015). doi: 10.1038/nphys3494
45. S. Redner, Citation statistics from 110 years of *Physical Review*. *Phys. Today* **58**, 49–54 (2005). doi: 10.1063/1.1996475


### ACKNOWLEDGMENTS

The *Physical Review* data set can be requested from the APS at http://journals.aps.org/datasets. Data with the result of the disambiguation procedure, described in section S1.1, can be found as supplementary materials. An interactive visualization of the data sets, focusing on the random-impact rule, can be found at www.barabasilab.com/scienceofsuccess/. We thank J. A. Evans, S. Fortunato, S. Lehmann, B. Uzzi, B. Coutinho, S. Gil, E. Guney, J. Huang, J. Menche, F. Simini, M. Szell, and all other colleagues at the Center for Complex Network Research (CCNR) for the valuable discussions and comments. We thank H. Shen for the help with the credit share analysis. R.S. and A.-L.B. were supported by Air Force Office of Scientific Research (AFOSR) grants FA9550-15-1-0077 and FA9550-15-1-0364. A.-L.B. was also supported by the Future and Emerging Technologies Project 317 532 "Multiplex" financed by the European Commission. D.W. was supported by AFOSR grant FA9550-15-1-0162 and a Young Investigator Award. P.D. acknowledges support by the National Foundation for Scientific Research and the Research Department of the Communauté française de Belgique (Large Graph Concerted Research Action). R.S. developed the majority of this work during her stay at the CCNR, supported by the J. S. McDonnell Foundation. All authors designed and did the research. R.S. analyzed the empirical data, developed the models and controls, and performed the calculations. A.-L.B. was the lead writer of the manuscript.

### SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/aaf5239/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S49
Supplementary data
References

1 November 2014; resubmitted 22 February 2016
Accepted 29 September 2016
10.1126/science.aaf5239

RESEARCH ARTICLE SUMMARY

CHEMICAL BIOLOGY

# Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity

**Tom H. Wright, Ben J. Bower, Justin M. Chalker, Gonçalo J. L. Bernardes, Rafal Wiewiora, Wai-Lung Ng, Ritu Raj, Sarah Faulkner, M. Robert J. Vallée, Anuchit Phanumartwiwath, Oliver D. Coleman, Marie-Laëtitia Thézénas, Maola Khan, Sébastien R. G. Galan, Lukas Lercher, Matthew W. Schombs, Stefanie Gerstberger, Maria E. Palm-Espling, Andrew J. Baldwin, Benedikt M. Kessler, Timothy D. W. Claridge, Shabaz Mohammed, Benjamin G. Davis***

**INTRODUCTION:** Natural posttranslational modifications (PTMs) to proteins expand the chemical groups available to proteins. The ability to expand posttranslational functional group diversity in an unbounded manner could, in principle, allow exploration and understanding of how these groups modulate biological function. Natural PTMs feature bonds to heteroatoms (non-carbon) made at the γ (Cys Sγ, Thr Oγ, Ser Oγ) or ω (Lys Nω, Tyr Oω) positions of side chains. However, one of the central features of biomolecules is C ($sp^3$)–C ($sp^3$) bond formation. Because all amino acid side chains contain this C–C bond, mastering its construction on proteins could allow free-ranging structural alteration of residues in proteins (both natural and unnatural).

ON OUR WEBSITE

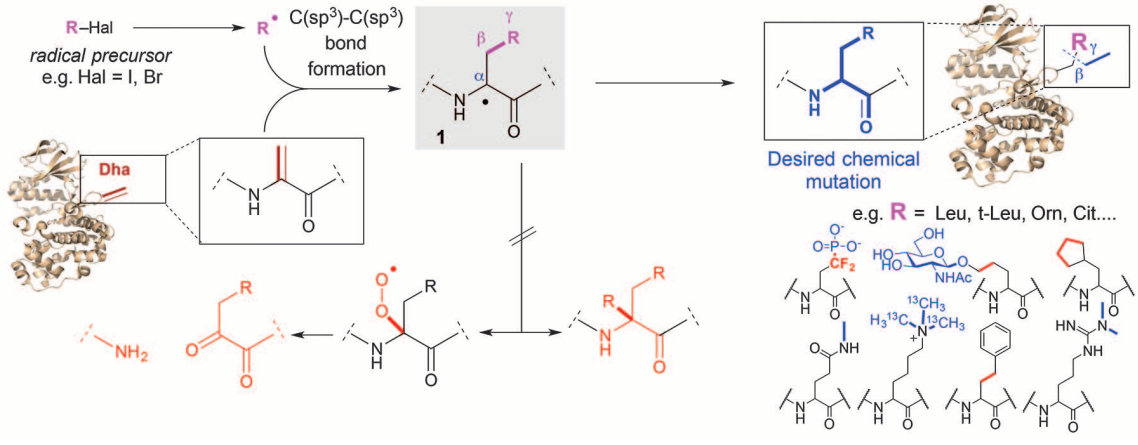
Read the full article at http://dx.doi.org/10.1126/science.aag1465

**RATIONALE:** In principle, C ($sp^3$)–C ($sp^3$) disconnections at the β,γ C–C bond would allow the chemical installation of a wide range of amino acid functionalities. Traditional two-electron chemistry (using nucleophiles and electrophiles) requires reagents that are often incompatible with biological substrates and/or water. Free radicals can be tolerant of aqueous conditions and unreactive (and thereby compatible) with the majority of functionality present in biomolecules. We therefore reasoned that mild, carbon-centered free radical chemistry would be enabled by matching free-radical reactivity with a suitable, uniquely reactive functional group partner that possesses a chemical affinity for such singly occupied molecular orbitals. The amino acid residue dehydroalanine (Dha) can be readily introduced in a site-selective manner genetically, biosynthetically, or chemically; upon reaction with a suitable radical, Dha would favorably generate a stabilized Cα radical **1**. Suitable "quenching" of the central Cα radical intermediate **1** generated after formation of the critical C–C bond would thus allow "chemical mutation" of the side chain.

**RESULTS:** A range of precursor halides (**R**-Hal, Hal = I or Br) allowed the creation of radicals **R•**. These radicals reacted selectively with Dha in peptides and proteins with excellent site selectivity and regioselectivity (>98% β,γ) and typically with a diastereoselectivity of ~1:1. Combined use of R-Hal with $NaBH_4$ under low-oxygen conditions suppressed competing oxidation and disubstitution side reactions of intermediates **1**. This allowed for rapid reactions (typically 30 min) with improved efficiency across a range of representative protein types and scaffolds (all α, α/β folds, all β, receptor, enzyme, antibody). The reactivity of primary, secondary, and tertiary alkyl halides allowed installation of natural, simple hydrophobic residue side chains. Charged or polar protic (e.g., OH, NH) functionality in amino acid side chains was also possible. Even the use of side-chain reagents in unprotected form proved possible, thus highlighting not only exquisite chemoselectivity but also compatibility with common biological functional groups. These transformations enabled the creation of a wide diversity of natural, unnatural, posttranslationally modified (methylated, glycosylated, phosphorylated, hydroxylated) and labeled (fluorinated, isotopically labeled) side chains, as well as difficult-to-access but important residues in proteins (e.g., methyl-Arg, citrulline, ornithine, methyl-Gln, phospho-Ser).

**CONCLUSION:** This approach to chemical editing of amino acid residues, outside of the rigid constraints of the ribosome and enzymatic processing, may prove to be a general technology for accessing diverse, previously unattainable proteins. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ben.davis@chem.ox.ac.uk
Cite this article as T. H. Wright *et al.*, *Science* **354**, aag1465 (2016). DOI: 10.1126/science.aag1465




**Posttranslational chemical mutagenesis through C ($sp^3$)–C ($sp^3$) bond-forming radical reactions.** Modification in a protein after translation using C–C bond formation allows construction of many side chains, not just the modification of existing natural amino acid residues. t-Leu, *tert*-leucine; Orn, ornithine; Cit, citrulline.

RESEARCH ARTICLE

CHEMICAL BIOLOGY

# Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity

**Tom H. Wright,[1] Ben J. Bower,[1] Justin M. Chalker,[1]* Gonçalo J. L. Bernardes,[1]† Rafal Wiewiora,[1]‡ Wai-Lung Ng,[1] Ritu Raj,[1] Sarah Faulkner,[1] M. Robert J. Vallée,[1] Anuchit Phanumartwiwath,[1] Oliver D. Coleman,[1] Marie-Laëtitia Thézénas,[2] Maola Khan,[1] Sébastien R. G. Galan,[1] Lukas Lercher,[1]§ Matthew W. Schombs,[1] Stefanie Gerstberger,[1] Maria E. Palm-Espling,[1] Andrew J. Baldwin,[1] Benedikt M. Kessler,[2] Timothy D. W. Claridge,[1] Shabaz Mohammed,[1] Benjamin G. Davis[1]||**

**Posttranslational modification of proteins expands their structural and functional capabilities beyond those directly specified by the genetic code. However, the vast diversity of chemically plausible (including unnatural but functionally relevant) side chains is not readily accessible. We describe C ($sp^3$)–C ($sp^3$) bond-forming reactions on proteins under biocompatible conditions, which exploit unusual carbon free-radical chemistry, and use them to form Cβ–Cγ bonds with altered side chains. We demonstrate how these transformations enable a wide diversity of natural, unnatural, posttranslationally modified (methylated, glycosylated, phosphorylated, hydroxylated), and labeled (fluorinated, isotopically labeled) side chains to be added to a common, readily accessible dehydroalanine precursor in a range of representative protein types and scaffolds. This approach, outside of the rigid constraints of the ribosome and enzymatic processing, may be modified more generally for access to diverse proteins.**

Natural posttranslational modifications (PTMs) to proteins partially expand the chemical groups available to proteins, modulating both structure and function (Fig. 1A) (*1*). For example, protein glycosylation tunes both physical (e.g., solubility, stability, folding) and biological (e.g., immune response, cell adhesion events, signaling) activity (*2*). Phosphorylation, one of the most frequently occurring posttranslational modifications, is widely used in nature as a powerful functional activation mechanism (as an "on switch") for proteins (*3*). Even relatively small modifications such as methylation have been shown to be critical in a range of pathways with diverse biological effects, such as the transcriptional regulation mediated by histone proteins (*4*). The ability to expand posttranslational functional group diversity in an unbounded manner could therefore, in principle, allow exploration and understanding of even greater and more diverse effects in modulation of biological function.

The vast majority of all known natural PTMs feature bonds to heteroatoms (non-carbon) made at the γ (Cys Sγ, Thr Oγ, Ser Oγ) or ω (Lys Nω, Tyr Oω) positions of side chains (*5*). Yet one of the central features of living "organic" matter is that it exploits carbon's ability as an element to catenate [typically through C ($sp^3$)–C ($sp^3$) bond formation], providing one of nature's most important structural motifs. Because all amino acid side chains contain this bond, mastering its construction on proteins could allow free-ranging structural alteration of residues in proteins (both natural and unnatural) and thence functional reprogramming. Such extension of the chemical space accessible to protein engineering could be considered a near-unlimited form of synthetic biology, a form of "chemical mutagenesis" (*6*, *7*).

Site-directed mutagenesis has revolutionized the study and understanding of proteins (*8*, *9*). This now long-standing technique, however, is generally restricted to the 20 natural amino acid building blocks by the high selectivity of natural aminoacyl tRNA synthetases and the limited plasticity of the ribosome (*10*), which creates an effective "filter" to translation. The incorporation of unnatural amino acids expands a protein's functional capacity and can provide insight into biochemical mechanisms (*11*–*13*). Some strategies for the incorporation of noncanonical residues have emerged as a powerful route to unnatural mutant proteins. Biological techniques such as amber codon suppression (*14*), while useful, remain limited in scope of structural variation (*13*) by the tolerance of the translational machinery and hence must be optimized on a case-by-case basis (*12*). The total or partial synthesis of proteins has been made possible by powerful native chemical ligation techniques (*15*). However, there remain restrictions on the size of the proteins that can be readily synthesized; such syntheses typically require many steps followed by correct refolding, and hence expertise is required for all but the simplest protein targets (*16*).

An alternative, divergent, and potentially unlimited approach would be to incorporate a single amino acid that can act as a general chemical precursor for any desired side chain, whether natural or unnatural (*17*). The introduction of various chemical "tags" (*18*) that allow selective protein modification via reactivity compatible with that of natural biomolecules (*19*) has been an important step toward this goal. However, current protein modification approaches rely on unnatural carbon-heteroatom linkages (*20*) that do not construct the C–C framework found in biology and thus cannot be used for the site-selective introduction of natural posttranslational modifications and their modified variants (*19*). Thus far, the formation of carbon (sp3)–carbon (sp3) bonds for protein modification has remained out of reach, despite the ubiquity of the C–C bond in amino acid side chains. Access to such reactivity would enable the rapid and divergent exploration of both natural and unnatural "side chain" space from a readily accessible precursor.

Here, we demonstrate that such C ($sp^3$)–C ($sp^3$) bond formation is possible in a protein-compatible manner by exploiting the selectivity of carbon-centered radicals to allow a form of general posttranslational mutagenesis (*7*, *17*, *21*).

## Design of a biocompatible C ($sp^3$)–C ($sp^3$) bond-forming reaction

We envisaged a unique strategy through retrosynthetic analysis (Fig. 1B). In principle, C ($sp^3$)–C ($sp^3$) disconnections at the β,γ C–C bond would allow the chemical installation of not only natural amino acid residues but also their posttranslationally modified variants and a wide range of unnatural amino acids. Traditional two-electron chemistry (using nucleophiles and electrophiles; Fig. 1B) results in consideration of reagents likely incompatible with biological substrates by virtue of their reactivity with water and/or the functional groups found in natural biomolecules. We considered that single-electron chemistry might prove more compatible with proteins, because suitable free radicals are tolerant of aqueous conditions (*22*) and are unreactive (and thereby compatible) with the majority of existing functionality present in biomolecules ["bio-orthogonal" (*19*)]. We reasoned that use of such mild, carbon-centered free-radical chemistry (Fig. 1, B and C) would be enabled by matching free-radical reactivity with a suitable, uniquely reactive functional group

[1]Department of Chemistry, University of Oxford, Oxford OX1 3TA, UK. [2]Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Headington, Oxford OX3 7FZ, UK. *Present address: School of Chemical and Physical Sciences, Flinders University, Bedford Park, SA 5042, Australia. †Present addresses: Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK, and Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal. ‡Present address: Sloan-Kettering Institute, New York, NY, USA. §Present address: BMWZ and Institute of Organic Chemistry, Leibniz University Hannover, 30167 Hannover, Germany. ||Corresponding author. Email: ben.davis@chem.ox.ac.uk

partner that possesses a chemical affinity for such singly occupied molecular orbitals (SOMOs). The amino acid residue dehydroalanine (Dha) can be readily introduced in a site-selective manner genetically (*23*, *24*), biosynthetically (*25*), or chemically (*26*, *27*) and is a potent "SOMOphile" (*28*), which, upon reaction with a suitable radical, would favorably generate a capto-dative stabilized Cα radical (Fig. 1C and Fig. 2).

### Development of a peptide-compatible C ($sp^3$)–C ($sp^3$) bond-forming reaction

Our attention focused on methods for the ready generation of suitable carbon-centered free radicals and the suitable productive "quenching" of the central Cα radical intermediate **1** generated after formation of the critical C–C bond (Fig. 2B). We considered that alkyl radicals might be derived from the corresponding alkyl halides (R-Hal) through processes that would generate single electron species either through direct homolytic bond fission (e.g., of the C–Hal bond) or through single-electron transfer from metals with suitable redox potentials in their low valence states (*29*, *30*) [e.g., zinc (*31*, *32*) $Zn^0$ or indium (*33*) $In^0$] followed by halide anion loss.

We chose leucine (Leu), an amino acid residue that has a widespread occurrence in proteins, as an initial test side-chain system, which we could potentially generate from the readily available simple organic compound isopropyl iodide as a precursor (Fig. 2A, R = iPr). Pleasingly, reaction of Dha-containing derivatives and peptides as small-molecule models under $Zn^0$-mediated conditions in aqueous buffer [pH 5 to 6, $NH_4Cl$ (aq)] afforded the corresponding Leu derivatives directly (fig. S1 and supplementary materials). Generation of Dha from cysteine (Cys) followed by conversion to Leu thus allowed the overall "chemical mutation" of a residue (Cys → Leu) inside an intact peptide backbone; under optimized conditions (see supplementary materials), this could be achieved in >90% yield and in less than 30 min.

Extension of this methodology to a range of alkyl halides (R-Hal) enabled the synthesis of a variety of natural and unnatural amino acid residues, including unnatural aliphatic and cyclic structures (fig. S2). Notably, not only were primary, secondary, and tertiary alkyl halides all tolerated, allowing installation of the natural simple and hydrophobic residue side chains, so too was the presence of polar protic (e.g., hydroxyl and amine) functionality common in amino acid side chains. The use of these side-chain reagents proved possible even without protection, thus highlighting not only exquisite chemoselectivity but also compatibility with common biological functional groups and hence biological compatibility (orthogonality). Full characterization of the adducts (see fig. S1 and supplementary materials) confirmed absolute regioselectivity (>98% with diastereomeric ratio 55:45) for the radical addition, consistent with the designed, matched polarities of the radicals and the corresponding radical acceptor Dha, respectively (*34*).

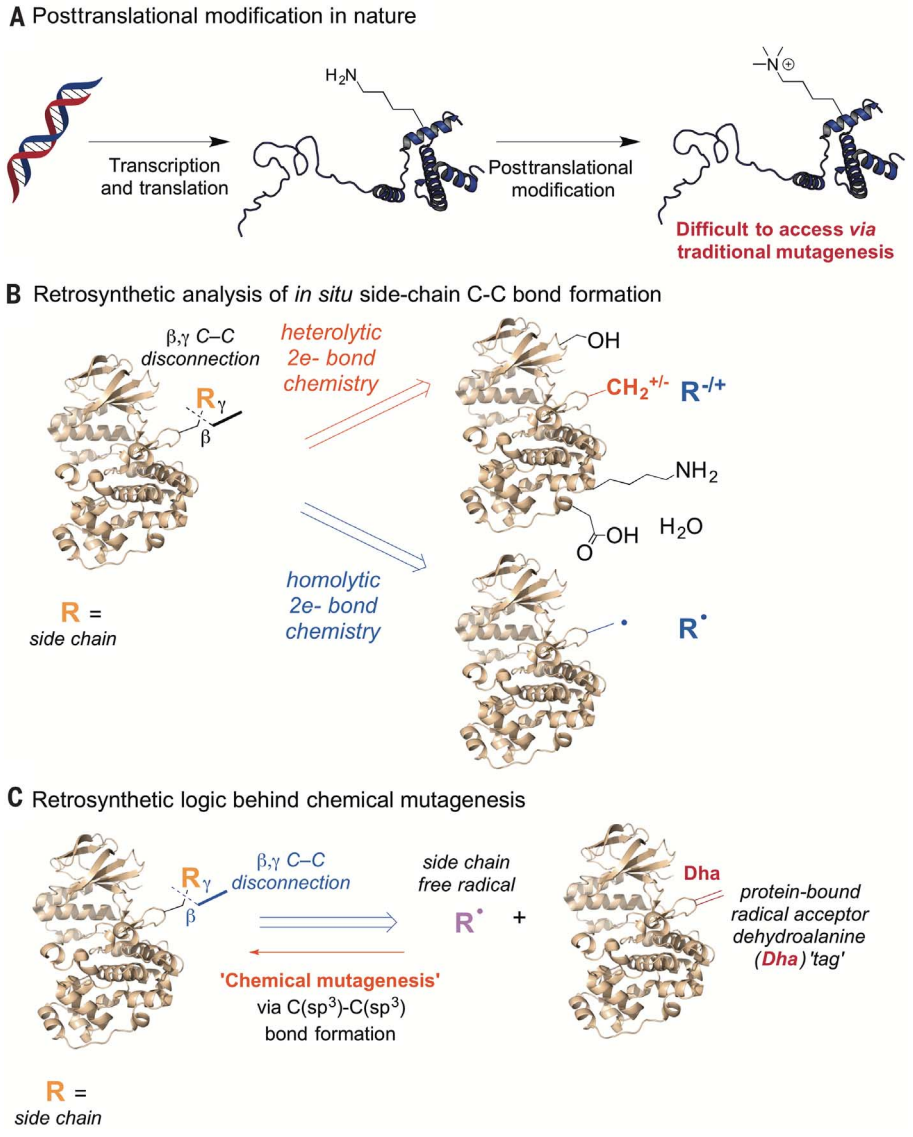



**Fig. 1. Design of a biocompatible C ($sp^3$)–C ($sp^3$) bond-forming reaction as part of a general strategy for chemical mutagenesis.** (**A**) Natural posttranslational modification exploits C-heteroatom bond formation, such as C–N bond formation in Lys methylation, to diversify protein structure and hence modulate function. (**B**) Modification in a protein after translation, akin to PTM formation, but using C–C bond formation would allow construction of many side chains, not just the modification of existing natural amino acid residues. Retrosynthetic analysis (blue arrow) suggests a C–C β,γ bond disconnection and reveals several possible "synthons." For heterolytic two-electron reactions (shown in red), the resulting disconnection yields "synthetic equivalents" $\mathbf{R^+}$ or $\mathbf{R^-}$ or reagents that would react and/or quench with the water in the necessary aqueous solvent or with example protein side chains shown. However, for the homolytic one-electron reaction (shown in blue), the resulting disconnection yields a free-radical "synthon" that would allow an equivalent $\mathbf{R^\bullet}$ for which compatibility can be envisaged. (**C**) A functional group with affinity for carbon-centered free radicals that was suitably polarized could act as the "synthetic equivalent" corresponding to the other protein-based "synthon" arising from this homolytic disconnection. The residue dehydroalanine (Dha) is the functional group proposed as a radical reactive "synthetic equivalent" that would allow corresponding C–C β,γ bond formation.

We observed that the predicted Cα radical intermediate **1** (Fig. 2B), as well as displaying advantageous stability that would favor initial reaction, was sufficiently long-lived to allow further reaction with other radicals in unproductive and unwanted termination reactions (fig. S3). For example, reaction with a second alkyl radical afforded disubstituted ("dialkylated") products, whereas reaction with molecular oxygen (which in its natively abundant state is the triplet form $^3O_2$ that may react with radicals) led to apparent oxidative protein cleavage (*35*). Notably, additional amounts of reagents were seemingly needed for full conversion because of such

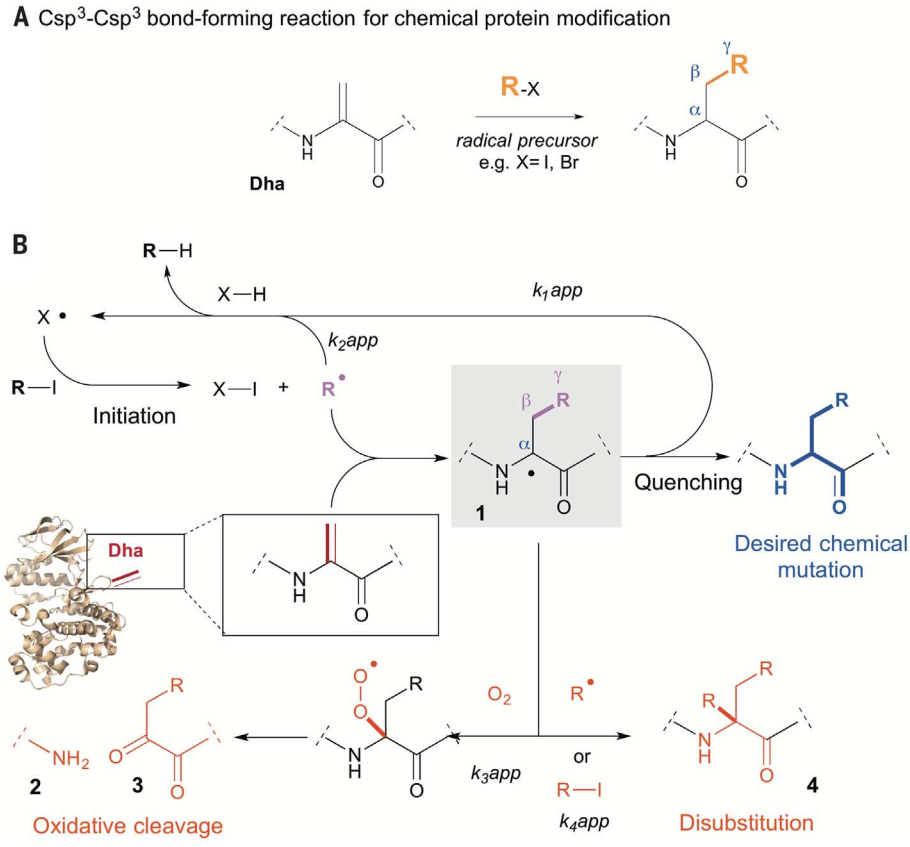



**Fig. 2. Proposed radical reaction to allow posttranslational mutagenesis and suggested mechanism.** (**A**) The required "side-chain free-radical" synthetic equivalent could be generated from a suitable radical precursor R–X and then reacted with the radical-reactive SOMOphile residue dehydroalanine (Dha) as a privileged unnatural amino acid "tag" to allow site-selective "chemical mutagenesis." (**B**) This would generate the capto-dative stabilized intermediates **1**. The proposed mechanism of free-radical R• generation illustrates the paths to efficient C–C bond-forming chain reaction and desired product (in blue) and competing side reactions and unwanted products (in red).

observed competing processes (reduction of alkyl halide to alkane as well as disubstitution and oxidative degradation). The successful development of a radical reaction for protein modification (Fig. 2A) therefore necessitated a means of eliminating and controlling these undesired pathways.

## Mechanism-guided reaction development allowed optimization of protein-compatible C ($sp^3$)–C ($sp^3$) bond formation

Next, two initial model proteins were selected to test radical reactivity on more complex extended polypeptides: a highly ordered three-layer α/β Rossman-fold serine protease [subtilisin from *Bacillus lentus* (SBL)] and a three-α-helix protein representative of the histone fold that contains both ordered and disordered motifs (histone H3). Dha was installed (*27*, *36*) site-selectively from corresponding single cysteine variants (see supplementary materials) to generate radical acceptor sites at position 156 in SBL (SBL-Dha$^{156}$) and at three separate sites (9, 27, and 64) in histone H3 (H3-Dha$^{9}$, H3-Dha$^{27}$, H3-Dha$^{64}$). These allowed us to test alterations of the protein scaffold as well as variation of the reaction site within the same protein scaffold.

Under essentially identical conditions to those used on small-molecule systems (isopropyl iodide, $Zn^0$ aqueous ammonium acetate buffer, pH 6), we observed direct Dha → Leu "chemical mutation" (fig. S4) in both SBL and histone H3. However, additional side products were also detected. Careful isolation, trapping, and characterization of these side products (figs. S5 to S8) revealed that they were the product of two competing pathways, both of which are consistent with the long-lived intermediacy of the Cα capto-dative radical formed after addition of **R•** to Dha. The observed oxidative cleavage products, C-terminal amide **2** and dicarbonyl **3** (figs. S5 to S7), would arise from the termination reaction of the Cα radical with triplet oxygen (Fig. 2B) (*37*). Disubstituted ("dialkylated") products **4** were observed to arise from the termination reaction of the Cα radical with **R•**. Consistent with these analyses, "peptide mapping" [tryptic tandem mass spectrometry (MSMS)] confirmed clean site selectivity of both the desired mutations and these side reactions: No residues other than that determined by the Dha "tag" site were identified (fig. S8). Notably, these results were also wholly consistent with the corresponding Cγ- and Cα-disubstituted products observed under comparable conditions from small-molecule models (see above) (fig. S3). These observations in proteins therefore were also consistent with radical addition at Cγ followed by termination of the Cα radical intermediate thus generated (Fig. 2B).

The generality of these parallel processes—chemical mutation with competing oxidative cleavage and disubstitution—was confirmed by its observation in both of the scaffolds (SBL and histone H3) and at several sites (e.g., Lys$^{9}$, Lys$^{27}$, and Lys$^{64}$ within histone H3). These results not only highlight the need for an improved reaction with better control of radical addition, they also provide compelling chemical evidence for the intermediacy of the proposed Cα radical intermediate (Fig. 2B) and hence the (partial and initial) success of our designed C–C radical-forming reaction in proteins.

These methodological observations and mechanistic rationalization allowed us in turn to optimize the balance between the desired radical side-chain addition and the unwanted competing side reactions (oxidative cleavage and disubstitution). In our initial reaction systems, our use of metal-mediated single-electron transfer exploited a system that relies on the redox potentials of the metals that were used not only in the initiation step but also in the subsequent second-electron transfer, which creates an enolate that is quenched by protonation (likely from solvent) (see fig. S9). In principle, more effective second-electron transfer to enhance enolate formation would allow more rapid formation of desired product at the expense of side reactions (quenching of the intermediate Cα radical before side reaction). Survey of redox potentials (*38*, *39*) suggested various metal potentials that might prove useful; of those that are compatible with water, indium suggested itself as a strong alternative candidate to zinc standard reduction potentials $E°/V$, $Zn^{(2/0)}$ –0.76; $In^{(1/0)}$ –0.14; $In^{(2/1)}$ –0.40; $In^{(3/2)}$ –0.49). This tuning of the single-electron donor, through the use of $In^0$ instead of $Zn^0$, proved partially successful, yielding cleaner and more effective chemical mutation through radical addition (see fig. S10 and supplementary materials) on several but not all substrates. In particular, certain primary iodides (e.g., the side chains of MeArg and MeLys, where Me = methyl) were so reactive that unwanted disubstitution side products remained.

Next, we considered alternative methods for "quenching" of the intermediate Cα radical. In principle, direct hydrogen atom ("radical hydride") transfer (Fig. 2B) would not only enhance desired product formation but, by building a suitable chain reaction, could prove more efficient and sustainable. However, analysis of the putative chain cycle (Fig. 2B) highlighted that this would require selective increase of the rate of this step (encompassed by $k_{1app}$) to a greater extent than that of not only the oxidative and disubstitution side reactions ($k_{3app}$ and $k_{4app}$, respectively) but also over that of the direct reduction of alkyl

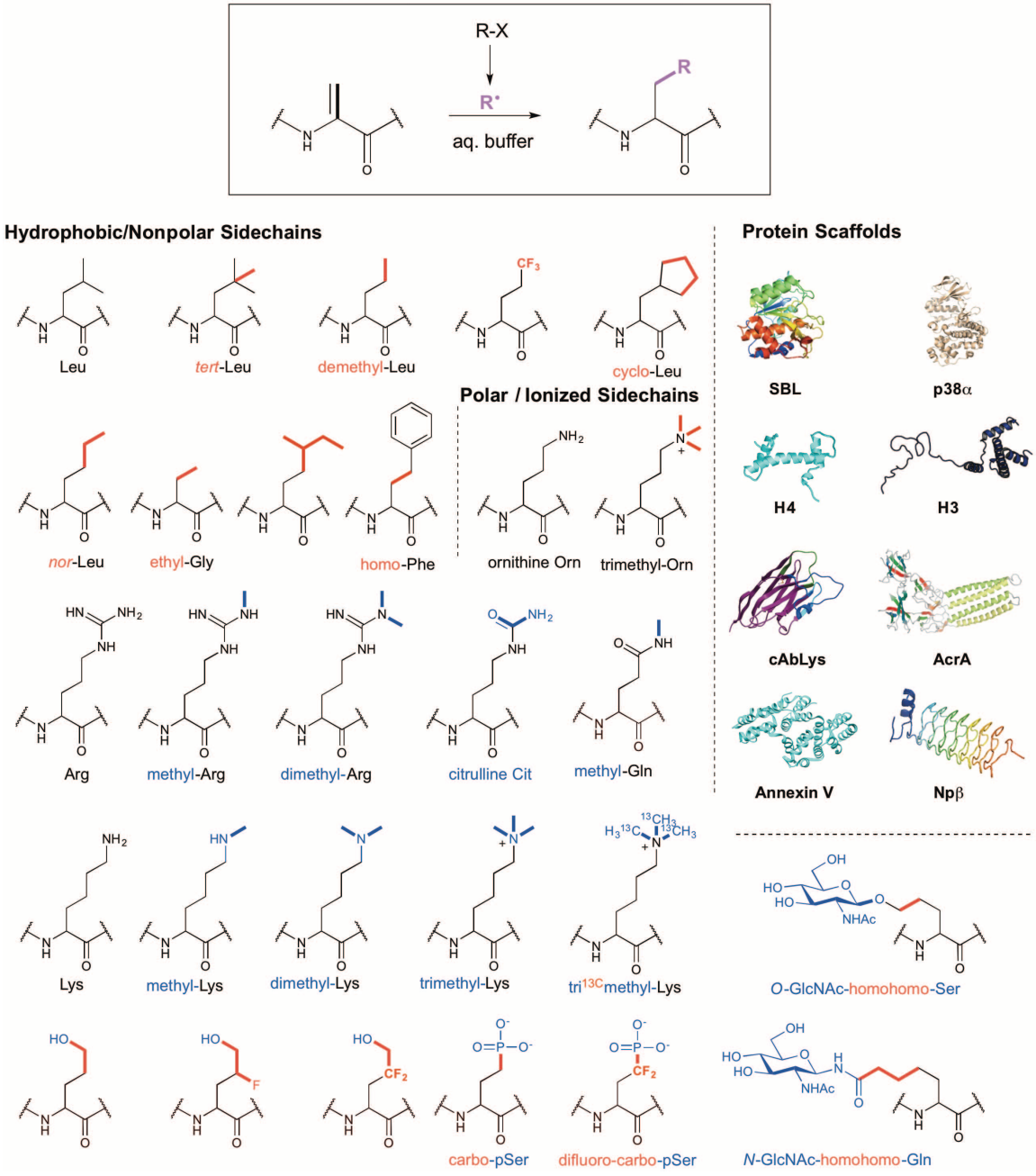



**Fig. 3. Biocompatible C ($sp^3$)–C ($sp^3$) bond-formation allows wide-ranging chemical mutagenesis.** Application of the optimized bond-forming reaction allows direct installation of side chains onto Dha tags found in multiple, representative protein scaffolds and at different sites within the same scaffolds. Attachment of polar, hydrophobic, nonpolar, ionized, and modified side chains with natural motifs (black), with natural modifications (blue) or with unnatural motifs (red) were all possible allowing the construction of >50 individual protein examples representative of >25 side chains on eight varied, representative protein scaffolds.

iodide (R–I → R–H, $k_{2app}$) (Fig. 2B). A range of traditional "radical hydride" sources were screened (e.g., $R_3SnH$, $R_3SiH$, RSH), yet none proved useful and all appeared to favor $k_{2app}$ over $k_{1app}$. We reasoned that these bulkier hydride sources preferentially transfer hydride H• to less bulky radicals (such as direct transfer to **R•**, thereby favoring $k_{2app}$), and so we next tested less hindered hydride sources (which might be able to access the more hindered intermediate Cα radical **1**, as desired). Although borohydrides $RBH_3^-$ are traditionally viewed as nucleophilic hydride sources, rare studies on these (*40*–*42*) and related aluminum hydrides (*43*) have previously suggested possible radical behavior under certain circumstances (*44*). Strikingly, we found that $NaBH_4$ in aqueous solution proved to be a highly effective reagent, allowing improved efficiency (as judged by the need for reduced equivalents of iodides) and by the quality of protein chemical mutation products (fig. S11). The radical nature of this controlled, clean, and efficient reaction was confirmed not only by direct observation of radicals by electron paramagnetic resonance (EPR) (fig. S12) but also through the use of radical trapping; the reaction was fully inhibited by substoichiometric 4-hydroxy-TEMPO (4-hydroxy-2,2,6,6-tetramethylpiperidin-1-oxyl) or acrylamide (fig. S13 and supplementary materials).

Finally, having suppressed competing disubstitution, we were able to efficiently suppress competing oxidative cleavage simply through the removal of molecular oxygen from the buffer solutions in which we conducted "chemical mutagenesis" reactions; controlled equilibration experiments at a variety of oxygen partial pressures (figs. S14 and S15) revealed that incubation at <6 ppm $O_2$ for 6 hours prior to reaction proved generally sufficient. Application of the combined optimized conditions led directly to clean "chemically mutated" proteins without side reactions that, when sequenced by MSMS, were interpreted directly as the intended mutation (fig. S16). Both whole-protein liquid chromatography–MS and post-digestion MSMS (figs. S17 and S18) confirmed that there was no nonselective alkylation or overalkylation by the halide reagents on other residues in proteins substrates. Additionally,

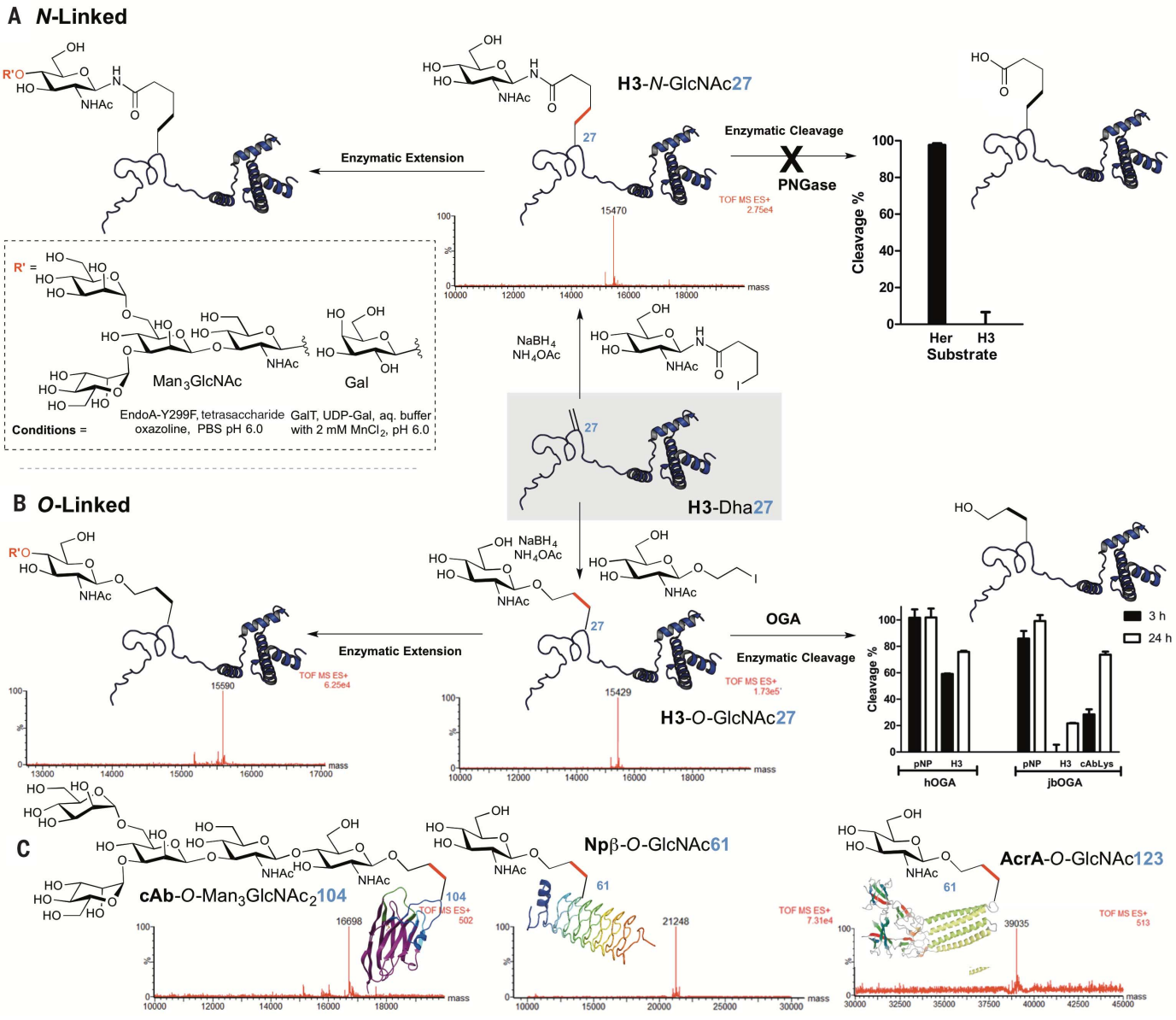



**Fig. 4. Use of C–C bond-forming mutagenesis to create N- and O-glycosylated proteins.** Using a common intermediate, H3-Dha[27], both forms of GlcNAc (O-linked and N-linked) could be readily introduced and their behavior in extension and cleavage by relevant glycan-processing enzymes tested. (**A**) Despite the position in histone H3, which is not normally glycosylated, N-linked GlcNAc was readily extended to either a disaccharide (Gal-GlcNAc, LacNAc) or the core pentasaccharide (found in all natural N-linked glycans) by appropriate enzymatic systems (GalT and EndoA, respectively). However, this site proved resistant to enzymatic cleavage under conditions that led to full cleavage in natural N-glycosylated sites (conditions, PNGase, 2 M urea, pH 8.0). (**B**) As for N-linked glycosylation, the disaccharide LacNAc (Gal-GlcNAc) and the core pentasaccharide were readily formed from O-linked GlcNAc, despite the unnatural site. Strikingly, O-GlcNAcase from a range of sources showed cleavage activity even at unnatural sites and notably with the human enzyme hOGA. (**C**) Glycosylation of varied protein platforms, antibody cAbLys, efflux protein AcrA, and pentapeptide repeat protein Nβ all proved possible. Experimental data reported in bar graphs represent the normalized percentage (mean ± SD) of cleaved substrate, taken in triplicate.

control experiments in which $NaBH_4$ was omitted from the reaction mixture showed that the halide reagents alone did not react with proteins under the standard conditions.

### Biocompatible C ($sp^3$)–C ($sp^3$) bond-formation enables a "toolbox" for building natural and unnatural proteins

With optimized conditions for C–C bond formation enabling chemical mutation in hand, we next explored the breadth of the side chains that could be introduced (and hence mutations that would be accessible) and the protein scaffolds that they could be introduced into (Fig. 3). Representative proteinogenic and nonproteinogenic; polar, nonpolar, aromatic, ionized, and modified amino acid residues, bearing both natural and unnatural motifs, were chosen and all readily incorporated. This demonstrated the tolerance of the reaction to many of the most common amino acid (and biological) functional groups: hydroxyl (OH), aminyl (NH), guanidine, amide, thereby highlighting its excellent biocompatibility (orthogonality). Choice of the corresponding organic iodide, typically readily available commercially or through chemical synthesis, allowed wide-ranging and systematic variations. Thus, not only could Leu be incorporated, so could a series of systematic variants of Leu: demethyl-Leu (smaller), *tert*-Leu (bulkier), *nor*-Leu

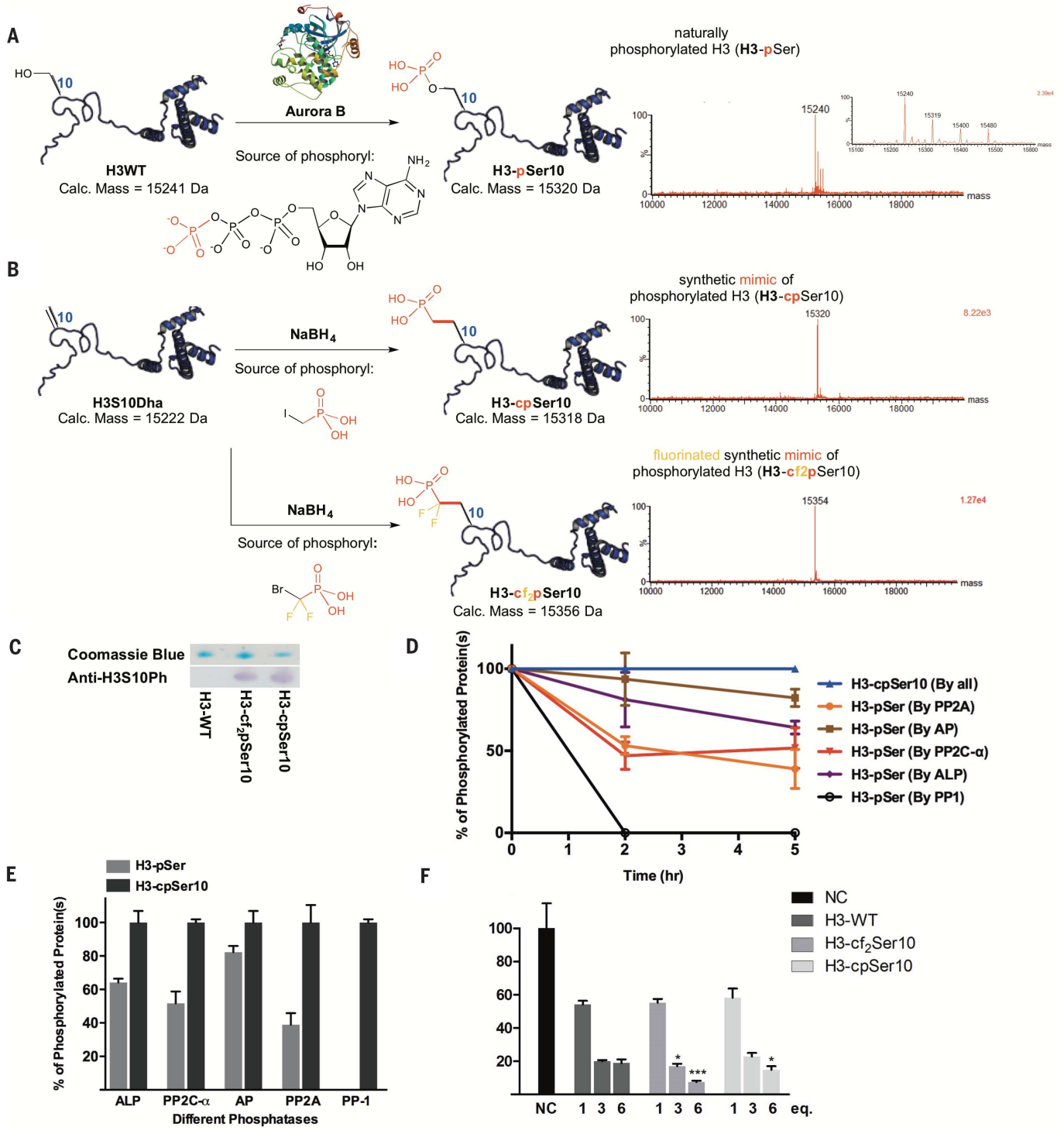



**Fig. 5. Use of C–C bond-forming mutagenesis to build phosphorylated histone proteins.** Formation of natural and synthetic phosphorylated histone H3 (H3-cpSer$^{10}$ and H3-*cf2*pSer$^{10}$) via enzymatic phosphorylation and chemical mutagenesis, respectively. (**A**) Natural phosphorylation is catalyzed by Aurora B kinase and gives rise to a mixture (inset to mass spectrum) of phosphoforms. (**B**) Chemical phosphorylation via C–C bond formation gives essentially homogeneous products H3-cpSer$^{10}$ and H3-*cf2*pSer$^{10}$. (**C**) Western blot analysis with antibody to H3-pSer$^{10}$ and SDS-PAGE gel stained with Coomassie blue (for full gel, see fig. S29) shows that the phosphomimics carba-pSer (cpSer) and difluorocarba-pSer installed by building a C–C bond are faithfully recognized as phosphorylated by the corresponding antibody raised to bind the natural modification (pSer). (**D** and **E**) A comparison of stability of H3-pSer$^{10}$ and H3-cpSer$^{10}$ under the actions of five different protein phosphatases shows that the C–C bond in H3-cpSer$^{10}$ provides full resistance even to enzymes that fully and readily degrade the naturally phosphorylated H3-pSer$^{10}$. The *y* axis shows normalized percentage (mean ± SD) of phosphorylated protein taken in triplicate. (**F**) Alpha-screen binding assay using phospho-reader protein MORC3 shows binding by H3-cpSer$^{10}$ and enhanced binding by H3-*cf2*pSer$^{10}$. Measurements were performed at least three times and were analyzed by Student's *t* test at 95% confidence interval [H3-*cf2*pSer$^{10}$ > H3-WT in binding, $P = 0.0285$ (3 eq.) and $P = 0.0004$ (6 eq.), respectively; H3-cpSer$^{10}$ > H3-WT, $P = 0.0291$ (6 eq.)]. Error bars indicate SEM.

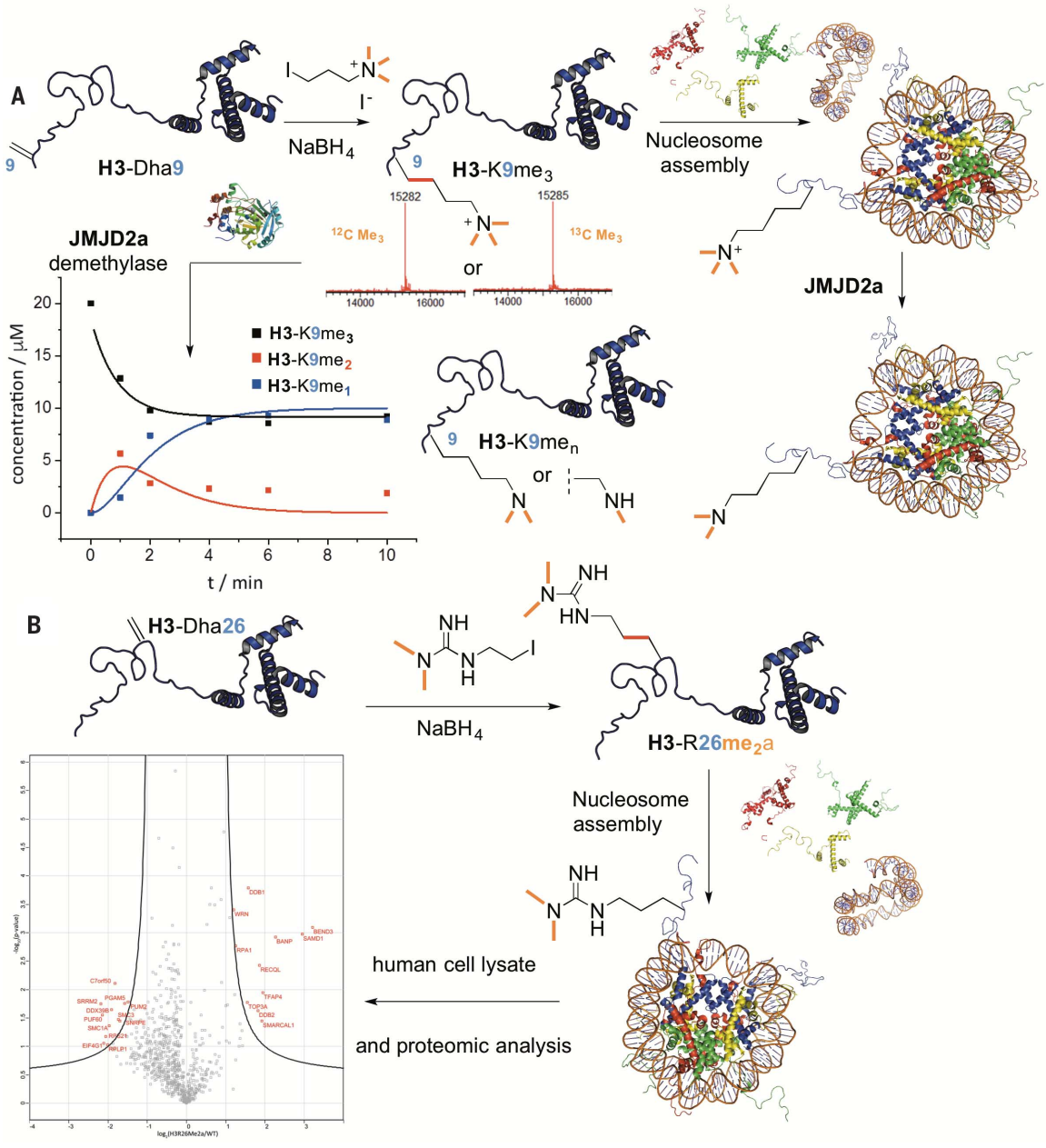



**Fig. 6. Use of C–C bond-forming mutagenesis to build methylated nucleosomes.** Formation of methylLys-modified (H3K9me$_3$) or methylArg-modified (H3R26me$_2$a) nucleosomes by chemical mutagenesis enables insight into the biological functions of key histone modifications. (**A**) Chemical methylation via C–C bond formation allows installation of not only K9me$_3$ with natural isotope distribution (primarily $^{12}$C) but also a $^{13}$C-enriched variant precisely placed at the methyl-group carbon atoms. These reveal the time course for demethylation by the demethylase enzyme KDM4a/JMJD2a, in both an isolated protein context (shown here by LC-MS), and, notably, in the context of intact nucleosomes. This suggests the use of such "isotope-PTMs" as probes of demethylase activity, for example, by $^{13}$C NMR. Time points represent the mean of three independent experiments; error bars (SEM) shown in SI are omitted here for clarity; curves were fitted by global least-squares regression algorithm to solutions of a simplified first-order model. (**B**) C-C bond-forming mutagenesis also enabled the direct site-specific installation of asymmetric dimethylarginine residues into intact histones. H3R26me$_2$a was cleanly installed via C-C bond formation (see supplementary materials), assembled into nucleosomes. When used to probe human (HeLa) cell extracts for interaction partners (three independent biological and two technical replicates), previously unanticipated protein partners implicated in rDNA repression and recognition of DNA damage were identified among 797 quantified proteins in nuclear extract, which suggests that ablation of DNA-to-H3 hydrogen bonding may be critically affected by such R26 methylation. Significant interacting protein partners (denoted by their gene name) identified upon label-free quantification (LFQ) on Student's *t* test analysis are shown in red in the "volcano plot" inset (*x* axis, logarithmized ratio of LFQ intensity difference among two groups; *y* axis, logarithmized *P* value from the test statistics).

(slimmer), and "cyclo"-Leu (conformationally restricted) (Fig. 3). Similarly, systematic variation of side chain length, methylene unit by methylene unit, was also possible (e.g., Ala, ethyl-Gly, demethyl-Leu, *nor*-Leu). We were also able to strategically replace, with atomic precision, methyl groups in residues with their labeled or precisely altered variants: thus, $CH_3 \rightarrow CF_3$ (e.g., in demethyl-Leu) or $CH_3 \rightarrow {}^{13}CH_3$ (e.g., in trimethyl-Lys). Current methods for isotopic labeling based on "feeding" experiments result in universal incorporation at every codon-determined site; here, isotopic labels can be installed at a single site. Such precisely fluorinated or isotopically labeled amino acids are not only powerful biophysical reporters (*45*–*47*), with use particularly in protein (e.g., $^{19}$F and $^{13}$C) nuclear magnetic resonance (NMR) methods, but can also act as modulators of protein structure and binding (*48*). Indeed, use of $^{19}$F NMR allowed us to further confirm both the regioselectivity and stereoselectivity for chemical mutations (fig. S19), which proved to be essentially identical to that found on peptide models (>98% and diastereomeric ratio ~1:1).

Key posttranslational modifications also proved accessible, including glycosylation (*2*) (in O- and N-linked form), Lys methylation (*49*) (in all three states: mono-, di-, and tri-, as well as labeled tri-$^{13}$C), Arg methylation (*50*) (mono- and di-), and Gln methylation (*51*). Notably, no other chemical methods exist (barring total protein synthesis) for the installation of the majority of these

residues. Moreover, several residues that have been previously biologically inaccessible in proteins were also readily introduced. These included ornithine (Orn), which by virtue of intramolecular cyclization chemistry cannot be loaded onto tRNAs and hence is incompatible with ribosomal incorporation (*52*), and di- and trimethylated Lys that cannot yet be incorporated into proteins [even indirectly, as mono-methyl-Lys currently is (*53*)] by cellular stop-codon suppression (*13*).

Finally, we surveyed the introduction of side chains to representative examples of protein functions (structural, channel, enzyme, glycoprotein) from differing protein folds (with varying levels of α, β, and unstructured secondary motifs) and species types. Thus, as well as SBL and histone H3, we also surveyed the variously structured histone H4; the transmembrane bacterial efflux component protein AcrA (*54*); p38α mammalian mitogen-activated MAP kinase (*55*); the mammalian antibody cAbLys3 (*56*); the apoptosis marker binding protein annexinV (*57*); and the pentapeptide-repeat protein Npβ (*58*). These also represent proteins that are variously associated with localization in different cellular environments: nuclear (histones H3 and H4), cytosolic (p38α, annexinV), transmembrane (AcrA, Npβ), and extracellular (cAbLys3, annexinV). They also allowed us to survey highly diverse architectures: α helix–coiled coil (AcrA); histone fold (mixed unfolded and α helix, histones H3 and H4); mixed α/β fold (SBL, p38α); variable-domain immunoglobulin fold (four-strand β sheet plus five-strand β sheet, cAbLys); α helix–rich globular annexin fold (four × 5–α-helix domains in a "super-helix," annexinV); and even a right-handed quadrilateral β helix (Npβ). Sites of modifications sat within various feature types including helix, sheets and loose loops (table S2). Notably, all proteins proved compatible with side-chain attachment at all targeted sites (see supplementary materials for full characterization details). Multiple sites in the same proteins were also surveyed: five different sites in histone H3 (sites 9, 10, 26, 27, and 64) and two in histone H4 (sites 16 and 17). Together, these cumulative variations of side chains in different protein substrates at different sites allowed access to >50 "chemical mutants."

Analysis of sequence (e.g., by tryptic MSMS peptide mapping; fig. S20), structure (e.g., secondary structural content by circular dichroism, amide- and aliphatic-resonance analysis and diffusional analysis by protein $^{1}$H-NMR; figs. S21 to S23) and retention of function (e.g., enzymatic activity, antibody-binding function and biomarker recognition, protein complex assembly), not only confirmed the site selectivity of the C–C bond-forming chemical mutation but also suggested no gross alteration of global structure (see supplementary materials), although it should be noted that localized structural changes cannot be discounted. Notably, too, test experiments in both model systems and proteins containing disulfides suggested good compatibility of reagents [including 2,5-dibromohexane diacetamide (DBHDA), consistent with prior results (*27*), and $NaBH_4$; see supplementary materials and figs. S24 and S25] with disulfides.

## C ($sp^3$)–C ($sp^3$) bond-formation chemical mutation enables diverse techniques for the study of protein methylation, glycosylation, and phosphorylation

With this ability to directly insert side-chain alterations and hence perform chemical mutagenesis on proteins with wide variation in protein and site, we chose to test differing proof-of-principle strategies that would allow insight into the biological function of posttranslationally modified residues (and their mimics) that are ordinarily difficult to install with fidelity into proteins by other means. To this end, we chose three of the most important PTMs: glycosylation, phosphorylation, and methylation.

Glycosylation is the most diverse of the posttranslational modifications (*1*, *59*), yet has been prominently absent as a readily accessible motif in proteins via chemical or genetic mutagenesis methods (*60*). Until now no general chemical method for convergently installing N- and O-linked glycans has been possible (*59*, *60*). We used C–C bond-forming mutagenesis along with corresponding (entirely unprotected) N- and O-linked glycosidic iodides to install *N*-acetylglucosamine (GlcNAc), a glycan that is found naturally in both N- and O-linked form, to create an unnatural glycosylation site at position 27 of histone H3 (using the same, common divergent histone H3–Dha$^{27}$ protein intermediate, Fig. 4). Despite the unnatural constitution of this site, enzymatic extension with either glycosyltransferase or endoglycosidase allowed the overall installation of more complex glycans onto both N- and O-linked GlcNAc, even up to the N-linked core pentasaccharide that is found in all N-linked glycoproteins (*1*, *59*). We discovered that despite an apparent plasticity with respect to protein scaffold (*61*), PNGase (peptide-*N*-glycosidase), a widely used *N*-glycosidase (*62*), did not cleave synthetic variants with extended side-chain length (Fig. 4A and fig. S26). In striking contrast, we discovered that a variety of synthetic O-GlcNAcylated glycoproteins were cleaved quite readily by O-GlcNAcases from different sources (Fig. 2B and fig. S27), including the human O-GlcNAcase (hOGA) enzyme. The latter, which is implicated in diabetes (*63*), dementia (*64*), and cancer (*65*), has until now been presumed selective, given that it is the sole encoded protein O-GlcNAcase in the human genome (*66*); our results suggest a previously unappreciated and surprising plasticity. In addition to the N- and O-glycosylation of histone H3, we were able to chemically N-glycosylate and O-glycosylate other sites and proteins, including the naturally glycosylated protein AcrA, the pentapeptide repeat protein Nβ, and notably the heavy-chain antibody cAbLys with the putative Fcγ receptor ligand glycan $Man_3GlcNAc_2$ (Fig. 4C and fig. S28).

Aurora B kinase, which is overexpressed in cancers (*67*), phosphorylates Ser$^{10}$ of histone H3 during mitosis (*68*) and is hence implicated as a controlling factor in cell division and proper distribution of genetic information. The lability of phosphorylation and the mixtures of phosphorylated proteins often formed from natural enzymatic phosphorylation greatly complicates the study of phosphoryl groups on given sites, such as H3-pSer$^{10}$ (*69*). We used C–C bond-forming mutagenesis along with corresponding (notably unprotected) iodophosphonates to create stable analogs of H3-pSer$^{10}$ in which a single oxygen atom was replaced by methylene or difluoromethylene units to create the carba-phosphoSer variants *c*pSer (*70*) and *cf2*pSer (*71*) (Fig. 5A); MS analysis revealed that this could be achieved with a purity that is not possible with current biological methods for phosphorylating histone H3 (Fig. 5B). Antibodies and appropriate binding proteins [such as the "reader" protein MORC3 (*72*)] not only recognized the resulting phosphomimic proteins (Fig. 5, C and F, and fig. S29) but also proved stable to either chemical or enzyme-catalyzed dephosphorylation (Fig. 5, D and E, and fig. S30), even with phosphatases that readily processed naturally phosphorylated histone H3 (e.g., protein phosphatases 1 or 2A). It also proved possible to readily install carba-pSer into other proteins (fig. S31). Although *c*pSer$^{10}$ proved to be a functioning mimic of pSer$^{10}$, it has been argued (*73*, *74*) that fluorophosphonates may act as more effective phosphate mimics by virtue of p$K_a$ (*71*), polarity, and shape (*75*). The C–C bond-forming mutagenesis allowed us to also install a difluorophosphonate-Ser variant (difluoro-carba-pSer, *cf2*pSer) at the same site of histone H3 to create H3-*cf2*pSer$^{10}$. Consistent with improved mimicry, this variant showed enhanced binding to MORC3, validating proposed (*71*, *73*–*75*) difluorophosphonate mimicry of phosphates in proteins (Fig. 5F).

Protein methylation (*76*) is a central biological process (e.g., in epigenetic regulation and cell signal transduction). Yet the precise elucidation of the functional mechanistic role of methylation at the molecular level and the ready delineation of proteins associated with the "methylproteome" (e.g., "writers," "readers," and "erasers") remains a grand challenge in biological science. We chose two methylated protein targets that have not been possible to create through other protein generation methods and created these through the site-selective C–C bond-forming mutagenesis reaction: a site-selectively $^{13}$C-labeled variant of trimethylated Lys in histone H3 (Fig. 6A) and a dimethylated Arg residue site-selectively installed into in an intact nucleosomal particle (Fig. 6B). We created H3-[$^{13}$C-Me]$_3$-Lys$^9$ (Fig. 6A) with precise trimethylation on Lys$^9$, which has been previously been observed and implicated in direct transcriptional regulation (*77*, *78*). The structure and function of this chemical mutant of histone H3 was explored by protein MS and NMR. These revealed ready observation of the $^{13}$C-labeled protein, by virtue of its isotopic label, and the readily observable processing and release of the $^{13}$C-labeled methyl groups ([$^{13}$C]Me) from ([$^{13}$C]Me)$_3$-Lys$^9$ in histone H3 by the known demethylation enzyme JMJD2A/KDM4a (Fig. 6A and fig. S32A) (*79*, *80*).

Having installed and explored precise Lys methylation in an isolated histone protein, we explored precise Lys and Arg methylation in the context of an entire, intact, nucleosomal particle (Fig. 6, A and B); both LysMe3 and ArgMe2a were readily introduced. Both anti-LysMe3 antibody recognition and JMJD2a/KDM4a-catalyzed demethylation were readily demonstrated using synthetically methylated nucleosome (fig. S32B). Arg methylation, and especially asymmetric dimethylation (ArgMe2a), remains only a partially understood alteration. Moreover, given the variant isomeric forms of Arg methylation and dimethylation, precise control of the installation of methylation to create representatively methylated proteins has also not proved possible. Using the C–C chemical mutagenesis method (Fig. 6B), we precisely installed ArgMe2a into site 26 of histone H3 (H3R26me$_2$a) [a site implicated (*81*) in so-called "cross-talk" epigenetic modifications with an order that is not clearly understood] in intact nucleosomal particles.

These synthetic nucleosomal probes bearing asymmetric-dimethylation at H3R26me$_2$a allowed the identification of key partner proteins through affinity enrichment proteomics in human cells (Fig. 6B and table S1). Notably, two of the strongest interacting partners, BEND3 and BANP, contain BEN domains, a recently characterized α-helical module found in chromatin-associated proteins (*82*); BEND3 (*83*), a novel rDNA transcription repressor, is the most enriched interacting protein partner. Interestingly, BEND3 can recruit PRC2 complex that promotes H3K27me$_3$ modification, a known transcription repression marker (*84*); this suggests potential cross-talk between these two modifications leading to generation of a repressive chromatin state. Intriguingly, a majority of the remaining significant interactors (SMARCAL1, RECQL, DDB1, DDB2, TOP3A) are annotated as being involved in DNA replication/repair. Although a link between histone arginine methylation and DNA repair has not been previously reported, the results of this experiment suggest increased accessibility of nucleosome-bound DNA to a range of relevant DNA-binding effectors; such a loss of DNA-to-arginine hydrogen bonding would be anticipated upon methylation. Such is the flexibility of the C–C chemical mutagenesis method that it also proved possible to readily install methylated Lys (LysMe, LysMe2, LysMe3), methylated Arg (ArgMe2a, ArgMe1), and methylated Gln into a wide range of other sites and proteins (fig. S33). On the basis of the ready discovery of previously unanticipated interacting partners for Arg26Me2a (see above), we anticipate that these too will prove to be powerfully precise probes of methylation function and direct "methyl-proteome" interactions.

Finally, we tested the scalability of the C–C chemical mutagenesis method; applications such as translational trials of biologics and structural biological studies can necessitate larger (multimilligram) amounts of proteins. Reaction of 10 mg of the Dha protein H3S10Dha with the relevant bromide and sodium borohydride allowed the installation of difluoromethylenephosphohomoalanine (*cf*2pSer) in useful "synthetic" (isolated, 65 to 70%) yield after desalting thereby suggesting utility even for larger-scale synthetic applications in protein chemistry (fig. S34).

## Discussion

The compatibility in the use of radicals as effective and yet benign reactive intermediates for protein modification suggests that other radical-based methods (*35*) may prove powerful in the field of protein chemistry. Indeed, the key implicated propagating intermediate in our C–C bond-forming mutagenesis, the capto-datively stabilized Cα backbone radical, is similar to other stabilized radicals suggested in natural processes (*35*); not only may nature have long been taking advantage of such methods, but other routes of access to such intermediates could allow similarly powerful bond-forming strategies. This protein radical chemistry is likely to require new reagents such as boron-based compounds, which have played an important role in radical chemistry (*85*, *86*). The associated mechanisms of these reagents remain a topic of active debate (*43*, *44*), and mechanisms other than the one we propose here cannot be discounted.

Given their benign application and compatibility, we can also envisage the ready combination of the C–C mutagenesis reaction with other protein chemistries or assembly methods. For example, thioester-mediated backbone assembly methods ("native chemical ligation") (*15*) typically use peptide fragments with N-terminal Cys residues that remain at the junction point after "ligation." Using the C–C mutagenesis reaction, these Cys residues could be readily converted to Dha and thence to almost any residue of choice. Indeed, in a proof-of-principle of such a combined strategy, we have been able to generate Thr- and Leu-containing peptides in which the Thr-Leu moiety is derived from thioester-mediated amide ligation with Cys followed by C–C chemical mutagenesis to Leu (see fig. S35 and supplementary materials).

Because the simple choice of reagent allows for the variation of single-atom substituents (e.g., O → $CH_2$ → $CF_2$ in pSer, *c*pSer, and *cf*2pSer, as we have shown here), our approach may enable a more molecular understanding of mechanism of proteins and a fine-tuning of function. The use of stable phosphoryl mimics (such as *c*pSer or *cf*2pSer) suggests itself as a promising way to "fish" not only for phospho-binders ("phospho-readers," as we have shown here) but for enzyme partners that would process their natural modification counterparts (e.g., phosphatases that would cleave pSer but, as we have shown here, cannot cleave *c*pSer). It should be noted in this context that although amber-codon suppression ("genetic code expansion") methods have proven highly powerful in certain cases, many useful and biologically relevant structures elude current approaches. For example, the residue types ($^{13}$C-trimethyl-Lys, dimethyl-Arg, O- and N-linked glycosylated residues, difluorocarba- or carba-pSer) installed in the proof-of-principle studies (see above) have proven intractable to direct installation into biologically relevant sites by amber-codon suppression (*87*). Indeed, residues bearing only small differences from their natural counterparts (methylated Lys, Arg, Gln), which we were able to incorporate chemically, sometimes prove difficult to incorporate by such genetic methods because of their strong structural resemblance in the translational machinery to their unmodified counterparts.

Note that although the C–C bond-forming mutagenesis method described here allows the creation of the relevant constitution (connectivity) found in proteins, our analysis of the stereoselectivity in both peptides and proteins (see above) suggests approximately equal formation of both D- and L-configurations at the mutation site. This would result in yet more complex mixtures of protein diastereomers for multiple mutation sites.

The improved access to relevant protein architectures will potentially reveal surprising biological functions. For instance, our data have revealed that the human enzyme hOGA is more plastic in its cleavage activity of O-GlcNAcylated proteins than had been previously anticipated. Coupled with inferred plasticity of the corresponding glycosyltransferase (OGT) that was recently suggested by studies on peptides (*88*), this suggests that any O-GlcNAcylation–associated regulatory mechanism may be much looser than previously realized. The elucidation of new binding partners for nucleosomes in previously inaccessible methylation states (e.g., asymmetrically dimethylated ArgMe2a) suggests that many other new interactions may be discoverable via C–C chemical mutagenesis. These in turn will allow us to identify and synthetically "program" into proteins exactly those residues that engender wanted functional (e.g., pharmaceutical) benefit in a truly broad manner. For instance, the chemical glycosylation (via C–C bond mutagenesis) of an antibody fragment with possible sugar ligands for the Fcγ receptor raises the possibility of new cell-killing strategies (*89*) mediated by synthetic antibody fragments. In this way, we envisage access to synthetic proteins that will allow application of newly elucidated functions in, for example, new protein drugs or "synthetic biologics" (*90*).

## Materials and methods

### *General Dha formation on histone protein substrate*

The relevant protein cysteine mutant (10 mg of lyophilized powder) was dissolved in 500 μL of reaction buffer (3M guanidinium hydrochloride, 100 mM sodium phosphate, pH 8.0) and 30 mg DTT was added. The resultant protein solution was shaken at room temperature for 30 min. Excess DTT was removed by passing the solution through a PD MiniTrap G25 (equilibrated with the same buffer). Eluting with 1 mL of the reaction buffer gave the protein at a final concentration of 10 mg/mL. To the reduced protein

mutant was added 50 μL of a 2,5-dibromohexane diacetamide (DBHDA) stock solution (0.5 mM, 100 mg in 610 μL of DMF). The reaction was gently shaken for 45 min at room temperature and then 90 min at 37°C. The reaction mixture was concentrated to 500 μL using a Vivaspin 500 concentrator (MWCO 10 kDa) and excess reagent was removed by passing the solution through a PD MiniTrap G25 (equilibrated with chemical mutagenesis reaction buffer). Eluting with 1 mL of the same buffer provided the dehydroalanine-containing protein.

### General chemical mutagenesis method

A 2 mL eppendorf tube containing a solution of protein (typically 0.2- 5 mg/mL, 200 μL, ~ 10-200 μM) in reaction buffer (pH 4-8, most common buffers tolerated) was ported into a $N_2$ atmosphere glovebox (Belle Technologies) and equilibrated to anaerobic conditions (cap open) overnight at 4°C in a standard benchtop eppendorf cooler. In parallel, a stock solution of the bromide or iodide of interest (prepared by dissolving solids in water and adjusting the pH to that of the reaction buffer), or the solid reagent itself, was also ported into the glovebox. In the case of liquid reagents, either a stock solution was prepared as above, or the neat reagent was ported into the glovebox and stored at the appropriate temperature. Solid sodium borohydride (0.3 – 1 mg, smaller amounts of sodium borohydride can be difficult to weigh accurately) was ported into the glovebox in a 2 mL eppendorf tube and placed in the cooler kept at 4°C. Following a minimum of 8 hours degassing, inside the glovebox, the volume of the protein solution was ascertained by pipette and "topped up" with reaction buffer to the desired volume (e.g., 200 μL). 1-20 μL of the previously prepared iodide or bromide stock solution (typically 100 – 2000 molar equivalents relative to protein) was then pipetted into the protein solution, and the resultant solution gently mixed by pipette. The entire volume (~200 μL) of the protein solution was then added to the 2 mL eppendorf tube kept at 4°C and containing solid $NaBH_4$. The tube was gently shaken by hand, causing effervescence/foaming to spread through the solution. Any discoloration from the iodide reagent should rapidly disappear (In some cases, particularly at high pH where the hydrolysis of borohydride is retarded, sustained effervescence is observed, whereas for most iodides a colorless and still solution is obtained after ~20 min incubation at 4°C). The reaction was incubated at 4°C for 30 min (cap open) before being removed from the glovebox. Shaking is not necessary but may assist in removing evolved gases. The eppendorf tube was then capped and removed from the glovebox. Outside the glovebox, the reaction solution was loaded onto SpinTrap G-25 desalting column, pre-equilibrated with the desired storage buffer. Elution according to manufacturers instructions afforded the desired chemical mutagenesis product, which was analyzed by LC-MS. For downstream applications, additional SEC may be necessary for sample desalting.

### REFERENCES AND NOTES


1. C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto Jr., Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44**, 7342–7372 (2005). doi: 10.1002/anie.200501023; pmid: 16267872
2. K. Ohtsubo, J. D. Marth, Glycosylation in cellular mechanisms of health and disease. *Cell* **126**, 855–867 (2006). doi: 10.1016/j.cell.2006.08.019; pmid: 16959566
3. P. Cohen, The origins of protein phosphorylation. *Nat. Cell Biol.* **4**, E127–E130 (2002). doi: 10.1038/ncb0502-e127; pmid: 11988757
4. E. I. Campos, D. Reinberg, Histones: Annotating chromatin. *Annu. Rev. Genet.* **43**, 559–599 (2009). doi: 10.1146/annurev.genet.032608.103928; pmid: 19886812
5. T. H. Wright, M. R. J. Vallée, B. G. Davis, From chemical mutagenesis to post-expression mutagenesis: A 50 year odyssey. *Angew. Chem. Int. Ed.* **55**, 5896–5903 (2016). doi: 10.1002/anie.201509310; pmid: 27119221
6. C. Zioudrou, M. Wilchek, A. Patchornik, Conversion of the L-serine residue to an L-cysteine residue in peptides. *Biochemistry* **4**, 1811–1822 (1965). doi: 10.1021/bi00885a018
7. K. E. Neet, D. E. Koshland Jr., The conversion of serine at the active site of subtilisin to cysteine: A "chemical mutation". *Proc. Natl. Acad. Sci. U.S.A.* **56**, 1606–1611 (1966). doi: 10.1073/pnas.56.5.1606; pmid: 5230319
8. C. A. Hutchison 3rd *et al*., Mutagenesis at a specific position in a DNA sequence. *J. Biol. Chem.* **253**, 6551–6560 (1978). pmid: 681366
9. J. A. Brannigan, A. J. Wilkinson, Protein engineering 20 years on. *Nat. Rev. Mol. Cell Biol.* **3**, 964–970 (2002). doi: 10.1038/nrm975; pmid: 12461562
10. C. R. Woese, G. J. Olsen, M. Ibba, D. Söll, Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol. Mol. Biol. Rev.* **64**, 202–236 (2000). doi: 10.1128/MMBR.64.1.202-236.2000; pmid: 10704480
11. L. Wang, P. G. Schultz, Expanding the genetic code. *Angew. Chem. Int. Ed.* **44**, 34–66 (2004). doi: 10.1002/anie.200460627; pmid: 15599909
12. J. W. Chin, Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* **83**, 379–408 (2014). doi: 10.1146/annurev-biochem-060713-035737; pmid: 24555827
13. A. Dumas, L. Lercher, C. D. Spicer, B. G. Davis, Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* **6**, 50–69 (2015). doi: 10.1039/C4SC01534G
14. C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182–188 (1989). doi: 10.1126/science.2649980; pmid: 2649980
15. P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994). doi: 10.1126/science.7973629; pmid: 7973629
16. L. R. Malins, R. J. Payne, Recent extensions to native chemical ligation for the chemical synthesis of peptides and proteins. *Curr. Opin. Chem. Biol.* **22**, 70–78 (2014). doi: 10.1016/j.cbpa.2014.09.021; pmid: 25285753
17. J. M. Chalker, B. G. Davis, Chemical mutagenesis: Selective post-expression interconversion of protein amino acid residues. *Curr. Opin. Chem. Biol.* **14**, 781–789 (2010). doi: 10.1016/j.cbpa.2010.10.007; pmid: 21075673
18. J. M. Chalker, G. J. L. Bernardes, B. G. Davis, A "tag-and-modify" approach to site-selective protein modification. *Acc. Chem. Res.* **44**, 730–741 (2011). doi: 10.1021/ar200056q; pmid: 21563755
19. E. M. Sletten, C. R. Bertozzi, Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **48**, 6974–6998 (2009). doi: 10.1002/anie.200900942; pmid: 19714693
20. C. D. Spicer, B. G. Davis, Selective chemical protein modification. *Nat. Commun.* **5**, 4740 (2014). doi: 10.1038/ncomms5740; pmid: 25190082
21. L. Polgár, M. L. Bender, Simulated mutation at the active site of biologically active proteins. *Adv. Enzymol. Relat. Areas Mol. Biol.* **33**, 381–400 (1970). pmid: 4916857
22. T. Perchyonok, *Radical Reactions in Aqueous Media* (Royal Society of Chemistry, 2009).
23. J. Wang, S. M. Schiller, P. G. Schultz, A biosynthetic route to dehydroalanine-containing proteins. *Angew. Chem. Int. Ed.* **46**, 6849–6851 (2007). doi: 10.1002/anie.200702305; pmid: 17685371
24. F. P. Seebeck, J. W. Szostak, Ribosomal synthesis of dehydroalanine-containing peptides. *J. Am. Chem. Soc.* **128**, 7150–7151 (2006). doi: 10.1021/ja060966w; pmid: 16734454
25. Y. O. You, M. R. Levengood, L. A. F. Ihnken, A. K. Knowlton, W. A. van der Donk, Lacticin 481 synthetase as a general serine/threonine kinase. *ACS Chem. Biol.* **4**, 379–385 (2009). doi: 10.1021/cb800309v; pmid: 19292452
26. G. J. L. Bernardes, J. M. Chalker, J. C. Errey, B. G. Davis, Facile conversion of cysteine and alkyl cysteines to dehydroalanine on protein surfaces: Versatile and switchable access to functionalized proteins. *J. Am. Chem. Soc.* **130**, 5052–5053 (2008). doi: 10.1021/ja800800p; pmid: 18357986
27. J. M. Chalker *et al*., Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2**, 1666–1676 (2011). doi: 10.1039/c1sc00185j
28. B. Giese, *Radicals in Organic Synthesis: Formation of Carbon-Carbon Bonds*, J. E. Baldwin, Ed. (Pergamon, 1986).
29. C. J. Li, Organic reactions in aqueous media with a focus on carbon-carbon bond formation. *Chem. Rev.* **93**, 2023–2035 (1993). doi: 10.1021/cr00022a004
30. C.-J. Li, Organic reactions in aqueous media with a focus on carbon-carbon bond formations: A decade update. *Chem. Rev.* **105**, 3095–3165 (2005). doi: 10.1021/cr030009u; pmid: 16092827
31. C. Petrier, C. Dupuy, J. L. Luche, Conjugate additions to α,β-unsaturated carbonyl compounds in aqueous media. *Tetrahedron Lett.* **27**, 3149–3152 (1986). doi: 10.1016/S0040-4039(00)84739-5
32. T. Huang, C. C. K. Keh, C.-J. Li, Synthesis of α-amino acid derivatives and amines via activation of simple alkyl halides by zinc in water. *Chem. Commun.* **2002**, 2440–2441 (2002). doi: 10.1039/B204585K; pmid: 12430480
33. H. Miyabe, M. Ueda, A. Nishimura, T. Naito, Indium-mediated intermolecular alkyl radical addition to electron-deficient C=N bond and C=C bond in water. *Org. Lett.* **4**, 131–134 (2002). doi: 10.1021/ol017013h; pmid: 11772108
34. J. M. Tedder, Which factors determine the reactivity and regioselectivity of free radical substitution and addition reactions? *Angew. Chem. Int. Ed. Engl.* **21**, 401–410 (1982). doi: 10.1002/anie.198204011
35. C. L. Hawkins, M. J. Davies, Generation and propagation of radical reactions on proteins. *Biochim. Biophys. Acta* **1504**, 196–219 (2001). doi: 10.1016/S0005-2728(00)00252-8; pmid: 11245785
36. J. M. Chalker, L. Lercher, N. R. Rose, C. J. Schofield, B. G. Davis, Conversion of cysteine into dehydroalanine enables access to synthetic histones bearing diverse post-translational modifications. *Angew. Chem. Int. Ed.* **51**, 1835–1839 (2012). doi: 10.1002/anie.201106432; pmid: 22247073
37. W. M. Garrison, Reaction mechanisms in the radiolysis of peptides, polypeptides, and proteins. *Chem. Rev.* **87**, 381–398 (1987). doi: 10.1021/cr00078a006
38. G. Milazzo, S. Caroli, V. K. Sharma, *Tables of Standard Electrode Potentials* (Wiley, 1978).
39. A. J. Bard, R. Parsons, J. Jordan, *Standard Potentials in Aqueous Solutions* (Dekker, 1985).
40. J. T. Groves, K. W. Ma, Dehalogenations with sodium borohydride. Evidence for a free radical reaction. *J. Am. Chem. Soc.* **96**, 6527–6529 (1974). doi: 10.1021/ja00827a053
41. E. J. Corey, J. W. Suggs, Method for catalytic dehalogenations via trialkyltin hydrides. *J. Org. Chem.* **40**, 2554–2555 (1975). doi: 10.1021/jo00905a039
42. G. Stork, N. H. Baine, Cyclization of vinyl radicals: A versatile method for the construction of five- and six-membered rings. *J. Am. Chem. Soc.* **104**, 2321–2323 (1982). doi: 10.1021/ja00372a042
43. E. C. Ashby, R. N. DePriest, A. B. Goel, B. Wenderoth, T. N. Pham, Occurrence of electron transfer in the reduction of organic halides by lithium aluminum hydride and aluminum hydride. *J. Org. Chem.* **49**, 3545–3556 (1984). doi: 10.1021/jo00193a019
44. J. A. Barltrop, D. Bradbury, Chain photoreaction of sodium borohydride with halogenated aromatic hydrocarbons. Initiation by aryl radicals. *J. Am. Chem. Soc.* **95**, 5085–5086 (1973). doi: 10.1021/ja00796a071
45. J. J. Liu, R. Horst, V. Katritch, R. C. Stevens, K. Wüthrich, Biased signaling pathways in β2-adrenergic receptor characterized by $^{19}$F-NMR. *Science* **335**, 1106–1110 (2012). doi: 10.1126/science.1215802; pmid: 22267580
46. G. S. Kubelka, J. Kubelka, Site-specific thermodynamic stability and unfolding of a de novo designed protein structural motif mapped by $^{13}$C isotopically edited IR spectroscopy. *J. Am. Chem. Soc.* **136**, 6037–6048 (2014). doi: 10.1021/ja500918k; pmid: 24684597
47. V. S. Lelyveld, A. Björkbom, E. M. Ransey, P. Sliz, J. W. Szostak, Pinpointing RNA-protein cross-links with site-specific stable

isotope-labeled oligonucleotides. *J. Am. Chem. Soc.* **137**, 15378–15381 (2015). doi: 10.1021/jacs.5b10596; pmid: 26583201
48. N. C. Yoder, K. Kumar, Fluorinated amino acids in protein design and engineering. *Chem. Soc. Rev.* **31**, 335–341 (2002). doi: 10.1039/b201097f; pmid: 12491748
49. C. Martin, Y. Zhang, The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Biol.* **6**, 838–849 (2005). doi: 10.1038/nrm1761; pmid: 16261189
50. M. T. Bedford, S. G. Clarke, Protein arginine methylation in mammals: Who, what, and why. *Mol. Cell* **33**, 1–13 (2009). doi: 10.1016/j.molcel.2008.12.013; pmid: 19150423
51. P. Tessarz *et al*., Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. *Nature* **505**, 564–568 (2014). doi: 10.1038/nature12819; pmid: 24352239
52. H. Jakubowski, Misacylation of tRNALys with noncognate amino acids by lysyl-tRNA synthetase. *Biochemistry* **38**, 8088–8093 (1999). doi: 10.1021/bi990629i; pmid: 10387054
53. D. P. Nguyen, M. M. Garcia Alai, P. B. Kapadnis, H. Neumann, J. W. Chin, Genetically encoding N(ϵ)-methyl-L-lysine in recombinant histones. *J. Am. Chem. Soc.* **131**, 14194–14195 (2009). doi: 10.1021/ja906603s; pmid: 19772323
54. H. I. Zgurskaya, H. Nikaido, AcrA is a highly asymmetric protein capable of spanning the periplasm. *J. Mol. Biol.* **285**, 409–420 (1999). doi: 10.1006/jmbi.1998.2313; pmid: 9878415
55. J. C. Lee *et al*., A protein kinase involved in the regulation of inflammatory cytokine biosynthesis. *Nature* **372**, 739–746 (1994). doi: 10.1038/372739a0; pmid: 7997261
56. A. Desmyter *et al*., Crystal structure of a camel single-domain VH antibody fragment in complex with lysozyme. *Nat. Struct. Biol.* **3**, 803–811 (1996). doi: 10.1038/nsb0996-803; pmid: 8784355
57. C. E. Creutz, The annexins and exocytosis. *Science* **258**, 924–931 (1992). doi: 10.1126/science.1439804; pmid: 1439804
58. M. W. Vetting, S. S. Hegde, K. Z. Hazleton, J. S. Blanchard, Structural characterization of the fusion of two pentapeptide repeat proteins, Np275 and Np276, from *Nostoc punctiforme*: Resurrection of an ancestral protein. *Protein Sci.* **16**, 755–760 (2007). doi: 10.1110/ps.062637707; pmid: 17384236
59. D. P. Gamblin, E. M. Scanlan, B. G. Davis, Glycoprotein synthesis: An update. *Chem. Rev.* **109**, 131–163 (2009). doi: 10.1021/cr078291i; pmid: 19093879
60. L.-X. Wang, B. G. Davis, Realizing the promise of chemical glycobiology. *Chem. Sci.* **4**, 3381–3394 (2013). doi: 10.1039/c3sc50877c; pmid: 23914294
61. F. Altmann, S. Schweizer, C. Weber, Kinetic comparison of peptide: N-glycosidases F and A reveals several differences in substrate specificity. *Glycoconj. J.* **12**, 84–93 (1995). doi: 10.1007/BF00731873; pmid: 7540902
62. A. L. Tarentino, C. M. Gómez, T. H. Plummer Jr., Deglycosylation of asparagine-linked glycans by peptide:N-glycosidase F. *Biochemistry* **24**, 4665–4671 (1985). doi: 10.1021/bi00338a028; pmid: 4063349
63. S. Hardivillé, G. W. Hart, Nutrient regulation of signaling, transcription, and cell physiology by O-GlcNAcylation. *Cell Metab.* **20**, 208–213 (2014). doi: 10.1016/j.cmet.2014.07.014; pmid: 25100062
64. Y. Zhu, X. Shan, S. A. Yuzwa, D. J. Vocadlo, The emerging link between O-GlcNAc and Alzheimer disease. *J. Biol. Chem.* **289**, 34472–34481 (2014). doi: 10.1074/jbc.R114.601351; pmid: 25336656
65. C. Slawson, G. W. Hart, O-GlcNAc signalling: Implications for cancer cell biology. *Nat. Rev. Cancer* **11**, 678–684 (2011). doi: 10.1038/nrc3114; pmid: 21850036
66. R. Hurtado-Guerrero, H. C. Dorfmueller, D. M. F. van Aalten, Molecular mechanisms of O-GlcNAcylation. *Curr. Opin. Struct. Biol.* **18**, 551–557 (2008). doi: 10.1016/j.sbi.2008.09.005; pmid: 18822376
67. J. R. Bischoff *et al*., A homologue of *Drosophila* aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J.* **17**, 3052–3065 (1998). doi: 10.1093/emboj/17.11.3052; pmid: 9606188
68. J.-Y. Hsu *et al*., Mitotic phosphorylation of histone H3 is governed by Ipl1/aurora kinase and Glc7/PP1 phosphatase in budding yeast and nematodes. *Cell* **102**, 279–291 (2000). doi: 10.1016/S0092-8674(00)00034-9; pmid: 10975519
69. R. Klingberg *et al*., Analysis of phosphorylation-dependent protein-protein interactions of histone H3. *ACS Chem. Biol.* **10**, 138–145 (2015). doi: 10.1021/cb500563n; pmid: 25330109
70. D. T. Rogerson *et al*., Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. *Nat. Chem. Biol.* **11**, 496–503 (2015). doi: 10.1038/nchembio.1823; pmid: 26030730
71. W. Zheng *et al*., Cellular stability of serotonin N-acetyltransferase conferred by phosphonodifluoromethylene alanine (Pfa) substitution for Ser-205. *J. Biol. Chem.* **280**, 10462–10467 (2005). doi: 10.1074/jbc.M412283200; pmid: 15632116
72. N. Kunowska, M. Rotival, L. Yu, J. Choudhary, N. Dillon, Identification of protein complexes that bind to histone H3 combinatorial modifications using super-SILAC and weighted correlation network analysis. *Nucleic Acids Res.* **43**, 1418–1432 (2015). doi: 10.1093/nar/gku1350; pmid: 25605797
73. G. M. Blackburn, Phosphonates as analogues of biological phosphates. *Chem. Ind.* 134–138 (1981).
74. G. M. Blackburn, D. E. Kent, F. Kolkmann, The synthesis and metal binding characteristics of novel, isopolar phosphonate analogues of nucleotides. *J. Chem. Soc., Perkin Trans. 1* **1984**, 1119–1125 (1984). doi: 10.1039/p19840001119
75. D. J. Burton, Z.-Y. Yang, W. Qiu, Fluorinated ylides and related compounds. *Chem. Rev.* **96**, 1641–1716 (1996). doi: 10.1021/cr941140s; pmid: 11848807
76. K. K. Biggar, S. S. Li, Non-histone protein methylation as a regulator of cellular signalling and function. *Nat. Rev. Mol. Cell Biol.* **16**, 5–17 (2015). doi: 10.1038/nrm3915; pmid: 25491103
77. W. Fischle *et al*., Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **17**, 1870–1881 (2003). doi: 10.1101/gad.1110503; pmid: 12897054
78. C. R. Vakoc, S. A. Mandat, B. A. Olenchock, G. A. Blobel, Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin. *Mol. Cell* **19**, 381–391 (2005). doi: 10.1016/j.molcel.2005.06.011; pmid: 16061184
79. J.-F. Couture, E. Collazo, P. A. Ortiz-Tello, J. S. Brunzelle, R. C. Trievel, Specificity and mechanism of JMJD2A, a trimethyllysine-specific histone demethylase. *Nat. Struct. Mol. Biol.* **14**, 689–695 (2007). doi: 10.1038/nsmb1273; pmid: 17589523
80. S. S. Ng *et al*., Crystal structures of histone demethylase JMJD2A reveal basis for substrate specificity. *Nature* **448**, 87–91 (2007). doi: 10.1038/nature05971; pmid: 17589501
81. S. Daujat *et al*., Crosstalk between CARM1 methylation and CBP acetylation on histone H3. *Curr. Biol.* **12**, 2090–2097 (2002). doi: 10.1016/S0960-9822(02)01387-8; pmid: 12498683
82. S. Abhiman, L. M. Iyer, L. Aravind, BEN: A novel domain in chromatin factors and DNA viral proteins. *Bioinformatics* **24**, 458–461 (2008). doi: 10.1093/bioinformatics/btn007; pmid: 18203771
83. A. Khan *et al*., BEND3 represses rDNA transcription by stabilizing a NoRC component via USP21 deubiquitinase. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8338–8343 (2015). doi: 10.1073/pnas.1424705112; pmid: 26100909
84. N. Saksouk *et al*., Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. *Mol. Cell* **56**, 580–594 (2014). doi: 10.1016/j.molcel.2014.10.001; pmid: 25457167
85. C. Ollivier, P. Renaud, Organoboranes as a source of radicals. *Chem. Rev.* **101**, 3415–3434 (2001). doi: 10.1021/cr010001p; pmid: 11840989
86. T. Kawamoto, I. Ryu, Radical reactions of borohydrides. *Org. Biomol. Chem.* **12**, 9733–9742 (2014). doi: 10.1039/C4OB01784F; pmid: 25349957
87. A. K. Antonczak, Z. Simova, E. M. Tippmann, A critical examination of *Escherichia coli* esterase activity. *J. Biol. Chem.* **284**, 28795–28800 (2009). doi: 10.1074/jbc.M109.027409; pmid: 19666472
88. C. Lizak *et al*., Unexpected reactivity and mechanism of carboxamide activation in bacterial N-linked protein glycosylation. *Nat. Commun.* **4**, 2627 (2013). doi: 10.1038/ncomms3627; pmid: 24149797
89. M. Kurogochi *et al*., Glycoengineered monoclonal antibodies with homogeneous glycan (M3, G0, G2, and A2) using a chemoenzymatic approach have different affinities for FcγRIIIa and variable antibody-dependent cellular cytotoxicity activities. *PLOS ONE* **10**, e0132848 (2015). doi: 10.1371/journal.pone.0132848; pmid: 26200113
90. B. G. Davis, in *Biotherapeutics: Recent Developments Using Chemical Biology*, L. H. Jones, A. J. McKnight, Eds. (Royal Society of Chemistry, 2013), pp. 130–144.

## ACKNOWLEDGMENTS

Supported by the Felix Foundation (R.R.), the Croucher Foundation (W.L.N.), the Rutherford Foundation (T.H.W., S.F.), Cancer Research UK (S.F., L.L.), BBSRC (S.R.G.G., B.J.B., A.J.B.), AstraZeneca (S.R.G.G.), UCB (B.J.B.), EPSRC (O.D.C., L.L., M.W.S., M.E.E.), and a Royal Society Wolfson Research Merit Award (B.G.D.). B.G.D. is a member of the Catalent Biologics Scientific Advisory Board. A pET12a PAPI plasmid encoding the gene for annexin V (originally deposited by J. Tait) was obtained from Addgene under a material transfer agreement. We thank the Wellcome Trust for funding NMR equipment, and W. Myers, T. Parsons, J. Wickens, H. Hailu, P. Mori, S. Gunnoo, A. Lawson, A. Henry, K. Chooi, T. Fleming for useful discussions and/or technical assistance. B.G.D., J.M.C., and G.J.B. are inventors on patent WO 2009103941 filed by the University of Oxford assigned to Catalent Biologics that covers, in one claim, reaction of the carbon-carbon double bond of dehydroalanine residues reacted with an organohalide in the presence of elemental zinc. Data reported in the paper are presented in the supplementary materials (see pdf and associated xlsx spreadsheet detailing affinity proteomic data; table S3). Associated data are also archived in the University of Oxford data archive (ORA-data) with accession doi: 10.5287/bodleian:dp1ypwOB6. Author contributions: T.H.W., B.J.B., J.M.C., G.J.B., R.W., W.L.N., S.F., M.R.J.V., A.P., O.D.C., M.E.E., S.R.G.G., L.L., and M.W.S. conceived, designed, and/or performed protein mutagenesis experiments; T.H.W., B.J.B., J.M.C., and S.G. performed model experiments; T.H.W., B.J.B., W.L.N., M.R.J.V., R.R., and M.W.S. prepared proteins; T.H.W., B.J.B., W.L.N., T.D.W.C., R.R., M.R.J.V., M.E.E., M.-L.T., A.J.B., and B.K. physically characterized protein products; T.H.W., B.J.B., W.L.N., S.F., M.K., R.R., and S.M. biochemically characterized protein products; T.H.W., B.J.B., J.M.C., G.J.B., R.W., W.L.N., R.R., M.-L.T., A.J.B., B.K., S.M., and B.G.D. analyzed data; and T.H.W., J.M.C., and B.G.D. wrote the paper. All authors read and commented on the paper.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/aag1465/suppl/DC1
Materials and Methods
Figs. S1 to S35
Tables S1 to S3
References (*91*–*104*)

19 May 2016; accepted 12 September 2016
Published online 22 September 2016
10.1126/science.aag1465

RESEARCH ARTICLE SUMMARY

SYSTEMS BIOLOGY

# A transcription factor hierarchy defines an environmental stress response network

**Liang Song, Shao-shan Carol Huang, Aaron Wise, Rosa Castanon, Joseph R. Nery, Huaming Chen, Marina Watanabe, Jerushah Thomas, Ziv Bar-Joseph, Joseph R. Ecker***

**INTRODUCTION:** Transcription factors (TFs) are often studied one by one or in clusters of a few related factors. However, the integration and networks of transcriptional changes to response to environmental stresses often involve many related TFs. In many organisms, such as plants, overlapping functions can make it difficult to understand how a biologically relevant end result can be achieved via the complex signaling networks controlled by these TFs. To better understand how the reference plant *Arabidopsis* deals with the stresses incurred by water limitation via the hormone abscisic acid (ABA), we characterized all DNA sequences that bind to the 21 ABA-related TFs in vivo.

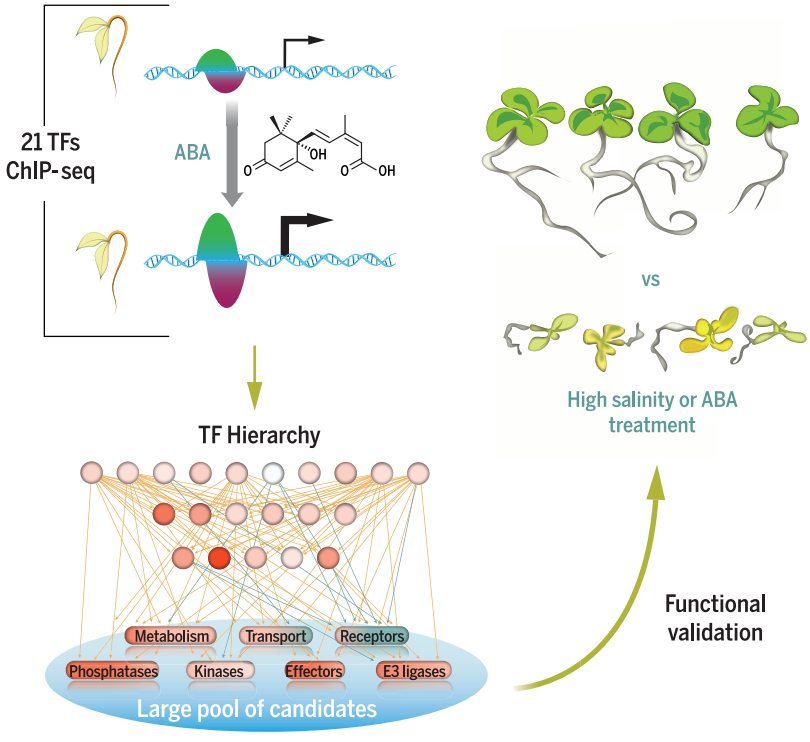
**RATIONALE:** There have been limited systematic studies of stress-responsive TF networks in multicellular organisms. We chose ABA, an essential plant hormone that is required for both development and responses to osmotic stress, as an elicitor to investigate complex gene regulatory networks under stress. Combining differential binding (DB) of 21 ABA-related TFs at a single time point measured by chromatin immunoprecipitation sequencing (ChIP-seq) with differentially expressed genes from a time-series RNA sequencing (RNA-seq) data set, we analyzed the relationship between DB of TFs and differential expression (DE) of target genes, the determinants of DB, and the combinatorial effects of multi-TF binding. These data sets also provide a framework to construct an ABA TF network and to predict genes and cis-regulatory elements important to ABA responses and related environmental stresses.

**RESULTS:** We found that, in general, DNA binding is correlated with transcript and protein levels of TFs. Most TFs in our study are induced by ABA and gain binding sites (termed "peaks") after the hormone treatment. ABA also increases the binding of the TFs at most peaks. However, in some peaks, TF binding may be static or even decrease after ABA exposure, revealing the complexity of locus-specific gene regulation. De novo motif discovery enabled us to identify distinct, primary motifs often centrally localized in the ChIP-seq peaks for most TFs. However, it is not uncommon to find motifs, such as the G-box, that are shared by peaks from multiple TFs and may contribute to binding dynamics at these sites. DB of multiple TFs is a robust predictor of both the DE and ABA-related functions of the target genes. Using the DB and DE data, we constructed a network of TFs and canonical ABA pathway genes and demonstrated a regulatory hierarchy of TFs and extensive feedback of ABA responses. On the basis of a "guilt-by-association" paradigm, we further predicted genes whose functions were previously not linked to ABA responses, and we thus functionally characterized a new family of transcriptional regulators.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aag1550

**CONCLUSION:** These data sets will provide the plant community with a roadmap of ABA-elicited transcriptional regulation by 21 ABA-related TFs. We propose that dynamic, multi-TF binding could be a criterion for prioritizing the characterization of TF binding events, cis-regulatory elements, and functionally unknown genes in both plants and other species. In our proof-of-principle experiments, ectopic expression of the transcriptional regulators ranked highly in our model results in altered sensitivity to both ABA and high salinity. Together with the fact that our modeling recovered genes related to seed development and osmotic stresses, we believe such predictions are likely applicable to a broad range of developmental stages and osmotic stresses. ■




**Transcriptional landscape of the ABA response.** ABA response pathway gene targets were identified by large-scale ChIP-seq and time-series RNA-seq experiments. A network model was built to reveal the hierarchy of TFs and the impact of multi-TF dynamic binding on gene expression. A new family of transcriptional regulators was predicted by the model and was functionally tested to evaluate the role of these regulators in osmotic stress in plants.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: ecker@salk.edu
Cite the article as: L. Song *et al.*, *Science* **354**, aag1550 (2016). DOI: 10.1126/science.aag1550

RESEARCH ARTICLE

SYSTEMS BIOLOGY

# A transcription factor hierarchy defines an environmental stress response network

**Liang Song,[1] Shao-shan Carol Huang,[1] Aaron Wise,[2] Rosa Castanon,[3] Joseph R. Nery,[3] Huaming Chen,[3] Marina Watanabe,[1] Jerushah Thomas,[1] Ziv Bar-Joseph,[2] Joseph R. Ecker[1,3,4]***

**Environmental stresses are universally encountered by microbes, plants, and animals. Yet systematic studies of stress-responsive transcription factor (TF) networks in multicellular organisms have been limited. The phytohormone abscisic acid (ABA) influences the expression of thousands of genes, allowing us to characterize complex stress-responsive regulatory networks. Using chromatin immunoprecipitation sequencing, we identified genome-wide targets of 21 ABA-related TFs to construct a comprehensive regulatory network in *Arabidopsis thaliana*. Determinants of dynamic TF binding and a hierarchy among TFs were defined, illuminating the relationship between differential gene expression patterns and ABA pathway feedback regulation. By extrapolating regulatory characteristics of observed canonical ABA pathway components, we identified a new family of transcriptional regulators modulating ABA and salt responsiveness and demonstrated their utility to modulate plant resilience to osmotic stress.**

Transcription is a key step in gene expression. There have been concerted efforts to map functional elements in human, fly, and worm (*1*–*3*), including a large number of cis-regulatory elements identified by profiling transcription factor (TF) binding using chromatin immunoprecipitation sequencing (ChIP-seq). One area that remains largely unexplored is how a stimulus modulates TF binding and subsequent transcriptome changes. Furthermore, compared with studies in animals, very few comprehensive in vivo TF binding data sets are available for the Plantae kingdom. To begin to address this knowledge gap, we generated more than 100 ChIP-seq and time-series RNA sequencing (RNA-seq) data sets to characterize a stimulus-influenced transcriptional network and map functional cis-regulatory elements in the reference plant *Arabidopsis thaliana*, with a focus on the phytohormone abscisic acid (ABA). The response to ABA provides a versatile model for the examination of stimulus-influenced transcriptional regulation. ABA triggers differential expression (DE) of thousands of genes and TFs, providing a robust response that enables modeling of complex gene regulatory networks. Moreover, ABA's role in a variety of plant processes is important for both fundamental biology and agriculture (*4*, *5*).

[1]Plant Biology Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [2]Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [3]Genomic Analysis Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [4]Howard Hughes Medical Institute (HHMI), Salk Institute for Biological Studies, La Jolla, CA 92037, USA.
*Corresponding author. Email: ecker@salk.edu

Abscisic acid plays a pivotal role in optimizing water use in plants and is required for both seed development and responses to multiple environmental stresses, such as drought and high salinity. In *A. thaliana,* ABA is recognized by the PYR/PYL/RCAR receptor proteins (*6*–*8*). Binding of ABA triggers the interaction of PYR/PYL/RCARs with group A PP2C protein phosphatases and derepresses the SnRK2 protein kinases (*7*, *9*). SnRK2s subsequently activate substrates such as TFs and elicit ABA responses (*10*, *11*). Whereas many TFs are known to contribute to the ABA responses (*12*), little is known about their target genes and the way these targets are combinatorially regulated. In vitro approaches, including the recently described *Arabidopsis* cistrome data set, have enabled identification of DNA motifs and inference of the associated TFs (*13*–*17*). However, accurate predictions are still challenging due to many large, multimember TF families. Furthermore, it is difficult to establish a direct link between TF binding and transcriptome changes or to address the dynamics of TF regulation through in vitro assays. Therefore, we applied ChIP-seq to unambiguously discover TF targets, examine stimulus-influenced TF binding dynamics, and link these findings to subsequent transcriptome changes.

## ChIP-seq analyses of ABA responses

We profiled the genome-wide binding dynamics of a diverse collection of TFs using ChIP-seq to develop an in planta ABA transcriptional regulatory network. We first surveyed ABA-responsive transcripts in *A. thaliana* by generating strand-specific RNA-seq libraries from 3-day-old etiolated whole seedlings treated with either 10 μM (±)-ABA or an ethanol-containing mock for 1, 4, 8, 12, 24, 36, and 60 hours (Fig. 1A). Among 18,310 expressed genes, 3061 are differentially expressed [false discovery rate (FDR) < 0.01] (table S1) (*18*) for at least one time point. One hour of ABA treatment leads to moderate DE of many genes, and most transcriptional responses plateau after 8 hours (Fig. 1B and fig. S1). On the basis of the gene expression data, we performed ChIP-seq experiments 4 hours after dosing with ABA. We selected TFs based on responsiveness to ABA and published evidence, aiming to provide a comprehensive representation of TF families (fig. S2 and tables S2 and S3). In general, highly expressed and responsive TFs were chosen in each representative TF family because in the context of an in planta experimental framework, the impact of these TFs on gene expression can be more effectively investigated compared with that of their weakly expressed homologs. All TF genes were epitope-tagged by a recombineering-based approach (*19*), mostly with large DNA transformable artificial chromosomes, allowing the TFs to be expressed under their native promoters and genomic context (table S3). The final data set consisted of 122 ChIP-seq experiments involving 21 TFs from 11 families, including mock- and ABA-treated conditions.

Overall, the number of binding sites ("peaks") varies greatly across TFs and between treatments (Fig. 1C). Most TFs gain binding sites across the genome after ABA treatment (Fig. 1C), consistent with the fact that these TFs are induced by ABA at both the transcript and protein level (Fig. 1D and figs. S2 and S3). As exemplified by *CYP707A1* and *HAI2* (*20*, *21*), two important genes regulating ABA catabolism and signaling, respectively, the dynamic binding of TFs elicited by ABA is often accompanied by altered transcript abundance of the target genes (Fig. 1D). Accounting for binding location and strength, a comparison of the genome-wide binding profiles of these TFs revealed that the TFs are generally grouped by family and known physical interactions (Fig. 1E) (*22*). The binding profiles between NAM, ATAF, and CUC (NAC) and other TF families become more similar after ABA treatment (Fig. 1E, box A versus M), which suggests that ABA prompts coordinated regulation of target genes by these TFs.

## Hormonal effects on TF binding and expression of target genes

We observed substantial changes in TF binding at promoter regions of several known components of the ABA signaling pathway (Fig. 1D), so we investigated whether dynamic binding may predict gene function in the ABA pathway. To quantify hormone-dependent, locus-specific changes of TF binding, we compared ChIP-seq peaks of each TF between ABA- and mock-treated conditions by performing differential binding analysis of the sequencing reads under the peaks (*23*). Three measures of differential binding were calculated for each peak: (i) normalized read count change (RCC) that measures absolute changes of binding, (ii) fold change (FC) that measures relative changes of binding, and (iii) statistically significant differential binding (FDR). Because there are limited

down-regulated binding events in our data set, we focused on up-regulated binding to determine the optimal cutoff of RCC, FC, and FDR to define the relationship between dynamic binding targets and genes involved in the ABA response. We extracted three groups of *A. thaliana* genes (table S4) on the basis of gene ontology (GO) annotation (*24*). Group 1 contains 493 genes involved in the ABA response. Group 2 contains 1452 genes involved in responses to either ABA or other related processes such as water deprivation, osmotic stress, salt stress, cold, seed development, and stomatal movement. Group 3 contains 999 genes involved in responses to other hormones after excluding genes shared with group 2. Three observations emerged from comparing these lists to the TF target gene lists defined by various thresholds on RCC, FC, and FDR. First, when group 1 and 2 genes are used as a reference set, the percentage of TF targets overlapping with the set increases with the number of bound TFs (Fig. 2A, left and middle panels). By contrast, there is very little, if any, increase when group 3 (other hormone genes) is used as the reference (Fig. 2A, right panel). Second, an increase of RCC and FC threshold beyond the top 20% boosted the percentage of target genes involved in ABA-related responses but not the percentage of genes related to other hormones (Fig. 2A, right panel). This improvement is even more obvious for genes targeted by multiple TFs. Last, FDR thresholds of 0.1 and 0.2 show few differences across all analyses. These results support the premise that dynamic binding by multiple TFs is an important

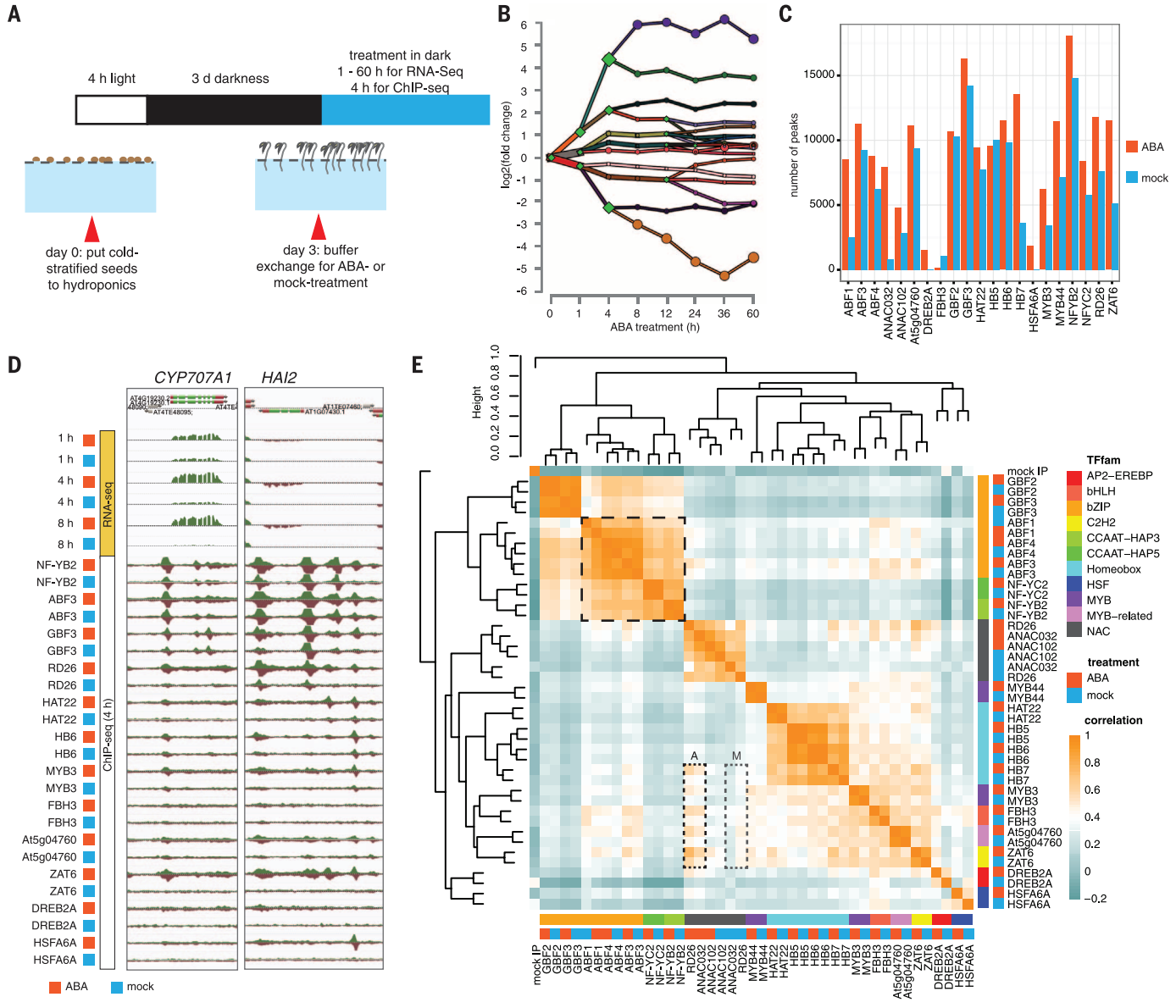



**Fig. 1. TF identity and hormone treatment determine genome-wide binding profiles.** (**A**) Growing *A. thaliana* in hydroponics allows convenient buffer exchange for hormone treatment. (**B**) DREM-reconstructed RNA expression paths 60 hours after exposure to ABA. Each path corresponds to a set of genes that are coexpressed. Split nodes (green diamonds) represent a temporal event where a group of genes coexpressed up to that point diverges in expression, most likely due to regulatory events. Most splits are observed up to and including the 4-hour time point, indicating that the majority of regulatory events occur during the first 4 hours. (**C**) The number of ChIP-seq peaks varies greatly between TFs and treatment conditions. (**D**) ABA mediated differential gene expression and altered dynamics of TF binding, as exemplified by the *CYP707A1* and *HAI2* genes. (**E**) Comparison of binding correlations based on normalized ChIP-seq read counts near binding sites shows that TFs from the same family often have similar binding profiles. TF-TF interaction (bZIPs and NF-Y, black dashed box) and hormone treatment (RD26 and ANAC032, dotted boxes A and M for ABA- and mock-treatment, respectively) also contribute to binding profile similarities between TFs.

feature to specifically recover genes involved in ABA-related responses. We thus chose the top 20% of RCCs and FCs, as well as FDR = 0.1, as the cutoffs for follow-up analyses. As shown in Fig. 2B and fig. S4, peaks passing all three thresholds were designated as top-ranked up-regulated ("top up") or down-regulated ("top down"), whereas those failing all thresholds were designated as static; all remaining peaks were classified as moderately up-regulated ("moderate up") or down-regulated ("moderate down"). For all tested TFs except FBH3 and HB5, peaks tend to gain binding instead of maintaining or losing binding after ABA treatment.

We then explored the relationship between dynamic TF binding triggered by ABA treatment and gene expression. The Dynamic Regulatory Event Miner (DREM) (*25*) reports 11 paths of DE genes for the first 8 hours of ABA treatment (Fig. 2C). As shown in fig. S5, combining DREM with DNA motifs from PBM (protein binding microarrays), AGRIS (Arabidopsis Gene Regulatory Information Server), and DAP-seq (DNA affinity purification sequencing) databases recovered few TFs in our data set, likely because of a low overlap of these TFs with the databases (*13*, *15*, *26*). The DREM model identified TFs from all chromatin immunoprecipitated families except for CCAAT-HAP3 and CCAAT-HAP5, which do not bind DNA as a monomer in in vitro assays (*27*). In addition, although TF binding was examined at a single time point, we observed a positive correlation between the number of dynamically bound TFs and the magnitude of DE across all time points (fig. S6), which suggests that TF binding data at 4 hours after dosing with ABA can explain a broad temporal span of gene expression. ABA-related GO terms such as "seed development" and "response to salt/osmotic stress/water deprivation" were enriched in up-regulated genes, whereas a few growth-related terms such as "response to auxin stimulus" and "cell wall organization" were enriched in down-regulated genes (Fig. 2D). We observed a distinct distribution of dynamic binding category across DREM paths (Fig. 2E, fig. S1, and table S5). The extent of multi-TF dynamic binding is associated with the magnitude of differential gene expression. For example, highly up-regulated genes are often targeted by multiple TFs through top-up binding. Moderately up-regulated genes are more commonly targeted by multiple TFs through moderate-up binding. Down-regulated genes are rarely associated with top-up binding. Instead, these genes are predominantly associated with either static binding by multiple TFs or down-regulated TF binding. These data suggest that DE at the whole-seedling level is often subject to a combinatorial regulation by multiple TFs. As an independent validation, we built a regression model of DE using peak signals in ABA- and mock-treated conditions as features without hard thresholds for the level of dynamic binding. The resulting model reveals that multiple TF binding features such as ZAT6, NF-YB2, and ABF factors in both ABA- and mock-treated conditions contribute to DE of target genes (Fig. 2F and fig. S7).

## Determinants of differential TF binding

With the discovery of tens of thousands of differential binding events, we were interested in whether we could identify features that may predict

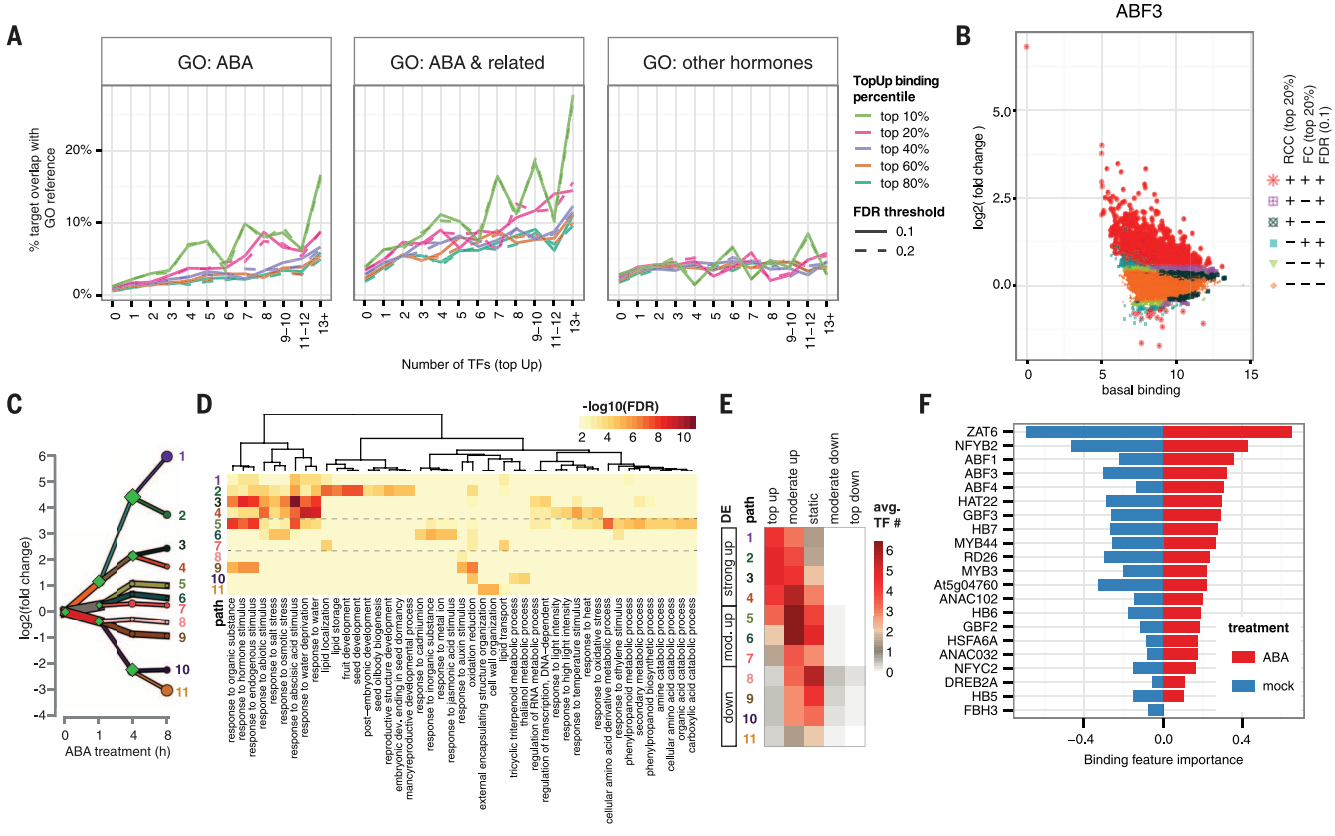



**Fig. 2. Dynamic TF binding triggered by ABA treatment correlates with gene function and expression.** (**A**) Genes targeted by higher numbers of TFs with dynamic binding events (*x* axis) have higher percentages of overlap (*y* axis) with genes annotated with ABA and ABA-related GO terms, but not with GO terms specific to other hormones. This positive correlation is stronger for target genes associated with stronger dynamics (different-colored lines). (**B**) Hormone-dependent, locus-specific TF binding dynamics vary greatly across the genome. Log2 (fold change) of TF binding upon ABA treatment (*y* axis) was plotted against basal binding measured as log2 (normalized read counts) under mock treatment (*x* axis). Peaks were classified by three criteria: read count change (RCC, within top 20%), fold change (FC, within top 20%), and DiffBind FDR (<0.1). Peaks satisfying all three criteria were designated as "top dynamic" (+++), and those failing all three were designated as "static" (– – –). The remaining peaks were designated as "moderately dynamic." (**C**) DREM analysis shows 11 paths of DE genes after 8 hours of ABA treatment. (**D**) Each DREM path is enriched for specific GO terms. (**E**) The level of DE is correlated with multi-TF dynamic binding. (**F**) Ridge regression model for DE at 4 hours, using binding strength in both ABA- and mock-treated conditions and including contributions from multiple TFs in both conditions. Regression coefficients are plotted as relative importance of the binding features.

binding dynamics. First, we performed motif discovery by MEME-ChIP (*28*) to identify enriched motifs of all 21 chromatin immunoprecipitated TFs from the strongest 600 peaks after either ABA or mock hormone treatment. A complete collection of the motifs is available at www.ABAtf.net. To investigate whether there are additional motifs that correlated with TF binding dynamics, we also performed motif discovery on both dynamic and static peaks for a handful of TFs. These factors (NF-YB2, ABF1, ABF4, FBH3, MYB3, RD26, ZAT6, and HB7) were selected to represent a variety of TF families. Pairwise comparison of primary and secondary motifs discovered from dynamic and static peaks across the selected TFs revealed three major clusters (Fig. 3A). Cluster 1 motifs are composed of $(AG)_n$ repeats. Cluster 2 motifs contain a (A/G)G(A/C)CC(A/C) consensus sequence, whereas cluster 3 comprises G-box motifs (Fig. 3B). To examine the contributors to binding dynamics, we used linear regression to model the FC of binding as a function of variables, including basal binding of the TF (under mock treatment) and the number of occurrences (counts) of a set of nonredundant sequence features that capture the diversity of motifs. These sequence features were selected from major clusters of all motifs found in the strongest 600 peaks (motifs of clusters A to I) and the dynamic and static peaks (motifs of clusters 1 to 3) (fig. S8). Examination of the regression coefficient *P* values (Fig. 3B) suggests that the primary motifs of ABF (which also represent cluster 3 motifs) and ANAC TFs are associated with enhanced dynamic binding, whereas basal binding and cluster 2 motifs are associated with a negative effect on binding dynamics for a broad range of TFs (Fig. 3B and table S6). Including cluster 3 or cluster 1 motifs in the regression results increases the explained variability by up to 20% (Fig. 3B). To visualize the effect of the cluster 3 G-box motif and the cluster 2 motif at the resolution of individual peaks, we plotted basal binding of TFs quantified by normalized read count against log2 FC of binding after ABA treatment and assigned a color to individual binding events on the basis of the count of motifs in the same peak (Fig. 3C). The proportion of peaks containing the cluster 3 motif increases along with log2

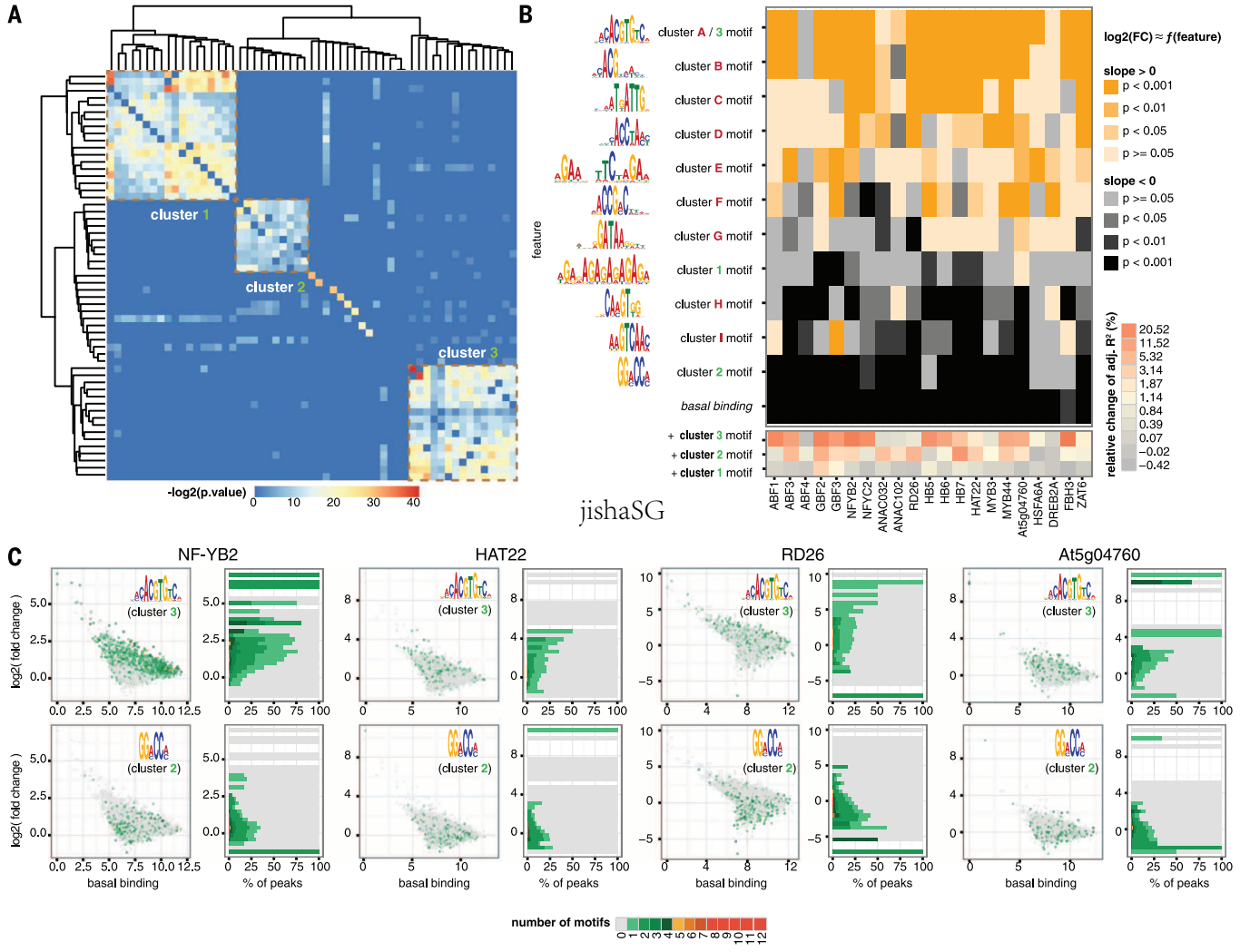



**Fig. 3. Determinants of TF binding dynamics.** (**A**) Hierarchical clustering of motifs enriched in dynamic and static peaks revealed three clusters. Each entry in the distance matrix is −log2(*P* value) of motif similarity reported by Tomtom (*28*). (**B**) Linear regression of differential binding using basal binding and nonredundant sequence features identified positive and negative determinants of dynamic TF binding. Heatmap colors map to two-tailed *t* test *P* values on the regression coefficients for the null hypothesis that the coefficient is zero. The sequence features were selected from motifs enriched in the strongest peaks in ABA- and mock-treated conditions, as well as dynamic and static peaks (fig. S8). (**C**) Scatterplots show basal binding of TFs quantified by normalized read count in the peak (*x* axis) against log2(FC) of TF binding after ABA treatment (*y* axis), with the color of each dot mapped to the number of indicated motifs in the same peak. The occurrence of cluster 3 and 2 motifs over the distributions of log2(FC) of binding is shown in the histograms, with the same color code as the scatter plot. The proportion of peaks containing the cluster 3 motif increases along with log2(FC) of TF binding for the indicated TFs, whereas the proportion of peaks containing the cluster 2 motif is negatively correlated with log2(FC) of TF binding.

FC of binding, whereas the proportion of peaks containing the cluster 2 common motif is negatively correlated with log2 FC of binding. These data suggest that the binding of a TF to the cluster 3 motif (likely the ABFs) and the binding of an unknown family of TFs to the cluster 2 motif regulate the binding dynamics of neighboring TFs positively and negatively, respectively.

### Construction of an ABA response network

To confirm that dynamic binding is more robust than total binding in predicting gene expression and genes involved in ABA and related responses, we compared the expression and functional composition of genes grouped by the number of targeting TFs through either any type of binding or top-up binding (Fig. 4A). Representation of both genes associated with ABA-related GO terms and ABA up-regulated genes increases more rapidly with the increasing number of TFs that have top-up binding. Therefore, we decided to use top-ranked dynamic TF binding triggered by ABA treatment to demonstrate the wiring of this ABA network using the core ABA metabolic and signaling genes and to calculate the hierarchical height of TFs in the network (Fig. 4, B and C, and fig. S9A). TFs in the network are organized into three tiers by their hierarchical height (Fig. 4C and fig. S9A). The level of DE of lower-tier TFs is often amplified compared with that of upper-tier TFs, resulting in greater changes in binding dynamics, likely because of greater protein accumulation (figs. S3, S4, and S9A). Negative regulators of ABA response—including genes encoding ABA catabolic enzymes, protein phosphatase 2Cs, and E3 ligases—are often induced by ABA and are heavily targeted by multiple TFs through highly up-regulated TF binding (Fig. 4C). By contrast, positive regulators of the ABA response can either be up-regulated due to increased TF binding or down-regulated due to reduced TF binding (Fig. 4C). These results point to a transcriptional feedback strategy in the ABA response, presumably to allow rapid restoration of normal growth once stress is lifted. Because some transcriptional responses triggered by ABA are similar to those triggered by natural stresses (fig. S9, A and B) such as high saline conditions, we also expect to see a similar organization of regulatory networks for other osmotic-related stresses.

Extensive targeting by ABA-responsive TFs appears specific to the ABA pathway, as the core ABA genes are targeted by a substantially greater number of TFs through top-up binding than are genes from other plant hormones (fig. S10 and table S7). However, instances of hormone crosstalk can be observed in dynamically targeted DE genes. For example, both *RGA-like 3* (*RGL3*), a master regulator of gibberellin response, and *ACC synthase 2* (*ACS2*), an ethylene biosynthesis gene, were reported to be ABA-responsive (*29*, *30*). We observed that dynamic binding is mainly contributed by the bZIP and NF-Y factors to the promoter of *RGL3* and by a diverse family of TFs to the gene body of *ACS2* (fig. S11). These results demonstrate the utility of these data to pinpoint regulatory regions that might modulate the expression of genes in one hormone response pathway by another.

No GO term besides the ABA-related ones was enriched in DE genes heavily targeted by the 21 TFs through top-up binding. This is partially because more than one-third (12,136/33,601) of the genes in the *A. thaliana* genome still have no information regarding their biological processes (BPs) (Fig. 5A and table S4). On the basis of a "guilt-by-association" paradigm (*31*), we speculate that many BP-unknown genes in Fig. 5A are also involved in ABA responses (table S8). As a proof of principle, we functionally characterized a family in which all members are BP-unknown and differentially expressed in response to ABA. In particular, three members in this family (*At3g48510*, *At5g50360*, and *At5g40790*) are heavily targeted by TFs through top-up binding (Fig. 5, A and B). Little is known about this family except that the proper expression of *At3g48510* relies on core ABA signaling (*32*). In addition, predicted proteins of this family contain no known domains. We generated dexamethasone (DEX)–inducible lines expressing green fluorescent protein (GFP) fusion of the two most heavily targeted genes (*At3g48510* and *At5g50360*). Analysis by RNA-seq showed that a few hundred DE genes were consistently identified from both short-term (4 hours) and long-term (10 days) DEX induction of the two genes (Fig. 5C and table S8). To reflect their regulation and function, these genes were named *Dynamic Influencer of Gene expression 1 (DIG1)* and *DIG2*, and their homologs were named *DIG-likes* (*DILs*). ABA-related GO terms, such as "response to water deprivation," were enriched in *DIG* down-regulated genes (Fig. 5C). Confocal imaging further showed that DIGs were localized to the nucleus (Fig. 5D). We then tested whether the DIGs are transcriptional regulators. ChIP-seq of DEX-inducible *GFP-DIGs* showed that the DIGs bind chromatin. Moreover, stronger binding was observed in the promoter of DIG down-regulated genes than up-regulated ones or non-DE genes (Fig. 5, E and F). De novo

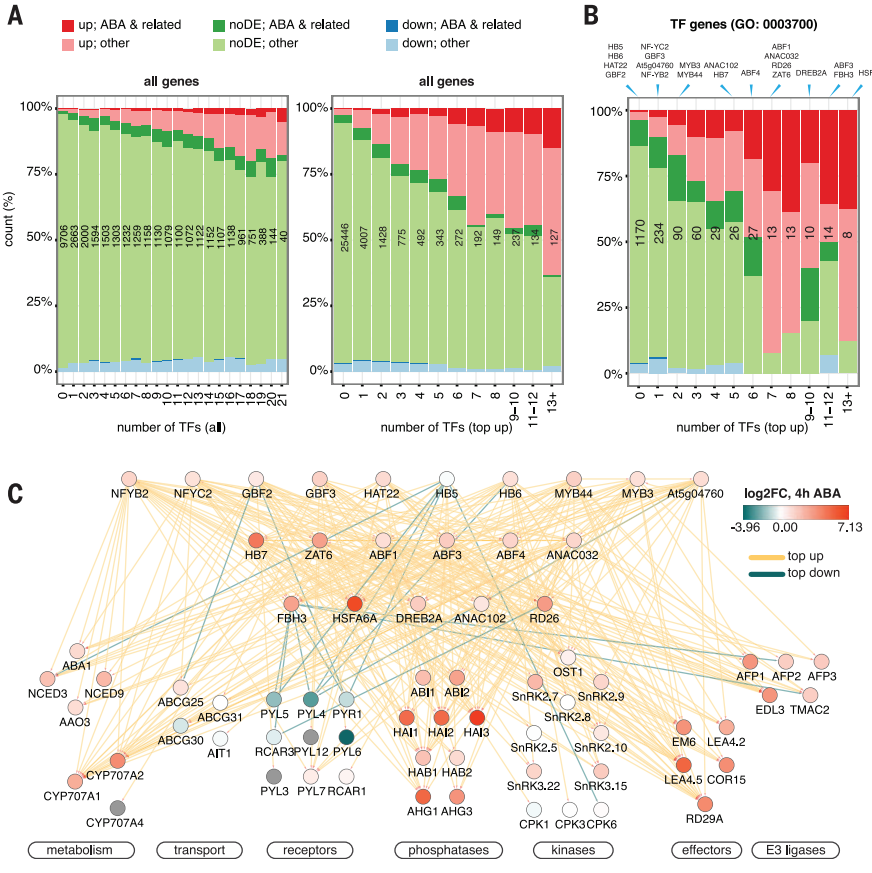



**Fig. 4. TF network integrates expression and connectivity features of genes in the ABA response.** (**A** and **B**) Expression and functional composition of all genes (A) and TF genes (B) are grouped by the number of targeting TFs through either any kind of binding or top-up binding. Compared with any type of binding, top-up binding is a better predictor for both ABA-related BP functions and DE. The number of genes in each bin is shown in black. The bins to which of the TFs included in this study belong are indicated at the top of (B). (**C**) ABA pathway genes are subject to extensive feedback regulations and multi-TF dynamic binding. Chromatin immunoprecipitated TFs are arranged in three tiers by normalized hierarchy height. Target genes are grouped by function. Node color depicts changes of transcript abundance after 4 hours of ABA treatment. Edge color corresponds to TF binding dynamic categories.

motif discovery allowed us to identify a CCAAT$(n)_8$ ABRE motif strongly enriched near the DIG1 binding sites within 1 kb of DIG down-regulated genes. By contrast, either a weaker motif or no similar motif was enriched near DIG binding sites in the corresponding regions of non-DE genes or DIG up-regulated genes (Fig. 5, G and H). Several ABA-responsive or developmental TFs are targeted by DIGs and differentially expressed

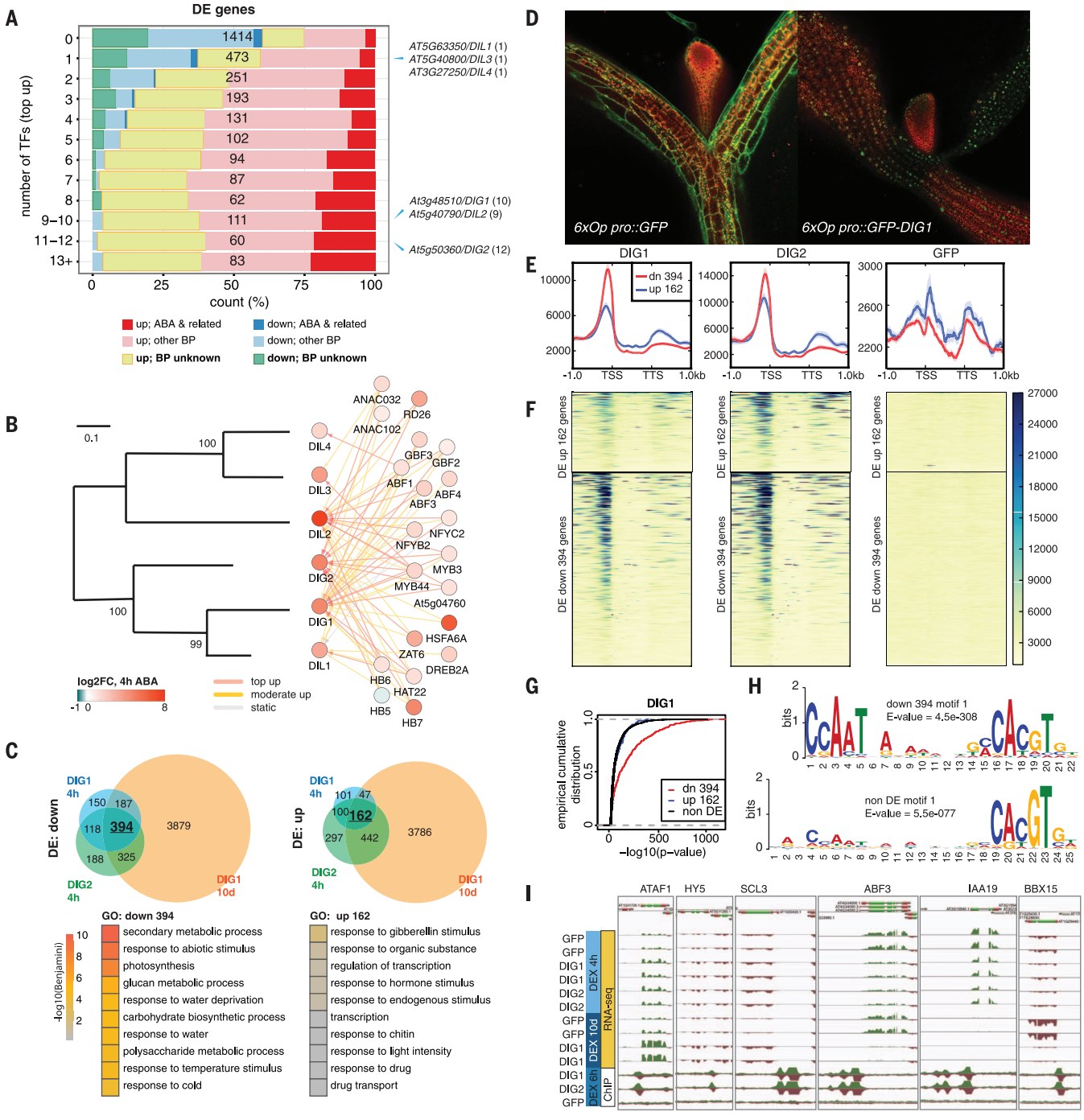



**Fig. 5. Network analysis identifies new transcriptional regulators of ABA response.** (**A**) Expression and functional composition of DE grouped by the number of targeting TFs through top-up binding. The number of genes in each bin is shown in black. The bins to which the *DIG* and *DIL* genes belong are indicated on the right, with the number of targeting TFs shown in parentheses. (**B**) *DIG* and *DILs* genes are regulated by multiple ABA-responsive TFs. (Left) Phylogram of *Arabidopsis* DIG and DIL proteins. (Right) TFs targeting the *DIG* and *DIL* genes. (**C**) DEX-induction of *DIGs* results in DE of stress- and water-related genes. (Top) DE genes by DIGs after the indicated period of DEX treatment. (Bottom) Top GO terms enriched in DIG DE genes. (**D**) Confocal imaging of 9-day-old DEX-treated transgenic seedlings shows that DIG1 is nuclear-localized. (**E** and **F**) Metagene profiles (E) and heatmaps (F) of normalized ChIP-seq read counts surrounding DIG DE genes. Down-regulated genes are often associated with strong DIG binding in their promoters. (**G**) Empirical cumulative distributions of −log10(*P* value) of ChIP-seq peaks of DIG1 show that it binds more strongly to DIG down-regulated genes than to up-regulated or non-DE genes. (**H**) A CCAAT$(n)_8$ ABRE motif is strongly enriched near DIG1 binding sites residing within 1 kb of DIG down-regulated genes. Either a weaker motif or no similar motif is enriched in the corresponding regions of non-DE genes or DIG up-regulated genes. (**I**) Induction of DIGs results in DE of ABA- and developmental-related TFs.

upon the induction of *DIG*s (table S9). Among these, ATAF1, HY5, and ABF3 have been linked to ABA sensitivity (*33*–*35*), whereas HY5, SCL3, and perhaps IAA19 have developmental roles (*34*, *36*, *37*). Sequence analysis revealed that *DIG*s are conserved between monocots and dicots (fig. S11). A remotely related clade of *DIG* contains the gene *Sdr4*, which regulates seed dormancy in rice (*38*) (fig. S12). The *Sdr4* paralog in *Arabidopsis* is also dynamically targeted by multiple ABA-responsive TFs and differentially expressed in response to ABA (table S1). However, the functionally important amino acid residues of *Sdr4* are not conserved in the *DIG*s and their homologs (fig. S13) (*38*). Therefore, genes in the *DIG* and *Sdr4* clades may exert ABA-related functions through distinct mechanisms. Finally, inducible expression of *DIG*s enhances ABA sensitivity, as assayed by cotyledon greening (Fig. 6, A and B) and lateral root growth (Fig. 6D). Similarly, enhanced growth inhibition of *DIG* lines can also be observed after prolonged growth under high-NaCl conditions (Fig. 6, C, E, and F). Combined, our results suggest that DIGs are a family of transcriptional regulators with broad roles that include regulating gene expression that affects ABA sensitivity and salt stress responses.

## Conclusions

We performed a systematic study of a transcriptional network by combining dynamic binding data of 21 TFs and time series RNA-seq data in response to a stimulus by the plant hormone ABA. We found that dynamic TF binding measured at a single time point correlated with the transcriptome changes over a prolonged span of time. Consistent with yeast and animals (*2*, *3*, *39*, *40*), transcription of genes in *Arabidopsis* is often subject to a complex regulation of multiple TFs. We further demonstrated that dynamic binding, especially by multiple TFs, is more functionally relevant than static TF binding in correlation with differential gene expression. We speculate that this is because an expression scheme with coordinated changes in the binding dynamics of multiple TFs would ensure robust responsiveness of target genes to a stimulus. This observation would have a broad application to both plants and other species, including prioritizing studies of (i) TF binding events and cis-regulatory elements and (ii) functionally unknown genes in a pathway.

In plants, studies of transcriptional regulation are often focused on master regulators (*33*, *41*, *42*). Our data confirmed the importance of master regulators in plants. For instance, we showed that ABFs and a physical interactor (NF-YB2) ranked as top contributors to explain gene expression. In addition, the primary binding motif of ABFs also enhances the binding dynamics of many other ABA-responsive TFs. However, more than just the master regulators are required to attain complex transcriptome changes to a stimulus, as many ABA-responsive genes are dynamically targeted by multiple TFs. Therefore, ABA response can be viewed as orchestrated by a handful of master regulators and facilitated by other TFs, where coordinated signaling and regulatory response lead to a rapid elicitation of transcriptome changes.

As indicated by GO annotation, network analyses of this study recovered genes affecting all aspects of ABA-related processes, such as seed development and response to osmotic stresses. In planta ectopic expression of several members of a newly discovered family of transcriptional

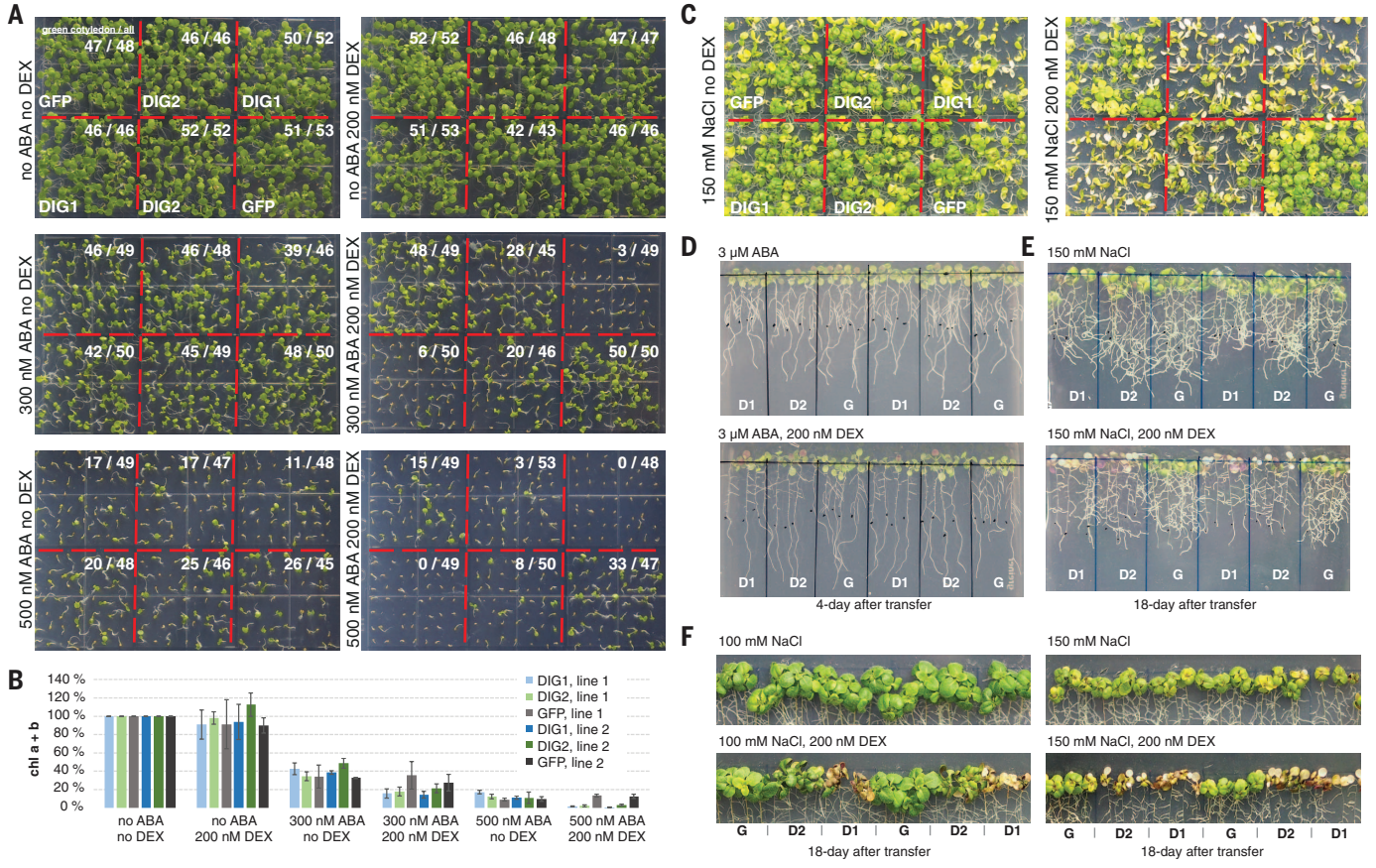



**Fig. 6. *DIG*-inducible lines exhibit enhanced sensitivity to ABA and salt.** (**A** and **B**) ABA-dependent delay of cotyledon greening in 8-day-old seedlings was further amplified upon DEX-mediated induction of *DIG1* and *DIG2* compared with *GFP* control quantified by count of green cotyledons (A) and measurement of relative chlorophyll content (B). In (B), each error bar reflects the 95% confidence interval around the mean estimate calculated from three biological replicates of ~50 8-day-old seedlings. (**C**) NaCl-dependent bleaching was observed in 4-week-old plants upon DEX-mediated induction of *DIG1* and *DIG2*. (**D** and **E**) DEX-mediated induction of *DIG1* (D1) and *DIG2* (D2) resulted in more severe inhibition of lateral root growth than *GFP* (G) control plants on ABA (D) and NaCl (E) plates. (**F**) DEX-mediated induction of *DIG1* and *DIG2* led to overaccumulation of pigments in leaves. In (D) to (F), seedlings were transferred to the indicated plates after being grown on Linsmaier and Skoog plates for 7 days.

regulators also exhibited altered response to both ABA and high salinity. Therefore, although we carried out the experiments using seedlings, the discoveries may be directly applicable to a broader range of development stages and stress scenarios. Emerging technologies to optimize plant water use have been developed based on the in-depth characterization of ABA perception (*43*). Knowledge derived from further studies of the genes uncovered in this study may also prove valuable to global agriculture, possibly enabling new strategies for plants to respond to the challenges of ongoing drought and groundwater depletion in changing environments.

## Materials and methods

### Plant materials

Recombineering lines for the ChIP-seq experiments were generated as previously described (*19*) with minor modifications. A YPet-6xHis-3xFLAG tag and a 3xFLAG-YPet tag were designed for C terminus and N terminus fusion to the TFs of interest (table S3 and fig. S14). To abolish weak dimerization of YPet (*44*), an A206K point mutation was introduced by primers 5′-ATCCTTGAAGAGCTTAGACTGGTAAGA-3′ and 5′-TCTTACCAGTCTAAGCTCTTCAAGGAT-3′. After floral dip of wild-type Col-0 plants (*45*), T1 seeds were pooled and transgenic plants were selected on plates containing 1x Linsmaier and Skoog (LS) pH buffered basal salts, pH 5.7 (Caisson laboratories, UT, USA, cat. # LSP03-1LT) with 0.7% agar and 15 μg/ml glufosinate ammonium (Fisher Scientific, NH, USA, cat. # N-12111-250MG). Single-insertional transgenic lines were selected by chi-square test from T2 plants on 1x LS plates containing 15 μg/ml glufosinate ammonium. The expression of the tagged TFs was confirmed by Western blotting. Homozygous transgenic lines were selected from the subsequent generation for bulking seeds. DEX-inducible lines for the functional characterization of *DIG*s were generated by cloning the coding sequence of *DIG*s into p35S::LhGR-p6xOP::mGFP-attL1-ccdB-attR1 cassette by LR combination. After floral dip of wild-type Col-0 plants, T1 seeds were pooled and transgenic plants were selected by hygromycin.

### ChIP-seq experiments and analysis

0.4 g seeds were surface sterilized by 50% bleach + 0.05% Triton-X100 for 10 min. After 4 days of stratification at 4°C, seeds were spread on nylon mesh (Component Supply, FL, USA, cat. # U-CMN-215) in 6 hydroponics (Sigma-Aldrich, MO, USA, cat. # P1552) containing 1x pH buffered LS basal salts. After exposure under light for 4 hours to enhance germination, seeds were grown in dark at 22°C for 3 days. Etiolated seedlings were then switched to 1x LS buffer containing either (±)-ABA (MP Biomedicals LLC, CA, USA, cat. # 190673) dissolved in 100% ethanol at a final concentration of 10 μM or ethanol alone as mock and treated for 4 hours in dark before ChIP as previously described (*46*). Briefly, harvested seedlings were cross-linked by 1% formaldehyde solution (Sigma-Aldrich, cat. # F8775) under vacuum for 20 min. After nuclei isolation, chromatin was sonicated to 100-400 bp fragments. Tagged TFs in the transgenic lines were immunoprecipitated by a rabbit polyclonal anti-GFP antibody (Thermo Fisher Scientific, MA, USA, cat. # A11122). After elution, reverse crosslinking and DNA purification, Illumina TruSeq libraries were constructed according to manufacturer's protocols. All ChIP-seq experiments in both ABA and ethanol mock-treatment conditions were done with biological replicates. Uniquely mapped sequencing reads to the TAIR10 genome assembly (Bowtie v0.12.7) (*47*, *48*) were used to call peaks by the IDR pipeline utilizing MACS2 (*49*) with mock IP of wild-type Col-0 ChIPped by the anti-GFP antibody as a control. Peaks with a $P \le 1 \times 10^{-16}$ were kept and differential binding of TFs were analyzed by DiffBind (v1.10.1 with edgeR v3.0.8) (*23*). To calculate TF binding similarity in Fig. 1E, the center of peaks (termed "summits") of all ChIPped TFs were pooled together to create a union list. Sequencing coverage within 50 base pairs of summits in the union list was counted and normalized by deepTools (v1.5.8) (*50*). Pairwise Pearson correlation between samples was used as entries in the distance matrix to plot the heat map in Fig. 1E. Hierarchy height of ChIPped TFs was calculated as previously described (*40*)

$$h = (O - I)/(O + I)$$

where $O$ and $I$ are out-degree and in-degree of examined TF through top-ranked dynamic binding. Peaks in each dynamic binding categories were associated to TAIR10 annotated genes within 1000 bp from the summit of the peaks, using the R BioConductor package ChIPpeakAnno (v2.12.1) (*51*).

### Motif discovery and modeling the contribution of individual motifs to TF binding dynamics

De novo motif discovery was carried out by MEME-ChIP (MEME 4.9.1) using a background file calculated from TAIR10 intergenic sequences (*47*). Top five enriched motifs identified within 50 base pairs of the summits were filtered at e-value cutoff of 1e-05. To model the contribution of individual features, a set of nonredundant sequence features were selected to represent the overall motif diversity. To do this we first assembled a set of 135 motifs in our data set, consisted of the two most enriched motifs in the top 600 peaks in ABA- and mock-treated conditions for each TF, as well as top five motifs enriched in dynamic and static peaks for all TFs. The motifs were clustered by applying hierarchical clustering using motif distances calculated by Pearson correlated coefficients as column comparison metric and Ungapped Smith-Waterman alignment method (*52*, *53*). Dynamic tree cut of the clustering dendrogram (*54*) identified 19 major clusters (color of dendrogram branch and the left of the annotation tracks in fig. S8). As several of the clusters contain similar motifs (for example, the G-boxes and the AG-rich motifs are split into multiple clusters), we selected 11 sequence features to capture the diversity in this set of motifs indicated by dark red color of motif name and dark red color in the right annotation track in fig. S8. Basal binding was measured as log2(normalized read counts) under mock treatment and occurrences of motifs were assessed by FIMO (*55*) at a $P$ value cutoff of 0.0004. These features were used in the lm() function of R (v3.2.2) to fit log2(fold change) of the binding of indicated TFs in Fig. 2D between ABA- and mock treatment. Relative changes of explained variance was calculated as

$$(R^{2'} - R^2)/R^2$$

where $R^{2'}$ and $R^2$ are the adjusted $R^2$ from lm() output that includes and excludes cluster II or cluster III motif as a feature, respectively.

### RNA-seq experiments and analysis

For ABA time series experiments, two biological replicates of 3-day-old etiolated, hydroponic-grown wild-type Col-0 seedlings were treated either by 10 μM (±)-ABA (MP biomedicals, LLC, cat. # 190673) dissolved in ethanol or ethanol-only mock control for 1, 4, 8, 12, 24, 36, and 60 hours. For DEX treatment, short-term experiment was carried out by treating 3-day-old etiolated, DEX-inducible *GFP-DIG1*, *GFP-DIG2*, or *GFP* lines with 10 μM DEX (Sigma Aldrich, cat. # D9184); long term experiment was carried out by growing the same lines of plants containing 500 nM DEX for 10 days. Total RNA was isolated using the RNeasy Plant Mini Kit (Qiagen, CA, USA, cat. # 74903), and cDNA libraries were constructed using the TruSeq Stranded Total RNA LT Sample Prep Kit (Illumina, CA, USA, cat. # 15032611) according to manufacturers' instructions. Single-end reads were generated by the HiSeq 2500 Sequencing System (Illumina) and mapped to TAIR10 genome assembly using TopHat 2 (v2.0.8) (*56*). Mapped reads with mapping score equal to or larger than 10 were counted by HTSeq (v0.5.4) (*57*) and analyzed by edgeR (v3.6.2) (*18*) to identify differentially expressed genes using contrasts between ABA- and mock-treated samples at each time point and false discovery rate 0.01 or 0.05 as thresholds.

### DREM

The Dynamic Regulatory Events Miner (DREM) (*25*, *58*), integrates TF-gene interactions from ChIP-seq experiments with time series gene expression data to identify patterns of temporal gene expression, the associated regulators and the dynamics of the interactions. Splits in the reconstructed network (green nodes in Figs. 1B and 2C and figs. S1 and S5) represent divergence of genes that are co-regulated up to that point and can be annotated by DREM with the TFs that are predicted to regulate them, allowing us to associate the temporal information (the timing of the splits) with the interaction information either directly measured by ChIP-seq (fig. S1) or inferred from the AGRIS database (*26*), PBM (*13*), and DAP-seq (*15*) data (fig. S5). The analysis performed here used the log fold change of 3061 DE genes (table S1) identified in the ABA time series RNA-seq data. DREM paths were created using all DE genes without further filtering.

For GO enrichment in DE genes targeted by categories of dynamic binding peaks (Fig. 2D),

we defined the genes in distinct DREM paths as foreground, and all expressed genes as background, and retrieved the Functional Annotation Chart with EASE score (modified *P* value) threshold of 0.1 and count threshold of 2, using functionalities provided by the R BioConductor package RDAVIDWebService (*59*) to query the DAVID web service (*60*). The GO terms in GO_TERM_BP_FAT with FDR ≤ 1% from all target gene sets are combined, and the enrichment *P* values of these terms are retrieved for each gene set to create the heatmap in Fig. 2D. If a term is not reported to be significant for a target set, its *P* value is set to 0.1 (the *P* value threshold).

### Modeling the contribution of individual TFs to gene expression

We adopted an approach similar to previous regression-based models that relate gene expression to TF binding (*61*, *62*). We first defined TF affinity score (TFAS), $A_{ij}$, for TF $j$ on gene $i$, using the peak closest to the TSS of the gene

$$A_{ij} = ge^{-\frac{d}{d_0}}$$

where $g$ is the log2 TMM normalized read counts of the peak, $d$ is the distance of the peak summit to TSS. $d_0$ is set to 1000. For $N$ genes and $M$ TF, we constructed one $N \times M$ TFAS matrix for the ABA treatment and one for the ethanol mock treatment, and concatenated these two matrices horizontally to create a final $N \times 2M$ matrix $A$. We centered and scaled each column of $A$ and fit a log-linear model

$$\log Y_i = \sum_{j=1}^{2M} \beta_j A_{ij} + \epsilon_i$$

where $Y_i$ is the fold change in expression of gene $i$ at 4 hours ABA treatment compared to mock. We limited the model training and testing to genes that are differentially expressed at 4 hours with FDR ≤ 0.01 and those are not differentially expressed at all time points (FDR > 0.7). A glmnet regression model (*63*) was trained on 75% of the genes by five repeats of 10-fold cross-validation using the caret package in R (*64*) with tuning metric set to RMSE and the elastic net mixing parameter $\alpha = 0$ to allow selection correlated TFAS features. The "best" rule was used to choose a value for the tuning parameter (in this case, the regularization parameter $\lambda$), i.e., a value that minimized the average RMSE of the regression on the 50 resampling of the training set. The glmnet model was then fitted using the chosen $\lambda$ value to arrive at the regression coefficients in the final model. The unscaled coefficients of the TFAS features are plotted as binding feature importance in Fig. 2F.

### Confocal imaging

Nine-day-old DEX inducible GFP and GFP-DIG1 seedlings grown on 1x LS plates containing 200 nM DEX and 300 nM ABA were imaged by Zeiss 710 confocal microscope under an Argon laser at 488 nm. GFP signal was captured within the 493- to 548-nm emission window and was pseudo-colored in green. Auto-fluorescence from chloroplasts was captured within the 569- to 695-nm emission window and was pseudo-colored in red.

### Sequence analysis of DIGs and their homologs

Protein sequence of DIG1 was used as a query to search for homologous protein sequences in *A. thaliana* by the BLASTP search tool on Ensembl Plants (*65*). The resulting six protein sequences (Q9FK36, Q9SMP6, Q9FGW7, Q9LK28, Q9FKS6, Q9FKS7) were used to query *A. thaliana*, *Glycine max*, *Solanum lycopersicum*, *Oryza sativa* japonica, and *Zea mays* by BLASP, resulting in 21 homologous sequences. These sequences were aligned by MEGA6 (*66*) using distance-based maximum likelihood method, and bootstrap values were generated from 1000 replications.

### Chlorophyll measurement

Chlorophyll content was determined as previously described (*67*). Briefly, each sample consisting of ~50 seeds were germinated and grown on LS plates supplemented with or without ABA and DEX for eight days. The seedlings were collected and ground in liquid nitrogen. Chlorophyll were extracted by 80% acetone until pellets were almost white. Absorbance was measured at 647 and 664 nm in a DU-730 spectrophotometer (Beckman Coulter, CA, USA). Chlorophyll content was determined as

$$chl\ a + b = 17.76 * A647 + 7.34 * A664$$

Chlorophyll content of each transgenic line was then normalized by the corresponding seedlings grown on LS plates containing no DEX or ABA. The 95% confidence interval around the mean estimate was calculated from three biological replicates.

## REFERENCES AND NOTES


1. E. Birney *et al.*, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007). doi: 10.1038/nature05874; pmid: 17571346
2. M. B. Gerstein *et al.*, Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010). pmid: 21177976
3. S. Roy *et al.*, Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**, 1787–1797 (2010). doi: 10.1126/science.1198374; pmid: 21177974
4. K. Yamaguchi-Shinozaki, K. Shinozaki, Transcriptional regulatory networks in cellular responses and tolerance to dehydration and cold stresses. *Annu. Rev. Plant Biol.* **57**, 781–803 (2006). doi: 10.1146/annurev.arplant.57.032905.105444; pmid: 16669782
5. S. R. Cutler, P. L. Rodriguez, R. R. Finkelstein, S. R. Abrams, Abscisic acid: Emergence of a core signaling network. *Annu. Rev. Plant Biol.* **61**, 651–679 (2010). doi: 10.1146/annurev-arplant-042809-112122; pmid: 20192755
6. S.-Y. Park *et al.*, Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* **324**, 1068–1071 (2009). pmid: 19407142
7. Y. Ma *et al.*, Regulators of PP2C phosphatase activity function as abscisic acid sensors. *Science* **324**, 1064–1068 (2009). pmid: 19407143
8. J. Santiago *et al.*, The abscisic acid receptor PYR1 in complex with abscisic acid. *Nature* **462**, 665–668 (2009). doi: 10.1038/nature08591; pmid: 19898494
9. J. Santiago *et al.*, Modulation of drought resistance by the abscisic acid receptor PYL5 through inhibition of clade A PP2Cs. *Plant J.* **60**, 575–588 (2009). doi: 10.1111/j.1365-313X.2009.03981.x; pmid: 19624469
10. T. Furihata *et al.*, Abscisic acid-dependent multisite phosphorylation regulates the activity of a transcription activator AREB1. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1988–1993 (2006). doi: 10.1073/pnas.0505667103; pmid: 16446457
11. H. Fujii *et al.*, In vitro reconstitution of an abscisic acid signalling pathway. *Nature* **462**, 660–664 (2009). doi: 10.1038/nature08599; pmid: 19924127
12. Y. Fujita, M. Fujita, K. Shinozaki, K. Yamaguchi-Shinozaki, ABA-mediated transcriptional regulation in response to osmotic stress in plants. *J. Plant Res.* **124**, 509–525 (2011). doi: 10.1007/s10265-011-0412-3; pmid: 21416314
13. M. T. Weirauch *et al.*, Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014). doi: 10.1016/j.cell.2014.08.009; pmid: 25215497
14. A. Jolma *et al.*, DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527**, 384–388 (2015). doi: 10.1038/nature15518; pmid: 26550823
15. R. C. O'Malley *et al.*, Cistrome and epicistrome features shape the regulatory DNA landscape. *Cell* **165**, 1280–1292 (2016). doi: 10.1016/j.cell.2016.04.038; pmid: 27203113
16. J. S. Cumbie, S. A. Filichkin, M. Megraw, Improved DNase-seq protocol facilitates high resolution mapping of DNase I hypersensitive sites in roots in *Arabidopsis thaliana*. *Plant Methods* **11**, 42 (2015). doi: 10.1186/s13007-015-0087-1; pmid: 26339280
17. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1–21.29.9 (2015). doi: 10.1002/0471142727.mb2129s109; pmid: 25559105
18. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010). doi: 10.1093/bioinformatics/btp616; pmid: 19910308
19. J. M. Alonso, A. N. Stepanova, in *Bacterial Artificial Chromosomes* (Springer Protocols: Methods in Molecular Biology, Humana Press, 2015), vol. 1227.
20. C. W. Lim, J. H. Kim, W. Baek, B. S. Kim, S. C. Lee, Functional roles of the protein phosphatase 2C, AtAIP1, in abscisic acid signaling and sugar tolerance in *Arabidopsis*. *Plant Sci.* **187**, 83–88 (2012). doi: 10.1016/j.plantsci.2012.01.013; pmid: 22404835
21. M. Okamoto *et al.*, CYP707A1 and CYP707A2, which encode abscisic acid 8′-hydroxylases, are indispensable for proper control of seed dormancy and germination in *Arabidopsis*. *Plant Physiol.* **141**, 97–107 (2006). doi: 10.1104/pp.106.079475; pmid: 16543410
22. R. W. Kumimoto *et al.*, NUCLEAR FACTOR Y transcription factors have both opposing and additive roles in ABA-mediated seed germination. *PLOS ONE* **8**, e59481 (2013). doi: 10.1371/journal.pone.0059481; pmid: 23527203
23. R. Stark, G. Brown, "DiffBind: Differential binding analysis of ChIP-Seq peak data" (2015); http://bioconductor.riken.jp/packages/3.2/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf.
24. The Gene Ontology Consortium, Gene Ontology Consortium: Going forward. *Nucleic Acids Res.* **43**, D1049–D1056 (2015). doi: 10.1093/nar/gku1179; pmid: 25428369
25. M. H. Schulz *et al.*, DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC Syst. Biol.* **6**, 104 (2012). doi: 10.1186/1752-0509-6-104; pmid: 22897824
26. F. Celli *et al.*, AGRIS: Providing access to agricultural research data exploiting open data on the web. *F1000 Research.* **4**, 110 (2015). pmid: 26339471
27. A. J. Oldfield *et al.*, Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. *Mol. Cell* **55**, 708–722 (2014). doi: 10.1016/j.molcel.2014.07.005; pmid: 25132174
28. T. L. Bailey, J. Johnson, C. E. Grant, W. S. Noble, The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015). doi: 10.1093/nar/gkv416; pmid: 25953851
29. U. Piskurewicz, L. Lopez-Molina, The GA-signaling repressor RGL3 represses testa rupture in response to changes in GA and ABA levels. *Plant Signal. Behav.* **4**, 63–65 (2009). doi: 10.4161/psb.4.1.7331; pmid: 19704711
30. I. C. Lee *et al.*, Age-dependent action of an ABA-inducible receptor kinase, RPK1, as a positive regulator of senescence in *Arabidopsis* leaves. *Plant Cell Physiol.* **52**, 651–662 (2011). doi: 10.1093/pcp/pcr026; pmid: 21382977

31. I. Lee, U. M. Blom, P. I. Wang, J. E. Shim, E. M. Marcotte, Prioritizing candidate disease genes by network-based boosting of genome-wide association data. *Genome Res.* **21**, 1109–1121 (2011). doi: 10.1101/gr.118992.110; pmid: 21536720
32. S. Lumba *et al*., A mesoscale abscisic acid hormone interactome reveals a dynamic signaling landscape in *Arabidopsis*. *Dev. Cell* **29**, 360–372 (2014). doi: 10.1016/j.devcel.2014.04.004; pmid: 24823379
33. T. Yoshida *et al*., AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. *Plant J.* **61**, 672–685 (2010). doi: 10.1111/j.1365-313X.2009.04092.x; pmid: 19947981
34. J. Lee *et al*., Analysis of transcription factor HY5 genomic binding sites revealed its hierarchical role in light regulation of development. *Plant Cell* **19**, 731–749 (2007). doi: 10.1105/tpc.106.047688; pmid: 17337630
35. M. K. Jensen *et al*., ATAF1 transcription factor directly regulates abscisic acid biosynthetic gene NCED3 in *Arabidopsis thaliana*. *FEBS Open Bio* **3**, 321–327 (2013). doi: 10.1016/j.fob.2013.07.006; pmid: 23951554
36. Z.-L. Zhang *et al*., Scarecrow-like 3 promotes gibberellin signaling by antagonizing master growth repressor DELLA in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 2160–2165 (2011). doi: 10.1073/pnas.1012232108; pmid: 21245327
37. M. Kohno, H. Takato, H. Horiuchi, K. Fujita, S. Suzuki, Auxin-nonresponsive grape *Aux/IAA19* is a positive regulator of plant growth. *Mol. Biol. Rep.* **39**, 911–917 (2012). doi: 10.1007/s11033-011-0816-0; pmid: 21562765
38. K. Sugimoto *et al*., Molecular cloning of *Sdr4*, a regulator involved in seed dormancy and domestication of rice. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5792–5797 (2010). doi: 10.1073/pnas.0911965107; pmid: 20220098
39. H. Yu, M. Gerstein, Genomic analysis of the hierarchical structure of regulatory networks. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14724–14731 (2006). doi: 10.1073/pnas.0508637103; pmid: 17003135
40. M. B. Gerstein *et al*., Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012). doi: 10.1038/nature11245; pmid: 22955619
41. C. Smaczniak *et al*., Characterization of MADS-domain transcription factor complexes in *Arabidopsis* flower development. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1560–1565 (2012). doi: 10.1073/pnas.1112871109; pmid: 22238427
42. M. A. Moreno-Risueno *et al*., Transcriptional control of tissue formation throughout root development. *Science* **350**, 426–430 (2015). pmid: 26494755
43. S.-Y. Park *et al*., Agrochemical control of plant water use using engineered abscisic acid receptors. *Nature* **520**, 545–548 (2015). doi: 10.1038/nature14123; pmid: 25652827
44. N. C. Shaner, P. A. Steinbach, R. Y. Tsien, A guide to choosing fluorescent proteins. *Nat. Methods* **2**, 905–909 (2005). doi: 10.1038/nmeth819; pmid: 16299475
45. S. J. Clough, A. F. Bent, Floral dip: A simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *Plant J.* **16**, 735–743 (1998). doi: 10.1046/j.1365-313x.1998.00343.x; pmid: 10069079
46. L. Song, Y. Koga, J. R. Ecker, Profiling of transcription factor binding events by chromatin immunoprecipitation sequencing (ChIP-seq). *Curr. Protoc. Plant Biol.* **1**, 293–306 (2016). doi: 10.1002/cppb.20014
47. P. Lamesch *et al*., The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210 (2012). doi: 10.1093/nar/gkr1090; pmid: 22140109
48. B. Langmead, Aligning short sequencing reads with Bowtie. *Curr. Protoc. Bioinformatics* **32**, 11.7.1–11.7.14 (2010). pmid: 21154709
49. S. G. Landt *et al*., ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012). doi: 10.1101/gr.136184.111; pmid: 22955991
50. F. Ramírez, F. Dündar, S. Diehl, B. A. Grüning, T. Manke, deepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014). doi: 10.1093/nar/gku365; pmid: 24799436
51. L. J. Zhu *et al*., ChIPpeakAnno: A Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010). doi: 10.1186/1471-2105-11-237; pmid: 20459804
52. S. Mahony, P. E. Auron, P. V. Benos, DNA familial binding profiles made easy: Comparison of various motif alignment and clustering strategies. *PLOS Comput. Biol.* **3**, e61 (2007). doi: 10.1371/journal.pcbi.0030061; pmid: 17397256
53. E. Mercier, R. Gottardo, "Motif identification and validation: MotIV" (2014); www.bioconductor.org/packages/devel/bioc/vignettes/MotIV/inst/doc/MotIV.pdf.
54. P. Langfelder, B. Zhang, S. Horvath, Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics* **24**, 719–720 (2008). doi: 10.1093/bioinformatics/btm563; pmid: 18024473
55. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: Scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011). doi: 10.1093/bioinformatics/btr064; pmid: 21330290
56. D. Kim *et al*., TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013). doi: 10.1186/gb-2013-14-4-r36; pmid: 23618408
57. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015). doi: 10.1093/bioinformatics/btu638; pmid: 25260700
58. A. Wise, Z. Bar-Joseph, cDREM: Inferring dynamic combinatorial gene regulation. *J. Comput. Biol.* **22**, 324–333 (2015). doi: 10.1089/cmb.2015.0010; pmid: 25844671
59. C. Fresno, E. A. Fernández, RDAVIDWebService: A versatile R interface to DAVID. *Bioinformatics* **29**, 2810–2811 (2013). doi: 10.1093/bioinformatics/btt487; pmid: 23958726
60. X. Jiao *et al*., DAVID-WS: A stateful web service to facilitate gene/protein list analysis. *Bioinformatics* **28**, 1805–1806 (2012). doi: 10.1093/bioinformatics/bts251; pmid: 22543366
61. Z. Ouyang, Q. Zhou, W. H. Wong, ChIP-Seq of transcription factors predicts absolute and differential gene expression in embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21521–21526 (2009). doi: 10.1073/pnas.0904863106; pmid: 19995984
62. R. C. McLeay, T. Lesluyes, G. Cuellar Partida, T. L. Bailey, Genome-wide in silico prediction of gene expression. *Bioinformatics* **28**, 2789–2796 (2012). doi: 10.1093/bioinformatics/bts529; pmid: 22954627
63. J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010). doi: 10.18637/jss.v033.i01; pmid: 20808728
64. M. Kuhn, Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008). doi: 10.18637/jss.v028.i05
65. P. J. Kersey *et al*., Ensembl Genomes 2013: Scaling up access to genome-wide data. *Nucleic Acids Res.* **42**, D546–D552 (2014). doi: 10.1093/nar/gkt979; pmid: 24163254
66. K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: Molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013). doi: 10.1093/molbev/mst197; pmid: 24132122
67. E. Petrillo *et al*., A chloroplast retrograde signal regulates nuclear alternative splicing. *Science* **344**, 427–430 (2014). doi: 10.1126/science.1250322; pmid: 24763593


## ACKNOWLEDGMENTS

We thank F. Turck and K. N. Chang for discussions of the ChIP procedures; J. Alonso for discussions of the recombineering procedures; J. Chory, N. Fedoroff, and M. Zander for comments; C. Serrano, J. P. Saldierna, and A. Nasamran for cloning and plant maintenance; U. Pedmale and M. Xie for sharing plasmids; and J. Simon for assistance with graphics. L.S. and S.C.H. were supported by Salk Pioneer Postdoctoral Fellowships. This work was supported by grants from the Gordon and Betty Moore Foundation (GBMF 3034 to J.R.E), NSF (MCB-1024999 to J.R.E. and DBI- 1356505 to Z.B.-J.), and NIH (U01 HL122626-01 to Z.B.-J.). J.R.E. is an investigator of the HHMI. All of the reported RNA-Seq and ChIP-seq data has been deposited at the National Center for Biotechnology Information (Gene Expression Omnibus accession no. GSE80568). Genome browser tracks and motifs of our data set can be viewed at www.ABAtf.net.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/aag1550/suppl/DC1
Figs. S1 to S14
Tables S1 to S9
References (*68*–*96*)

16 May 2016; accepted 28 September 2016
10.1126/science.aag1550

RESEARCH ARTICLE SUMMARY

YEAST GENETICS

# Exploring genetic suppression interactions on a global scale

**Jolanda van Leeuwen, Carles Pons, Joseph C. Mellor, Takafumi N. Yamaguchi, Helena Friesen, John Koschwanez, Mojca Mattiazzi Ušaj, Maria Pechlaner, Mehmet Takar, Matej Ušaj, Benjamin VanderSluis, Kerry Andrusiak, Pritpal Bansal, Anastasia Baryshnikova, Claire E. Boone, Jessica Cao, Atina Cote, Marinella Gebbia, Gene Horecka, Ira Horecka, Elena Kuzmin, Nicole Legro, Wendy Liang, Natascha van Lieshout, Margaret McNee, Bryan-Joseph San Luis, Fatemeh Shaeri, Ermira Shuteriqi, Song Sun, Lu Yang, Ji-Young Youn, Michael Yuen, Michael Costanzo, Anne-Claude Gingras, Patrick Aloy, Chris Oostenbrink, Andrew Murray, Todd R. Graham, Chad L. Myers,* Brenda J. Andrews,* Frederick P. Roth,* Charles Boone***

**INTRODUCTION:** Genetic suppression occurs when the phenotypic defects caused by a mutated gene are rescued by a mutation in another gene.

These genetic interactions can connect genes that work within the same pathway or biological process, providing new mechanistic insights into cellular function, or they can correct defects in gene expression or protein production. More generally, suppression interactions may play an important role in the genetics underlying human diseases, such as the diverse penetrance of Mendelian disease variants. Our ability to interpret personal genome sequences remains limited, in part, because we lack an understanding of how sequence variants interact in nonadditive ways to generate profound phenotypes, including genetic suppression.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aag0839

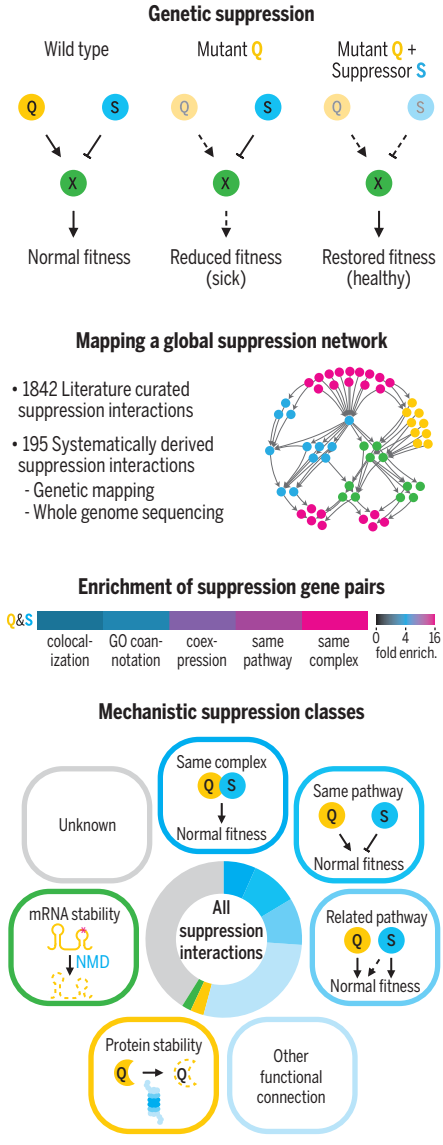
**RATIONALE:** Genetic interactions, in which mutations in two different genes combine to generate an unexpected phenotype, may underlie a significant component of trait heritability. Although genetic interactions that compromise fitness, such as synthetic lethality, have been mapped extensively, suppression interactions have not been explored systematically. To understand the general principles of genetic suppression and to examine the extent to which these interactions reflect cellular function, we harnessed the powerful genetics of the budding yeast *Saccharomyces cerevisiae* to assemble a global network of genetic suppression interactions.

**RESULTS:** By analyzing hundreds of published papers, we assembled a network of genetic suppression interactions involving ~1300 different yeast genes and ~1800 unique interactions. Through automated genetic mapping and whole-genome sequencing, we also isolated an unbiased, experimental set of ~200 spontaneous suppressor mutations that correct the fitness defects of deletion or hypomorphic mutant alleles. Integrating these results yielded a global suppression network.

The majority of suppression interactions identified novel gene-gene connections, thus providing new information about the functional wiring diagram of a cell. Most suppression pairs connected functionally related genes, including genes encoding members of the same pathway or complex. The functional enrichments observed for suppression gene pairs were several times as high as those found for other types of genetic interactions; this highlighted their discovery potential for assigning gene function. Our systematic suppression analysis also identified a prevalent allele-specific mechanism of suppression, whereby growth defects of hypomorphic alleles can be overcome by mutations that compromise either protein or mRNA degradation machineries.

From whole-genome sequencing of suppressor strains, we also identified additional secondary mutations, the vast majority of which appeared to be random passenger mutations. However, a small subset of genes was enriched for secondary mutations, several of which did not affect growth rate but rather appeared to delay the onset of the stationary phase. This delay suggests that they are selected for under laboratory growth conditions because they increase cell abundance within a propagating population.

**CONCLUSION:** A global network of genetic suppression interactions highlights the major potential for systematic studies of suppression to map cellular function. Our findings allowed us to formulate and quantify the general mechanisms of genetic suppression, which has the potential to guide the identification of modifier genes affecting the penetrance of genetic traits, including human disease. ■




**Global analysis of genetic suppression.** Genetic suppression interactions occur when the detrimental effects of a primary mutation can be overcome by a secondary mutation. Both literature-curated and experimentally derived suppression interactions were collected and yielded a genetic suppression network. This global network was enriched for functional relationships and defined distinct mechanistic classes of genetic suppression.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: cmyers@cs.umn.edu (C.L.M.); brenda.andrews@utoronto.ca (B.J.A.); fritz.roth@utoronto.ca (F.P.R); charlie.boone@utoronto.ca (C.B.)
Cite this article as J. van Leeuwen *et al.*, *Science* **354**, aag0839 (2016). DOI: 10.1126/science.aag0839

RESEARCH ARTICLE

YEAST GENETICS

# Exploring genetic suppression interactions on a global scale

Jolanda van Leeuwen,[1]* Carles Pons,[2,3]* Joseph C. Mellor,[1,4]† Takafumi N. Yamaguchi,[1,4,5] Helena Friesen,[1] John Koschwanez,[6] Mojca Mattiazzi Ušaj,[1] Maria Pechlaner,[7] Mehmet Takar,[8] Matej Ušaj,[1] Benjamin VanderSluis,[2]‡ Kerry Andrusiak,[1,5] Pritpal Bansal,[1,4] Anastasia Baryshnikova,[9] Claire E. Boone,[1] Jessica Cao,[1] Atina Cote,[1,4] Marinella Gebbia,[1,4] Gene Horecka,[1] Ira Horecka,[1] Elena Kuzmin,[1,5] Nicole Legro,[1] Wendy Liang,[1] Natascha van Lieshout,[1,4,5] Margaret McNee,[1] Bryan-Joseph San Luis,[1] Fatemeh Shaeri,[1,4] Ermira Shuteriqi,[1] Song Sun,[1] Lu Yang,[1] Ji-Young Youn,[4] Michael Yuen,[1] Michael Costanzo,[1] Anne-Claude Gingras,[4,5] Patrick Aloy,[3,10] Chris Oostenbrink,[7] Andrew Murray,[6] Todd R. Graham,[8] Chad L. Myers,[2,11]§ Brenda J. Andrews,[1,5]§ Frederick P. Roth,[1,4,5,11,12]§ Charles Boone[1,5,11]§

Genetic suppression occurs when the phenotypic defects caused by a mutation in a particular gene are rescued by a mutation in a second gene. To explore the principles of genetic suppression, we examined both literature-curated and unbiased experimental data, involving systematic genetic mapping and whole-genome sequencing, to generate a large-scale suppression network among yeast genes. Most suppression pairs identified novel relationships among functionally related genes, providing new insights into the functional wiring diagram of the cell. In addition to suppressor mutations, we identified frequent secondary mutations, in a subset of genes, that likely cause a delay in the onset of stationary phase, which appears to promote their enrichment within a propagating population. These findings allow us to formulate and quantify general mechanisms of genetic suppression.

Although causative variants have been identified for many Mendelian disorders, challenges remain in understanding how genetic variants combine to generate phenotypes. Great progress has been made in mapping and interpreting genetic interactions in yeast, by using growth rate as a proxy for fitness. High-throughput genetic interaction studies have identified hundreds of thousands of negative and positive interactions, in which the fitness defect of a yeast double mutant is either more or less severe, respectively, than the expected effect of combining the single mutants (Fig. 1A) (*1*, *2*). Positive interactions indicate that the phenotypic effects associated with detrimental mutations can be masked or overcome and may explain why certain individuals are healthy despite carrying severe disease-causing mutations (*3*).

Positive interactions can be further classified by their relative strength, ranging from masking, in which the double mutant fitness is higher than expected but less than or equal to that of the slowest growing single mutant, to suppression, in which the double mutant is healthier than the slowest growing single mutant and possibly has a fitness that is comparable to that of wild type (Fig. 1A) (*1*, *4*). These classes of positive interactions can represent biologically distinct functional relationships (*4*, *5*). Most positive interactions identified by systematic genetic interaction screens in yeast, based on synthetic genetic array (SGA) analysis with loss-of-function mutations (*2*, *6*), are relatively weak masking interactions (fig. S1A), such as the positive interactions that occur among genes within the same nonessential complex or pathway (*7*). By contrast, stronger suppression interactions remain largely unexplored.

Spontaneous suppressor mutations can be selected to overcome the fitness defect associated with a specific mutant allele. Extragenic suppressor mutations encompass two basic classes: (i) informational suppressors that change the protein translational or mRNA transcriptional machinery, such that the primary mutation is reinterpreted, and (ii) functional suppressors in which a mutation in a second gene functionally compensates for the defect associated with the primary mutation (*8*). Here, our major goal was to investigate the general principles of functional suppression by assembling a global network of these interactions, which should provide new mechanistic insights about protein function and enable the ordering of components of biological pathways.

## A network of literature-curated suppression interactions

To capture existing suppression interactions in *Saccharomyces cerevisiae*, we examined ~6000 potential interactions in ~1700 published papers derived from the BioGRID's "synthetic rescue" data set (*9*). From each interaction, we annotated the type of suppressor mutation (e.g., spontaneous mutation or deletion allele); the type of mutation that is being suppressed, which we refer to as a "query" mutation; and the use of specific conditions (e.g., a drug or specialized carbon source). Suppression interactions that were intragenic, involved a specific phenotype other than growth, or included more than two genes were excluded from the final data set. We also removed suppression interactions derived from high-throughput experiments or dosage interactions in which either the query or the suppressor was overexpressed. The resulting literature-curated network encompassed 1304 genes and 1842 unique suppression interactions (table S1). We visualized this network using a force-directed layout (*10*), so that query genes that share a common suppressor tend to be positioned together (Fig. 1B). Most query genes (69%) are suppressed by one or two suppressor genes, whereas a small subset of queries (5%) have numerous (10 to 27) reported interactions (fig. S1B). Despite the relatively low average network degree, genes involved in highly studied processes, such as DNA replication and repair or chromatin and transcription, tend to group together because of their shared suppression interactions (Fig. 1B).

Combining data from multiple studies can reveal suppression mechanisms between pathways or protein complexes that may not be apparent from any individual study. Indeed, a subnetwork focused on DNA replication and repair pathways showed that many of the interactions appear to represent the activation of alternative DNA repair pathways (Fig. 1C). For example, mutations that perturb Rad51-dependent homologous recombination (HR) often lead to toxic chromosomal deletions or rearrangements due to increased repair of double-strand DNA breaks by nonhomologous end joining (NHEJ) (*11*). In this case, suppression can occur through



[1]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada. [2]Department of Computer Science and Engineering, University of Minnesota-Twin Cities, 200 Union Street, Minneapolis, MN 55455, USA. [3]Institute for Research in Biomedicine (IRB Barcelona), the Barcelona Institute for Science and Technology, Barcelona, Catalonia, Spain. [4]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada. [5]Department of Molecular Genetics, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada. [6]Department of Molecular and Cellular Biology, Harvard University, 52 Oxford Street, Cambridge, MA 02138, USA. [7]Institute of Molecular Modeling and Simulation, University of Natural Resources and Life Sciences, Muthgasse 18, A-1190 Vienna, Austria. [8]Department of Biological Sciences, Vanderbilt University, 1161 21st Avenue South, Nashville, TN 37232, USA. [9]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA. [10]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain. [11]Canadian Institute for Advanced Research, 180 Dundas Street West, Toronto, Ontario M5G 1Z8, Canada. [12]Department of Computer Science, University of Toronto, 160 College Street, Toronto, Ontario M5S 3E1, Canada.
*These authors contributed equally to this work. †Present address: seqWell Inc., 376 Hale Street, Beverly, MA 01915, USA. ‡Present address: Simons Center for Data Analysis, Simons Foundation, 160 Fifth Avenue, New York, NY 10010, USA. §Corresponding author. Email: cmyers@cs.umn.edu (C.L.M.); brenda.andrews@utoronto.ca (B.J.A.); fritz.roth@utoronto.ca (F.P.R); charlie.boone@utoronto.ca (C.B.)

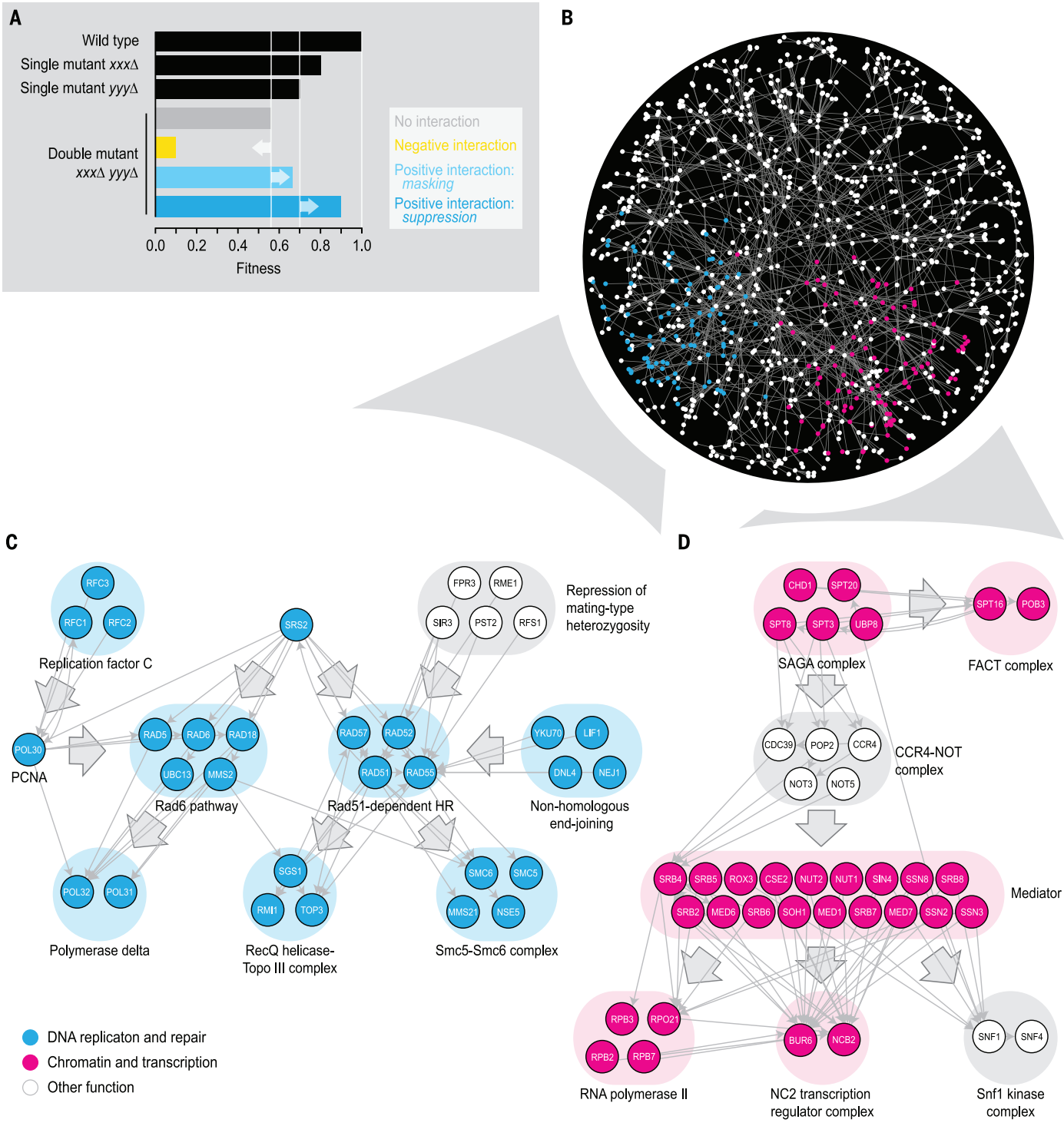



**Fig. 1. A global network of literature-curated suppression interactions for *S. cerevisiae*.** (**A**) Genetic interaction classes. When two single mutants (*xxx*Δ and *yyy*Δ) have a relative fitness of 0.8 and 0.7, the expected fitness of the resultant double mutant (*xxx*Δ *yyy*Δ) based on a multiplicative model is 0.8 × 0.7 = 0.56. A negative genetic interaction occurs when the observed double mutant fitness is lower than this expected fitness. A masking positive interaction occurs when the fitness of the double mutant is greater than expected, but lower or equal to that of the slowest growing single mutant. Suppression positive interactions occur when the double mutant fitness is greater than that of the slowest growing single mutant. (**B**) A global network of literature-curated suppression interactions for *S. cerevisiae*. Genes are represented as nodes and interactions as edges. The nodes were distributed using a force-directed layout, such that genes that share a suppressor tend to be close together on the network. Genes involved in chromatin and transcription or DNA replication and repair are highlighted in magenta and cyan, respectively. (**C** and **D**) Regions of the global network highlighting suppression interactions between complexes and pathways involved in chromatin and transcription (C) or DNA replication and repair (D) are shown. Arrows point from the suppressor to the query. PCNA, proliferating cell nuclear antigen.

NHEJ inactivation, which favors double-strand break repair by the compromised, but more accurate, HR machinery (*11*). Similar trends are observed for genes involved in transcription, for which suppression interactions between pathways mainly represent activation or repression of transcription (Fig. 1D). For example, mutations in genes encoding Mediator or RNA polymerase II subunits can reduce transcription efficiency, which can suppress the toxic effects of derepressed transcription caused by loss-of-function mutations in the NC2 transcription regulator complex (*12*). Thus, by integrating data from hundreds of papers, we derived a suppression network that provides insight on general suppression relationships and the ordering of pathways and complexes within a biological process.

### Suppression interactions within and across cellular processes

Consistent with other biological networks (*2*, *13*–*15*), many suppression interactions occurred between functionally related genes, such that a query mutant tended to be suppressed by another gene annotated to the same biological process (Fig. 2A). Genes connected by suppression interactions also tended to be coexpressed and encode proteins that function in the same subcellular compartment and/or belong to the same pathway or protein complex (Fig. 2B). The extent of functional relatedness between suppression gene pairs did not depend on the conditions under which the interaction was identified (e.g., a specific drug or carbon source), or whether the suppressor was isolated as a spontaneous suppressor mutation as opposed to an engineered allele that was directly tested for an interaction (fig. S2A). However, the frequency of shared complex membership was significantly higher for gene pairs in which the suppressor gene carried a gain-of-function mutation compared with gene pairs involving loss-of-function suppressor mutations ($P$ = 0.01, Fisher's exact test). Thus, when a query mutation perturbs a subunit of a complex, compensating mutations in another subunit can be gain of function—for example, by stabilizing the complex.

Notably, the functional enrichment observed in the genetic suppression network was substantially stronger than in a global network of negative and positive genetic interactions generated with SGA (*6*) (Fig. 2B). In fact, most positive genetic interactions identified in the global SGA network, especially among loss-of-function alleles of essential genes, do not overlap with other functional interaction data. Suppression interactions thus constitute a special class of positive genetic interaction that captures highly specific functional relationships between gene pairs (fig. S2B).

Despite their tendency to connect functionally related genes, suppression interactions also connect different biological processes. These interactions often occurred between genes involved in related processes, such as Golgi, endosome, or vacuole sorting and ER-Golgi traffic (Fig. 2A). Note that genes involved in protein degradation suppress growth defects associated with mutation of genes involved in many different biological


**Fig. 2. Properties of the suppression network.** (**A**) Frequency of suppression interactions connecting genes within and across indicated biological processes. Node size reflects fold enrichment for interacting gene pairs observed for a given pair of biological processes. Significance of the enrichment was determined by Fisher's exact test, comparing the observed frequency of suppression interactions between two given functional categories with the global frequency. The total number of suppression interactions involving genes annotated to a particular process is indicated. Kinet., kinetochore. (**B** and **C**) Fold enrichment for (B) colocalization, GO coannotation, coexpression, same pathway membership, and same complex membership for gene pairs involved in different types of genetic interaction (GI); and (C) overlap of literature-curated suppression interactions with dosage suppression interactions (*13*), or with negative and positive genetic interactions identified by SGA analysis using either an intermediate or a stringent interaction score threshold (*6*). A Fisher's exact test was performed to determine statistical significance of the results. (**D**) An example of a gene pair showing suppression, dosage suppression, and negative genetic interactions.

processes. This central role for protein turnover in the suppression network likely reflects a more general mechanism whereby growth defects of conditional temperature-sensitive (TS) alleles of essential query genes, which are often hypomorphic (partially functional) even at a permissive temperature, can be overcome by additional mutations that weaken the protein degradation machinery and elevate protein levels.

### Overlap with other genetic networks

The suppression network shows significant overlap with a dosage suppression network (*13*) ($P = 2 \times 10^{-101}$, Fisher's exact test) and with SGA-derived positive and negative genetic interaction networks (*2*, *6*) ($P = 5 \times 10^{-87}$ and $P = 1 \times 10^{-33}$, respectively, Fisher's exact test). The overlap with positive genetic interactions (fivefold enrichment) (Fig. 2C) is expected, as suppression interactions are an extreme type of positive interaction. Indeed, this overlap increases (11-fold enrichment) for stronger positive genetic interactions. The overlap of the suppression network with dosage-suppression interactions associated with gene overexpression reflects that overexpression may lead to a gain-of-function phenotype (*16*) and suppression can involve gain-of-function alleles (Fig. 2C and fig. S2C). Gain-of-function suppressor mutations also explain the 2.5-fold enrichment for negative genetic interactions between loss-of-function alleles (Fig. 2C and fig. S2C). For example, whereas the growth defect associated with loss-of-function mutations in *CDC25*, which encodes the guanine nucleotide exchange factor that activates Ras2, can be suppressed by gain-of-function mutations in *RAS2*, loss-of-function mutations in *RAS2* exacerbate the *cdc25* growth defect, thereby causing a synthetic lethal negative genetic interaction (Fig. 2D). Despite the overlap with other genetic networks, most suppression interactions (78%) are specific to the suppression network and thus provide novel insights into the functional wiring diagram of a cell.

### Systematic identification of spontaneous suppressor mutations

Literature-curated data can come from specific hypothesis-driven experiments and may thus be biased (*15*, *17*). We therefore compared the curated suppression network to an independent experimental set of spontaneous suppressor mutations identified through the large-scale application of SGA analysis. In SGA, a specific *natMX*-marked query mutation is crossed to an array of ~5000 *kanMX*-marked deletion mutants, to systematically construct a complete set of haploid *natMX*- and *kanMX*-marked double mutants (*18*, *19*). This also represents a genome-wide set of two-factor crosses, enabling us to scan the query strain genome for the presence of an unmarked extragenic suppressor locus, which SGA analysis reveals as a collinear set of small colonies spanning the genomic location of the suppressor mutation, which we refer to as a linkage group (*20*, *21*) (fig. S3A). In total, we completed 7056 full-genome SGA screens, involving mutant strains carrying deletion or hypomorphic alleles of 5845 different genes (*2*, *6*). In 251 SGA screens (~4%), we identified a linkage group that suggested the presence of a spontaneous extragenic suppressor mutation (tables S2 and S3).

The 251 candidate suppressor strains were analyzed by whole-genome sequencing, and for 216 (86%) of these, a mutation was discovered within the suppressor locus identified by SGA (fig. S3A and table S2). Almost all (98%) of these mutations were subsequently confirmed by Sanger sequencing (table S2). For 24 genes, multiple independently generated query strains carried a potential extragenic suppressor mutation (table S2). In 13 (54%) of these 24 cases, the extragenic suppressor mutations were in the same gene, whereas in the remaining 11 cases, two different suppressor genes were identified. In three instances, these different suppressor genes encoded known members of the same complex.

We next validated candidate suppressor genes using several genetic tests, including plasmid-based complementation assays and tetrad analysis of meiotic progeny derived from crossing each suppressor strain to a wild-type strain, a strain with a marked deletion that was genetically linked to the candidate suppressor, or a strain carrying a deletion or hypomorphic allele of the suppressor gene (fig. S3A) (*21*). Of the suppressor interactions, 88% gave a positive result in at least one assay (table S2). Based on these assays and the type of mutation, one-third (33%) of the suppressor mutations appeared to be gain-of-function, and two-thirds (67%) appeared to be loss-of-function mutations. We also randomly selected four potential loss-of-function and five potential gain-of-function suppressor alleles and introduced those into a diploid strain that was heterozygous for the corresponding query mutation. In all cases, sporulation and tetrad analysis of the resulting diploids confirmed the genetic interaction and identity of the suppressor mutation (table S2 and fig. S3A). Thus, we identified 216 unbiased mutations that arose spontaneously to suppress severe growth defects associated with 146 deletion mutants of nonessential genes and 70 hypomorphic alleles of essential genes (table S2).

Although we observed significant overlap with the literature-curated data set (15 shared interactions, $P = 1 \times 10^{-29}$, Fisher's exact test), most of the spontaneous suppression interactions identified through SGA (92%) have not been reported previously; this indicates that the yeast genetic suppression network has remained largely unexplored. The experimentally derived suppression interactions showed similar significant enrichments as the literature-curated set for different types of genetic interactions, as well as for functionally related gene pairs, suggesting that suppression interactions in both networks define close functional relationships between genes and share the same basic properties (fig. S3, B and C).

### Suppression interaction magnitude correlates with functional relatedness

Given that suppression interactions tend to connect functionally related genes, we examined whether the relative magnitude of a given suppression interaction was indicative of the extent of functional overlap. We estimated the relative magnitude of suppression for our systematic interactions (table S4) (*21*), ranked the suppression pairs by suppression magnitude, and calculated the fraction of functionally related pairs for the 33% strongest and weakest suppression interactions (fig. S4). Gene pairs exhibiting more severe suppression interactions showed stronger enrichments for various measures of functional relatedness (fig. S4), in line with what has been described for positive and negative genetic interactions (*2*). Thus, large improvements in fitness appear to be caused by mutations in genes that are functionally similar to the query, whereas weaker suppression may be achieved by more general or diverse mechanisms.

### Systematic analysis identifies suppressor hubs

The literature-curated network is enriched for genes involved in highly studied processes, such as chromatin and transcription, as well as DNA replication and repair (Fig. 3A). In contrast, in the experimentally derived network, queries and suppressors were more evenly spread over the various biological processes. As we found for the literature network (Fig. 2), genes involved in protein degradation were specifically overrepresented as suppressors in the systematic study (Fig. 3A), which mainly reflects suppression of point-mutation alleles of essential queries. Although no significant functional enrichment was found for genes involved in RNA processing, nonsense-mediated mRNA decay genes suppressed several DAmP alleles of essential genes (generated by decreased abundance by mRNA perturbation, DAmP) (*22*), which affect mRNA stability through disruption of their 3′ untranslated region. Thus, restoring protein or mRNA levels may represent a widespread mechanism to overcome growth defects caused by hypomorphic alleles.

It is noteworthy that suppressed queries with roles in ribosome biogenesis and translation were underrepresented in the literature but overrepresented in our systematic data set (Fig. 3A). This enrichment was driven by a set of 34 query genes, each encoding a component of the mitochondrial translation machinery. All 34 queries were suppressed by missense mutations in the α, β, or γ subunits of the $F_1$ domain of the mitochondrial adenosine triphosphate (ATP) synthase, and the majority of the substituted residues were located at the interfaces between these subunits (Fig. 3B). Mutations in the same mitochondrial ATP synthase subunits also suppressed deletion alleles of mitochondrial DNA and RNA polymerase genes, as well as three relatively uncharacterized genes: *IRC19*, *PET130*, and *YPR117W* (table S2). All of these query mutations led to loss of the mitochondrial genome (mtDNA), which results in decreased growth due to a defect in the import of proteins into the mitochondria (*23*) (Fig. 3, C and D, and fig. S5A). The ATP synthase suppressor mutations could restore both fitness and mitochondrial protein import in the absence of mtDNA (Fig. 3, C and D, and fig. S5B). Note that an

activity of the ATP synthase other than ATP synthesis was required for this suppression phenotype (fig. S5, C and D). Although the mechanism by which the suppressor mutations increase protein import is unclear, one possibility is that the mutations reverse ATP synthase activity to generate $ADP^{3-}$ instead of $ATP^{4-}$. The charge difference between these two nucleotide phosphates could be exploited by adenine nucleotide translocators to rebuild the mitochondrial membrane potential, which is lost in the absence of mtDNA and is thought to be required for protein import into the mitochondria (Fig. 3D) (*24*).

### Suppressor identification can predict novel gene function

The functional relationship observed between a query mutant and its suppressor can be exploited to assign gene function to previously uncharacterized genes. For example, in our systematically mapped suppressor network, we found that loss-of-function mutations in an uncharacterized gene, *YMR010W*, suppressed the growth defect of *mon2*Δ mutants (Fig. 4, A and B). Ymr010w belongs to the family of PQ-loop proteins, some of which function as membrane transporters (*25*), and localizes to both the Golgi and late endosomes (fig. S6A). Mon2 is distantly related to the Sec7 family of guanine nucleotide exchange factors and physically interacts with Dop1, a

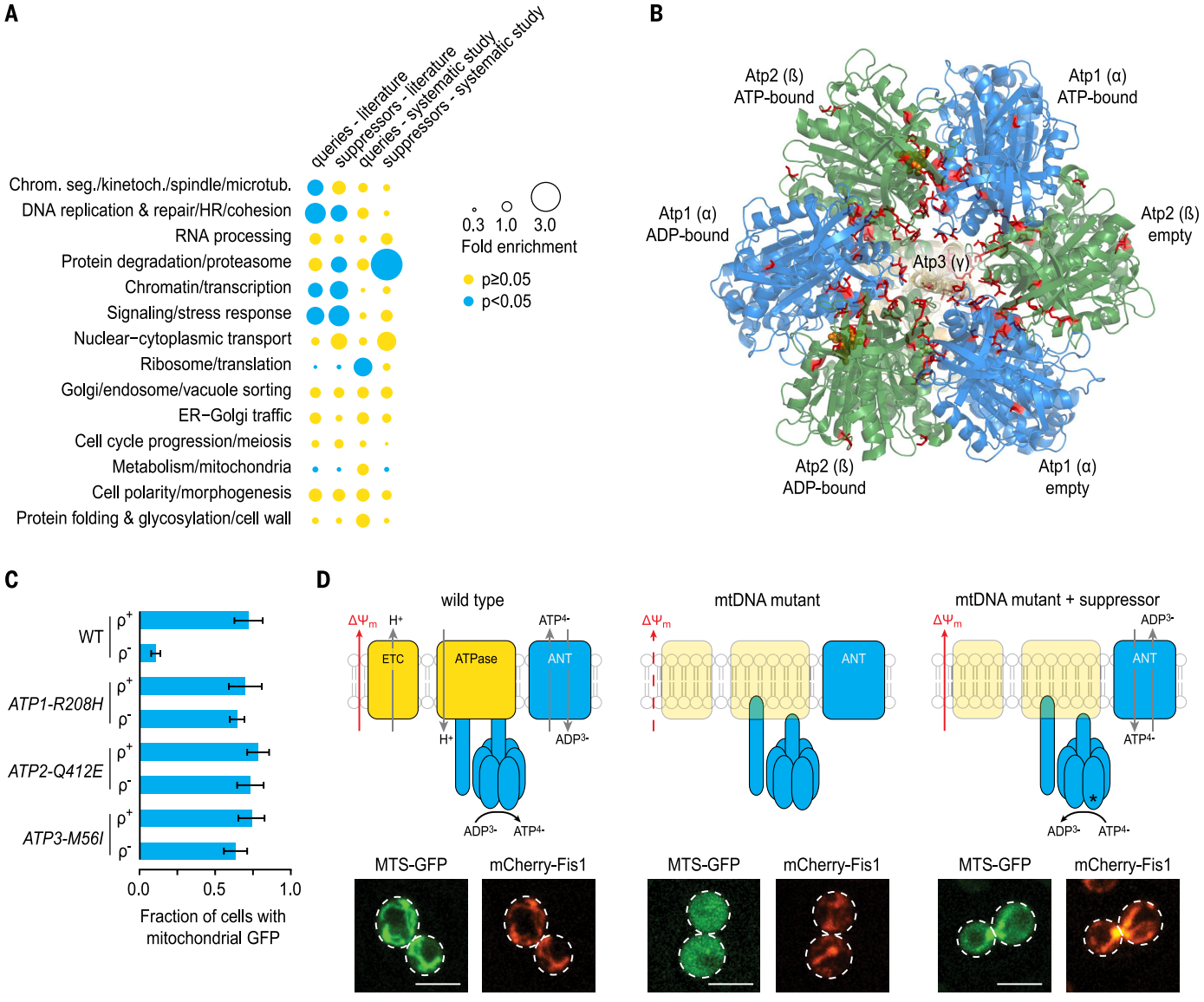



**Fig. 3. The mitochondrial $F_1$ ATPase (adenosine triphosphatase) is a suppressor hub in the systematic suppression network.** (**A**) The distribution of query and suppressor mutants in both the literature-curated and the systematic experimental network across different biological processes. Node size reflects fold enrichment or depletion for query and suppressor mutants observed for a given biological processes. Significant enrichment or depletion was determined by Fisher's exact test, comparing the observed to the expected proportion of genes in each functional category. Bonferroni-corrected *P* values are indicated. (**B**) Bottom view, facing the inner membrane from the mitochondrial matrix, of the yeast mitochondrial $F_1$ ATPase structure 2HLD. Residues that were found to suppress the growth defect of mitochondrial transcription or translation mutants are highlighted in red. Orange spheres represent the nucleotides bound to the catalytic sites. (**C**) Fraction of wild-type and ATP synthase-mutant cells either with intact ($\rho^+$) or (partially) deleted ($\rho^-$) mtDNA that show mitochondrial localization of GFP fused to a mitochondrial-targeting signal (MTS-GFP). Averages ($n = 4$) and SD are shown. (**D**) Model of ATP synthase–dependent suppression of mitochondrial mutants (top) and corresponding representative images of MTS-GFP import (bottom). Localization of outer mitochondrial membrane protein mCherry-Fis1 shows the presence and position of mitochondria. ETC, electron transport chain; $\Delta\Psi_m$, inner mitochondrial membrane potential; ANT, adenine nucleotide translocator. Scale bar, 5 μm.

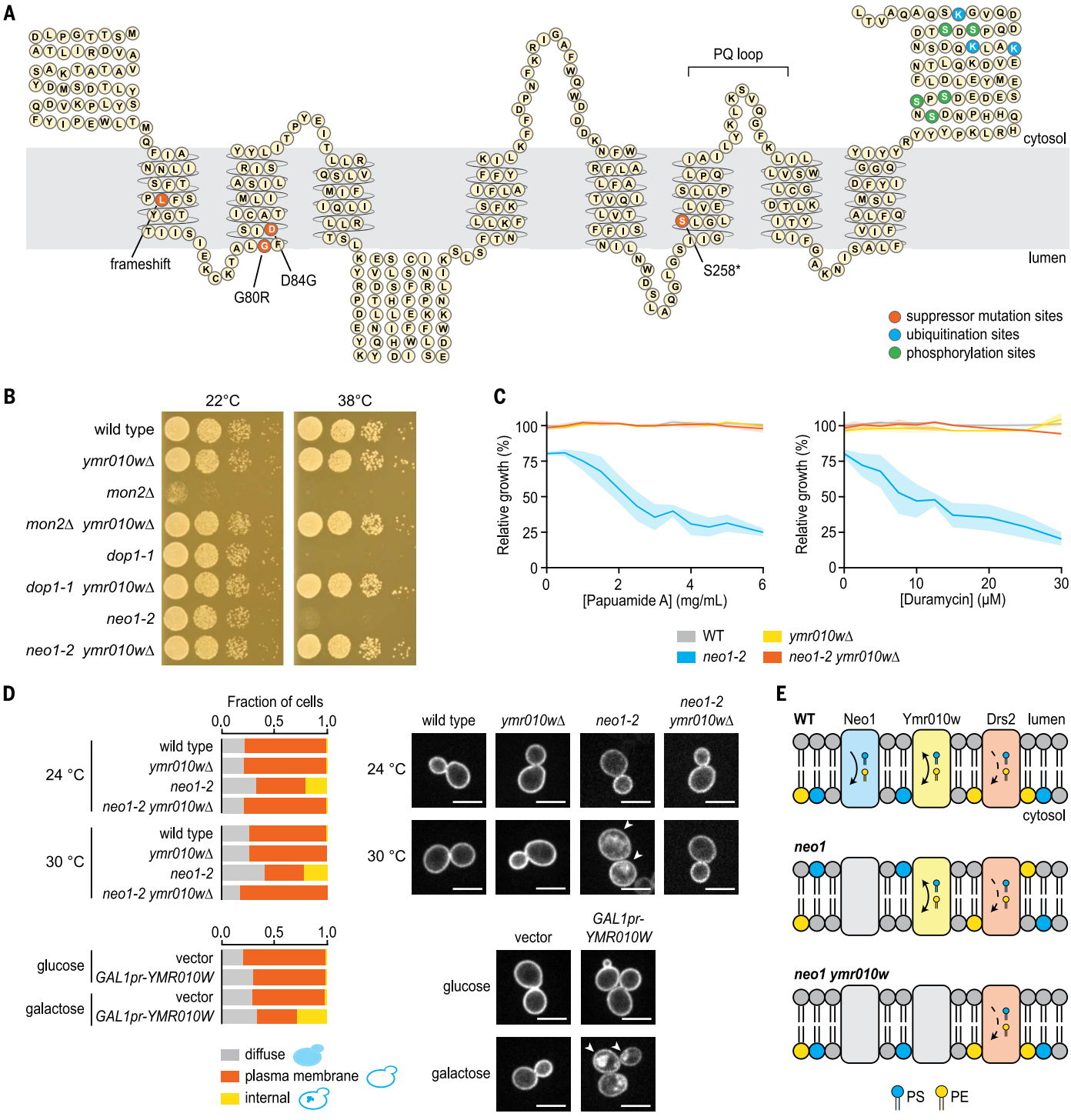



**Fig. 4. Characterization of *YMR010W* (*ANY1*).** (**A**) Predicted membrane topology of Ymr010w. Sites of suppressor mutations, ubiquitination, and phosphorylation are indicated. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (**B**) Suppression of the growth defect caused by a *mon2*Δ deletion allele, or TS alleles *dop1-1* and *neo1-2*, by deletion of *YMR010W*. Series of 10-fold dilutions of exponentially growing cultures of the indicated strains were spotted on plates with YPD medium and incubated at either 22°C or 38°C for 2 days. (**C**) Deletion of *YMR010W* restores membrane asymmetry in *neo1-2* cells. Wild-type, *ymr010w*Δ, *neo1-2* and *neo1-2 ymr010w*Δ cells were grown at 34°C in the presence of the phosphatidylserine (PS) targeting peptide papuamide A, or the phosphatidylethanolamine (PE) targeting peptide duramycin. Growth relative to vehicle-treated wild-type strain is plotted. SEM is indicated by shading ($n = 2$ to 3). (**D**) Intracellular distribution of PS, visualized using GFP-LactC2 (*31*). Shown are representative confocal fluorescent micrographs of exponentially growing cells of the indicated strains. The fraction of cells was calculated for each of the following groups: those (i) that showed diffuse cytosolic fluorescence or (ii) localization of GFP-LactC2 to the plasma membrane, or (iii) in which GFP-LactC2 was partially localized to distinct internal structures. Measurements were performed in triplicate on at least 100 cells, and averages are shown. (**E**) Model of suppression of flippase mutants by loss of Ymr010w.

conserved membrane protein involved in endosome to Golgi transport, as well as Neo1, an essential member of the phospholipid flippase family (*26*, *27*). When tested directly, we found that a *ymr010w*Δ deletion allele also suppressed the growth defects of *neo1-2* and *dop1-1* TS mutants (Fig. 4B). Moreover, a *ymr010w*Δ deletion allele suppressed the lethality associated with deletion alleles of the essential genes *NEO1* and *DOP1* (fig. S6B). Loss of *YMR010W* function can thus bypass the requirement for the Mon2/Dop1/Neo1 module.

The essential function of the Mon2/Dop1/Neo1 module is likely performed by Neo1, which is thought to flip phosphatidylserine (PS) and phosphatidylethanolamine (PE) from the exoplasmic to the cytoplasmic leaflet of membrane bilayers and thereby to establish an asymmetric distribution of these lipids (*28*). A *neo1-2* TS mutant is defective in establishing membrane asymmetry. This leads to hypersensitivity to papuamide A and duramycin, bioactive peptides that disrupt membranes through the binding of exposed PS and PE, respectively (*28*–*30*), and reduced plasma membrane localization of green fluorescent protein (GFP)–LactC2, a probe for visualizing the distribution of PS over cytoplasmic membrane leaflets (*31*) (Fig. 4, C and D). Overexpression of *YMR010W* also led to reduced levels of PS at the cytoplasmic leaflet of the plasma membrane, and accumulation of GFP-LactC2 in internal structures (Fig. 4D), thus mimicking the phenotype of a neo1-2 mutant. We found that a *ymr010w*Δ deletion allele suppressed both the sensitivity of a *neo1-2* TS mutant to papuamide A and duramycin (Fig. 4C), and the *neo1-2* GFP-LactC2 localization defect (Fig. 4D). The absence of these phenotypes suggests that the *neo1-2* phospholipid distribution defects are corrected in the double mutant.

In addition to suppressing loss of Neo1 function, a *ymr010w*Δ deletion allele suppressed the cold sensitivity caused by loss of the flippase Drs2 (fig. S6C). Moreover, *neo1*Δ lethality was no longer suppressed by *ymr010w*Δ in the absence of Drs2 (fig. S6C). An intriguing possibility is that Ymr010w functions as a scramblase that transports PS and PE bidirectionally to at least partially collapse the membrane asymmetry established by Neo1 and other flippases (Fig. 4E). Deletion of *YMR010W* would then allow Drs2, possibly with the help of other flippases, to more easily establish membrane asymmetry in the absence of Neo1. We named the *YMR010W* open reading frame *ANY1* for antagonizes Neo1 yeast phospholipid flippase.

## Frequent secondary mutations delay the onset of stationary phase

Whole-genome sequencing revealed that, besides the suppressor mutation, each suppressor strain carried on average eight additional secondary mutations (table S5). Unlike the suppressor mutations, none of these secondary mutations affected exponential cell growth enough to be detected by SGA mapping analysis (table S3), suggesting the majority are random mutations that arose during DNA replication. We therefore refer to these additional secondary mutations as "passenger" mutations. We identified a similar number of passenger mutations in a control set of 72 strains that did not carry a suppressor mutation that affects growth of the query mutant (table S5). Of the 304 strains that were sequenced at a coverage >10 times, only one query strain, deleted for *PMS1* that encodes a mismatch repair protein, displayed a mutator phenotype, exhibiting a relatively large number (76) of passenger mutations. In total, we identified 2024 unique passenger mutations, of which 996 were in coding regions, affecting 744 protein- or RNA-encoding genes. The fraction of missense, nonsense, and frameshift mutations was substantially smaller among the passenger mutations than among the suppressor mutations (Fig. 5A). In fact, most of the passenger mutations (64%) resulted in synonymous changes or mapped to intergenic regions (Fig. 5A). Furthermore, passenger missense mutations occurred less frequently in essential genes, were predicted to be less deleterious, were less often at protein-protein interaction interfaces, and occurred more often in disordered protein regions than suppressor missense mutations (Fig. 5B). Thus, the majority of the passenger mutations, which have no effect on exponential growth of the query strain, have a lower putative functional impact than the suppressor mutations that do affect query strain cell growth.

A previous study suggested that deletion of a particular query gene may select for further genetic changes, such as the occurrence of specific secondary nonsuppressor mutations (*32*). However, we did not observe a correlation between the number of passenger mutations and the fitness of the query strain (fig. S7A). Moreover, genes carrying passenger mutations do not tend to be coannotated or coexpressed with the corresponding query or suppressor gene (fig. S7B). In addition, we did not find any enrichment for particular GO terms among query genes that shared the same passenger mutation, or for shared passenger mutant genes among multiple, independent isolates of a particular query mutant strain. However, we found that 10 strains that all carried a suppressor mutation in *ATP2* but had different query mutations involved in mitochondrial transcription or translation, also harbored a third mutation in *HEM1*, *TPN1*, or *HAP1*. These three genes are important for heme biosynthesis, and these mutations may thus be selected for to maintain heme homeostasis in the absence of mitochondrial transcription, translation, or ATP synthase activity. Still, in most cases, different isolates of the same query suppressor strains did not contain mutations in the same passenger genes, and most passenger genes were not functionally related to either the query or the suppressor gene, indicating that passenger mutations are not generally dependent on preexisting mutations.

We did find several genes that were mutated in a large fraction of the sequenced strains; this suggested that they may be adaptive and may not represent innocuous passenger mutations (Fig. 5C). Of all sequenced strains, including wild-type controls, 29% carried unique mutations in *WHI2*, *IRA1*, *IRA2*, *RIM15*, *CUP9*, and/or *UBC13*. Multiple experimental evolution studies have identified a similar set of frequently mutated yeast genes (*33*–*35*). Most of the mutations were frameshift or nonsense mutations, suggesting a selection for loss-of-function of these genes (Fig. 5C). Exponential growth rates of *whi2*Δ, *ira2*Δ, *rim15*Δ, and *ubc13*Δ deletion mutants were not enhanced relative to a *his3*Δ deletion mutant control. It thus appears that there was no selection for these frequent secondary mutations on the basis of an increased maximum growth rate (fig. S7C). Note that Whi2, Ira1, Ira2, and Rim15 are all negative regulators of the RAS/cyclic adenosine monophosphate (cAMP)/protein kinase A (PKA) pathway, which, in response to glucose, stimulates population expansion (*36*–*39*). When glucose levels become limited, the RAS/cAMP/PKA pathway is repressed, thereby causing cells to stop dividing and enter stationary phase. Disruptive mutations in *WHI2*, *IRA1*, *IRA2*, or *RIM15* may cause a delayed response to low glucose levels that enables a few additional rounds of cell division before cells enter stationary phase and, thereby, lead to increased representation of these mutants after serial passaging under laboratory conditions. We constructed mixed populations consisting of a strain deleted for one of the frequently mutated genes and a wild-type strain and followed their ratio for six rounds of serial passaging under conditions with a relatively prolonged stationary phase (*21*). Indeed, the relative abundance of strains deleted for *WHI2*, *IRA2*, *RIM15*, or *UBC13* increased with each round of serial passaging, whereas five control mutant strains maintained abundances similar to or lower than the wild-type reference strain (Fig. 5D and fig. S7D). Similar results were obtained for *IRA1* and *IRA2* in another strain background, W303 (fig. S7, E and F). Thus, our data suggest that the vast majority of passenger mutations are random and not dependent on the query or suppressor mutation and, further, that a few additional secondary mutations arise at high frequency because of a selection for mutants that delay the onset of the stationary phase.

## Mechanistic categories of suppression interactions

We classified the suppression interactions into distinct mechanistic categories on the basis of the functional relationship between query and suppressor. Most queries (54%) reported in the literature or identified by our systematic analysis are suppressed by mutations in functionally related genes (class "A") (Fig. 6, A and B). These functional connections can be further divided into four subclasses. Subclass "A1" includes 135 interactions from the literature and systematic networks, in which both the query and the suppressor genes encode members of the same protein complex. These particular interactions can reflect a mechanism whereby the suppressor represents a gain-of-function mutation (fig. S2A). Subclass "A2," to which 201 interactions from our

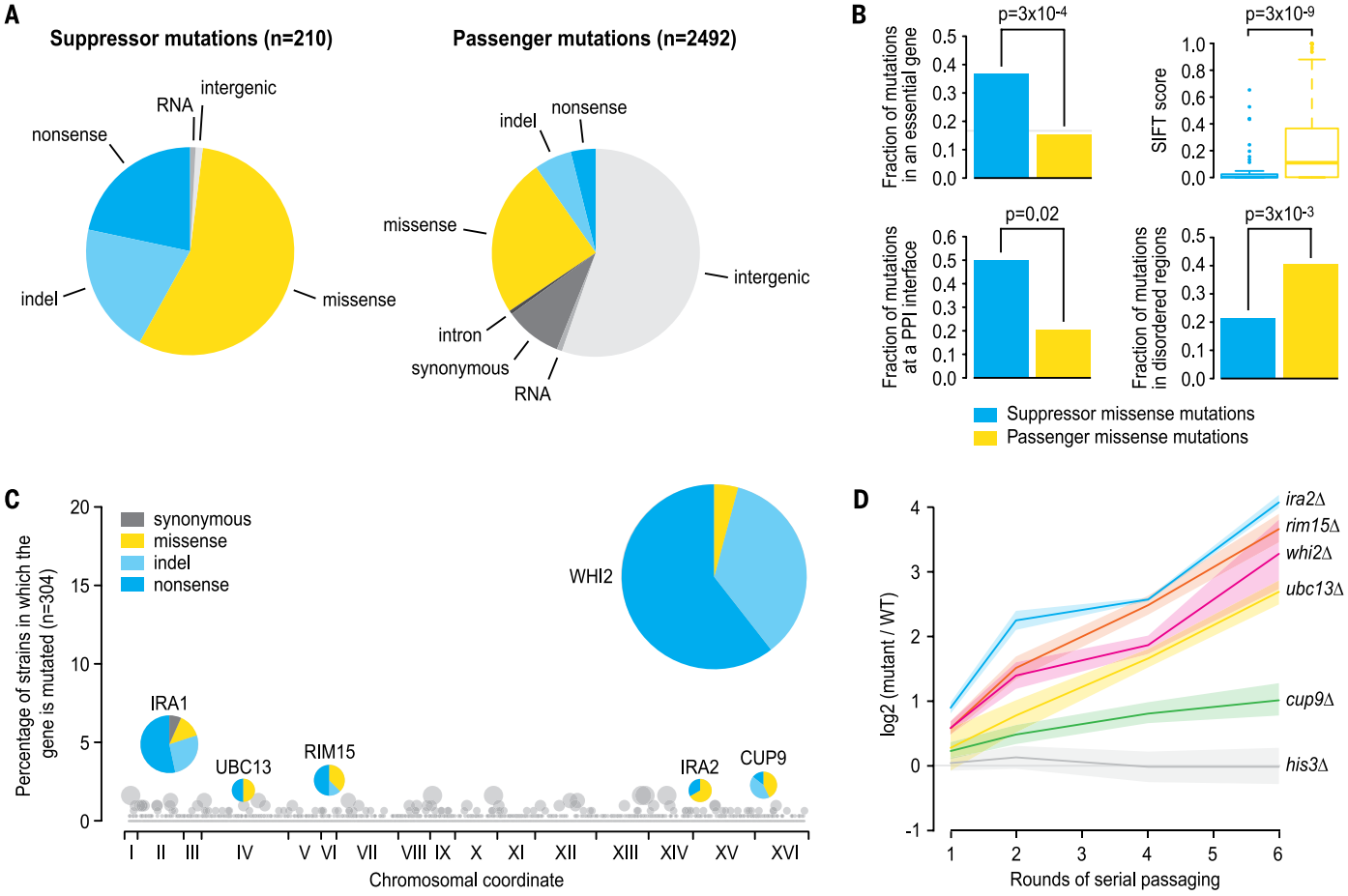



**Fig. 5. Characterization of potential passenger mutations.** (**A**) Distribution of suppressor and potential passenger mutations over variant effect classes. Only SNPs are considered, as reliable structural variant calls (deletions, insertions, or inversions involving >5 base pairs) were only available for suppressor mutations. The RNA class refers to mutations in an RNA species such as a noncoding, ribosomal, or transfer RNA. (**B**) The fraction of all suppressor or potential passenger missense mutations that map to an essential gene, at a protein-protein interaction (PPI) interface, or at a disordered region of a protein, and the predicted deleteriousness of these mutations (SIFT scores: 0 = extremely deleterious and 1 = benign). *P* values were calculated using Fisher's exact test, except for the SIFT analysis, in which a Mann-Whitney test was used. (**C**) The percentage of strains in which a particular gene carries a passenger mutation is plotted against the chromosomal position of the gene. Genes that are recurrently mutated in >2% of the sequenced strains are highlighted, and the distribution of the mutations over variant effect classes is shown. (**D**) Differentially fluorescently labeled cells of the indicated mutants [labeled with red fluorescent protein (RFP)] and wild type (GFP) were mixed, and the ratio of RFP to GFP was followed for six rounds of serial passaging on agar plates. Shading represents the SD, $n = 12$.

network were assigned, describes cases where the query mutant growth defect is suppressed by a mutation in a gene that is annotated to the same pathway. In the case of loss-of-function suppressor mutations, the suppressor gene often has antagonistic effects compared with the query gene (e.g., Fig. 4). Subclass "A3" involves suppression by a different, but related, pathway and explains 195 interactions in our networks. In this scenario, the growth phenotype caused by absence of a specific cellular function required for normal cell growth is suppressed when an alternative pathway is rewired to re-create the missing activity (e.g., Fig. 3). Finally, subclass "A4" consists of gene pairs that are annotated to the same biological process but for which pathway or complex annotation data were not available for both genes.

In addition to suppression interactions between functionally related genes, suppression interactions involving hypomorphic (partial loss-of-function) alleles—such as conditional TS alleles of essential genes—revealed a different and more general class of suppressors that affect expression of the query gene. This type of suppression (Figs. 2A and 3A) can be achieved by stabilizing a mutant mRNA or protein through the perturbation of pathways or complexes that regulate mRNA or protein turnover (Fig. 6A, class "B" and "C"). Although this type of suppression is rarely described in the literature, 48% of the hypomorphic queries in our experimental data set are suppressed by mutations in protein degradation or mRNA decay genes (Fig. 6B); this indicates that this type of allele-specific suppression is one of the main routes through which partial loss-of-function alleles can be suppressed. Of the suppression interactions, 60 to 70% fall into one of these mechanistic classes, as compared with only 34% of positive genetic interactions identified by SGA (*6*) and 11% of passenger-query pairs. Thus, positive genetic interactions that are true suppression interactions often show high functional and mechanistic specificity.

## Discussion

A global, literature-curated network of genetic suppression interactions (Fig. 1) showed that the majority of suppression interactions linked functionally related genes. Moreover, suppression interactions overlapped significantly with other types of genetic interactions (Fig. 2). Systematic suppression analysis confirmed these general properties of suppression and further showed that suppression of hypomorphic alleles often occurs via loss of protein or mRNA degradation, a finding that was less obvious in literature-curated data (Fig. 6). The underrepresentation of this class of interactions in the literature is consistent with what has been reported for dosage suppression interactions (*13*) and may reflect that mechanistic studies focused on the functional analysis of a particular gene or pathway are less likely to

report nonspecific suppressors. Nevertheless, an understanding of the prevalence of this form of suppression could be important when interpreting a genotype-to-phenotype relationship. Even though the genes encoding proteasome or mRNA-decay components are essential in human cell lines (*40*–*42*), we anticipate that genetic variation that subtly modulates the activity of these modules may exhibit genetic interactions associated with a decreased disease risk for a variety of human disorders. As in yeast, these processes may thus buffer a range of detrimental mutations in humans and, thereby, modify numerous different disease phenotypes.

Despite the prevalence of these general suppression mechanisms, most suppression gene pairs showed a close functional relationship (Fig. 6), so that genetic suppression can be used to assign function to a previously uncharacterized gene (Fig. 4). The suppressor interactions identified in our systematic screen resulted from the direct selection of spontaneous mutations during standard laboratory growth of a query mutant whose fitness was compromised. In total, ~3% of strains in the yeast nonessential deletion mutant collection and ~4% of the strains in the hypomorphic essential gene mutant collections showed evidence of a suppressor locus when screened by SGA. Whole-genome sequencing of 251 potential suppressor strains did not reveal any instances of suppression via aneuploidy, a mutational event involving copy number variation of many genes, possibly because aneuploidies are not necessarily revealed by SGA genetic mapping or because these events come at a fitness cost (*43*). Although SGA suppressor mapping can theoretically identify multiple suppressor mutations within one strain (*20*), no query strains with multiple suppressor linkages were identified. This suggests that the direct selection for spontaneous suppressors does not mimic adaptive evolution of wild-type strains in nutrient-limited conditions, in which aneuploidies and mutations in multiple genes, each contributing small fitness increases, combine to collectively produce a robust suppression phenotype (*35*, *44*). In contrast, we found that there is often a single direct suppression strategy because for most (~67%) of the queries for which we isolated several independent suppressor mutations, these recurrently occurred within the same single suppressor gene or within genes that encode subunits of the same complex. In addition, we found that large increases in fitness are mainly achieved by mutations in genes that have a close functional relation to the query gene (fig. S4). Thus, only a few, very specific mutational events appear to be able to substantially increase the fitness of a particular query mutant.

Besides the suppressor mutation, each strain also carried, on average, eight additional passenger mutations that did not have a measurable effect on exponential growth rate. In a previous, but relatively limited, study, it was suggested that the deletion of a query gene in the deletion mutant collection often selects for further genetic changes (*32*). Although this is true for suppressor

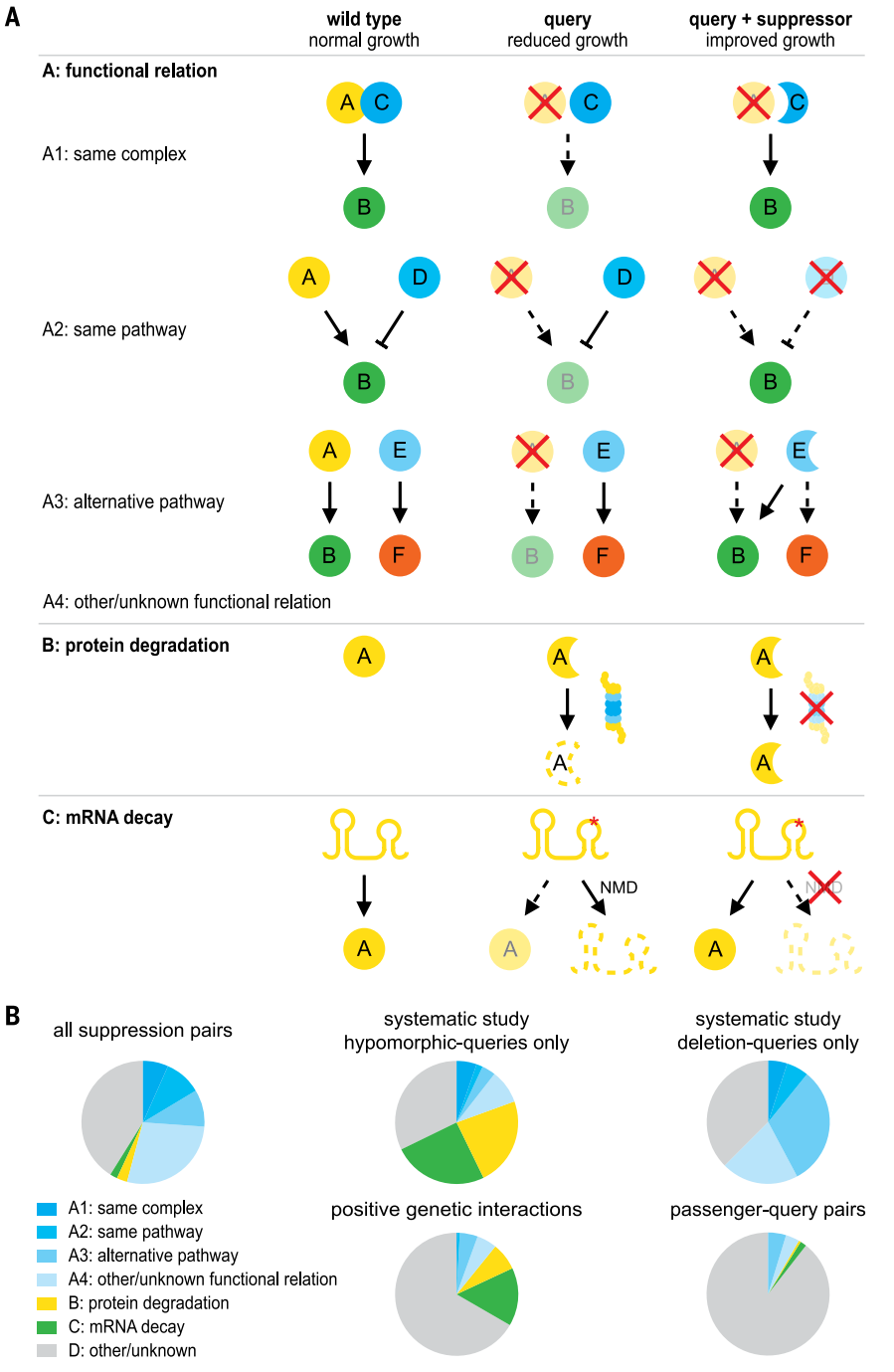



**Fig. 6. Mechanistic classes of suppression.** (**A**) Suppressor and query genes often have a functional relationship (class "A"). In a situation where the query (protein A) activates a protein B, which is required for normal growth, suppression can take place in multiple ways. For example, the suppressor (protein C) can be part of the same complex as the query, and gain-of-function mutations in C can restore the activation of B (class "A1"). Alternatively, the suppressor and query may be members of the same pathway, and the suppressor (protein D) may inactivate or inhibit B. Loss of D may thus suppress by partially restoring the activity of B (class "A2"). The suppressor (protein E) can also function in an alternative, but related, pathway, whose activity can be slightly altered to restore the activity of B (class "A3"). Suppression interactions can also occur among pairs of genes that do not share a close functional relationship. For example, partial loss-of-function query alleles may carry mutations that destabilize the protein or mRNA, leading to a fitness defect caused by reduced levels of the query protein. This can be suppressed by a loss-of-function mutation in a member of the protein degradation (class "B") or mRNA decay (class "C") pathway, which may partially restore the levels of the query protein. NMD, nonsense-mediated mRNA decay. (**B**) Distribution of suppression interactions, positive genetic interactions (*6*), and passenger-query pairs across different mechanistic suppression classes.

mutations, we could not find any substantial evidence connecting the query or suppressor mutation to the occurrence of most passenger mutations. Because we did not observe a significant enrichment for functionally related gene pairs among queries and passengers (fig. S7B), we conclude that the occurrence of query-driven non-suppressor mutations is likely rare.

In a mathematical model of bacterial serial passaging, de novo mutations that delay the onset of stationary phase were more likely to fix in a population than mutations that decrease lag time or increase growth or survival rates (*45*). This may be true for yeast as well, as the growth history of laboratory-grown yeast strains follows a similar pattern of repeating cycles of lag phase, exponential growth, and stationary phase. Indeed, we observe selection for mutations that likely delay the onset of stationary phase in 26% of the sequenced strains (Fig. 5C). These stationary-delay mutations are thus not true "passenger" mutations but are adaptive. However, in contrast to suppressor mutations that cause adaptation to the query mutation, the stationary-delay mutations are adaptive to laboratory passaging. These mutations could come at a cost, as they probably decrease viability during longer periods of starvation (*35*, *36*).

As most (78%) suppression interactions did not overlap with any previously identified genetic interactions, additional suppression mapping will help complete the yeast genetic interaction landscape. Conditional alleles have been developed for nearly all essential yeast genes (*6*), and thus, suppression interactions could be mapped for the full set of essential genes. Similarly, suppressors of nonessential genes could be identified in a conditional or synthetic lethal context in which the nonessential query has a fitness defect. Although we focused on mapping suppression interactions in yeast, similar suppression studies should be possible in mammalian cells and model systems and may identify new drug targets for query mutations related to human disease (*46*). As ~6% of human pathogenic variants are fixed in other mammalian species (*47*), compensatory mutations may be present at a high frequency in natural populations. Understanding genetic suppression may provide insight in how genetic variance accumulates during evolution and more specifically how modifier genes determine the severity of genetic traits, including human disease.

## Materials and methods

Detailed materials and methods are available in the supplementary information.

### *Literature curation*

The *Saccharomyces cerevisiae* "synthetic rescue" data set was downloaded from the BioGRID (*9*) on 9 November 2012 (version 3.1.49) and on 31 March 2014 (version 3.2.110). In total, these data sets consisted of 5985 interactions described in 1667 papers. Each paper was read in detail, and an interaction was considered a suppression interaction if the double mutant grew substantially better than at least one of the single mutants. For each interaction, suppressor and query allele type and specific conditions were annotated (*21*). The final data set consisted of 1842 unique interactions, involving 1304 genes (table S1).

### *Systematic suppressor identification*

All suppressor strains were part of either the BY4741 nonessential deletion mutant collection (*MAT***a** *xxxΔ::kanMX4 his3Δ1 leu2Δ0 ura3Δ0 met15Δ0;* Euroscarf), the SGA nonessential deletion mutant collection [*MATα xxxΔ::natMX4 can1Δ::Ste2pr-Sp_his5 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0;* (*2*)] or the corresponding *MAT***a** and *MATα* collections of DAmP or TS mutants of essential genes (*6*). The presence and genomic location of a spontaneous suppressor mutation were identified by the occurrence of a suppressor linkage group upon screening strains in these collections by SGA analysis (*20*) (table S3). Potential suppressor strains were subsequently sequenced whole-genome on the Illumina HiSeq 2500 platform using paired-end 100-bp reads. Read mapping and single-nucleotide polymorphism (SNP), as well as indel calling were performed by using standard methods (*21*). Candidate suppressor mutations were confirmed by amplifying the corresponding gene and flanking sequences by polymerase chain reaction, followed by Sanger sequencing (table S2). Suppression interactions were confirmed using plasmid-based complementation assays and tetrad analysis of meiotic progeny derived from crossing each suppressor strain to either a wild-type strain, a strain with a marked deletion that was genetically linked to the candidate suppressor, or a strain carrying a deletion or hypomorphic allele of the suppressor gene (table S2).

## REFERENCES AND NOTES


1. S. J. Dixon, M. Costanzo, A. Baryshnikova, B. Andrews, C. Boone, Systematic mapping of genetic interaction networks. *Annu. Rev. Genet.* **43**, 601–625 (2009). doi: 10.1146/annurev.genet.39.073003.114751; pmid: 19712041
2. M. Costanzo *et al.*, The genetic landscape of a cell. *Science* **327**, 425–431 (2010). doi: 10.1126/science.1180823; pmid: 20093466
3. R. Chen *et al.*, Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat. Biotechnol.* **34**, 531–538 (2016). doi: 10.1038/nbt.3514; pmid: 27065010
4. R. P. St Onge *et al.*, Systematic pathway analysis using high-resolution fitness profiling of combinatorial gene deletions. *Nat. Genet.* **39**, 199–206 (2007). doi: 10.1038/ng1948; pmid: 17206143
5. D. K. Breslow *et al.*, A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. *Nat. Methods* **5**, 711–718 (2008). doi: 10.1038/nmeth.1234; pmid: 18622397
6. M. Costanzo *et al.*, A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016). doi: 10.1126/science.aaf1420; pmid: 27708008
7. A. Baryshnikova *et al.*, Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods* **7**, 1017–1024 (2010). doi: 10.1038/nmeth.1534; pmid: 21076421
8. D. Botstein, *Decoding the Language of Genetics* (Cold Spring Harbor Laboratory Press, 2015).
9. C. Stark *et al.*, BioGRID: A general repository for interaction datasets. *Nucleic Acids Res.* **34**, D535–D539 (2006). doi: 10.1093/nar/gkj109; pmid: 16381927
10. P. Shannon *et al.*, Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003). doi: 10.1101/gr.1239303; pmid: 14597658
11. M. Valencia-Burton *et al.*, Different mating-type-regulated genes affect the DNA repair defects of *Saccharomyces* RAD51, RAD52 and RAD55 mutants. *Genetics* **174**, 41–55 (2006). doi: 10.1534/genetics.106.058685; pmid: 16782999
12. L. Peiró-Chova, F. Estruch, Specific defects in different transcription complexes compensate for the requirement of the negative cofactor 2 repressor in *Saccharomyces cerevisiae*. *Genetics* **176**, 125–138 (2007). doi: 10.1534/genetics.106.066829; pmid: 17339209
13. L. Magtanong *et al.*, Dosage suppression genetic interaction networks enhance functional wiring diagrams of the cell. *Nat. Biotechnol.* **29**, 505–511 (2011). doi: 10.1038/nbt.1855; pmid: 21572441
14. E. L. Huttlin *et al.*, The BioPlex network: A systematic exploration of the human interactome. *Cell* **162**, 425–440 (2015). doi: 10.1016/j.cell.2015.06.043; pmid: 26186194
15. T. Rolland *et al.*, A proteome-scale map of the human interactome network. *Cell* **159**, 1212–1226 (2014). doi: 10.1016/j.cell.2014.10.050; pmid: 25416956
16. R. Sopko *et al.*, Mapping pathways and phenotypes by systematic gene overexpression. *Mol. Cell* **21**, 319–330 (2006). doi: 10.1016/j.molcel.2005.12.011; pmid: 16455487
17. M. E. Cusick *et al.*, Literature-curated protein interaction datasets. *Nat. Methods* **6**, 39–46 (2009). doi: 10.1038/nmeth.1284; pmid: 19116613
18. A. H. Tong *et al.*, Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001). doi: 10.1126/science.1065810; pmid: 11743205
19. A. H. Tong *et al.*, Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813 (2004). doi: 10.1126/science.1091317; pmid: 14764870
20. P. Jorgensen *et al.*, High-resolution genetic mapping with ordered arrays of *Saccharomyces cerevisiae* deletion mutants. *Genetics* **162**, 1091–1099 (2002). pmid: 12454058
21. Detailed materials and methods are available as supplementary materials in *Science* Online.
22. M. Schuldiner *et al.*, Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005). doi: 10.1016/j.cell.2005.08.031; pmid: 16269340
23. C. D. Dunn, R. E. Jensen, Suppression of a defect in mitochondrial protein import identifies cytosolic proteins required for viability of yeast cells lacking mitochondrial DNA. *Genetics* **165**, 35–45 (2003). pmid: 14504216
24. G. D. Clark-Walker, Kinetic properties of F1-ATPase influence the ability of yeasts to grow in anoxia or absence of mtDNA. *Mitochondrion* **2**, 257–265 (2003). doi: 10.1016/S1567-7249(02)00107-1; pmid: 16120326
25. A. Jézégou *et al.*, Heptahelical protein PQLC2 is a lysosomal cationic amino acid exporter underlying the action of cysteamine in cystinosis therapy. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E3434–E3443 (2012). doi: 10.1073/pnas.1211198109; pmid: 23169667
26. A. K. Gillingham, J. R. Whyte, B. Panic, S. Munro, Mon2, a relative of large Arf exchange factors, recruits Dop1 to the Golgi apparatus. *J. Biol. Chem.* **281**, 2273–2280 (2006). doi: 10.1074/jbc.M510176200; pmid: 16301316
27. S. Barbosa, D. Pratte, H. Schwarz, R. Pipkorn, B. Singer-Krüger, Oligomeric Dop1p is part of the endosomal Neo1p-Ysl2p-Arl1p membrane remodeling complex. *Traffic* **11**, 1092–1106 (2010). doi: 10.1111/j.1600-0854.2010.01079.x; pmid: 20477991
28. M. Takar, Y. Wu, T. R. Graham, The essential Neo1 protein from budding yeast plays a role in establishing aminophospholipid asymmetry of the plasma membrane. *J. Biol. Chem.* **291**, 15727–15739 (2016). doi: 10.1074/jbc.M115.686253; pmid: 27235400
29. A. B. Parsons *et al.*, Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell* **126**, 611–625 (2006). doi: 10.1016/j.cell.2006.06.040; pmid: 16901791
30. K. Iwamoto *et al.*, Curvature-dependent recognition of ethanolamine phospholipids by duramycin and cinnamycin. *Biophys. J.* **93**, 1608–1619 (2007). doi: 10.1529/biophysj.106.101584; pmid: 17483159
31. G. D. Fairn, M. Hermansson, P. Somerharju, S. Grinstein, Phosphatidylserine is polarized and required for proper Cdc42 localization and for development of cell polarity.

*Nat. Cell Biol.* **13**, 1424–1430 (2011). doi: 10.1038/ncb2351; pmid: 21964439
32. X. Teng *et al.*, Genome-wide consequences of deleting any single gene. *Mol. Cell* **52**, 485–494 (2013). doi: 10.1016/j.molcel.2013.09.026; pmid: 24211263
33. G. I. Lang *et al.*, Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* **500**, 571–574 (2013). doi: 10.1038/nature12344; pmid: 23873039
34. S. Kryazhimskiy, D. P. Rice, E. R. Jerison, M. M. Desai, Global epistasis makes adaptation predictable despite sequence-level stochasticity. *Science* **344**, 1519–1522 (2014). doi: 10.1126/science.1250939; pmid: 24970088
35. D. J. Kvitek, G. Sherlock, Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment. *PLOS Genet.* **9**, e1003972 (2013). doi: 10.1371/journal.pgen.1003972; pmid: 24278038
36. E. Cameroni, N. Hulo, J. Roosen, J. Winderickx, C. De Virgilio, The novel yeast PAS kinase Rim 15 orchestrates G0-associated antioxidant defense mechanisms. *Cell Cycle* **3**, 462–468 (2004). doi: 10.4161/cc.3.4.791; pmid: 15300954
37. A. Smith, M. P. Ward, S. Garrett, Yeast PKA represses Msn2p/Msn4p-dependent gene expression to regulate growth, stress response and glycogen accumulation. *EMBO J.* **17**, 3556–3564 (1998). doi: 10.1093/emboj/17.13.3556; pmid: 9649426
38. D. Kaida, H. Yashiroda, A. Toh-e, Y. Kikuchi, Yeast Whi2 and Psr1-phosphatase form a complex and regulate STRE-mediated gene expression. *Genes Cells* **7**, 543–552 (2002). doi: 10.1046/j.1365-2443.2002.00538.x; pmid: 12090248
39. K. Tanaka *et al.*, S. cerevisiae genes *IRA1* and *IRA2* encode proteins that may be functionally equivalent to mammalian *ras* GTPase activating protein. *Cell* **60**, 803–807 (1990). doi: 10.1016/0092-8674(90)90094-U; pmid: 2178777
40. T. Hart *et al.*, High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015). doi: 10.1016/j.cell.2015.11.015; pmid: 26627737
41. T. Wang *et al.*, Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015). doi: 10.1126/science.aac7041; pmid: 26472758
42. V. A. Blomen *et al.*, Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092–1096 (2015). doi: 10.1126/science.aac7557; pmid: 26472760
43. E. M. Torres *et al.*, Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317**, 916–924 (2007). doi: 10.1126/science.1142210; pmid: 17702937
44. A. B. Sunshine *et al.*, The fitness consequences of aneuploidy are driven by condition-dependent gene effects. *PLOS Biol.* **13**, e1002155 (2015). doi: 10.1371/journal.pbio.1002155; pmid: 26011532
45. L. M. Wahl, A. D. Zhu, Survival probability of beneficial mutations in bacterial batch culture. *Genetics* **200**, 309–320 (2015). doi: 10.1534/genetics.114.172890; pmid: 25758382
46. C. M. Buchovecky *et al.*, A suppressor screen in Mecp2 mutant mice implicates cholesterol metabolism in Rett syndrome. *Nat. Genet.* **45**, 1013–1020 (2013). doi: 10.1038/ng.2714; pmid: 23892605
47. D. M. Jordan *et al.*, Identification of cis-suppression of human disease mutations by comparative genomics. *Nature* **524**, 225–229 (2015). doi: 10.1038/nature14497; pmid: 26123021

## ACKNOWLEDGMENTS

We thank G. Fairn, C. Dunn, N. Pascoe, and E. Styles for reagents and technical assistance. This work was primarily supported by grants from the NIH (R01HG005853, R01HG005084, P50HG004233, and U01HG001715) (C.B., B.J.A., C.L.M., and F.P.R.); the Canadian Institutes of Health Research (FDN-143264 and FDN-143265) (C.B. and B.J.A.); the Ontario Research Fund (Research Excellence Grant) (F.P.R.); the Canada Excellence Research Chairs Program (F.P.R.); and a postdoctoral fellowship from the Canadian Institutes of Health Research (J.v.L.). Additional support was provided by the Canadian Institutes of Health Research (FDN-143301) (A.-C.G.), the Vienna Science and Technology Fund (LS08-QM03) (C.O. and M.P.), and the NIH (R01GM107978) (T.R.G.). C.B., B.J.A., F.P.R., and C.L.M. are Senior Fellows in the Canadian Institute for Advanced Research Genetic Networks program. All whole-genome sequencing data are publicly available at NCBI's Sequencing Read Archive, under accession number SRP067030. All suppression interaction data are included in tables S1 and S2.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/aag0839/suppl/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S7
References (*48*–*62*)

9 May 2016; accepted 4 October 2016
10.1126/science.aag0839

REPORTS

OPTICS

# Microresonator soliton dual-comb spectroscopy

Myoung-Gyun Suh,* Qi-Fan Yang,* Ki Youl Yang, Xu Yi, Kerry J. Vahala†

**Measurement of optical and vibrational spectra with high resolution provides a way to identify chemical species in cluttered environments and is of general importance in many fields. Dual-comb spectroscopy has emerged as a powerful approach for acquiring nearly instantaneous Raman and optical spectra with unprecedented resolution. Spectra are generated directly in the electrical domain, without the need for bulky mechanical spectrometers. We demonstrate a miniature soliton-based dual-comb system that can potentially transfer the approach to a chip platform. These devices achieve high-coherence pulsed mode locking. They also feature broad, reproducible spectral envelopes, an essential feature for dual-comb spectroscopy. Our work shows the potential for integrated spectroscopy with high signal-to-noise ratios and fast acquisition rates.**

Since their demonstration in the late 1990s (*1*–*3*), optical frequency combs have revolutionized precision measurements of time and frequency and have enabled new technologies such as optical clocks (*3*). One remarkable method they make possible is dual-comb spectroscopy, which leverages the coherence properties of combs for rapid broadband spectral analysis with high accuracy (*4*–*10*). Frequency comb systems exist across a broad spectral range spanning ultraviolet to infrared, making this method well suited for measurement of diverse molecular species (*10*).

In parallel with advancements in frequency comb applications, the past decade has witnessed the appearance of miniature optical frequency combs or microcombs (*11*, *12*). These microcombs have been demonstrated using several dielectric materials across a range of emission bands (*11*, *13*–*17*). Under continuous-wave laser pumping, the combs are initiated by way of parametric oscillation (*18*, *19*) and are broadened by cascaded four-wave mixing (*11*, *12*) to spectral widths that can encompass an octave of spectrum (*15*). Four-wave mixing in the ultrafast intraband gain medium of quantum cascade lasers has also been shown to create frequency modulation (FM) combs (*20*). These FM systems have been applied to demonstrate dual-comb spectroscopy in the mid-infrared (*21*). Also, heterodyne of two conventional microcombs in the mid-infrared has been demonstrated, a key step toward dual-comb spectroscopy (*22*).

A major advancement in microcombs has been the realization of soliton mode locking (*23*–*27*). Soliton microcombs feature dissipative Kerr solitons that leverage the Kerr nonlinearity to both compensate dispersion and overcome cavity loss by way of parametric gain (*23*). Unlike earlier microcombs, this new device provides phase-locked femtosecond pulses with well-defined, repeatable spectral envelopes, which is important for dual-comb spectroscopy. The pulse repetition rate of the device is detectable, and it has excellent phase noise characteristics (*24*). In this work, we demonstrate dual-comb spectroscopy using this new platform. The dual-comb source spans >30 nm and the interferogram spectra feature high signal-to-noise ratios (SNRs). Also, precise microfabrication control enables close matching of the repetition rates so that >4 THz of optical bandwidth is measured within 500 MHz of electrical bandwidth.

A schematic view of the dual-comb experimental setup (Fig. 1) shows two soliton trains having different repetition rates ($\Delta f_r = f_{r1} - f_{r2}$) generated from distinct microresonators and then combined using a directional coupler. One of the combined streams is coupled through a gas cell of molecules (test sample in the figure) whose absorption spectrum is to be measured. The other combined stream provides a reference. The slight difference in repetition rates of the soliton streams creates a periodically time-varying interferogram in the detected current with a period $1/\Delta f_r$. Fourier transform of this time-varying signal reveals the interfering soliton comb spectra, now shifted to radio-frequency rates. The signal spectrum containing the molecular absorption information is then normalized using the reference spectrum to reveal the spectral absorption of the gas cell. [See (*28*) for details of the experimental setup, methods, and equipment.]

Solitons are generated and stabilized in two microresonators by means of the active-capture/locking technique (fig. S1) (*24*, *28*). The locking makes it possible for the two combs to remain stable indefinitely. Typical soliton optical spectra (Fig. 2, A and B) feature the characteristic $\mathrm{sech}^2$ envelope, observed in this case over a 60-nm wavelength span. The detected electrical spectrum for each soliton source was measured using a spectrum analyzer with a bandwidth of 26 GHz (Fig. 2, C and D). The narrow spectral lines (resolution bandwidth of 500 Hz) have SNRs greater than 75 dB, which shows that the corresponding repetition rates are extremely stable. The high-Q resonators used in this work are silica whispering-gallery devices fabricated on a silicon wafer by a combination of lithography and wet/dry etching (*29*). The unloaded quality factor of the microresonators is approximately 300 million, and the solitons have repetition rates of approximately 22 GHz that were determined primarily by the diameter of the devices (3 mm).

The optical outputs from the stabilized soliton sources are combined and coupled into two paths (Fig. 1 and fig. S1). One path contains a $H^{13}CN$ gas cell (*28*), which functions as the test sample in the

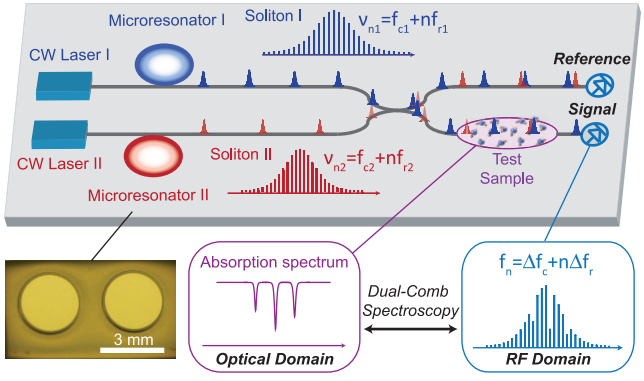



**Fig. 1. Microresonator-based dual-comb spectroscopy.** Two soliton pulse trains with slightly different repetition rates are generated by continuous-wave (CW) laser pumping of two microresonators. The pulse trains are combined in a fiber bidirectional coupler to produce a signal output path that passes through a test sample as well as a reference output path. The output of each path is detected (on Reference and Signal photo detectors) to generate an electrical interferogram of the two soliton pulse trains. The interferogram is Fourier-transformed to produce comb-like radio-frequency (RF) electrical spectra having spectral lines spaced by the repetition rate difference of the soliton pulse trains. The absorption features of the test sample can be extracted from this spectrum by normalizing the signal spectrum by the reference spectrum. Also shown are two silica disk resonators. The disks have a diameter of 3 mm and are fabricated on a silicon chip. The *n*th optical comb frequency ($\nu_{n1}$ and $\nu_{n2}$) for each soliton pulse train is given in terms of the respective repetition rate ($f_{r1}$ and $f_{r2}$) plus an offset frequency ($f_{c1}$ and $f_{c2}$). In the RF domain spectrum, the *n*th line occurs at a multiple of the difference in the repetition rates ($\Delta f_r = f_{r2} - f_{r1}$) plus an offset frequency ($\Delta f_c = f_{c2} - f_{c1}$).

T. J. Watson Laboratory of Applied Physics, California Institute of Technology, Pasadena, CA 91125, USA.
*These authors contributed equally to this work.
†Corresponding author. E-mail: vahala@caltech.edu

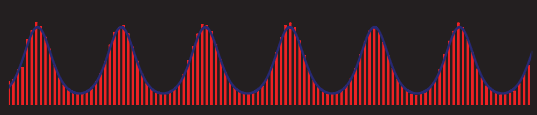


**Fig. 3. Measured el**
nterferogram of th
ectrum obtained b
pectra each are recorded over a time of 20 μs and averaged to obtain
layed spectra. (**C**) Resolved (multiple and individual) comb lines of
rum in (B) are equidistantly separated by 2.6 MHz, the difference
repetition rate of the two microresonators. The linewidth of each
absorber (WaveShaper; see fig. S1) w
(spectrally flat and sine-wave). The
(red) are compared with the pro
to 1565 nm.

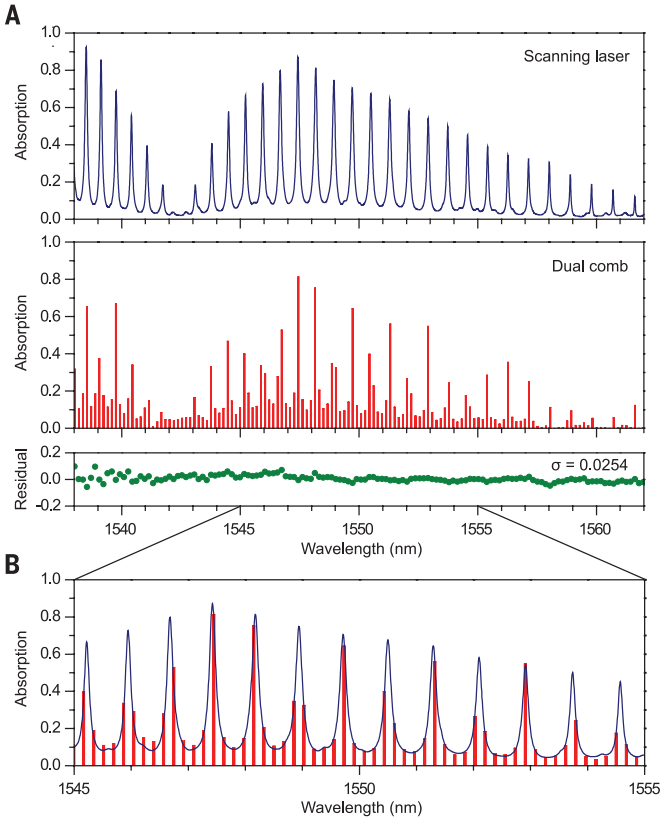



**Fig. 4. Measured molecular absorption spectra.** (**A**) Absorption spectrum of $2\nu_3$ band of $H^{13}CN$ measured by direct power transmission using a wavelength-calibrated scanning laser and comparison to the microresonator-based dual-comb spectrum. The residual difference between the two spectra is shown in green. (**B**) Overlay of the directly measured optical spectrum and the dual-comb spectrum, showing line-by-line matching. The vertical positions of the two spectra are adjusted to compensate for insertion loss.

measurement. The other path is coupled directly to a photodetector and functions as the reference. The test sample path also includes an alternate path (shown in fig. S1) in which a WaveShaper is inserted (*28*). Temperature control of one of the microresonators is used to tune the relative optical frequency difference of the two soliton streams. In our measurements, this difference was held below 1 GHz, allowing the observation of the temporal interferogram on an oscilloscope (bandwidth 1 GHz).

The reference interferogram (Fig. 3A) has a period of 386 ns, corresponding to a soliton repetition rate difference of 2.6 MHz. This relatively small repetition rate difference was made possible by precise control of the resonator diameter (and hence the resonator free spectral range) using calibrated wet etching of the silica (*29*). It was possible to fabricate disks with even more closely matched repetition rates (<100 kHz). Figure 3B shows the calculated Fourier transform of the interferogram. The small repetition rate difference on the much larger 22-GHz soliton repetition rate makes it possible to compress an optical span of 4 THz (1535 to 1567 nm) into 500 MHz of electrical spectrum. The measured wavelength span is actually narrower than the observable wavelength span of the original soliton pulse streams and is limited by the photodetector noise. The interferogram spectrum has a SNR around 30 dB near the central lines. A zoom-in of the spectrum (multi- and single-line) is shown in Fig. 3C. The electrical comb lines are equidistantly separated by 2.6 MHz and have a linewidth (full width at half maximum) of <50 kHz, limited by the mutual coherence of the independent fiber pump lasers. The pump laser line in a dissipative Kerr soliton is also a comb tooth in the soliton optical spectrum. As a result, the frequency jitter in each pump is transferred as an overall shift on the resulting soliton comb. Externally locking the two combs should reduce the observed linewidth in the interferogram spectrum.

The pump lines are placed toward the high-frequency side (near 550 MHz) of the spectral maximum in the interferogram spectrum (see Fig. 3B). In the optical spectra (Fig. 2, A and B), the pump is blue-detuned relative to the soliton spectral maximum [this occurs on account of the Raman self-frequency shift of the soliton (*24*)]. This spectral landmark shows that the relative spectral placement of the soliton combs is such that high optical frequencies are mapped to high interferogram frequencies. The way in which certain nonidealities in the soliton spectra map into the interferogram spectrum is also of interest. Specifically, there are avoided mode crossing–induced Fano-like spurs (*24*) in the soliton optical spectra (Fig. 2, A and B) occurring near 1535 nm, and this generates a corresponding feature at 750 MHz in Fig. 3B.

As an initial test of the dual-comb source, synthetic absorption spectra were programmed in a WaveShaper and then measured as dual-comb spectra (*28*). In Fig. 3, D and E, Fourier transforms of the signal interferograms produced using the synthetic absorber are shown. The two programmed functions are a spectrally flat 3-dB absorption and a sine-wave absorption having a 4-dB amplitude. The synthetic absorption spectra, obtained by normalizing the signal and reference electrical spectra, are compared with the programmed functions in Fig. 3, D and E. The ability to reconstruct these synthetic spectral profiles clearly demonstrates the reproducibility of the soliton spectral profile.

Finally, we studied the absorption spectrum of the $H^{13}CN$ $2\nu_3$ band. In Fig. 4A, the measured dual-comb absorption spectrum from 1538 to 1562 nm is shown in red and compared with a directly measured absorption spectrum shown in blue. Both absorption spectra are normalized. A laser locked to a molecular absorption line enabled absolute frequency calibration of the spectra (*28*). Sampling-induced choppiness of the dual-comb spectrum is caused by the relatively coarse spectral resolution of the solitons in comparison to the spectral scale of the $H^{13}CN$ absorption lines. Nonetheless, the characteristic envelope of the $H^{13}CN$ $2\nu_3$ band is clearly resolved. The residual difference between the two absorption spectra is shown in green in Fig. 4A; the calculated standard deviation is 0.0254. Furthermore, a line-by-line overlay of the measured optical and dual-comb spectra (Fig. 4B) confirms the wavelength precision and absorption intensity accuracy of the dual-comb source. The directly measured $H^{13}CN$ absorption spectrum is obtained by coupling an external cavity diode laser (ECDL) into the $H^{13}CN$ gas cell and then scanning the laser while monitoring the transmitted optical power. A separate signal is also tapped from the ECDL to function as a reference (*28*).

In principle, a soliton source with finer comb spacing (i.e., lower repetition frequency) is possible. Nonsoliton microcombs with mode spacings as narrow as 2.4 GHz have been demonstrated using the silica resonator platform (*16*). The use of electro-optical modulators to modulate the microcombs by a fraction of the repetition frequency is another possible way to create a finer spectral comb grid. Also, tuning of the combs is, in principle, possible so as to create a nearly continuous high-resolution absorption measurement.

The dual-comb source is centered near 1550 nm (in the C-band) for this work; however, operation at other wavelengths within the transmission window of silica is also possible. Moreover, using fiber nonlinear broadening or internal (resonator) dispersive wave generation (*25*), it should be possible to greatly extend the comb spectral span. Resonator dispersion engineering (*30*) can also be used to extend comb bandwidth. More generally, a wide range of materials are available for microcombs enabling access to mid-infrared spectra. With further improvements, it should also be possible to realize chip-based dual-comb coherent anti-Stokes Raman spectroscopy (CARS). The potential for monolithic integration with other devices makes soliton-based microcombs well suited for possible realization of a dual-comb spectroscopic system-on-a-chip.

## REFERENCES AND NOTES

1. D. J. Jones *et al.*, *Science* **288**, 635–640 (2000).
2. R. Holzwarth *et al.*, *Phys. Rev. Lett.* **85**, 2264–2267 (2000).
3. S. A. Diddams *et al.*, *Science* **293**, 825–828 (2001).
4. S. Schiller, *Opt. Lett.* **27**, 766–768 (2002).
5. F. Keilmann, C. Gohle, R. Holzwarth, *Opt. Lett.* **29**, 1542–1544 (2004).

6. I. Coddington, W. C. Swann, N. R. Newbury, *Phys. Rev. Lett.* **100**, 013902 (2008).
7. P. Giaccari, J.-D. Deschênes, P. Saucier, J. Genest, P. Tremblay, *Opt. Express* **16**, 4347–4365 (2008).
8. B. Bernhardt *et al.*, *Nat. Photonics* **4**, 55–57 (2010).
9. T. Ideguchi, A. Poisson, G. Guelachvili, N. Picqué, T. W. Hänsch, *Nat. Commun.* **5**, 3375 (2014).
10. I. Coddington, N. Newbury, W. Swann, *Optica* **3**, 414 (2016).
11. P. Del'Haye *et al.*, *Nature* **450**, 1214–1217 (2007).
12. T. J. Kippenberg, R. Holzwarth, S. A. Diddams, *Science* **332**, 555–559 (2011).
13. I. S. Grudinin, N. Yu, L. Maleki, *Opt. Lett.* **34**, 878–880 (2009).
14. S. B. Papp, S. A. Diddams, *Phys. Rev. A* **84**, 053833 (2011).
15. Y. Okawachi *et al.*, *Opt. Lett.* **36**, 3398–3400 (2011).
16. J. Li, H. Lee, T. Chen, K. J. Vahala, *Phys. Rev. Lett.* **109**, 233901 (2012).
17. B. Hausmann, I. Bulu, V. Venkataraman, P. Deotare, M. Lončar, *Nat. Photonics* **8**, 369–374 (2014).
18. T. J. Kippenberg, S. M. Spillane, K. J. Vahala, *Phys. Rev. Lett.* **93**, 083904 (2004).
19. A. A. Savchenkov *et al.*, *Phys. Rev. Lett.* **93**, 243905 (2004).
20. A. Hugi, G. Villares, S. Blaser, H. C. Liu, J. Faist, *Nature* **492**, 229–233 (2012).
21. G. Villares, A. Hugi, S. Blaser, J. Faist, *Nat. Commun.* **5**, 5192 (2014).
22. M. Yu, Y. Okawachi, A. Griffith, M. Lipson, A. L. Gaeta, in *CLEO: Science and Innovations 2016* (Optical Society of America, 2016), paper JTh4B.5.
23. T. Herr *et al.*, *Nat. Photonics* **8**, 145–152 (2014).
24. X. Yi, Q.-F. Yang, K. Y. Yang, M.-G. Suh, K. Vahala, *Optica* **2**, 1078 (2015).
25. V. Brasch *et al.*, *Science* **351**, 357–360 (2016).
26. P.-H. Wang *et al.*, *Opt. Express* **24**, 10890–10897 (2016).
27. C. Joshi *et al.*, *Opt. Lett.* **41**, 2565–2568 (2016).
28. See supplementary materials on *Science* Online.
29. H. Lee *et al.*, *Nat. Photonics* **6**, 369–373 (2012).
30. K. Y. Yang *et al.*, *Nat. Photonics* **10**, 316–320 (2016).



**ACKNOWLEDGMENTS**

We thank N. Newbury and G. Scalari for helpful comments on this manuscript. Supported by the Defense Advanced Research Projects Agency under the PULSE and SCOUT programs, NASA, and the Kavli Nanoscience Institute.


**SUPPLEMENTARY MATERIALS**

www.sciencemag.org/content/354/6312/600/suppl/DC1
Materials and Methods
Supplementary Text
Fig. S1
References (*31*–*34*)


27 July 2016; accepted 29 September 2016
Published online 13 October 2016
10.1126/science.aah6516


OPTICAL PROCESSING

# A coherent Ising machine for 2000-node optimization problems


**Takahiro Inagaki,[1]* Yoshitaka Haribara,[2,3,4] Koji Igarashi,[5] Tomohiro Sonobe,[4,6] Shuhei Tamate,[4] Toshimori Honjo,[1] Alireza Marandi,[7] Peter L. McMahon,[7] Takeshi Umeki,[8] Koji Enbutsu,[8] Osamu Tadanaga,[8] Hirokazu Takenouchi,[8] Kazuyuki Aihara,[2,3] Ken-ichi Kawarabayashi,[4,6] Kyo Inoue,[5] Shoko Utsunomiya,[4] Hiroki Takesue[1]***



**The analysis and optimization of complex systems can be reduced to mathematical problems collectively known as combinatorial optimization. Many such problems can be mapped onto ground-state search problems of the Ising model, and various artificial spin systems are now emerging as promising approaches. However, physical Ising machines have suffered from limited numbers of spin-spin couplings because of implementations based on localized spins, resulting in severe scalability problems. We report a 2000-spin network with all-to-all spin-spin couplings. Using a measurement and feedback scheme, we coupled time-multiplexed degenerate optical parametric oscillators to implement maximum cut problems on arbitrary graph topologies with up to 2000 nodes. Our coherent Ising machine outperformed simulated annealing in terms of accuracy and computation time for a 2000-node complete graph.**


Combinatorial optimization is an important task for various applications such as drug discovery, resource optimization in wireless networks, and machine learning (*1*). Many such problems are known to be nondeterministic polynomial time (NP)–hard or NP-complete problems, which are considered difficult to solve efficiently with modern digital computers. Several heuristic or "greedy" algorithms have been proposed in the field of computer science to find the optimal or near-optimal solutions to those problems. Among such algorithms, simulated annealing (SA) has performed well in solving various types of combinatorial optimization problems (*2*). It is known that combinatorial optimization problems can be mapped onto ground-state search problems of the Ising model with polynomial resources (*3*). Several approaches have been proposed and demonstrated to find solutions to Ising problems using networks of artificial spins based on various physical systems (*4*–*7*).

A coherent Ising machine (CIM) is one such artificial Ising machine, in which a laser (*8*) or a degenerate optical parametric oscillator (DOPO) (*9*) represents an artificial spin. This proposal was motivated by early studies on the quantum behaviors of a DOPO (*10*–*14*), and proof-of-principle CIM experiments have already been reported using time-multiplexed DOPOs (*15*, *16*). To date, physical Ising machines (*4*–*7*) have suffered from a limited number of spin-spin couplings due to localized spin implementations. For example, in a recent quantum annealing system based on 1152 superconducting qubits, the number of spin-spin couplings was 3360, constrained by the chimera graph structure (*17*). This limited number of spin-spin couplings has made it difficult to map various real-world optimization problems to the hardware. Here, we report a CIM that consists of 2048 DOPOs with full spin-spin couplings. A measurement and feedback (MFB) scheme enabled us to implement all possible connections among 2048 spins, which amounts to 2,096,128 spin-spin couplings, making it possible to solve arbitrary graph problems.

An Ising Hamiltonian without an external magnetic field is given by

$$H(\sigma) = -\sum_{i<j} J_{ij}\sigma_i\sigma_j \quad (1 \le i, j \le N) \qquad (1)$$

(*18*), where $\sigma_i \in \{-1,+1\}$ denotes the $i$th Ising spin, $J_{ij}$ is the interaction strength between the $i$th and $j$th spins, and $N$ is the total number of spins. The Ising machines described above are designed to find the ground state or low-energy states of the Ising Hamiltonian. A DOPO takes only a 0 or $\pi$ phase relative to the pump phase above the threshold, so it can be used as a stable artificial spin realized with photonics technologies (*19*–*22*). A CIM solves the Ising problem according to the minimum-gain principle (*8*). According to this principle, if the optical couplings between the DOPOs are adjusted to implement $\{J_{ij}\}$, the DOPO network oscillates in the phase configuration with the lowest loss, which corresponds to the ground state of the Ising Hamiltonian in Eq. 1.

The size of the problem that a CIM can handle is determined by the number of DOPOs and their couplings. Delayed interferometers were used to couple the DOPOs in the previous CIM experiments (*15*, *16*); however, this scheme does not scale as the problems become more complex. To avoid this issue, Haribara *et al.* proposed an MFB scheme that makes it possible to couple any combination of $N$ DOPOs (*23*).

Figure 1 shows a schematic of a CIM with MFB (*24*). The CIM uses a periodically poled lithium niobate (PPLN) waveguide, which is placed in a


[1]NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan. [2]Department of Mathematical Informatics, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo 113-8656, Japan. [3]Institute of Industrial Science, University of Tokyo, 4-6-1, Komaba, Meguro-ku, Tokyo 153-8505, Japan. [4]National Institute of Informatics, Hitotsubashi 2-1-2, Chiyoda-ku, Tokyo 101-8403, Japan. [5]Division of Electrical, Electronic and Information Engineering, Osaka University, Osaka 565-0871, Japan. [6]ERATO Kawarabayashi Large Graph Project, Japan Science and Technology Agency, Hitotsubashi 2-1-2, Chiyoda-ku, Tokyo 101-8403, Japan. [7]E. L. Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA. [8]NTT Device Technology Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan.
***Corresponding author. Email: inagaki.takahiro@lab.ntt.co.jp (T.I.); takesue.hiroki@lab.ntt.co.jp (H.T.)**

1-km fiber cavity, as a phase-sensitive amplifier (PSA) that amplifies only the 0 or π phase components relative to the pump phase (*25*, *26*). Other components in the 1-km fiber cavity include a 9:1 coupler (coupler 1) for extracting the DOPO pulses for balanced homodyne detection (BHD), a 1-km dispersion-shifted fiber, and another 9:1 coupler (coupler 2) for injecting feedback signal pulses. When we start pumping the PSA, quadrature-squeezed noise pulses are generated as a result of spontaneous parametric downconversion. The noise pulses undergo repeated phase-sensitive amplification, which leads to the formation of DOPO pulses. Because the cavity round-trip time is 5 μs and the pump pulse interval is 1 ns, the cavity can accommodate a maximum of 5082 DOPOs, of which 2048 DOPOs are used as Ising spins (henceforth "signal DOPOs"). Another 2834 DOPOs, which we call "training DOPOs," are always being oscillated and are used to obtain error signals for phase-locking the 1-km fiber cavity, the local oscillator for the BHD and the coupling pulses (*24*). As a result, we can perform sequential computations stably for a few seconds. The remaining 200 DOPOs are turned off (i.e., the pump pulses are turned off at the corresponding time slots) to separate the signal and training DOPO pulses.

To perform a computation, we turn the signal DOPOs on and off by controlling the power of the pump pulses for the signal DOPOs in a 5-ms period (fig. S2A). During the computation, the in-phase components $\{\tilde{c}_i\}$ of the signal DOPOs from coupler 1 are measured using the BHD for every circulation of the DOPOs in the cavity. The power of the local oscillator for the BHD is 2.5 mW, with which the shot noise power is ~8 dB larger than the thermal noise power, and thus a shot noise–dominant coherent measurement is achieved. The measured $\{\tilde{c}_i\}$ is supplied to a field-programmable gate array (FPGA) module. The given problem, specified by the spin-spin interaction terms $\{J_{ij}\}$ (a 2048 × 2048 adjacency matrix), is input in advance into the FPGA module. The current FPGA module only accepts $J_{ij}$ values of {−1, 0, 1}. With $\{\tilde{c}_i\}$ and $\{J_{ij}\}$, the FPGA module calculates the feedback signal to the $i$th pulse $f_i = -r\sum_j J_{ij}\tilde{c}_j$ for every DOPO circulation, where $r$ denotes a constant that determines the coupling strength and was set at ~0.01 in the following experiments. A push-pull modulator imposes the feedback signal $f_i$ on the coupling pulses, which are extracted from the pump pulses. Then, the modulated coupling

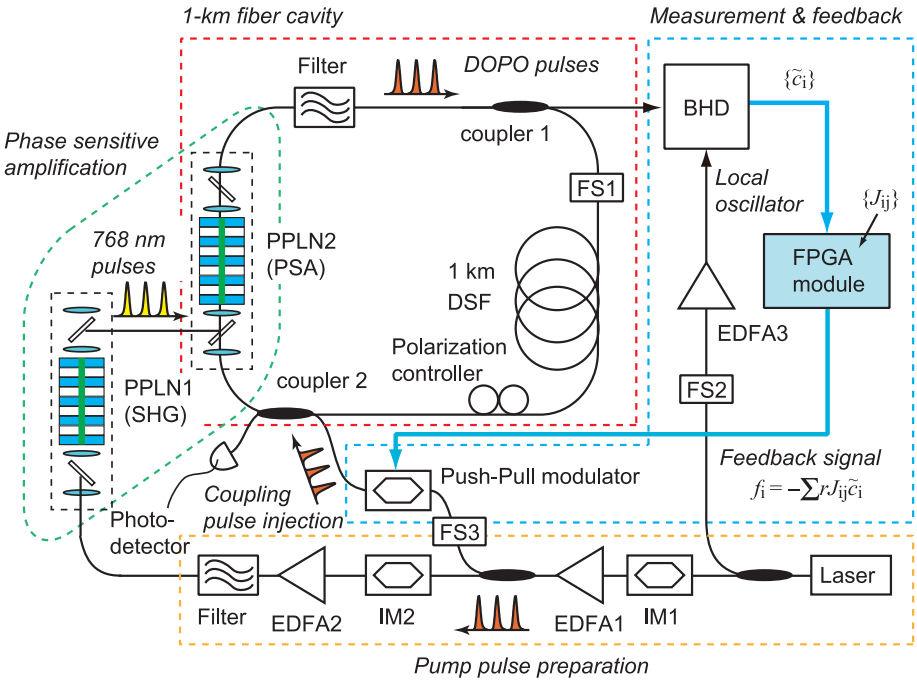



**Fig. 1. Setup of a coherent Ising machine (CIM) with measurement and feedback (MFB).** IM, intensity modulator; EDFA, erbium-doped fiber amplifier; SHG, second harmonic generation; FS, piezo-based fiber stretcher; DSF, dispersion-shifted fiber.

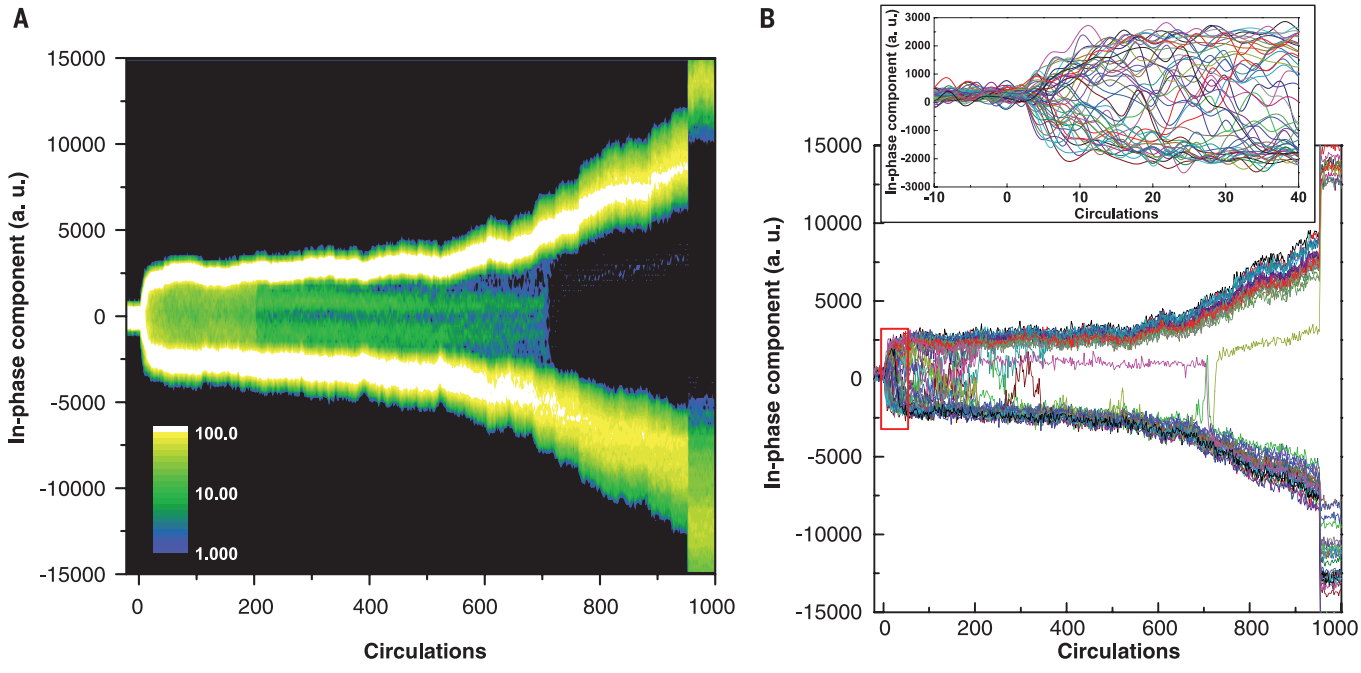



**Fig. 2. DOPO time evolutions when solving the MAX-CUT problem for complete graph $K_{2000}$.** (**A**) Histogram of 2000 signal DOPO time evolutions. (**B**) In-phase component dynamics of the 50 sampled DOPOs. Inset shows the magnification of the area outlined by the red rectangle.

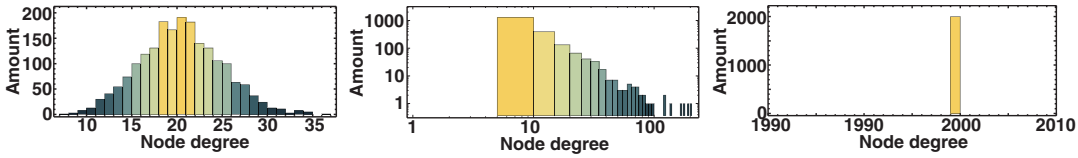
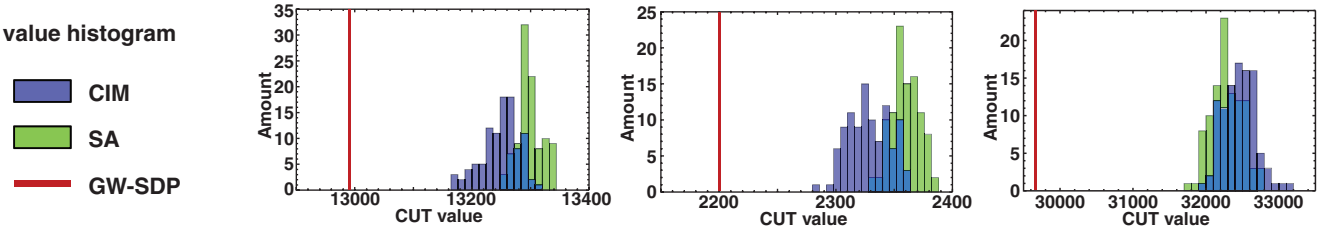
| | G22: random<br>2000 nodes, 19990 edges<br>edge weight: w∈{0, +1} | G39: scale-free<br>2000 nodes, 11778 edges<br>edge weight: w∈{-1, 0, +1} | $K_{2000}$: complete<br>2000 nodes, 1999000 edges<br>edge weight: w∈{-1, +1} |
|---|---|---|---|
| Node degree distribution |  | | |
| Best known | 13359 | 2408 | — |
| SA 50 ms (Best in 100) | 13336 | 2384 | 32781 |
| SA 50 ms (Mean in 100) | 13298 | 2359 | 32314 |
| CIM 5 ms (Best in 100) | 13313 | 2361 | 33191 |
| CIM 5 ms (Mean in 100) | 13248 | 2328 | 32457 |
| GW-SDP | 12992 | 2200 | 29619 |
| Cut value histogram |  | | |



**Fig. 3. Cut values obtained for the MAX-CUT problem on 2000-node graphs with the CIM, SA, and GW-SDP.** Best known refers to the highest cut value yet reported with a conventional algorithm.

pulses are launched into the cavity through coupler 2 so that a coupling pulse that conveys $f_i$ is injected into the $i$th DOPO. Note that this MFB sequence is repeated throughout the entire computational sequence. With the current FPGA module, it takes about ~60 s to transfer the problem matrix and the computation result data between the FPGA module and the computer, because of the use of a slow serial communication interface (RS-232C) that is the only available port in the module. The data transfer time can be much shorter than the computation trial time if we make use of high-speed interfaces (*24*).

We studied the computational accuracy of the CIM and compared it with SA run on a single core of an Intel Xeon X5650 (2.67 GHz) processor for 2000-node graphs. With the CIM and SA, we solved maximum-cut (MAX-CUT) problems (*24*) for three types of 2000-node undirected graphs: G22, G39, and $K_{2000}$. G22 and G39 are sparsely connected graphs taken from the G-set (*27*). G22 is a random graph with a 1% edge density, and G39 is a scale-free graph in which the node degree distribution follows power-law scaling. $K_{2000}$ is a fully connected complete graph with 1,999,000 undirected edges, which is randomly weighted by {–1, +1}. These three instances were randomly constructed with a machine-independent graph generator (*28*). We also solved the same graphs using the semidefinite programming relaxation algorithm proposed by Goemans and Williamson (GW-SDP), because this algorithm guarantees to deliver solutions to MAX-CUT problems with an expected value of at least 87.8% of the optimal value (*29*). We used the cut values obtained by GW-SDP as an index for the computational accuracy.

An example of the temporal evolution of the signal DOPO amplitudes observed while solving the $K_{2000}$ graph is shown in Fig. 2A, which portrays the cooperative behavior of the DOPOs leading to the final stable states after frequent changes of their signs at the beginning of the computation. To show the detailed dynamics of each signal DOPO, we plot the measured in-phase components $\tilde{c}_i$ of 50 of the 2000 signal DOPOs in Fig. 2B. As shown in the inset of Fig. 2B, we observed the evolution of the DOPO amplitudes searching for a lower-energy state almost immediately after the pumping started; the dynamics was complex, especially in the first 200 circulations. This result confirmed a reduction in the Ising energy (fig. S4).

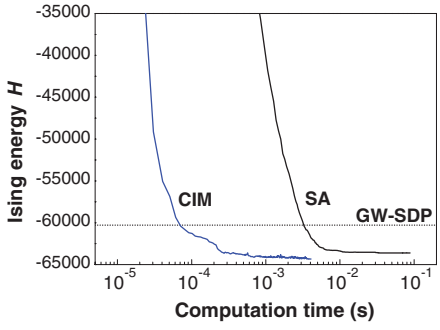



**Fig. 4. Time evolutions of Ising energy obtained with CIM (blue curve) and SA (black curve) when solving the complete graph $K_{2000}$.** The dotted line corresponds to the target Ising energy of –60,278 obtained with GW-SDP.

Using the CIM, we performed 100 sequential computations for MAX-CUT on the three graphs. The cut values obtained with the CIM and SA for each graph are summarized in Fig. 3. Here, the final DOPO states for each computation trial were stably recorded sequentially 100 times, thanks to the phase stabilization of the optical system. The SA result was obtained in 50 ms for each trial without the data read and write time. With GW-SDP, a computation trial took about 1 s. For all three graphs, the CIM achieved larger cut values than those obtained by GW-SDP, which confirmed that a physical system based on the CIM concept successfully worked to find solutions to the large-scale Ising problems with as many as 2000 spins. SA produced better values than the CIM when solving sparse graphs G22 and G39, as confirmed by the cut histograms in Fig. 3. In contrast, the cut obtained by the CIM for the $K_{2000}$ graph exceeded that obtained by SA with one-tenth of the computation time needed for SA.

We then compared the computation time needed to find an approximate solution for the MAX-CUT of the $K_{2000}$ graph. Because the exact solution of this problem is not known, we cannot evaluate the computation time to find the optimal solution. Hence, we measured the computation times needed with the CIM and SA to reach the target Ising energy obtained with GW-SDP, namely –60,278 (or 29,619 cuts). The temporal evolution curves for the CIM and SA are shown in Fig. 4 by

blue and black curves, respectively. Here, we performed the computation 100 times; the curves indicating the shortest times needed to reach the target energy are shown for both the CIM and SA. The times needed to reach the target Ising energy were about 70 μs for the CIM and 3.2 ms for SA, implying that the CIM could obtain the benchmark Ising energy more quickly than with SA by a factor of nearly 50. This result suggests that a CIM can outperform SA on a CPU when solving dense graph problems.

According to the numerical study of Maruo *et al.* (*14*), the formation of a superposed state is observed when an optically coupled CIM is operated in the shot noise–limited regime. This suggests that the CIM may exhibit quantum behavior, although experimental confirmation is yet to be achieved. A more detailed discussion on the quantumness of the CIM is provided in (*24*).

## REFERENCES AND NOTES


1. R. M. Karp, *Complexity of Computer Computations* (Plenum, 1972), pp. 85–103.
2. S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi, *Science* **220**, 671–680 (1983).
3. F. Barahona, *J. Phys. A* **15**, 3241–3253 (1982).
4. M. W. Johnson *et al.*, *Nature* **473**, 194–198 (2011).
5. K. Kim *et al.*, *Nature* **465**, 590–593 (2010).
6. I. Mahboob, H. Okamoto, H. Yamaguchi, *Sci. Advances* **2**, e1600236 (2016).
7. M. Yamaoka *et al.*, *IEEE J. Solid-State Circuits* **51**, 303–309 (2015).
8. S. Utsunomiya, K. Takata, Y. Yamamoto, *Opt. Express* **19**, 18091–18108 (2011).
9. Z. Wang, A. Marandi, K. Wen, R. L. Byer, Y. Yamamoto, *Phys. Rev. A* **88**, 063853 (2013).
10. M. Wolinsky, H. J. Carmichael, *Phys. Rev. Lett.* **60**, 1836–1839 (1988).
11. L. Krippner, W. J. Munro, M. D. Reid, *Phys. Rev. A* **50**, 4330–4338 (1994).
12. P. D. Drummond, K. Dechoum, S. Chaturvedi, *Phys. Rev. A* **65**, 033806 (2002).
13. K. Takata, A. Marandi, Y. Yamamoto, *Phys. Rev. A* **92**, 043821 (2015).
14. D. Maruo, S. Utsunomiya, Y. Yamamoto, *Phys. Scr.* **91**, 083010 (2016).
15. A. Marandi, Z. Wang, K. Takata, R. L. Byer, Y. Yamamoto, *Nat. Photonics* **8**, 937–942 (2014).
16. T. Inagaki *et al.*, *Nat. Photonics* **10**, 415–419 (2016).
17. J. King, S. Yarkoni, M. M. Nevisi, J. P. Hilton, C. C. McGeoch, http://arxiv.org/abs/1508.05087 (2015).
18. E. Ising, *Z. Phys. A* **31**, 253–258 (1925).
19. C. D. Nabors, S. T. Yang, T. Day, R. L. Byer, *J. Opt. Soc. Am. B* **7**, 815–820 (1990).
20. A. Marandi, N. C. Leindecker, V. Pervak, R. L. Byer, K. L. Vodopyanov, *Opt. Express* **20**, 7255–7262 (2012).
21. D. K. Serkland, G. D. Bartolini, A. Agarwal, P. Kumar, W. L. Kath, *Opt. Lett.* **23**, 795–797 (1998).
22. Y. Okawachi *et al.*, *Opt. Lett.* **40**, 5267–5270 (2015).
23. Y. Haribara, S. Utsunomiya, K. Kawarabayashi, Y. Yamamoto, https://arxiv.org/abs/1501.07030 (2015).
24. See supplementary materials on *Science* Online.
25. M. E. Marhic, C. H. Hsia, J. M. Jeong, *Electron. Lett.* **27**, 210–211 (1991).
26. R.-D. Li, P. Kumar, W. L. Kath, *J. Lightwave Technol.* **12**, 541–549 (1994).
27. C. Helmberg, F. Rendl, *SIAM J. Optim.* **10**, 673–696 (2000).
28. G. Rinaldy, rudy graph generator, www-user.tu-chemnitz.de/~helmberg/rudy.tar.gz (1996).
29. M. X. Goemans, D. P. Williamson, *J. ACM* **42**, 1115–1145 (1995).


## ACKNOWLEDGMENTS

Supported by the Impulsing Paradigm Change Through Disruptive Technologies (ImPACT) Program of the Council of Science, Technology and Innovation (Cabinet Office, Government of Japan). We thank H. Nishimori for fruitful discussions, K. Inaba for fruitful discussions, and H. Tamura for various types of support during this research. S.U. and H. Takesue are inventors on patent application PCT/JP2015/059057 submitted by the National Institute of Informatics (NII) and Nippon Telegraph and Telephone (NTT) Corporation that covers the coherent Ising machine based on the measurement and feedback scheme. A.M. and S.U. are inventors on patent application PCT/US2014/046025 submitted by Stanford University and NII that covers the implementation of a coherent Ising machine using degenerate optical parametric oscillators. T.U., K.E., and H. Takenouchi are inventors on patent application PCT/JP1012/000360 submitted by NTT that covers the configurations of phase-sensitive amplifiers based on periodically poled lithium niobate waveguides.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/603/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S5
References (*30*, *31*)

8 July 2016; accepted 26 September 2016
Published online 20 October 2016
10.1126/science.aah4243

QUANTUM GASES

# Universal space-time scaling symmetry in the dynamics of bosons across a quantum phase transition

**Logan W. Clark,* Lei Feng, Cheng Chin**

**The dynamics of many-body systems spanning condensed matter, cosmology, and beyond are hypothesized to be universal when the systems cross continuous phase transitions. The universal dynamics are expected to satisfy a scaling symmetry of space and time with the crossing rate, inspired by the Kibble-Zurek mechanism. We test this symmetry based on Bose condensates in a shaken optical lattice. Shaking the lattice drives condensates across an effectively ferromagnetic quantum phase transition. After crossing the critical point, the condensates manifest delayed growth of spin fluctuations and develop antiferromagnetic spatial correlations resulting from the sub-Poisson distribution of the spacing between topological defects. The fluctuations and correlations are invariant in scaled space-time coordinates, in support of the scaling symmetry of quantum critical dynamics.**

Critical phenomena near a continuous phase transition reveal fascinating connections between seemingly disparate systems that can be described via the same universal principles. Such systems can be found in the contexts of superfluid helium (*1*), liquid crystals (*2*), biological cell membranes (*3*), the early universe (*4*), and cold atoms (*5*, *6*). An important universal prediction is the power-law scaling of the topological defect density with the rate of crossing a critical point, as first discussed by T. Kibble in cosmology (*4*) and extended by W. Zurek in the context of condensed matter (*1*). Their theory, known as the Kibble-Zurek mechanism, has been the subject of intense experimental study that has largely supported the scaling laws (*7*). Recent theoretical works further propose the so-called universality hypothesis, according to which the collective dynamics across a critical point should be invariant in the space and time coordinates that scale with the Kibble-Zurek power law (*8*–*10*).

Atomic quantum gases provide a clean, well-characterized, and controlled platform for studying critical dynamics (*6*, *11*, *12*). They have enabled experiments on the formation of topological defects across the Bose-Einstein condensation transition (*13*–*16*) as well as critical dynamics across quantum phase transitions (*17*–*23*). Recent experiments using cold atoms in shaken optical lattices (*24*–*26*) have provided a vehicle for exploring phase transitions in spin models (*27*–*29*).

Here we study the critical dynamics of Bose condensates in a shaken optical lattice near an effectively ferromagnetic quantum phase transition. The transition occurs when we ramp the shaking amplitude across a critical value, causing the atomic population to bifurcate into two pseudospinor ground states (*28*). We measure the growth of spin fluctuations and the spatial spin correlations for ramping rates varied over two orders of magnitude. Beyond the critical point, we observe delayed development of ferromagnetic spin domains with long-range antiferromagnetic correlations due to the bunching of the domain sizes, which is not expected in a thermal distribution of ferromagnets. The times and lengths characterizing the critical dynamics agree with the scaling predicted by the Kibble-Zurek mechanism. The measured fluctuations and correlations collapse onto single curves in scaled space and time coordinates, supporting the universality hypothesis.

Our experiments use elongated three-dimensional (3D) Bose-Einstein condensates (BECs) of cesium atoms. We optically confine the condensates with trap frequencies of $(\omega_{x'}, \omega_{y'}, \omega_z) = 2\pi \times (12, 30, 70)$

James Franck Institute, Enrico Fermi Institute and Department of Physics, University of Chicago, Chicago, IL 60637, USA.
*Corresponding author. Email: lwclark@uchicago.edu

Hz, where the long ($x'$) and short ($y'$) axes are oriented at 45° with respect to the $x$ and $y$ coordinates (Fig. 1A). The tight confinement along the vertical $z$ axis suppresses nontrivial dynamics in that direction (see the discussion on the dynamics in the $y$ direction below), which is also the optical axis of our imaging system. We adiabatically load the condensates into a 1D optical lattice (*11*) along the $x$ axis with a lattice spacing of 532 nm and a depth of 8.86 $E_R$, where $E_R = h \times 1.33$ kHz is the recoil energy and $h$ is Planck's constant.

To induce the ferromagnetic quantum phase transition, we modulate the phase of the lattice beam to periodically translate the lattice potential by $\Delta x(t) = (s/2)\sin(\omega t)$, where $s$ is the shaking amplitude and the modulation frequency ω is tuned to mix the ground and first excited lattice bands (fig. S1) (*28*, *30*). The hybridized single-particle ground band energy ε can be modeled for small quasimomentum $\mathbf{q} = (q_x, q_y, q_z)$ by

$$\varepsilon(\mathbf{q}; s) = \alpha(s)q_x^2 + \beta(s)q_x^4 + \frac{q_y^2 + q_z^2}{2m} \quad (1)$$

where $m$ is the atomic mass, and the coefficients of its quadratic (α) and quartic (β) terms depend on the shaking amplitude (Fig. 1B). For shaking amplitudes below the critical value, the coefficient α is positive and the BEC occupies the lone ground state at momentum $\mathbf{q} = 0$. The quantum phase transition occurs when the quadratic term crosses zero at $s = s_c$, where $\alpha = 0$ and $\beta > 0$. At this point, the speed of sound for superfluid excitations, formally studied in (*31*), drops to zero along $x$ but remains nearly constant along $y$ and $z$. Even stronger shaking converts the ground band into a double well with $\alpha < 0$, yielding two degenerate ground states with $q_x = \pm q^*$. Repulsively interacting bosons with this double-well ground band are effectively ferromagnetic, having two degenerate many-body ground states with all atoms either pseudo-spin up ($q_x = q^*$) or down ($q_x = -q^*$) (*28*). Notably, transitioning to one of these two ground states requires the system to spontaneously break the symmetry of its Hamiltonian. Describing the dynamics across the critical point presents a major challenge because of the divergence of the correlation length of quasimomentum and the relaxation time (critical slowing).

The Kibble-Zurek mechanism provides a powerful insight into quantum critical dynamics. According to this theory, when the time remaining to reach the critical point inevitably becomes shorter than the relaxation time, the system becomes effectively frozen (Fig. 1C). The system only unfreezes at a delay time $t_{KZ}$ after passing the critical point, when relaxation becomes faster than the ramp. At this time, topological defects become visible, and the typical distance $d_{KZ}$ between neighboring defects is proportional to the equilibrium correlation length. The Kibble-Zurek mechanism predicts that $t_{KZ}$ and $d_{KZ}$ depend on the quench rate $\dot{s}$ as

$$t_{KZ} \propto \dot{s}^{-a}, a = \frac{z\nu}{1 + z\nu} \quad (2)$$

$$d_{KZ} \propto \dot{s}^{-b}, b = \frac{\nu}{1 + z\nu} \quad (3)$$

where $z$ and ν are the equilibrium dynamical and correlation length exponents given by the universality class of the phase transition. Although the details of this picture may not apply to every phase transition, the general scaling arguments are very robust, and similar predictions hold for a variety of quench types across the transition (*12*) and for phase transitions that break either continuous or discrete symmetries (*7*).

For slow ramps, $t_{KZ}$ and $d_{KZ}$ diverge and become separated from other scales in the system, making them the dominant scales for characterizing the collective critical dynamics (*8*–*10*). This idea motivates the universality hypothesis, which can be expressed as

$$f(x, t; \dot{s}) \propto F\left(\frac{x}{d_{KZ}}, \frac{t}{t_{KZ}}\right) \quad (4)$$

indicating that the critical dynamics of any collective observable $f$ obeys the scaling symmetry and can be described by a universal function $F$ of the scaled coordinates $x/d_{KZ}$ and $t/t_{KZ}$. The only effect of the quench rate is to modify the length and time scales.

We test the scaling symmetry of time by monitoring the emergence of quasimomentum fluctuations at different quench rates. Here, fluctuations refer to deviations of quasimomentum from zero, which vary across space and between individual samples; fluctuations should saturate to a large value when domains having $q = \pm q^*$ are fully formed. After loading the condensates into the lattice, we ramp the shaking amplitude linearly from $s = 0$ to values well above the critical amplitude $s_c = 13.1$ nm (*32*) and interrupt the ramps at various times to perform a brief time-of-flight (TOF) before detection. After TOF, we measure the density deviation $\delta n(\mathbf{r}) = n_{-1}(\mathbf{r}) - n_1(\mathbf{r}) - \langle n_{-1}(\mathbf{r}) - n_1(\mathbf{r})\rangle$ (*32*), which is nearly proportional to the quasimomentum distribution (fig. S2), where $n_i(\mathbf{r})$ is the density profile of the $i$th Bragg peak and the angle brackets denote averaging over multiple images. This detection method is particularly sensitive near the critical point when the quasimomentum just starts deviating from zero, indicating the emergence of fluctuations in the ferromagnetic phase where the ground states have nonzero quasimomentum. The spin density measurement used later to study spatial correlations is viable only when atoms settle to $q = \pm q^*$.

Over a wide range of quench rates, the evolution of quasimomentum fluctuation can be described in three phases (Fig. 2A). First, below the critical point, quasimomentum fluctuation does not exceed its baseline level. Second, just after passing the critical point, critical slowing keeps the system "frozen," and fluctuation remains low. Finally, the system unfreezes and quasimomentum fluctuation quickly increases and saturates, indicating the emergence of ferromagnetic domains. We quantify this progression by investigating the fluctuation of contrast $\Delta c = \langle \delta n^2 / n^2 \rangle$ (Fig. 2B) that tracks quasimomentum fluctuation in our condensates via the fluctuation of $\delta n$ (fig. S2), where $n$ is the total density and the angle brackets denote averaging over space and over multiple images. For comparison between different quench rates, we calculate the normalized fluctuation $\Delta C = (\Delta c - \Delta c_i)/(\Delta c_f - \Delta c_i)$, where

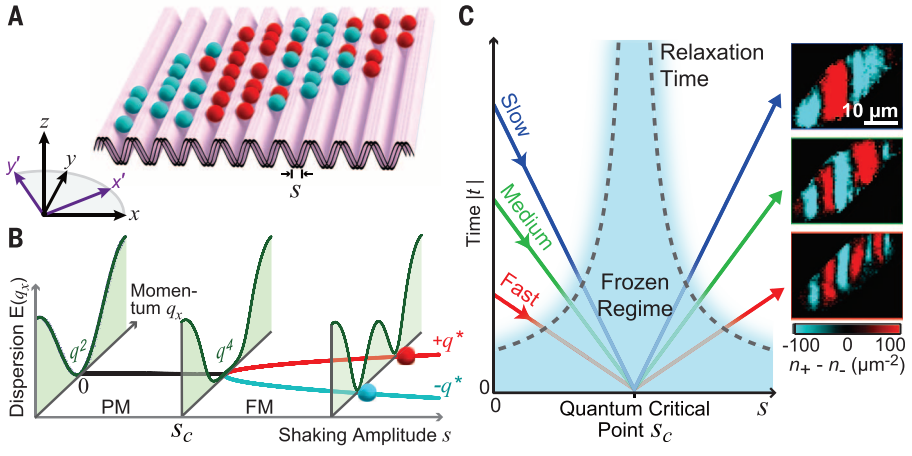



**Fig. 1. Ferromagnetic quantum phase transition of bosons in a shaken optical lattice.** (**A**) A BEC of cesium atoms (spheres) in a 1D optical lattice (pink surface) shaking with peak-to-peak amplitude $s$ can form ferromagnetic domains (blue and red regions). The elliptical harmonic confinement has principal axes rotated 45° from the lattice. (**B**) The transition occurs when the ground band evolves from quadratic for $s < s_c$ [paramagnetic (PM) phase], through quartic at the quantum critical point $s = s_c$, to a double well for $s > s_c$ [ferromagnetic (FM) phase] with two minima at $q_x = \pm q^*$ (*28*). (**C**) Kibble-Zurek picture. Evolution of the condensate crossing the phase transition becomes diabatic in the frozen regime (cyan) when the time $t$ remaining to reach the critical point is less than the relaxation time. Faster ramps cause freezing farther from the critical point, limiting the system to smaller domains. Sample domain images are shown for slow, medium, and fast ramps.

subscripts i and f indicate the fluctuation at early and late times, respectively (*32*). We find empirically that the growth of normalized fluctuations is well fit by the function

$$\Delta C(t) = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{t - t_{\mathrm{d}}}{t_{\mathrm{f}}}\right) \qquad (5)$$

where the time $t$ is defined relative to when the system crosses the critical point at $t = 0$, $t_{\mathrm{d}}$ characterizes the delay time when the system unfreezes, and $t_{\mathrm{f}}$ is the formation time over which the fluctuation grows. The measurement of fluctuation over time provides a critical test for both the Kibble-Zurek scaling and the universality hypothesis. First, both $t_d$ and $t_f$ exhibit clear power-law scaling with the quench rate $\dot{s}$ varied over more than two orders of magnitude (Fig. 2C). Power-law fits yield the exponents of $a_{\mathrm{d}} = 0.50(2)$ and $a_{\mathrm{f}} = 0.50(6)$, respectively. The nearly equal exponents are suggestive of the universality hypothesis, which requires all times to scale identically. Indeed, the growth of contrast fluctuation $\Delta C$ follows a universal curve when time is scaled by $t_{\mathrm{d}}$ (Fig. 2D), strongly supporting the universality hypothesis (Eq. 4). Note that any observable time characterizing the collective dynamics can be chosen as $t_{\mathrm{KZ}}$ in Eq. 4, including $t_{\mathrm{d}}$ and $t_{\mathrm{f}}$.

We next test the spatial scaling symmetry based on the structures of pseudo-spin domains that emerge after the system unfreezes. Here, we cross the critical point with two different protocols: The first is a linear ramp starting from $s = 0$, whereas the second begins with a jump to $s = s_c$, followed by a linear ramp. We detect domains near the time $t = 1.4t_{\mathrm{d}}$ in the spin density distribution $j_z(\mathbf{r}) = n_+(\mathbf{r}) - n_-(\mathbf{r})$ based on the density $n_{+/-}$ of atoms with spin up/down (fig. S3). At this time, the spin domains are fully formed and clearly separated by topological defects (domain walls), as shown in Fig. 3A. Furthermore, choosing this time just after domain formation minimizes the time available for nonuniversal relaxation processes. We characterize the domain distribution with the spin correlation function (*17*, *28*)

$$G(\mathbf{r}) = \left\langle \int j_z(\mathbf{R} + \mathbf{r}) j_z(\mathbf{R})\, d\mathbf{R} \right\rangle \qquad (6)$$

averaged over multiple images (Fig. 3B). Both ramping protocols lead to similar correlation functions, suggesting that the domain distribution is insensitive to increases in the quench rate below the critical point.

The spin correlations reveal rich domain structure that strongly depends on the quench rate. For slower ramps $\dot{s} < 1.3$ nm/ms, the structures are predominantly one-dimensional and the density of topological defects increases with the quench rate. The tighter confinement and finite speed of sound near the critical point along the $y$ and $z$ axes allow spin correlations to span the gas in those directions. The dynamics thus appear one-dimensional. When the quench rate exceeds 1.3 nm/ms, defects start appearing along the $y$ axis, and the domain structures become multidimensional. We attribute this dimensional crossover to the unfreezing time becoming too short to establish correlation along the $y$ axis. For the remainder of this work, we focus on the slower quenches and investigate the spin correlations along the $x$ axis.

We examine the 1D correlations using line cuts of the density-weighted correlation functions $g(\mathbf{r}) = G(\mathbf{r}) / \left\langle \int n(\mathbf{R} + \mathbf{r}) n(\mathbf{R})\, d\mathbf{R} \right\rangle$ (*17*, *28*). The results exhibit prominent decaying oscillation (Fig. 3C). We extract two essential length scales from the correlation functions: the average domain size $d$, or equivalently, the distance between neighboring topological defects, and the correlation length $\xi$, indicating the width of the envelope function. These two scales are determined from the position and width of the peak in the Fourier transform of $g(x)$ (*32*).

These length scales enable us to test the spatial scaling symmetry. The lengths $d$ and $\xi$ both display power-law scaling consistent with the Kibble-Zurek mechanism (Fig. 3D), with fits yielding exponents $b_{\mathrm{d}} = 0.26(2)$ for the domain size and $b_{\xi} = 0.26(5)$ for the correlation length. Similarly, the correlations, measured at the same scaled time, collapse to a single curve in spatial coordinates scaled by the domain size $d$ (Fig. 3E). This result strongly supports the spatiotemporal scaling from

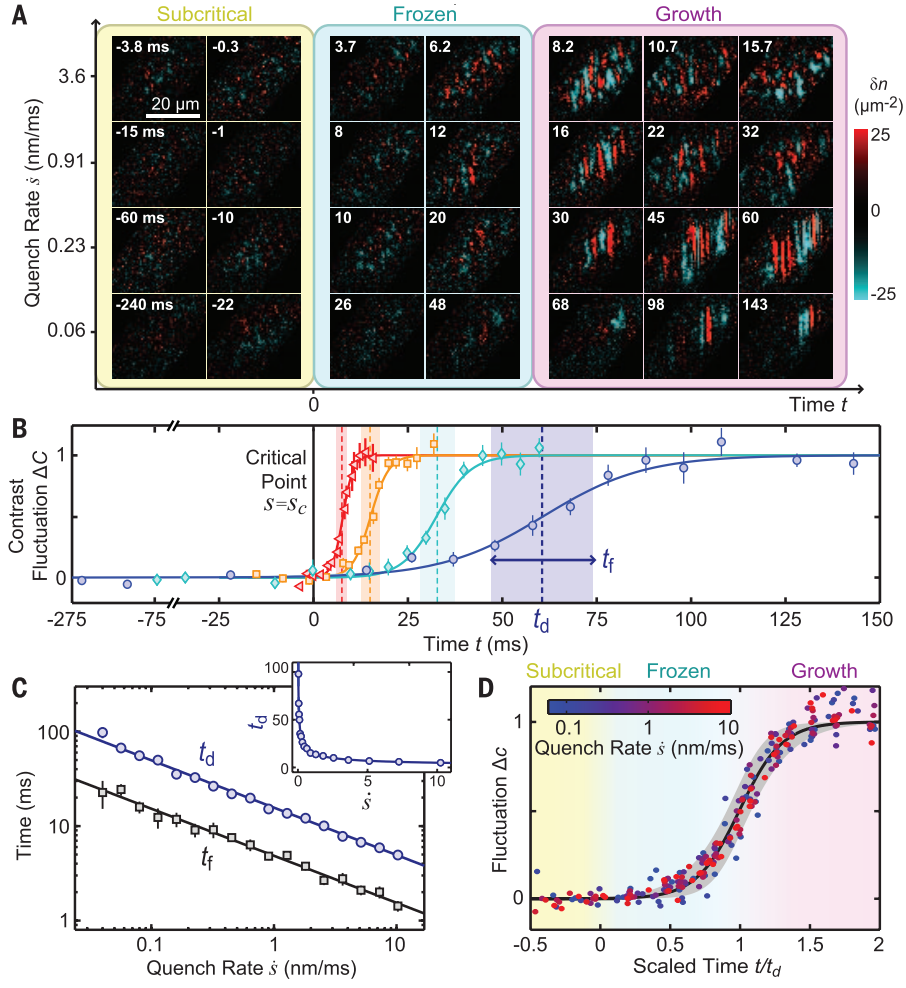



**Fig. 2. Growth of quasimomentum fluctuation in quantum critical dynamics.** (**A**) Sample images show the emergence of nonzero local quasimomentum via the density deviation between Bragg peaks $\delta n$ as the system is linearly ramped across the ferromagnetic phase transition; four ramps with increasing quench rates (bottom to top) are shown. Each ramp exhibits three regimes: a subcritical regime before the transition; a frozen regime beyond the critical point where the fluctuation remains low; and a growth regime in which fluctuation increases and saturates, indicating domain formation. Time $t = 0$ corresponds to the moment when the system reaches the critical point. (**B**) Quasimomentum fluctuation for ramp rates $\dot{s} = 3.6$ (triangles), 0.91 (squares), 0.23 (diamonds), and 0.06 nm/ms (circles) arises at a delay time $t = t_{\mathrm{d}}$ over a formation time $t_{\mathrm{f}}$. Fluctuation is normalized for each ramp rate to aid comparison (*32*). The solid curves show fits based on Eq. 5. (**C**) The dependence of $t_{\mathrm{d}}$ (circles) and $t_{\mathrm{f}}$ (squares) on the quench rate is well fit by power laws (solid curves) with scaling exponents of $a_{\mathrm{d}} = 0.50(2)$ and $a_{\mathrm{f}} = 0.50(6)$, respectively. The inset shows $t_{\mathrm{d}}$ on a linear scale. (**D**) Fluctuations measured for 16 ramp rates from 0.06 to 10.3 nm/ms collapse to a single curve when time is scaled by $t_{\mathrm{d}}$ based on the power-law fit. The solid curve shows the best fit based on the empirical function (Eq. 5), and the gray shaded region covers one standard deviation. Error bars in (B) and (C) indicate one standard error.

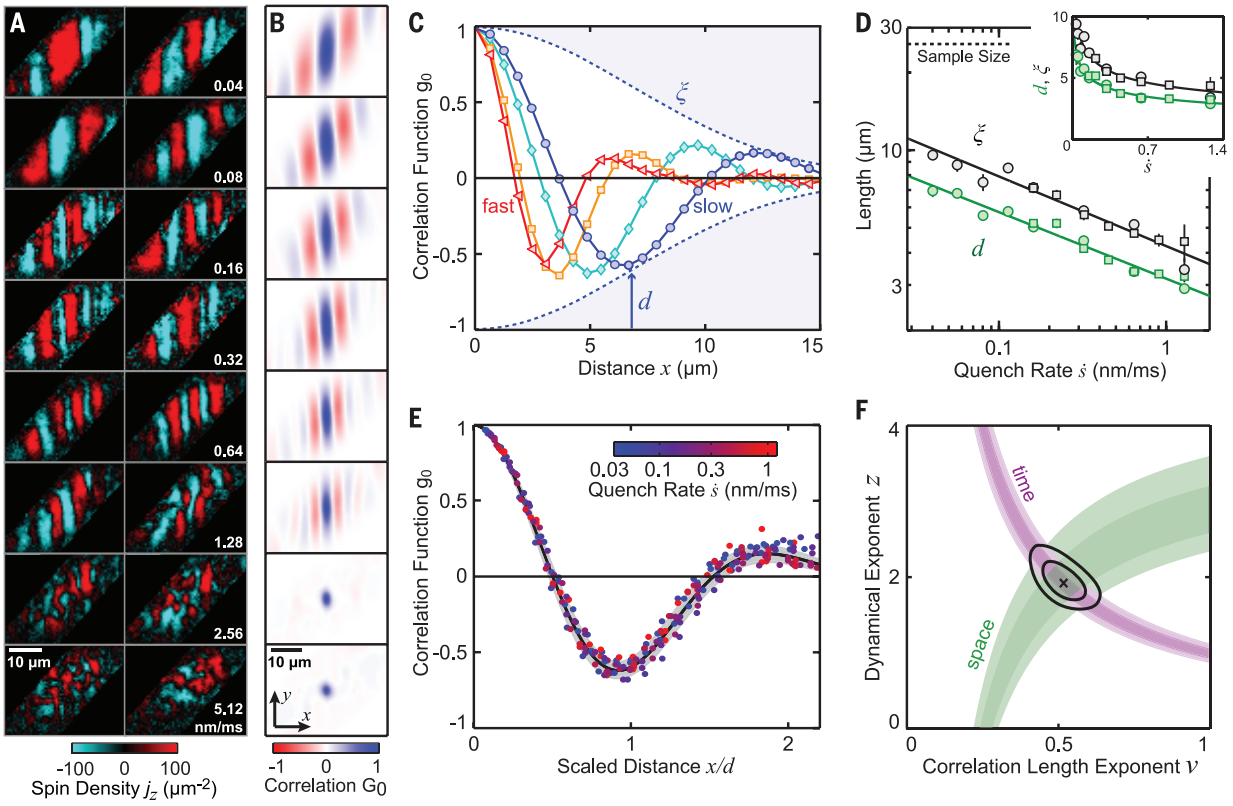



**Fig. 3. Antiferromagnetic spatial correlations from quantum critical dynamics.** (**A**) Two sample images at each quench rate exemplify spin domains measured near the time $t = 1.4t_q$ after crossing the phase transition. These images correspond to linear ramps starting at $s = 0$ ($\dot{s} = 0.32$ to 5.12 nm/ms) and $s = s_c$ ($\dot{s} = 0.04$ to 0.16 nm/ms). (**B**) Spin correlation functions $G_0(\mathbf{r}) = G(\mathbf{r})/G(0)$ (Eq. 6) are calculated from ensembles of 110 to 200 images. (**C**) Cuts across the density-weighted correlation functions $g_0(x) = g(x)/g(0)$ are shown for quench rates $\dot{s} = 1.28$ (triangles), 0.45 (squares), 0.16 (diamonds), and 0.056 nm/ms (circles). Solid curves interpolate the data to guide the eye. The typical domain size $d$ and the correlation length ξ are illustrated for 0.056 nm/ms by the arrow and dashed envelope, respectively (*32*). (**D**) The dependence of $d$ (green) and ξ (black) on the quench rate is well fit by power laws (Eq. 3) with spatial scaling exponents of $b_d = 0.26(2)$ and $b_\xi = 0.26(5)$, respectively. Marker shape indicates linear ramps starting at $s = 0$ (squares) or at $s = s_c$ (circles). The inset shows the results on a linear scale. Error bars indicate 1 SE. (**E**) Correlation functions for $\dot{s} = 0.04$ to 1.28 nm/ms collapse to a single curve when distance is scaled by the domain size extracted from the power-law fit in (D). The solid curve shows the fit based on Eq. 7; the gray shaded area covers 1 SD. (**F**) The temporal scaling exponents $a_d$ and $a_\xi$ from Fig. 2C (magenta) and the spatial scaling exponents $b_d$ and $b_\xi$ from (D) (green) constrain the critical exponents ν and $z$ according to Eqs. 2 and 3 with 68% (dark) and 95% (light) confidence intervals. The cross marks the best values with contours of 68 and 95% overall confidence (*32*).

the universality hypothesis (Eq. 4). An empirical curve

$$g_0(x) = \exp\left(-\frac{1}{2\sigma^2}\frac{x^2}{d^2}\right)\cos\left(\frac{\pi x}{\gamma d}\right) \qquad (7)$$

provides a good fit to the universal correlation function, yielding σ = 1.01(1) and γ = 1.04(1), indicating that the width of the envelope is close to the typical domain size.

The most striking feature of the universal correlation function is the emergence of oscillatory, antiferromagnetic order in the ferromagnetic phase. In thermal equilibrium, ferromagnets are expected to have a finite correlation length but no anticorrelation. The appearance of strong anticorrelation at $x = d$ suggests that domains of size $d$ form preferentially during the quantum critical dynamics. A statistical analysis of the topological defect distribution reveals that the domain sizes are bunched with their standard deviation $\sigma_d = 0.31(2)d$ well below their mean, indicating that the topological defects are created by a sub-Poisson process (fig. S4).

Finally, the combined scaling exponents of space and time allow us to extract the equilibrium critical exponents based on the Kibble-Zurek mechanism (*33*) (Fig. 3F). Solving Eqs. 2 and 3, we obtain the dynamical exponent $z = 1.9(2)$ and correlation length exponent $\nu = 0.52(5)$, which are close to the mean-field values $z = 2$ and $\nu = 1/2$ up to our experimental uncertainty. Note that the dynamical critical exponent $z = 2$ results from the unique quartic kinetic energy $\varepsilon = \beta q_x^4$ of our system at the critical point (*32*).

Direct identification of domain walls presents intriguing possibilities for future studies of the topological defects generated during critical dynamics. These opportunities would be particularly interesting if the shaking technique were extended to higher dimensions in such a way that the transition breaks a continuous symmetry. In addition, the scaling of the correlation functions suggests that the antiferromagnetic order may be a shared feature of quantum critical dynamics for phase transitions in the same universality class, meriting future experiments.

## REFERENCES AND NOTES


1. W. H. Zurek, *Nature* **317**, 505–508 (1985).
2. I. Chuang, R. Durrer, N. Turok, B. Yurke, *Science* **251**, 1336–1342 (1991).
3. S. L. Veatch, O. Soubias, S. L. Keller, K. Gawrisch, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17650–17655 (2007).
4. T. W. B. Kibble, *J. Phys. Math. Gen.* **9**, 1387–1398 (1976).
5. Q. Zhou, T. L. Ho, *Phys. Rev. Lett.* **105**, 245702 (2010).
6. A. Polkovnikov, K. Sengupta, A. Silva, M. Vengalattore, *Rev. Mod. Phys.* **83**, 863–883 (2011).
7. A. del Campo, W. H. Zurek, *Int. J. Mod. Phys. A* **29**, 1430018 (2014).
8. M. Kolodrubetz, B. K. Clark, D. A. Huse, *Phys. Rev. Lett.* **109**, 015701 (2012).
9. A. Chandran, A. Erez, S. S. Gubser, S. L. Sondhi, *Phys. Rev. B* **86**, 064304 (2012).
10. A. Francuz, J. Dziarmaga, B. Gardas, W. H. Zurek, *Phys. Rev. B* **93**, 075134 (2016).
11. I. Bloch, J. Dalibard, W. Zwerger, *Rev. Mod. Phys.* **80**, 885–964 (2008).
12. J. Dziarmaga, *Adv. Phys.* **59**, 1063–1189 (2010).
13. C. N. Weiler *et al.*, *Nature* **455**, 948–951 (2008).
14. G. Lamporesi, S. Donadello, S. Serafini, F. Dalfovo, G. Ferrari, *Nat. Phys.* **9**, 656–660 (2013).
15. L. Corman *et al.*, *Phys. Rev. Lett.* **113**, 135302 (2014).
16. N. Navon, A. L. Gaunt, R. P. Smith, Z. Hadzibabic, *Science* **347**, 167–170 (2015).
17. L. E. Sadler, J. M. Higbie, S. R. Leslie, M. Vengalattore, D. M. Stamper-Kurn, *Nature* **443**, 312–315 (2006).

18. K. Baumann, R. Mottl, F. Brennecke, T. Esslinger, *Phys. Rev. Lett.* **107**, 140402 (2011).
19. D. Chen, M. White, C. Borries, B. DeMarco, *Phys. Rev. Lett.* **106**, 235304 (2011).
20. E. Nicklas *et al.*, *Phys. Rev. Lett.* **115**, 245301 (2015).
21. S. Braun *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 3641–3646 (2015).
22. C. Meldgin *et al.*, *Nat. Phys.* **12**, 646–649 (2016).
23. M. Anquez *et al.*, *Phys. Rev. Lett.* **116**, 155301 (2016).
24. N. Gemelke, E. Sarajlic, Y. Bidel, S. Hong, S. Chu, *Phys. Rev. Lett.* **95**, 170404 (2005).
25. H. Lignier *et al.*, *Phys. Rev. Lett.* **99**, 220403 (2007).
26. J. Struck *et al.*, *Phys. Rev. Lett.* **108**, 225304 (2012).
27. J. Struck *et al.*, *Science* **333**, 996–999 (2011).
28. C. V. Parker, L.-C. Ha, C. Chin, *Nat. Phys.* **9**, 769–774 (2013).
29. G. Jotzu *et al.*, *Nature* **515**, 237–240 (2014).
30. W. Zheng, B. Liu, J. Miao, C. Chin, H. Zhai, *Phys. Rev. Lett.* **113**, 155303 (2014).
31. L.-C. Ha, L. W. Clark, C. V. Parker, B. M. Anderson, C. Chin, *Phys. Rev. Lett.* **114**, 055301 (2015).
32. Materials and methods are available on *Science* Online.
33. J. Xu *et al.*, *Europhys. Lett.* **113**, 50003 (2016).

**ACKNOWLEDGMENTS**

We thank L.-C. Ha, C. V. Parker, B. M. Anderson, B. J. DeSalvo, A. Polkovnikov, S. Sachdev, and Q. Zhou for helpful discussions. L.W.C. was supported by a National Defense Science and Engineering Graduate Fellowship and a Nambu Fellowship. This work was supported by NSF Materials Research Science and Engineering Centers (DMR-1420709), NSF grant PHY-1511696, and Army Research Office–Multidisciplinary Research Initiative grant W911NF-14-1-0003. The data presented in this paper are available upon request to C.C. (cchin@uchicago.edu).

**SUPPLEMENTARY MATERIALS**

www.sciencemag.org/content/354/6312/606/suppl/DC1
Materials and Methods
Figs. S1 to S4
References (*34*, *35*)

28 April 2016; accepted 26 September 2016
10.1126/science.aaf9657

PHYSICS

# Buffer-gas cooling of antiprotonic helium to 1.5 to 1.7 K, and antiproton-to–electron mass ratio

**Masaki Hori,[1]* Hossein Aghai-Khozani,[1] Anna Sótér,[1] Daniel Barna,[2] Andreas Dax,[3]† Ryugo Hayano,[3] Takumi Kobayashi,[3]‡ Yohei Murakami,[3] Koichi Todoroki,[3]§ Hiroyuki Yamada,[3] Dezső Horváth,[2,4] Luca Venturelli[5]**

**Charge, parity, and time reversal (CPT) symmetry implies that a particle and its antiparticle have the same mass. The antiproton-to-electron mass ratio $M_{\bar{p}}/m_e$ can be precisely determined from the single-photon transition frequencies of antiprotonic helium. We measured 13 such frequencies with laser spectroscopy to a fractional precision of $2.5 \times 10^{-9}$ to $16 \times 10^{-9}$. About $2 \times 10^9$ antiprotonic helium atoms were cooled to temperatures between 1.5 and 1.7 kelvin by using buffer-gas cooling in cryogenic low-pressure helium gas; the narrow thermal distribution led to the observation of sharp spectral lines of small thermal Doppler width. The deviation between the experimental frequencies and the results of three-body quantum electrodynamics calculations was reduced by a factor of 1.4 to 10 compared with previous single-photon experiments. From this, $M_{\bar{p}}/m_e$ was determined as 1836.1526734(15), which agrees with a recent proton-to-electron experimental value within $8 \times 10^{-10}$.**

Techniques to cool atoms containing antiparticles such as antihydrogen (*1*, *2*), positronium (*3*, *4*), or antiprotonic helium ($\bar{p}He^+$) (*5*, *6*) allow their spectral lines to be measured with narrow thermal Doppler widths and strong signal intensities. Their transition frequencies can then be determined with the highest precision, thus increasing the sensitivity of the experiments to the consistency of charge, parity, and time reversal (CPT) symmetry (*7*), which implies that particles and antiparticles have equal masses and absolute values of the charge. Spectral data on $\bar{p}He^+$ have also been used to set upper limits on a possible fifth force at angstrom-length scales (*8*, *9*). Buffer-gas cooling techniques (*10*, *11*) have been used to cool many ordinary atoms and molecules that lack the closed-cycle transitions suitable for laser cooling. It relies on elastic collisions between the atom of interest with a cryogenic buffer gas to allow the system to reach thermal equilibrium. Here, we report on the application of this method to cool large ($2 \times 10^9$) numbers of $\bar{p}^4He^+$ and $\bar{p}^3He^+$ isotopes to temperatures $T$ = 1.5 to 1.7 K.

Metastable $\bar{p}He^+$ is a three-body exotic atom (fig. S1) that contains a helium nucleus, an electron occupying the 1s state, and an antiproton in a Rydberg state of principal- and angular-momentum quantum numbers $n \sim l + 1 \sim 38$ (*5*, *12*–*15*). The antiproton that occupies these circular orbitals that lie ~80 keV above the 1s antiprotonic state retains microsecond-scale lifetimes against annihilation in the nucleus because it is localized away from the nucleus. Before this experiment, it was not known whether collisions with normal-matter atoms would cool $\bar{p}He^+$ to such low (1.5 K) temperatures because the corresponding multibody calculations were complicated. In fact, other exotic atoms such as pionic hydrogen (*16*) were heated by collisions that deexcite the atom; $\bar{p}He^+$ atoms colliding with other helium atoms annihilated (*5*) the antiprotons occupying some states. We adjusted the density of the buffer helium gas so that the $\bar{p}He^+$ atoms, once formed, rapidly underwent a few hundred or more cooling collisions before being interrogated by the laser beam. The 1s electron protected most of the $\bar{p}He^+$ atoms during this cooling.

The $\bar{p}He^+$ transition frequencies have recently been calculated (*12*–*15*) to a precision of ~$10^{-10}$ by evaluating the complete set of quantum electrodynamics (QED) corrections up to order $m_e\alpha^7$ in atomic units. These a priori calculations used the International Council for Science Committee on Data for Science and Technology (CODATA) 2010–recommended values of the fundamental constants (*17*), including the fine structure constant $\alpha$, the $^3$He- and $^4$He-to-electron mass ratios, the Bohr radius, and the Rydberg constant. By comparing the calculated and experimental $\bar{p}He^+$ frequencies, the $M_{\bar{p}}/m_e$ ratio can in principle be determined to a fractional precision of $<1 \times 10^{-10}$ (*12*–*15*), which may rival the best measurements of the proton-to-electron mass ratio $M_p/m_e$ (*18*–*22*). Previous measurements (*23*) of the $\bar{p}He^+$ single-photon frequencies, however, differed by up to $8 \times 10^{-8}$ from the calculations available at the time, which corresponded in some cases to 2 to 3 standard experimental deviations. A major source of experimental uncertainty was the Doppler broadening of the spectral lines. We later measured a different set of two-photon resonances using a Doppler-cancellation technique (*6*) in which $\bar{p}He^+$ atoms were irradiated by a pair of counterpropagating laser beams. Three transition frequencies involving the $n < 36$ states were measured with this complicated method. Assuming CPT invariance, the results were included in the CODATA2010 data set in order to determine the electron mass $m_e$ in atomic mass units; however, further improvements in the precision of the determination of $M_{\bar{p}}/m_e$ were limited by the poor experimental precision of the single-photon frequencies (*23*) that had higher sensitivities $\kappa_M$ (*5*) to $M_{\bar{p}}/m_e$ and $m_e$, as compared with those of the two-photon lines (Table 1). The Doppler-broadened single-photon

[1]Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, 85748 Garching, Germany. [2]MTA Wigner Research Centre for Physics, H-1525 Budapest, Hungary. [3]Department of Physics, University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. [4]Institute of Nuclear Research (ATOMKI), H-4001 Debrecen, Hungary. [5]Dipartimento di Ingegneria dell'Informazione, Università di Brescia, Istituto Nazionale di Fisica Nucleare, I-25133 Brescia, Italy.
***Corresponding author. Email: masaki.hori@mpq.mpg.de**
†Present address: Paul Scherrer Institut, CH-5232 Villigen, Switzerland. ‡Present address: National Metrology Institute of Japan (NMIJ), National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8563, Japan. §Present address: NEC Corporation, Shiba, Minato-Ku, Tokyo, Japan.

lines of buffer-gas–cooled $\overline{p}He^+$ observed in the present work—together with improved lasers, detectors, and a hybrid magnetic and electrostatic beamline—yielded $M_{\overline{p}}/m_e$ and $m_e$ values with a precision beyond that of the previous sub-Doppler two-photon experiment (*6*).

In our experiment, $\overline{p}He^+$ atoms were formed by allowing antiprotons to come to rest in a helium buffer gas in thermal contact with an open-cycle Joule-Thomson cryocooler at $T$ = 1.3 K (Fig. 1). About 3% of the antiprotons replaced one of the 1s electrons of helium and formed metastable $\overline{p}He^+$. Targets of low density were used to reduce the effects of collisional shifting of the spectral lines; for this, antiproton beams of sufficiently low energy and emittance that could be made to come to rest in the targets were needed. The experiments were carried out at the Antiproton Decelerator of the European Organization for Nuclear Research (CERN), which provided a pulsed beam containing $2 \times 10^7$ to $3 \times 10^7$ antiprotons of kinetic energy $E$ = 5.3 MeV and repetition rate 0.01 Hz. About 25% of these were decelerated to $E$ = 75 keV by allowing them to traverse a radiofrequency quadrupole decelerator. The emerging beam had a large emittance $\varepsilon$ = 50 to 100 $\pi$ mm mrad and an energy spread of $\Delta E$ = 10 keV. To transport this to the position of the experimental target while preventing the growth of beam halos, we constructed a 2-m-long achromatic beamline (*24*). This reduced the beam diameter (<15 mm) in the target by a factor of 2 compared with previous experiments (*6*, *23*). Each arrival of an antiproton pulse synthesized $10^5$ $\overline{p}He^+$ atoms. Some $3 \times 10^3$ to $15 \times 10^3$atoms populated the resonance parent state at the moment of laser irradiation, depending on the transition. Data from $>1.5 \times 10^9$ cold $\overline{p}^4He^+$ atoms and $>5 \times 10^8$ cold $\overline{p}^3He^+$ atoms were collected over a 3-year period.

The target chamber was machined out of a 150 by 150 by 150 mm block of oxygen-free high-conductivity copper (Fig. 1). The antiprotons entered through a 1.4-μm-thick, 28-mm-diameter window made of biaxially oriented polyethylene terephthalate (BoPET) supported by a grid of gold-coated tungsten wires. Target $^4$He gas of purity >99.999% introduced into the chamber condensed into a superfluid film on the inner walls cooled to a constant temperature of $T \approx 1.3$ K; cold gas that evaporated from this was allowed to collide with the antiprotons. No such film formed for the $^3$He, or when the pressure of $^4$He in the target (measured by capacitance manometers) was low. By controlling the amount of gas introduced into the sealed chamber, the pressure was stabilized at several settings between $P$ = 40 and 170 Pa, corresponding to atomic target densities between $\rho = 2 \times 10^{18}$ and $9 \times 10^{18}$ cm$^{-3}$. A 30-mm-diameter laser beam entered through a fused silica window with a major diameter of 70 mm, which was fixed on one side of the chamber at Brewster's angle. The laser beam intersected the antiproton beam at a 90° angle and exited through a second window on the opposite side. This prevented any retroreflected laser light from irradiating the $\overline{p}He^+$ atoms and deforming the spectral line shape, which was a problem in the previous single-photon experiment (*23*) based on collinear antiproton and laser beams. Three 1-m-long copper baffles cooled to $T$ < 7 K reduced the thermal radiation entering through the large windows. Within the cryocooler, liquid helium was cooled to $T$ ~ 2.2 K in a recuperative heat exchanger, before being expanded through a throttle valve, which provided further cooling to $T$ ~ 1.3 K. The coolant condensed in a heat exchanger that was evacuated though a 150-mm-diameter cryogenic pumping line by means of a roots blower and rotary vane pump with respective speeds of 2000 and 800 m$^3$/hour. A 1-m-thick concrete wall isolated the target from the 80 dB acoustic noise of the pumps. The temperature was stabilized to within 0.03 K by controlling this throttle valve using a stepping motor.

Dye and Ti:sapphire laser systems generated the 40- to 100-ns-long laser pulses with peak powers $P$ = 0.5 to 10 kW and wavelengths $\lambda$ = 264 to 841 nm needed to excite the $\overline{p}He^+$ atoms within their microsecond-scale lifetimes. Their

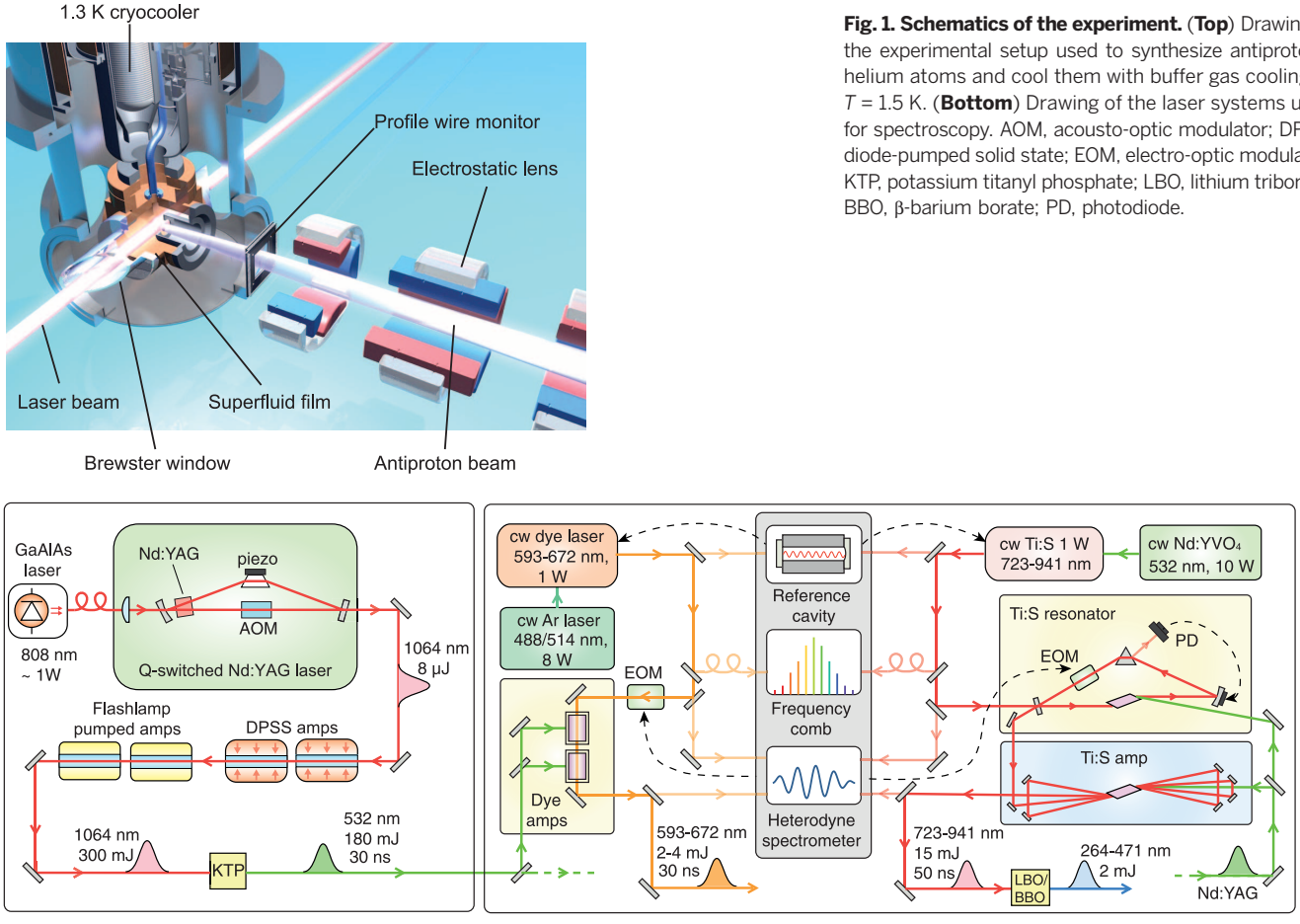



**Fig. 1. Schematics of the experiment.** (**Top**) Drawing of the experimental setup used to synthesize antiprotonic helium atoms and cool them with buffer gas cooling to $T$ = 1.5 K. (**Bottom**) Drawing of the laser systems used for spectroscopy. AOM, acousto-optic modulator; DPSS, diode-pumped solid state; EOM, electro-optic modulator; KTP, potassium titanyl phosphate; LBO, lithium triborate; BBO, β-barium borate; PD, photodiode.

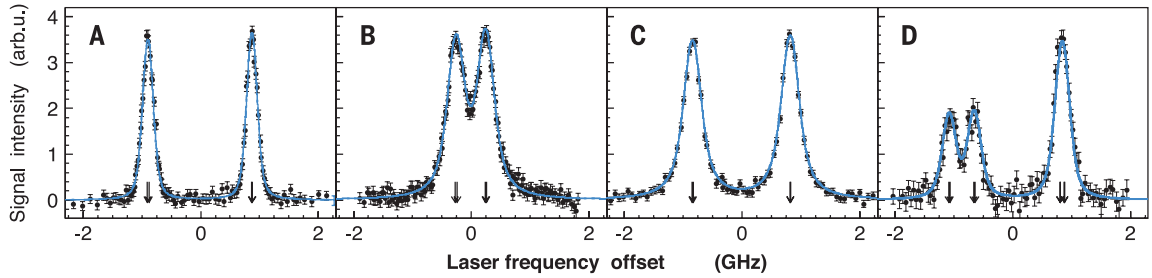



**Fig. 2. Resonance profiles of the single-photon transitions.** (**A**) $(n, l) = (37, 35) \rightarrow (38, 34)$, (**B**) $(39, 35) \rightarrow (38, 34)$, and (**C**) $(38, 35) \rightarrow (39, 34)$ of buffer-gas cooled $\overline{p}^4He^+$ atoms. The narrow, intense spectral lines are compatible with the Doppler widths of atoms cooled to $T = 1.5$ to 1.7 K, relative to the corresponding transition frequencies between $\nu = 356$ and 502 THz and when the Auger widths of the resonance daughter states are subtracted. The arrows indicate the theoretical positions of the hyperfine lines caused by the spin-spin interaction between the antiproton and electron (*26*); the (B) and (C) splittings had not been observed previously. (**D**) The profile $(36, 34) \rightarrow (37, 33)$ of cooled $\overline{p}^3He^+$ shows a further splitting, due to the $^3$He nuclear spin. The *x*-abscissa indicates the offset of the optical frequency of the laser relative to the resonance centroid. Blue curves indicate the best fit of an ab initio model based on the optical Bloch equation (*30*).

linewidths $\Gamma = 6$ to 20 MHz were a factor of 4 to 10 narrower than those used in previous single-photon experiments (*23*). In the first system, a ring dye laser pumped by an Ar-ion laser produced continuous-wave beams of $P = 1$ W, $\lambda = 593$ to 672 nm, and $\Gamma < 0.1$ MHz. Its frequency was stabilized by using the Pound-Drever-Hall technique to a temperature-stabilized, 75-mm-long Fabry-Perot cavity that was suspended in a vacuum chamber. The seed laser frequencies were measured to a precision of $< 2 \times 10^{-10}$ by using a Ti:sapphire femtosecond optical comb that was stabilized to a global positioning satellite–disciplined, quartz oscillator. This seed beam was amplified to a pulse energy $E = 2$ to 4 mJ in two dye cells pumped by a Q-switched, single-longitudinal-mode Nd:YAG laser of $E = 180$ mJ and $\lambda = 532$ nm. In the previous single-photon experiment (*23*), a systematic uncertainty arose from the spectral broadening and shift during the dye amplification; this was caused by rapid modulations in the intensity of the pump laser. We reduced this uncertainty by developing a pump laser (*24*) with a long ($\Delta t = 40$ to 60 ns), smooth temporal profile with a single longitudinal mode.

In the second system, a continuous-wave Ti:sapphire laser pumped by a Nd:$YVO_4$ laser generated seed beams of $P = 1$ W, $\lambda = 723$ to 941 nm, and $\Gamma < 0.1$ MHz. This beam was injection-seeded into a 1-m-circumference Ti:sapphire cavity (*6*, *25*), which produced 40- to 100-ns-long laser pulses of $E = 4$ to 13 mJ. For measuring the blue and ultraviolet (UV) resonances, the beam was further amplified to $E = 15$ to 40 mJ in a Ti:sapphire multipass amplifier, before being frequency-doubled or -tripled in lithium triborate (LBO) and β-barium borate (BBO) crystals. The absolute precision of the pulsed dye and Ti:sapphire laser systems of $< 2 \times 10^{-9}$ and $< 1.4 \times 10^{-9}$, respectively, was verified by measuring some transition frequencies in Cs continuously over a 2-week period.

The laser pulses excited transitions between $\overline{p}He^+$ states with 1- to 2-μs lifetimes and states that led to Auger emission of the 1s electron with nanosecond lifetimes. The $\overline{p}He^{2+}$ ion that remained after Auger decay was rapidly destroyed by Stark effects, which mixed the Rydberg ionic states with *s*, *p*, and *d* states at high *n* during collisions with helium atoms. The charged pions that emerged from the resulting antiproton annihilations in the nucleus were detected with a Cherenkov detector (*24*), thus revealing the resonance condition between the laser and $\overline{p}He^+$ as a sharp spike in the annihilation rate (fig. S2).

**Table 1. Transition frequencies of $p^4He^+$ and $p^3He^+$ atoms.** For experimental values, total, statistical, and systematic uncertainties are given in parentheses. Theoretical frequencies have uncertainties from numerical errors of the calculation and uncalculated QED terms. $\kappa_M$ and $\kappa_Q$ are the sensitivities of different transitions to antiproton mass and charge, respectively. ppb, parts per billion.

| Transition $(n, l) \rightarrow (n', l')$ | Experiment frequency (MHz) | Theor. frequency (MHz) | $\kappa_M$ (ppb) | $\kappa_Q$ (ppb) |
|---|---|---|---|---|
| $\overline{p}^4He^+$ | | | | |
| (40, 35) → (39, 34) | 445608573(5)(4)(1) | 445608572.3(4) | 1.9 | 3.1 |
| (38, 35) → (39, 34) | 356155990.1(2.1)(2.0)(0.8) | 356155990.5(4) | 2.7 | 4.1 |
| (39, 35) → (38, 34) | 501948753.4(2.1)(1.9)(0.8) | 501948755.1(2) | 1.8 | 3.0 |
| (37, 35) → (38, 34) | 412885133.1(1.0)(0.8)(0.6) | 412885132.4(2) | 2.6 | 4.0 |
| (37, 34) → (36, 33) | 636878154.3(2.2)(1.9)(1.1) | 636878152.09(5) | 1.6 | 2.8 |
| (34, 33) → (35, 32) | 655062100(10)(10)(1) | 655062101.92(7) | 2.1 | 3.4 |
| (35, 33) → (34, 32) | 804633058.2(2.1)(1.8)(1.2) | 804633058.46(6) | 1.5 | 2.6 |
| (32, 31) → (31, 30) | 1132609226.7(2.8)(2.5)(1.4) | 1132609224.01(8) | 1.3 | 2.4 |
| $\overline{p}^3He^+$ | | | | |
| (38, 34) → (37, 33) | 505222282(4)(4)(1) | 505222281.0(3) | 1.8 | 3.1 |
| (36, 34) → (37, 33) | 414147510.4(2.6)(2.3)(1.2) | 414147508.9(3) | 2.6 | 4.0 |
| (36, 33) → (35, 32) | 646180416(5)(4)(1) | 646180412.58(5) | 1.6 | 2.8 |
| (34, 32) → (33, 31) | 822809167(5)(5)(1) | 822809172.30(7) | 1.5 | 2.6 |
| (32, 31) → (31, 30) | 1043128581(5)(4)(1) | 1043128580.64(8) | 1.3 | 2.5 |

The profile of the $\overline{p}^4He^+$ transition $(n, l) = (37, 35) \rightarrow (38, 34)$ is shown in Fig. 2A, obtained by plotting the intensities of the annihilation signals induced at laser frequencies between −1 and 1 GHz around the resonance centroid. Each data point was collected from 6 to 10 antiproton pulses. Pairs of fine structure sublines arise from the dominant interaction between the orbital angular momentum of the antiproton and the electron spin; the positions of the four hyperfine sublines (*26*) that arise from the additional spin-spin interaction between the antiproton and electron are indicated in Fig. 2 with arrows. This single-photon resolution exceeds those of previous sub-Doppler two-photon spectroscopy experiments (*6*) and is consistent with the Doppler width $\approx v\sqrt{8kT\log 2/Mc^2} = 160$ to 170 MHz of $\overline{p}^4He^+$ thermalized to $T = 1.5$ to 1.7 K when the contributions of power broadening (<30 MHz), natural width (20 MHz), laser line width (6 MHz)—which includes the Fourier transform limit (*25*) due to the pulse length of the laser —and the hyperfine structure (<30 MHz) are subtracted. Collisional broadening effects are small (*27*). Here, $k$ denotes the Boltzmann constant, $M$ is the atom's mass, and $c$ is the speed of light. The strong signal intensity (3× larger than in the previous high-resolution single-photon experiment relative to the total number of $\overline{p}^4He^+$ atoms formed in the target) arises from the high density of cold $\overline{p}He^+$ that lies within the thermal velocity class excited by the laser. The profile of a previously undetected $\overline{p}^4He^+$ resonance $(38, 35) \rightarrow (39, 34)$ is shown in Fig. 2C. Its spectral resolution is

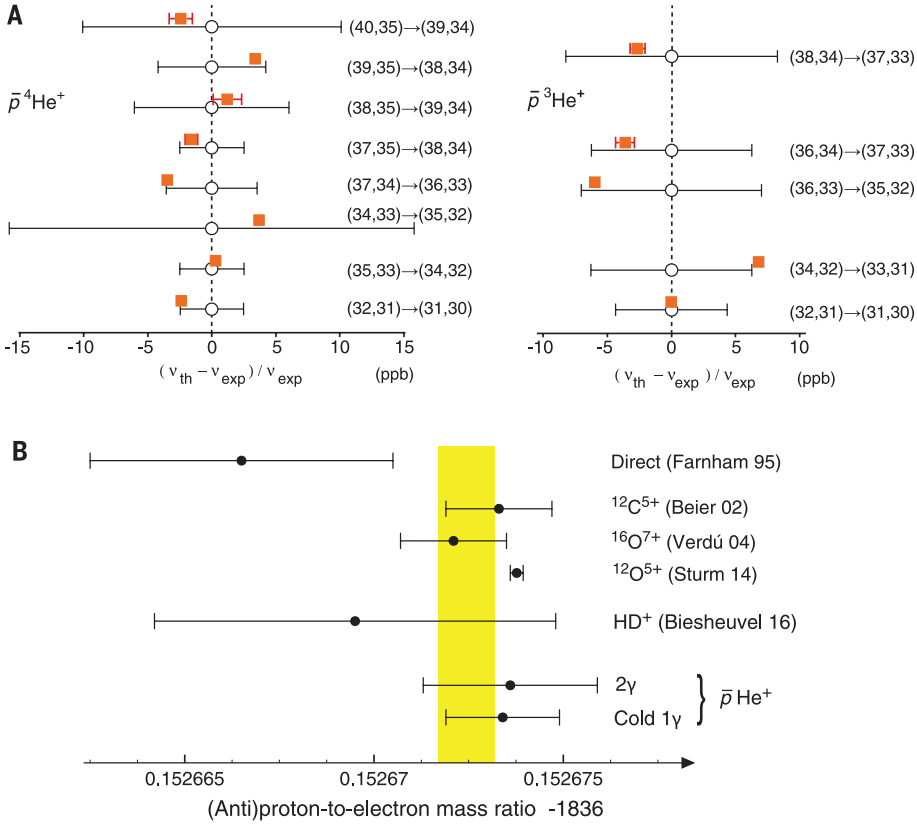



**Fig. 3. Single-photon transition frequencies of antiprotonic helium, and the antiproton-to-electron mass ratio determined from them.** (**A**) Comparison of experimental and calculated transition frequencies. Shown are the fractional differences between the experimental (open circles) and theoretical (squares) values of 13 transition frequencies of $\bar{p}^4$He$^+$ and $\bar{p}^3$He$^+$ atoms cooled to $T$ = 1.5 to 1.7 K. (**B**) Antiproton-to-electron mass ratio compared with proton-to-electron mass ratios. From top to bottom are the most precise direct measurement (*18*) of the proton-to-electron mass ratio obtained by contrasting the cyclotron frequencies of a proton and an electron confined in a Penning trap; indirect determinations (*19*–*21*) by measuring the magnetic moment of the bound electron in $^{12}C^{5+}$ and $^{16}O^{7+}$ ions in a Penning trap; laser spectroscopy of HD$^+$ molecules (*22*); antiproton-to-electron mass ratio obtained from two-photon laser spectroscopy of $\bar{p}^4$He$^+$ (*6*); and single-photon spectroscopy of cold atoms in this work. The yellow band represents the CODATA2010-recommended value (*17*). The CODATA2014 value is similar to the value of (*21*).

limited by the >1-GHz Auger width of the daughter state. Previous experiments could not resolve the fine structure in favored transitions of the type $(n, l) \rightarrow (n-1, l-1)$, but it is now revealed in the (39, 35) → (38, 34) resonance of cold $\bar{p}^4$He$^+$ (Fig. 2B). We also resolved the three-peak structure in the $\bar{p}^3$He$^+$ resonance (36, 34) → (37, 33) (Fig. 2D) arising from the eight unequally spaced hyperfine sublines caused by the additional interaction of the $^3$He nuclear spin. Altogether, we measured 13 resonances, which represent most of the metastable $\bar{p}$He$^+$ population (*28*). The fact that they were much narrower than the single-photon resonances measured previously (*23*) with inferred Doppler widths of 160 to 470 MHz relative to their transition frequencies of $\nu$ = 356 to 1133 THz, and no substantial broadened background component, indicates that nearly all of the atoms had cooled to $T$ = 1.5 to 1.7 K. Some of the transitions [ (34, 33)→(35, 32) of $\bar{p}^4$He$^+$ (fig. S3)] were broadened because of the short ($\tau$ = 170 ps) Auger lifetime of the resonance daughter state (*14*).

The spin-independent transition frequencies $\nu_{exp}$ (Table 1) were determined by fitting the profiles with a theoretical line shape (Fig. 2, blue lines). The line shape was obtained by solving the optical Bloch equations of the single-photon transitions between the $(2l + 1) \sim 70$ magnetic substates and taking into account power-broadening effects, thermal motion of the atoms, the spurious frequency modulation in the laser, and the spatial and temporal profiles of the laser beam. The positions of the hyperfine sublines were fixed to the theoretical values (*26*), the uncertainty of which affects $\nu_{exp}$ by 0.1 MHz (table S1).

Cold collisions with helium atoms deform the $\bar{p}$He$^+$ orbitals and shift the resonance frequencies (*27*, *29*). As shown in fig. S4, A and C, the frequency shifts of two $\bar{p}^4$He$^+$ transitions, (38, 35) → (39, 34) and (37, 35) → (38, 34), have the largest observed gradients $d\nu/d\rho$, measured at target densities between $\rho = 1.8 \times 10^{18}$ and $8 \times 10^{18}$ cm$^{-3}$. The much smaller shifts of other transitions, such as (32, 31) → (31, 30), were not resolvable within the 2-MHz experimental uncertainty. The measured shifts agree with quantum chemistry calculations (*27*) within ~20%. A linear extrapolation to $\rho$ = 0 yielded the zero-density frequencies; for transitions with small $d\nu/d\rho$, the gradients measured at higher densities $\rho > 1 \times 10^{20}$ cm$^{-3}$ were used (*23*).

The strong resonance signal from the large density of cold $\bar{p}$He$^+$ occupying the thermal distribution depopulated by the laser yielded a small statistical uncertainty $\sigma_{stat}$ = ±1 MHz on $\nu_{exp}$ of the $\bar{p}^4$He$^+$ transition (37, 35) → (38, 34), and $\sigma_{stat}$ = ±10 MHz for (34, 33) → (35, 32), which has a daughter state of large (1 GHz) natural width (fig. S3 and table S1). A systematic uncertainty of 0.4 to 3 MHz arose from the fitting function (*30*) itself. The magnetic Zeeman shifts in these transitions between Rydberg states are small (<0.1 MHz) under our experimental conditions. The spurious frequency modulation of the laser pulses (*24*, *25*) was measured with a precision of 0.4 to 1.0 MHz. The shift in $\nu_{exp}$ arising from laser fields that induce ac Stark effects (*30*) was ≤5 MHz in our earlier two-photon experiment carried out at higher laser fluences (*6*); the shifts were smaller (<0.2 MHz) for the single-photon transitions studied here.

All measured transition frequencies $\nu_{exp}$ (Fig. 3A, open circles with error bars) agree with theoretical $\nu_{th}$ values (filled squares) within the experimental uncertainties of $2.5 \times 10^{-9}$ to $15 \times 10^{-9}$. This agreement is 1.4 to 10 times better than in previous single-photon experiments (*23*). The uncertainties for most of the theoretical frequencies $\nu_{th}$ are due to uncalculated QED contributions of orders higher than $m_e\alpha^7$, whereas for some transitions [(37, 35) → (38, 34) of $\bar{p}^4$He$^+$] numerical uncertainties in the calculation of the relativistic Bethe logarithm (table S1) dominate. The corrections on $\nu_{th}$ due to the finite charge radii (*12*, *31*) of the helium nucleus (4 to 7 MHz) and of the antiproton (<1 MHz) are small because the Rydberg antiproton orbital has negligible overlap with the nucleus and is polarized away from the 1s electron.

The frequencies $\nu_{th}$ for the unfavored $(n, l) \rightarrow (n+1, l-1)$ transitions with low transition probabilities (*30*) between the $n \geq 36$ states changed by $2.6 \times 10^{-9}$ to $2.7 \times 10^{-9}$ when the $M_{\bar{p}}/m_e$ ratio used in the calculations was changed by $1 \times 10^{-9}$. These sensitivities $\kappa_M$ = 2.6 to 2.7 (Table 1) of the transitions (*5*) measured here are >2× larger than those of the low-$n$ or two-photon transitions studied earlier. The $M_{\bar{p}}/m_e$ value that yielded the best agreement between $\nu_{th}$ and $\nu_{exp}$ of the $\bar{p}^4$He$^+$ transition (37,35) → (38,34) deviated from the CODATA2010-recommended value for $M_p/m_e$ by $\delta_{M_p/m_e} = (7 \pm 10) \times 10^{-10}$. By minimizing $\sum[\nu_{th}(M_{\bar{p}}/m_e) - \nu_{exp}]^2_{\bar{p}}/\sigma^2_{stat}$ over the 13 $\bar{p}^4$He$^+$ frequencies, and considering the above systematic uncertainties ($\sigma_{sys}$), we obtained the ratio $M_{\bar{p}}/m_e$ = 1836.1526734(15). The 1-SD uncertainty in the parenthesis includes the contributions $9 \times 10^{-7}$, $11 \times 10^{-7}$, and $3 \times 10^{-7}$ of the experimental statistical and

systematic uncertainties and the theoretical uncertainty, respectively. This agrees with the determination $M_{\overline{p}}/m_e$ = 1836.1526736(23) obtained from sub-Doppler two-photon laser spectroscopy (*6*).

The atomic mass of the electron was recently determined (*21*) with a precision of $3 \times 10^{-11}$ by measuring the ratio between the cyclotron frequency and the precession frequency of the electron spin of a $^{12}C^{5+}$ ion confined in a Penning trap and comparing the results with QED calculations of its *g*-factor. From this and the known proton mass, the proton-to-electron mass ratio was determined as $M_p/m_e$ = 1836.15267377(17), which is larger by $4 \times 10^{-9}$ and $8 \times 10^{-10}$ than the ratios determined by comparing the cyclotron frequencies of protons and electrons in a Penning trap and the CODATA2010-recommended value. Our $M_{\overline{p}}/m_e$ value agrees with the result from (*21*) (Fig. 3B). The precision of our measurements here is a factor of 3.5 better than the recent determination of $M_p/m_e$ by means of laser spectroscopy of $HD^+$ molecules (*22*).

We used the analysis method of Hughes and Deutch (*32*–*36*) to constrain the equality between antiproton and proton masses $\delta_M = (M_p - M_{\overline{p}})/M_p$ and charges $\delta_Q = (Q_p + Q_{\overline{p}})/Q_p$ from these measured $\overline{p}^4He^+$ transition frequencies. The $\overline{p}^4He^+$ transition frequencies are proportional to $Q_{\overline{p}}^2 M_{\overline{p}}$ and have linear dependencies $\delta_M \kappa_M + \delta_Q \kappa_Q < |v_{exp} - v_{th}|/v_{th}$; the calculated sensitivities $\kappa_M$ and $\kappa_Q$ are shown in Table 1. The right side of the equation was evaluated as $<(3 \pm 15) \times 10^{-10}$ by averaging over the transitions and considering the statistical and systematic uncertainties. The Antihydrogen Trap (ATRAP) and Baryon Antibaryon Symmetry Experiment (BASE) collaborations (*34*–*36*) have, on the other hand, measured the cyclotron frequency ($\propto Q_{\overline{p}}/M_{\overline{p}}$) of antiprotons confined in Penning traps as $(Q_{\overline{p}}/M_{\overline{p}})/(Q_p/M_p) + 1 = 1.6(9) \times 10^{-10}$ (*34*, *35*) and $1(69) \times 10^{-12}$ (*36*), which implies that $\delta_Q \sim \delta_M$. By combining the two types of experimental results, we conclude that any deviation between the proton and antiproton masses and charges is $<5 \times 10^{-10}$ at the 90% confidence level. Recently, the ALPHA experiment has provided a limit of $<7 \times 10^{-10}$ on any residual charge of antihydrogen (*37*) by analyzing its movement in a magnetic bottle trap; when combined with the $\overline{p}He^+$ results, this constrains the equality of the electron and positron charges to a similar level.

## REFERENCES AND NOTES


1. C. Amole *et al.*, *Nature* **483**, 439–443 (2012).
2. G. Gabrielse *et al.*, *Phys. Rev. Lett.* **108**, 113002 (2012).
3. S. Mariazzi, P. Bettotti, R. S. Brusa, *Phys. Rev. Lett.* **104**, 243401 (2010).
4. D. B. Cassidy *et al.*, *Phys. Rev. A* **81**, 012715 (2010).
5. R. S. Hayano, M. Hori, D. Horváth, E. Widmann, *Rep. Prog. Phys.* **70**, 1995–2065 (2007).
6. M. Hori *et al.*, *Nature* **475**, 484–488 (2011).
7. R. Bluhm, V. A. Kostelecký, N. Russell, *Phys. Rev. Lett.* **82**, 2254–2257 (1999).
8. E. J. Salumbides, W. Ubachs, V. I. Korobov, *J. Mol. Spec.* **300**, 65–69 (2014).
9. J. Murata, S. Tanaka, *Class. Quantum Gravity* **32**, 033001 (2015).
10. W. N. Hardy, A. J. Berlinsky, L. A. Whitehead, *Phys. Rev. Lett.* **42**, 1042–1045 (1979).
11. S. C. Doret, C. B. Connolly, W. Ketterle, J. M. Doyle, *Phys. Rev. Lett.* **103**, 103005 (2009).
12. V. I. Korobov, L. Hilico, J.-P. Karr, *Phys. Rev. Lett.* **112**, 103003 (2014).
13. V. I. Korobov, L. Hilico, J.-P. Karr, *Phys. Rev. A* **89**, 032511 (2014).
14. V. I. Korobov, *Phys. Rev. A* **89**, 014501 (2014).
15. V. I. Korobov, L. Hilico, J.-P. Karr, *Hyperfine Interact.* **233**, 75–82 (2015).
16. A. Badertscher *et al.*, *Phys. Lett. B* **392**, 278–282 (1997).
17. P. J. Mohr, B. N. Taylor, D. B. Newell, *Rev. Mod. Phys.* **84**, 1527–1605 (2012).
18. D. L. Farnham, R. S. Van Dyck Jr., P. B. Schwinberg, *Phys. Rev. Lett.* **75**, 3598–3601 (1995).
19. T. Beier *et al.*, *Phys. Rev. Lett.* **88**, 011603 (2002).
20. J. Verdú *et al.*, *Phys. Rev. Lett.* **92**, 093002 (2004).
21. S. Sturm *et al.*, *Nature* **506**, 467–470 (2014).
22. J. Biesheuvel *et al.*, *Nat. Commun.* **7**, 10385 (2016).
23. M. Hori *et al.*, *Phys. Rev. Lett.* **96**, 243401 (2006).
24. Materials and methods are available as supplementary materials on *Science* Online.
25. M. Hori, A. Dax, *Opt. Lett.* **34**, 1273–1275 (2009).
26. V. I. Korobov, *Phys. Rev. A* **73**, 022509 (2006).
27. D. Bakalov, B. Jeziorski, T. Korona, K. Szalewicz, E. Tchoukova, *Phys. Rev. Lett.* **84**, 2350–2353 (2000).
28. M. Hori *et al.*, *Phys. Rev. Lett.* **89**, 093401 (2002).
29. V. I. Korobov, Z.-X. Zhong, Q.-L. Tian, *Phys. Rev. A* **92**, 052517 (2015).
30. M. Hori, V. I. Korobov, *Phys. Rev. A* **81**, 062508 (2010).
31. R. Pohl *et al.*, *Nature* **466**, 213–216 (2010).
32. R. J. Hughes, B. I. Deutch, *Phys. Rev. Lett.* **69**, 578–581 (1992).
33. K. A. Olive *et al.*, *Chin. Phys. C* **38**, 090001 (2014).
34. G. Gabrielse *et al.*, *Phys. Rev. Lett.* **82**, 3198–3201 (1999).
35. J. K. Thompson, S. Rainville, D. E. Pritchard, *Nature* **430**, 58–61 (2004).
36. S. Ulmer *et al.*, *Nature* **524**, 196–199 (2015).
37. M. Ahmadi *et al.*, *Nature* **529**, 373–376 (2016).


## ACKNOWLEDGMENTS

We are deeply indebted to the CERN Antiproton Decelerator and Proton Synchrotron operational staff, the CERN cryogenics laboratory, the Atomic Spectroscopy And Collisions Using Slow Antiprotons (ASACUSA) collaboration, D. Bakalov, L. Dufay-Chanat, T. W. Hänsch, T. Koetting, V. I. Korobov, J. Mouleyre, W. Pirkl, and O. Pirotte. This work was supported by the Max-Planck-Gesellschaft, the European Research Council (ERC-Stg), Monbukagakusho, and the Hungarian Research Foundation.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/610/suppl/DC1
Materials and Methods
Figs. S1 to S4
Table S1
References (*38*–*41*)

17 May 2016; accepted 3 October 2016
10.1126/science.aaf6702

OPTICAL PROCESSING

# A fully programmable 100-spin coherent Ising machine with all-to-all connections

**Peter L. McMahon,[1,2]*† Alireza Marandi,[1]*† Yoshitaka Haribara,[2,3,4] Ryan Hamerly,[1] Carsten Langrock,[1] Shuhei Tamate,[2] Takahiro Inagaki,[5] Hiroki Takesue,[5] Shoko Utsunomiya,[2] Kazuyuki Aihara,[3,4] Robert L. Byer,[1] M. M. Fejer,[1] Hideo Mabuchi,[1] Yoshihisa Yamamoto[1,6]**

**Unconventional, special-purpose machines may aid in accelerating the solution of some of the hardest problems in computing, such as large-scale combinatorial optimizations, by exploiting different operating mechanisms than those of standard digital computers. We present a scalable optical processor with electronic feedback that can be realized at large scale with room-temperature technology. Our prototype machine is able to find exact solutions of, or sample good approximate solutions to, a variety of hard instances of Ising problems with up to 100 spins and 10,000 spin-spin connections.**

Combinatorial optimization problems, including many nondeterministic polynomial-time–hard (NP-hard) problems, are central in numerous important application areas, including operations and scheduling, drug discovery, finance, circuit design, sensing, and manufacturing. Despite large advances in both algorithms and digital computer technology, even typical instances of NP-hard problems that arise in practice may be very difficult to solve on conventional computers. There is a long history of attempts to find alternatives to current von Neumann–computer–based methods for solving such problems, including use of neural networks realized with analog electronic circuits (*1*, *2*) and by using molecular computing (*3*–*5*). Both lines of investigation continue to inspire related work (*6*, *7*). A major topic of contemporary interest is the study of adiabatic quantum computation (AQC) (*8*) and quantum annealing (QA) (*9*, *10*). Sophisticated AQC/QA devices are already under study (*11*–*14*), but providing dense connectivity

[1]E. L. Ginzton Laboratory, Stanford University, Stanford, CA 94305, USA. [2]National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan. [3]Department of Mathematical Informatics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. [4]Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan. [5]NTT Basic Research Laboratories, 3-1 Morinosato, Wakamiya, Atsugi, Kanagawa 243-0198, Japan. [6]ImPACT Program, Japan Science and Technology Agency, Gobancho 7, Chiyoda-ku, Tokyo 102-0076, Japan.
***Corresponding author. Email: pmcmahon@stanford.edu (P.L.M.); marandi@stanford.edu (A.M.)** †These authors contributed equally to this work.

between qubits remains a major challenge (*15*), with important implications for the efficiency of AQC/QA systems (*16*).

Networks of coupled optical parametric oscillators (OPOs) are an alternative physical system, with an unconventional operating mechanism (*17*–*20*), for solving the Ising problem (*21*, *22*) and by extension many other combinatorial optimization problems (*23*). Formally, the $N$-spin Ising problem is to find the configuration of spins $\sigma_i \in \{-1, +1\}$ ($i$ = 1, ..., $N$) that minimizes the energy function $H = -\sum_{1 \le i < j \le N} J_{ij}\sigma_i\sigma_j - \sum_{1 \le i \le N} h_i\sigma_i$, where the particular problem instance being solved is specified by the $N \times N$ matrix $J$ (with elements $J_{ij}$) and the length-$N$ vector $h$ (with elements $h_i$).

We have realized a system with a scalable architecture that uses measurement feedback in place of a network of optical delay lines [which were used in initial, low-connectivity, nonreprogrammable demonstrations of the concept (*18*, *24*, *25*)]. Our 100-spin Ising machine allows connections between any spin and any other spin and is fully programmable. We show that measurement-feedback–based OPO Ising machines can solve many different Ising problems, and in cases in which exact solutions are not easy to obtain, we can find good approximate solutions.

The schematic of our experimental setup (Fig. 1) shows that our Ising machine is formed by the combination of time-division–multiplexed OPOs (*18*) in a single fiber-ring cavity, with measurement and feedback (injection) stages that act to couple the pulses in the cavity such that the Ising Hamiltonian is realized. Details are provided in the supplementary materials (*26*).

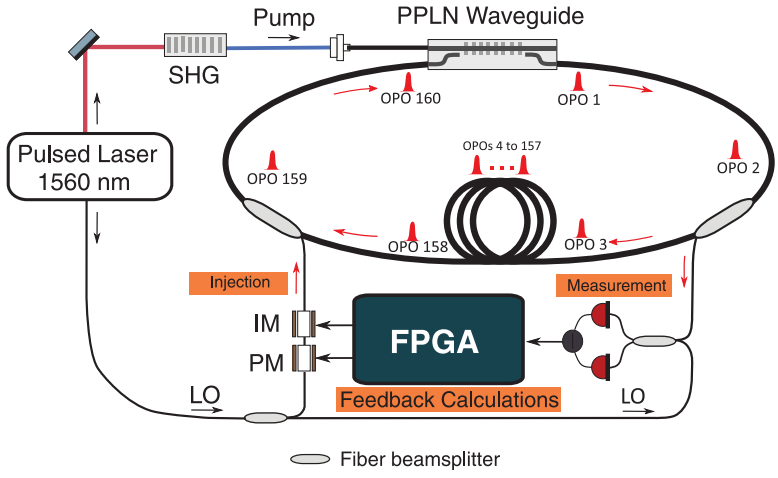



**Fig. 1. Experimental schematic of a measurement-feedback–based coherent Ising machine.** A time-division–multiplexed pulsed degenerate optical parametric oscillator is formed by a nonlinear crystal [periodically poled lithium niobate (PPLN)] in a fiber ring cavity containing 160 pulses. A fraction of each pulse is measured and used to compute a feedback signal that effectively couples the otherwise-independent pulses in the cavity. IM, intensity modulator; PM, phase modulator; LO, local oscillator; SHG, second-harmonic generation; FPGA, field-programmable gate array.

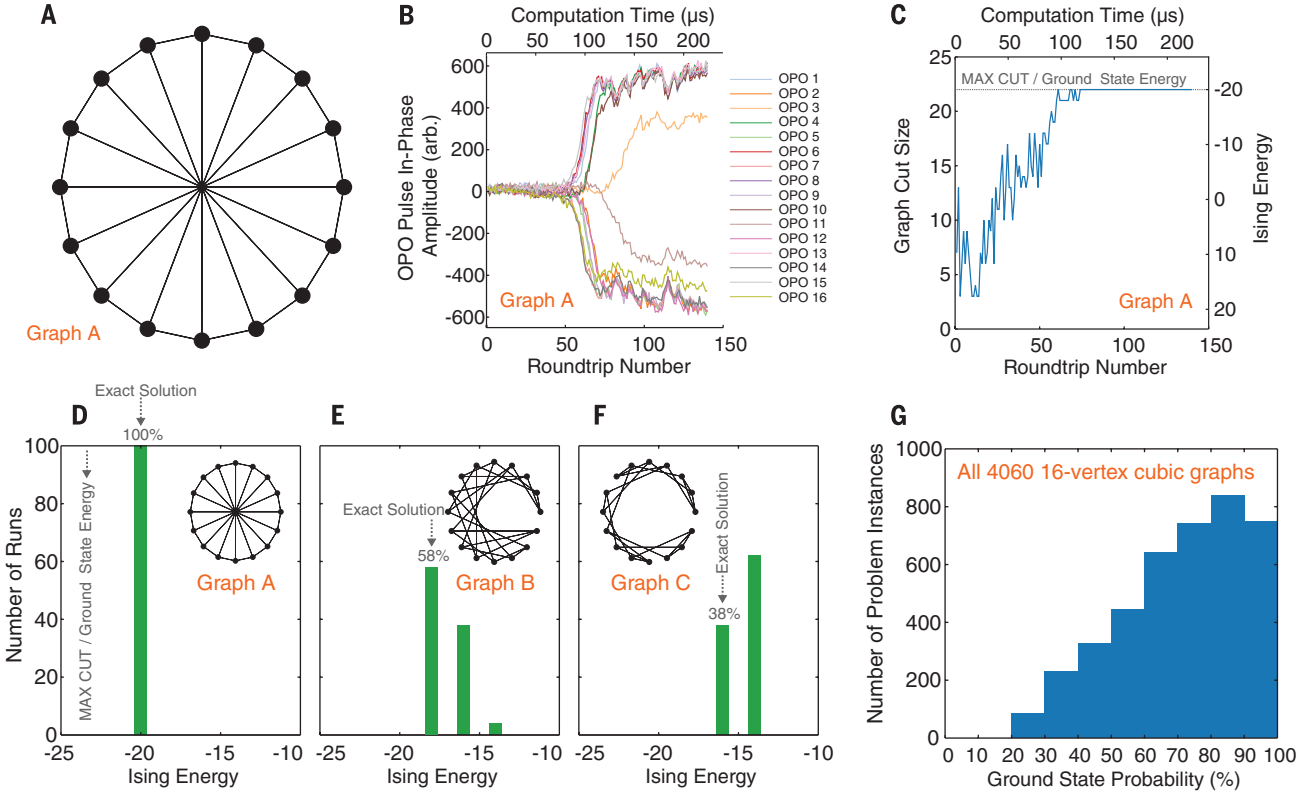



**Fig. 2. Results with $N$ = 16 cubic graphs.** (**A**) A Möbius ladder graph with $N$ = 16 vertices. (**B**) The evolution of the in-phase components $c_i$ of the $N$ = 16 OPO pulses as a function of the computation time. (**C**) The graph cut size achieved as a function of the computation time. (**D** to **F**) Histograms of obtained solutions in 100 runs for the graphs shown in the insets. (**G**) Histogram of the observed probabilities of obtaining a ground state in a single run, for all 4060 unweighted, undirected cubic graphs with $N$ = 16 vertices.

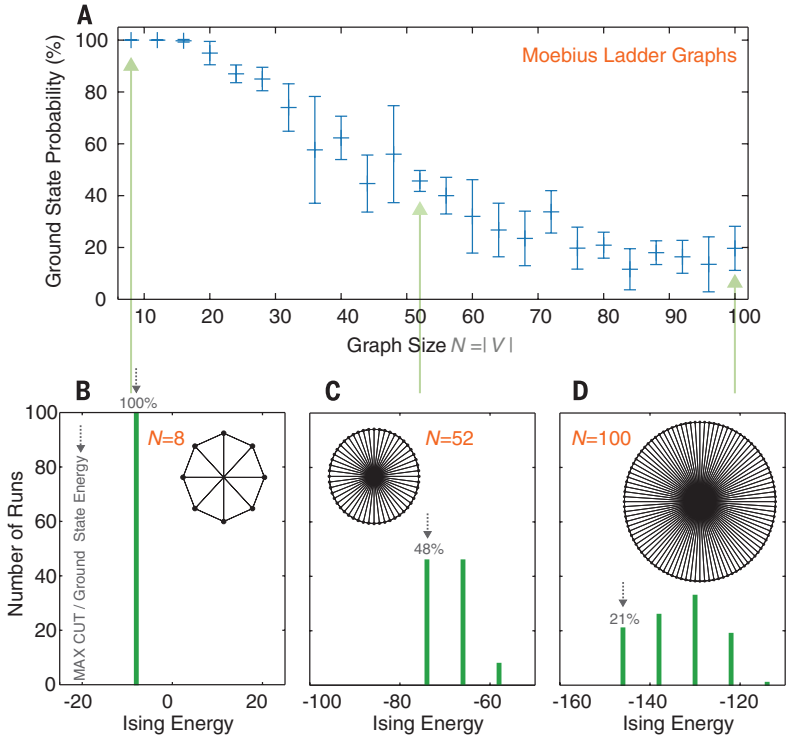



**Fig. 3. Results with various-size Möbius ladder graphs.** (**A**) Observed probability of obtaining a ground state of the Möbius ladder graph in a single run, as a function of the size *N* of the graph. Multiple 100-run batches were performed for each graph size to obtain the standard deviations, which are shown as error bars. (**B** to **D**) Histograms of obtained solutions in 100 runs for the graphs shown in the insets.

We have focused our experiments on Ising problems on undirected, unweighted graphs ($V$, vertices; $E$, edges), where $J_{ij} = -1$ when spin $i$ and spin $j$ are connected, and $J_{ij} = 0$ otherwise, and for which the linear (Zeeman) terms $h_i$ are zero. An Ising problem of this form is equivalent to the Max-Cut problem on the underlying graph (*17*). Max-Cut is the problem of partitioning the vertices of a graph into two disjoint subsets such that the number of edges between the two subsets is maximized; the partition is called a cut. Max-Cut remains NP-hard even when the input is restricted to unweighted cubic graphs (*27*). We refer interchangeably to the Max-Cut problem and to the Ising problem, and the solutions thereof. The energy of a particular spin configuration $\{\sigma_i\}_{i=1,\ldots,N}$ for an Ising problem is given by $H(\{\sigma_i\}) = -\sum_{1\le i<j\le N} J_{ij}\sigma_i\sigma_j - \sum_{1\le i\le N} h_i\sigma_i$. The same spin configuration $\{\sigma_i\}$ represents a cut of size $C(\{\sigma_i\}) = -\frac{1}{2}\sum_{1\le i<j\le N} J_{ij} - \frac{1}{2}H(\{\sigma_i\})$; there is a direct mapping between the energies in the Ising problem and the cut sizes in the Max-Cut problem, and minimizing the Ising energy maximizes the cut.

The unweighted, undirected Möbius ladder (cubic) graph with $N = 16$ vertices is shown in Fig. 2A. We programmed the corresponding $J$ matrix into our feedback electronics and ran

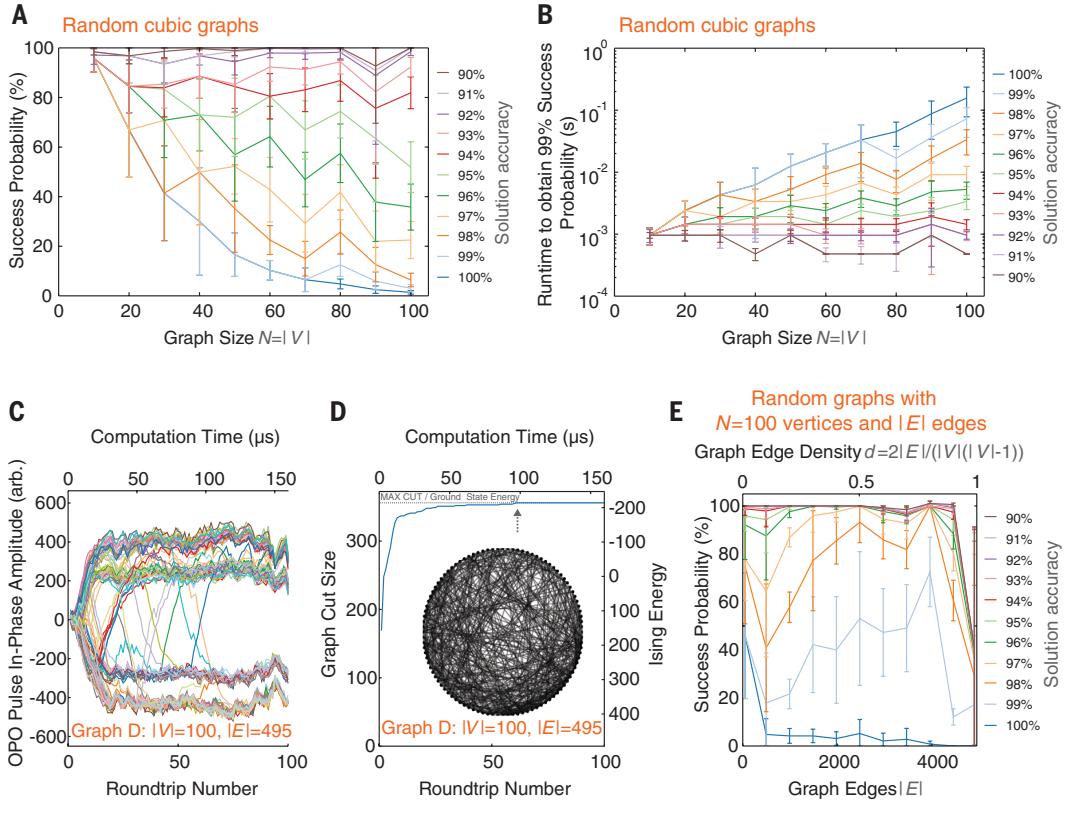



**Fig. 4. Results with various-size and various-density random graphs.** (**A**) Observed probability of obtaining a solution whose cut size is at least *x*% of the global optimum (maximum cut), as a function of graph size *N*, for random cubic graph instances. Error bars indicate 1 SD, which is dominated by the difference in difficulty between the various problem instances. (**B**) The runtime that would be required to obtain a solution of a particular accuracy with 99% probability. (**C**) The evolution of the in-phase components $c_i$ of the $N = 100$ OPO pulses as a function of the computation time, for a single run with the graph shown in the (D) inset. (**D**) The graph cut size achieved as a function of the computation time. (Inset) The graph being solved. (**E**) Observed success probability of obtaining a solution with a particular accuracy as a function of the density of edges in the graph. Experiments were performed on randomly generated $N = 100$-vertex graphs with fixed numbers of edges. Error bars indicate 1 SD.

the problem on our experimental apparatus. Figure 2B shows, from a single run of this single problem instance, the evolution of the in-phase component of each OPO pulse ($c_i$) as a function of the number of times each pulse circulated around the cavity (the number of round trips). The computation time is given by $T_{comp} = T_{rt}N_{rt}$, where $T_{rt}$ = 1.6 μs is the single round trip time and $N_{rt}$ is the number of round trips. Each spin $\sigma_i$ in the Ising problem is represented by a single OPO pulse; if the in-phase component $c_i$ of the $i$th OPO pulse is less than zero ($c_i < 0$), we make the spin assignment $\sigma_i = -1$, and if the in-phase component is greater than zero ($c_i > 0$), we make the spin assignment $\sigma_i = +1$. As the feedback signal level is gradually increased, the OPO amplitudes increase. Most of the OPOs obtain their ultimate signs by round trip 60. By round trip 100, all the OPOs have reached a steady state. Shown in Fig. 2C is the graph cut size, or equivalently the Ising energy, represented by the spin configuration after each roundtrip for the run shown in Fig. 2B. We see that the system evolves toward solutions representing larger graph cuts, or equivalently lower Ising energies, and the steady-state configuration is that of a ground state. The system finds the ground state within 120 μs. If we rerun the computation many times, we find that our system always returns the optimal solution for this particular problem instance (Fig. 2D). The distributions of obtained solutions for two other $N$ = 16 cubic graphs are shown in Fig. 2, E and F. In both cases, the system finds the ground state in a large fraction of the runs, and when the system does not return a ground state, it returns a state that has energy close to that of the ground state, a good approximate solution. To illustrate that these instances are not special cases for which the Ising machine finds exact solutions (ground states), we attempted to solve all possible problem instances of $N$ = 16 cubic graphs, of which there are 4060. We were able to find ground states with probability greater than 20% for every single instance (Fig. 2G). In this experiment, and all that follow, every run was set to proceed for $N_{rt}$ = 300 round trips, a computation time of $T_{comp}$ = 480 μs.

We next explored how the Ising machine performs as a function of the size of the problem. In this first scaling study, we used the Möbius ladder graphs with varying size. This is a convenient choice because there is a closed-form solution for the maximum cut and Ising energy for every instance in this family of graphs. The Möbius ladder graphs with $N$ = 8, 12, ..., 100 were solved (Fig. 3A). For each instance, we performed multiple sets of 100 runs. Each 100-run set resulted in a solution-energy histogram (an example for $N$ = 8 is shown in Fig. 3B). For the runs in which the ground state is not found, the system again finds good approximate solutions (Fig. 3, C and D).

To understand how the performance of our system scales as a function of problem size for more general problems, 10 random instances of cubic graphs were generated for each graph size $N$ = 10, 20, ..., 100, and each instance run on our apparatus. The success probabilities for each instance are shown in Fig. 4A, aggregated by graph size. In contrast to the previous results, not only the probability of obtaining a ground state but also the probability of obtaining a solution that is within $x$% of the optimal (maximum cut) solution is shown. The error bars indicate a spread in success probabilities primarily due to the fact that different random instances of each size may be more or less difficult to solve.

Figure 4B shows, for each graph size, the total computation time required to obtain a solution with the given accuracy, with probability 99% within that time, and is derived directly from Fig. 4A. The total time is given by $T_{comp} \times \lceil \log(1 - 0.99)/\log(1 - p_s) \rceil$, where $p_s$ is the corresponding success probability from Fig. 4A. The total computation time required to obtain ground states (100% accuracy) grows rapidly with problem size $N$, although it is still small in absolute terms for $N$ = 100: less than 200 ms. The growth in total computation time is far less severe when the required solution accuracy is reduced; for 95% accuracy, the required time increases from ~1 ms when $N$ = 10 to ~4 ms when $N$ = 100; for 90% accuracy, the required time remains roughly constant at ~1 ms.

The OPO amplitudes and cut sizes/energies are shown in Fig. 4, C and D, respectively, as a function of the number of roundtrips for a single run on a single nonregular random graph instance with $N$ = 100 vertices and $|E| = 495$ edges. We see that the global optimum is reached within 100 μs, and the system reaches a steady state after ~120 μs. This graph instance is one of 10 instances that were generated having 495 edges. We generated 10 random graphs with a fixed number of edges, for each of $|E| = 50, 495, 990, 1485, 1980, 2475, 2970, 3465, 3960, 4455, 4900$, and ran these instances on our apparatus. The results of this study of the performance on random graphs as a function of edge density $d = 2|E|/[|V|(|V| - 1)]$ are shown in Fig. 4E; the system was able to find good solutions for all the attempted densities.

OPO Ising machines have the potential to harness a number of quantum features, including squeezing and superposition (*20*). Pulsed OPOs already have a substantial quantum-mechanical nature (*28*), and networks of OPOs can generate spatial multimode entanglement (*19*). The experimental system reported in this work is well-described by a quantum-mechanical model (*20*) [with very good agreement between our experimental results and numerical simulation results (*26*)]. However, the extent to which classical models can capture the behavior of OPO Ising machines is not yet known, and follow-up studies are needed to determine the fundamental computational power of OPO Ising machines. Two future modifications to the experiment that could increase the relevance of quantum mechanics are reducing the cavity round-trip loss and injecting squeezed vacuum states into the open port of the measurement beamsplitter (*29*), which is predicted to improve the success probability (*26*).

Although we find the overall direction of OPO Ising machines promising, the techniques we have demonstrated are not necessarily restricted to OPO Ising machines, and photonic-AQC and boson-sampling (*30*) experiments, among others, may be able to adapt our methods.

## REFERENCES AND NOTES


1. J. J. Hopfield, D. W. Tank, *Science* **233**, 625–633 (1986).
2. D. Tank, J. Hopfield, *IEEE Trans. Circ. Syst.* **33**, 533–541 (1986).
3. L. M. Adleman, *Science* **266**, 1021–1024 (1994).
4. R. J. Lipton, *Science* **268**, 542–545 (1995).
5. R. S. Braich, N. Chelyapov, C. Johnson, P. W. K. Rothemund, L. Adleman, *Science* **296**, 499–502 (2002).
6. M. Ercsey-Ravasz, Z. Toroczkai, *Nat. Phys.* **7**, 966–970 (2011).
7. L. Qian, E. Winfree, J. Bruck, *Nature* **475**, 368–372 (2011).
8. E. Farhi *et al.*, *Science* **292**, 472–475 (2001).
9. T. Kadowaki, H. Nishimori, *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* **58**, 5355–5363 (1998).
10. J. Brooke, D. Bitko, T. F. Rosenbaum, G. Aeppli, *Science* **284**, 779–781 (1999).
11. M. W. Johnson *et al.*, *Nature* **473**, 194–198 (2011).
12. T. F. Rønnow *et al.*, *Science* **345**, 420–424 (2014).
13. S. Boixo *et al.*, *Nat. Phys.* **10**, 218–224 (2014).
14. R. Barends *et al.*, *Nature* **534**, 222–226 (2016).
15. P. I. Bunyk *et al.*, *IEEE Trans. Appl. Supercond.* **24**, 1–10 (2014).
16. E. G. Rieffel *et al.*, *Quant. Inform. Process.* **14**, 1–36 (2014).
17. Z. Wang, A. Marandi, K. Wen, R. L. Byer, Y. Yamamoto, *Phys. Rev. A* **88**, 063853 (2013).
18. A. Marandi, Z. Wang, K. Takata, R. L. Byer, Y. Yamamoto, *Nat. Photonics* **8**, 937–942 (2014).
19. K. Takata, A. Marandi, Y. Yamamoto, *Phys. Rev. A* **92**, 043821 (2015).
20. Y. Haribara, S. Utsunomiya, Y. Yamamoto, *Entropy* **18**, 151 (2016).
21. F. Barahona, *J. Phys. Math. Gen.* **15**, 3241–3253 (1982).
22. S. Istrail, *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing* (STOC, 2000), pp. 87–96.
23. A. Lucas, *Front. Phys.* **2**, 5 (2014).
24. K. Takata *et al.*, *Sci. Rep.* **6**, 340897 (2016).
25. T. Inagaki *et al.*, *Nat. Photonics* **10**, 415–419 (2016).
26. Materials and methods are available as supplementary materials on *Science* Online.
27. M. R. Garey, D. S. Johnson, L. Stockmeyer, *Theor. Comput. Sci.* **1**, 237–267 (1976).
28. J. Roslund, R. M. de Araújo, S. Jiang, C. Fabre, N. Treps, *Nat. Photonics* **8**, 109 (2014).
29. W. J. Munro, M. D. Reid, *Phys. Rev. A* **52**, 2388 (1995).
30. K. R. Motes, A. Gilchrist, J. P. Dowling, P. P. Rohde, *Phys. Rev. Lett.* **113**, 120501 (2014).


## ACKNOWLEDGMENTS

This research was funded by the Impulsing Paradigm Change through Disruptive Technologies (ImPACT) Program of the Council of Science, Technology and Innovation (Cabinet Office, Government of Japan). P.L.M. was partially supported by a Stanford Nano- and Quantum Science and Engineering Postdoctoral Fellowship. We thank K. Leedle, A. Ceballos, K. Wen, and Z. Wang for technical assistance and M. Digonnet, B. Lantz, T. Onodera, E. Ng, T. Leleu, C. Limouse, D. Gray, G. Tabak, and N. Tezak for helpful discussions.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/614/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S5
References (*31*–*58*)

8 July 2016; accepted 26 September 2016
10.1126/science.aah5178

CANCER ETIOLOGY

# Mutational signatures associated with tobacco smoking in human cancer

Ludmil B. Alexandrov,[1,2,3]* Young Seok Ju,[4] Kerstin Haase,[5] Peter Van Loo,[5,6] Iñigo Martincorena,[7] Serena Nik-Zainal,[7,8] Yasushi Totoki,[9] Akihiro Fujimoto,[10,11] Hidewaki Nakagawa,[10] Tatsuhiro Shibata,[9,12] Peter J. Campbell,[7,13] Paolo Vineis,[14,15] David H. Phillips,[16] Michael R. Stratton[7]*

**Tobacco smoking increases the risk of at least 17 classes of human cancer. We analyzed somatic mutations and DNA methylation in 5243 cancers of types for which tobacco smoking confers an elevated risk. Smoking is associated with increased mutation burdens of multiple distinct mutational signatures, which contribute to different extents in different cancers. One of these signatures, mainly found in cancers derived from tissues directly exposed to tobacco smoke, is attributable to misreplication of DNA damage caused by tobacco carcinogens. Others likely reflect indirect activation of DNA editing by APOBEC cytidine deaminases and of an endogenous clocklike mutational process. Smoking is associated with limited differences in methylation. The results are consistent with the proposition that smoking increases cancer risk by increasing the somatic mutation load, although direct evidence for this mechanism is lacking in some smoking-related cancer types.**

Tobacco smoking has been associated with at least 17 types of human cancer (Table 1) and claims the lives of more than 6 million people every year (*1*–*4*). Tobacco smoke is a complex mixture of chemicals, among which at least 60 are carcinogens (*5*). Many of these are thought to cause cancer by inducing DNA damage that, if misreplicated, leads to an increased burden of somatic mutations and, hence, an elevated chance of acquiring driver mutations in cancer genes. Such damage often occurs in the form of covalent bonding of metabolically activated reactive species of the carcinogen to DNA bases, termed DNA adducts (*6*). Tissues directly exposed to tobacco smoke (e.g., lung), as well as some tissues not directly exposed (e.g., bladder), show elevated levels of DNA adducts in smokers and, thus, evidence of exposure to carcinogenic components of tobacco smoke (*7*, *8*).

Each biological process causing mutations in somatic cells leaves a mutational signature (*9*). Many cancers have a somatic mutation in the *TP53* gene, and catalogs of *TP53* mutations compiled two decades ago enabled early exploration of these signatures (*10*), showing that lung cancers from smokers have more C>A transversions than lung cancers from nonsmokers (*11*–*14*). To investigate mutational signatures using the thousands of mutation catalogs generated by systematic cancer genome sequencing, we recently described a framework in which each base substitution signature is characterized using a 96-mutation classification that includes the six substitution types together with the bases immediately 5′ and 3′ to the mutated base (*15*). The analysis extracts mutational signatures from mutation catalogs and estimates the number of mutations contributed by each signature to each cancer genome (*15*). Using this approach, more than 30 different base substitution signatures have been identified (*16*–*18*).

In this study, we examined 5243 cancer genome sequences (4633 exomes and 610 whole genomes) of cancer classes for which smoking increases risk, with the goal of identifying mutational signatures and methylation changes associated with tobacco smoking (table S1). Of the samples we studied, 2490 were reported to be from tobacco smokers and 1063 from never-smokers (Table 1). Thus, we were able to investigate the mutational consequences of smoking by comparing somatic mutations and methylation in smokers versus nonsmokers for lung, larynx, pharynx, oral cavity, esophageal, bladder, liver, cervical, kidney, and pancreatic cancers (Fig. 1 and table S2).

We first compared total numbers of base substitutions, small insertions and deletions (indels), and genomic rearrangements. The total number of base substitutions was higher in smokers compared with nonsmokers for all cancer types together ($q$-value < 0.05) and, for individual cancer types, in lung adenocarcinoma, larynx, liver, and kidney cancers (table S2). Total numbers of indels were higher in smokers compared with nonsmokers in lung adenocarcinoma and liver cancer (table S2). The whole-genome–sequenced cases allowed comparison of genome rearrangements between smokers and nonsmokers in pancreatic and liver cancer, where no differences were found (table S2). However, subchromosomal copy-number changes entail genomic rearrangement and can serve as surrogates for rearrangements. Lung adenocarcinomas from smokers exhibited more copy-number aberrations than those from nonsmokers (table S2).

We then extracted mutational signatures, estimated the contributions of each signature to each cancer, and compared the numbers of mutations attributable to each signature in smokers and nonsmokers. Increases in smokers compared with nonsmokers were seen for signatures 2, 4, 5, 13, and 16 [the mutational signature nomenclature is that used in the Catalogue of Somatic Mutations in Cancer (COSMIC) and in (*16*–*18*)]. There was sufficient statistical power to show that these increases were of clonal mutations (mutations present in all cells of each cancer) for signatures 4 and 5 ($q$ < 0.05), as expected if these mutations are due to cigarette smoke exposure before neoplastic change (supplementary text).

Signature 4 is characterized mainly by C>A mutations with smaller contributions from other base substitution classes (Fig. 2B and fig. S1). This signature was found only in cancer types in which tobacco smoking increases risk and mainly in those derived from epithelia directly exposed to tobacco smoke (figs. S2 and S3). Signature 4 is very similar to the mutational signature induced in vitro by exposing cells to benzo[$a$]pyrene (cosine similarity = 0.94) (Fig. 2B and fig. S3), a tobacco smoke carcinogen (*19*). The similarity extends to the presence of a transcriptional strand bias indicative of transcription-coupled nucleotide excision repair (NER) of bulky DNA adducts on guanine (fig. S1), the proposed mechanism of DNA damage by benzo[$a$]pyrene. Thus, signature 4 is likely the direct mutational consequence of misreplication of DNA damage induced by tobacco carcinogens.

Most lung and larynx cancers from smokers had many signature 4 mutations. Signature 4 mutations occurred more often in cancers from

[1]Theoretical Biology and Biophysics (T-6), Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [2]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA. [3]University of New Mexico Comprehensive Cancer Center, Albuquerque, NM 87102, USA. [4]Graduate School of Medical Science and Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea. [5]The Francis Crick Institute, 1 Midland Road, London NW1 1AT, UK. [6]Department of Human Genetics, University of Leuven, 3000 Leuven, Belgium. [7]Cancer Genome Project, Wellcome Trust Sanger Institute, Hinxton CB10 1SA, Cambridgeshire, UK. [8]Department of Medical Genetics, Addenbrooke's Hospital National Health Service Trust, Cambridge, UK. [9]Division of Cancer Genomics, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan. [10]Laboratory for Genome Sequencing Analysis, RIKEN Center for Integrative Medical Sciences, Tokyo, Japan. [11]Department of Drug Discovery Medicine, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan. [12]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Minato-ku, Tokyo, Japan. [13]Department of Haematology, University of Cambridge, Cambridge CB2 0XY, UK. [14]Human Genetics Foundation, 10126 Torino, Italy. [15]Department of Epidemiology and Biostatistics, Medical Research Council (MRC)–Public Health England (PHE) Centre for Environment and Health, School of Public Health, Imperial College London, Norfolk Place, London W2 1PG, UK. [16]King's College London, MRC-PHE Centre for Environment and Health, Analytical and Environmental Sciences Division, Franklin-Wilkins Building, 150 Stamford Street, London SE1 9NH, UK.
*Corresponding author. Email: lba@lanl.gov (L.B.A.); mrs@sanger.ac.uk (M.R.S.)

smokers compared with nonsmokers in all cancer types together (table S2) and in lung squamous, lung adenocarcinoma, and larynx cancers (table S2). This finding largely accounts for differences in total numbers of base substitutions (Table 1). In nonsmokers, 13.8% of lung cancers showed many signature 4 mutations (Fig. 2A; >1 mutation per megabase), which may be due to passive smoking, misreporting of smoking habits, or annotation errors. Signature 4 mutations were also detected in cancers of the oral cavity, pharynx, and esophagus, albeit in much smaller numbers than in lung and larynx cancers, perhaps because of reduced exposure to tobacco smoke or more efficient clearance. Differences in mutation burden attributed to signature 4 between smokers and nonsmokers were not observed in these cancer types (Fig. 1). Signature 4 mutations were found at low levels in cancers of the liver, an organ not directly exposed to tobacco smoke, and were elevated in smokers versus nonsmokers (Fig. 1).

Signature 4 was not extracted from bladder, cervical, kidney, or pancreatic cancers, despite the known risks conferred by smoking and the presence of many smokers in these series. Additionally, this mutational signature was not extracted from cancers of the stomach, colorectum, and ovary, nor from acute myeloid leukemia (in the analyzed series, the smoking status of patients with these cancers was unknown, but it is likely that many have been smokers). The tissues from which all of these cancer types are derived are not directly exposed to tobacco smoke. Simulations indicate that the lack of signature 4 is not due to statistical limitations (supplementary text and fig. S4). The absence of signature 4 suggests that misreplication of direct DNA damage due to tobacco smoke constituents does not contribute substantially to mutation burden in these cancers, even though DNA adducts indicative of tobacco-induced DNA damage are present in the tissues from which they arise (*7*).

Signatures 2 and 13 are characterized by C>T and C>G mutations, respectively, at TpC dinucleotides and have been attributed to overactive DNA editing by APOBEC deaminases (*20*, *21*). The cause of the overactivity in most cancers has not been established, although APOBECs are implicated in the cellular response to the entrance of foreign DNA, retrotransposon movement, and local inflammation (*22*). Signatures 2 and 13 showed more mutations in smokers versus nonsmokers with lung adenocarcinoma (table S2). Because these signatures are found in many other cancer types, where they are apparently unrelated to tobacco smoking, it seems unlikely that the signature 2 and 13 mutations associated with smoking in lung adenocarcinoma are direct consequences of misreplication of DNA damage induced by tobacco smoke. More plausibly, the cellular machinery underlying signatures 2 and 13 is activated by tobacco smoke, perhaps as a result of inflammation arising from the deposition of particulate matter or by indirect consequences of DNA damage.

Signature 5 is characterized by mutations distributed across all 96 subtypes of base substitution, with a predominance of T>C and C>T mutations (Fig. 2B) and evidence of transcriptional strand bias for T>C mutations (*18*). Signature 5 is found in all cancer types, including those unrelated to tobacco smoking, and in most cancer samples. It is "clocklike" in that the number of mutations attributable to this signature correlates with age at the time of diagnosis in many cancer types (*17*). Signature 5, together with signature 1, is thought to contribute to mutation accumulation in most normal somatic cells and in the germline (*17*, *23*). The mechanisms underlying signature 5 are not well understood, although an enrichment of signature 5 mutations was found in bladder cancers harboring inactivating

**Table 1. Mutational signatures and cancer types associated with tobacco smoking.** Information about the age-adjusted odds ratios for current male smokers to develop cancer is taken from (*2*–*4*). Odds ratios for small cell lung cancer, squamous cell lung cancer, and lung adenocarcinoma are for an average daily dose of more than 30 cigarettes. Odds ratios for cervical and ovarian cancers are for current female smokers. Detailed information about all mutation types, all mutational signatures, and DNA methylation is provided in table S2. Nomenclature for signature identification numbers is consistent with the COSMIC database (http://cancer.sanger.ac.uk/cosmic/signatures). The numbers of smokers and nonsmokers are unknown (i.e., not reported in the original studies) for acute myeloid leukemia, stomach, ovarian, and colorectal cancers. The patterns of all mutational signatures with elevated mutation burden in smokers are displayed in Fig. 2B. N/A denotes lack of smoking annotation for a given cancer type. Asterisks indicate that a signature correlates with pack years smoked in a cancer type. N.S. reflects cancer types without statistically significant elevation of mutational signatures. The odds ratio for all cancer types is not provided.

| Cancer type | Odds ratio | Nonsmokers | Smokers | Total number of mutational signatures found in the cancer type | Signature 4 found in cancer type | Mutational signatures with elevated mutation burden in smokers versus nonsmokers ($q < 0.05$) |
|---|---|---|---|---|---|---|
| All cancer types | ND | 1062 | 2490 | 26 | Y | 4, 5* |
| Small cell lung cancer | 111.3 | 3 | 145 | 6 | Y | N.S. |
| Lung squamous | 103.5 | 7 | 168 | 8 | Y | 4*, 5 |
| Lung adenocarcinoma | 21.9 | 120 | 558 | 7 | Y | 2*, 4*, 5*, 13* |
| Larynx | 13.2 | 6 | 117 | 5 | Y | 4*, 5 |
| Pharynx | 6.6 | 27 | 49 | 5 | Y | 5* |
| Oral cavity | 4.2 | 98 | 265 | 5 | Y | 5* |
| Esophagus squamous | 3.9 | 99 | 193 | 9 | Y | 5 |
| Esophagus adenocarcinoma | 3.9 | 67 | 175 | 9 | Y | N.S. |
| Bladder | 3.8 | 111 | 288 | 5 | N | 5* |
| Liver | 2.9 | 157 | 235 | 19 | Y | 4*, 5, 16 |
| Stomach | 2.1 | 472 | | 13 | N | N/A |
| Acute myeloid leukemia | 2.0 | 202 | | 2 | N | N/A |
| Ovary | 1.9 | 458 | | 3 | N | N/A |
| Cervix | 1.8 | 94 | 74 | 8 | N | N.S. |
| Kidney | 1.7 | 154 | 103 | 6 | N | 5 |
| Pancreas | 1.6 | 119 | 120 | 11 | N | N.S. |
| Colorectal | 1.3 | 559 | | 4 | N | N/A |

mutations in *ERCC2*, which encodes a component of NER (*24*).

Signature 5 (or a similar signature that is difficult to differentiate from signature 5 because of the relatively flat profiles of these signatures) was increased by a factor of 1.3 to 5.1 ($q < 0.05$; table S2) in smokers versus nonsmokers in all cancer types together and in lung squamous, lung adenocarcinoma, larynx, pharynx, oral cavity, esophageal squamous, bladder, liver, and kidney cancers. The association of smoking with signature 5 mutations across these nine cancer types therefore includes some for which the risks conferred by smoking are modest and for which normal progenitor cells are not directly exposed to cigarette smoke (Table 1). Given the clocklike nature of signature 5 (*17*), its presence in the human germline (*23*), its ubiquity in cancer types unrelated to tobacco smoking (*18*), and its widespread occurrence in nonsmokers, it seems unlikely that signature 5 mutations associated with tobacco smoking are direct consequences of misreplication of DNA damaged by tobacco carcinogens. It is more plausible that smoking affects the machinery generating signature 5 mutations (*24*). Presumably as a consequence of the effects of smoking, signature 5 mutations correlated with age at the time of diagnosis in nonsmokers ($P = 0.001$) but not in smokers ($P = 0.59$).

Signature 16 is predominantly characterized by T>C mutations at ApT dinucleotides (Fig. 2B); exhibits a strong transcriptional strand bias consistent with almost all damage occurring on adenine (fig. S5); and, thus far, has been detected only in liver cancer. The underlying mutational process is currently unknown. Signature 16 exhibited a higher mutation burden in smokers versus nonsmokers with liver cancer (table S2).

For smokers with lung, larynx, pharynx, oral cavity, esophageal, bladder, liver, cervical, kidney, and pancreatic cancers, quantitative data on cumulative exposure to tobacco smoke were available (table S1). Total numbers of base substitution mutations were positively correlated with pack years smoked (1 pack year is defined as smoking one pack per day for 1 year) for all cancer types together ($q < 0.05$) and for lung adenocarcinoma (table S3). For individual mutational signatures, correlations with pack years smoked were found in multiple cancer types for signatures 4 and 5 (table S3). Signature 4 correlated with pack years in lung squamous, lung adenocarcinoma, larynx, and liver cancers. Signature 5 correlated with pack years in all cancers together, as well as in lung adenocarcinoma, pharynx, oral cavity, and

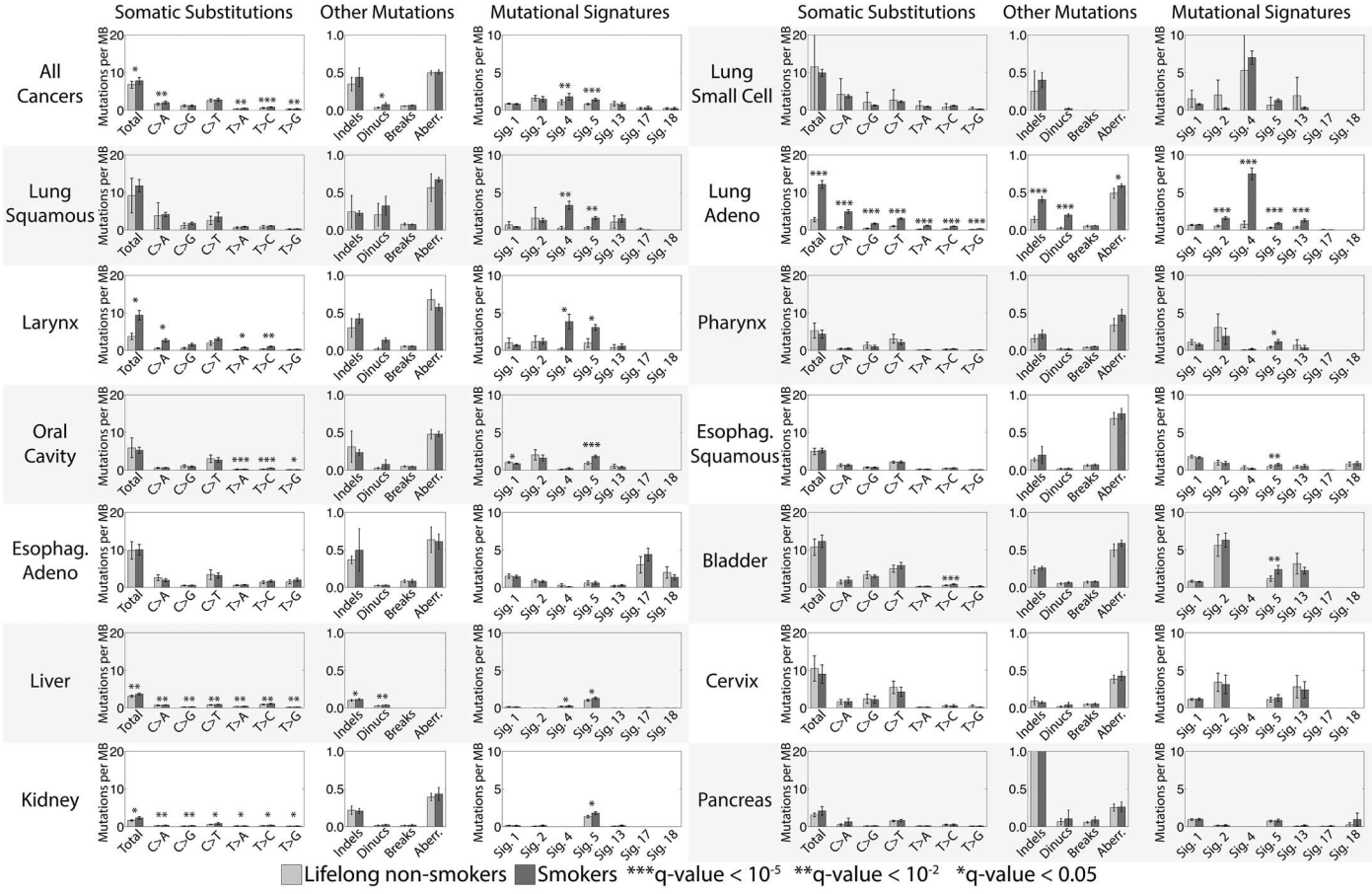



**Fig. 1. Comparison between tobacco smokers and lifelong nonsmokers.** Bars are used to display average values for numbers of somatic substitutions per megabase (MB), numbers of indels per megabase, numbers of dinucleotide mutations per megabase (Dinucs), numbers of breakpoints per megabase (Breaks), fraction of the genome that shows copy-number changes (Aberr.), and numbers of mutations per megabase attributed to mutational signatures found in multiple cancer types associated with tobacco smoking. Light gray bars represent nonsmokers, whereas dark gray bars are for smokers. Comparisons between smokers and nonsmokers for all features, including mutational signatures specific for a cancer type and overall DNA methylation, are provided in table S2. Error bars correspond to 95% confidence intervals for each feature. Each $q$ value is based on a two-sample Kolmogorov-Smirnov test corrected for multiple hypothesis testing for all features in a cancer type. Cancer types are ordered based on their age-adjusted odds ratios for smoking, as provided in Table 1. Data for numbers of breakpoints per megabase and fraction of the genome that shows copy-number changes were not available for liver cancer and small cell lung cancer. Adeno, adenocarcinoma; Esophag., esophagus. Note that the presented data include only a few cases (<10) of nonsmokers for small cell lung cancer, squamous cell lung cancer, and cancer of the larynx.

bladder cancers (table S3). In lung adenocarcinoma, correlations with pack years smoked were also observed for signatures 2 and 13. The rates of these correlations allow estimation of the approximate numbers of mutations accumulated in a normal cell of each tissue due to smoking a pack of cigarettes a day for a year: lung, 150 mutations; larynx, 97; pharynx, 39; oral cavity 23; bladder, 18; liver, 6 (table S3).

Consistent with our results, previous studies have reported higher numbers of total base substitutions in lung adenocarcinoma in smokers versus nonsmokers (mainly due to C>A substitutions) (*25*, *26*). The same is true of signatures 4 and 5 in lung adenocarcinoma (*18*), signature 4 in liver cancer (*27*), and signature 5 in bladder cancer (*24*).

Differential methylation of the DNA of normal cells of smokers compared to nonsmokers has been reported (*28*). Using data from methylation arrays, each containing ~470,000 of the ~28 million CpG sites in the human genome, we evaluated whether differences in methylation are found in cancers. Overall levels of CpG methylation in DNA from cancers were similar in smokers and nonsmokers for all cancer types (fig. S6). Individual CpGs were differentially methylated (>5% difference) in only two cancer types: 369 CpGs were hypomethylated and 65 were hypermethylated in lung adenocarcinoma, with five hypomethylated and three hypermethylated in oral cancer (Fig. 3 and fig. S7). CpGs exhibiting differences in methylation clustered in certain genes but were not associated with known cancer genes more than expected by chance, nor with genes hypomethylated in normal blood or buccal cells of tobacco smokers (fig. S8 and tables S4 and S5) (*28*). Therefore, with the exception of lung cancer, CpG methylation showed limited differences between the cancers of smokers and nonsmokers (Fig. 3).

The genomes of smoking-associated cancers permit reassessment of our understanding of how tobacco smoke causes cancer. Consistent with the proposition that an increased mutation load caused by tobacco smoke contributes to increased cancer risk, the total mutation burden is elevated in smokers versus nonsmokers with lung adenocarcinoma, larynx, liver, and kidney cancers. However, differences in total mutation burden were not observed in the other smoking-associated cancer types and, in some, there were no statistically significant smoking-associated differences in mutation load, signatures, or DNA methylation. Caution should be exercised in the interpretation of the latter observations. In addition to limitations of statistical power, multiple rounds of clonal expansion over many years are often required for development of a symptomatic cancer. It is thus conceivable that, in the normal tissues from which smoking-associated cancer types originate, there are more somatic mutations (or differences in methylation) in smokers than in nonsmokers but that these differences become obscured during the intervening clonal evolution. Moreover, some theoretical models predict that relatively small differences in mutation burden caused by smoking in preneoplastic cells could account for the observed increases in cancer

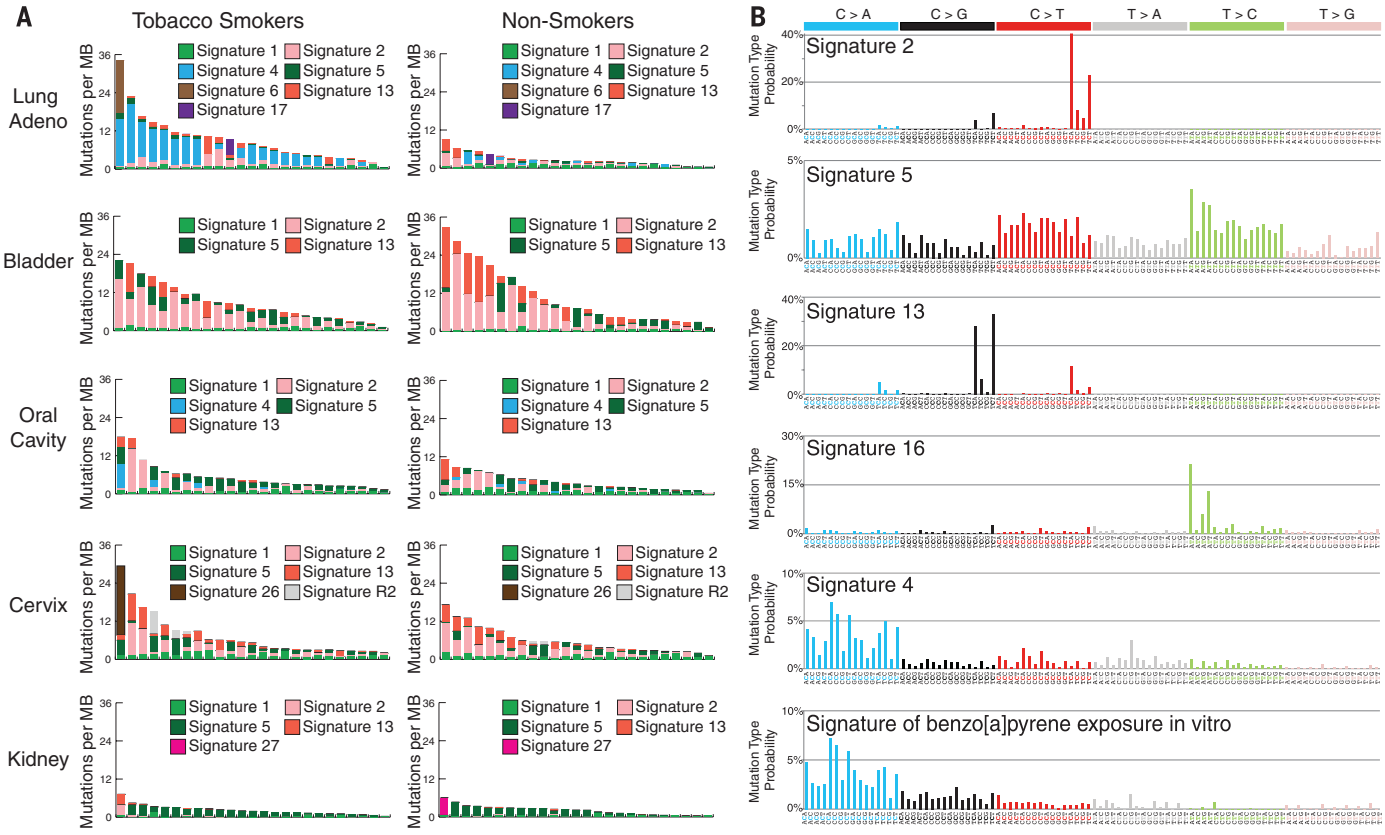



**Fig. 2. Mutational signatures associated with tobacco smoking.** (**A**) Each panel contains 25 randomly selected cancer genomes (represented by individual bars) from either smokers or nonsmokers in a given cancer type. The *y* axes reflect numbers of somatic mutations per megabase. Each bar is colored proportionately to the number of mutations per megabase attributed to the mutational signatures found in that sample. The naming of mutational signatures is consistent with previous reports (*16*–*18*). (**B**) Each panel contains the pattern of a mutational signature associated with tobacco smoking. Signatures are depicted using a 96-substitution classification defined by the substitution type and sequence context immediately 5′ and 3′ to the mutated base. Different colors are used to display the various types of substitutions. The percentages of mutations attributed to specific substitution types are on the *y* axes, whereas the *x* axes display different types of substitutions. Mutational signatures are depicted based on the trinucleotide frequency of the whole human genome. Signatures 2, 4, 5, 13, and 16 are extracted from cancers associated with tobacco smoking. The signature of benzo[*a*]pyrene is based on in vitro experimental data (*19*). Numerical values for these mutational signatures are provided in table S6.

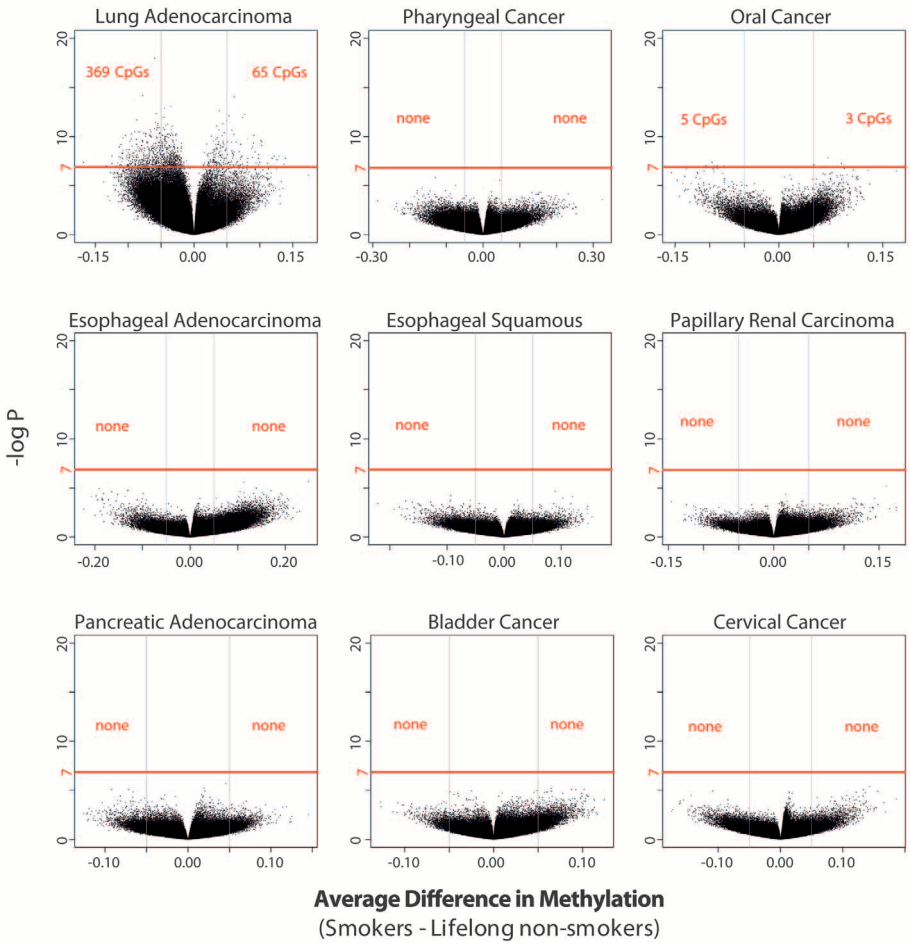



**Fig. 3. Differentially methylated individual CpGs in tobacco smokers across cancers associated with tobacco smoking.** Each dot represents an individual CpG. The *x* axes reflect differences in methylation between lifelong nonsmokers and smokers, where positive values correspond to hypermethylation and negative values to hypomethylation. The *y* axes depict levels of statistical significance. Results satisfying a Bonferroni threshold of $10^{-7}$ (above the red line) are considered statistically significant.

risks (*29*). Other models indicate that differences in mutation burden between smokers and nonsmokers need not be observed in the final cancers (supplementary text and fig. S6). Thus, increased somatic mutation loads in precancerous tissues may still explain the smoking-induced risks of most cancers, although other mechanisms have been proposed (*30*, *31*).

However, the generation of increased somatic mutation burden by tobacco smoking appears to be mechanistically complex. Smoking correlates with increases in base substitutions of multiple mutational signatures, together with increases in indels and copy-number changes. The extent to which these distinct mutational processes operate differs between tissue types (at least partially depending on the degree of direct exposure to tobacco smoke), and their mechanisms range from misreplication of DNA damage caused by tobacco smoke constituents to activation of more generally operative mutational processes. Although we cannot exclude roles for covariate behaviors of smokers or differences in the biology of cancers arising in smokers compared with nonsmokers, smoking itself is most plausibly the cause of these differences.

## REFERENCES AND NOTES


1. B. Secretan *et al.*, *Lancet Oncol.* **10**, 1033–1034 (2009).
2. S. S. Lim *et al.*, *Lancet* **380**, 2224–2260 (2012).
3. B. Pesch *et al.*, *Int. J. Cancer* **131**, 1210–1219 (2012).
4. A. Agudo *et al.*, *J. Clin. Oncol.* **30**, 4550–4557 (2012).
5. S. S. Hecht, *Nat. Rev. Cancer* **3**, 733–744 (2003).
6. D. H. Phillips, in *The Cancer Handbook*, M. R. Allison, Ed. (Macmillan, 2002), pp. 293–306.
7. D. H. Phillips, *Carcinogenesis* **23**, 1979–2004 (2002).
8. D. H. Phillips, S. Venitt, *Int. J. Cancer* **131**, 2733–2753 (2012).
9. L. B. Alexandrov, M. R. Stratton, *Curr. Opin. Genet. Dev.* **24**, 52–60 (2014).
10. P. Hainaut, M. Hollstein, *Adv. Cancer Res.* **77**, 81–137 (1999).
11. M. F. Denissenko, A. Pao, M. Tang, G. P. Pfeifer, *Science* **274**, 430–432 (1996).
12. G. P. Pfeifer, M. F. Denissenko, *Environ. Mol. Mutagen.* **31**, 197–205 (1998).
13. L. E. Smith *et al.*, *J. Natl. Cancer Inst.* **92**, 803–811 (2000).
14. F. Le Calvez *et al.*, *Cancer Res.* **65**, 5076–5083 (2005).
15. L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, P. J. Campbell, M. R. Stratton, *Cell Reports* **3**, 246–259 (2013).
16. L. B. Alexandrov, *Science* **350**, 1175–1177 (2015).
17. L. B. Alexandrov *et al.*, *Nat. Genet.* **47**, 1402–1407 (2015).
18. L. B. Alexandrov *et al.*, Australian Pancreatic Cancer Genome InitiativeICGC Breast Cancer ConsortiumICGC MMML-Seq ConsortiumICGC PedBrain, *Nature* **500**, 415–421 (2013).
19. S. Nik-Zainal *et al.*, *Mutagenesis* **30**, 763–770 (2015).
20. S. Nik-Zainal *et al.*, *Cell* **149**, 979–993 (2012).
21. S. A. Roberts *et al.*, *Mol. Cell* **46**, 424–435 (2012).
22. C. Swanton, N. McGranahan, G. J. Starrett, R. S. Harris, *Cancer Discov.* **5**, 704–712 (2015).
23. R. Rahbari *et al.*, *Nat. Genet.* **48**, 126–133 (2016).
24. J. Kim *et al.*, *Nat. Genet.* **48**, 600–606 (2016).
25. R. Govindan *et al.*, *Cell* **150**, 1121–1134 (2012).
26. M. Imielinski *et al.*, *Cell* **150**, 1107–1120 (2012).
27. A. Fujimoto *et al.*, *Nat. Genet.* **48**, 500–509 (2016).
28. A. E. Teschendorff *et al.*, *JAMA Oncol.* **1**, 476–485 (2015).
29. C. Tomasetti, L. Marchionni, M. A. Nowak, G. Parmigiani B. Vogelstein, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 118–123 (2015).
30. M. Sopori, *Nat. Rev. Immunol.* **2**, 372–377 (2002).
31. H. Rubin, *Oncogene* **21**, 7392–7411 (2002).


## ACKNOWLEDGMENTS

This work was supported by the Wellcome Trust (grant 098051). S.N.-Z. is a Wellcome-Beit Prize Fellow and is supported through a Wellcome Trust Intermediate Fellowship (grant WT100183MA). P.J.C. is personally funded through a Wellcome Trust Senior Clinical Research Fellowship (grant WT088340MA). M.R.S. is a paid advisor for GRAIL, a company developing technologies for sequencing of circulating tumor DNA for the purpose of early cancer detection. L.B.A. is personally supported through a J. Robert Oppenheimer Fellowship at Los Alamos National Laboratory. This research used resources provided by the Los Alamos National Laboratory Institutional Computing Program, which is supported by the U.S. Department of Energy (DOE) National Nuclear Security Administration under contract no. DE-AC52-06NA25396. Research performed at Los Alamos National Laboratory was carried out under the auspices of the National Nuclear Security Administration of the DOE. This work was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (grant FC001202), the UK MRC (grant FC001202), and the Wellcome Trust (grant FC001202). P.V.L. is a Winton Group Leader in recognition of the Winton Charitable Foundation's support toward the establishment of The Francis Crick Institute. D.H.P. is funded by Cancer Research UK (grant C313/A14329), the Wellcome Trust (grants 101126/Z/13/Z and 101126/B/13/Z), the National Institute for Health Research (NIHR) Health Protection Research Unit in Health Impact of Environmental Hazards at King's College London in partnership with PHE [the views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health, or PHE], and by the project EXPOSOMICS (grant agreement 308610-FP7) (European Commission). P.V. was partially supported by the project EXPOSOMICS (grant agreement 308610-FP7) (European Commission). Y.T. and T.S. are supported by the Practical Research for Innovative Cancer Control from Japan Agency for Medical Research and Development (grant 15ck0106094h0002) and National Cancer Center Research and Development Funds (26-A-5). We thank The Cancer Genome Atlas, the International Cancer Genome Consortium, and the authors of all studies cited in table S1 for providing free access to their somatic mutational data.

## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/618/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 to S6
References (*32*–*54*)

2 May 2016; accepted 23 September 2016
10.1126/science.aag0299

CHEMICAL BIOLOGY

# A chemical biology route to site-specific authentic protein modifications

Aerin Yang,[1] Sura Ha,[1] Jihye Ahn,[2] Rira Kim,[1] Sungyoon Kim,[1] Younghoon Lee,[1] Jaehoon Kim,[2] Dieter Söll,[3]* Hee-Yoon Lee,[1]* Hee-Sung Park[1]*

**Many essential biological processes are controlled by posttranslational protein modifications. The inability to synthetically attain the diversity enabled by these modifications limits functional studies of many proteins. We designed a three-step approach for installing authentic posttranslational modifications in recombinant proteins. We first use the established *O*-phosphoserine (Sep) orthogonal translation system to create a Sep-containing recombinant protein. The Sep residue is then dephosphorylated to dehydroalanine (Dha). Last, conjugate addition of alkyl iodides to Dha, promoted by zinc and copper, enables chemoselective carbon-carbon bond formation. To validate our approach, we produced histone H3, ubiquitin, and green fluorescent protein variants with site-specific modifications, including different methylations of H3K79. The methylated histones stimulate transcription through histone acetylation. This approach offers a powerful tool to engineer diverse designer proteins.**

Posttranslational modifications (PTMs) play vital roles in expanding protein functional diversity and critically affect numerous biological processes (*1*). The availability of proteins with specific modifications at selected residues is essential for experimental strategies to investigate fundamental biological mechanisms. Methods to generate diverse native protein covalent modifications do not exist at present. Genetic code expansion approaches are useful in producing recombinant proteins with specific modifications (*2*, *3*) but rely on the availability of an orthogonal transfer RNA (tRNA) synthetase•tRNA pair for acylation of a specific noncanonical amino acid. Despite much progress, the creation of many important protein modifications (e.g., trimethyllysine) is not yet feasible. Among chemical conjugation approaches (*4*), Cys-based strategies have been widely applied to generate protein conjugates (*5*) and mimics of PTMs (*6*, *7*). Yet the final products are PTM analogs whose value for finding unidentified properties of the natural system may be questionable (*8*). Thus, despite extensive efforts, synthetic approaches for many authentic PTMs are not available, because no carbon-carbon bond–forming reactions have thus far been successfully applied to protein modifications, despite the prevalence of such reactions in organic chemistry (*9*).

Here we propose a three-step strategy (Fig. 1A) that, in principle, is applicable to generating diverse forms of authentic and selective protein modifications. (i) The site of the intended modification is established by cotranslational *O*-phosphoserine (Sep) incorporation into a recombinant protein by using the Sep orthogonal *Escherichia coli* translation system (*10*, *11*). (ii) Then, the Sep residue of the purified recombinant protein is converted by phosphate removal to dehydroalanine (Dha), which in turn serves as a radicalophile, enabling a bioorthogonal chemical reaction. (iii) Last, PTM moieties are directly coupled to Dha through metal-mediated conjugate additions of alkyl iodides in aqueous solution, facilitating chemoselective carbon-carbon bond formation in proteins (Fig. 1B).

To demonstrate the feasibility and versatility of our strategy, we set out to generate recombinant *Xenopus laevis* histone H3K79 with five different modifications: monomethylation, dimethylation, trimethylation, formylation, and acetylation. Reversible lysine methylation in proteins presents the most complex and dynamic modification (*12*). Several approaches have attempted lysine methylation (*13*), but the trimethylated product has proved elusive. Histone H3K79 appears to be dynamically regulated (*14*) and associated with diverse cellular processes; its exact role has not been fully examined because such authentically modified histones have not been generated.

First, the Sep-containing histone H3Sep79 was made by expressing a *X. laevis* histone H3 mRNA containing a UAG codon at position 79 and a coding sequence for a C-terminal $His_6$ tag in *E. coli* containing the orthogonal Sep translation system [the engineered SepRS9•$tRNA^{Sep}$ pair and the evolved elongation factor (EF)–Sep21] (*11*). H3Sep79 was routinely obtained in good yield (~20 mg/liter of culture) (fig. S1). The incorporation of Sep at the intended position (79) was confirmed by matrix-assisted laser desorption/ionization–time-of-flight mass spectrometry (MALDI-TOF MS) analysis of the purified recombinant histone (fig. S2) and its tryptic peptides (Fig. 2A and tables S1 and S2).

Because phosphoamino acids are known to be labile under alkaline conditions (*15*), we attempted

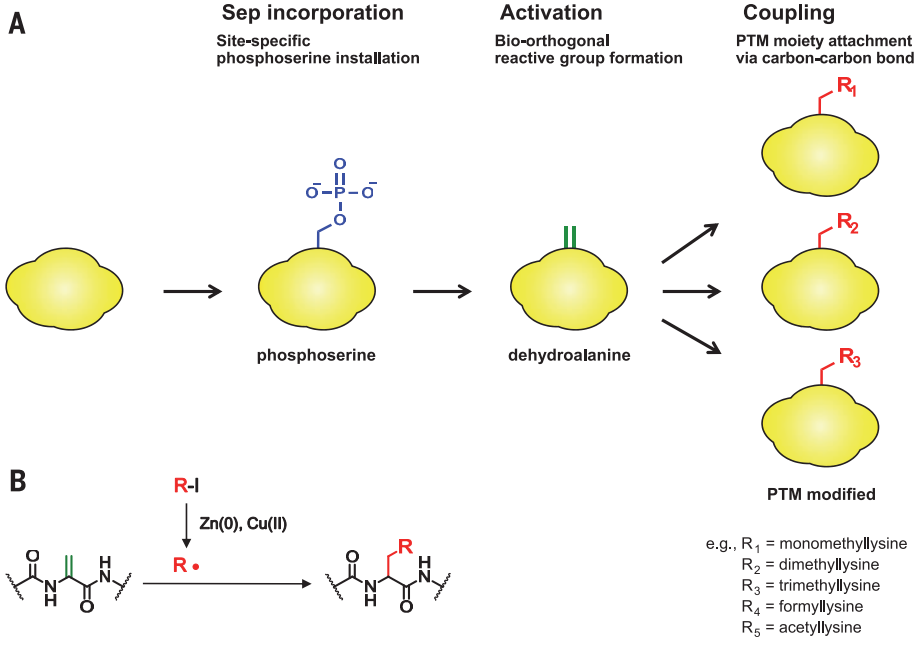



**Fig. 1. Approach for protein chemical modifications.** (**A**) Schematic representation of the three-step synthesis of proteins with authentic PTMs. (**B**) Our coupling scheme enables chemoselective carbon-carbon bond formation in proteins. R, PTM moiety.

[1]Department of Chemistry, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea. [2]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea. [3]Department of Molecular Biophysics and Biochemistry, Yale University, 266 Whitney Avenue, New Haven, CT 06511, USA.
*Corresponding author. Email: dieter.soll@yale.edu (D.S.); hylee@kaist.ac.kr (H.-Y.L.); hspark@kaist.ac.kr (H.-S.P.)

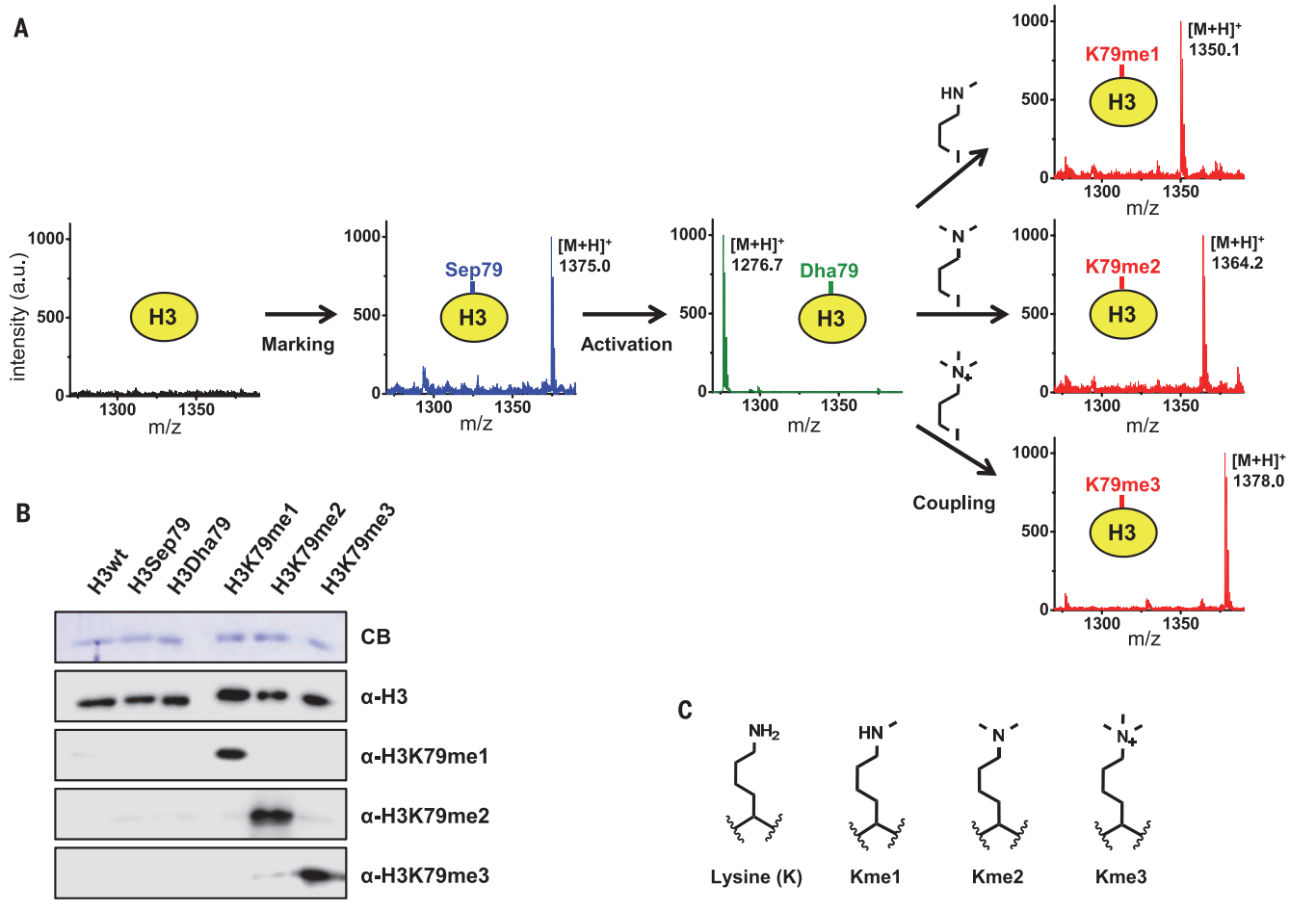



**Fig. 2. Synthesis of proteins with selective differential methylations.** (**A**) MALDI-TOF MS analysis of proteins after trypsin digestion. Formation of H3Sep79 is evidenced by the Sep-containing peptide (Sep79, blue). Synthesis of H3Dha79 generates a new Dha-containing peptide (Dha79, green) but eliminates the Sep-containing peptide. Coupling of the methyl iodides onto H3Dha79 produces new methylated lysine-containing peptides (K79me1, K79me2, and K79me3, red). *m/z*, mass/charge ratio; $[M+H]^+$, monoisotopic mass; a.u., arbitrary units. (**B**) Western blot analysis of the modified proteins, using antibodies (α) against H3K79me1, H3K79me2, and H3K79me3. CB, Coomassie blue. (**C**) Chemical structures of lysine and differentially methylated lysine residues.

to convert Sep to Dha by mild alkali treatment. To determine the optimal procedure, purified H3Sep79 was incubated with LiOH, $Ba(OH)_2$, or $Sr(OH)_2$ under different conditions, followed by neutralization with acetic acid and dialysis against distilled water (fig. S3). The reaction progress was analyzed by MALDI-TOF MS after in-gel trypsin digestion (tables S1 and S2). Near-complete disappearance of the Sep residue and concurrent generation of Dha took 30 min at room temperature in 40 mM $Ba(OH)_2$ solution. Mass analysis of whole proteins (fig. S4) and tryptic peptides (Fig. 2A) demonstrated that these conditions led to highly selective chemical transformation of H3Sep79 into H3Dha79 with no noticeable protein oxidation or other side reactions (*16*).

Dha-mediated conjugation approaches have been successfully applied to the generation of thiol-linked PTM analogs or mimics (*17*, *18*). To produce authentic PTMs from Dha, we needed a new coupling scheme enabling chemoselective carbon-carbon bond formation in proteins (Fig. 1B). We were encouraged by reports of water-based organic reactions in which an alkene group functions as a radical acceptor (*19*). Knowing that Dha could function as a radical acceptor, we chose alkyl radicals that can be generated from alkyl iodides by transition metals (*19*). Inspired by recent metal-mediated conjugate addition reactions (*20*), we reasoned that organozinc species would form from the precursor halides and Zn metal. Subsequently, zinc-to-copper transmetalation would occur, generating organocopper reagents, which would lead to the formation of alkyl radical species. Last, conjugate coupling of the alkyl radical with the Dha residue of a protein would take place (Fig. 1B). To demonstrate the proposed coupling scheme, we tested first the addition of trimethyl iodide (3-iodo-*N*,*N*,*N*,-trimethylpropan-1-amine, **3**; fig. S5) to H3Dha79 to generate the expected product H3K79me3. The reaction products were analyzed by Western blotting using an antibody against H3K79me3. Our initial attempts to directly use water-based organic reactions were unsuccessful. After testing a large array of conditions (described in the supplementary materials, figs. S6 to S16)—including different buffers and pH ranges, surfactants for protein stabilization, essential reagents for the metal-mediated coupling reaction (Zn metals and copper salts), and possible auxiliary reagents [e.g., tetramethylethylenediamine (TMEDA)]—we arrived at conditions that reproducibly led to the formation of carbon-carbon bonds in high yield. The deduced optimal conditions were H3Dha79 (10 μM), alkyl iodide (30 mM), Zn powder (0.4 mg), $Cu(OAc)_2$ (1 mM), Triton X-100 (2.0 weight %), and TMEDA (10 mM) in sodium acetate (pH 4.5, 0.5 M). Reactions (20 to 50 μl) were incubated at room temperature.

With optimized reaction conditions in hand, we attempted to synthesize H3K79me1, H3K79me2, and H3K79me3 by incubating H3Dha79 with three different methyl iodides: monomethyl iodide (3-iodo-*N*-methylpropan-1-amine, **1**), dimethyl iodide (3-iodo-*N*,*N*-dimethylpropan-1-amine, **2**), or trimethyl iodide (Fig. 2C and fig. S5). MALDI-TOF MS analysis revealed the disappearance of Dha and the concomitant generation of methylated lysine residues (Fig. 2A). The coupling reactions were highly selective and efficient (normally >80%), as demonstrated by mass analyses of tryptic peptides (Fig. 2A and table S2) and whole proteins

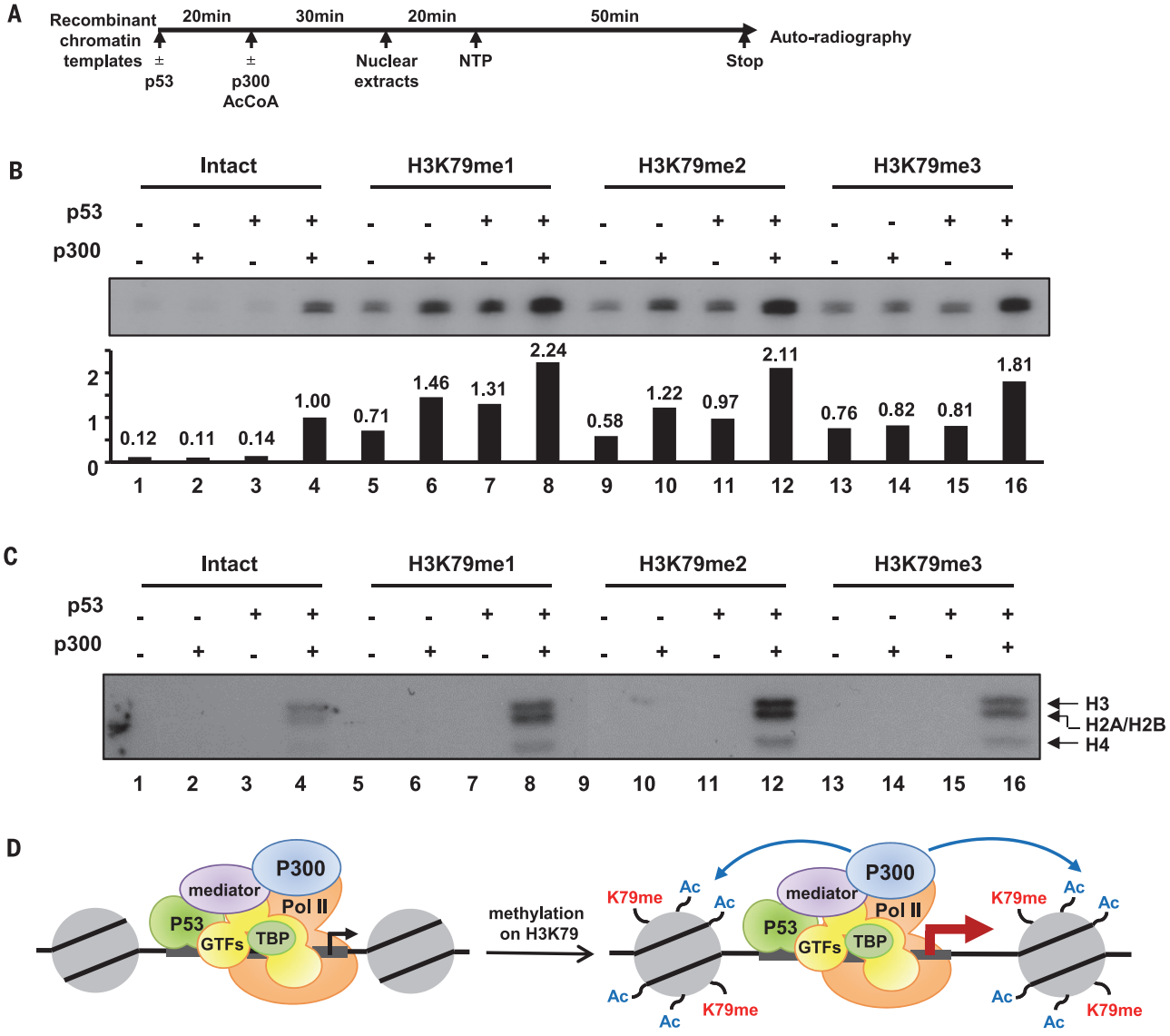



**Fig. 3. Effects of H3K79 methylation on chromatin transcription.** (**A**) Schematic of the standard in vitro transcription assay. AcCoA, acetyl coenzyme A; NTP, nucleoside triphosphate. (**B**) Effect of methylated H3K79 on chromatin transcription (*x* axis, lane number; *y* axis, relative transcription activity. (**C**) Effect of methylated H3K79 on p300-mediated chromatin acetylation. Histone acetylation status was monitored by fluorography. (**D**) Schematic representation of transcriptional activation by H3K79 methylation. Pol, polymerase; GTF, general transcription factor; TBP, TATA-binding protein.

(fig. S17), and had a good recovery yield (between 50 and 70%). Selective and differential lysine methylations were also demonstrated by Western blot analysis using antibodies against H3K79me1, H3K79me2, and H3K79me3 (Fig. 2B). To determine whether other PTM moieties could be coupled onto H3Dha79, we synthesized formyl iodide [*N*-(3-iodopropyl)formamide, **4**] and acetyl iodide [*N*-(3-iodopropyl)-acetamide, **5**; fig. S5] and used them for producing histones H3K79$N^{\varepsilon}$-formyl and H3K79$N^{\varepsilon}$-acetyl, respectively, as verified by MALDI-TOF MS analysis (fig. S18).

To demonstrate whether our three-step synthesis can be used for modification of other proteins, we set out to generate ubiquitin variants with site-specific modifications. Eight ubiquitin variants with various lysine modifications (acetylation and differential methylations) at position 33 or 48 were efficiently generated (fig. S19 and tables S3 to S5). Also, we found that different alkyl iodides (iodoethane, **6**; 2-iodopropane, **7**; 2-iodo-2-methylpropane, **8**; 1-iodobutane, **9**; iodocyclopentane, **10**; 3-iodopronionic acid, **11**; and *tert*-butyl-iodobutoxydimetylsilane, **12**) (fig. S5) were efficiently conjugated to Ub33Dha (fig. S20 and table S4). In particular, coupling of dansyl iodide (*5*-{dimethylamino}-*N*-{2-[(3-iodopropyl)amino]ethyl}naphthalene-1-sulfonamide, **13**; fig. S5) onto Ub33Dha led to site-specific fluorescent dye labeling via carbon-carbon bond formation (figs. S20 and S21 and table S4). Last, site-specific $Lys^{204}$ trimethylation of green fluorescence protein (GFP) by our three-step synthetic route led to the desired product GFPK204me3 but with lower conversion efficiency (~20%) and recovery yield (~30%) compared with histone H3 and ubiquitin (fig. S22 and tables S6 and S7). Thus, the structural context within the protein may affect the ease of modification. These data show that our approach is applicable to a variety of proteins. Because each protein behaves differently with the diverse chemical reagents, individualized optimization should lead to efficient implementation of our synthetic strategy with other proteins.

Next, we performed biochemical assays with the modified histones generated by our three-step approach to examine whether they are fully functional in the biological context. Genome-wide analyses of chromatin has revealed that H3K79 methylation is enriched in actively transcribing regions (*14*), but its detailed role is not yet well understood. A transcription assay using in vitro assembled chromatin templates and a highly purified RNA polymerase II transcription apparatus

provides a biochemically defined system to study the functions of individual histone modifications. The modified histones H3K79me1, H3K79me2, and H3K79me3 were first assembled with H2A, H2B, and H4 to form histone octamers. The methylated histones showed no noticeable difference from the unmodified (intact) histone during octamer assembly and nucleosome reconstitution (fig S23). Recombinant chromatin templates were reassembled with the histone octamers and a p53ML plasmid and were applied to p53 (activator)– and p300 (coactivator)–dependent in vitro transcription assays (Fig. 3A) (*21*). Intact chromatin lacking methylation showed a low level of transcription only in the presence of both the activator and coactivator, as expected (Fig. 3B). Notably, the levels of basal (activator- and/or coactivator-independent) transcription from chromatins with all three methylation states were greatly enhanced (Fig. 3B). Also, activator- or coactivator-dependent transcription from H3K79-methylated chromatin was elevated (Fig. 3B), demonstrating a direct stimulatory effect of H3K79 methylation on chromatin transcription. More importantly, we also found that H3K79 methylation increased histone acetylation mediated by p300 (Fig. 3C). Histone acetylation was differentially affected by the level of H3K79 methylation (Fig. 3C). Thus, H3K79 methylation is indeed positively associated with transcription activation through p300-mediated histone acetylation, which is differently affected by the various methylation states (Fig. 3D). These results illustrate that the methylated histones are fully active and, depending on the modification level, functionally distinct, demonstrating the utility and gentle nature of our three-step approach.

An open question concerns the diastereoselectivity of our coupling reaction. It is known that radical and thiol conjugate additions to Dha lead to epimeric mixtures (*22*–*24*); a definitive analysis of the epimeric ratio will require high-resolution crystallography of the protein. However, the steric context of the local protein conformation may substantively affect the final diastereomeric ratio of the products. The fact that our synthetic proteins H3K79me1, H3K79me2, and H3K79me3 were well recognized by antibodies, and that these modified histones could assemble into octamers with biological in vitro activity, underscores the utility of our approach.

We anticipate that with well-tailored alkyl iodides [e.g., (*25*)], our approach can be extended to produce designer proteins with diverse forms of chemical modifications (e.g., glycosylated amino acids and phosphotyrosine). Such efforts will markedly expand the available chemical diversity in proteins and facilitate the study of many PTM-mediated biological processes.

## REFERENCES AND NOTES


1. C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto Jr., *Angew. Chem. Int. Ed. Engl.* **44**, 7342–7372 (2005).
2. K. Lang, J. W. Chin, *Chem. Rev.* **114**, 4764–4806 (2014).
3. C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **79**, 413–444 (2010).
4. C. D. Spicer, B. G. Davis, *Nat. Commun.* **5**, 4740 (2014).
5. E. V. Vinogradova, C. Zhang, A. M. Spokoyny, B. L. Pentelute, S. L. Buchwald, *Nature* **526**, 687–691 (2015).
6. M. D. Simon *et al.*, *Cell* **128**, 1003–1012 (2007).
7. S. I. van Kasteren *et al.*, *Nature* **446**, 1105–1109 (2007).
8. D. P. Nguyen, M. M. Garcia Alai, S. Virdee, J. W. Chin, *Chem. Biol.* **17**, 1072–1076 (2010).
9. D. J. Hart, *Science* **223**, 883–887 (1984).
10. H. S. Park *et al.*, *Science* **333**, 1151–1154 (2011).
11. S. Lee *et al.*, *Angew. Chem. Int. Ed. Engl.* **52**, 5771–5775 (2013).
12. E. L. Greer, Y. Shi, *Nat. Rev. Genet.* **13**, 343–357 (2012).
13. W. R. Liu, Y. S. Wang, W. Wan, *Mol. Biosyst.* **7**, 38–47 (2011).
14. A. T. Nguyen, Y. Zhang, *Genes Dev.* **25**, 1345–1358 (2011).
15. Y. Oda, T. Nagasu, B. T. Chait, *Nat. Biotechnol.* **19**, 379–382 (2001).
16. J. Wang, S. M. Schiller, P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* **46**, 6849–6851 (2007).
17. J. Guo, J. Wang, J. S. Lee, P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* **47**, 6399–6401 (2008).
18. J. M. Chalker, L. Lercher, N. R. Rose, C. J. Schofield, B. G. Davis, *Angew. Chem. Int. Ed. Engl.* **51**, 1835–1839 (2012).
19. A. Postigo, N. S. Nudelman, *Coord. Chem. Rev.* **255**, 2991–3030 (2011).
20. B. H. Lipshutz, S. Huang, W. W. Leong, G. Zhong, N. A. Isley, *J. Am. Chem. Soc.* **134**, 19985–19988 (2012).
21. J. Kim, M. Guermah, R. G. Roeder, *Cell* **140**, 491–503 (2010).
22. D. Crich, J. W. Davies, *Tetrahedron* **45**, 5641–5654 (1989).
23. Y. Zhu, W. A. van der Donk, *Org. Lett.* **3**, 1189–1192 (2001).
24. M. M. Müller, T. W. Muir, *Chem. Rev.* **115**, 2296–2349 (2015).
25. J. M. Garcia Fernandez, C. O. Mellet, M. A. Pradera Adrian, J. F. Mota, *Carbohydr. Res.* **216**, 21–32 (1992).



## ACKNOWLEDGMENTS

We thank M. Simon and S. Miller for stimulating discussions. This work was supported by grants from the National Research Foundation of Korea (KAIST Systems Healthcare, 2011-0020322 and 2014M3A6A4075060 to H.-S.P. and 2012M3A9B4027956 and 2012M3A9C6049937 to J.K.) and from the U.S. National Institute for General Medical Sciences (GM22854-42 to D.S.). All data are available in the supplementary materials. H.-S.P., H.-Y.L., A.Y., and S.H. are inventors on patent application no. 62/383,245, submitted by the Korea Advanced Institute of Science and Technology, which covers protein chemical modifications mediated by carbon-carbon bond formation. H.-S.P., H.-Y.L., and A.Y. conceived and designed the study; H.-S.P. and A.Y. performed and analyzed all experiments, except chemical synthesis and in vitro transcription assays, which were conducted by S.H., R.K, Y.L., J.A., and J.K; H.-S.P., H.-Y.L., and D.S. supervised the project; and H.-S.P. and D.S. wrote the manuscript.


## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/623/suppl/DC1
Materials and Methods
Figs. S1 to S23
Tables S1 to S7
References (*26*–*28*)


27 June 2016; accepted 20 September 2016
Published online 29 September 2016
10.1126/science.aah4428


OCEAN CHEMISTRY

# Ocean mixing and ice-sheet control of seawater $^{234}U/^{238}U$ during the last deglaciation


**Tianyu Chen,[1]* Laura F. Robinson,[1] Matthew P. Beasley,[1] Louis M. Claxton,[1] Morten B. Andersen,[2,3] Lauren J. Gregoire,[4] Jemma Wadham,[5] Daniel J. Fornari,[6] Karen S. Harpp[7]**



**Seawater $^{234}U/^{238}U$ provides global-scale information about continental weathering and is vital for marine uranium-series geochronology. Existing evidence supports an increase in $^{234}U/^{238}U$ since the last glacial period, but the timing and amplitude of its variability has been poorly constrained. Here we report two seawater $^{234}U/^{238}U$ records based on well-preserved deep-sea corals from the low-latitude Atlantic and Pacific Oceans. The Atlantic $^{234}U/^{238}U$ started to increase before major sea-level rise and overshot the modern value by 3 per mil during the early deglaciation. Deglacial $^{234}U/^{238}U$ in the Pacific converged with that in the Atlantic after the abrupt resumption of Atlantic meridional overturning. We suggest that ocean mixing and early deglacial release of excess $^{234}U$ from enhanced subglacial melting of the Northern Hemisphere ice sheets have driven the observed $^{234}U/^{238}U$ evolution.**


The last deglaciation [18.0 to 10.5 thousand years ago (ka)] saw massive changes in Earth's surficial environments, including in temperature and precipitation, as well as the retreat of the Northern Hemisphere (NH) ice sheets and sea-level rise (*1*, *2*). These processes have the potential to induce large variability in the weathering of the continents and the composition of chemical fluxes to the ocean. The ratio $^{234}U/^{238}U$ is one of the isotopic signatures with the potential to have recorded global changes in continental weathering during this critical climate transition (*3*).

The activity ratio of $^{234}U$ to $^{238}U$ in the modern seawater is ~15% higher than secular equilibrium (*4*) because of the relatively mobile nature of $^{234}U$ induced by α-recoil effects (*5*) in the weathered host rocks. $^{234}U$ is enriched relative to $^{238}U$ at


[1]Bristol Isotope Group, School of Earth Sciences, University of Bristol, Bristol, UK. [2]Institute of Geochemistry and Petrology, Department of Earth Sciences, ETH Zürich, Zürich, Switzerland. [3]School of Earth and Ocean Sciences, Cardiff University, Cardiff, UK. [4]School of Earth and Environment, University of Leeds, Leeds, UK. [5]Bristol Glaciology Centre, School of Geographical Sciences, University of Bristol, Bristol, UK. [6]Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, MA 02543, USA. [7]Department of Geology, Colgate University, 13 Oak Drive, Hamilton, NY 13346, USA.
***Corresponding author. Email: tc14502@bristol.ac.uk**

the deep-sea corals in this study) would act to reduce the effective U residence time and allow its $\delta^{234}U$ to change more rapidly than in the Pacific. We applied a two-box model, consisting of an upper Atlantic-Arctic box and a "rest of the ocean" box, to study the influence of changing external sources and changing ocean circulation on seawater $\delta^{234}U$ (*19*). Modern high-latitude riverine inputs have considerably higher $\delta^{234}U$ than middle- to low-latitude inputs and mainly supply the Arctic and polar North Atlantic (table S1) (*19*). If the $\delta^{234}U$ and U fluxes of all external sources are kept constant throughout the past 25 ky, our model shows that a slowdown of ocean circulation during HS1 can result in a resolvable difference in $\delta^{234}U$ between the upper Atlantic and the rest of the ocean, depending on the degree of reduction in exchange flux (Fig. 3A). This result is consistent with the difference in $\delta^{234}U$ between the Pacific and Atlantic records (Fig. 1) and contrasts with the general assumption that U isotopes are homogenous throughout the global ocean [i.e., differences no larger than 0.4‰ (*4*)]. Our result also raises the possibility that other isotope systems with relatively long residence times might exhibit differences between different ocean basins during periods of reduced ocean mixing.

Ocean circulation, however, cannot account for the overall glacial-interglacial increase of ~3‰ or more in the $\delta^{234}U$ of both ocean basins (Fig. 3A). External $^{234}U$ inputs must have increased relative to $^{238}U$ during the early deglaciation. With three times the modern riverine U flux, a 3‰ shift in oceanic $\delta^{234}U$ is possible during HS1 (fig. S4). However, given that the hydrological cycle was probably weaker during the late last glacial period and early deglaciation than at present (*2*), increased dissolved U flux from global rivers alone as a driver for the $\delta^{234}U$ change is unlikely. An increase in $\delta^{234}U$ in the continental inputs is therefore likely to have been important in seawater $\delta^{234}U$ evolution. There are no direct measurements of the $\delta^{234}U$ of past watersheds, so we compiled $\delta^{234}U$ data from speleothems that grew during the relevant time period. Although speleothem $\delta^{234}U$ does not necessarily reflect the primary surface-weathering signal, because it includes influences such as the percolation of meteoric waters from the surface down into the cave, it may still provide a first-order indication of the variability in the hydrological cycle that controlled the overall $^{234}U$ budget available for weathering (*26*). The data come from the middle to low latitudes (fig. S5), and they do not exhibit any distinctive $\delta^{234}U$ shifts from 30 to 15 ka, suggesting that weathering variability in these areas was not large enough to account for the oceanic $\delta^{234}U$ observations.

Instead, high-latitude processes could have played a key role in driving the increase in global seawater $\delta^{234}U$ from the last glacial period to 15 ka. In the early part of the last deglaciation, the bases of North American ice sheets are thought to have become increasingly wet (*10*). We hypothesize that $^{234}U$-enriched water was released from subglacial melt reservoirs with a prolonged residence time and from leaching of the previously frozen

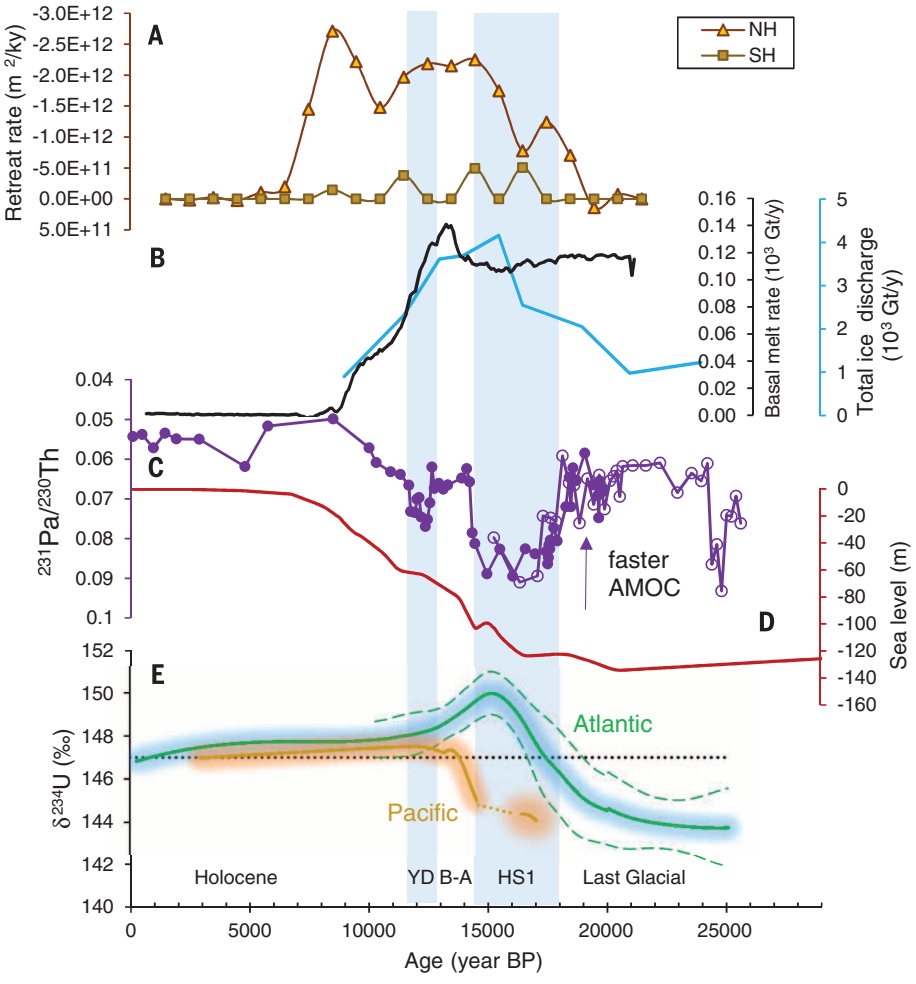



**Fig. 2. Seawater $\delta^{234}U$ evolution compared with other climate records.** (**A**) Modeled retreat rate of the Northern Hemisphere (NH) and Southern Hemisphere (SH) ice sheets (*33*). (**B**) Modeled basal melting rate of the North American ice sheets (*28*) and ice discharge of the Laurentide Ice Sheet (*29*). (**C**) $^{231}Pa/^{230}Th$ from North Atlantic deep sediment cores (*21*, *23*). (**D**) Sea-level history (*34*). (**E**) Our reconstructed $\delta^{234}U$ evolution in the upper Atlantic and Pacific. Green solid line, Atlantic mean; green dashed line, Atlantic 2σ error; gold line, Pacific mean (dotting, HS1 data gap); gold shading, estimated range of Pacific $\delta^{234}U$. black dotted line, modern seawater $\delta^{234}U$; Gt, gigaton; y, year.

subglacial sediments by basal meltwater over this period. In both cases, subglacial meltwaters are likely to be enriched in recoiled $^{234}U$ (*7*, *11*). In addition, sediments (and their former porewaters) frozen within the bases of icebergs might also have contributed to releasing U and nutrients to the ocean (*27*) during this period. The reconstructed number of ice streams based on field evidence (*8*) and the modeled subglacial melt rate (*10*, *28*) (Fig. 2B) of the North American ice sheets were much higher during the early deglaciation (before 15 ka) than later. It is notable that the modeled ice discharge (*29*) of the Laurentide Ice Sheet was also high during the early deglaciation, reaching a peak at late HS1 (Fig. 2). This timing is consistent with the enhanced release and transport of excess $^{234}U$ to the ocean during the early deglaciation. In this case, the peak oceanic $\delta^{234}U$ indicates that basal water from the whole ice-sheet interior may have been active in exporting to the margin during HS1. In another modeling experiment (*19*), the sensitivity of oceanic responses to inputs with different $\delta^{234}U$ ratios was tested. With an average $\delta^{234}U$ of ~800‰ for the input to the upper Atlantic during HS1, the model is able to reproduce the observed amplitude of the increase in $\delta^{234}U$ from the last glacial period to 15 ka (Fig. 3B). A more realistic case might be an increase in both the U flux and the $\delta^{234}U$ of the high-latitude continental inputs, but deconvolving these two factors is difficult. There is a hint from the compiled data that the surface ocean had lower $\delta^{234}U$ than the intermediate-depth ocean (fig. S3). Our extended modeling experiments (table S2) are unable to replicate this feature, even with all high-latitude excess $^{234}U$ routed to the intermediate ocean via the polar surface Atlantic and no deep-ocean overturning and mixing (*19*). This indicates that other mechanisms, such as local influences or diagenetic processes, are responsible for those differences.

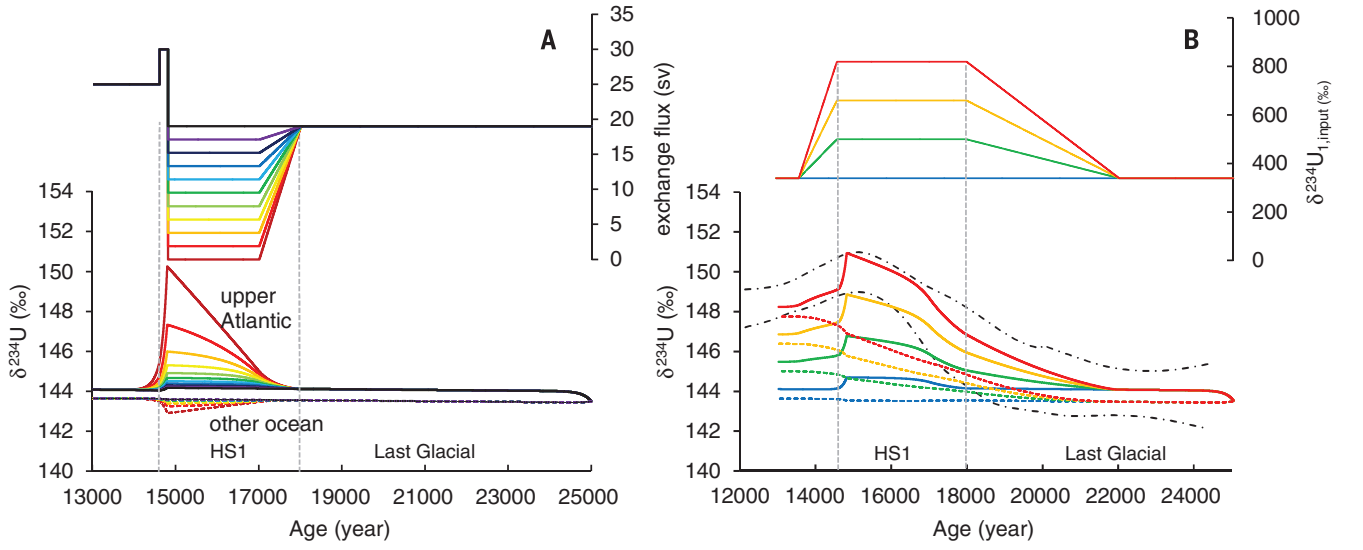



**Fig. 3. Sensitivity experiment assessing the response of seawater $\delta^{234}U$ to ocean circulation and external inputs.** (**A**) Effect of ocean mixing slowdown alone (sv, sverdrup). (**B**) Effect of variable $\delta^{234}U$ in inputs to the upper Atlantic, applying a 50% reduction of the glacial exchange flux during HS1. The area between the dashed-dotted curves is the range of low-latitude Atlantic $\delta^{234}U$. External U fluxes through time in both experiments are kept at the modern inputs. The spectrum of colors denotes different experiments (solid lines, upper Atlantic; dashed lines, other ocean). Further details are in (*19*).

The transition from HS1 to the B-A is marked by a distinct decrease in $\delta^{234}U$ in the low-latitude Atlantic by ~3‰ from about 150‰ to the modern seawater signature. The Pacific $\delta^{234}U$ appears to have risen at the same time, converging with the Atlantic $\delta^{234}U$ during the mid–B-A. We suggest that this convergence is due to the abrupt increase in overturning of the Atlantic at the end of HS1 (*21*, *23*) (Figs. 2C and 3A), although the depletion of the subglacial excess $^{234}U$ pool might also have played a role (*19*). Stabilized Holocene seawater $\delta^{234}U$ afterward implies that the U cycle in the modern ocean is likely close to a steady state (*30*). Nevertheless, perturbation of oceanic U isotopes by polar processes might have continued to occur even during the climatically and oceanographically more stable late Holocene. For example, widespread collapse of the Ross Ice Shelf and the export of old materials from inland Antarctica is inferred to have occurred at ~5 to 1.5 ka in response to regional warming (*31*). These processes may have been accompanied by a $^{234}U$-rich flux to the Southern Ocean, although any shift in the whole-ocean $\delta^{234}U$ was probably limited. On longer orbital time scales, seawater $\delta^{234}U$ is thought to be higher during past interglacial periods and lower during glacial periods [e.g., (*3*, *16*, *17*)], with the decreases potentially associated with low sea level (*16*). By comparison with the deglacial mechanism, our study implies that progressive NH glaciation could have reduced the weathering input from high-latitude continental areas, leading to lower glacial oceanic $\delta^{234}U$.

The retreat of the NH ice sheets started about 20 ka and continued through HS1 (Fig. 2A) (*1*), considerably earlier than the enhanced surface melting that dominated ice-sheet mass loss and sea-level rise in the late deglaciation and early Holocene (*29*). Our data provide evidence for enhanced subglacial melting of the NH ice-sheet interior during the early deglaciation, supporting the notion that basal melting and sliding represents one of the feedbacks involved in enhancing early deglaciation as a result of the buildup of very large NH ice sheets (*10*). An interesting consequence of the basal melt inputs may be the associated release of nutrients to the ocean. Recent work from the Greenland Ice Sheet indicates that dissolved phosphorus yields are at least equal to those from the Mississippi or Amazon rivers (*32*). In this regard, nutrients from direct subglacial weathering should be considered in future research as a potential source fueling productivity in the North Atlantic during HS1.

## REFERENCES AND NOTES


1. G. H. Denton *et al.*, *Science* **328**, 1652–1656 (2010).
2. P. U. Clark *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E1134–E1142 (2012).
3. L. F. Robinson, G. M. Henderson, L. Hall, I. Matthews, *Science* **305**, 851–854 (2004).
4. M. B. Andersen, C. H. Stirling, B. Zimmermann, A. N. Halliday, *Geochem. Geophys. Geosyst.* **11**, Q12003 (2010).
5. K. Kigoshi, *Science* **173**, 47–48 (1971).
6. F. Chabaux, J. Riotte, O. Dequincey, *Rev. Mineral. Geochem.* **52**, 533–576 (2003).
7. M. B. Andersen, Y. Erel, B. Bourdon, *Geochim. Cosmochim. Acta* **73**, 4124–4141 (2009).
8. C. R. Stokes, M. Margold, C. D. Clark, L. Tarasov, *Nature* **530**, 322–326 (2016).
9. A. M. Le Brocq *et al.*, *Nat. Geosci.* **6**, 945–948 (2013).
10. S. J. Marshall, P. U. Clark, *Geophys. Res. Lett.* **29**, 67-1–67-4 (2002).
11. G. M. Henderson, B. L. Hall, A. Smith, L. F. Robinson, *Chem. Geol.* **226**, 298–308 (2006).
12. J. A. Mikucki *et al.*, *Nat. Commun.* **6**, 6831 (2015).
13. G. M. Henderson, *Earth Planet. Sci. Lett.* **199**, 97–110 (2002).
14. R. M. Dunk, R. A. Mills, W. J. Jenkins, *Chem. Geol.* **190**, 45–67 (2002).
15. K. B. Cutler *et al.*, *Radiocarbon* **46**, 1127–1160 (2004).
16. T. M. Esat, Y. Yokoyama, *Geochim. Cosmochim. Acta* **70**, 4140–4150 (2006).
17. W. G. Thompson, H. A. Curran, M. A. Wilson, B. White, *Nat. Geosci.* **4**, 684–687 (2011).
18. T. M. Esat, Y. Yokoyama, *Geochim. Cosmochim. Acta* **74**, 7008–7020 (2010).
19. Materials and methods are available as supplementary materials on *Science* Online.
20. S. L. Jaccard, E. D. Galbraith, *Nat. Geosci.* **5**, 151–156 (2012).
21. J. F. McManus, R. Francois, J. M. Gherardi, L. D. Keigwin, S. Brown-Leger, *Nature* **428**, 834–837 (2004).
22. A. Ganopolski, S. Rahmstorf, *Nature* **409**, 153–158 (2001).
23. E. Böhm *et al.*, *Nature* **517**, 73–76 (2015).
24. J. Lynch-Stieglitz *et al.*, *Nat. Geosci.* **7**, 144–150 (2014).
25. M. H. Simon *et al.*, *Paleoceanography* **30**, 1318–1327 (2015).
26. J. Z. Zhou *et al.*, *Chem. Geol.* **221**, 1–20 (2005).
27. R. Raiswell, L. G. Benning, M. Tranter, S. Tulaczyk, *Geochem. Trans.* **9**, 7 (2008).
28. L. J. Gregoire, A. J. Payne, P. J. Valdes, *Nature* **487**, 219–222 (2012).
29. D. J. Ullman, A. E. Carlson, F. S. Anslow, A. N. LeGrande, J. M. Licciardi, *Nat. Geosci.* **8**, 534–537 (2015).
30. F. L. H. Tissot, N. Dauphas, *Geochim. Cosmochim. Acta* **167**, 113–143 (2015).
31. Y. Yokoyama *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 2354–2359 (2016).
32. J. Hawkings *et al.*, *Global Biogeochem. Cycles* **30**, 191–210 (2016).
33. W. R. Peltier, D. F. Argus, R. Drummond, *J. Geophys. Res. Solid Earth* **120**, 450–487 (2015).
34. K. Lambeck, H. Rouby, A. Purcell, Y. Sun, M. Sambridge, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15296–15303 (2014).


## ACKNOWLEDGMENTS


Data related to this work are available in the supplementary materials. Insightful comments from two anonymous reviewers helped us improve the manuscript. We thank C. Taylor and C. Coath for help in the laboratory. This study was funded by the European Research Council, the Natural Environment Research Council, the Philip Leverhulme Trust, the U.S. National Science Foundation, a Marie Curie Reintegration Grant, and the NOAA (National Oceanic and Atmospheric Administration) Ocean Exploration Trust. We also thank the Charles Darwin Foundation, Galápagos National Park, and INOCAR (Instituto Oceanográfico de la Armada) for making the Galápagos sampling possible.


## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/626/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Tables S1 to S8
References (*35*–*90*)


9 May 2016; accepted 22 September 2016
Published online 6 October 2016
10.1126/science.aag1015

NEUROREGENERATION

# Injury-induced *ctgfa* directs glial bridging and spinal cord regeneration in zebrafish

**Mayssa H. Mokalled,[1] Chinmoy Patra,[2]* Amy L. Dickson,[1] Toyokazu Endo,[1] Didier Y. R. Stainier,[2] Kenneth D. Poss[1]†**

**Unlike mammals, zebrafish efficiently regenerate functional nervous system tissue after major spinal cord injury. Whereas glial scarring presents a roadblock for mammalian spinal cord repair, glial cells in zebrafish form a bridge across severed spinal cord tissue and facilitate regeneration. We performed a genome-wide profiling screen for secreted factors that are up-regulated during zebrafish spinal cord regeneration. We found that *connective tissue growth factor a* (*ctgfa*) is induced in and around glial cells that participate in initial bridging events. Mutations in *ctgfa* disrupted spinal cord repair, and transgenic *ctgfa* overexpression or local delivery of human CTGF recombinant protein accelerated bridging and functional regeneration. Our study reveals that CTGF is necessary and sufficient to stimulate glial bridging and natural spinal cord regeneration.**

The spinal cord of adult zebrafish recovers spontaneously after injury (Fig. 1A and fig. S1, A and B). Efficient axon growth, adult neurogenesis, and absence of glial scarring distinguish this injury response from that in mammals (*1*, *2*). After initial inflammation, ependymal cells proliferate and glia form a bridge that is thought to provide a scaffold for axonal growth (*3*). The severed cord reconnects, and new neuronal connections lead to functional recovery (Fig. 1A) (*4*). Here, we analyzed extracellular factors up-regulated in the regenerating zebrafish spinal cord.

To identify potential secreted, pro-regenerative signaling molecules, we screened zebrafish transcriptomes for genes induced after spinal cord injury (Fig. 1B and table S1). Our screen identified seven genes encoding secreted extracellular proteins, including *fibronectin 1 a* (*fn1a*), previously implicated in axon growth promotion (*5*, *6*). Transcripts for *fn1a*, *fn1b*, *connective tissue growth factor a* (*ctgfa*), *myostatin b* (*mstnb*), and *stanniocalcin 1 like* (*stc1l*) increased in ependymal cells at 1 and 2 weeks post-injury (wpi) (Fig. 1C and fig. S1C). Whereas *fn1a* was expressed around the entirety of the central canal at 2 wpi, *mstnb* and *ctgfa* were predominantly expressed in the dorsal and ventral ependyma, respectively (Fig. 1C).

CTGF is a matricellular, multifunctional protein that can influence the activity of multiple major signaling pathways, affecting cell adhesion, migration, proliferation, and differentiation. CTGF expression increases after central nervous system (CNS) trauma in rodents, but its function after spinal cord injury has not been elucidated (*7*–*9*). Because *ctgfa* is induced upon spinal cord injury in zebrafish, we hypothesized that it may have pro-regenerative roles. We analyzed *ctgfa* expression along the rostrocaudal spinal cord axis (Fig. 1D). At 1 wpi, *ctgfa* transcription was induced in multiple cell types across the lesion site and in ependymal cells at the central canal near the lesion. At this time point, we detected strongest *ctgfa* expression in the ventral ependyma. By 2 wpi, *ctgfa* expression localized to ventral ependymal cells and marked the cellular bridge that had formed at the lesion site. Expression declined beyond 3 wpi (fig. S2A). Thus, *ctgfa* expression correlates with formation of the glial bridge.

To identify the cell populations that express *ctgfa* during spinal cord repair, we generated transgenic reporter zebrafish with a 5.5-kb genomic sequence upstream of the *ctgfa* translational start site fused to an enhanced green fluorescent protein (EGFP) reporter cassette. *ctgfa*:EGFP fluorescence resembled endogenous *ctgfa* mRNA expression in spinal cord tissue at 2 wpi (fig. S2, B and C). As early as 5 days post-injury (dpi), domains of *ctgfa*:EGFP and glial fibrillary acidic protein (GFAP), a marker of multiple glial cells in the CNS, overlapped within a subpopulation of glial cells at the injury site (Fig. 1E). We interpret these to be bridging cells. Similarly, *ctgfa*:EGFP colocalized with the GFAP$^+$ bridge at 2 wpi (Fig. 1E). Comparison of *ctgfa*:EGFP and GFAP expression on serial cross sections revealed that ~97% of GFAP$^+$ bridging glia at the lesion core were also EGFP$^+$ at 1 wpi (fig. S3). Further away from the lesion, *ctgfa*:EGFP was mainly present in ventral ependymal cells (fig. S3). We also detected *ctgfa*:EGFP in skeletal muscle, bone cells, and reactive fibroblast-like cells around the lesion site. *ctgfa*-driven expression suggested delineation of "pioneer" bridging glia.

To determine whether *ctgfa* is required for spinal cord regeneration, we generated a *ctgfa* mutant allele (*ctgfa$^{bns50}$*; referred to as *ctgfa$^-$*) that harbors a frameshift-causing 7-nucleotide deletion within the third exon of the *ctgfa* locus (fig. S4, A to C). *ctgfa$^{-/-}$* animals are adult viable and appear to have unaffected motor function capacity (*10*) (fig. S4D). However, *ctgfa$^{-/-}$* animals showed diminished swim capacity after spinal cord injury, with no significant functional recovery between 2 and 6 wpi (Fig. 2A). Heterozygous (*ctgfa$^{+/-}$*) animals showed partial recovery of swim capacity by 6 wpi (Fig. 2A). At 4 wpi, anterograde axon tracing indicated that axon regeneration across the lesion site was reduced by ~25% in *ctgfa$^{+/-}$* and ~60% in *ctgfa$^{-/-}$* spinal cords proximal to the lesion, and by ~40% in *ctgfa$^{+/-}$* and ~80% in *ctgfa$^{-/-}$* cords distally (Fig. 2B). Thus, *ctgfa* is required for spinal cord regeneration.

Deficits in glial bridging might underlie the defects in spinal cord regeneration displayed by *ctgfa* mutants. Using GFAP and acetylated α-tubulin immunohistochemistry, we observed robust glial bridges in wild-type and *ctgfa$^{+/-}$* animals at 4 wpi (Fig. 2C and fig. S5A). However, *ctgfa$^{-/-}$* animals displayed ~71% less bridging than did wild-type clutchmates (Fig. 2C and fig. S5A). At 2 wpi, glial cells within *ctgfa$^{-/-}$* injury sites often failed to extend projections into the lesion (fig. S5B). 5-ethynyl-2′-deoxyuridine (EdU) incorporation assays showed that *ctgfa$^{-/-}$* zebrafish displayed a ~48% reduction in glial cell proliferation at 1 and 2 wpi (Fig. 2D and fig. S6). EdU$^+$GFAP$^-$ cell numbers were comparable in wild-type, *ctgfa$^{+/-}$*, and *ctgfa$^{-/-}$* tissues at 1 wpi (Fig. 2D), which suggests that the effects of *ctgfa* mutations were preferential to glial cells. Thus, *ctgfa* is required for the changes in proliferation and morphology of glial cells during spinal cord regeneration.

To examine the effects of excess *ctgfa* on spinal cord regeneration, we generated and injured transgenic fish that express full-length *ctgfa* under control of a heat-inducible promoter (*hsp70*:*ctgfa-FL*) (fig. S7). Recovery of swim capacity was improved in *ctgfa*-overexpressing animals given daily heat shocks and assessed at 2, 4, and 6 wpi (Fig. 3A). Swim capacity was comparable between sham-injured *hsp70*:*ctgfa-FL* and wild-type clutchmates at 2 and 6 weeks after heat shock (Fig. 3A), indicating contextual effects of *ctgfa* overexpression. Histology indicated increased bridging and axon regeneration in *ctgfa*-overexpressing fish relative to controls at 2 and 4 wpi, respectively (Fig. 3, B and C, and fig. S8A). Thus, whole-animal *ctgfa* overexpression promotes regeneration after spinal cord injury.

CTGF harbors four protein interaction domains and a protease domain that self-cleaves CTGF into profibrotic N-terminal and proliferative C-terminal peptides (*11*, *12*). To determine the active portion of zebrafish Ctgfa during spinal cord regeneration, we created transgenic fish with heat-inducible expression of either N-terminal or C-terminal *ctgfa* fragments (*hsp70*:*ctgfa-NT* and *hsp70*:*ctgfa-CT*) (fig. S7A). Only *ctgfa-CT* overexpression recapitulated the pro-regenerative

[1]Department of Cell Biology, Duke University Medical Center, Durham, NC 27710, USA. [2]Department of Developmental Genetics, Max Planck Institute for Heart and Lung Research, 61231 Bad Nauheim, Germany.
*Present address: Agharkar Research Institute, Pune, India.
†Corresponding author. Email: kenneth.poss@duke.edu

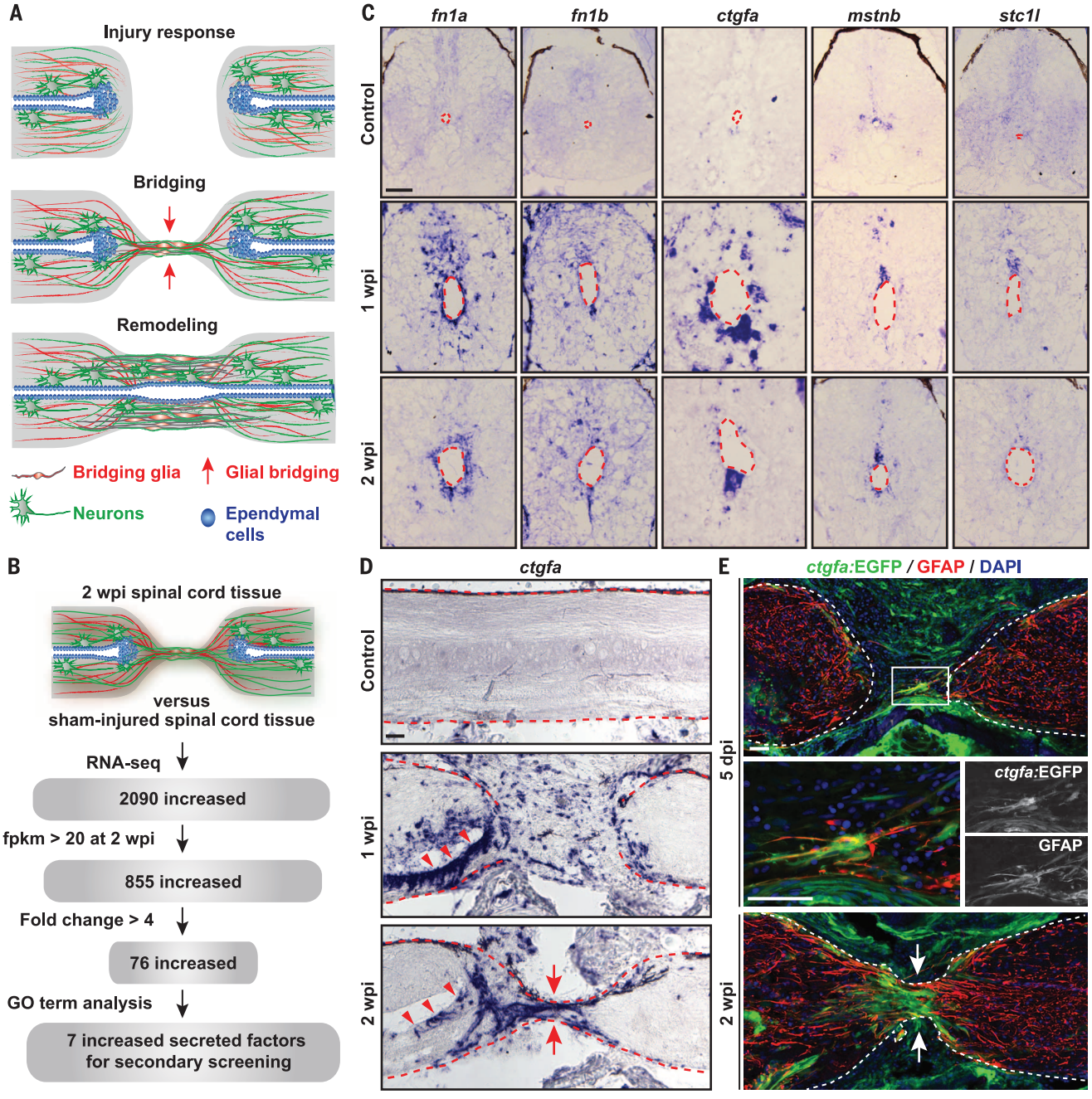



**Fig. 1. Identification of *ctgfa* from a screen for regulators of spinal cord regeneration.** (**A**) Schematic of the multistep process of spinal cord regeneration in zebrafish. (**B**) A screen for secreted factors expressed during spinal cord regeneration (fpkm, fragments per kilobase of transcript per million; GO, Gene Ontology). (**C**) In situ hybridization on spinal cord cross sections at 1 and 2 wpi and in uninjured control tissue. Sections proximal to the lesion from the rostral side are shown; dashed lines delineate the central canals. The canal dilates after injury. (**D**) *ctgfa* in situ hybridization on longitudinal spinal cord sections at 1 and 2 wpi and in uninjured control tissue. (**E**) *ctgfa*:EGFP reporter expression and GFAP immunohistochemistry during early bridging events at 5 dpi (top) and after bridge formation at 2 wpi (bottom). The middle panel shows a high-magnification view of the boxed area in the top panel. In (D) and (E), dashed lines delineate spinal cord edges, arrows point to sites of bridging, and arrowheads point to ventral ependymal cells. Scale bars, 50 μm.

effects of *ctgfa-FL* (Fig. 3, D to F). By 6 wpi, swim capacity was markedly increased in *ctgfa-CT*–overexpressing animals relative to *ctgfa-NT*–overexpressing or wild-type clutchmates (Fig. 3D). Anatomically, *ctgfa-CT* overexpression resulted in a factor of >3 increase in glial bridging at 2 wpi and a factor of >2 increase in axon growth at 4 wpi (Fig. 3, E and F, and fig. S8B). These experiments indicate that the pro-regenerative activity of Ctgfa maps to its C-terminal domains.

To examine whether the effects of Ctgfa augmentation could be reproduced by localized delivery into the spinal cord lesion site, we injured wild-type animals and applied human recombinant CTGF (HR-CTGF-FL and HR-CTGF-CT) peptides adjacent to the lesion site at 5 and 10 dpi using a gelfoam sponge. We then assessed regeneration at 2 and 4 wpi, corresponding to 9 and 23 days after initial treatment. Human CTGF and zebrafish Ctgfa are 81% identical and 87% similar at the amino acid level within the four

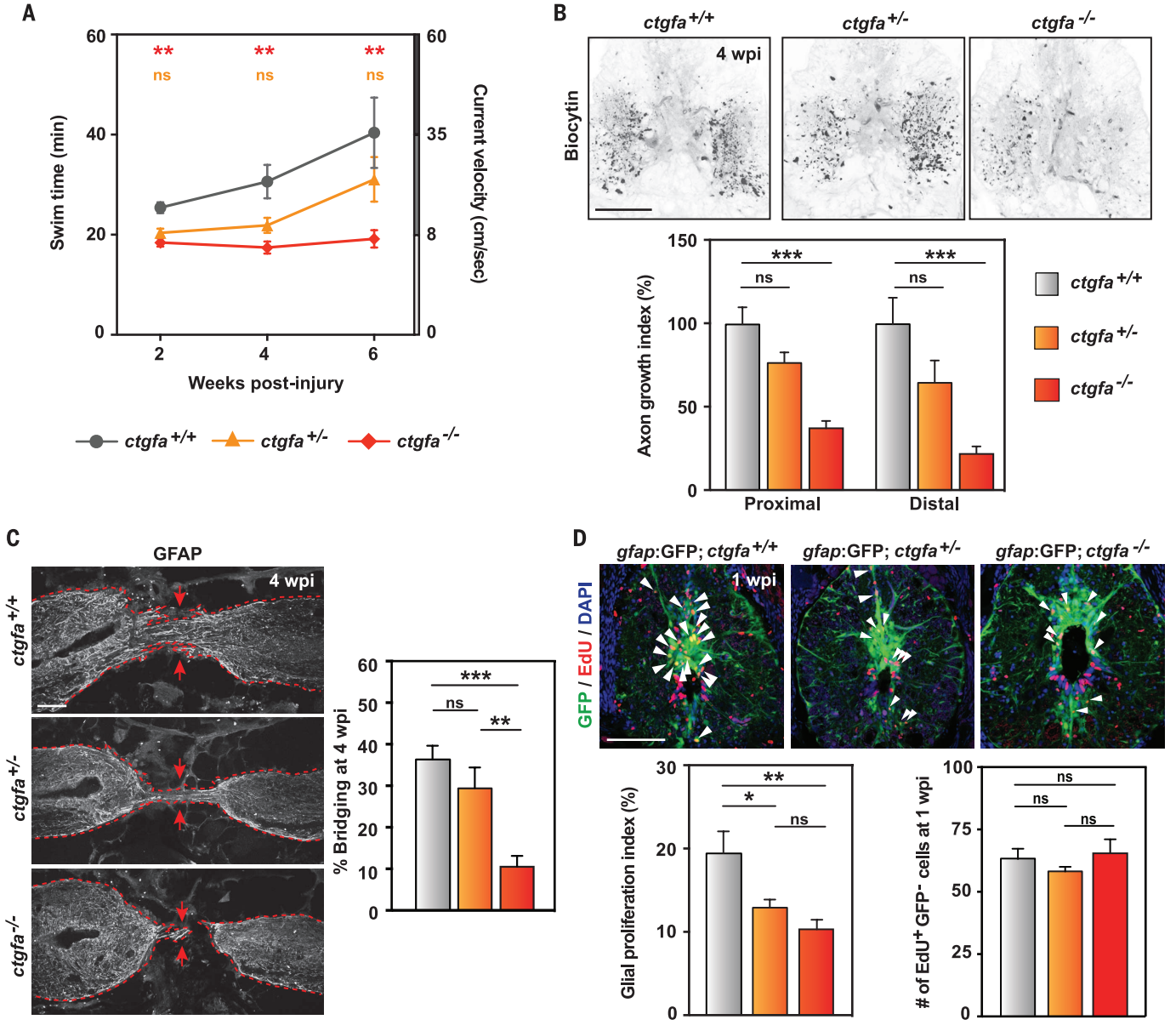



**Fig. 2. *ctgfa* is necessary for glial bridging and spinal cord regeneration.** (**A**) Swim assays assessed animals' capacity to swim against increasing water current inside an enclosed swim tunnel. Seven wild-type (*ctgfa*$^{+/+}$), 10 *ctgfa* heterozygous (*ctgfa*$^{+/-}$), and 10 mutant (*ctgfa*$^{-/-}$) clutchmates were assayed at 2, 4, and 6 wpi. Statistical analyses of swim times are shown for *ctgfa*$^{-/-}$ (red) and *ctgfa*$^{+/-}$ (orange) relative to wild type. Recovery of *ctgfa*$^{-/-}$ animals was not significant between 2 and 6 wpi. (**B**) Anterograde axon tracing in *ctgfa* mutant animals at 4 wpi. For quantification of axon growth at areas proximal (shown in images) and distal to the lesion core, 16 wild-type, 17 *ctgfa*$^{+/-}$, and 20 *ctgfa*$^{-/-}$ zebrafish from two independent experiments were used. (**C**) GFAP immunohistochemistry in *ctgfa* mutant spinal cords at 4 wpi. Percent bridging was quantified for 10 wild-type, 9 *ctgfa*$^{+/-}$, and 10 *ctgfa*$^{-/-}$ clutchmates from three independent experiments. Dashed lines delineate glial GFAP staining; arrows point to sites of bridging. (**D**) Glial cell proliferation in wild-type, *ctgfa*$^{+/-}$, and *ctgfa*$^{-/-}$ spinal cords at 1 wpi. Arrowheads indicate EdU-positive *gfap*:GFP-positive cells. For quantification of glial proliferation indices (left) and number of EdU-positive *gfap*:GFP-negative cells (right), 10 wild-type, 12 *ctgfa*$^{+/-}$, and 15 *ctgfa*$^{-/-}$ animals from two independent experiments were used. *$P$ < 0.05, **$P$ < 0.01, ***$P$ < 0.001; ns, not significant. Scale bars, 100 μm.

protein interaction domains (fig. S9). Treatment with either HR-CTGF-FL or HR-CTGF-CT enhanced zebrafish spinal cord regeneration (Fig. 3, G to I, and fig. S8C). Swim capacity was improved in HR-CTGF-FL–treated or HR-CTGF-CT–treated animals by 2 wpi (Fig. 3G), near that of uninjured animals. As early as 1 wpi, we observed increased GFAP expression and a factor of >10 enhancement of glial bridging in HR-CTGF-CT–treated animals relative to vehicle-treated controls (fig. S8D). Bridging remained greater at 2 wpi by a factor of ~3 (Fig. 3H). At 4 wpi, axon regeneration was increased by a factor of ~1.5 to 3 in HR-CTGF-FL–treated or HR-CTGF-CT–treated animals relative to vehicle controls (Fig. 3I). Application of exogenous CTGF protein at the lesion site rescued functional and anatomical spinal cord regeneration in *ctgfa* mutants, indicating specificity of both the phenotype and treatment (fig. S10). Thus, human CTGF protein, ostensibly via its C-terminal domains, enhances spinal cord regeneration in zebrafish.

Glial cell responses are thought to dictate the outcomes of spinal cord injury across species. We

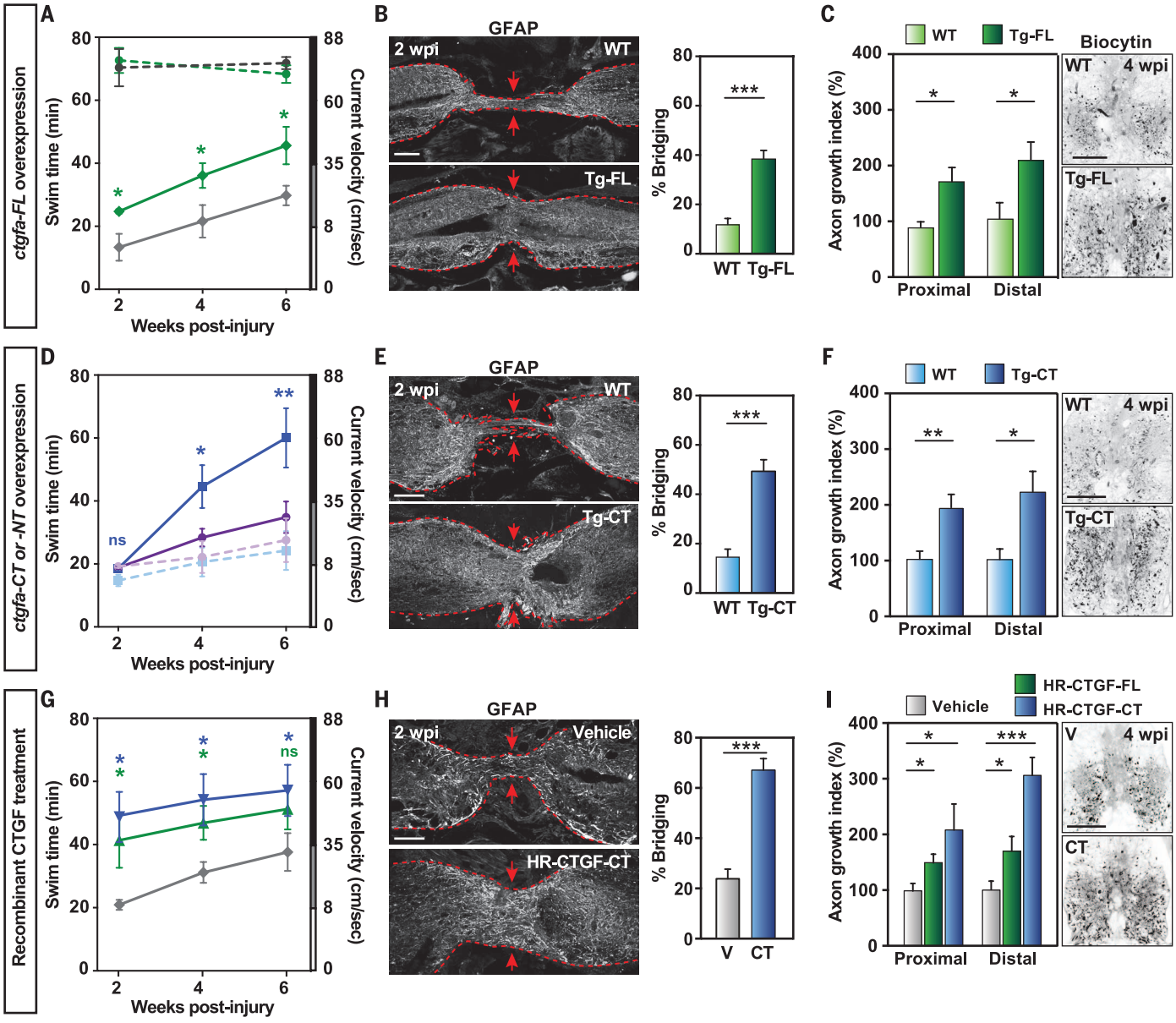



**Fig. 3. *ctgfa* promotes glial bridging and spinal cord regeneration.** (**A**) Swim assays determined motor function recovery of 10 *hsp70:ctgfa-FL* (green) and 10 wild-type (gray) clutchmates at 2, 4, and 6 wpi. For sham controls, 8 *ctgfa-FL*–overexpressing (dashed green) and 7 wild-type (dashed gray) zebrafish were analyzed. Statistical analyses of swim times are shown for injured *ctgfa-FL* (green) relative to wild type. (**B**) GFAP immunohistochemistry was used to quantify glial bridging at 2 wpi in 18 *ctgfa-FL*–overexpressing and 16 wild-type zebrafish from three independent experiments. (**C**) Anterograde axon tracing at 4 wpi after *ctgfa-FL* overexpression. Quantification at areas proximal (shown in images) and distal to the lesion core represents 12 *ctgfa-FL*–overexpressing and 10 wild-type zebrafish from two independent experiments. (**D**) Swim assays for 8 *ctgfa-CT*–overexpressing (blue), 10 *ctgfa-NT*–overexpressing (violet), and 9 wild-type clutchmate animals (wild-type controls for CT in dashed blue and for NT in dashed violet). Statistical analyses of swim times are shown for *ctgfa-CT* (blue) relative to wild type. (**E**) Glial bridging at 2 wpi in 19 *ctgfa-CT*–overexpressing and 20 wild-type animals from two independent experiments. (**F**) Anterograde axon tracing at 4 wpi after *ctgfa-CT* overexpression. Quantification represents 16 *ctgfa-CT*–overexpressing and 16 wild-type animals from two independent experiments. (**G**) Swim capacity was assessed for 9 vehicle-treated (gray), 8 HR-CTGF-FL–treated (green), and 9 HR-CTGF-CT–treated (blue) animals. Statistical analyses are shown for HR-CTGF-FL (green) and HR-CTGF-CT (blue) treatments relative to vehicle controls. (**H**) Glial bridging at 2 wpi in 18 HR-CTGF-CT–treated and 15 vehicle-treated animals from three independent experiments. (**I**) Anterograde axon tracing at 4 wpi after HR-CTGF-CT treatment. Quantification represents 18 vehicle-treated, 16 HR-CTGF-FL–treated, and 14 HR-CTGF-CT–treated animals from two independent experiments. For histology in (B), (E), and (H), dashed lines delineate glial GFAP staining and arrows point to sites of bridging. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. Scale bars, 100 μm.

found that Ctgfa promotes zebrafish spinal cord regeneration, at least in part by facilitating the proliferation and bridging activity of pioneer glial cells that also express *ctgfa*. Other cell types also induce *ctgfa* in the injured zebrafish spinal cord and might contribute additional functions during regeneration. In mammals, reactive gliosis causes scarring and inhibits regeneration (*13*, *14*), although evidence indicates that the lesion contains a heterogeneous pool of glial cells that release both axon growth–inhibiting and axon growth–permissive factors (*6*, *15*–*19*). We suggest

that identifying the mammalian glial subtype that can produce CTGF, or that is competent to respond to it, could reveal a pro-regenerative mammalian counterpart to the zebrafish bridging glia.

### REFERENCES AND NOTES


1. T. Becker, M. F. Wullimann, C. G. Becker, R. R. Bernhardt, M. Schachner, *J. Comp. Neurol.* **377**, 577–595 (1997).
2. S. P. Hui, A. Dutta, S. Ghosh, *Dev. Dyn.* **239**, 2962–2979 (2010).
3. Y. Goldshmit *et al.*, *J. Neurosci.* **32**, 7477–7492 (2012).
4. M. M. Reimer *et al.*, *J. Neurosci.* **28**, 8510–8516 (2008).
5. C. Y. Lin, Y. S. Lee, V. W. Lin, J. Silver, *J. Neurotrauma* **29**, 589–599 (2012).
6. V. J. Tom, C. M. Doller, A. T. Malouf, J. Silver, *J. Neurosci.* **24**, 9282–9290 (2004).
7. M. Hertel, Y. Tretter, C. Alzheimer, S. Werner, *Eur. J. Neurosci.* **12**, 376–380 (2000).
8. Y. Liu *et al.*, *Diagn. Pathol.* **9**, 141 (2014).
9. S. Conrad, H. J. Schluesener, M. Adibzahdeh, J. M. Schwab, *J. Neurosurg. Spine* **2**, 319–326 (2005).
10. J. Wang *et al.*, *Development* **138**, 3421–3430 (2011).
11. G. R. Grotendorst, M. R. Duncan, *FASEB J.* **19**, 729–738 (2005).
12. P. M. Robinson *et al.*, *Invest. Ophthalmol. Vis. Sci.* **53**, 8093–8103 (2012).
13. M. V. Sofroniew, *Trends Neurosci.* **32**, 638–647 (2009).
14. C. Raposo, M. Schwartz, *Glia* **62**, 1895–1904 (2014).
15. L. Waselle, X. Quaglia, A. D. Zurn, *Mol. Cell. Neurosci.* **42**, 315–327 (2009).
16. B. D. White *et al.*, *Stem Cells* **28**, 297–307 (2010).
17. M. A. Anderson *et al.*, *Nature* **532**, 195–200 (2016).
18. K. Meletis *et al.*, *PLOS Biol.* **6**, e182 (2008).
19. S. J. Hill, E. Barbarese, T. K. McIntosh, *J. Neuropathol. Exp. Neurol.* **55**, 1221–1229 (1996).


### ACKNOWLEDGMENTS

We thank A. Johnson, C. Eroglu, and M. Bagnat for discussions; N. Lee and K. Jones for technical and bioinformatics help; and the Duke University School of Medicine Zebrafish Shared Resource for animal care. Supported by NIH training grant T32HL007101 (M.H.M.), NIH grant R01 HL081674 (K.D.P.), the Max Planck Society (C.P. and D.Y.R.S.), and Duke University School of Medicine (K.D.P.). RNA sequencing data are archived at GEO (accession number GSE77025). The supplementary materials contain additional data. M.H.M. and K.D.P. are inventors on patent applications (62/159,413 and 62/398,781) submitted by Duke University that cover the use of Ctgf for spinal cord regeneration.

### SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/630/suppl/DC1
Materials and Methods
Figs. S1 to S10
Table S1
References (*20*–*26*)

20 January 2016; resubmitted 15 August 2016
Accepted 27 September 2016
10.1126/science.aaf2679

PLANT SCIENCE

# A selective insecticidal protein from *Pseudomonas* for controlling corn rootworms

**Ute Schellenberger,[1]* Jarred Oral,[1] Barbara A. Rosen,[1] Jun-Zhi Wei,[1] Genhai Zhu,[1] Weiping Xie,[1] Mark J. McDonald,[1] David C. Cerf,[1] Scott H. Diehn,[2] Virginia C. Crane,[2] Gary A. Sandahl,[2] Jian-Zhou Zhao,[2] Timothy M. Nowatzki,[2] Amit Sethi,[2] Lu Liu,[2] Zaiqi Pan,[2] Yiwei Wang,[2] Albert L. Lu,[2] Gusui Wu,[2] Lu Liu[1]†**

**The coleopteran insect western corn rootworm (WCR) (*Diabrotica virgifera virgifera* LeConte) is a devastating crop pest in North America and Europe. Although crop plants that produce *Bacillus thuringiensis* (*Bt*) proteins can limit insect infestation, some insect populations have evolved resistance to *Bt* proteins. Here we describe an insecticidal protein, designated IPD072Aa, that is isolated from *Pseudomonas chlororaphis*. Transgenic corn plants expressing IPD072Aa show protection from WCR insect injury under field conditions. IPD072Aa leaves several lepidopteran and hemipteran insect species unaffected but is effective in killing WCR larvae that are resistant to *Bt* proteins produced by currently available transgenic corn. IPD072Aa can be used to protect corn crops against WCRs.**

The coleopteran insect western corn rootworm (WCR) (*Diabrotica virgifera virgifera* LeConte) and related species cause extensive damage to corn (*Zea mays* L.) in North America and Europe, resulting in crop yield losses and management costs (*1*–*3*). Farmer adoption of corn plants that produce insecticidal proteins for controlling major corn pests, including WCRs, has increased productivity in areas with insect pressure (*4*). Commercial transgenic crops introduced to date for insect control have been developed with genes from *Bacillus thuringiensis* (*Bt*) (*5*, *6*). These genes encode insect-specific crystal proteins (Cry) or vegetative insecticidal proteins (Vip). Four of them—mCry3A, Cry3Bb1, eCry3.1Ab, and the binary protein complex Cry34Ab1/Cry35Ab1—have been commercialized for WCR control in the United States (*5*, *7*). Emergence of WCR resistance to mCry3A, Cry3Bb1, and eCry3.1Ab proteins (*8*–*10*) under field conditions demonstrates the need for broader, more durable strategies to control insect infestations and limit the development of resistance. The genetic diversity of *Bt* has already been explored in a search for additional insecticidal proteins (*5*). An alternative approach involves disabling WCR genes by RNA interference through the expression of double-stranded RNA (*11*). We hypothesized that, like *Bt*, other soil bacteria that live with soil-dwelling coleopteran larvae may produce proteins that kill corn rootworms. Here we describe a potent and selective insecticidal protein from a Gram-negative bacterium, *Pseudomonas chlororaphis,* which we identified through screening soil-isolated microbial strains. This protein protects transgenic corn plants from WCR injury under field conditions.

We collected soil samples from the U.S. states of Iowa, Illinois, and Wisconsin, and isolated microbial strains under growth conditions to favor species other than *Bt*. Crude protein extracts from these isolates were screened for activity against WCRs in artificial diet bioassays. One isolate—identified as *P. chlororaphis* by 16*S* ribosomal DNA sequence analysis (*12*)—showed a potent inhibitory effect on WCRs. We used multistep chromatography to isolate the insecticidal protein and achieved high-purity (>80%) enrichment in the final step (fig. S1). The major protein band with an apparent molecular weight of ~10 kDa was digested with trypsin and subjected to identification by liquid chromatography–tandem mass spectrometry (LC-MS/MS). We used the LC-MS/MS data of tryptic peptides to search against a protein database including protein sequences translated from the in-house genome sequence of this specific *P. chlororaphis* isolate. A high-confidence match to a hypothetical protein encoded by a gene from our specific *P. chlororaphis* isolate was obtained. This gene encodes a small protein of 86 amino acids, which we designated IPD072Aa (GenBank accession number KT795291). IPD072Aa was cloned and expressed in *Escherichia coli* to confirm that it was responsible for the insecticidal activity. Affinity-purified recombinant IPD072Aa was highly potent against WCR larvae and also active against northern corn rootworm (NCR) (*Diabrotica barberi*) larvae in diet bioassays (Fig. 1).

To evaluate the activity spectrum, we tested purified recombinant IPD072Aa protein at concentrations up to 875 μg/ml in diet assays against several lepidopteran and hemipteran species—including black cutworm (*Agrotis ipsilon* Hufnagel), corn earworm (*Helicoverpa zea* Boddie), European corn borer (*Ostrinia nubilalis* Hübner), fall armyworm (*Spodoptera frugiperda* J. E. Smith), soybean looper (*Pseudoplusia includens* Walker), and western tarnished plant bug (*Lygus hesperus* Knight)—and observed no mortality or inhibitory effect on any of those insect species. We also used an *Agrobacterium*-mediated transient gene expression system (*13*) to express IPD072Aa under control of a double-enhanced mirabilis mosaic virus viral promoter (*14*) in common bean (*Phaseolus vulgaris*). Sets of leaf disks transiently expressing

[1]DuPont Pioneer, Hayward, CA 94545, USA. [2]DuPont Pioneer, Johnston, IA 50131, USA.
*Present address: TeneoBio, 1490 O'Brien Drive, Menlo Park, CA 94025, USA. †Corresponding author: Email lu.liu@pioneer.com

IPD072Aa, Cry1F (GenBank accession number AAE8100) as a positive control, or no insecticidal protein (vector backbone as a negative control) were infested with two or three neonates from each of the same five lepidopteran species used in the diet assays. Leaf disks expressing Cry1F were fully protected from feeding damage across all of the lepidopteran species tested. In contrast, leaf disks expressing IPD072Aa were totally consumed and comparable to the negative control. Western blot analysis (fig. S2) indicated accumulation of IPD072Aa protein at 3067 ± 594 ng/mg of total extracted protein. Thus, IPD072Aa has a limited pest spectrum that includes corn rootworms.

We analyzed the primary sequence of IPD072Aa with protein prediction tools within InterProScan (comprising PROSITE, PRINTS, Pfam, ProDom, SMART, TIGRFAMs, and PIR SuperFamilies) (*15*). No known peptide motif, domain, or signature was identifiable. We used the alignment tool BLASTP (*16*) to identify proteins homologous to IPD072Aa in sequence databases and found 18 matches with primary sequence identity ranging from 38 to 83% (table S1 and Fig. 1). All homologs identified from public databases were annotated as hypothetical or uncharacterized proteins.

The majority of these proteins (12 of 18) were encoded in *Pseudomonas* spp. Three sequences were deposited from metagenomics analysis of microbial communities populating the switch grass rhizosphere (Joint Genome Institute: http://jgi.doe.gov), and three more IPD072 homologs came from *Burkholderia*, *Xenorhabdus*, and *Photorhabus* spp., respectively. Certain species of the latter two genera are known to produce high–molecular weight and multicomponent insecticidal protein complexes active against some of the lepidopteran species (*17*). Our findings suggest that additional proteins that can control rootworms might exist in these bacteria. To test this, a subset of IPD072 homologs was cloned, expressed, affinity-purified, and evaluated against WCR in diet bioassays. Several of these homologs, including distantly related IPD072Aa protein family members, possess insecticidal activity (Fig. 1C).

Transgenic corn plants were generated by *Agrobacterium*-mediated transformation (*18*) to assess IPD072Aa's potential for controlling corn rootworm in planta. Transformation was performed using a plasmid vector (ZmIPD072Aa) containing an expression cassette of an IPD072Aa gene under the control of a banana streak virus (BSV) promoter (see supplementary materials). Stable single-copy parental ($T_0$) transgenic plants (called events) expressing IPD072Aa were assayed in the greenhouse for root protection against WCR larvae, using 400 WCR eggs per plant. Root damage was assessed using the Iowa State University 0-3 Node-Injury Scale (*19*). Plants showing low node-injury scores (<0.10) were retested in the first generation ($T_1$) and demonstrated excellent root protection compared with controls (Fig. 2). In addition, expression of IPD072Aa in root tissue from these plants was detected by Western blot analysis (fig. S3).

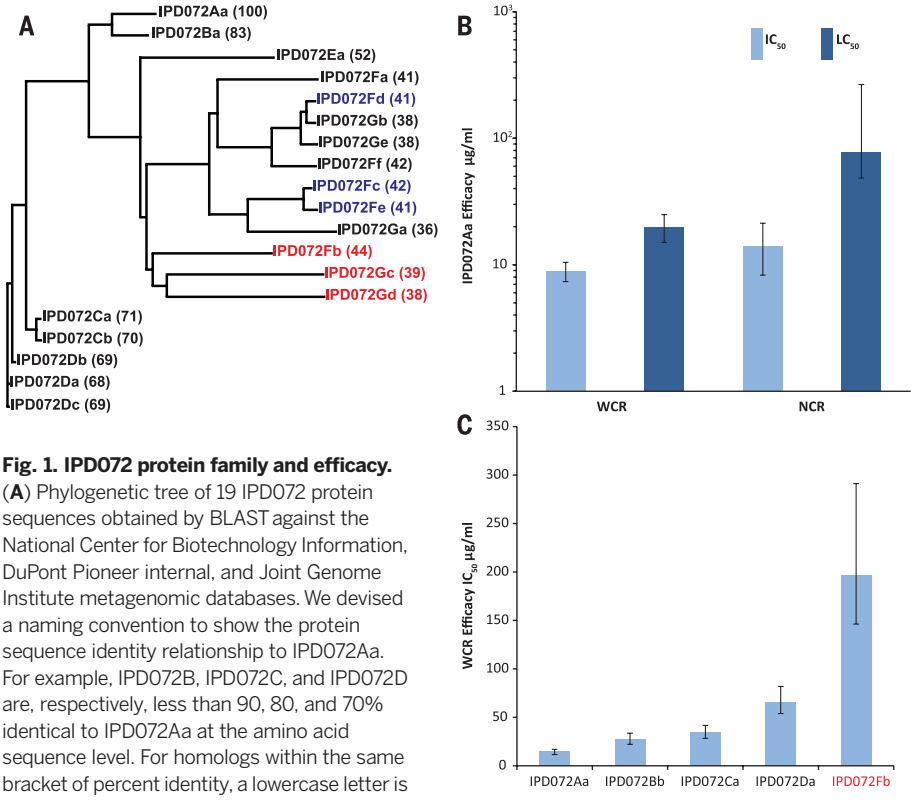



**Fig. 1. IPD072 protein family and efficacy.** (**A**) Phylogenetic tree of 19 IPD072 protein sequences obtained by BLAST against the National Center for Biotechnology Information, DuPont Pioneer internal, and Joint Genome Institute metagenomic databases. We devised a naming convention to show the protein sequence identity relationship to IPD072Aa. For example, IPD072B, IPD072C, and IPD072D are, respectively, less than 90, 80, and 70% identical to IPD072Aa at the amino acid sequence level. For homologs within the same bracket of percent identity, a lowercase letter is added (e.g., IPD072Fa, IPD072Fb, IPD072Fc, etc.). Numbers in parentheses represent the percent identity compared to IPD072Aa. Black, IPD072 homologs from *Pseudomonas* spp.; blue, proteins identified in metagenomes from microbial communities; red, proteins derived from *Burkholderia* (IPD072Fb), *Xenorhabdus* (IPD072Gc), and *Photorhabdus* (IPD072Gd). Detailed information about each protein can be found in table S1. (**B**) IPD072Aa activity on WCRs (8-day assay) and NCRs (7-day assay). Light blue bars, concentration to inhibit 50% of the larvae ($IC_{50}$) from survival (as dead) or normal growth (>60% reduction in size compared with control larvae); dark blue bars, concentration to kill 50 percent of the larvae ($LC_{50}$). Shown are values with 95% fiducial limits (FLs) (error bars); $n$ = 77 to 228 larvae, not for side-by-side direct comparison. (**C**) WCR insecticidal activity charted as the values of $IC_{50}$ from a 6-day assay of several IPD072 proteins. Shown are values with 95% FLs (error bars); $n$ = 144 to 209 larvae. IPD072Ea and IPD072Ga were also tested, and no activity was observed at a concentration of 400 µg/ml.

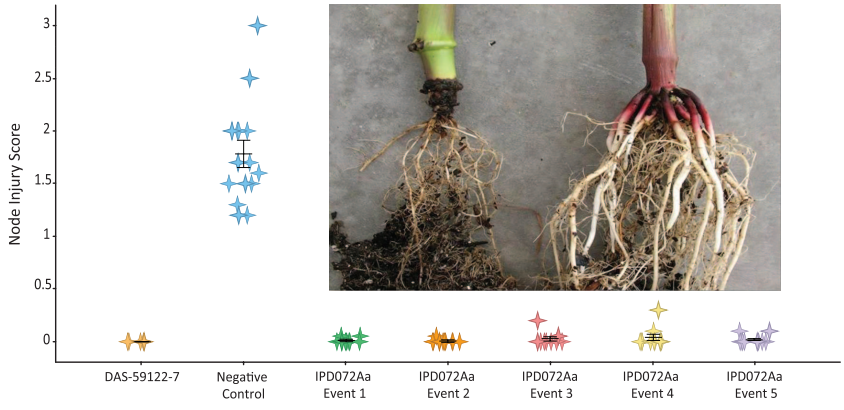



**Fig. 2. Root protection from WCR injury by transgenic expression of IPD072Aa in $T_1$ corn plants grown under greenhouse conditions.** Node-injury scores in five transgenic events expressing IPD072Aa under the control of the BSV promoter compared with nontransgenic corn (negative control) and the commercial corn line DAS-59122-7 expressing Cry34Ab1/35Ab1. The scatter plot shows mean values ± SE (error bars); $n$ = 10 plants each. (**Inset**) Roots of a plant expressing IPD072Aa (right), compared with a negative control (left).

For evaluation in the field, second-generation ($T_2$) nonsegregating hybrid maize seeds, derived from five original $T_0$ transgenic events of construct ZmIPD072Aa, were planted at five locations across the U.S. corn belt in 2014 (table S3). The field trials were conducted on land that contained late-planted conventional corn in the previous season to attract corn rootworm beetles for egg laying and to enhance natural infestations. In addition, plots at all locations were manually infested with 900 WCR eggs per plant between growth stages V2 and V4. On the basis of node-injury scores obtained from the negative control plants (table S3), corn rootworm larval feeding pressure was classified as high at Janesville (Wisconsin), Johnston (Iowa), and Readlyn (Iowa); moderate at Rochelle (Illinois); and low at Volga (South Dakota). General observations from all locations indicated that the predominant corn rootworm species was the WCR. At one location (Janesville, Wisconsin), however, the corn rootworm species composition was 44% WCR and 56% NCR, as indicated by beetle collections from emergence cages. Another location (Readlyn, Iowa) showed substantial root injury (mean ± SE node-injury score of 1.6 ± 0.15, $n$ = 30 plants) to a commercial hybrid corn line (MON88017) expressing only Cry3Bb1, which was planted in a separate comparative trial and exposed to natural corn rootworm infestations. The commercial hybrid corn line DAS-59122-7 (*20*), which expresses Cry34Ab1/Cry35Ab1, showed very low injury at this location (Table 1).

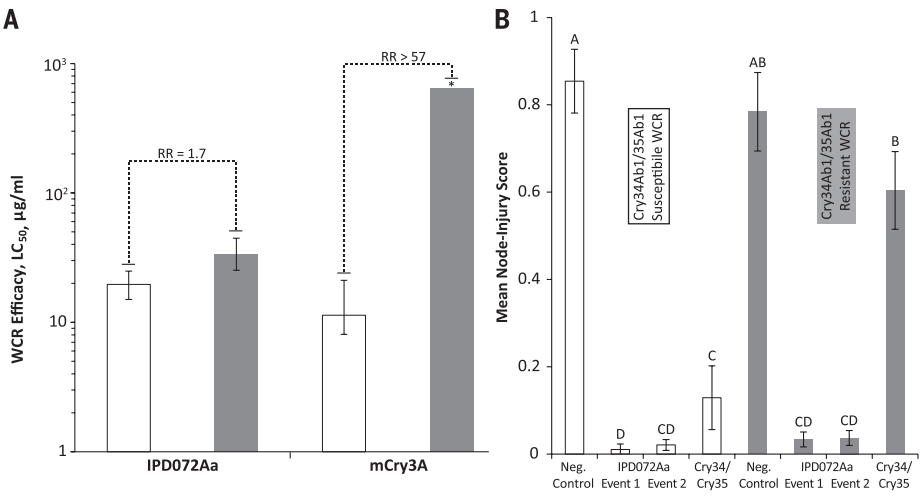



**Fig. 3. Cross-resistance evaluation of IPD072Aa against mCry3Aa and Cry34Ab1/Cry35Ab1.** (**A**) mCry3A- susceptible (white bars) and -resistant (gray bars) WCR larvae are inhibited by diet incorporation of recombinant IPD072Aa protein. The graph shows $LC_{50}$ values (in micrograms per milliliter) with 95% FLs (error bars) and resistance ratios (RRs), defined as the ratio of $LC_{50}$ values of resistant over susceptible colonies. A RR of 1.7 for IPD072Aa is unlikely to be biologically important. The asterisk indicates that no mortality is observed at the top concentration (647 µg/ml) for mCry3A on the resistant colony ($LC_{50}$ > 647) and that the exact RR cannot be calculated but is much higher than 57. (**B**) Expression of IPD072Aa provides root protection from both Cry34Ab1/Cry35Ab1-susceptible (white bars) and -resistant (gray bars) WCR larvae. Two independent transgenic events expressing IPD072Aa were analyzed in Rootrainer plant-based assays. A corn plant expressing Cry34Ab1/Cry35Ab1 was used as a positive control; nontransgenic corn served as a negative control. Results are mean values with 95% confidence intervals (error bars). Bars topped with different letters differ significantly at the 0.05 significance level (Tukey-Kramer test).

**Table 1. Comparison of node-injury scores among treatments at all five locations in 2014.** Estimated node-injury score: Injury from corn rootworm larval feeding was assessed with root scores on the Iowa State University 0-3 Node-Injury Scale (*17*). Within each location, estimated node-injury scores followed by different letters are significantly different ($P$ < 0.05). Locations were assigned a corn rootworm pressure level based on estimated node-injury scores of negative controls (low ≤ 0.75; moderate = 0.75 to 1.50; high ≥ 1.50).

| Location | Treatment | Estimated node-injury score | Standard error | 95% confidence intervals | |
|---|---|---|---|---|---|
| | | | | Lower | Higher |
| Volga, South Dakota | ZmIPD072Aa | 0.07 a | 0.04 | 0.00 | 0.14 |
| | DAS-59122-7 | 0.06 a | 0.05 | −0.04 | 0.16 |
| | Negative control | 0.73 b | 0.05 | 0.64 | 0.82 |
| Rochelle, Illinois | ZmIPD072Aa | 0.04 a | 0.13 | −0.22 | 0.30 |
| | DAS-59122-7 | 0.12 a | 0.16 | −0.19 | 0.43 |
| | Negative control | 1.17 b | 0.14 | 0.90 | 1.43 |
| Janesville, Wisconsin | ZmIPD072Aa | 0.05 a | 0.16 | −0.26 | 0.36 |
| | DAS-59122-7 | 0.04 a | 0.20 | −0.34 | 0.43 |
| | Negative control | 2.05 b | 0.17 | 1.72 | 2.39 |
| Johnston, Iowa | ZmIPD072Aa | 0.26 a | 0.26 | −0.25 | 0.77 |
| | DAS-59122-7 | 0.12 a | 0.31 | −0.49 | 0.72 |
| | Negative control | 2.45 b | 0.28 | 1.90 | 2.99 |
| Readlyn, Iowa | ZmIPD072Aa | 0.09 a | 0.13 | −0.17 | 0.35 |
| | DAS-59122-7 | 0.12 a | 0.16 | −0.19 | 0.44 |
| | Negative control | 2.33 b | 0.14 | 2.05 | 2.61 |

At all five testing locations, plants expressing the IPD072Aa protein provided good root protection from feeding injury by corn rootworm larvae (Table 1). Details on statistics for the fixed and random effects are summarized separately in table S4A and table S4B. The node-injury score for plants from construct ZmIPD072Aa was not significantly different from the root protection provided by the commercial hybrid corn line DAS-59122-7 (*20*) at all five locations ($P$ > 0.05, Table 1). Both constructs provided good protection compared with the negative control ($P$ < 0.05, Table 1). As reference, a node-injury score of 1.0 under field conditions has been estimated to cause a 15 to 18% reduction in corn grain yield (*21*, *22*). The presence of NCRs at Janesville did not negatively influence the efficacy of plants from construct ZmIPD072Aa, which suggests that the IPD072Aa protein may be effective against NCRs as well. Additionally, results at Readlyn indicated a low likelihood of cross-resistance between IPD072Aa and the Cry3Bb1 protein expressed by MON88017. Plants from construct ZmIPD072Aa and DAS-59122-7 each provided good root protection at this location.

Our field data from the Readlyn location suggest that WCRs resistant to Cry3Bb1 were not cross-resistant to IPD072Aa. We used bioassays to determine whether laboratory-selected colonies of WCRs resistant to mCry3A or Cry34Ab1/Cry35Ab1 (*23*, *24*) were cross-resistant to IPD072Aa. mCry3A-resistant and -susceptible WCR larvae were evaluated using IPD072Aa protein in artificial diet bioassays (Fig. 3A). The mCry3A protein was used as a positive control. The resistance ratio (RR), defined as $LC_{50}$ values (the concentration needed to kill 50% of the larvae) of resistant over susceptible colonies, was >57 for mCry3A but only 1.7 for IPD072Aa. A small RR of 1.7 for IPD072Aa is unlikely to be biologically important. To further evaluate any potential of weak cross-resistance

between this pair of insect actives, we conducted an additional probit analysis to assess the main effect of the WCR population and the interaction between bioassay dose and WCR population type. The results support no cross-resistance between mCry3A and IPD072Aa ($P > 0.05$, table S5). We then tested the effects of IPD072Aa on Cry34Ab1/Cry35Ab1-resistant and -susceptible WCR larvae in a greenhouse Rootrainer plant-based assay (*23*). Compared with Cry34Ab1/Cry35Ab1 plants, plants from the ZmIPD072Aa construct expressing IPD072Aa provided good root protection from feeding injury by Cry34Ab1/Cry35Ab1-resistant WCR larvae (Fig. 3B). No significant difference was found in the node-injury scores caused by Cry34Ab1/Cry35Ab1-resistant and -susceptible WCR larvae on the ZmIPD072Aa plants. Results from the Rootrainer assay suggest no cross-resistance between Cry34Ab1/Cry35Ab1 and IPD072Aa. Thus, IPD072Aa is effective in killing WCR insects that are resistant to either of the two types of *Bt* proteins used in current commercial transgenic corn lines.

The discovery of the potent and selective WCR active protein IPD072Aa from an isolate of *P. chlororaphis* indicates that soil microbes, other than *Bt*, have potential as sources of insecticidal proteins. Though its insecticidal mechanism is under further investigation, IPD072Aa appears to have functional similarities to *Bt*-derived insecticidal proteins. IPD072Aa is orally active against WCR larvae, suggesting an effect through the insect digestive system. When expressed in transgenic plants, IPD072Aa shows protection of corn roots from WCR larval injury. The ability of IPD072Aa to kill WCR larvae resistant to mCry3A or Cry34Ab1/Cry35Ab1 implies that its target site is different from each of these two *Bt* proteins. As resistance evolves among insect pests targeted by *Bt* corn seed products, IPD072Aa offers promise to effectively combat the challenges posed by continuing corn rootworm infestations.

## REFERENCES AND NOTES


1. R. L. Metcalf, in *Methods for the Study of Pest Diobrotica*, J. L. Krysan, T. A. Miller, Eds. (Springer, 1986), pp. vii–xv.
2. D. A. Andow *et al*., *J. Econ. Entomol.* **109**, 1–12 (2016).
3. M. E. Gray, T. W. Sappington, N. J. Miller, J. Moeser, M. O. Bohn, *Annu. Rev. Entomol.* **54**, 303–321 (2009).
4. J. Fernandez-Cornejo, J. J. Wechsler, M. Livingston, A. Mitchell, "Genetically engineered crops in the United States" (Economic Research Report no. ERR-162, U.S. Department of Agriculture, 2014).
5. K. E. Narva, B. D. Siegfried, N. P. Storer, *Adv. Biochem. Eng. Biotechnol.* **136**, 135–162 (2013).
6. G. Sanahuja, R. Banakar, R. M. Twyman, T. Capell, P. Christou, *Plant Biotechnol. J.* **9**, 283–300 (2011).
7. D. L. Frank, A. Zukoff, J. Barry, M. L. Higdon, B. E. Hibbard, *J. Econ. Entomol.* **106**, 2506–2513 (2013).
8. A. J. Gassmann *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 5141–5146 (2014).
9. S. N. Zukoff *et al*., *J. Econ. Entomol.* **109**, 1387–1398 (2016).
10. S. N. Zukoff, A. L. Zukoff, R. W. Geisert, B. E. Hibbard, *J. Econ. Entomol.* **109**, 1834–1845 (2016).
11. J. A. Baum *et al*., *Nat. Biotechnol.* **25**, 1322–1326 (2007).
12. W. G. Weisburg, S. M. Barns, D. A. Pelletier, D. J. Lane, *J. Bacteriol.* **173**, 697–703 (1991).
13. J. Kapila, R. De Rycke, M. Van Montagu, G. Angenon, *Plant Sci.* **122**, 101–108 (1997).
14. N. Dey, I. B. Maiti, *Plant Mol. Biol.* **40**, 771–782 (1999).
15. P. Jones *et al*., *Bioinformatics* **30**, 1236–1240 (2014).
16. S. F. Altschul *et al*., *Nucleic Acids Res.* **25**, 3389–3402 (1997).
17. R. H. ffrench-Constant, A. Dowling, N. R. Waterfield, *Toxicon* **49**, 436–451 (2007).
18. M. J. Cho *et al*., *Plant Cell Rep.* **33**, 1767–1777 (2014).
19. J. D. Oleson, Y. L. Park, T. M. Nowatzki, J. J. Tollefson, *J. Econ. Entomol.* **98**, 1–8 (2005).
20. D. J. Moellenbeck *et al*., *Nat. Biotechnol.* **19**, 668–672 (2001).
21. Z. Dun, P. D. Mitchell, M. Agosti, *J. Appl. Entomol.* **134**, 409–419 (2010).
22. N. A. Tinsley, R. E. Estes, M. E. Gray, *J. Appl. Entomol.* **137**, 161–169 (2013).
23. S. A. Lefko *et al*., *J. Appl. Entomol.* **132**, 189–204 (2008).
24. J. Z. Zhao *et al*., *J. Econ. Entomol.* **109**, 1369–1377 (2016).

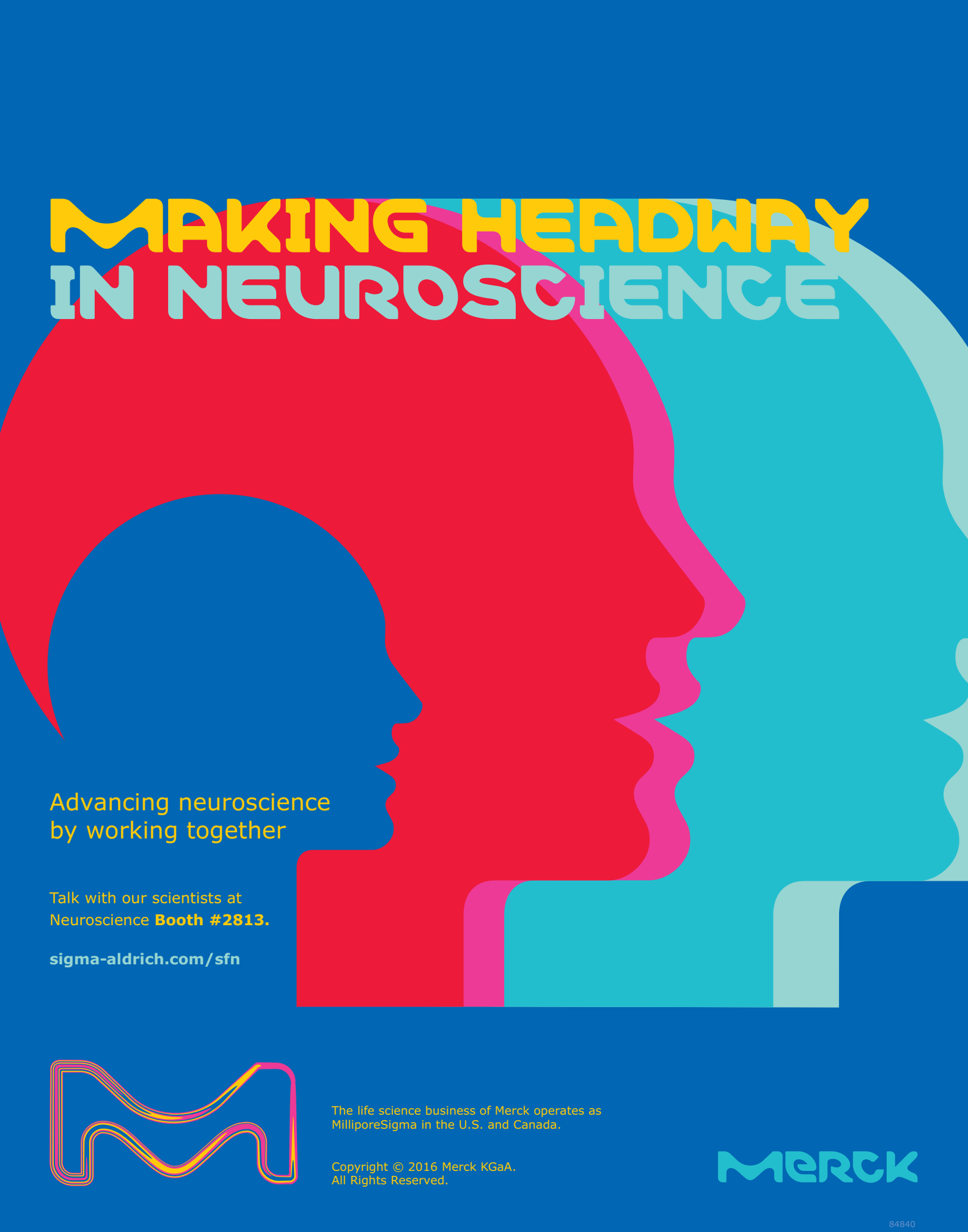

## ACKNOWLEDGMENTS

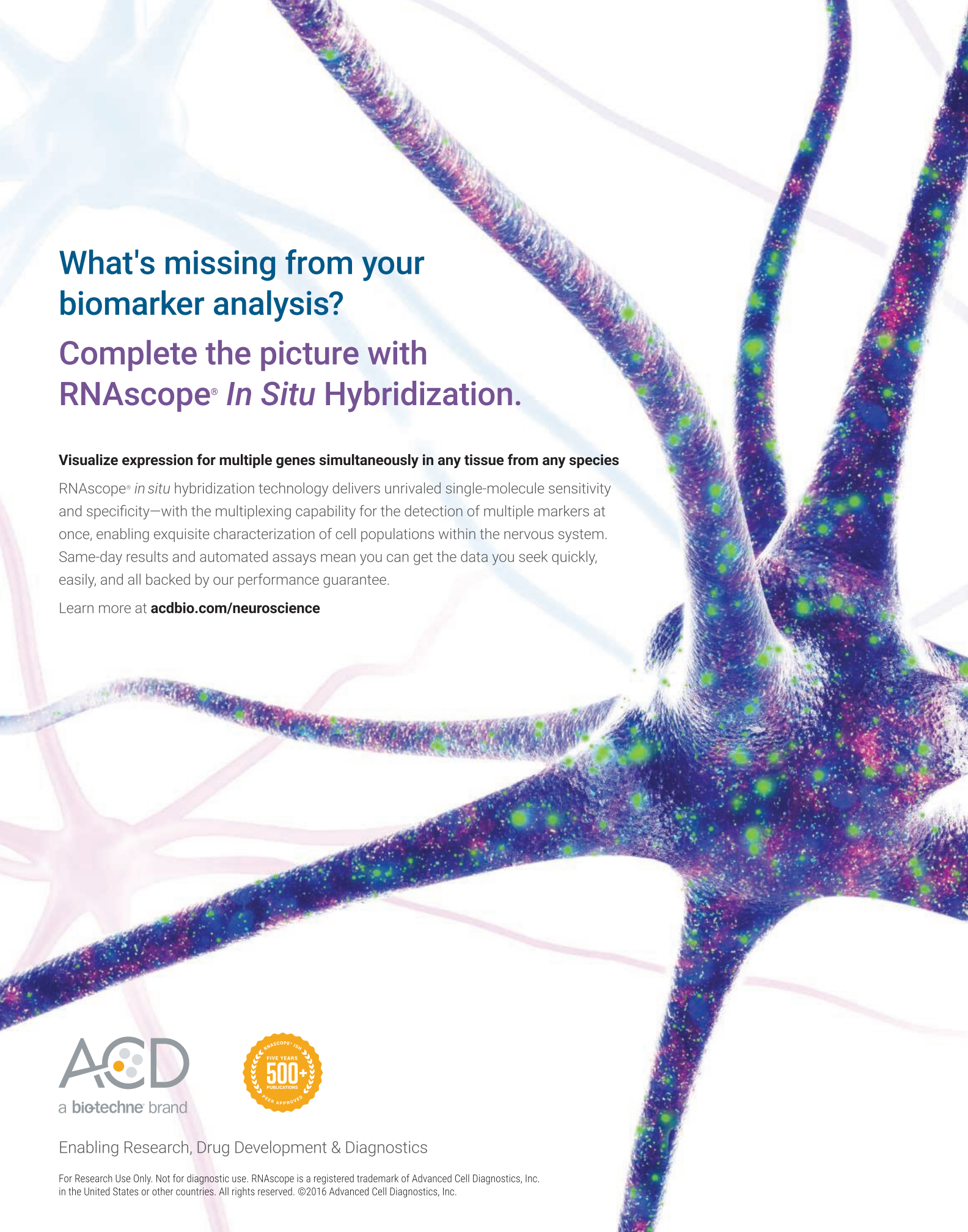

We thank N. Yalpani, A. Kassa, M. Heckert, J. Lee, D. Gillespie, and C. Hendrick for technical assistance and M. Nelson, L. Flexner, L. Higgins, S. Bass, K. McBride, I. Lamb, and D. Siehl for comments and suggestions. U.S., J.-Z.W., A.L.L., G.W., and L.L.[1] designed the experiments. J.O., B.A.R., G.Z., W.X., M.J.M., D.C.C., S.H.D., V.C.C., G.A.S., J.-Z.Z., T.M.N., and A.S. performed the experiments. Z.P., L.L.[2], and Y.W. analyzed the data. U.S., G.W., and L.L.[1] wrote the manuscript. A patent application (US2014/055128) entitled "Insecticidal proteins and methods for their use" has been filed. The materials reported in this paper may be subject to third-party ownership and/or governmental regulations. Availability of the materials described in this manuscript to academic investigators for noncommercial research purposes under an applicable material transfer agreement will be subject to the requisite permission from any third-party owners of all or parts of the material and to governmental regulation considerations. Obtaining permission from third parties will be the responsibility of the requestor. DNA sequences of IPD072Aa and its homologs were deposited in GenBank with accession numbers spanning KT95291 to KT95301. The supplementary materials contain additional information and data.


## SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/354/6312/634/suppl/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 to S5
References (*25*–*33*)


7 March 2016; accepted 9 September 2016
Published online 22 September 2016
10.1126/science.aaf6056


MAKING HEADWAY
IN NEUROSCIENCE
Advancing neuroscience
by working together
Talk with our scientists at
Neuroscience **Booth #2813.**
**sigma-aldrich.com/sfn**
The life science business of Merck operates as
MilliporeSigma in the U.S. and Canada.
Copyright © 2016 Merck KGaA.
All Rights Reserved.
MERCK
84840



# What's missing from your biomarker analysis?

## Complete the picture with RNAscope® *In Situ* Hybridization.

**Visualize expression for multiple genes simultaneously in any tissue from any species**

RNAscope® *in situ* hybridization technology delivers unrivaled single-molecule sensitivity and specificity—with the multiplexing capability for the detection of multiple markers at once, enabling exquisite characterization of cell populations within the nervous system. Same-day results and automated assays mean you can get the data you seek quickly, easily, and all backed by our performance guarantee.

Learn more at **acdbio.com/neuroscience**

ACD
a bio-techne brand

RNASCOPE ISH
FIVE YEARS
500+
PUBLICATIONS
PEER APPROVED

Enabling Research, Drug Development & Diagnostics

# Unique opportunities for fellows at SciLifeLab

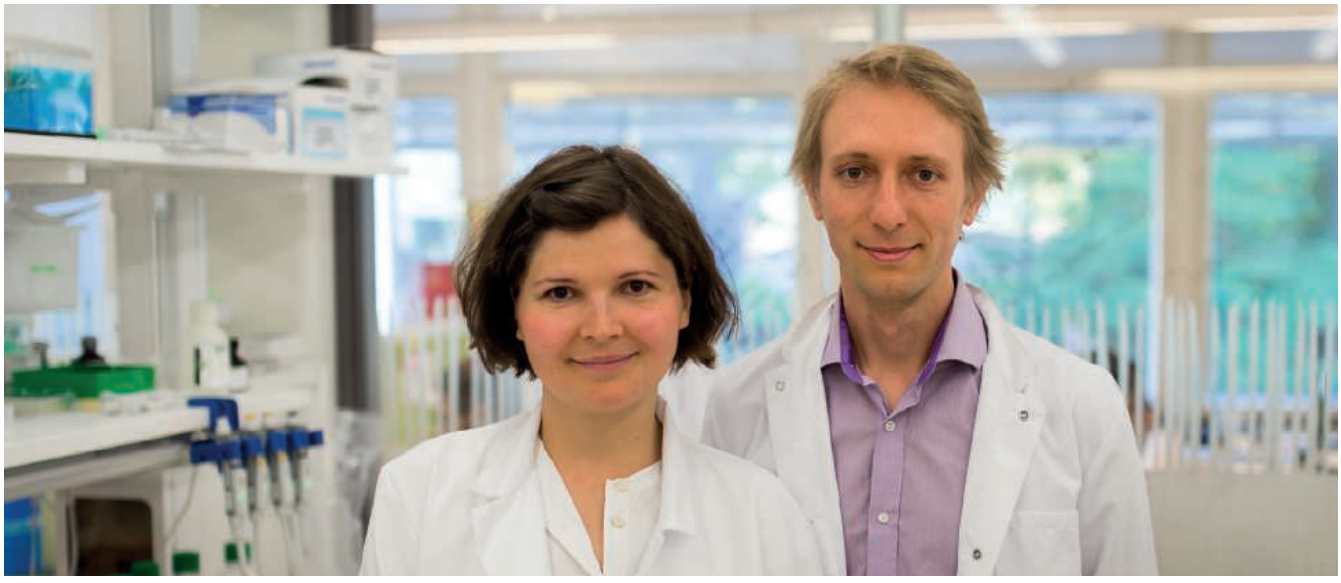


Photo: Daniel Rosik

"SciLifeLab is a beautiful workplace, both as a building and as an organization", says Simon Elsässer (right). "Yes, I like the spirit and the freshness here, that it is all new and there is a lot of enthusiasm", says Magda Bienko (left).

**SciLifeLab, Science for Life Laboratory, is a Swedish research center within molecular biosciences with a focus on health and environment. To further strengthen the research environment at SciLifeLab the center regularly recruits young, talented research leaders to become SciLifeLab fellows.**

Each fellow is recruited by one of the center's host universities and receives funding from them. The first call for SciLifeLab Fellows was opened in 2013 and 17 fellows have now been recruited, mainly internationally. Three of those also received ERC starting grants for 2016. And it seems that being part of the SciLifeLab fellows program played a key role in attracting that funding.

"Thanks to the funding from the SciLifeLab fellows program, we have been able to perform preliminary experiments that enabled us to show other granting agencies that we have potential. It really made me an independent fellow from the start", says Simon Elsässer, fellow at Karolinska Institutet/SciLifeLab.

Sebastian Deindl, fellow at Uppsala University/SciLifeLab agrees:
"The SciLifeLab fellows program is really quite unique in the sense that it provided a complete package, including funding and access to advanced infrastructure, which gives stability to perform ambitious and challenging science."

## Multidisciplinary environment

"It was very appealing for me that SciLifeLab brings multiple research areas from several universities together in one building", says Magda Bienko, fellow at Karolinska Institutet/SciLifeLab. "It was also comforting to know that I would start at the same time as other young group leaders. I liked the idea of being able to contribute to shaping the environment at the center", says Magda Bienko.

"We are all ambitious, young scientists with different backgrounds, and I am excited and humbled to be given the chance to do something meaningful," says Sebastian Deindl.

Photo: Mikael Wallerstedt

"The SciLifeLab fellows program is quite unique in the sense that it provides a complete package compared to most other places", says Sebastian Deindl.

SciLifeLab

The fellows were recruited from three different universities; Magda Bienko from Massachusetts Institute of Technology, Sebastian Deindl from Harvard University, and Simon Elsässer from University of Cambridge. They complement each others research since they are all interested in chromatin and developing techniques to study it. Their interest spans from the architecture of the genome and its dynamic regulation by histone proteins to the mechanisms of chromatin-associated complexes studied at the single-molecule and atomic level.

"We have invested the first few months at SciLifeLab to set up our respective methods, now there is a lot of room for collaboration and combining these methods", says Magda Bienko.

### Access to infrastructure

All researchers in Sweden have access to the SciLifeLab infrastructure. For example, Magda Bienko collaborates with the superresolution microscopy facility at the center.

"Getting access to this facility was one of the very appealing aspects of Sweden for me since my research rely very much on this", says Magda Bienko. "The expertise within the SciLifeLab facilities is really good, not only in terms of help with actual experiments but also with the theory behind. For example there is a bioinformatics drop-in support every week which we use and that is working really well for us."

"I have gotten so used to having access to the SciLifeLab infrastructure that I don't even think about it", says Sebastian Deindl. "I am collaborating with the facility for cryo-electron microscopy as well as the ones for protein science and chemical biology. It would completely change the character of my research did I not have access to this type of infrastructure."

### Looking forward to a bright future

It actually seems that the three fellows have come to like SciLifeLab and Sweden so much that they plan to stay here.

"I think that the future here will be very bright and I would like to stay in Sweden and in Stockholm, it is really good to live here", says Simon Elsässer. "I hope we are going to make some great discoveries the following years, leading to exciting and unexpected findings."

Magda Bienko agrees:
"I want to stay at SciLifeLab and I wish that it remains very dynamic and I am willing to contribute to that. Science is always the top priority and I can do the science I want here very well so this is where I see myself in the future."

"We are making some really exciting progress and it is more fun than ever in the lab," says Sebastian Deindl.

### SciLifeLab – a national resource

Being a joint effort by four Swedish universities (Karolinska Institutet, KTH Royal Institute of Technology, Stockholm University and Uppsala University) SciLifeLab offers a cross-disciplinary research setting that interacts with healthcare, authorities and industry to meet the need for new clinical methods and a better environment. Founded in 2010, the center toady encompasses more than 1 200 researchers mainly located in and around the two center nodes in Stockholm and Uppsala.

SciLifeLab is also a Swedish national center with the mission to develop, use and provide advanced technologies for applications in health and environmental research. The center infrastructure encompasses a multitude of biomolecular technologies and bioinformatics services. National funding makes SciLifeLab's services and expertise available to researchers in all of Sweden. In addition, SciLifeLab provides education for students and researchers at all levels.

# Any Sample Type, Any Application, FastPrep-24™ 5G has You covered!

## Powerful • Intuitive • Flexible


## Sample Preparation for Nucleic Acid and Protein Extraction

FastPrep-24™ 5G is the market leader in sample preparation. Complete and thorough grinding of up to 48 samples in 40 seconds or less, guaranteed. The system is intuitive with an interactive user-friendly touchscreen that includes more than 70 recommended protocols.

Interchangeable sample holders give you flexibility in sample tube processing sizes (2 ml, 4.5 ml, 15 ml or 50 ml) at either cryogenic or ambient temperature. Optimized Lysing Matrix and all-inclusive Purification Kits make the FastPrep-24™ 5G truly the most advanced sample preparation system ....Period!

# Often Imitated, Never Duplicated!

*Visit www.mpbio.com/sampleprep-offer for current special promotions on FastPrep Instruments, purification kits and lysing matrix.*

## MP Biomedicals - It's What's Inside that Counts!

**www.mpbio.com**

**MP Biomedicals USA: 800.854.0530 • MP Biomedicals Europe: 00800.7777.9999**

MP™

# AAAS 2017 ANNUAL MEETING

FEBRUARY 16-20 • BOSTON

SERVING SOCIETY THROUGH

# SCIENCE POLICY

## Register Now

Register and book your hotel now. Reduced rates are available until January 27, 2017. AAAS members are eligible for further discounts.

**aaas.org/meetings**

AAAS, publisher of *Science*, thanks the sponsors and supporters of the 2017 Annual Meeting

European Commission

RESEARCH COUNCILS UK

SUBARU

HELMHOLTZ | ASSOCIATION

THE KAVLI FOUNDATION

for its generous support of the Science Journalism Awards

NANYANG TECHNOLOGICAL UNIVERSITY

SNOLAB

Bristol-Myers Squibb

*As of October 31, 2016*

AACR
American Association for Cancer Research

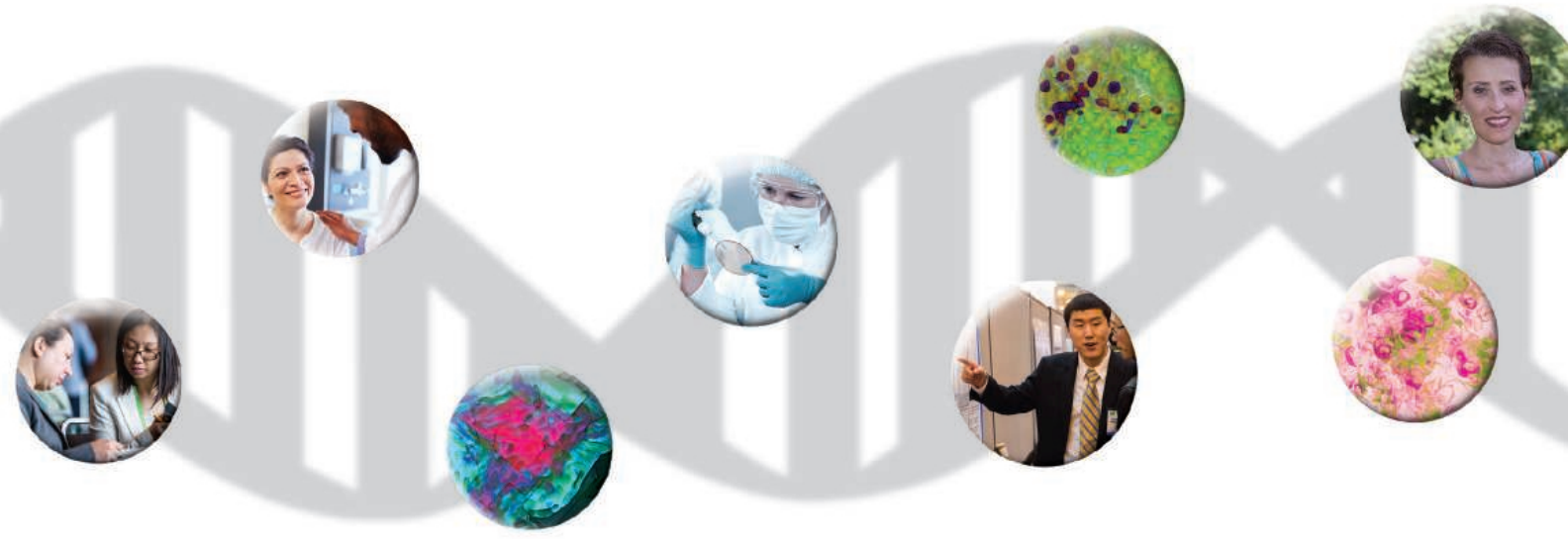
# ANNUAL MEETING

2017 WASHINGTON, DC

RESEARCH PROPELLING CANCER PREVENTION AND CURES

**April 1-5, 2017 • Walter E. Washington Convention Center • Washington, DC**

**AACR.org • #AACR17**



# Join us in Washington, DC

## for the latest innovative and inspiring cancer research from around the world... the AACR Annual Meeting 2017!

You won't want to miss this five-day, comprehensive program offering world-class opportunities to learn, collaborate, and share not only the progress made in cancer research, but also the promise of what is to come. This new era of cancer discovery is unprecedented with respect to rapidly emerging cancer science and new and effective molecularly targeted therapies.

The best and brightest researchers in the world will attend this meeting that covers every aspect of cancer – from molecular biology, clinical studies, epidemiology, and prevention to survivorship. And you'll want to be among them.

**Submit your abstract by THURSDAY, NOVEMBER 17.**

**Register and save by JANUARY 29.**

**Become a Member!**
Join the AACR and receive a discount on registration.

**We look forward to seeing you in Washington, DC!**

110
1907 • YEARS • 2017

AACR
American Association for Cancer Research

# NEUROSCIENCE IN YOUR HANDS

Eugenia, Principal Scientist

With a strong focus on Neuroscience, Atlas Antibodies is proud to offer a wide selection of PrecisA Monoclonals and Triple A Polyclonals for this research area. Take a look at our **Neuroscience Marker Panel**, targeting the main anatomical and neurochemical cell types of the nervous system of humans and rodents.

And don't miss The Atlas of the Mouse Brain, showing antibody stainings of consecutive sections of the entire mouse brain: *www.proteinatlas.org/humanproteome/mouse+brain*

**Welcome to Atlas Antibodies.**
The world leading provider of antibodies for immunohistochemistry.

MADE IN SWEDEN

ATLAS ANTIBODIES
Learn more today at atlasantibodies.com

eppendorf & Science

# PRIZE FOR NEURO BIOLOGY

Congratulations
to Gilad Evrony, M.D., Ph.D.
Harvard Medical School /
Boston Children's Hospital
and Mount Sinai Hospital

# And the Winner Is...

**Eppendorf & Science Prize for Neurobiology**
Congratulations to Gilad Evrony on winning the 2016 Eppendorf & Science Prize for his work on developing technologies to sequence and analyze the genomes of single cells from the human brain. Dr. Evrony's research has revealed a diversity of mutations in neuronal genomes indicating that every neuron in the brain carries a unique fingerprint of somatic mutations. Such mutations can cause focal brain malformations and may have a role in other unsolved neurologic diseases. The technology also allows, for the first time, reconstruction of developmental lineage trees in the human brain to study how cells proliferate and migrate to build the brain.

The annual US$25,000 Eppendorf & Science Prize for Neurobiology honors scientists, like Dr. Evrony, for their outstanding contributions to neurobiological research. Gilad Evrony is the 15th recipient of this international award. He will be presented with the Prize at a ceremony held during the week of the 2016 Annual Meeting of the Society for Neuroscience in San Diego.

**You Could be Next Year's Winner.**
If you are 35 years of age or younger and currently performing neurobiological research, you could be next to win the 2017 Prize. Deadline for entries is June 15, 2017.

Learn more at: www.eppendorf.com/prize

eppendorf Science AAAS

# Mouse with a milkshake: Behavioral windows into brain function

Neuroscientists are employing new techniques and technologies to use animal behavior as a window into brain function. For instance, researchers are teaching mice to nose-poke a touchscreen by doling out strawberry milkshakes as part of cognitive testing. It's a light-hearted reminder of the similarity between mice and humans, perhaps—but hopes are rising that new therapies developed using mouse models of Alzheimer's disease, for example, can be translated to humans.
**By Caitlin Smith**

For years, neuroscientists have gained new revelations about how the brain works by separately studying genes, neurons, or animal behavior. Yet they are surely far from understanding the entirety of brain function. Today, neuroscientists are making strides by developing new tools and combining existing techniques to yield fresh insights.

One strategy is to use behavior as a specific entryway into learning how the brain operates. A simple behavior—like the patellar or knee-jerk reflex your doctor tests by tapping the patellar tendon in your knee—indicates the healthy functioning of a specific region of your central nervous system. But studying behaviors that involve multiple neural and motor pathways and different brain areas—such as eating and physical activity—remains challenging. And researchers encounter even more obstacles when trying to analyze complex behaviors like social interactions. Neuroscientists are now harnessing new technologies and advanced computing capabilities that analyze both simple and complex behaviors to increase our knowledge of brain function.

## Body in motion, mind in motion

As any neurologist can tell you, small movements speak volumes about what's going on inside the brain. For example, supination—turning the hand and forearm to the palm-up position—can indicate the health of the corticospinal tract, a band of nerve fibers in the brain that ferries information about limb movements to and from the spinal cord.

Neurologists can detect brain or spinal cord damage—sustained during a stroke, for example—by evaluating a patient's supination ability. "This is a classic neurological sign for corticospinal tract injury," says Jason Carmel, director of the Motor Recovery Laboratory at the **Burke Medical Research Institute.** Patients with damage to the corticospinal tract usually show arm drift when asked to hold their arms out palms-up, as if holding up a pizza box. But researchers can't ask rats to pretend to hold a pizza box. So Carmel, also an attending neurologist, developed a supination task for rats to quantify the effectiveness of therapies—ultimately to find the best therapies to help his patients regain motor function.

Carmel worked with scientists at **Vulintus,** whose forelimb assessment system MotoTrak was designed to replace the near-constant human supervision required for traditional motor assessments with precision sensors and computer monitoring for behavioral reinforcement. Pretraining animals for motor assessment tasks can take weeks or months, via traditionally one-on-one training, which can monopolize the lab's time. MotoTrak automates almost all of an animal's training and testing. Now, says Carmel, "we can get studies done more quickly and use a larger cohort of animals."

MotoTrak uses an adaptive training protocol, automatically adjusting the task parameters required for rewards to keep the animal motivated. In MotoTrak's supination module, "animals are trained to reach for a doorknob-like manipulandum [object] and turn the knob a specified number of degrees to trigger a [food pellet] reward," says Andrew Sloan, director at Vulintus. Once rats are fluent in the supination task, MotoTrak automatically tests how well they supinate before and after different therapies for corticospinal tract injury. These therapies include standard neurorehabilitation exercises, and future work may incorporate electrical stimulation of the brain and/or spinal cord. Carmel hopes to use this new device to develop therapies for corticospinal tract injuries that will benefit his patients.

## Home sweet home

Rodents are the most common models for behavioral studies, and the advent of knockout technology has made mice especially interesting subjects. Traditionally, lab mice live in home cages—not their natural habitat, but as "normal" as possible for mice who live in a lab—with researchers removing them temporarily to test devices such as mazes or social recognition chambers. But increasingly, researchers are recognizing the advantages of testing animals in their home cages whenever possible, as it reduces stress on the animals. Testing mice in a home cage environment is especially important when measuring behavioral parameters that might be more easily confounded by the effects of human handling or changes in environment. Several companies offer home cages that double as behavioral testing chambers.

The PhenoTyper home cage from **Noldus Information Technology** (NIT) was specifically developed for behavioral tracking—a video camera and lights on top are optimized for video recording, eliminating shadows and illuminating the cage floor uniformly. "In the PhenoTyper, animals can move about, burrow



*Upcoming Features*
*Cell Analysis—November 25* ■ *Genomics—January 13* ■ *Digital Lab Management—February 3*

in bedding, eat and drink, dig holes, and play—[it's cozier] than the older test cages, which are bare plastic and far from a natural environment," says Lucas Noldus, founder and managing director of NIT.

PhenoTyper is currently used in many behavioral studies of knockout mice. NIT is a partner in the Innovative Medicines Initiative project **EU-AIMS** (European Autism Interventions—A Multicentre Study for Developing New Medications), a consortium of academia and pharmaceutical companies searching for therapeutic drugs for autism spectrum disorder. NIT is providing the PhenoTyper cage and software for preclinical studies at Utrecht University Medical Center, the Netherlands. Hoffman-La Roche, Pfizer, and other partners are testing the effects of drugs on behaviors in mouse models for autism.

## Putting the "move" in movie

In conjunction with the PhenoTyper cage, NIT offers EthoVision XT automatic behavior recognition software, which monitors rodents in their normal environment around the clock by video tracking. Unlike other systems that track an animal as one center-of-mass point, EthoVision finds the animal's contour and follows it as three points—nose point, center of mass, and tail base—connected by two vectors, and tracks many other body features extracted from the video image.

Tracking an animal using points and vectors allows EthoVision to derive much more information automatically. For example, the angle formed by the two vectors (which intersect at the center of mass) tells you whether the animal's body is stretched out straight, curled up, or turned to the left or right. The direction of the front vector indicates the direction an animal is heading (if moving), and its field of view. If the back vector is stationary and the front vector is moving side-to-side, then the animal is moving its head back and forth, scanning its environment.

EthoVision, which automatically recognizes 10 different behaviors, has been used to develop animal models for Parkinson's and Alzheimer's diseases, and to screen for therapeutic drugs by looking for behavioral changes in the animal models, says Noldus. NIT also works with other technology companies, such as **Data Sciences International,** which manufactures wireless telemetry sensors. A small microchip-like sensor injected into an animal can transmit vital signs such as body temperature and heart rate to EthoVision, which synchronizes the physiological vital signs to the behavioral data. "In a resident–intruder paradigm, this lets you ask if the heart rate is elevated when the intruder is confronted by an angry resident animal that is defending its territory," says Noldus, "and whether the stress on the animal is suppressed by anxiolytic drugs."

## Mind over food: Latest behavior tracking technology

Researcher Catherine Kotz, a professor in the **Department of Food Science and Nutrition at the University of Minnesota** and the Minneapolis VA Medical Center, along with collaborators Charles Billington and Jennifer Teske, are also applying new behavior tracking technology in obesity research. The team applies multiple techniques in studying the role of the neuropeptide orexin in the brain's regulation of eating behaviors and energy expenditure in rodents.

Kotz uses Promethion system home cages from **Sable Systems,** because they are identical to the animals' home cages and do not require an adaptation period—which she emphasizes is especially important for energy balance measurements. "A characteristic sign of stress for these animals during acclimatizing to a new environment is weight loss, so you can't study weight loss," says Kotz. "All the assumptions that go into measuring energy being used by metabolic rate and physical activity must be met, and this requires a stable energy balance." And the Promethion system offers more sensitive measurements of food and water intake, body weight, physical activity, and energy expenditure than any system she has previously encountered. In addition, it allows for measurement of total energy expenditure instead of energy expended due to metabolic rate only. "This system enables us to get much better data," she says.

Orexin affects eating behavior as well as physical activity level, so Kotz wants to measure the energy expended during particular activities. "Other systems are OK for measuring basic metabolic rate, but if you want to measure the calories burned during a certain behavior, the other systems don't work," she says. The Promethion system lets her do this, because it can return synchronized behavioral and metabolic information in real time. It uses light beams in the X, Y, and Z planes so that beam breaks record the animal's movement path inside the cage. In addition, the cage contains a balance to measure body weight, a running wheel, and sensors on the food and water dispensers. This information can be paired with time-stamped energy expenditure data to determine the calories burned during a certain behavior.

Orexin is produced in the hypothalamus and projects throughout the brain. To study its effects, Kotz is applying the technique of optogenetics, using mice engineered to express light-sensitive ion channels in orexin-producing neurons. When light is pulsed onto the brains of these mice via a small cable attached to the head, the neurons release this peptide. Kotz is also applying the cont.>

### Featured Participants

**Actual Analytics**
www.actualanalytics.com

**Burke Medical Research Institute**
www.burke.org/research

**Cambridge University**
www.cam.ac.uk

**Campden Instruments**
www.campdeninstruments.com

**Charles River Discovery**
discovery.criver.com

**Data Sciences International**
www.datasci.com

**Department of Food Science and Nutrition, University of Minnesota**
fscn.cfans.umn.edu

**EU-AIMS**
www.eu-aims.eu

**International Mouse Phenotyping Consortium**
www.mousephenotype.org

**Lafayette Neuroscience**
www.lafayetteneuroscience.com

**Med Associates**
www.med-associates.com

**MRC Harwell**
www.har.mrc.ac.uk

**Noldus Information Technology**
www.noldus.com

**Sable Systems**
www.sablesys.com

**Stoelting**
www.stoeltingco.com

**Vulintus**
www.vulintus.com


pharmacogenetic DREADD (designer receptor exclusively activated by designer drugs) approach to define orexin action for physical activity energy expenditure. Kotz's lab is verifying the electrophysiology of the engineered mice in collaboration with University of Minnesota neuroscientist Mark Thomas, and it soon plans to study their behaviors in response to light stimulation using the Promethion system.

Kotz is especially interested in the behavioral outcomes of the optogenetics and DREADD approaches because in previous work, they introduced orexin into the brain by injection. The hypothalamic neurons that release orexin also release other substances at the same time, possibly leading to different behaviors than would be attributable to orexin alone. With optogenetics and DREADD, "we'll be able to look at the physiology of the whole orexin neuron and everything it releases when stimulated—and the behaviors that follow," she says.

## Of mice and men: Social behavioral monitoring

Researchers from **MRC Harwell**, a Medical Research Council institute in the United Kingdom, are analyzing behaviors of knockout mice in conjunction with the **International Mouse Phenotyping Consortium.** Their goal is to look at the pleiotropic effects of knocking out every protein-coding gene in mice—a feat that involves characterizing over 20,000 knockouts with many hours of behavioral observation. "We are hoping that 24-hour monitoring will allow us to see subtle changes to the behavior, feeding patterns, activity, and social interactions of many of our mutant lines," says Sara Wells, director of the Mary Lyon Centre at MRC Harwell.

MRC Harwell recently adopted the **Actual Analytics** system for behavioral studies of knockout mice, using the ActualHCA home cage environment that fits into their high-density cage racks. According to Patrick Nolan, neurobehavioral group leader at MRC Harwell's Mammalian Genetics Unit, the system has two advantages: the ability to monitor behaviors in an undisturbed home cage environment, and the opportunity for group housing. Social animals by nature, mice are healthier and less stressed in a group-house setting than when housed alone, which also allows researchers to study social interactions.

Automatic behavioral tracking of more than one animal has always been tricky, because the monitoring system cannot always distinguish animals. Actual Analytics makes this possible using radio frequency identification (RFID) technology, in which a microchip is placed just under the skin of each animal (just like microchipping your pets at the vet). Underneath the ActualHCA cage is an RFID reader that sends information to a computer to help track the identity of each animal. An array of detectors and high-definition video also tracks the animals' behaviors. "Our analytical tools measure a range of spontaneous behaviors such as locomotion, circadian rhythms, climbing, and interactions between animals," says Douglas Armstrong, chief science officer of Actual Analytics. "These are under constant development, with new behaviors regularly being added."

For Nolan's group, which studies the genetic basis of neuropsychiatric disorders, the ActualHCA (home cage analysis) system lets them record social behaviors more accurately, and tests whether mixing genotypes in a single cage influences behaviors. He hopes that studying mouse mutants will yield insight into behaviors that might "provide benefits not only for our specific research domains, but also for neurological and neurodegenerative disease research in general."

## Touchscreen technology taps into brain behavior

At the forefront of cognitive testing in mice is touchscreen technology, which is generating excitement because such testing is easily translated from mouse models to humans. Mice are trained in a Bussey-Saksida touchscreen chamber, named for its developers Tim Bussey and Lisa Saksida, both researchers at **Cambridge University** in the United Kingdom. Inside the chamber is a touch-sensitive video screen that displays images. Mice learn that when they touch the correct image(s) on the screen with their noses (so called "nose-pokes"), they are rewarded. Each correct nose-poke earns them something tasty, like a droplet of strawberry milkshake. Several companies offer Bussey-Saksida chambers, including **Campden Instruments, Lafayette Neuroscience, Med Associates,** and **Stoelting.**

At **Charles River Discovery** (part of Charles River Laboratories), head of translational biology Maksym Kopanitsa is using these touchscreen chambers to study cognition in mouse models of brain disorders such as Alzheimer's and Huntington's diseases. Kopanitsa's group wants to know which cognitive parameters change the most in mouse disease models, so that Charles River's clients can then use those parameters to test novel drug treatments—perhaps finding a drug that reverses the altered parameter back to that of normal mice.

The touchscreen tests used with mouse models of Alzheimer's disease resemble those used in human Alzheimer's patients, making translation easier. "Touchscreen technology allows one to evaluate mouse models in tasks that are directly relevant to human learning and memory, and to test how well drugs improve performance on those tasks," says Kopanitsa. He hopes to eventually perform the touchscreen experiments simultaneously with electrophysiological measurements of brain activity, in order to observe neuronal circuits engaged during cognition. "Having these measurements will essentially close the circle by providing evidence that a mouse model reproduces a human brain disorder behaviorally and physiologically."

Armed with touchscreen tests and other techniques, today's neuroscientists are opening new behavioral windows into how the brain functions. As researchers come to better understand the link between function and behavior, discoveries may be translated into better therapeutics for some of our most devastating medical disorders.

*Caitlin Smith is a freelance science writer based in Portland, Oregon.*

DOI: 10.1126/science.opms.p1600109

### CRISPR Knockout Kit

OriGene offers genome-wide gene knockout/knockin kits via CRISPR (human and mouse). A kit is available against every human gene locus, containing two guide RNAs (gRNAs) and a compatible donor vector with a functional cassette. In addition to gene knockout, these kits can also be used for promoter strength study for each gene locus. The predesigned nature of the kit greatly reduces the researcher's time and effort. CRISPR/Cas9-based genome editing has become a popular tool for targeted genome manipulation because of its simplicity and high cutting efficiency. This system requires a functional Cas9 protein and a gRNA for effective double-stranded breakage at a desired site. OriGene has developed the pCas-Guide system, a dual-function vector with both gRNA and Cas9 expression. OriGene has also designed a set of donor cassettes for construction of donor vectors. These cassettes include *Luciferase-Loxp-Puro-Loxp*, *tGFP-Loxp-Puro-Loxp*, and *tRFP-Loxp-BSD-Loxp*.

**OriGene Technologies**
For info: 888-267-4436
www.origene.com

### Small Animal Physiological Monitoring System

The Harvard Apparatus Small Animal Physiological Monitoring System integrates monitoring of rectal temperature, electrocardiogram, respiration, oxygen saturation, blood pressure, and exhaled $CO_2$. Designed to make surgery and other manipulations on small animals easier, the system provides superior monitoring of multiple physiological parameters on a single small platform. It features a real-time display of numeric data and user-defined graphic waveforms. An advanced wireless interface and touch-screen display allow the user to view, add notes, and save and export data for easy analysis. The system's controlled heating platform maintains the animal's body temperature at the desired level. A stereotaxic adaptor option is also available.

**Harvard Apparatus**
For info: 800-272-2775
www.harvardapparatus.com


### Neural Basal Medium

PRIME-XV Neural Basal Medium is a chemically defined medium that supports long-term cultures of human and rodent neuronal stem cells for neurological research. Neural progenitors are sensitive to culture conditions and difficult to obtain. Compared to media containing serum or undefined components, PRIME-XV's serum-free formula is designed to increase experimental reproducibility and culture consistency to maximize the use of these cells. The ability to culture primary and reprogrammed (induced pluripotent stem cell–derived) neural progenitors is offering neurological researchers better ways to screen drugs for toxicity, understand brain cancer, and use patient cells for disease modeling to advance the treatment of neurological diseases such as Alzheimer's, Parkinson's, and epilepsy. PRIME-XV Neural Basal Medium can be used with PRIME-XV IS21, a serum-free supplement for neuronal culture that is prescreened for efficacy and avoids the risks that serum presents to neuronal cultures.

**Irvine Scientific**
For info: 800-577-6097
www.irvinesci.com

### iPSC-Derived Astrocyte Kits

Axol Bioscience has expanded its range of neuroscience research tools to include an induced pluripotent stem cell (iPSC)-Derived Astrocyte Progenitor Kit and iPSC-Derived Mature Astrocyte Kit. These kits include cryopreserved cells, optimized media, and supplements. The new kits offer researchers a complete culture system that is easy to use and promotes long-term cell viability for use in disease modeling and drug discovery applications. Astrocytes are capable of lipid uptake and serve to support, repair, and maintain the central nervous system by providing essential nutrients that encourage neuronal maturation. Astrocyte dysfunction has been implicated in a number of neurological conditions such as Alzheimer's and Parkinson's diseases, autism, amyotrophic lateral sclerosis, Rett syndrome, and schizophrenia. Axol iPSC-Derived Astrocyte Progenitor Kits are designed to generate mature astrocytes in under 3 weeks, and iPSC-Derived Mature Astrocytes express prototypical phenotype markers such as glial fibrillary acidic protein.

**Axol Bioscience**
For info: 800-678-2965
www.axolbio.com

### Laser Beam Combining Filters

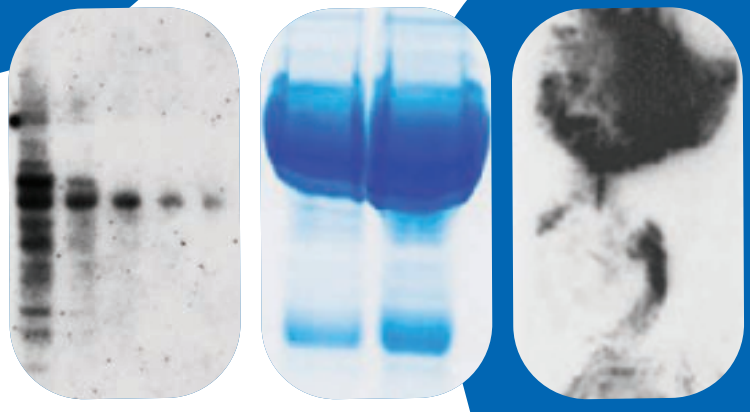
New Multiphoton LaserMUX beam combiners enable deeper tissue imaging and improved contrast in multicolor and multimodal fluorescence microscopy. The filters set new performance standards by simultaneously achieving high transmission, high reflection, and low group delay dispersion (GDD) over both reflection and transmission, while maintaining minimal wavefront distortion. These beam-combining filters allow users to achieve higher transmission and reflection levels over an expanded wavelength range, with wavefront performance optimized for deep tissue and in vivo imaging. Ideal for combining two femtosecond pulsed laser beams, they are perfect for optogenetics and other life science applications.

**IDEX Health & Science**
For info: +49-(0)-1801-808-800
www.idex-hs.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/products/newproducts.dtl 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.

Millipore®
Filtration, Separation
& Preparation



# it's not you. it's your tools.

## The right Western choices can take the "**ugly**" out of your next blot.

**Western blotting tools from MilliporeSigma help tailor your choices to your target:**

- Ultradurable, tear-resistant TruPAGE™ precast gels
- Complete selection of PVDF and nitrocellulose membranes to help you make the right choice for transferring your protein of interest
- 30 minute immunodetection with the SNAP i.d.® 2.0 system for cleaner, more consistent data
- Ready-to-use reagents like the ultrasensitive Luminata™ substrate
- Application-specific antibody manufacturing expertise, with over 70,000 tested in Western blot

**Show us your ugliest blots.** We'll provide tips and tricks for revealing beautiful Western data in our Protein Blotting Handbook, 6th edition.

**sigma-aldrich.com/westernblot**

P.S. If your blot has a big fingerprint in the middle…
it might actually be you.

The life science business of Merck KGaA, Darmstadt, Germany
operates as MilliporeSigma in the U.S. and Canada.



MilliporeSigma

*By* **Pierfrancesco Riccardi**

# *The beauty of outreach*

After about 20 years as an academic researcher in experimental condensed matter physics, I got involved in outreach quite by chance. It happened a couple years ago, when some friends were writing a proposal for their company to run "edutainment" activities for the children in our town. Teaching at the university level had been an important part of my job for years, but I had never been involved in anything like what they were proposing. Nonetheless, it sounded like fun, so I suggested a partnership between their company and my university's physics department. This is how I discovered the joy of outreach, which has ended up taking my career in an unexpected direction.

When the proposal was funded, I offered to help set up and supervise the activities—simple experiments that explored basic concepts such as light and energy using balloons, plastic bottles, batteries, and other everyday materials. With no experience in this kind of environment, I didn't know quite what to expect. As I worked with the children, I was interested to discover that they gained as much from their interactions as from the content of the activities. Their excitement grew as they developed a community by collaborating or, more often, engaging in friendly competition. I found that I greatly enjoyed facilitating this kind of fun and socializing through physics.

***"It has helped me find a new level of community and connection."***

After hearing about my work on these edutainment activities, the director of the physics department asked me to coordinate our departmental outreach program's main project, which aims to enrich physics education for students in their final years of high school. I was grateful for this opportunity, which allowed me to make outreach part of my professional life. I have also found, unexpectedly, that it has been even more important on a personal level. It has helped me find a new level of community and connection with students, which has reignited my own excitement about physics and my appreciation of the relationships that can emerge from sharing in the scientific inquiry process.

Working with the graduate students involved in the project has shown me how collaboration can create a vibrant, creative community, which simultaneously makes the work more fun and leads to a better final result. When I started on the project, graduate students in the department merely helped conduct established educational activities, but soon, on their own initiative, many of them started to propose new activities related to their research. In turn, we started to work more as a team. Now, our meetings feature lively discussions about the status of the outreach project and possible directions for future development. Everyone contributes, leading to warm, meaningful personal interactions. Seeing the students thrive in this environment has made me realize that our outreach program serves not merely to explain scientific concepts, but also to share the idea that participating in science is a positive, enriching personal experience.

A few months ago, when the new director of the department asked me to take on the responsibility for the entire program, I accepted enthusiastically, effectively choosing outreach as my primary activity. I still plan to continue my research, because these activities are complementary and I like the way they complete my work life, but I know that doing both outreach and research will be a challenge. Nonetheless, I'm confident that it will be worthwhile, and I look forward to continuing to develop our outreach program. I want to engage the high school students as well as educate them so that they can experience the joy of sharing their scientific enthusiasm in a communal environment.

It all started for fun, but in just a couple years, my work life has completely changed. Now, when school students ask for advice about pursuing a career in science, I tell them that the beauty of science is engagement and collaboration with a diverse group of individuals. Thanks to outreach, I rediscovered this beauty. ■

*Pierfrancesco Riccardi is a researcher at the University of Calabria in Italy.* *He thanks the students and postdocs involved in the Progetto Lauree Scientifiche. Send your career story to SciCareerEditor@aaas.org.*