Why Europe rules the field of ancient DNA *p. 1384*

iPhone, FBI, and tech-saavy law enforcement *p. 1398*

Habits control the behavior of cocaine users *p. 1468*

Science

$15
17 JUNE 2016
sciencemag.org

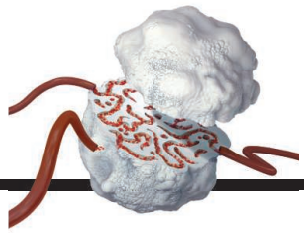
AAAS

SPECIAL ISSUE

SIGNALS IN RNA

CONTENTS

17 JUNE 2016 • VOLUME 352 • ISSUE 6292



SPECIAL SECTION

Signals in RNA

SEE ALSO

▸ PERSPECTIVES PP. 1390 & 1391

▸ RESEARCH ARTICLES PP. 1425 & 1430

ON THE COVER

Artist's interpretation of RNA activity (the train represents a ribosome). RNA serves many biological purposes: It carries genetic information, regulates gene expression, and functions as a structural component of cells' molecular building machinery. RNA is replete with signals that control its activity, and its pivotal role in the cell has made it an attractive candidate for development as a therapeutic agent. For more on these topics, see the special section beginning on page 1406. *Illustration: Charles Williams/ madeup.org*

NEWS

IN BRIEF

IN DEPTH

FEATURES

INSIGHTS

PERSPECTIVES

CONTENTS

1396, 1436, & 1474
Restoring stem cell maintenance

17 JUNE 2016 • VOLUME 352 • ISSUE 6292

RESEARCH

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* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

Published by AAAS

Science
AAAS
1200 New York Avenue, NW, Washington, DC 20005
Bateman House, 82-88 Hills Road, Cambridge, UK CB2 1LQ

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

New leadership for Science

Next month, the American Association for the Advancement of Science (AAAS) welcomes a new editor-in-chief for the *Science* family of journals, Jeremy M. Berg, as Marcia McNutt steps aside to assume her role as president of the U.S. National Academy of Sciences. This passing of the baton is always an exciting time for AAAS, because it presents an opportunity to refresh views on *Science*'s trajectory in the increasingly complex landscape of science's place in society. In this regard, *Science* and the other *Science* family journals will be in good hands with Jeremy Berg.

A scientist of great breadth and accomplishment in structural chemistry and biochemistry, science education, and science policy, Berg brings creative ideas and good judgment about scientific publishing. Currently the Associate Senior Vice Chancellor for Science Strategy and Planning at the University of Pittsburgh, he has been, among other things, chair of the Department of Biophysics and Biophysical Chemistry at Johns Hopkins University and director of the National Institute of General Medical Sciences. This distinguished background, along with his passion for data-driven analysis of science policy, dedication to increasing the diversity of perspectives among scientists, and enthusiasm about leading the *Science* family of journals, makes Berg well suited for the job.

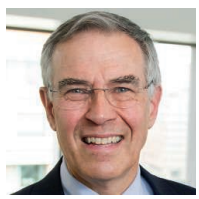
Berg joins a line of distinguished editors. Since its beginning 168 years ago, AAAS has had a mission to foster communication among scientists and with the public. At the turn of the 20th century, AAAS addressed this mission by taking over a struggling journal called *Science*, founded two decades earlier by Thomas Edison and kept alive by Alexander Graham Bell. The journal grew and thrived over the decades, with each editor-in-chief bringing something different to *Science*. To mention a few, Philip Abelson (1962–1984) broadened the publication's scope, modernized its peer-review process, and created the News section as we know it today; Daniel Koshland Jr. (1985–1995) solidified *Science*'s stature internationally in the emerging fields of cell and molecular biology and expanded commentaries; Floyd Bloom (1995–2000) ushered *Science* into the digital age; Donald Kennedy (2000–2008) gave the science of sustainability new prominence; and Bruce Alberts (2008–2013) brought emphasis to science education and oversaw the first expansion of research journals, with *Science Signaling* and *Science Translational Medicine*. Marcia McNutt not only launched the first online, open-access family member, *Science Advances*, but also the forthcoming digital journals *Science Immunology* and *Science Robotics*. And she has led the charge in directing international attention to higher standards in peer review and reproducibility in all science journals.

"...the job of the editor-in-chief may seem easy. That is far from the truth."

With such a history of success, it's understandable how the job of the editor-in-chief may seem easy. That is far from the truth. Over the time that AAAS has published *Science*, these leaders have had to set policy on a variety of matters, ranging from the fair balance of scientific disciplines covered in Research and in News; to government restrictions on data; to the content and accuracy of articles and even advertising; to the efficiency, thoroughness, and fairness of refereeing and editing; to issues of prior publication and conflicts of interest. We are fortunate that *Science*'s editors-in-chief have been remarkably wise in making such decisions and in advancing the scientific enterprise.

With so much ferment in the world of scientific publishing today, there will surely be more important decisions ahead. Editor-in-chief Berg (along with AAAS's new publisher, Bill Moran) takes leadership of a family of journals that is healthy and influential at a time when there is both a need and an opportunity to make changes to better serve science and society in the modern day. AAAS is fortunate to have such leadership in its publishing.

– Rush Holt



Rush Holt is Chief Executive Officer of AAAS and Executive Publisher of the Science *journals.*

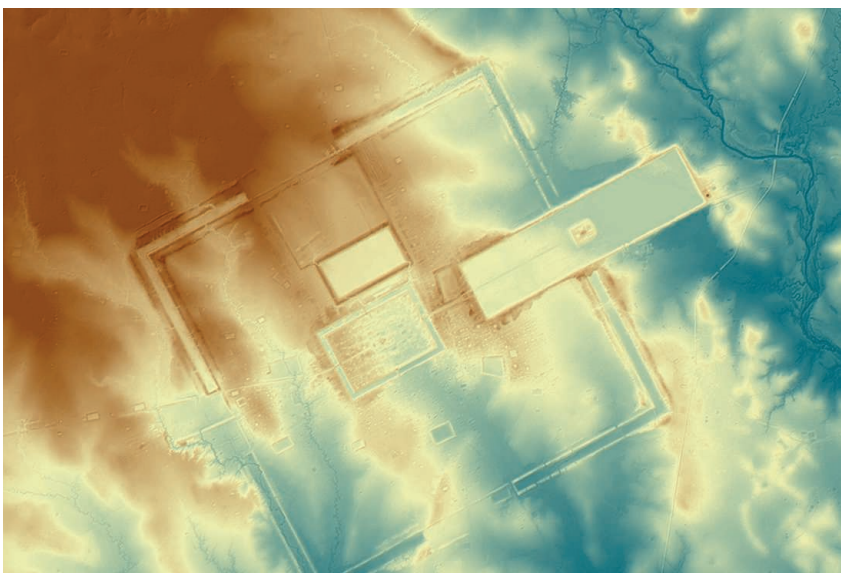
10.1126/science.aag3342

NEWS

> "It is like if the trans-Atlantic phone cable is cut by half, and someone wants to put it together using Krazy Glue."
>
> **University of Pennsylvania transplant surgeon Abraham Shaked**
> to *The New York Times*, criticizing a group's plan to transplant a head onto another body using a gluelike substance to connect spinal cord nerves.

IN BRIEF

Lost cities around Angkor Wat



This 120-square-kilometer digital terrain model shows where a moat encircled a city at Preah Khan.

From the ninth to the 15th century, the Khmer empire spanned much of what is now Cambodia, central Thailand, and southern Vietnam. Modern tourists flock to the massive 12th century temple complex at Angkor Wat in Cambodia. But in recent years, scientists have used airborne laser scanning technology, or LiDAR, to reveal that the site's true medieval urban footprint spanned a much larger area, some 1000 square kilometers. Now, after expanding the scanning area to some 1900 square kilometers, the team reports this week that they have discovered multiple cities beneath the forest between 900 and 1400 years old. Unlike the stone temples, most Khmer structures were built of perishable materials such as wood. But LiDAR reveals where they once stood by highlighting humanmade changes in topography, such as the earthen mounds used as foundations for the wooden structures. Among the most startling finds in the study, in press in the *Journal of Archaeological Science*, is the hidden city of Preah Khan (shown), where a 12th century temple still stands. Archaeologists had long suspected that it was also a major metropolis—but spent a decade searching fruitlessly for signs of it in the jungle. The LiDAR data showed that it had been hiding in plain sight.

PHOTO: DAMIAN EVANS/CAMBODIAN ARCHAEOLOGICAL LIDAR INITIATIVE

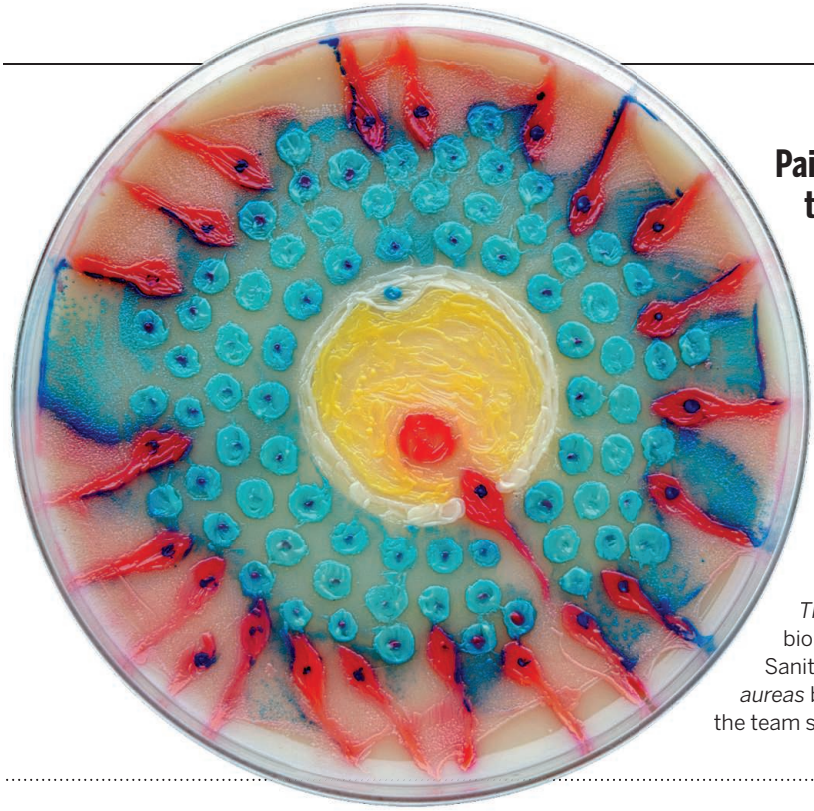
AROUND THE WORLD

A hard look at Canada's research

OTTAWA | Canadian Minister of Science Kirsty Duncan on 13 June named a nine-member panel of experts to do a "holistic assessment" of a decade of policies in the country aimed at converting university labs into tools for industrial development and commercialization. The policies were set under the previous prime minister, Stephen Harper. To lead the study, Duncan appointed former University of Toronto President David Naylor. Naylor has earned a reputation for pulling no punches. Previously, he chaired a task force appointed by Harper on health care innovation; the ensuing report so angered the Conservative government that it refused to allow him to hold a press conference to announce his findings. Naylor says he expects the panel to review "the whole ecosystem, without drawing sharp lines between what is basic and applied." The panel, he says, has "no preconceptions about either the level of funding or the optimum architecture." But he personally thinks that the research councils have been buffeted by unreasonable expectations to be both supporters of fundamental research and engines for commercialization of new technologies. The panel is expected to report by the end of this year.
http://scim.ag/Canreview

French panel resigns in protest

PARIS | The entire 20-member panel tasked with reviewing mathematics and informatics proposals for the French National Research Agency (ANR) has resigned in protest over the low success rate for proposals in its discipline. The current success rate for all areas of science, including mathematics, stands at about 11%. Before 2014, however, that rate was higher for disciplines like mathematics, because of ANR's Programme Blanc, a funding program that had tended to reward more theoretical fields. That program was eliminated as ANR established a new funding system in line with national and European ambitions to tackle big societal challenges. The protest, which hit the national press on 9 June, came just as the



Paint-by-microbes: Winners of the 2016 Agar Art contest

The canvas? Agar, a jellylike substance released by certain algae, ubiquitous in biological laboratories to nurture bacteria in a petri dish. The paint? Microbes. Armed with these, and with marching orders to create magic out of the mundane, researchers submitted more than a hundred works of art to the American Society for Microbiology's second annual Agar Art competition. This year's winner, titled *The First Race*, was created by Md Zohorul Islam, a graduate student at the University of Copenhagen. Using four species of bacteria as paint, Islam re-created the competition between spermatozoa to be the first to reach an egg. The second place winner, titled *This is not a beer!*, was created by a team led by microbiologist Mariarosaria Marinaro of the Istituto Superiore di Sanità in Rome. Using methicillin-resistant *Staphylococcus aureas* bacteria—aka MRSA—and a dash of hydrogen peroxide, the team sketched out a "foaming" mug of beer.

panel was to hand over its list of winning proposals to ANR after 3 days of deliberation. The panel is demanding more wiggle room to select winning proposals and transparency in the overall selection process, and has received broad support from the mathematical community in France. Going back to the old system is not on the table, but the French government previously pledged to increase the ANR budget by €65 million starting this month to support more projects across all disciplines. **http://scim.ag/Frpanelresig**

In the path of the monsoon

CHENNAI, INDIA | U.K. and Indian scientists are preparing to embark this month on a joint £8 million project to improve forecasts of the Indian monsoon. The summer monsoon can provide as much as 80% of the annual rainfall to a billion people in India; better forecasts could help farmers plan crop planting and residents prepare for both floods and droughts. While this year's monsoon's weather systems form, the research team for the Bay of Bengal Boundary Layer Experiment plans to deploy seven underwater robots to collect data on ocean conditions; meanwhile, a partner project will fly aircraft to study simultaneous atmospheric conditions. The data will be used to improve computer models of the ocean and atmosphere.

PHOTOS: (TOP TO BOTTOM) MD ZOHORUL ISLAM; AP PHOTO/ARUN SANKAR K

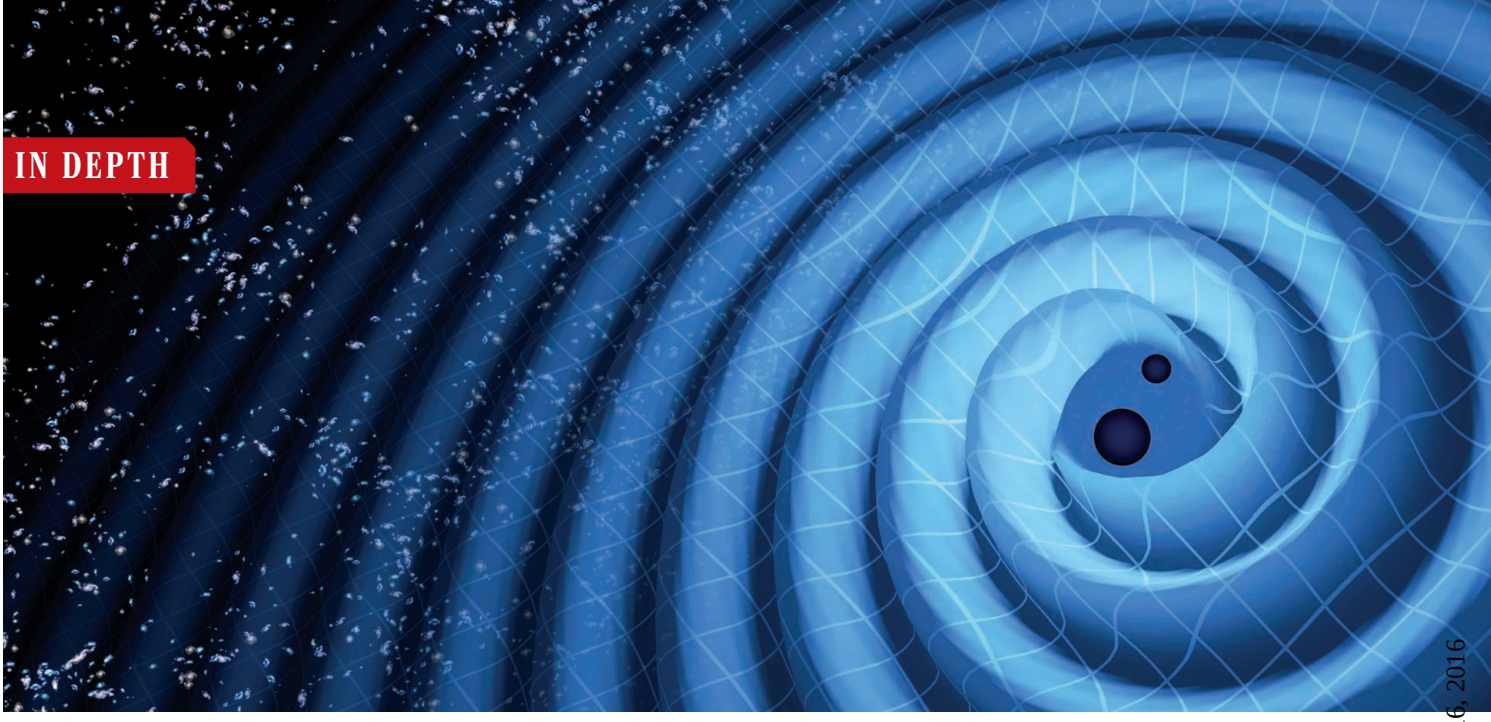
Monsoon-driven floods inundated southern India in late 2015.

NIH to launch grants bias study

BETHESDA, MARYLAND | Do reviewers at the National Institutes of Health (NIH) give lower scores to research proposals from African-American applicants? NIH officials are gearing up for a long-awaited test of that disturbing possibility, which arose after a 2011 report found a significant racial disparity in who receives NIH's bread-and-butter R01 awards. The bias study would draw from a pool of recently rejected grant applications that have been anonymized to remove any hint of the applicant's race, home institution, and training. Reviewers would be asked to score them on a one-to-nine scale using NIH's normal rating system. The exercise is one of three "experimental interventions" to be launched in the coming months that were described last week at a meeting of a high-level NIH advisory council. **http://scim.ag/NIHracebias**

New element names on the table

RESEARCH TRIANGLE PARK, NORTH CAROLINA | Time to throw out that old copy of the periodic table: New names have just been penciled in for four elements officially recognized back in December 2015. Nihonium, moscovium, tennessine, and oganesson will grace the blocks assigned to atomic Nos. 113, 115, 117, and 118, the International Union of Pure and Applied Chemistry (IUPAC) announced last week. Following a 5-month public comment period, the new names will be added to the table. Nihonium, discovered by a Japanese team, means "the land of the rising sun," whereas moscovium and tennessine are named after places near the labs where they were discovered. Oganesson recognizes the work of Russian chemist Yuri Oganessian. By tradition, the right to suggest a name for an element is granted to its discoverer, although IUPAC has the final say. This round of naming seems to have gone off without a hitch, but dramatic battles over element names have been waged in the past, from the hundred-year-long battle over the element now called niobium to the Transfermium Wars of the 1960s.



IN DEPTH

ASTROPHYSICS

LIGO detects another black hole crash

Second gravitational-wave detection augurs bumper crop of sightings

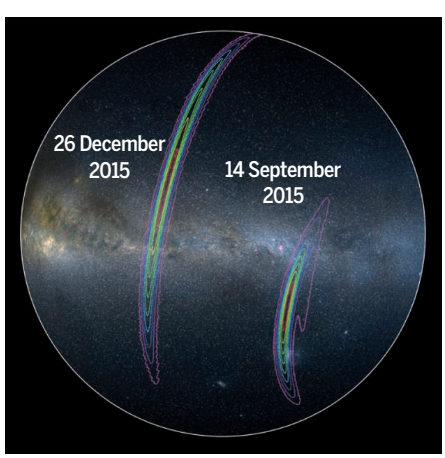
By **Adrian Cho**

The biggest discovery in science this year—the observation of ripples in space-time called gravitational waves—was no fluke. For a second time, physicists working with the two massive detectors in the Laser Interferometer Gravitational-Wave Observatory (LIGO) have detected a pulse of such waves, the LIGO team reported on 15 June at a meeting of the American Astronomical Society in San Diego, California. Once again the waves emanated from the merger of two black holes, the ultraintense gravitational fields left behind when massive stars collapse into infinitesimal points. The new observation suggests that after fine-tuning, LIGO will spot dozens or even hundreds of the otherwise undetectable events each year.

"It's very reassuring," says Gabriela González, a physicist at Louisiana State University, Baton Rouge, and chair of the 1000-member LIGO Scientific Collaboration. "You need another one to be completely convinced and this is it." Cole Miller, an astrophysicist at the University of Maryland, College Park, who is not a member of the collaboration, says that the new find shows that LIGO "is genuinely a new window on the universe." Astrophysicists' catalog of stellar-mass black holes, he says, "is going to be overwhelmingly increased by LIGO."

The new observation came at 3:38.53 Coordinated Universal Time on 26 December 2015—late on Christmas day at LIGO's detectors in Livingston, Louisiana, and Hanford, Washington. As in the first event, the detectors sensed an oscillating stretching of space-time, the signal, according to Einstein's general theory of relativity, of massive objects in violent motion. Computer modeling indicated that its source was two black holes spiraling together about 1.4 billion light-years away. (LIGO researchers had seen a weaker signal on 12 October 2015 that may be a third black hole merger.)

The first signal LIGO spied, recorded in September 2015 and unveiled to the world in February (*Science*, 12 February, p. 645), emanated from surprisingly heavy black holes, with masses 36 and 29 times that of the sun. It lasted only 0.2 seconds, and physicists glimpsed only the last 10 cycles of the black holes' spiraling motion before their collision. The December 2015 sighting involved smaller black holes, 14 and 7.5 times as heavy as the sun. Physicists witnessed 55 cycles of the death spiral, a full second.



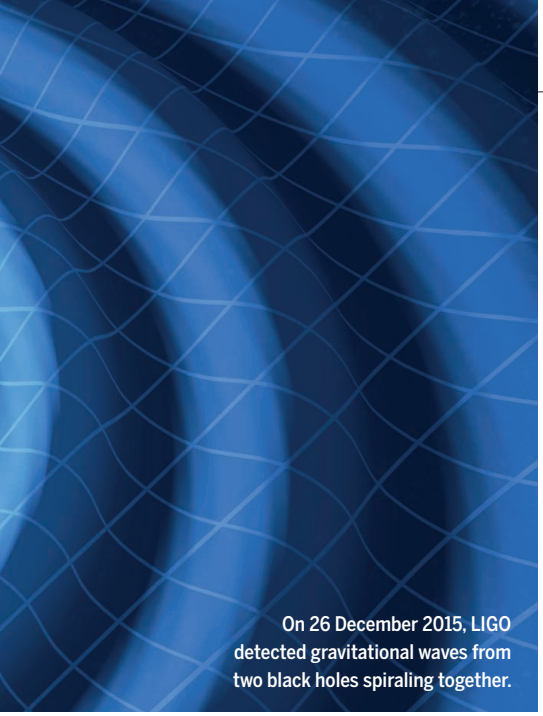
The gravitational waves that LIGO has detected originated in the southern sky. The purple and yellow lines define the likely signal source regions.

The first observation remains a riddle. Those two black holes were twice as massive as the relatively nearby stellar-mass black holes identified from the x-rays emitted by hot gas swirling into them. Astrophysicists don't know how such jumbo black holes formed. The black holes in the new event "are much more garden variety," says Sebastian Heinz, an astrophysicist at the University of Wisconsin, Madison.

Nevertheless, the latest event yields new insights. For example, physicists determined that one of the black holes was spinning frenetically, at roughly 20% of the maximum spin rate allowed by general relativity. And because the new event includes many more cycles, it tests predictions of general relativity somewhat more stringently than the first event, González says. (Einstein's theory passes the test.)

Most important, the second observation shows that going forward, LIGO should reap a vast harvest of black hole mergers. Rebuilt from 2010 to 2015, the new LIGO detectors have not yet reached their design sensitivity. If they do, they should see as many as

IMAGES: (TOP TO BOTTOM) LIGO/T. PYLE; LIGO/AXEL MELLINGER



On 26 December 2015, LIGO detected gravitational waves from two black holes spiraling together.

PHOTO: AP PHOTO/FELIPE DANA

one black hole merger per day, estimates Stephen Fairhurst, a gravitational-wave astrophysicist and LIGO team member at Cardiff University in the United Kingdom. The resulting sample should lay bare the mysterious evolution of black hole binaries, he says, showing whether they start out as pairs of stars that turn into black holes or as black holes that form independently and ultimately find each other.

Reaching the design sensitivity—a factor of 2.5 better than today's—is a challenge, as the instruments currently suffer from a "mystery noise" at low frequencies. David Reitze, LIGO's executive director at the California Institute of Technology in Pasadena, says he's cautiously optimistic that physicists can eliminate the noise and reach design sensitivity by 2019. "I won't say with 100% confidence that we will get there, but I won't say that we won't either," he says.

Physicists hope LIGO will eventually detect waves from other kinds of cosmic collisions. Mergers involving neutron stars, for example, would plumb the mysterious physics of these objects, which are essentially gigantic atomic nuclei with masses between 1.5 and three times that of the sun. LIGO physicists are combing their data for such signals, but finding them is a long shot at the present sensitivity, González says.

LIGO's next big splash will likely come in 2017. The detectors, now restarting after tune-ups, are expected to begin a second data run this fall. They should also be joined by the revamped VIRGO detector, an interferometer near Pisa, Italy, that will help pinpoint sources in the sky and measure their distances. LIGO may well have more surprises in store, Heinz says: "There's always that great possibility of finding something unexpected." ■

INFECTIOUS DISEASE

Experts fear Zika's effects may be even worse than thought

Doctors and researchers are scrambling to define "Zika congenital syndrome"

By **Gretchen Vogel**

Zika keeps unveiling menacing new talents. For many years the virus was thought to cause only mild symptoms. Then, in July 2015 doctors in northeast Brazil noticed a spike in cases of microcephaly, or unusually small heads, showing up in ultrasound examinations of pregnant women. Many of the women reported having had a fever or the rash typical of Zika infection, which had started spreading in the region a few months earlier. In April, the U.S. Centers for Disease Control and Prevention concluded that Zika virus causes this brain defect, which can result in developmental and learning disabilities. Now, case reports make it increasingly clear that babies infected in utero also have problems in their eyes, ears, limbs, and perhaps other organs.

"It seems there are a broader range of effects that go beyond microcephaly," says Tarun Dua, who is helping cooordinate the response to Zika at the World Health Organization (WHO) in Geneva, Switzerland. Earlier this month WHO announced that it is launching an effort to characterize what doctors are now calling Zika congenital syndrome. "We are looking at many thousands of babies, and we need a systematic process for defining their symptoms," Dua says.

Animal- and cell-based studies have shown that Zika virus readily infects the cells of the nervous system, which would explain most of the symptoms identified so far. "We've been learning very bad things on a weekly basis," says Paolo Zanotto, a virologist at the University of São Paulo in Brazil. He notes that the damage is so extensive in some cases that the skull has filled mostly with fluid, and "there is no brain." Those babies die shortly after birth. Other infants may have only mild brain damage that will not be apparent until learning disabilities or other problems show up years later.

Doctors focusing on microcephaly may miss significant problems. On 7 June, doctors in Recife, Brazil, reported in *The Lancet* that they had diagnosed brain damage and scars on the retina of a 2-month-old baby that did not have microcephaly but did have limb spasms shortly after it was born. A brain scan revealed significant defects, including ventricles that were larger than normal and brain tissue that was smoother than it should be. The mother reported no symptoms of Zika infection during pregnancy. However, further tests showed that the baby had antibodies to the virus in its cerebrospinal fluid, strong evi-

A physician in Recife, Brazil, examines the brain scans of a baby born with microcephaly.

dence that it was infected before birth.

Several cases suggest that the damage may go well beyond the nervous system, says Pilar Ramón-Pardo, coordinator for Zika emergency response at the Pan American Health Organization in Washington, D.C. Doctors in Panama and Colombia are reporting that a handful of mothers who had Zika symptoms during pregnancy have given birth to children with problems in the heart, digestive system, and genitals, Ramón-Pardo says. "We need to investigate whether those cases are related to Zika" or have some other cause.

In July, WHO will gather about 50 clinicians and researchers in Recife to draw up a preliminary definition of Zika congenital syndrome. The participants will also develop guidelines for the tests that doctors should use, both at birth and during several years of follow-up, to try to capture the full spectrum of problems the virus might cause, Ramón-Pardo says. The work should help countries better understand what kinds of treatment and support affected children and their families will need, Dua says.

One continuing challenge is the difficulty of diagnosing Zika itself. Symptoms are often mild, or even absent, and a blood test is accurate for only a few days after infection. Previous infections with dengue virus, very common across Latin America, can cloud results on antibody tests, the standard for detecting past infections in people who have already recovered. And many symptoms can have multiple causes. For example, microcephaly can be caused by genetic disorders, alcohol or drug exposure, and infections such as cytomegalovirus, toxoplasmosis, and syphilis.

To try to get a clearer picture of the risk for pregnant women and their babies, researchers have started several prospective studies, enrolling thousands of women in areas where Zika is circulating. The researchers plan to test the women regularly for infection and then follow them and their babies for at least 2 years. Such studies "will provide some information, but it is going to take some time," Dua says.

In the meantime, there is a glimmer of good news from the Recife region. A few months ago, Manoel Sarno, a fetal medicine doctor at the Federal University of Bahia in Salvador, says he saw as many as 10 cases of microcephaly per week. He now sees only about one a month. That may only reflect a seasonal drop in Zika infections during last year's dry season, but there are other signs that the outbreak in the region is receding. It's possible, he says, that so many people were infected last year that the population's "herd immunity" is high enough to stop the virus's spread—and make Zika congenital syndrome a rare occurrence in the region. ■

EUROPE

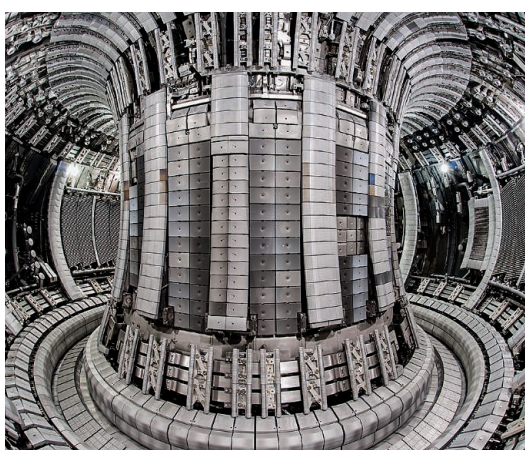
Brexit casts pall on fusion

ITER backers nervously await vote that could undermine U.K. research ties with continental Europe

By **Daniel Clery**, *in Cambridge, U.K.*

If the United Kingdom votes to leave the European Union on 23 June, the exit will break up cross-border collaborations and cut off E.U. funding for U.K. scientists. For fusion research, the possibility of a Brexit is particularly worrying. Europe's largest fusion facility, the Joint European Torus (JET), is sited just south of Oxford, U.K.; a vote to leave would put it in a legal limbo that could halt vital research supporting the International Thermonuclear Experimental Reactor (ITER), now under construction in southern France. JET dominates much of the work at its host institution, the Culham Centre for Fusion Energy. A Brexit "certainly will make us very vulnerable," says Steve Cowley, the center's director.

Polls suggest the referendum's outcome is too close to call, and a vote for a Brexit would not affect some pan-European research facilities because the European Union does not control them. These include the CERN particle physics laboratory in Geneva, Switzerland; the Paris-based European Space Agency, which has a technology center at Culham; the European Southern Observatory in Garching, Germany; and the European Molecular Biology Laboratory, which has a bioinformatics institute in Cambridge, U.K.



A U.K. departure from the European Union could halt vital research at the Joint European Torus, a fusion facility in England.

Fusion is different. The nuclear arm of the European Union, known as Euratom, pays a consortium of national labs and university groups—dubbed EUROfusion—to carry out fusion research, most of which is aimed at supporting ITER. EUROfusion will receive €424 million for the work over 5 years (2014–18), and another €283 million will go to the Culham Centre for operating JET and hosting ITER-related research.

JET is the world's largest tokamak—a doughnut-shaped fusion reactor like ITER. Its innards have been coated with the same beryllium and tungsten that will line ITER, and its heating power has been boosted to make it as ITER-like as possible. JET is also the only tokamak in the world equipped to use the fuel that will eventually power ITER: a mix of the hydrogen isotopes deuterium and tritium. Tritium is radioactive and hard to handle, but this mixture is the easiest to "burn," meaning fusing its nuclei can release excess energy. JET researchers are currently testing how the machine behaves when filled with hydrogen, deuterium, and tritium individually before attempting deuterium-tritium burns in 2019.

A Brexit could halt those experiments. "We would be less well prepared for ITER operation," says Lorne Horton, JET's exploitation manager. "From a scientific point of view, I hope it doesn't happen." Tony Donné, head of EUROfusion in Garching, thinks that after a "leave" vote, the United Kingdom would operate the machine at least until the end of the current contract in 2018. But what might happen after that "is really crystal ball–gazing." The United Kingdom could opt to buy into the Euratom research program, as some non-E.U. nations like Switzerland have done, but that process could take years and is likely far down the list of negotiations that the U.K. government will need to complete following a Brexit vote.

Donné says EUROfusion has been working on plans to turn JET into a fully international facility so that the other ITER partners—China, India, Japan, Russia, South Korea, and the United States—can also use it for ITER preparation. "If it flies, this would be a fantastic project," Donné says. By cutting ties to Euratom, a Brexit could scuttle that plan, too. ■

PHOTO: © EUROFUSION

Summer melting turns Greenland into a land of lakes.

CLIMATE

Sea ice retreat said to accelerate Greenland melting

New claim intensifies "Arctic amplification" debate

By **Eli Kintisch**

As Arctic temperatures have crept upward over the last 3 decades, prodigious surface melting of the Greenland Ice Sheet has removed hundreds of billions of tons of ice. But last summer a peculiar pattern of melting captivated Marco Tedesco, a climate scientist at Columbia University. Although the northern parts of the ice sheet experienced record melting as summer temperatures rose as high as 19°C (66°F), southern Greenland was unusually cold, with just average melting. To Tedesco, the contrast suggested that the Greenland melt was being shaped by another Arctic phenomenon: the retreat of floating sea ice across the Arctic Ocean.

Last week Tedesco published a paper that blamed the melt pattern on a 500-kilometer-wide, high-pressure vortex, known as a block, that loitered north of the island during June and July 2015, wreaking weather havoc. "How the atmosphere is behaving is driving new records of melting on Greenland," he says. Such atmospheric blocks are just what you'd expect in a world of retreating sea ice, Tedesco says—a claim that has added fuel to a contentious dispute over the global influence of the warming Arctic.

Until now, that argument has focused on how disappearing sea ice might be favoring extreme midlatitude weather, such as floods in Texas or heat waves in Russia. Tedesco's study, in *Nature Communications*, is one of a trio of recent papers that have expanded the scope of the debate. They suggest that "continued human-induced loss of Arctic sea ice is likely to contribute to additional melt of the Greenland Ice Sheet," says meteorologist Jennifer Francis of Rutgers University, New Brunswick, in New Jersey, one of the first to propose that Arctic warming could be transforming weather at midlatitudes.

There's little doubt that sea ice retreat has helped speed Arctic warming: The melting has revealed dark, heat-absorbing water and released prodigious water vapor, a strong greenhouse gas. What some researchers question is the proposal that the disproportionate warming, known as Arctic amplification, has affected the jet stream, the high-altitude river of air that circles the Northern Hemisphere, separating cold air from warm. The jet stream is propelled by the temperature contrast between the poles and the midlatitudes, and Francis and others think that as Arctic amplification has diminished that contrast, the jet stream has grown slower, wavier, and more prone to allowing blocks of anomalously cold or warm air to form over the midlatitudes.

Longer lasting weather events in the Arctic, like last summer's block over Greenland, are another effect, Tedesco says. In June 2015 the jet stream was very wavy, meandering as far as 76° north, the farthest since 1948. The vortex it spun off swirled clockwise, pulling warm air from the south along the island's western side while cold air from the Arctic flooded into southern Greenland.

Tedesco also showed that over the last 6 decades, the northern extent of the jet stream has ventured an average of 3° farther north. The trend has led to more blocks, climatologist Edward Hanna of the University of Sheffield in the United Kingdom concluded in a January paper in the *International Journal of Climatology*. Hanna and colleagues calculated that seven of the top 11 summer blocking events over Greenland since 1851 have occurred since 2007, just as Arctic amplification has ramped up. June 2012 was the worst of any June, they calculated. The next month, 90% of the ice sheet's surface melted for the first time in more than a century.

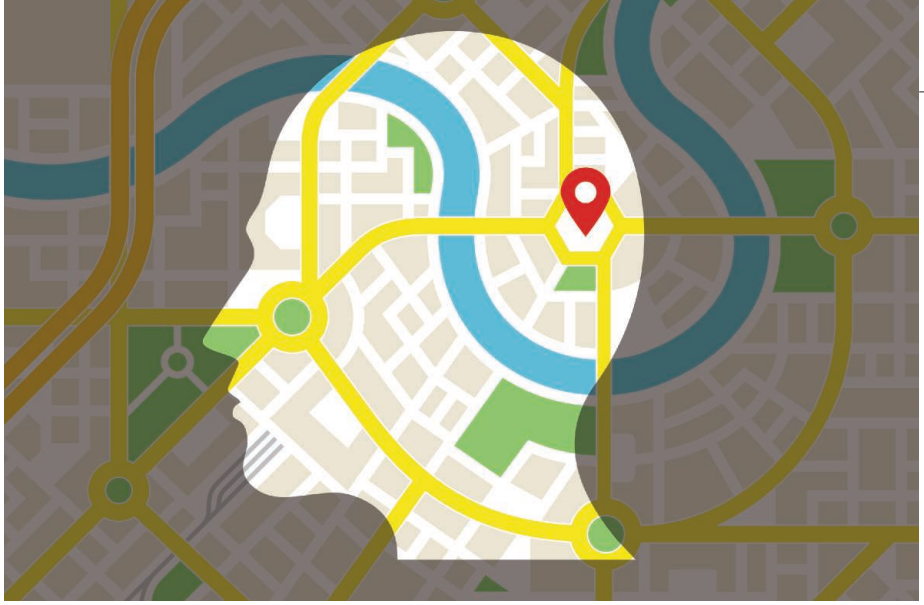
A third recent paper, published in the *Journal of Climate* in January by Jiping Liu of the State University of New York at Albany, used climate models to compare the effects of low and average sea ice. Modeling runs with low ice saw more blocks over Greenland, warmer temperatures, and wind patterns that resemble recent conditions.

Correlation is not causation, warns James Booth, a climatologist at the City College of New York in New York City. Booth was a co-author on Tedesco's paper, but the two differ on how strongly the paper supports the Arctic amplification hypothesis. "What's clear is that when there's the jet stream waviness, you can have effects like blocking. But the waviness could be caused by something other than Arctic amplification," Booth says.

Kevin Trenberth of the National Center for Atmospheric Research in Boulder, Colorado, for example, emphasizes variations in the flow of energy out of the mighty tropical Pacific, a key source of the heat that spills into the Arctic. "I don't buy the mainly local forcing of blocking [in the Arctic]," he says. Other researchers have linked Greenland's warming to seesaws of the North Atlantic Oscillation, a fluctuating pattern of atmospheric pressure that influences weather across the Northern Hemisphere.

Meanwhile, the argument over the global role of the Arctic rages on. In a webinar next month, climate scientist Judah Cohen of the Massachusetts Institute of Technology in Cambridge will debate the topic with Lantao Sun of the University of Colorado, Boulder. Cohen sits squarely in the Arctic camp. "Not only is the Arctic comparable to the tropics in global influence, it's much greater," he says. "It's heresy, I know." ■

PHOTO: © JUSTINE EVANS/ALAMY STOCK PHOTO



Neurons called grid cells help make mental maps of physical locations, but a study in which people watch bird silhouettes morph (below) suggests similar brain cells fire when we explore conceptual spaces.

NEUROSCIENCE

Expanding our mental maps

Grid cells that represent physical space in the human brain may also organize more conceptual forms of knowledge

By **Emily Underwood**

The brain is a mapmaker. As you navigate a landscape, neurons in a region called the entorhinal cortex fire at multiple locations, marking out a hexagonal grid on a mental map. The discovery of these so-called grid cells, and their role as a neuronal GPS for spatial navigation, won the 2014 Nobel Prize in Physiology or Medicine for Norwegian scientists Edvard Moser and May-Britt Moser. Now, it seems that the brain may make maps of abstract realms, too.

On p. 1464, a team at the University of Oxford in the United Kingdom provides evidence that gridlike neuronal activity throughout the brain helps people organize nonnavigation knowledge—for the purposes of the new study, differences in body shape between various types of birds. The work "points out the ways in which the spatial system is actually quite useful for something else," says Lynn Nadel, a cognitive scientist at the University of Arizona in Tucson.

The Mosers first saw grid cells at work in lab rats exploring an open area, and others later confirmed the finding in humans—first by using implanted electrodes in people with epilepsy and then by using functional magnetic resonance imaging (fMRI) to noninvasively watch brain activity in people playing video games. Tim Behrens, an Oxford neuroscientist, wondered whether grid cells worked in other realms, so he recently asked a Ph.D. student, Alexandra Constantinescu, to investigate. The ancient memory "palace" technique, in which people remember facts by arranging them in an imaginary structure, was one "inspiration" for the study, Constantinescu says.

She chose birds as the basis of the experiment in part because she has two pet finches. More important, a bird's 2D silhouette can be morphed into many different, discrete shapes, creating a conceptual "space" for the brain to navigate, she says. Simply by stretching or shortening a bird's neck and legs along the x and y axes, for example, she can make the silhouette resemble birds ranging from swans to ducks to storks. If people showed grid cell–like brain activity on fRMI scans while exploring this 2D "bird space," the Oxford team reasoned, the same organizational brain system used in navigating real space could be in play.

The team recruited 28 people to play a computer game in which each morphed a bird silhouette using buttons on a keyboard. To provide memorable landmarks in bird space, specific Christmas-themed icons—think gingerbread men and teddy bears—popped up in association with certain leg-to-neck ratios. After several hours of training, participants watched a video of the same bird morphing in different ways while researchers monitored their brain activity in an fMRI scanner. Every time the video paused, participants were asked to imagine which Christmas icon would pop up if the bird continued to stretch with the same neck-to-leg ratio. (Imagining the outcome was important to keep participants actively engaged in the task, Constantinescu says.)

An fMRI machine's resolution is too coarse to detect the hexagonal firing patterns of individual grid cells, but previous studies have shown that the aggregate activity of such neurons produces an fMRI signature that exhibits some hexagonal symmetry. With this as a proxy, Behrens's team found grid cell–like activity that reliably varied as participants changed directions in 2D "bird space," depending on how much the birds' necks or legs had stretched. Seeing such a robust signal in fMRI is "very unusual," Constantinescu says. "You're unlikely to see anything like this by chance."

The crisper and easier to distinguish someone's grid cell–like signal was, the better they were at remembering which bird shape corresponded with which Christmas icon, Constantinescu says. The signal was located not only in the entorhinal cortex, but throughout a network of brain regions associated with conceptual thinking and known to use a grid code when people navigate through actual spaces.

Scientists once thought grid cells operated only in the entorhinal cortex, because of their specific role in spatial processing, says Howard Eichenbaum, a neuroscientist at Boston University. "All of a sudden you see them in a handful of cortical regions, and you think, what are these cells really doing?" Recently, he notes, researchers have discovered hints that similar brain maps may exist for other types of information, such as social relationships, color, sound, and even time.

Researchers need to directly record large numbers of neurons in people using electrodes to confirm that grid cells are really at work for more than spatial navigation, but the study is "extremely clever," says May-Britt Moser at the Norwegian University of Science and Technology, Trondheim. To neuroscientist György Buzsáki at New York University in New York City, it makes "perfect sense" that the same mechanism that the brain evolved to navigate space has been co-opted over time to make sense of more abstract realms. "To me these things are identical," Buzsáki says. In one system, "one is navigating in the real world," whereas in others, "one is navigating in the mental world." ■

IMAGES: (TOP TO BOTTOM) G. GRULLON/*SCIENCE*; A. CONSTANTINESCU *ET AL.* *SCIENCE*

FUNDING

A peek at peer review helps young scientists

Service on NIH study sections boosts success rates on grants but not diversity

By **Jeffrey Mervis**

Winning a grant from the National Institutes of Health (NIH) is hard, especially if it's your first. Sitting in judgment of other grant applicants before applying for your own can help, new data from a pilot project called the Early Career Reviewer (ECR) Program suggest. Yet the program may not be meeting another key goal: helping close a stark racial disparity in success on applications for NIH grants.

NIH created the ECR Program after a disturbing 2011 report found that from 2000 to 2006, African-American applicants were 10% less likely than whites to receive an R01 grant, NIH's bread-and-butter funding mechanism for academic researchers (*Science*, 19 August 2011, p. 925). The disparity persists, NIH officials told a top-level advisory council last week: A new in-house analysis of a more recent cohort of 1054 matched pairs of applicants found that African-American grant seekers were 35% less likely to be successful than whites.

Observers say one explanation could be that minority applicants are unfamiliar with how NIH study sections work. That possibility motivated officials to create the ECR Program, which allows young scientists to attend one meeting of a study section in their field of expertise. (Regular members serve 4-year terms on one of the 150 study sections, which meet three times a year.) They rate up to four proposals and join in the discussion of all applications.

Being part of the review process can be an eye-opener, senior scientists say. "One of the best education experiences for me as a junior investigator learning how to write a successful grant was being on a study section," notes advisory council member Deborah Griffith, a biomedical engineer at the Massachusetts Institute of Technology in Cambridge.

New data presented to the advisory panel by Richard Nakamura, director of NIH's Center for Scientific Review in Bethesda, Maryland, which runs the program and oversees the peer-review process, support that idea. He said ECR alumni have been more than twice as successful as the typical new investigator in winning an R01 grant. But Nakamura adds an important caveat: The pool of successful ECR alumni includes those who revised their application, sometimes more than once, after getting feedback on a declined proposal. That extra step greatly improves the odds of winning a grant. In contrast, the researchers in the comparison group hadn't gone through the resubmission process.

It's also hard to interpret the fact that 18% of the successful ECRs were underrepresented minorities because NIH did not report the fraction of minorities among ECR alumni applicants. So it is not clear whether African-Americans participating in the program did any better than the cohort as a whole—suggesting that the program might begin to close the racial gap—or better than a comparable group of minority scientists who were not ECR alumni.

The program has attracted many fewer underrepresented minorities than diversity advocates had hoped when it was launched. Minorities comprise only 13% of the roughly 5100 researchers accepted into the program (6% African-American and 7% Hispanic), a percentage that roughly matches their current representation on study sections.

Nakamura says NIH has had a hard time finding minority candidates who meet the eligibility requirements—being a senior author on two recent peer-reviewed papers and a full-time faculty member for 2 years—and who want to participate. One-quarter of researchers in ECR's first cohort were from minority groups, he notes. "But as we've gone along, there are fewer underrepresented minorities coming into the pool."

Meanwhile, interest remains high among white and Asian scientists, who account for roughly 80% of the ECRs. To cope with the high demand for limited slots—each study section is allowed only one ECR scientist—NIH decided last year to exclude scientists who had already received an R01.

Serving on a study section "was a great experience," says Blake Riggs, an associate professor of cell biology at San Francisco State University in California and one of the first scientists to be chosen for the program. But Riggs, the only African-American in his department, thinks the program is too brief to help minority scientists truly become part of the mainstream, and may even exacerbate their sense of being marginalized.

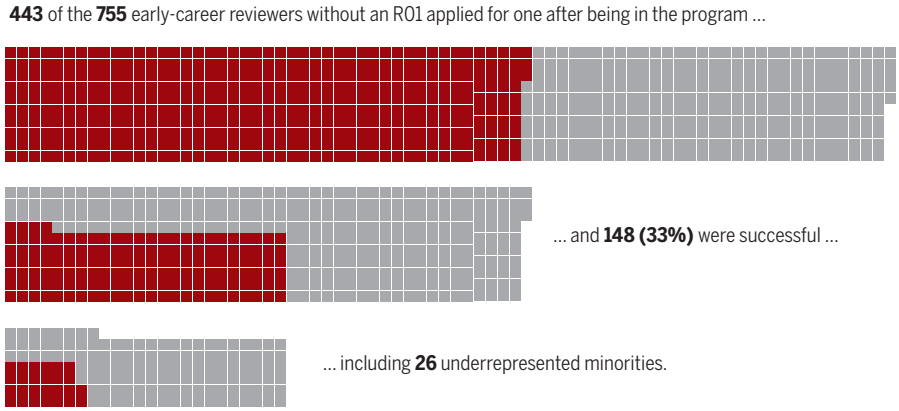
"After I sat on the panel, I realized there was a real network that exists, and I wasn't part of that network," he says. "My comments as a reviewer weren't taken as seriously. And the people who serve on these panels get really nervous about having people … that they don't know, or who they think are not qualified, or who are not part of the establishment."

If NIH "wants this to be real," Riggs suggests having early-career researchers "serve as an ECR and then call them back in 2 years and have them serve a full cycle. I would have loved to do that."

But Nakamura worries that asking minority scientists to play a bigger role in NIH's grantsmaking process could distract them from building up their lab, finding stable funding, and earning tenure. Serving on a study section, he says, means that "those individuals will have less time to write applications. So we need to strike the right balance." ■

Gaining an edge

Preliminary data show that young scientists who have sat in on a National Institutes of Health study section do better than their peers in subsequent competitions for grants.



CREDITS: (DATA) CENTER FOR SCIENTIFIC REVIEW, NIH; (GRAPHIC) A. CUADRA/*SCIENCE*

POLITICAL SCIENCE

How to attack the Islamic State online

Study of social media sites suggests women are key recruiters for terrorist group

By **John Bohannon**

For 6 months last year, Yulia Vorobyeva tracked terrorists and their sympathizers online. A Ph.D. student in political science at the University of Miami in Coral Gables, Florida, she daily browsed the web pages of social media groups that support the terrorist group that calls itself the Islamic State (IS). "It was very gruesome," she says. "Blood, decapitated and dismembered bodies. Everything." The worst were the images of suffering children, which the group blames on its enemies and exploits to win converts. "I am a mother myself," Vorobyeva says.

Vorobyeva's harrowing exercise in data-gathering has helped her understand how IS wages an online war of propaganda. In a pair of studies, one published last week in *Science Advances* and the other on p. 1459, she and her colleagues reveal in fine detail the growth and evolution of the online networks that supply the terrorist group with converts and support. One surprise is that women play a key role, possibly more important than men, in the group's recruitment and propaganda machine.

The team's analysis of how pro-IS groups expand on social media also suggests a strategy for thwarting it, says Karl Rethemeyer, a political scientist at the State University of New York at Albany, who was not involved in the new study. "Disruption of these [online] communities affects the spread of ideas and information," and could stymie a group's recruitment of new followers, he says.

The studies were led by physicist Neil Johnson of the University of Miami, who models human organization and conflict. "Up to now, it [was] assumed that support for IS arises from online contagion," he says, with information and beliefs spreading from person to person. That has led to a bumper crop of studies focusing on the millions of individual terror-related messages on the internet. Yet Johnson had a hunch that what matters is not individual people but aggregates of them connected through pro-IS social media groups. "Find the aggregates," he says, "and you have your hand on the pulse of the entire organization."

Easier said than done. Social media sites such as Facebook and Twitter have cracked down on hate speech and terrorist propaganda. But pro-IS online activity is alive and well on Vkontakte, a Russia-based social media site with 360 million users around the world. So between January and August last year, Johnson's team, including Vorobyeva, meticulously explored the site. They uncovered 196 pro-IS groups with a total of more than 100,000 members.

Given the often harsh treatment of women endorsed by the terrorist organization, the team was startled to find that about 40% of participants declare they are female. To find out how important a role those women are playing, Johnson's team created a network graph based on the first 2 months of data. If any two people were a member of the same pro-IS group, they were considered to be connected. The result was a network with about 1 million links. Using statistical tools from an area of mathematics known as graph theory, the team then measured how information and resources flow through the network.

A key measurement of importance for people in a network is something called their betweenness centrality (BC). Those with high BC serve as key points of communication that connect far-flung parts of the network. The women had BC scores twice as high as those of men, on average, the team reports in *Science Advances*. That is striking, Johnson says. "In typical everyday situations, such as corporate settings, the BC ratio of women-to-men is known to be less than one." But when it comes to promoting IS terrorism, he says, "a woman will on average be a conduit for twice as many pieces of information."

That suggests that winning the hearts and minds of women may be an effective countermeasure, Rethemeyer says. "Counterterrorism authorities may want to use gender" as a shortcut for identifying people important in the IS propaganda network.

The second study by Johnson's team tracked the growth of pro-IS groups over the 6-month period, capturing the process with a mathematical model. Members of small groups, which may only have dozens of members, merge into larger ones in what seems to be a predictable "coalescence" process. Monitoring the size and number of the groups supporting terrorism could give an early warning that a global movement is emerging, Johnson says.

But how do you prevent such online movements from gathering momentum in the first place? "We find mathematically that you can wipe out all large [groups] simply by eliminating those of 10% their size," he says. Johnson notes that proactively disrupting such small hubs isn't risk-free. "[There is the] cost of making mistakes and hence being seen as censoring free speech," he notes.

Jacob Shapiro, a political scientist at Princeton University and a former U.S. Navy officer, notes that there aren't many options for attacking social media hubs: Either legally force the social media company to shut down a group, or covertly infiltrate and spread misinformation within the group to break down trust. Both are time-consuming. Johnson's model could help counterterrorism agencies "calibrate how much effort you put to shutting these things down," Shapiro says.

At this point, Shapiro adds, the data set itself may be the most useful yield from the research. It could help train computer algorithms to automatically scrape online networks night and day for signs of terrorist recruitment and attack planning. That would save researchers like Vorobyeva from this gruesome task. ■

An image from a training video made by the terrorist group calling itself the Islamic State. Such videos are posted on social media sites as recruiting tools.

PHOTO: BALKIS PRESS/SIPA USA/AP IMAGES

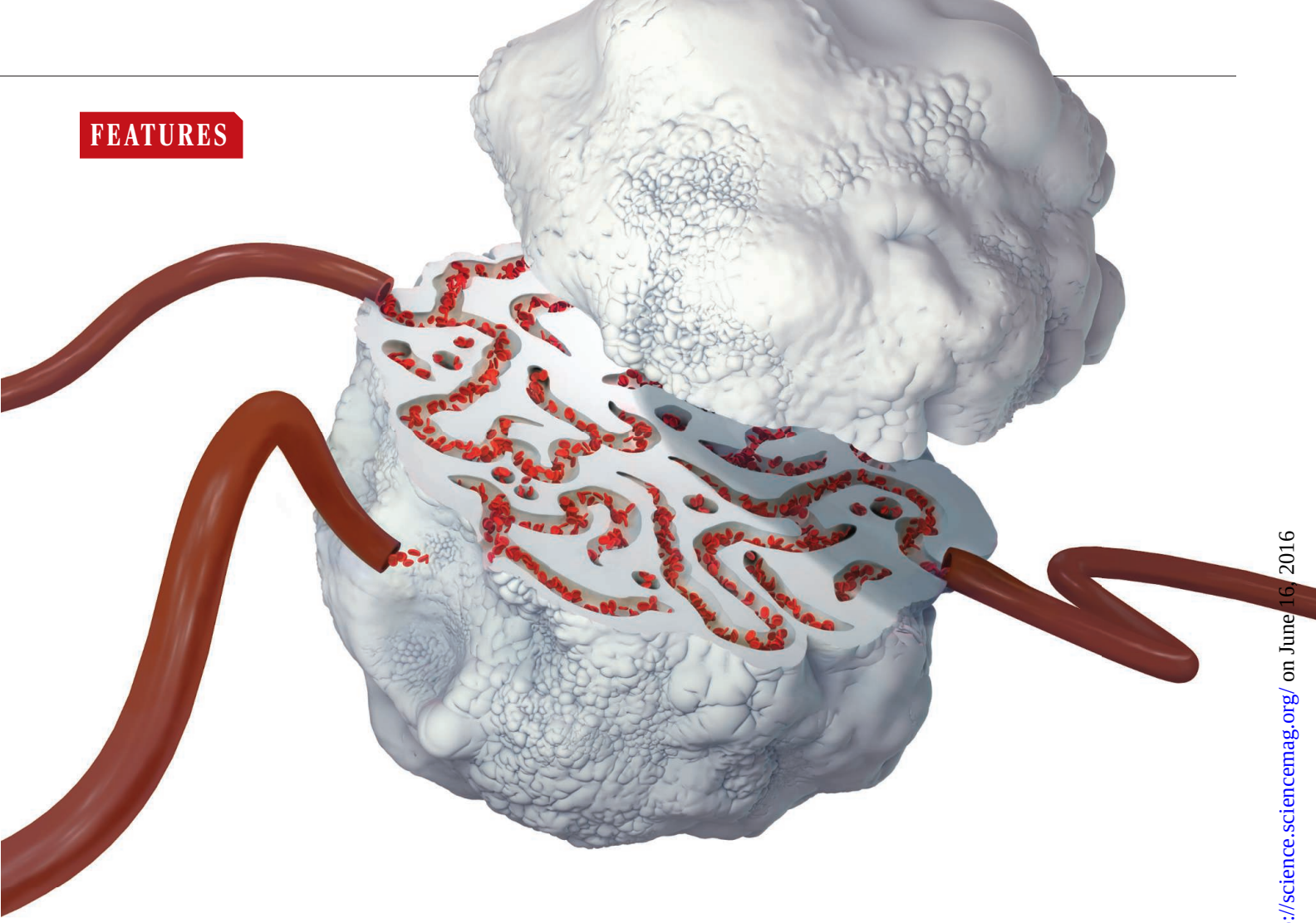


ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

TUMORS' DO-IT-YOURSELF BLOOD VESSELS

The unusual supply lines built by cancer cells may explain why some treatments fail and offer new targets for drugs

By **Mitch Leslie**

On a cool day in March 2000, several hundred researchers jammed into a hotel auditorium in Salt Lake City, eager to see a showdown over what had become one of the most controversial ideas in cancer research. On one side stood cancer biologist Mary Hendrix of the University of Iowa Cancer Center in Iowa City, whose team the year before had reported an unusual, seemingly new way through which tumor cells can tap into the blood supply and obtain nutrients. Facing off against her was tumor vascular biologist Donald McDonald of the University of California, San Francisco, who was certain that she and her colleagues had misinterpreted their data. "This debate had the feeling of a boxing match—with the championship belt hanging in the balance," Hendrix recalls.

Researchers knew at the time that tumors can induce the endothelial cells of normal blood vessels to form new supply lines into a tumor, a process called angiogenesis. But Hendrix and her colleagues contended that tumor cells themselves sometimes create their own blood-delivering tubes, a mecha-

nism they dubbed vasculogenic mimicry (also known as vascular mimicry). Their 1999 paper "started lots of upheaval," says histopathologist Francesco Pezzella of the University of Oxford in the United Kingdom. The Utah debate, held at a Keystone meeting, was the first public discussion of the concept.

In the end, neither side scored a knockout. Hendrix asserted that the loops and networks her team had observed represented a mini–circulatory system produced by the tumors themselves. McDonald countered that the patterns were folds of connective tissue, not tubes that carried blood. In the years since, the controversy has waned, and Hendrix and other researchers have pieced together a picture of how tumors build their own blood vessels and how they can affect prognosis and treatment. But some scientists continue to find the idea deeply unsettling.

Now, vasculogenic mimicry faces another big test. The first clinical trial of a drug to block the process—and thus potentially limit tumor growth—has begun in the United States and Taiwan. If the drug succeeds, it would bolster what Hendrix and other researchers have been saying about these do-it-yourself blood vessels for nearly 17 years. And it might also explain why some of the most hyped drugs in cancer therapy—angiogenesis inhibitors—have underperformed.

VASCULOGENIC MIMICRY roiled the cancer field because it undermined the leading idea for how tumors obtain their blood supply. In the early 1970s, Judah Folkman of Harvard Medical School in Boston proposed that tumors can grow large because they trigger angiogenesis, inducing new blood vessels that speed nutrients and oxygen to fast-dividing cancer cells. Halting the growth of these vessels into tumors, he suggested, would starve the masses. Folkman himself famously struggled to convince the many angiogenesis skeptics among cancer researchers. But by the late 1990s, drug companies were bustling to develop compounds that curtail angiogenesis, and DNA pioneer James Watson announced that Folkman's approach would "cure cancer in 2 years."

Yet if tumors have an alternative way to secure the blood they require, Hendrix and colleagues argued, antiangiogenic compounds might fail. "That put a big target on us," she says. Later clinical trials of angiogenesis inhibitors have confirmed their doubts, however. Although several of the drugs have received U.S. approval for use in cancer patients, including Avastin and Nexavar, they only temporarily slow tumor growth; the tumors often become resistant. Pezzella says the evidence now suggests that Hendrix was right. "Vasculogenic mimicry is one of the ways in which tumors develop a blood supply independently from classical angiogenesis."

Two ways to feed a tumor

Vasculogenic mimicry, in which cancer cells themselves form blood-carrying channels, may supplement traditional angiogenesis in supplying nutrients to tumors.

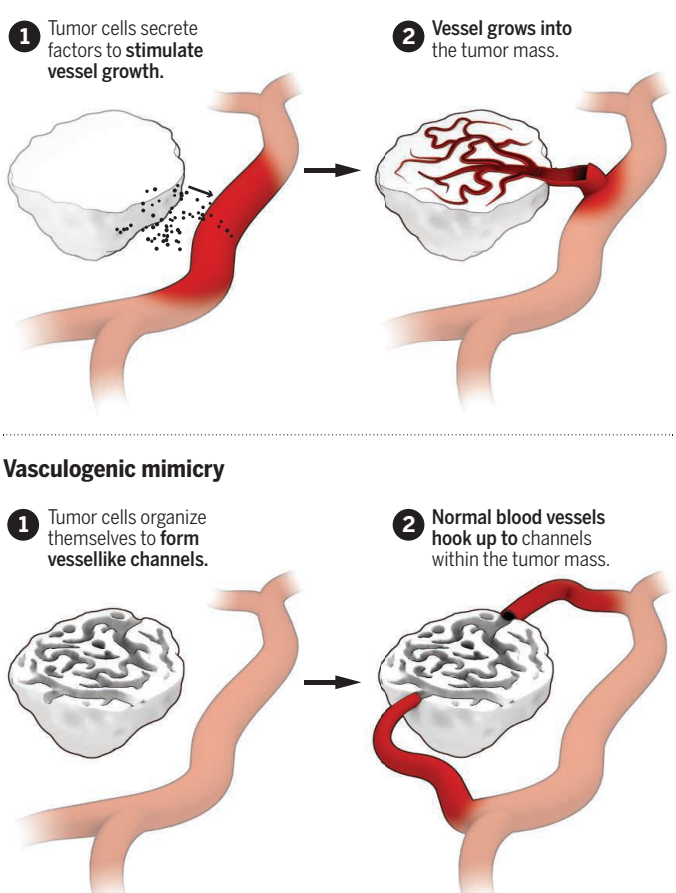


DIAGRAM: V. ALTOUNIAN/*SCIENCE*

Hendrix and her colleagues, including University of Iowa pathologist Robert Folberg, didn't set out to deflate the excitement over angiogenesis. They were studying why some melanomas are virulent, spreading rapidly and often killing patients, whereas others are less dangerous. In one revealing experiment, they planted human melanoma cells on a gel that mimics the extracellular matrix, the fibrous material that surrounds cells. Aggressive tumor cells "migrate through the matrix and scrunch it up," Hendrix says, resulting in networks of channels. "At low [microscope] power, they look like chicken-wire meshwork," she says.

Folberg had been seeing similar patterns in melanomas that grew in patients' eyes. "We naïvely assumed they were blood vessels," says Folberg, dean of the Oakland University William Beaumont School of Medicine in Rochester, Michigan.

Taking a closer look at eye melanomas, the researchers noticed another similarity to blood vessels: Some of the channels contained red blood cells. But to the scientists' surprise, the networks lacked the endothelial cells that line normal blood vessels, suggesting that the cancers themselves formed the channels. In a commentary on their paper, which appeared in 1999 in *The American Journal of Pathology*, noted cell biologist Mina Bissell of the Lawrence Berkeley National Laboratory in Berkeley, California, called the clinical implications of the findings "far-reaching."

Yet in a rebuttal published a few months later in the same journal, McDonald and two other researchers knocked the paper on several grounds, such as failing to demonstrate that blood flowed through the networks. The blood cells inside the channels could have leaked from conventional tumor vessels, McDonald and his co-authors proposed. They deemed the evidence for vasculogenic mimicry "neither persuasive nor novel."

Hendrix and colleagues acknowledge that other researchers had previously suggested the phenomenon. She also admits that during those early years, "we did not have all the answers." But she says that the case for vasculogenic mimicry has become much stronger. For one thing, researchers have identified the tubes in many more tumor varieties, including breast, prostate, kidney, lung, and bone cancers. Moreover, growing circumstantial evidence suggests that when they appear, a cancer patient's odds of survival plummet. A 2016 meta-analysis by researchers in China, which combined results from 36 studies on more than a dozen cancer types, estimated that patients' chances of dying were roughly doubled if their tumors showed evidence of vasculogenic mimicry.

Researchers have also furnished evidence

that the channels do transport blood. In a 2008 study, for instance, Folberg and colleagues injected a fluorescent dye into the arms of patients who had melanomas in their eyes. Within 30 seconds, the dye had traveled through the patients' circulatory systems to their eyes and had appeared in the channels in their tumors.

McDonald, however, remains skeptical. Other researchers, including vascular biologist Drew Dudley of the University of North Carolina, Chapel Hill, accept that vasculogenic mimicry can occur but want further evidence of its relevance. "It isn't clear whether this occurs substantially in human patients with cancer and whether this has anything to do with antiangiogenic therapies not working very well," Dudley says.

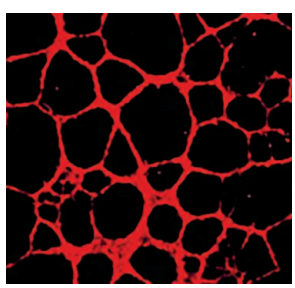
"If we shook up the field enough to get people to think about a new approach, I'll be happy."

Mary Hendrix, Shepherd University

HENDRIX AND OTHER scientists are pushing ahead to learn how vasculogenic mimicry might work. They're still not sure how the tumor-spawned networks hook up to the normal circulation, but they have determined that as tumors build blood vessels from their cells, the cells switch on many of the same genes that normally define endothelial cells. For example, in normal blood vessels, clots could form and cause blockages if endothelial cells didn't release anticoagulant compounds. "That's a real problem that cancer cells [forming their own blood vessels] would have to solve," Dudley says. They apparently have. Hendrix and colleagues discovered that cancer cells involved in vasculogenic mimicry release some of the same anticlotting molecules endothelial cells do.

Only certain cells in tumors seem to have the ability to produce the blood-transporting channels, and they may overlap with so-called cancer stem cells, rare cells in tumors thought by many researchers to fuel the overall growth of the cancerous masses. In melanomas, for instance, skin pathologist George Murphy of Brigham and Women's Hospital in Boston and colleagues reported that the cells that are capable of vasculogenic mimicry show cancer stem cell characteristics such as chemotherapy drug resistance and the ability to specialize into different cell types. "There appears to be a subpopulation of cells within a cancer that are very smart," Murphy says.

The tubes that these smart cells build could be dangerous not just because they allow tumors to receive needed blood. Vasculogenic mimicry may also promote metastasis, the migration of tumor cells to new parts of the body, which is responsible for most cancer deaths. In a study reported last year in *Nature*, molecular biologist Greg Hannon of the University of Cambridge in the United Kingdom and colleagues tagged individual breast cancer cells with a specific nucleotide sequence, a DNA barcode, and injected them into mice. Some of the cells gave rise to tumors, and some of these tumors spawned metastases. The barcodes allowed the researchers to track metastases to their ancestral cells.



A tubular network formed by glioblastoma cells in a 3D culture.

Hannon and colleagues found that the cells most likely to grow into metastases had cranked up two genes that inhibit blood clotting. To the team's surprise, tumors derived from these wayward cells also showed vasculogenic mimicry. When the researchers inhibited the two genes, the number of vasculogenic mimicry channels in the tumors declined, suggesting that the genes induce formation of these blood-carrying networks.

Hannon says that the genes may benefit a tumor in two ways—by spurring it to grow the ersatz blood vessels and by preventing clots from blocking them, thus ensuring a steady supply of oxygen and nutrients. But the vasculogenic mimicry channels would also make it easier for tumor cells to make a getaway into the circulation. "To my knowledge, this is the first time we've discovered that a process that promotes metastasis provides a selective benefit to the primary tumor," Hannon says.

IF VASCULOGENIC mimicry does pave the way for metastasis, blocking it could save lives. Researchers have tested whether several standard angiogenesis inhibitors curb vasculogenic mimicry, but they seem to do the opposite. By stalling the formation of normal blood vessels and starving tumors of oxygen, the drugs appear to trigger cancer cells to build their own blood highways.

TaiRx, a biotech firm in Taipei, is exploring an alternative approach. The company originally developed the drug CVM-1118, a derivative of a plant compound, to block cancer cell growth. Yi-Wen Chu, TaiRx's senior vice president, sent the drug to Hendrix, her former Ph.D. supervisor, to determine whether it would halt vasculogenic mimicry. It did, curbing the activity of *Nodal*, a gene that drives vasculogenic mimicry by making cancer cells more like stem cells. This year the company launched a phase I trial to evaluate the safety of CVM-1118 in people with a variety of untreatable cancers and to assess its effectiveness.

Although CVM-1118 is the first drug that targets vasculogenic mimicry to reach clinical trials, pharmaceutical companies are trying to develop others, Hendrix says; several have sent her candidates to test, although she can't disclose the firms' names.

Hendrix recently became president of Shepherd University in Shepherdstown, West Virginia, a liberal arts institution that is her alma mater, and has arranged to move her lab to West Virginia University, Morgantown, where she plans to continue searching for agents that block tumors' do-it-yourself vessels. Looking back on the controversy she and her colleagues stirred, she says it was traumatic. But "if we shook up the field enough to get people to think about a new approach, I'll be happy." ■

ANCIENT DNA DIVIDE

While Europe forges ahead on a transformative technique, U.S. researchers struggle for funding

By **Ann Gibbons**

Christina Warinner was in demand. She had developed a new method to study bacterial DNA in plaque from ancient teeth, and researchers in Europe wanted her to apply it to their samples. Could she trace the spread of dairying in Europe 3000 to 4000 years ago? Or figure out why Stone Age hunter-gatherers in Morocco got so many cavities?

Warinner, 35, was eager to collaborate, especially because her European colleagues had something she didn't: ample funding. Based at the University of Oklahoma (OU), Norman, she publishes in leading journals and is widely considered a rising star, but she has only been able to cobble together about $100,000 a year in short-term grants from the National Science Foundation (NSF) in Arlington, Virginia, and OU. Her would-be collaborators had won European grants worth €1.5 million each—for research featuring her method. So this spring she did what other U.S. talents in the field of ancient DNA have done: crossed the ocean. She's taking a leave from Oklahoma to move to Germany, where the new Max Planck Institute for the Science of Human History in Jena will underwrite her research for the next 5 years.

"I want to stay in the U.S.," says Warinner, a Kansas native. But "they are offering me the equivalent of multiple NSF grants for the next 5 years. … What I see is this opportunity to do amazing science in Europe that is hard to do here because of funding limitations."

As a technological revolution has led to what Warinner and others call a golden age of paleogenomics (*Science*, 24 July 2015, p. 358), U.S. researchers are struggling to keep pace. In northern Europe, the epicenter of this fast-paced field, "ancient DNA labs are popping up like mushrooms," says Greger Larson, a U.S. evolutionary biologist who is director of paleogenomics at the University of Oxford in the United Kingdom. "A number of young Americans are coming over here."

By contrast, the structure and politics of science funding in the United States have put ancient DNA research—the epitome of curiosity-driven science—at a serious disadvantage. The interdisciplinary nature of the method is part of its power but also makes it prone to fall through the cracks in the U.S. system. And most human evolution research in the Unites States is considered social science, which has low priority.

As a result, the United States has fallen far behind in a burgeoning field that has transformed our understanding of the past, and also turns out to have unexpected medical and environmental applications. European labs dominate lists of the top papers in the field, and researchers based in Europe are getting funding to crack some of the most tempting scientific problems in the Americas, as well. "I was talking with my colleagues today and I asked them: Is there any other area of science where the U.S. is so far behind the rest of the world?" says population geneticist David Reich of Harvard Medical School in Boston.

THE FIELD OF ANCIENT DNA actually got its start in the United States. In the mid-1980s, researchers at the University of California (UC), Berkeley, first extracted DNA from a museum specimen of an extinct type of zebra and, soon after, from ancient insects trapped in amber (*Science*, 25 September 1992, p. 1860). But funding dried up in North America as researchers realized that contamination with modern DNA could throw off results.

By contrast, in Europe government agencies and foundations invested in a handful of key researchers who bet that those problems could be overcome. Germany's Max Planck Society invited Swedish paleogeneticist Svante Pääbo to build an institute in Leipzig in 1997 and paid for state-of-the-art labs. Max Planck does not divulge what it spends. But in 2006, when Pääbo asked for an extra $6 million for his risky—but ultimately successful—project to sequence the entire genome of a Neandertal, Max Planck immediately wrote a check. "What we try to do is not look for research areas, but for excellent people," says neuroscientist Angela Friederici, a vice president of the Max Planck Society in Leipzig. "We'll build an institute around you."

This summer, Max Planck is opening another paleogenomics institute—the Max Planck Institute for the Science of Human History—around Pääbo's former student,

PHOTO: ALEXANDER BABIC PHOTOGRAPHY

Christina Warinner, shown here in her lab at the University of Oklahoma, Norman, can't afford to stay in the United States and is moving to Germany.

Johannes Krause, and two other scientists. This is the institute that has drawn Warinner and other Americans to Germany.

Across the English Channel, New Zealander Alan Cooper established the Henry Wellcome Ancient Biomolecules Centre at Oxford back in 1999, and the trust spent about $5.6 million in 2003 on a new building that served as incubator for a wave of rising stars. One Wellcome Trust postdoc there was Eske Willerslev, who later returned to his native Denmark in 2005 to open a lab. As co-director of the Centre of Excellence in GeoGenetics at the University of Copenhagen, he and his collaborators have won $8.35 million from 2011 to 2019 for ancient DNA and other research, including $7.6 million from the Danish National Research Foundation (DNRF) alone. He's now seeking funds to build a lab at the University of Cambridge in the United Kingdom as well.

The European Research Council (ERC), which aims to fund cutting-edge research in Europe, has also pitched in, offering generous grants for young investigators. Larson, for example, who trained with Cooper at Oxford, received about $1.7 million over 5 years.

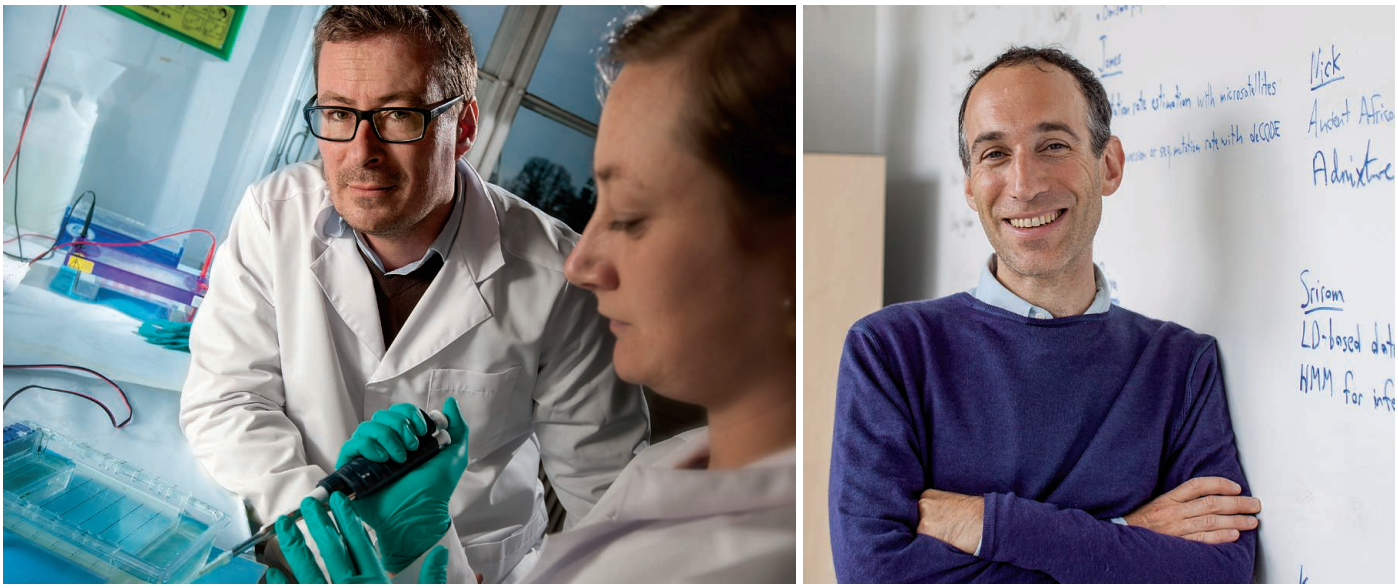
The return on this European investment has been spectacular. Hardly a week goes by without a new paleogenomics paper published in *Science*, *Nature*, or other leading journals. Thomson Reuters's Web of Science, which indexes more than 12,000 of the most influential peer-reviewed journals, recently identified a cluster of highly cited ancient DNA papers as a "research front"—a rapidly growing area of science (*Science*, 24 July 2015, p. 362). And 18 of the 20 most cited papers in the field over the past 5 years had lead authors based in Europe, notably Pääbo and Willerslev; one came from Cooper, now based in Australia, and one paper had a lead author—Reich—based in the United States.

Although Europe funds ancient DNA work as basic research, it is beginning to pay medical and technological dividends. For example, by sequencing the Neandertal genome—a blue skies project if ever there was one—researchers have discovered a host of Neandertal immune and other genes in living people that profoundly affect risks of disease (*Science*, 12 February, p. 648). And studies of ice age plant and animal DNA are helping researchers predict the fate of populations threatened by climate change (*Science*, 24 July 2015, p. 367).

All the same, money is getting tighter even in Europe, Willerslev says, because "the trend has moved from funding basic science to applied sciences." DNRF, which underwrites basic research, only funds 4% of its proposals, down from 5% in 2014, he says. Meanwhile, expenses are rising. Willerslev spends 38% of his center's grant income on salaries alone. "Denmark has the most expensive Ph.D. and postdoc salaries in the world," he says.

Still, Larson says, Europe has done a far better job of nurturing the field. "If what you want is top scientists publishing in top journals, the U.S. is missing out on a huge possibility," he says.

IN THE UNITED STATES, only one major governmental source funds most ancient DNA research: NSF. Its awards are not generous. "One ERC grant would be five NSF grants," says evolutionary biologist Beth Shapiro at UC Santa Cruz.



David Reich (right) of Harvard Medical School in Boston gets less funding than European stars like Eske Willerslev (left) of the University of Copenhagen.

Whereas ERC awards money on the basis of the investigator, NSF slices research into fields and earmarks funds for each one. Because ancient DNA is a multidisciplinary tool, it spans more than two dozen programs within NSF, and it is difficult to track money spent on it. However, many of the key European projects, such as sequencing the Neandertal genome, would fall under NSF's biological anthropology program. This program, part of NSF's Social, Behavioral and Economic Sciences (SBE) directorate, has a total budget of $4.4 million this year (out of a $7.7 billion NSF budget) to fund all research in human and primate evolution—from hunting for fossils to studying primate behavior to imaging methods—says program director Rebecca Ferrell. Awards to individual investigators in the program typically range from $100,000 to $350,000 a year for 3 years.

Other ancient DNA proposals fall under the archaeology program of SBE, such as one from archaeologist Dennis Jenkins of the University of Oregon, Eugene, to trace the peopling of the Americas by studying DNA in ancient human feces in Oregon. This program funded about $7.2 million worth of individual awards last year on everything from tree ring dating to pre-Columbian America, says program director John Yellen. Jenkins, working with Willerslev, got a total of $180,000 between 2009 and 2012, and used Danish funding to complete his project.

Part of the funding difference may be that Europe tends to put a higher priority on funding research into prehistory. "There's more archaeology in Europe and a longer history of funding it," Larson notes. Another factor is politics. In Congress, some legislators have challenged the idea of curiosity-driven research, instead pushing for science "in the national interest." SBE, the smallest NSF directorate, has come under fire from legislators who mock social science studies (*Science*, 1 January, p. 10).

The European community's funding is less vulnerable to political agendas, researchers say. "ERC grants are rarely tied up by politics in a way that prioritizes fields to be funded," says U.K. evolutionary biologist Tom Gilbert of the University of Copenhagen, who's gotten about $10 million from ERC over the past decade. "They are more about supporting excellence of top researchers and training."

Exacerbating the money crunch in the United States is the fact that research in paleogenomics has become more ambitious and technically demanding, pricing itself out of biological anthropology. "It's almost like ancient DNA isn't a subfield of anthropology anymore—it's really now genomics applied to ancient samples," Warinner says. "Our budget needs have really changed."

Researchers who want to sequence full human genomes are hard-pressed to do so on NSF grants that often yield less than $100,000 a year after indirect costs—the 40% to 70% share that goes to universities to pay for overhead—are deducted. Lab techs and postdocs cost $60,000 to $80,000 a year, so researchers have to make hard choices. "Hiring people is almost impossible. I can't afford a lab tech, and with lab costs, it's a nightmare," says Warinner, who relies mostly on graduate students.

A handful of Americans have found ways around this system. Shapiro bypassed NSF's small biological anthropology and archaeological programs, instead approaching NSF's Arctic Natural Sciences Program, which recently awarded her $423,000 over 4 years to study ancient DNA of mammoths and other extinct animals in the Yukon (*Science*, 24 July 2015, p. 369). She feels "lucky" to have two NSF grants, plus a "genius" grant from the MacArthur Foundation, and $1.2 million from the Gordon and Betty Moore Foundation.

Reich is funded by the private Howard Hughes Medical Institute (HHMI); like the Max Planck Society, this foundation awards grants to people rather than by field. HHMI grants reportedly average about $1 million a year for 5 years for salary, benefits, and research (*Science*, 25 July 2014, p. 366). Reich also was the only ancient DNA researcher to benefit from a one-time competition to fund large-scale research on human origins at NSF, which gave him $3 million over 6 years. That funding ends in August, and even with it, Reich's funding pales in comparison with what European governments and foundations invest in their top ancient DNA researchers.

Some paleogenomics researchers have applied to the National Institutes of Health (NIH), which is the largest funder of biomedical and human genetics research in the world and whose mission includes basic research, but they haven't had much luck. Anthropological geneticist Anne Stone of Arizona State University, Tempe, recently applied to the National Institute for Allergy and Infectious Diseases to follow up on the discovery that 1000-year-old mummies in Peru had contracted tuberculosis (TB) from seals, which she reported with Krause last year in *Nature*. But the NIH panel's response was: "Why do we care about what strains are in seals? Most TB is caused by human strains," Stone says.

PHOTOS: (LEFT TO RIGHT) MIKAL SCHLOSSER; PRESIDENT AND FELLOWS OF HARVARD COLLEGE

Ancient DNA researchers need to make sure they apply to the right NIH division, suggests Dorit Zuk, director of the National Institute of General Medical Sciences's Division of Genetics and Developmental Biology in Bethesda, Maryland. They might have better luck at her institute, which supports basic research—and funds 29% of the proposals it receives, she says. "What we fund is what we get." The agency supports about a half-dozen teams working on methods to analyze variation in living and ancient human genomes. But it has funded only two researchers' work in a wet lab with ancient human DNA, according to a search of the NIH database.

"It's anomalous that our funding agencies don't have a bucket of money for ancient DNA," says Reich, who argues that "NSF is the natural place to support this … [because it] is a revolutionary new scientific instrument like a microscope or a telescope," and the agency supports such new technologies.

But at NSF, the decision to spend more would have to be made at the highest levels. "I think ancient DNA research is important … but decisions about broader NSF initiatives are not within my purview," Ferrell says. The heads of the social and biological directorates both declined to comment for this story. Amanda Greenwell, head of NSF's Office of Legislative and Public Affairs, would speak only generally, saying that defining funding priorities takes months of discussion with the National Science Board, NSF advisory committees, and program staff, and must address congressional requirements.

NSF's specific, finely drawn funding priorities simply leave less room for "risky" science, some analysts say. "The more NSF and Congress focus on short-term results and things that are obviously tied to narrowly defined measures of success, the less you're going to get curiosity-driven science," says economist Julia Lane at New York University in New York City, who developed tools to measure the impact of funding on "transformative research" while she was at NSF from 2009 to 2012. "You want researchers to take risks … or you risk killing the goose that lays the golden egg."

That's what European researchers appreciate: "The amazing thing about the ERC is they actually like 'high-risk' projects," says Oxford archaeologist Michael Petraglia, an American headed for the Max Planck Institute in Jena.

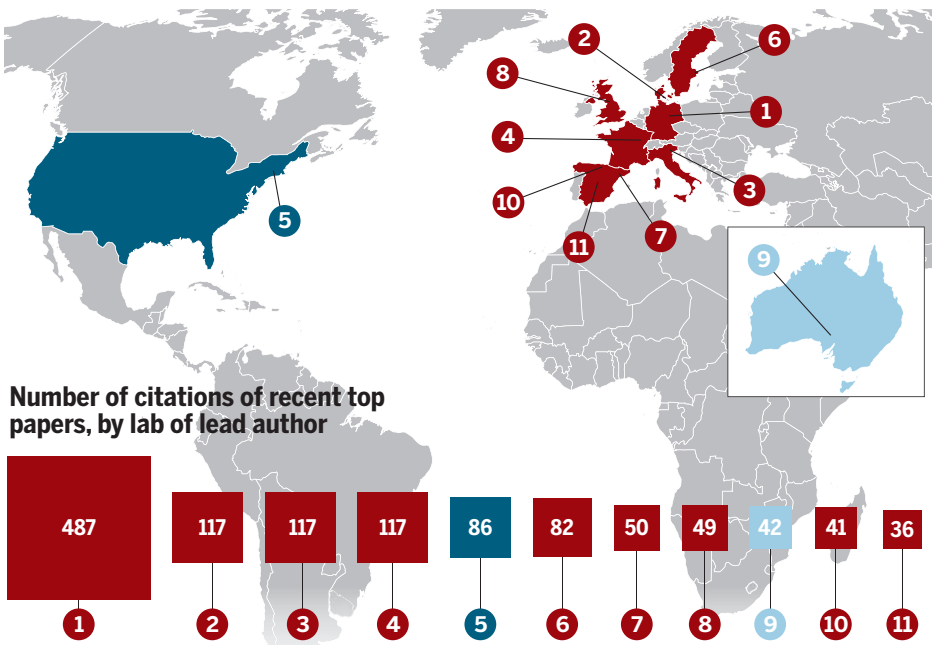
That, and the more generous grants, boosts the odds for those U.S. ancient DNA researchers who move to Europe. Evolutionary virologist Alex Greenwood, who studies ancient viruses in extinct sloths and other animals, struggled to survive on a $30,000 grant at Old Dominion University in Norfolk, Virginia. Now, he is at the Leibniz Institute for Zoo and Wildlife Research in Berlin, and has won seven of nine grants he's applied for from the German government, ERC, and the Leibniz society, worth almost $1.5 million.

Other Americans work to find European collaborators. "I work at one of the best museums of archaeology and anthropology in the country and we're farming out samples," says anthropologist Theodore Schurr of the University of Pennsylvania, who has been unable to raise the $1 million to $2 million he needs to build an ancient DNA lab.

Indeed, it is Willerslev of the University of Copenhagen who uses his European funding to address key paleogenomics questions in the United States, recently sequencing the genomes of ancient Native Americans such as Kennewick Man and the Anzick child. "When I think of Native American ancient DNA genetics, it's Eske who comes to mind," Greenwood says. "It doesn't have to be." ■

Geography of high-impact ancient DNA research

Analysis of the most cited ancient human DNA papers in the past 5 years shows the effect of European funding. Only one Australian lab and one U.S. lab had a paper in the top 18 (excluding review papers).



1 Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany
2 University of Copenhagen
3 European Academy of Bozen/Bolzano, Italy
4 University of Toulouse, France
5 Harvard Medical School, Boston
6 Uppsala University, Sweden
7 Institut de Biologia Evolutiva, Barcelona, Spain
8 University of Huddersfield, Queensgate, U.K.
9 University of Adelaide, Australia
10 University of the Basque Country, Bizkaia, Spain
11 Universidad Complutense de Madrid

GRAPHIC: A. CUADRA/*SCIENCE*

Richer and poorer

Europeans tend to receive larger, more open-ended grants, whereas U.S.-based researchers receive smaller, more narrowly defined awards, as shown by a sampling of current public grants.

LEAD INVESTIGATOR	FUNDING AGENCY	AMOUNT IN U.S. DOLLARS
Alex Greenwood	Deutsche Forschungsgemeinschaft	221,000
Johannes Krause	European Research Council (ERC)	1,700,000
Greger Larson	ERC/U.K. Natural Environment Research Council	1,700,000/480,000
Eske Willerslev	Danish National Research Foundation	7,600,000
Christina Warinner	National Science Foundation (NSF), Biological Anthropology	188,000
Anne Stone	NSF, Biological Anthropology	242,176
Beth Shapiro	NSF, Arctic Natural Science	423,000
David Reich	NSF, one-time special Human Origins award	3,000,000

PERSPECTIVES

MARINE ENVIRONMENT

Persistent pollutants, persistent threats

Polychlorinated biphenyls remain a major threat to marine apex predators such as orcas

By **Paul D. Jepson**[1] *and* **Robin J. Law**[1,2]

Persistent organic pollutants (POPs) are chemical substances that persist in the environment, accumulate in the food web, and pose a risk of adverse effects in humans and wildlife (*1*). Rachel Carson first identified the potentially devastating effect of POPs on wildlife in the early 1960s (*2*). In the late 1960s, polychlorinated biphenyls (PCBs) were detected in high concentrations in wildlife in Sweden (*3*). After PCB use and manufacture were banned in 1979 (US), 1981 (UK), and 1987 (EU), levels started to decline slowly in all biota around the world (*4*–*6*). In 2004, the Stockholm Convention committed more than 90 signatory countries to phasing out or eliminating large stocks or other sources of POPs, including PCBs (*1*). Yet, PCBs continue to threaten the survival of marine predators. Concerted efforts are thus still needed to mitigate PCB pollution.

Since legislative restrictions on different POPs have been introduced, tissue concentrations of POPs, including PCBs and DDT-related compounds (DDTs), have declined, and many wildlife populations have recovered. For example, populations and reproductive indices of grey seal, otter, and white-tailed sea eagle recovered in Sweden during the 1980s as tissue PCB and DDT concentrations fell substantially from the 1960s and 1970s to 2010 (*5*, *7*). Most avian marine apex predators, including herons, gulls, ospreys, petrels, and skuas, are no longer listed as threatened on the International Union for Conservation of Nature (IUCN) Red List, although half of the remaining sea eagle species are still threatened with extinction due to a range of threats, potentially including PCBs (*8*). Among pinnipeds, only the Hawaiian and Mediterranean monk seals and several spe-

[1]*Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK.* [2]*Centre for Environment, Fisheries and Aquaculture Science (Lowestoft Laboratory), Pakefield Road, Lowestoft, Suffolk, NR33 0HT, UK. Email: paul.jepson@ioz.ac.uk*

PHOTO: PAUL NICKLEN/NATIONAL GEOGRAPHIC CREATIVE

A pod of orcas in the northern fjords of Norway. Although polychlorinated biphenyl (PCB) concentrations have declined in many wildlife species, they remain elevated in orcas and other marine apex predators, potentially threatening their survival.

cies in restricted habitats close to humans, such as Saimaa ringed seals, Ungava harbor seals, and Caspian seals, are still threatened with extinction (*8*).

Banned organochlorine pesticides like DDT and dieldrin have also declined significantly in marine apex predators in Europe. However, PCBs have stopped declining and still persist at excessively high concentrations in some cetaceans, including killer whales (orcas) (see the photo) and bottlenose dolphins, in the northeast Atlantic (*6*, *9*) and many cetacean species in the Mediterranean Sea (*6*, *10*). High PCB concentrations in European cetaceans from 1990 to 2012 were widely associated with long-term population declines and low or zero rates of reproduction, consistent with severe PCB-induced population-level effects (*6*). Further research is needed to assess the full effect of PCB exposure on cetaceans in Europe, particularly in Iberian (northeast Atlantic), Mediterranean, and Black Sea countries, where exposures are highest.

So, are high PCB concentrations just a European problem? Not necessarily. The high trophic level and longevity of most marine apex predators globally mean that accumulation of PCBs and other POPs is inevitable (*4*). The relatively long lactation period in cetaceans also enables considerable transmission of PCBs from mother to calf. The killer whale remains the most highly PCB-contaminated species on Earth, with very high concentrations found in individual killer whales throughout their range (*4*, *6*, *11*). After humans, killer whales may once have been the most widely distributed mammalian species on Earth, with a historic distribution of "pole to pole and everywhere in between" (*8*). Yet, in recent years, high killer whale numbers have only been found near the less polluted Arctic and Antarctic regions (*4*, *8*, *11*, *12*).

Other long-lived marine apex predator species that may still be at risk from PCB pollution include false killer whales and coastal bottlenose dolphins, ringed seals in the Baltic Sea, all marine mammal species in the Mediterranean and Black Seas, beluga in the Saint Lawrence River, Canada, polar bears in the Arctic, and some sea eagle species in the Northern Hemisphere (*8*). In Southeast Asia, China's recent industrialization and agricultural expansion is thought to have increased concentrations of PCBs and other POPs and may have played a role in the probable extinction of the Yangtze River dolphin and in threatening the critically endangered Yangtze finless porpoise (*8*). PCBs and DDTs may also play a role in ongoing population declines of Indo-Pacific finless porpoises, Indo-Pacific humpback dolphins, Ganges River dolphins in India and Bangladesh, and Indus River dolphins in Pakistan (*8*).

Although experiments have shown that PCBs have toxic effects in fish, empirical evidence of PCB toxicity in wild fish is lacking (*13*). Fish also lack maternal transfer of POPs through fat-rich milk (*8*, *13*). Nonetheless, some apex predator sharks are relatively long-lived species that feed at a similar trophic level to killer whales but have not been rigorously assessed for PCB (or other POP) exposures (*8*). Further research is needed on PCB exposures in apex predator sharks like the great white shark to assess whether PCB exposures are likely to have significant population-level effects. There has been a clear lack of sightings, bycatch, and strandings of great white sharks in the past 20 to 30 years, particularly in the highly PCB-contaminated northeast Atlantic and Mediterranean Sea regions (*8*). Other apex predator sharks potentially affected by PCBs include bull sharks, great hammerheads, shortfin mako, and tiger sharks (*8*).

POPs, including PCBs, are not static. They transfer over long distances from industrialized to nonindustrialized regions, including the Arctic, mainly through cycles of atmospheric volatilization and condensation called global distillation (*14*), with killer whales and polar bears accumulating the highest Arctic exposures (*11*). In East Greenland polar bears, blubber PCBs increased unexpectedly between 2010 and 2013, resulting in PCB concentrations that were as high in 2013 as in 1983 (*15*). Whatever the cause, the recent increase in PCBs and other POPs in polar bear blubber is a worrying trend if it were to continue, emphasizing the importance of continued monitoring of legacy POPs in Arctic biota.

"...there is an urgent need to review current methods of PCB mitigation in the marine environment—both in Europe and elsewhere."

Future research should investigate pathways of PCB contamination of the marine environment. Which sources are PCB inputs from land and which are recycled in the marine environment? Future monitoring should also assess PCB levels in marine sediments, discharges, and freshwater outflow. At a European policy level, PCB exposures in comparison to established marine mammal PCB toxicity thresholds should be used to assess whether marine mammal populations reach "Favourable Conservation Status" under the European Commission's Habitats Directives. PCBs should also be included as a cetacean "indicator" for populations under the EU's Marine Strategy Framework Directive (under Descriptor 8).

Perhaps most important, there is an urgent need to review current methods of PCB mitigation in the marine environment—both in Europe and elsewhere. This should include full compliance with the Stockholm Convention to significantly reduce PCB contamination of the marine and terrestrial environment by 2028. This is the critical issue for future conservation success. PCB mitigation measures include the safe disposal or destruction of large stocks of PCBs and PCB-containing equipment, limiting the dredging of PCB-laden rivers and estuaries, reducing PCB leakage from old landfills (*6*), limiting PCB mobilization in marine sediments, and regulating demolition of PCB-containing precast buildings such as tower blocks built in the 1950s to 1980s (*6*). ■

REFERENCES

1. United Nations Environment Programme, *The Stockholm Convention on Persistent Organic Pollutants* (UNEP, Nairobi, 2001); http://chm.pops.int/default.aspx.
2. R. Carson, *Silent Spring* (Houghton Mifflin Mariner Books, 1962).
3. S. Jensen, A. G. Johnels, M. Olsson, G. Otterlind, *Nature* **224**, 247 (1969).
4. M. Houde *et al.*, *Rev. Environ. Contam. Toxicol.* **184**, 1 (2005).
5. A. M. Roos *et al.*, *Environ. Pollut.* **170**, 268 (2012).
6. P. D. Jepson *et al.*, *Sci. Rep.* **6**, 18573 (2016).
7. Latin species names for all species mentioned in the main text: grey seal, *H. grypus*; otter, *L. lutra*; white-tailed sea eagle, *H. albicilla*; Hawaiian monk seal, *N. schauinslandi*; Mediterranean monk seal, *M. monarchus*; Saimaa ringed seal, *P. hispida saimensis*; Ungava harbor seal, *P. vitulina melonae*; Caspian seal, *P. caspica*; killer whale, *O. orca*; bottlenose dolphin, *T. truncates*; false killer whale, *P. crassidens*; ringed seal, *P. hispida*; beluga, *D. leucas*; polar bear, *U. maritimus*; Yangtze River dolphin, *L. vexillifer*; Yangtze finless porpoise, *N. asiaeorientalis asiaeorientalis*; Indo-Pacific finless porpoise, *N. phoceonoides/N. asiaeorientalis*; Indo-Pacific humpback dolphin, *S. chinensis*; Ganges River dolphin, *P. gangetica gangetica*; Indus River dolphin, *P. gangetica minor*; great white shark, *C. charcharias*; bull shark, *C. leucas*; great hammerhead, *S. mokarran*; shortfin mako, *I. oxyrinchus*; tiger shark, *S. cuvier*.
8. International Union for Conservation of Nature, *The IUCN Red List of Threatened Species, Version 2015.4* (IUCN, Gland, 2015); www.iucnredlist.org.
9. International Council for the Exploration of the Sea, *Report of the Working Group on Marine Mammal Ecology (WGMME)*, 8–11 February 2016, Madrid, Spain (ICES, Copenhagen, 2016).
10. M. Pinzone *et al.*, *Environ. Res.* **142**, 185 (2015).
11. R. J. Letcher *et al.*, *Sci. Total Environ.* **408**, 2995 (2010).
12. K. A. Forney, P. R. Wade, in *Whales, Whaling, and Ocean Ecosystems*, J. A. Estes *et al.*, Eds. (Univ. of California Press, Berkeley/Los Angeles, 2006), pp. 145–162.
13. T. B. Henry, *Crit. Rev. Tox.* **45**, 643 (2015).
14. J. Ma., H. Hung, C. Tian, R. Kallenborn, *Nat. Climate Change* **1**, 255 (2011) .
15. F. Rigét *et al.*, *Environ. Pollut.* 10.1016/j.envpol.2015.11.006 (2015).

10.1126/science.aaf9075

GENE EXPRESSION

Single-cell variability guided by microRNAs

Variability in miRNA activity, and therefore gene networks, may define cell state

By Salil Garg[1,2] and Phillip A. Sharp[2]

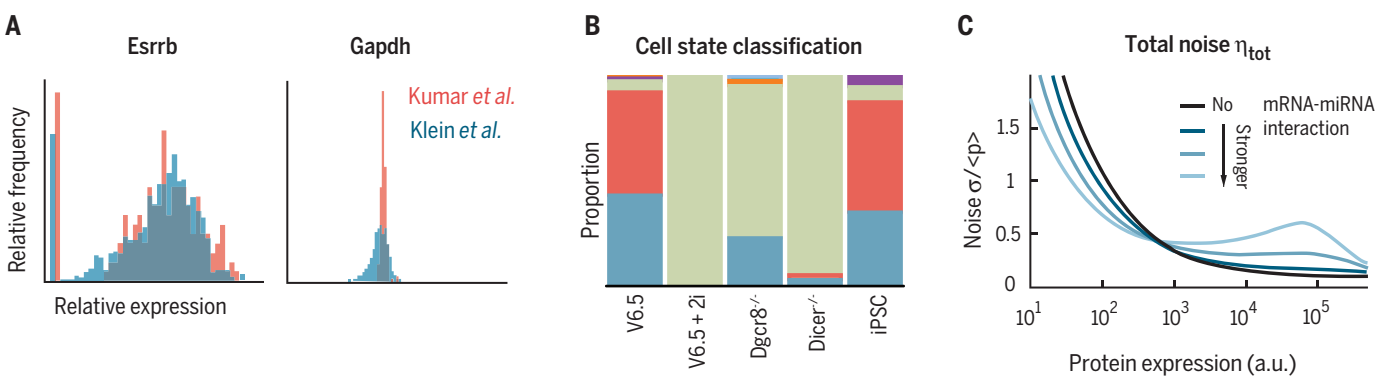
A single zygote with a defined DNA sequence gives rise to all the varied cells of the mammalian body. This variety of cell fates may arise in part from cell-to-cell variability in the gene expression programs of embryonic stem cells (ESCs). Three recent studies have taken different approaches to characterizing this variability in gene expression in stem cells (*1*–*3*). These results suggest that microRNAs (miRNAs) could play an important role in controlling and generating this variability.

Pluripotent stem cells move irreversibly down specific differentiation pathways during development. Yet the effective concentrations of key pluripotency factors fluctuate over time, and seemingly uniform populations of stem cells can exhibit multiple cell fates. Differentiated cells can also be driven to dedifferentiate, as seen in induced pluripotent stem cells (iPSCs) (*4*). Thus, cells may exist in a continuum of "cell states" that can interconvert with each other and have varying potential to differentiate or to self-renew. However, the factors that drive interconversion of cell states have remained poorly defined, and variability is often attributed to "noise"—the random fluctuations of biochemical events. This raises the question of whether such variability is stochastic or coherently organized.

Kumar *et al.* (*1*) used single-molecule RNA fluorescence in situ hybridization (FISH), single-cell reverse transcription polymerase chain reaction (RT-PCR), immunofluorescence, and single-cell RNA sequencing to explore the gene variability landscape of pluripotent ESCs. Klein *et al.* (*2*) performed deep RNA sequencing using a droplet-microfluidic approach to identify genes displaying variable transcript levels among single ESCs. Both reports find that the majority of genes, such as those encoding housekeeping and metabolic factors, display stable normal distributions of expression. However, RNA levels for a subset of genes, such as signaling factors and developmental regulators, are more variable cell-to-cell and display bimodal or heterogeneous expression patterns. Among these variable genes are pluripotency factors known to fluctuate cell-to-cell across the population (e.g., Nanog, Esrrb) and whose expression is key to pluripotency (*5*, *6*). Strikingly, for many genes, both Kumar *et al.* and Klein *et al.* report similar distribution of expression and extent of variability (see the figure, panel A).

Determining the set of genes with variable expression within an ESC population represents an important step in understanding what drives the interconversion between cell states. A cell state can be defined as the relatively stable expression of a group of genes at (or near) specific levels. Both Kumar *et al.* and Klein *et al.* undertook principal components analysis to identify gene expression relationships. In the case of Kumar *et al.*, the range of cell types analyzed included ESCs lacking mature miRNAs and lacking signaling through the ERK/GSK3 kinase pathway (see the figure, panel B). Similar discrete cell states were observed across the various experimental conditions, although the portion of cells in each cell state varied depending on the nature of the perturbation. For example, inhibiting the ERK/GSK3 pathway (the 2i condition) or removing miRNA regulation (Dicer$^{-/-}$, Dgcr8$^{-/-}$) resulted in a common "ground" state with little variation in the expression of key pluripotency genes (green in the figure, panel B). Other ESCs were classified into "transition" states according to their expression of key components (red and blue in the figure, panel B). Labeling these as transition states is reasonable given the overlap in expression with the ground state, but the experimental methodologies did not allow testing of the interconversion of individual cells between states.

Klein *et al.* identified a similar continuum of heterogeneous cell states within wild-type ESC populations. Further, they examined gene expression covariation to identify a set of genes closely correlating with a given gene and with each other to infer regulatory information. This topological network analysis of covariation patterns reveals that pluripotency factors such as Nanog are strongly enriched for interaction with other pluripotency factors, which



Taming the noise. (**A**) Comparing the single-cell RNA sequencing data of Kumar *et al.* (*1*) and Klein *et al.* (*2*) reveals that the pluripotency factor Esrrb is heterogeneously expressed, whereas Gapdh is narrowly expressed. (**B**) Embryonic stem cells under chemical (2i) and genetic (Dgcr8$^{-/-}$, Dicer$^{-/-}$) perturbations show similar cell states (green, ground; blue, transition 1; red, transition 2; purple, primed; orange, fibroblast states) within the total population relative to the wild type (V6.5). (**C**) In the model proposed by Schmiedel *et al.* (*3*), at low protein expression (a.u., arbitrary units), miRNA activity decreases gene expression variability (noise = standard deviation/mean); at high protein expression, miRNA activity increases gene expression variability through transmission of variability in the miRNA pool.

[1]Massachusetts General Hospital Department of Pathology, Harvard Medical School, Boston, MA, USA. [2]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. Email: sharppa@mit.edu

in turn form a gene neighborhood. Thus, pluripotency factors likely fluctuate together as a network to generate transitions between cell states.

The third study, by Schmiedel *et al.* (*3*), may provide a key insight. Combining a reporter construct for monitoring miRNA activity with theoretical and mathematical modeling, they proposed that miRNAs increase protein expression noise for highly expressed genes, and that they decrease such noise for genes that are not highly expressed (see the figure, panel C). Reduction in gene expression noise was previously proposed as a hallmark of miRNA function in the context of transcriptional dynamics. Yet viewed in the context of findings from single-cell RNA sequencing in ESCs, a role for miRNAs in generating gene expression variability is noteworthy. Perhaps miRNA activity increases the cell-to-cell variability of key pluripotency factors through the transmission of variability in the miRNA pool itself. This is consistent with the cell-to-cell variation in miRNA activity observed by Schmiedel *et al.* (*3*). Variability is then transmitted throughout the pluripotency network according to the topology of gene regulatory interactions. Downstream action by pluripotency factors may then establish miRNA expression patterns through a virtuous cycle that determines a cell state. Because miRNAs have the potential to control expression of large numbers of genes, variability in miRNA activity would provide a potential mechanism for generating transitions between cell states.

The three studies raise more questions than they answer. For example, it is not known what accounts for the network of genes that covary between cell states, nor the process that produces variability in key pluripotency genes. Variation in the expression of miRNAs may provide one answer but raises the issue of what accounts for miRNA variation. Further, it is not clear whether similar mechanisms are important for the activity of other stem cells, such as cancer stem cells. A major limitation of all three studies is the inability to follow variation within individual cells and relate this to their biochemical state. Nonetheless, if correct, miRNA-guided variability will have implications for intrinsically generated heterogeneity in tumorigenesis, development, and immune function. ■

REFERENCES

1. R. M. Kumar *et al.*, *Nature* **516**, 56 (2014).
2. A. M. Klein *et al.*, *Cell* **161**, 1187 (2015).
3. J. M. Schmiedel *et al.*, *Science* **348**, 128 (2015).
4. K. Takahashi, S. Yamanaka, *Cell* **126**, 663 (2006).
5. I. Chambers *et al.*, *Nature* **450**, 1230 (2007).
6. H. Niwa *et al.*, *Nature* **460**, 118 (2009).

10.1126/science.aag1097

ILLUSTRATION: K. SUTLIFF/*SCIENCE*

PROTEIN TRANSLATION

Seeing translation

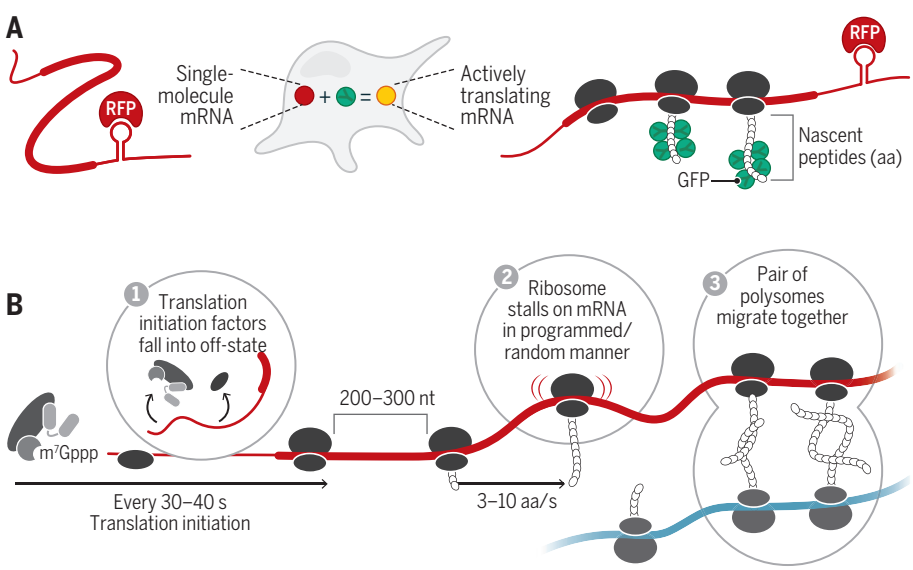
Monitoring individual messenger RNAs as they make protein reveals the tricks of translation

By **Shintaro Iwasaki** *and* **Nicholas T. Ingolia**

By controlling gene expression at the level of translation, cells can fine-tune protein abundance, respond quickly to environmental changes, and restrict synthesized proteins to specific places within the cell. Whereas the basic biochemistry of translation is understood in great detail, questions remain about translation in the context of the cell: how ribosomes load onto a messenger RNA (mRNA), how fast and how far translating ribosomes move, and how rapidly translation reacts to the environment. On pages 1425 and 1430 of this issue, studies by Morisaki *et al.* (*1*) and Wu *et al.* (*2*), respectively, as well as papers in *Cell* by Yan *et al.* (*3*) and Wang *et al.* (*4*), provide a new view of translation by developing real-time imaging of nascent peptides being synthesized from single mRNAs in cells. These novel technologies provided versatile frameworks for quantitative in vivo translation studies and revealed hidden aspects of proteins synthesis.

Although it might seem straightforward to visualize the translation of an intrinsically fluorescent protein such as green fluorescent protein (GFP), this approach is infeasible because maturation of the fluorophore is much slower than protein synthesis, and a single fluorescent protein is far too faint to be detected. To reveal nascent peptides emerging from translating ribosomes, all four studies used large tandem repeats of an epitope that is rapidly bound by preexisting fluorescent antibodies once it emerges from the ribosome. Simultaneously, the authors track the reporter mRNAs with fluorescent RNA-binding proteins, such that colocalized nascent peptides report on the translation status of a single mRNA (see the figure, panel A).

Strikingly, all four studies observed similar kinetics of translation in spite of notable differences in the labeling strategies, cell types, and analysis methods used. Ribosomes occur every 200 to 300 nucleotides (nt) on an mRNA, and elongation speeds (3 to 10 amino acids per second) were fairly close to those measured by ribosome profiling (5.6 amino



Translating the message. (**A**) Fluorescence-based translational reporters label nascent proteins (antibodies, green) and transcripts (RNA-binding proteins, red). An actively translated mRNA then appears yellow (green + red) in the cell. (RFP, RNA-binding red fluorescent protein.) (**B**) Translation kinetics in vivo. (1) Initiation. Translation typically initiates every 30 to 40 s, but this can be interrupted by a translationally silent state lasting minutes or hours. (2) Elongation. Translation usually proceeds at 3 to 10 amino acids (aa) per second, but ribosomes can stall in response to programmed signals or random events. (3) Diffusion. Most polysomes undergo free, independent diffusion, but a small fraction move together as a pair.

acids per second) (*5*). In addition, these studies provided the first direct observation of translation initiation frequency in cells, occurring every 30 to 40 s on an actively translated mRNA; this is suggested to be the rate-limiting step for most protein production (*6*) (see the figure, panel B).

These averaged features agreed well with bulk biochemical measurements, but single-molecule analyses can also reveal how translation is variable among mRNAs. It has been widely accepted that mRNAs continually bind initiation factors through their 5′ cap and recruit ribosomes. By tracking a single mRNA over several hours, Yan *et al*. and Wu *et al*. revealed fluctuations between nontranslating and translating states of mRNA. The time scale for switching translation on and off ranged from 15 min (seen as "bursts") to 3 hours across different cell types and reporters. Although the cause of the fast translation bursts remains unknown, they may be coupled with the localization of mRNAs, or reflect cycles of mRNA decapping and cytoplasmic recapping (*7*).

Translation kinetics vary widely between genes, depending on the mRNA sequence and RNA binding proteins bound to the transcript, and are subject to dynamic control (*8*). Wang *et al*. shed new light on the classic translation regulation pathway triggered by the response to oxidative stress and by the unfolded protein response in endoplasmic reticulum. These stress responses induce phosphorylation of the eukaryotic translation initiation factor eIF2α, generally reducing overall initiation while activating a subset of mRNAs, including the transcript encoding the stress-induced transcription factor activating transcription factor 4 (Atf4). Surprisingly, this translation activation is transient, only lasting ~150 s, despite the sustained increase in the amount of Atf4 protein seen during stress.

Locally controlled translation can restrict protein production to a specific place within the cytoplasm and thereby plays many roles in the internal spatial organization of the cell (*9*). For example, regulated protein synthesis at specific neuronal synapses is required for long-term potentiation of those synaptic connections, and thus for learning and memory. Wu *et al*. and Wang *et al*. directly visualized translation occurring within dendrites, the branched projections of neurons. Dendritic mRNAs are thought to be transported from the cell body to the synapses in a translationally repressed state. The authors observed a trend of higher translation efficiency toward the proximal end of dendrites, and lower efficiency at the distal end. Intriguingly, translation is not completely silenced during transport; ~20% of mRNAs are actively translated while exhibiting rapid (~2 μm/s) and directional movement along dendrites.

An mRNA being actively translated by a number of ribosomes forms a complex known as a polysome. Polysome mobility varied between subcellular compartments and between individual mRNAs in both neurons and less elaborated cells. Most cytosolic polysomes showed freely diffusive movement, although some showed constrained, "subdiffusive" movement, and a rare few were actively transported.

The protein encoded by the mRNA also influenced polysome mobility—longer reading frames produce larger, slower-moving polysomes at steady state. Nascent actin protein being synthesized by the ribosome yielded subdiffusive polysome movement, perhaps mediated by actin protein-protein interactions. Secreted proteins undergoing cotranslational membrane insertion restrained polysome movement even more.

Such nascent protein interactions might also explain why, although the majority of polysomes behaved independently, a detectable fraction (~5%) of polysomes from two distinct mRNAs encoding the same protein comigrated in cytoplasm. This compartmentalized "dipolysome" may be explained by the cotranslational assembly of a protein complex, as observed in bacteria (*10*).

Although these studies relied on exogenous reporters to detect translation, tagging endogenous loci by genome editing could illuminate translation in a yet more physiological context. Indeed, following in the footsteps of single-molecule mRNA tracking, measurement of translation in diverse tissues and developmental conditions will uncover critical information on the dynamics of translation and its regulation in time and space. ■

Department of Molecular and Cell Biology, Center for RNA Systems Biology, University of California, Berkeley, CA 94720, USA. Email: ingolia@berkeley.edu

REFERENCES

1. T. Morisaki *et al.*, *Science* **352**, 1425 (2016).
2. B. Wu, C. Eliscovich, Y. J. Yoon, R. H. Singer, *Science* **352**, 1430 (2016).
3. X. Yan, T. A. Hoek, R. D. Vale, M. E. Tanenbaum, *Cell* **165**, 976 (2016).
4. C. Wang, B. Han, R. Zhou, X. Zhuang, *Cell* **165**, 990 (2016).
5. N. T. Ingolia, L. F. Lareau, J. S. Weissman, *Cell* **147**, 789 (2011).
6. P. Shah, Y. Ding, M. Niemczyk, G. Kudla, J. B. Plotkin, *Cell* **153**, 1589 (2013).
7. C. Mukherjee *et al.*, *Cell Rep.* **2**, 674 (2012).
8. R. J. Jackson, C. U. Hellen, T. V. Pestova, *Nat. Rev. Mol. Cell Biol.* **11**, 113 (2010).
9. H. Jung, C. G. Gkogkas, N. Sonenberg, C. E. Holt, *Cell* **157**, 26 (2014).
10. Y. W. Shieh *et al.*, *Science* **350**, 678 (2015).

10.1126/science.aag1039

ECOLOGY

Soil immune responses

Soil microbiomes may be harnessed for plant health

By **Jos M. Raaijmakers**[1,2] *and* **Mark Mazzola**[3]

Soil microorganisms are central to the provision of food, feed, fiber, and medicine. Engineering of soil microbiomes may promote plant growth and plant health, thus contributing to food security and agricultural sustainability (*1*, *2*). However, little is known about most soil microorganisms and their impact on plant health. Disease-suppressive soils offer microbiome-mediated protection of crop plants against infections by soil-borne pathogens. Understanding of the microbial consortia and mechanisms involved in disease suppression may help to better manage plants while reducing fertilizer and pesticide inputs.

There are two types of disease suppression in soils. General suppression is based on competitive activities of the overall micro- and macroflora and is universal to all soils.

> ***"The complexity of soil microbiome–plant interactions argues for [taking] a community perspective."***

Specific suppression is attributed to the enrichment of specific subsets of soil microorganisms. Specific suppression has been reported for plant pathogenic fungi, fungal-like oomycetes, bacteria, nematodes, and parasitic weeds. It is eliminated by soil pasteurization or biocides and can be transferred to conducive soils, in which only general suppression is operative, via soil transplantations. When Henry first reported transplantation of disease-suppressive soils 85 years ago, he elegantly showed that specific suppression of *Helminthosporium* foot rot of wheat was most likely

[1]*Netherlands Institute of Ecology, Royal Netherlands Academy of Arts and Sciences, Wageningen, Netherlands.* [2]*Institute of Biology, Leiden University, Leiden, Netherlands.* [3]*U.S. Department of Agriculture–Agricultural Research Service, Wenatchee, WA, USA. Email: j.raaijmakers@nioo.knaw.nl*

caused by the combined action of soil bacteria and fungi (*3*).

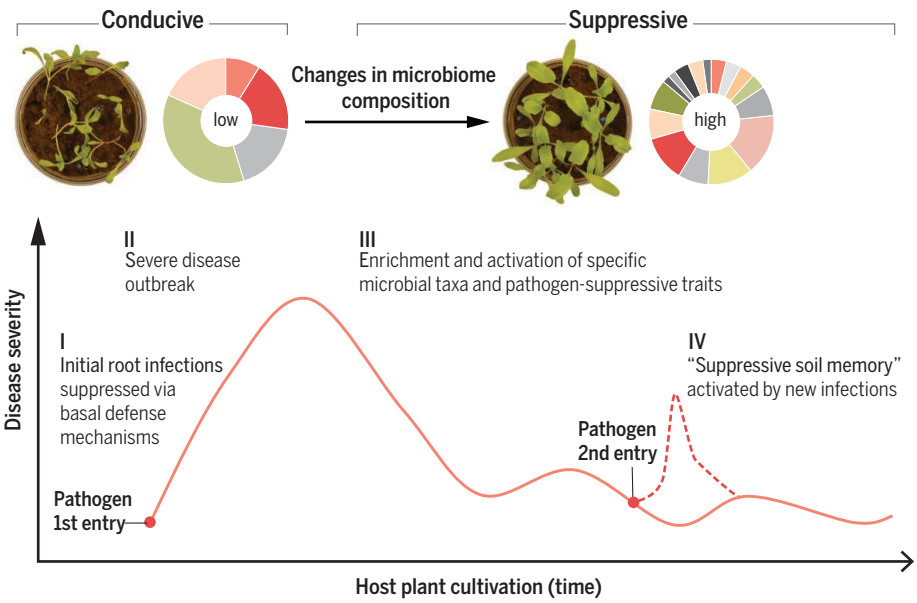
Specific suppression of various fungal root pathogens is typically induced by a disease outbreak that occurs in field soils during continuous cultivation of a susceptible host plant. Once established, specific suppression can dissipate if nonhost plants are grown or other root diseases emerge. It is rapidly regained in the presence of the original host plant and inducing pathogen (see the figure).

The characteristics of general and specific suppression of soils are comparable to those described for innate and adaptive immunity in animals. Both general suppression of soils and innate immunity in animals provide a fast, nonspecific line of defense against an invading pathogen. Both specific suppression of soils and the adaptive immune response in animals require time to react to the invading pathogen, are specific to the pathogen, and have a memory of the previously encountered pathogen (see the figure).

Specific suppression is mechanistically complex, requiring enrichment and activation of select microbial consortia and antagonistic traits that interfere with the infection cycle. Eliciting specific suppression requires multilateral interactions between pathogen, host plant, and soil microbiome. The initial interaction between pathogen and plant, leading to a disease outbreak, may cause the release of metabolites from the pathogen and the plant, which in turn enrich and activate pathogen-suppressive microbes (*4*).

Specific suppression of several fungal root pathogens has been attributed, in part, to the production of antifungal metabolites by different bacterial genera (*5*, *6*) and to carbon competition and induced systemic resistance by nonpathogenic fungi (*7*, *8*). Kinkel *et al.* have implicated *Streptomyces* species in suppression of scab, a bacterial disease of potato (*9*). Olatinwo *et al.* have proposed parasitism by the fungus *Dactylella oviparasitica* as a key mechanism in suppression of a plant pathogenic nematode (*10*). Although the interactions in soils suppressive to a specific pathogen are biologically complex, the mechanisms appear to be the same in different soils from geographically distinct regions (*11*). This functional similarity across many agroecosystems suggests that it may be possible to develop a universal approach to engineer disease-suppressive soil microbiomes.

Molecular and chemical technologies now allow identification of differences in microbiome composition between suppressive and conducive soils beyond the description of select microbial genera. They further enable comprehensive analyses of the temporal changes in microbiome activities as the soil shifts from the conducive to the disease-suppressive state. This knowledge also allows elucidation of the mechanisms that lead to the onset of specific disease suppression.



Lines of defense. If a pathogen can circumvent the basal defenses in both soil and plant, a severe disease outbreak may occur. This disease outbreak can last for years but will ultimately enrich for specific microbial consortia and pathogen-suppressive traits in the soil and plant microbiome. This specific suppression can dissipate but is rapidly regained in the presence of the original host plant and inducing pathogen. The images show plants exposed to a fungal pathogen in disease-conducive and -suppressive soils. In the conducive soil with a low abundance of antagonistic microbial consortia, the fungal pathogen causes disease (left), whereas in the suppressive soil with a high abundance of antagonistic microbial consortia, most seedlings remain healthy (right).

Studies of disease-suppressive soils have not yet yielded far-reaching solutions to soil-borne disease management and enhancing crop productivity. Rather, the main outcome has been the isolation of single microbial species subsequently applied to soil or plant seeds as biological agents for pathogen control. Many of these microbial strains fail to establish or survive in soil or on plant roots because of competition with the indigenous soil microbiome. As a result, this approach has met with limited success in large-scale agriculture. The complexity of soil microbiome–plant interactions argues for new strategies that go beyond "one-microbe-at-a-time" approaches and take a community perspective. This includes the design and application of mixtures of different microbial species, referred to as synthetic communities or syncoms (*12*). A second strategy involves augmenting indigenous disease-suppressive consortia native to the soil ecosystem. Engineering such indigenous microbial consortia could yield a more stable soil memory that limits pathogen infestations.

Practical means to attain this outcome in sustainable disease management include selection of or breeding for plant genotypes with specific root traits that recruit or activate pathogen-suppressive microbial populations (*12*, *13*). Agricultural production system inputs, including soil amendments such as compost and seed meal, can also be used, like prebiotics in humans, to selectively drive the microbiome to a composition and active state that limits proliferation of soil-borne pathogens (*14*). To this end, fundamental knowledge of coevolutionary trajectories in plant-pathogen-microbiome interactions is needed (*10*). Mechanistic understanding of specific plant metabolites and pathogen effectors that trigger, like vaccines in animals, the adaptive immune response of soils may provide practical means to engineer the indigenous soil microbiome for enhancing plant health and securing future crop yields. ■

REFERENCES

1. U. G. Mueller, J. L. Sachs, *Trends Microbiol.* **23**, 606 (2015).
2. S. F. Bender, C. Wagg, M. G. A. van der Heijden, *Trends Ecol. Evol.* **31**, 440 (2016).
3. A. W. Henry, *Can. J. Res.* **4**, 69 (1931).
4. E. Chapelle, R. Mendes, P. A. H. M. Bakker, J. M. Raaijmakers, *ISME J.* **10**, 265 (2015).
5. R. Mendes *et al.*, *Science* **332**, 1097 (2011).
6. J.-Y. Cha *et al.*, *ISME J.* **10**, 119 (2016).
7. P. Lemanceau, C. Alabouvette, *Biocontrol Sci. Technol.* **3**, 219 (1993).
8. R. P. Larkin, D. L. Hopkins, F. N. Martin, *Phytopathology* **86**, 812 (1996).
9. L. L. Kinkel *et al.*, *Res. Microbiol.* **163**, 490 (2012).
10. R. Olatinwo, B. Yin, J. O. Becker, J. Borneman, *Phytopathology* **96**, 111 (2006).
11. D. M. Weller *et al.*, *Annu. Rev. Phytopathol.* **40**, 309 (2002).
12. S. L. Lebeis *et al.*, *Science* **349**, 860 (2015).
13. M. Mazzola, D. L. Funnell, J. M. Raaijmakers, *Microbial Ecol.* **48**, 338 (2004).
14. M. Mazzola *et al.*, *Phytopathology* **105**, 460 (2015).

10.1126/science.aaf3252

GRAPHIC: N. CARY/*SCIENCE*

NANOSCIENCE

Designing a robust single-molecule switch

A single-molecule switch works at room temperature

By **C. Daniel Frisbie**

In molecular electronics, researchers combine atomic control of molecular structure with high-precision nanofabrication techniques to connect individual molecules to tiny electrodes (*1*). The aim is to understand electrical transport, such as conductivity in individual molecules, and to uncover phenomena with practical utility in nanoelectronic devices. On page 1443 of this issue, Jia *et al.* (*2*) report robust, room-temperature optoelectronic switching of a single molecule connected to conducting carbon (graphene) contacts. In terms of the magnitude of the switching effect, its reversibility, and stability at room temperature, the results represent the state of the art for single-molecule electronics.

The electronic properties of all materials, molecules included, reflect their atomic-scale architecture. Molecules thus offer potentially limitless opportunities for tailoring electronic function. In addition, their conduction mechanisms (such as quantum tunneling) can minimize heat dissipation. Combined with the compact size of most molecules, this offers the potential for higher numbers of devices per area, meaning faster computation in smaller computer chips.

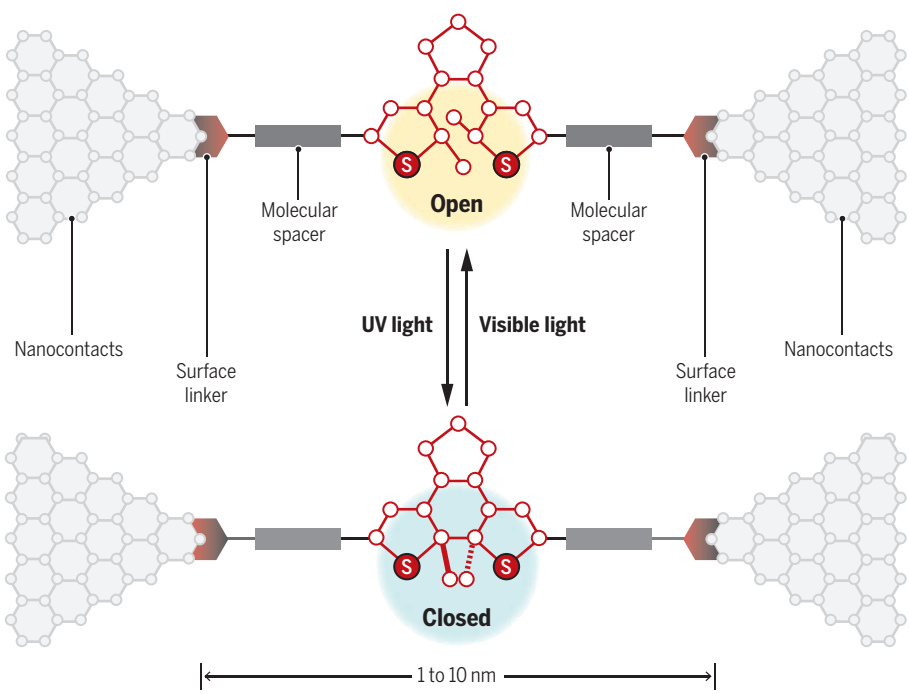
However, unlike in silicon microelectronics, there are no firmly established design rules for electronic functions in molecular electronics. Rational design means knowing how structure predicts function. To make headway, molecular electronics researchers must uncover rules for how to assemble functional molecules and how to connect them to electrodes to form junctions with predictable electrical behavior, such as current rectification or switching. Structural design parameters include the choice and location of all atoms in the molecule, the pattern of chemical bonding, the total molecular length, and the type of functional groups used to link the molecule to the electrical contacts (see the figure). Symmetry considerations are also crucial, as is the strategy for making nanoscale contacts.

Jia *et al.* report a convincing step toward rational design in molecular electronics by describing the creation of a high-performance molecular switch. The main switching component of their junction is the well-known chromophore diarylethene. The open (less conjugated) form of this molecule is insulating. Exposure to ultraviolet light drives a reversible isomerization to a more conjugated—and more conductive—closed structure; irradiation of the conductive closed state with visible light brings the molecule back to the insulating open state. Thus, incorporation of diarylethene into a solid-state molecular junction produces a light-activated toggle switch. Repeatedly exposing the junction to ultraviolet and visible light turns the current on and off.

"To make headway, molecular electronics researchers must uncover rules for how to assemble functional molecules and how to connect them to electrodes to form junctions with predictable electrical behavior, such as current rectification or switching."

Previous studies have shown this switching behavior (*3*–*5*); however, the junctions were not stable, and the on-to-off current ratio degraded over time. In the present study, Jia *et al.* achieve both high on-to-off current ratios and remarkable device stability at room temperature. They do so by using strong covalent bonds to link the molecule to the graphene nanocontacts and molecular spacers to precisely position the chromophore within the junction (see the figure). Many prior molecular junction studies have reported switching by mechanisms other than light exposure, but these were performed at cryogenic temperatures (*6*), which limits application possibilities.

The study complements a string of recent



GRAPHIC: K. SUTLIFF/*SCIENCE*

Design features for a single-molecule switch. Jia *et al.* create a room-temperature switch by linking a diarylethene chromophore to graphene nanocontacts with strong covalent bonds and positioning it precisely within the junction with molecular spacers. The switch can be turned on through irradiation with ultraviolet light; visible light switches it off again.

Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA. Email: frisbie@umn.edu

advances toward rational design in molecular electronics. These include demonstration of quantum interference effects (*7*), strong current rectification (*8*), and commercial application (*9*). In these and other reports, researchers were successful in intentionally creating molecules to produce desired electrical behaviors—a crucial step toward rational design. Theoretical models of molecular junction behavior have also converged, and the current-voltage characteristics of many simple, nonswitching junctions are well understood (*10*).

The work of Jia *et al.* also raises questions. As the authors note, the precise mechanism by which electrons are transported through the light-switchable junction is not entirely clear. The mechanisms may even be different for closed and open forms of the molecule. Understanding the mechanistic details of the transport is crucial for further improvements not just in this particular study but more broadly in molecular electronics.

As scientists become more adept at creating molecular junctions with prescribed electrical functions, are there new behaviors that can be envisioned that differ fundamentally from those seen in silicon? Recent efforts to understand spin transport (*11*), thermoelectric transport (*12*), and biomolecular electron tunneling (*13*) in molecular junctions provide possible inspiration. In the broadest sense, however, molecular electronics need not be simply about creating the next tiny switch.

"...Jia* et al. *achieve both high on-to-off current ratios and remarkable device stability at room temperature."

The science as exemplified by Jia *et al.*'s study represents exquisite control over matter at nanometer length scales and is a worthy intellectual pursuit in its own right with broad, long-term benefits.

REFERENCES

1. T. A. Su *et al.*, *Nat. Rev. Mater.* **1**, 16002 (2016).
2. C. Jia *et al.*, *Science* **352**, 1443 (2016).
3. F. Meng *et al.*, *Nat. Commun.* **5**, 3023 (2014).
4. N. Katsonis *et al.*, *Adv. Mater.* **18**, 1397 (2006).
5. X. Zhang, L. Hou, P. Samorì, *Nat. Commun.* **7**, 11118 (2016).
6. H. Song *et al.*, *Nature* **462**, 1039 (2009).
7. D. Z. Manrique *et al.*, *Nat. Commun.* **6**, 6389 (2015).
8. L. Yuan *et al.*, *Nat. Commun.* **6**, 6324 (2015).
9. A. J. Bergren *et al.*, *J. Phys. Condens. Matter* **28**, 094011 (2016).
10. I. Bâldea, Z. Xie, C. D. Frisbie, *Nanoscale* **7**, 10465 (2015).
11. S. Schmaus *et al.*, *Nat. Nanotechnol.* **6**, 185 (2011).
12. Y. Kim, W. Jeong, K. Kim, W. Lee, P. Reddy, *Nat. Nanotechnol.* **9**, 881 (2014).
13. L. Xiang *et al.*, *Nat. Chem.* **7**, 221 (2015).

10.1126/science.aag0827

APPLIED PHYSICS

Tuning organic band structures with Coulomb interactions

The smooth change of band gaps in blends of organic semiconductors arises from long-range electronic interactions

By **Nobuo Ueno**

The working principles of semiconductor devices are crucially determined by their band gap—the amount of energy needed to excite immobile charge carriers into ones that conduct current. In many inorganic semiconductors, band gaps can be tuned in a systematic way by alloying or inducing strain in the material. Although devices based on organic semiconductors are already in commercial use, there are few rational approaches for similar "band structure engineering" of these materials. On page 1446 of this issue, Schwarze *et al.* (*1*) now demonstrate, using long-range Coulomb interactions, a tuning effect of the band structure for organic semiconductors that are weakly bound by van der Waals forces. This effect, which has been totally neglected in discussions of the electronic states, is rather closely related to universal features of organic molecular crystals (*2*).

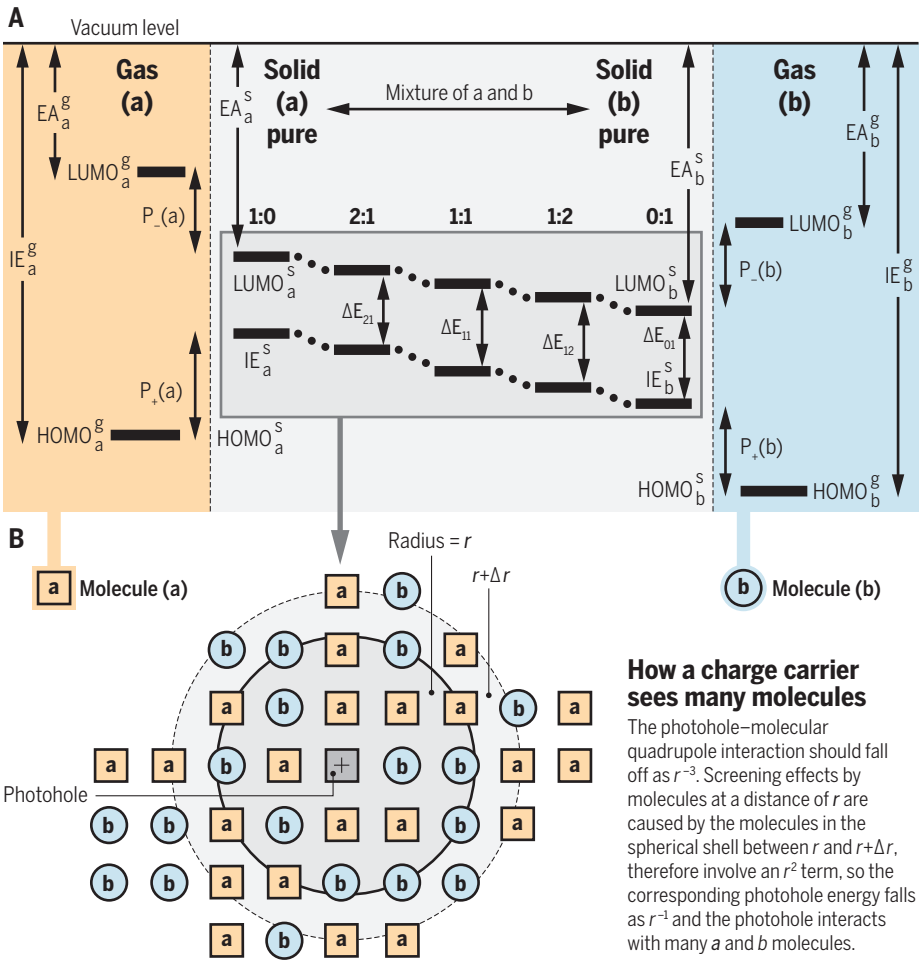
Organic semiconductors have gained increasing attention for their application in flexible large-area devices, including organic solar cells (OSCs), light-emitting diodes (OLEDs), and field-effect transistors (OFETs). An accurate energy-level tuning of electron- or hole-transporting states, which are represented by the frontier orbitals—the highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO)—is necessary for optimal device efficiency (*3*, *4*). The relative energy position of the HOMO of the donor and the LUMO of the acceptor material controls the open-circuit voltage V_{oc} in OSCs. Also, the tuning of the energy level alignment at the electrode–organic semiconductor interface is crucial for efficient charge injection in OLEDs and OFETs and extraction in OSCs (*5*).

As depicted in the figure, Schwarze *et al.* discovered a simple method of energy-level tuning of organic semiconductors over a wide energy by long-range Coulomb interactions. Charge-quadrupole intermolecular interactions provide the key for tuning the ionization energy (IE) and electron affinity (EA); the figure shows how these properties are related to the energies of the HOMO and LUMO levels. When they blended zinc phthalocyanine (ZnPc) with different fluorinated derivatives (F_xZnPc, where x can be 4, 8, or 16), they observed by ultraviolet photoemission spectroscopy that the IE shifted from 5.55 eV for ZnPc continuously with blending (up to 7.17 eV for x = 16, an energy width is as large as 1.62 eV).

In the absence of intermolecular interactions, the density of states (DOS) of the binary mixture should consist only of two peaks corresponding to the gas-phase ionization energies of ZnPc and the particular F_xZnPc used. Instead, as seen in the figure, the HOMO levels of two different molecules, ZnPc and F_{16}ZnPc, are observed at very close position on the ionization energy scale and give approximately a "single HOMO band" with greater width in the mixed film. This single HOMO moves from the ZnPc HOMO position to the F_{16}ZnPc HOMO position linearly as its fraction in the film increases. Schwarze *et al.* theoretically show that the LUMO represents a behavior similar to that of the HOMO.

Because the IE tuning originates from a superposition of long-range Coulomb interaction from two compounds, the ideal molecular alloy would be the most beneficial for the tuning. The photogenerated hole interacts with molecular quadrupoles, so all of the molecules surrounding the photohole contribute the polarization energy and shift the IE. Contributions of the molecular quadrupole to the polarization have been discussed in the study of polarization energy P_+ for the hole (IE) and P_- for the electron (EA) (*6*–*8*).

From theoretical simulations, Schwarze *et al.* found that the long-range Coulomb interactions tend to smooth out electrostatic potentials contributed from even larger domains of neat phases where the molecules have not mixed completely. Moreover, the simulations revealed that the microscopic mechanism responsible for the linear IE-tuning behavior is the superposition of quadrupolar molecular fields of two mixed compounds that are acting on the excess hole. They obtained similar simulation results for the EA, and further confirmed



Smooth changes. (**A**) Schwarze *et al*. found that mixtures of organic semiconductor molecules, 1 and 2, change their electron affinities (EA) and ionization energies (IE) smoothly; the solid-state values are offset by polarization terms. (**B**) The origin of this smooth change can be explained by considering a photogenerated charge carrier. It interacts with the quadrupole moment of many nearby molecules, sampling a population of both 1 and 2.

these phenomena with different molecular systems, such as mixtures of subphthalocyanine and its hexachlorinated derivative.

This effect can likely be exploited in a number of other molecules. The present IE tuning may have generality as a band-structure engineering method for organic semiconductors. This phenomenon is unexpected but was confirmed by the authors, who measured V_{oc} and current density–voltage curves of organic solar cells and field-effect mobility for the ZnPc-F_xZnPc blends.

Although the spatial range of this tuning effect and the required molecules for intermixing need further investigation, the present phenomena are the result of weak intermolecular interactions in organic semiconductors and the quadrupole of large-size molecules, both of which are well-known features of organic molecular solids. These studies may motivate entirely new designs of device architectures for organic semiconductors that take advantage of gap states–mediated Fermi-level tuning (*9*, *10*), doping effects (*5*), and effects of intermolecular charge transfer (*11*, *12*). ■

Graduate School of Advanced Integration Science, Chiba University, Inage-ku, Chiba 263-8522, Japan. Email: uenon@faculty.chiba-u.jp

REFERENCES

1. M. Schwarze *et al*., *Science* **352**, 1446 (2016).
2. N. Ueno, in *Electronic Processes in Organic Electronics: Bridging Nanostructure, Electronic States and Device Properties*, H. Ishii, K. Kudo, T. Nakayama, N. Ueno, Eds. (Springer, 2015), pp. 3–9.
3. A. W. Hains *et al*., *Chem. Rev.* **110**, 6689 (2010).
4. A. J. Heeger, *Adv. Mater.* **26**, 10 (2014).
5. K. Walzer *et al*., *Chem. Rev.* **107**, 1233 (2007).
6. B. J. Bounds, R. W. Munn, *Chem. Phys.* **59**, 41 (1981).
7. B. J. Topham, Z. G. Soos, *Phys. Rev. B* **84**, 165405 (2011).
8. H. Yoshida *et al*., *Phys. Rev. B* **92**, 075145 (2015).
9. T. Sueyoshi *et al*., *Appl. Phys. Lett.* **96**, 093303 (2010).
10. S. Olthof *et al*., *Phys. Rev. Lett.* **109**, 176601 (2012).
11. I. Salzmann *et al*., *Phys. Rev. Lett.* **108**, 035502 (2012).
12. H. Méndez *et al*., *Nat. Commun.* **6**, 8560 (2015).

10.1126/science.aaf8937

CELL METABOLISM

The resurgence of NAD+

Restoring a mitochondrial metabolite slows stem cell loss and aging

By **Leonard Guarente**

Interventions that can slow mammalian aging have been rare. On pages 1436 and 1474 of this issue, Zhang *et al*. (*1*) and Cambronne *et al*. (*2*), respectively, highlight nicotinamide adenine dinucleotide (NAD^+) as a major intervention point to slow or ameliorate phenotypes of aging.

NAD was discovered over a century ago, and its role in cells as a redox conduit in metabolism was subsequently established. More recently, its oxidized form, NAD^+, resurfaced as a key molecule in aging in organisms ranging from yeast to mammals by the finding that the antiaging proteins, sirtuins, are NAD^+-dependent deacylases (*3*). These proteins play a key role in mitochondrial function. Indeed, aging is also associated with loss of sirtuin and mitochondrial function.

This NAD^+-sirtuin axis plays a crucial role in maintaining health and staving off diseases of aging (*4*). The amount of cellular NAD^+ declines during normal aging, as revealed in mice engineered to overexpress the sirtuin SIRT1 in pancreatic β-cells (*5*). The novel phenotypes conferred in young mice by this overexpression, such as glucose tolerance and increased insulin secretion, were lost in old mice, but could be restored by supplementing their diet with the NAD^+ precursor, nicotinamide mononucleotide. Numerous subsequent studies of aging in worms and mice have shown that NAD^+ replenishment is associated with better metabolic health and restored mitochondrial function (*6*). One possible explanation for the NAD^+ loss during aging is that getting old is associated with the accumulation of DNA damage, which triggers chronic activation of poly(ADP-ribose) polymerases (PARPs) and the resulting depletion of the substrate used by PARP for protein PARylation, NAD^+ (*7*).

Mitochondrial dysfunction in aging mammals may be due to a disharmony between nuclear and mitochondrial gene expression, and this phenotype could be rescued by NAD^+

Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA. Email: leng@mit.edu

replenishment (*8*). This dysfunction was also demonstrated in premature aging diseases resulting from genetic defects in DNA repair, and this deficit appears to be due to NAD^+ depletion and the resulting SIRT1 inactivation. Metabolic maladies in these mice could also be corrected by supplementation with an NAD^+ precursor, nicotinamide riboside (*9*). Thus, a model emerges in which aging is associated with PARP activation, NAD^+ depletion, sirtuin inactivation, mitochondrial dysfunction, and degeneration of cells and tissues, and where this calamity can be corrected or forestalled by supplementation with NAD^+ precursors (see the figure).

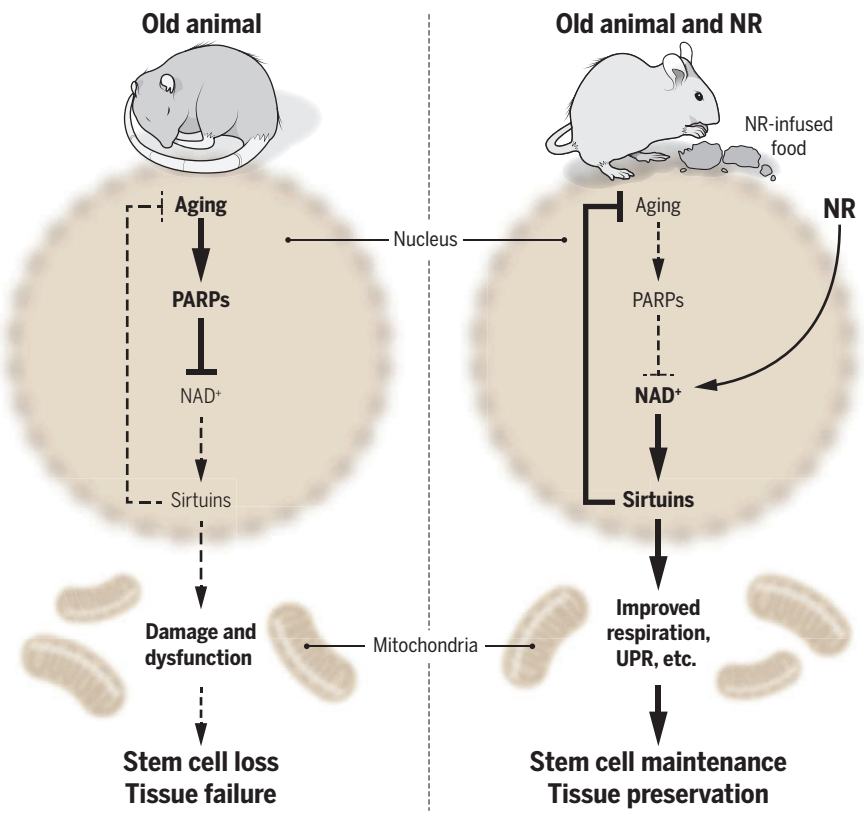
Zhang *et al.* show that adult mouse stem cells (i.e., muscle stem cells) are lost during normal aging, and this could be ameliorated by supplementing their diet with nicotinamide riboside. This salutary effect of nicotinamide riboside involved an improvement of mitochondrial function in these stem cells, including respiration, membrane potential, production of adenosine triphosphate (ATP), and the mitochondrial unfolded protein response, and did not occur in stem cells lacking SIRT1. Nicotinamide riboside also suppressed dysfunction in a mouse model of muscular dystrophy, which is consistent with earlier findings that overexpressing SIRT1 in muscle also rescued this disease model (*10*). Generalizing the findings in muscle stem cells, Zhang *et al.* also found that nicotinamide riboside countered senescence in adult neural stem cells and adult melanocyte stem cells. The loss of adult stem cells may be an important cause of organ dysfunction,

"...NAD^+ replenishment may also provide health benefits in people."

and this process may limit the life span of animals. Indeed, the authors show that nicotinamide riboside supplementation extended the life span of wild-type mice fed the normal chow diet, joining a short list of compounds with this ability (SIRT1 activators, metformin, and rapamycin). These findings suggest that preserving mitochondrial function by way of nicotinamide riboside may be a key to maintaining adult stem cells, which in turn may slow or counter degenerative effects of aging and, possibly, diseases.

Cambronne *et al.* describe an NAD^+ sensor consisting of a fusion of Venus fluorescence protein to the NAD^+ binding domain of bacterial DNA ligase. This ligase uses NAD^+ instead of ATP to activate the DNA ends for joining. The binding of NAD^+ to this hybrid sensor protein partially quenches the fluorescence signal, and this decline can be quantitatively measured. The authors targeted this NAD^+ sensor to different cellular compartments of mammalian cells and could determine the amount of NAD^+ in the local environment. As a further proof of principle, Cambronne *et al.* demonstrate that they could detect an inhibition of NAD^+ synthesis due to reducing the expression of specific biosynthetic enzymes. It will be important to determine whether the sensor can also detect the perhaps more subtle changes in NAD^+ amounts that occur as a function of diet, aging, or disease. The possibility that the NAD^+ sensor may respond to aging is exciting because a reliable and versatile biomarker for aging has long been sought without success.

As to whether NAD^+ replenishment can improve health maintenance in humans, it has been reported that cellular NAD^+ amounts decline during human aging (*11*). Also, the strict conservation in the relevant pathways of NAD^+ synthesis, sirtuins, and PARPs suggests that NAD^+ replenishment may also provide health benefits in people. Still, it will be important to test in humans whether dietary supplementation with NAD^+ precursors will raise cellular NAD^+ concentrations sufficiently to compensate for the loss due to aging. If so, it will also be necessary to test, in rigorously controlled trials, whether raising NAD^+ concentrations improves health parameters, such as blood glucose and lipid profile, as well as inflammation. More expanded trials could measure effects on bone density, endothelial cell function, muscle mass, or even cognition. If NAD^+ precursors can positively affect health parameters, it is reasonable to anticipate their place at the table alongside more traditional pharmaceutical drugs in the treatment of diseases. ■



Restorative intervention. In old mice, aging may trigger NAD^+ loss by activating poly(ADP-ribose) polymerases (PARPs). NAD^+ reduction lowers the activity of the antiaging proteins, sirtuins, leading to a feedforward cycle of aging. Nicotinamide riboside (NR) restores NAD^+ amounts and reverses this cycle, resulting in better stem cell maintenance and tissue function. UPR, unfolded protein response.

GRAPHIC: C. BICKEL/*SCIENCE*

REFERENCES AND NOTES

1. H. Zhang *et al.*, *Science* **352**, 1436 (2016).
2. X. A. Cambronne *et al.*, *Science* **352**, 1474 (2016).
3. S. Imai, C. M. Armstrong, M. Kaeberlein, L. Guarente, *Nature* **403**, 795 (2000).
4. L. Guarente, *Genes Dev.* **27**, 2072 (2013).
5. K. M. Ramsey, K. F. Mills, A. Satoh, S. Imai, *Aging Cell* **7**, 78 (2008).
6. J. Yoshino, K. F. Mills, M. J. Yoon, S. Imai, *Cell Metab.* **14**, 528 (2011).
7. L. Mouchiroud *et al.*, *Cell* **154**, 430 (2013).
8. A. P. Gomes *et al.*, *Cell* **155**, 1624 (2013).
9. E. F. Fang *et al.*, *Cell* **157**, 882 (2014).
10. A. Chalkiadaki, M. Igarashi, A. S. Nasamu, J. Knezevic, L. Guarente, *PLOS Genet.* **10**, e1004490 (2014).
11. H. Massudi *et al.*, *PLOS ONE* **7**, e42357 (2012).

ACKNOWLEDGMENTS

L.G. is a founder of Elysium Health and on the scientific advisory boards of GlaxoSmithKline, Sibelius, and InsideTracker.

10.1126/science.aag1718

TECHNOLOGY, SECURITY, AND LAW

The real security issues of the iPhone case

Law enforcement needs 21st-century investigative savvy

By **Susan Landau**

The U.S. Federal Bureau of Investigation (FBI) pitched its recent court battle with Apple as one of national security versus privacy. But in the conflict over public use of cryptography, decades in the making, this latest battle is actually a collision over differing notions of security in a digital age (*1*). Law enforcement contends that there should be no "warrant-proof" spaces: When there is a valid court order, agents should be able to access communications devices and systems. Technologists (*2*) and many former government security officials (*3*–*5*) see this differently: Weakening smartphones' security provided by encrypted communications is counterproductive to long-term security. The ability to secure data on smartphones is crucial not just for the private information present on phones but also because of the ability of smartphones to securely authenticate users to online accounts. Rather than rely on out-of-date approaches to law enforcement, the FBI must develop 21st-century investigative capability.

POLICY

Worcester Polytechnic Institute, Worcester, MA 01690, USA.
Email: susan.landau@privacyink.org

UPDATES AND AUTHENTICATION. The FBI wanted Apple to use a software update to suppress security protections for a terrorist's phone. The FBI stated that Apple would control the update software, which would be tailored to work solely on the one phone and deleted after use. But the FBI had 11 more phones to unlock, and the Manhattan district attorney had 200. It was unlikely that the software would be developed, used, and deleted. Updates to unlock phones may have become routine.

Thus, the software that the FBI wanted Apple to create would provide a key to open not just a single house but millions of homes (*6*). Apple cryptographically "signs" updates to prevent others from presenting malware masquerading as an update. A process that happens rarely—currently the case for signing updates—can be carefully scrutinized; the chance for malfeasance is low. If it becomes routine to use signed updates to remove security features of a particular phone, the process is prone to be less careful. Some neglect in the process or the collaboration of a rogue employee would make it easy for false requests to be slipped into the update queue.

There are other problems as well. Misusing the update process is apt to create user distrust. Might people stop using automatic updates if these were a surreptitious technique to search devices, not for terrorist activity but for tax fraud? Could installed malware steal bank account information? An FBI "solution" that decreases the use of automatic patch updates would have devastating security effects.

These security concerns pale in the face of the authentication issue. A National Security Agency (NSA) official noted that login credentials, especially of those with high levels of network access, are attackers' most valuable data (*7*). Once an adversary has those credentials, he or she has a beachhead from which they can exfiltrate data or attack your system. Indeed, in December 2015, attackers used system operators' login credentials to bring down the Ukrainian power grid (*8*).

Although static passwords are easily stolen, and typically discovered only after an account has been compromised, smartphones don't easily go missing without being noticed. They enable dynamic responses, making the authentication process more robust. Smartphones are already being used to provide two-factor authentication. You log onto your account with a password, and a text message to your phone provides a one-time personal identification number that you type in. Because your smartphone is a device that you carry all the time, the process is simple. The combination of convenience and security is winning; Facebook (*9*), Google (*10*), and some high-placed U.S. government agencies are using smartphone solutions for authentication.

But both the authentication applications and the phones must be secure. If phones' protection mechanisms are weakened, an attacker could more easily alter data anywhere in the phone. Smartphones' viability as trusted authenticators would be sharply reduced, and "interesting targets"—the secretary to the vice president of operations of an oil and gas company; the heating, ventilation, air conditioning supplier to a power plant; and the like—would be at risk.

OUT-OF-DATE LAW ENFORCEMENT. In the 1990s, FBI and NSA interests in limiting use of encryption were aligned; export-control restrictions limited foreign usage and slowed domestic deployment. The agencies had differing views, in part because of their differing missions (NSA's includes strong information assurance, whereas the FBI's does not) and differing levels of technical expertise. In the late 1990s, the NSA was unprepared for changing communications technologies—for example, the volume of email, the use of fiber-optic cables (harder to tap than radio transmissions), and the use of strong cryptography by European, Asian, and third-world governments (*11*). The agency adapted.

Meanwhile, the FBI fought hard to maintain investigative capabilities honed during analog telephony's heyday. Not surprisingly, such an approach has created problems. The 1994 Communications Assistance for Law Enforcement Act (CALEA) requires that digitally switched telephone networks be built "wiretap enabled." This mandates vulnerabilities, providing multiple ways for nefarious sorts to take advantage—for example, wiretapping of the cell phones of senior members of the Greek government in 2004–2005 (*12*). Several Cisco systems were deployed with configurations for wiretapping Internet Protocol–based communications that would enable a hacker to wiretap at the switch (*13*). In tests of CALEA-compliant switches before their use in Department of Defense systems, NSA discovered security problems in every single switch submitted (*14*).

CALEA is part of a larger "Going Dark" scheme for digital devices under which law enforcement has sought exceptional access (strong security with access solely for law enforcement) (*15*). Such design requirements preclude strong security, eliminating forward secrecy (decryption keys are deleted after a single use) and authenticated encryption (encryption of the communication and authentication of the sender are done in one step) (*2*).

> ***"Technology that helps secure the many does not preclude a targeted attack against a specific device...."***

U.S. restrictions on encryption would accomplish little. A recent survey revealed 865 hardware and software encryption products available in various jurisdictions; not all prevent law-enforcement access (*16*). Controlling cryptography's use would not prevent sophisticated bad actors from communicating securely but would make it harder for the rest of us to secure ourselves.

OLD PROBLEMS, NEW SOLUTIONS. I, and many others, argued that the FBI did not need Apple to undo the protections of the phone and that the FBI could find other ways of accessing the information (*3*). The FBI strongly disputed that point but ultimately paid a third party who broke through Apple's security protections, providing access to the phone's contents. Technology that helps secure the many does not preclude a targeted attack against a specific device, particularly one of which you have physical possession.

My coauthors and I described how law enforcement can use vulnerabilities in software—including phones and computers—to install a remote wiretap, even when communications are encrypted from sender to receiver (*17*). Indeed, the FBI has used "lawful hacking" since 2003 (*18*).

In seeking to have Apple develop software enabling third-party access to a secured iPhone 5c, the FBI was seeking to weaken security protections. Had Apple developed such software, every iPhone 5c would have been at risk (and perhaps other iPhones as well). The FBI would have been creating a CALEA applied to devices, not only weakening the phones' protections but also eliminating our best possibility for secure authentication. That's a very poor approach to security.

The FBI has some excellent capabilities in cyber investigations, but not at the scale and level for solving today's problems. The FBI's Going Dark program, responsible for lawful hacking and technical challenges posed by encryption, anonymization, and the like, currently has 39 positions (11 agents) and a budget of $31 million; the 2017 budget requests an increase to $38.3 million but no additional positions (*19*). By comparison, the FBI's physical surveillance effort has 549 agents and a $297.8 million budget (*19*). The inadequacy of the Going Dark effort may go a long way toward explaining the FBI's current view of encrypted communications and secured devices. One forensics researcher has observed that in recent years, "an uncomfortable number [of investigators] have regressed to a state of 'push button forensics'" (*20*).

Law enforcement's solution to an inadequate effort is that technology companies should weaken security. In congressional testimony, I recommended vastly improved FBI tools and capabilities (*1*), a suggestion since picked up by the Congressional Research Service (*21*). The FBI needs an investigative center with agents who have deep technical understanding of modern telecommunications technologies and computer science. The FBI will need applied researchers to handle various types of fielded devices. The NSA has this expertise, but the FBI brings cases to court, which risks the tools being revealed. Thus, the FBI will have to develop its own. This FBI center should develop surveillance technologies to match the direction of new communications technologies.

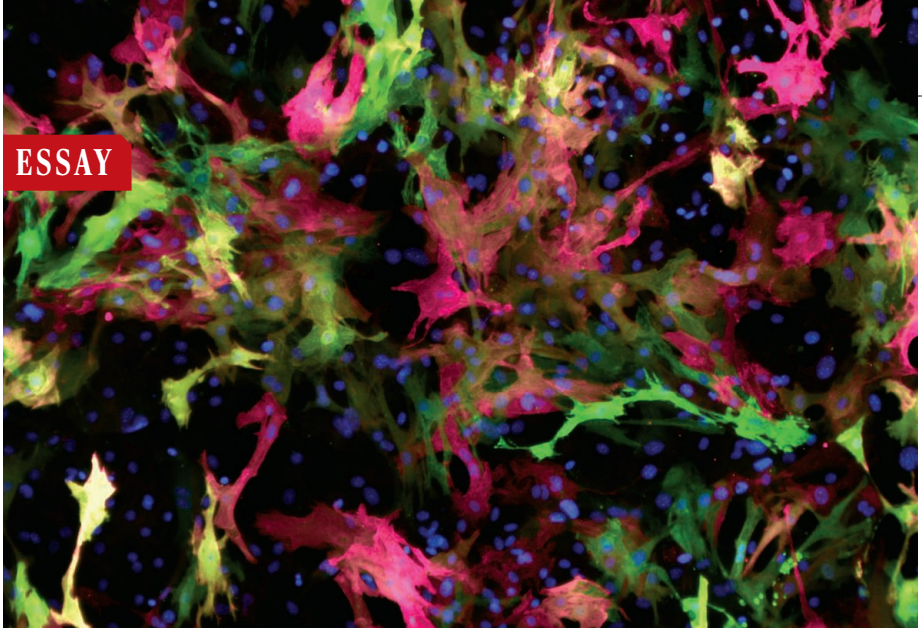
Outside consultants may prove useful. But unlike the Apple case, where the FBI does not know how the tool it purchased works (*22*), the FBI should have full technical understanding of tools developed by its consultants. Anything less undermines trust in the evidence and risks leaving the vulnerability open for others to exploit. NSA has a "vulnerabilities equities process" that determines under which circumstances vulnerabilities are shared with manufacturers so that systems can be patched (*23*). The process for the FBI, which operates domestically and thus under a different set of equities, must be developed (*17*). Also to be resolved is the process through which FBI capabilities are shared with state and local police, who are largely not in a position to develop their own.

The FBI must develop 21st-century investigative savvy. This will require government investment, but the alternative, of permitting bad actors access to our systems, is unacceptable. The FBI should be urging manufacturers to increase the security of their devices. As for the Apple phone, the FBI should not be undermining the best security that any consumer device has to date. ■

REFERENCES

1. This article is based on S. Landau, Testimony, House of Representatives Committee on the Judiciary, "The encryption tightrope: Balancing Americans' security and privacy," 1 March 2016.
2. H. Abelson *et al.*, *J. Cybersecurity* **1**, 69 (2015).
3. "Encryption, privacy are larger issues than fighting terrorism, Clarke says," National Public Radio, 14 March 2016.
4. M. Hayden, "General Michael Hayden on Apple, the FBI, and data encryption" [blog] (AEIdeas, American Enterprise Institute, 23 March 2016); https://www.aei.org/publication/gen-michael-hayden-on-apple-the-fbi-and-data-encryption/.
5. M. McConnell, M. Chertoff, W. Lynn, "Why the fear over ubiquitous data encryption is overblown," *Washington Post*, 28 July 2015.
6. Apple, Answers to your questions about Apple and Security; www.apple.com/customer-letter/answers/.
7. R. Joyce, Talk at USENIX ENIGMA (2016); https://www.youtube.com/watch?v=bDJb8WOJYdA.
8. Department of Homeland Security, "Alert (IR Alert-H-16-056-01): Cyberattack against Ukrainian critical infrastructure" (DHS, Washington, DC, 25 February 2016).
9. Duo Security, *Facebook's Security Philosophy, and How Duo Helps* (Duo Security, Ann Arbor, MI, 2016).
10. S. Perez, "Google begins experimenting with password-free logins" (TechCrunch.com, 2015); http://techcrunch.com/2015/12/22/google-begins-testing-password-free-logins/.
11. S. Hersh, "The intelligence gap," *The New Yorker*, 6 December 1999.
12. V. Prevelakis, D. Spinellis, *IEEE Spectrum* **44**, 26 (2007).
13. T. Cross, "Exploiting lawful intercept to wiretap the Internet" [slides], Black Hat DC 2010, 31 January to 3 February 2010, Crystal City, VA (www.blackhat.com, 2010).
14. S. Landau, *J. Telecom. High Tech. Law* **11**(1), 1 (2013).
15. J. Comey, "Going dark: Are technology, privacy, and public safety on a collision course?" [transcript], Brookings Institution, 16 October 2014, Washington, DC, 2014.
16. B. Schneier, K. Seidel, S. Vijaykumar, "A worldwide survey of encryption products: Version 1.0" (Schneier.com, 2016).
17. S. M. Bellovin, M. Blaze, S. Clark, S. Landau, *Nw. J. Tech. Intell. Prop.* **12**(1), art. 1 (2014).
18. M. Apuzzo, "FBI used hacking software decade before iPhone fight," *New York Times*, 13 April 2016, p. B1.
19. Federal Bureau of Investigation, "FY 2017 Budget request at a glance" (FBI, Washington, DC, 2016).
20. J. Zdziarski, "Open letter to Congress on encryption back doors" (2016); www.zdziarski.com/blog/?p=6058#more-6058.
21. R. M. Thompson II, C. Jaikaran, "Encryption: Selected legal issues" (R44407, Congressional Research Service, Washington, DC, 2016).
22. E. Nakashima, "Comey defends FBI's purchase of iPhone hacking tool," *Washington Post*, 11 May 2016.
23. M. Daniels, "Heartbleed: Understanding when we disclose cyber vulnerabilities" [blog] (Whitehouse.gov, 2014); https://www.whitehouse.gov/blog/2014/04/28/heartbleed-understanding-when-we-disclose-cyber-vulnerabilities.

10.1126/science.aaf7708



ESSAY

Reprogrammed iCMs in a dish. Green, cardiac reporter αMHC-GFP; red, cardiac TroponinT (cTnT); blue, DAPI-labeled nuclei.

REGENERATIVE MEDICINE

Hope for the brokenhearted

Cellular reprogramming improves cardiac function in a mouse model of myocardial infarction

By **Li Qian**

Cellular reprogramming, the conversion of one type of somatic cell into another, has substantially affected the field of stem cells and regenerative medicine in the past decade. Today, it holds great promise as a novel approach for the treatment of various human diseases and as a tool for disease modeling to advance personalized medicine.

Inspired by the paradigm-shifting method, developed by Shinya Yamanaka, of creating induced pluripotent stem cells (iPSCs), and with the full support of my postdoctoral mentor Deepak Srivastava, in the fall of 2009 I undertook a study to determine whether resident cardiac fibroblasts could be converted into functional cardiomyocytes. After overcoming numerous technical and conceptual obstacles, I demonstrated that endogenous cardiac fibroblasts can be reprogrammed into induced cardiomyocytes (iCMs) in their native environment, and that this conversion leads to a functional improvement associated with scar size reduction in an infarcted heart (*1*). This research was ranked number 2 in the 2012 list of Top 10 Advances in Heart Disease and Stroke Research, compiled by the American Heart Association. From this research, we developed a system in which cardiac reprogramming could be rigorously studied and implemented (*1*–*3*).

Trained as a basic developmental and cell biologist under the mentorship of Rolf Bodmer, I have always believed in the power of basic science to advance translational research. I started my own laboratory at the University of North Carolina (UNC)–Chapel Hill in the fall of 2012. Now as a principal investigator of my own team, I rely on my training in both translational and basic science to tackle reprogramming questions by exploring the fundamental molecular mechanisms of this fascinating process.

Leveraging the knowledge that faithful cell fate conversion requires a precise dosage and temporal expression of transcriptional factors, my laboratory recently identified the optimal ratio of cardiac reprogramming factors Gata4 (G), Mef2c (M), and Tbx5 (T) for more complete and efficient iCM generation (*4*). In this study, we generated a complete set of polycistronic constructs containing G, M, and T in all possible splicing orders with identical 2A sequences in a single mRNA. We found that each splicing order gave rise to a distinct ratio of G, M, and T protein expression and resulted in significantly different reprogramming efficiencies. The most desirable combination promoted a significant increase in the generation of beating iCMs when compared with G, M, and T delivered via pooled viruses. Importantly, at the molecular level the most optimal G, M, and T stoichiometry, defined by higher protein expression level of M and lower levels of G and T, promoted significantly increased expression of mature cardiomyocyte markers. We went on to demonstrate that when the polycistronic vector that induced the optimal ratio of G, M, and T—which we named MGT—is delivered in vivo to an infarcted heart, it improves iCM reprogramming efficiency and results in a further improvement in ventricular contractile function (*5*). This work resulted in a series of recent publications and numerous inquiries from around the world to obtain our new tool set for cardiac reprogramming (*4*–*6*).

With the optimal MGT system, we continued our efforts to identify roadblocks of iCM reprogramming. Because cell reprogramming involves substantial chromatin reorganization, it will be of major importance to identify the epigenetic mechanisms that govern this process. Using our polycistronic system, we were able to determine the dynamics of histone marks such as H3K27me3 and H3K4me3 in parallel with gene expression at a set of carefully selected cardiac and fibroblast loci during iCM reprogramming. Additionally, we determined the DNA methylation states of representative cardiac promoters and found that CpG sites were unequally demethylated during early stages of iCM reprogramming. Thus, we propose that there are specific CpGs, whose demethylation states correlate with transcription activation, that play more prominent roles than others. Our study also revealed that the reprogramming fibroblasts rapidly activated the cardiac program while progressively suppressing fibroblast fate at both epigenetic and transcriptional levels. This research provides the first insight into epigenetic repatterning events during cardiac reprogramming (*7*).

Efficiency of cardiac reprogramming can be further boosted by removing a key epigenetic barrier, which we recently identified through a novel loss-of-function screen. Knockdown of the polycomb ring finger oncogene *Bmi1* early in the reprogramming process resulted in the most significant enhancement of iCM induction. This inhibitory role was not associated with its well-known function in regulating cell proliferation through genes such as *p16Ink4a*, *p19Arf*, and *p53*. Instead, we found that Bmi1 directly bound to the regulatory regions of several cardiogenic genes, including *Gata4*, *Nkx2-5*, *Isl1*, *Pitx2*, and *Tbx20*. Knockdown of *Bmi1* resulted in a marked decrease in repressive histone mark H2AK119ub and an increase in active histone mark H3K4me3 at these loci

Department of Pathology and Laboratory Medicine, University of North Carolina, Chapel Hill, NC 27599, USA. Email: li_qian@med.unc.edu.

PHOTO: L. WANG

and derepressed these cardiogenic genes during iCM reprogramming. We further demonstrated that Bmi1 depletion substituted for Gata4 in reprogramming fibroblasts into beating iCMs. Thus, our findings implicate Bmi1 as a critical epigenetic barrier to cardiac reprogramming. Furthermore, our studies reveal a previously uncharacterized function of Bmi1 in repressing the cardiogenic program and demonstrate that removing a specific epigenetic barrier simplifies iCM generation and increases yield, potentially streamlining iCM production for therapeutic purposes (*8*).

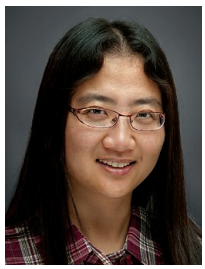
My team and I will continue to develop and use new reprogramming approaches to advance personalized medicine. I believe that our efforts and the efforts of others will one day lead to tailored therapies designed for individual patients.

PHOTOS: (TOP TO BOTTOM) C. GOODFELLOW; SHAY HERMAN

REFERENCES

1. L. Qian *et al.*, *Nature* **485**, 593–598 (2012).
2. L. Qian *et al.*, *Nature Protoc.* **8**, 1204 (2013).
3. L. Qian, D. Srivastava, *Circ. Res.* **113**, 915 (2013).
4. L. Wang *et al.*, *Circ. Res.* **116**, 237 (2015).
5. H. Ma *et al.*, *Cardiovasc. Res.* **108**, 217 (2015).
6. L. Wang *et al.*, *J. Vis. Exp.* **105**, e53426 (2015).
7. Z. Liu *et al.*, *Stem Cell Res.* **16**, 507 (2016).
8. Y. Zhou *et al.*, *Cell Stem Cell* **18**, 382 (2016).

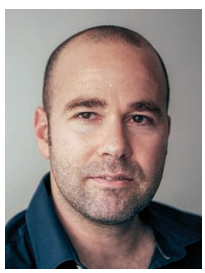
10.1126/science.aag1213



WINNER

Li Qian

Li Qian received her undergraduate degree from Fudan University and a Ph.D. from the University of Michigan, Ann Arbor. She completed her postdoctoral training at Gladstone Institutes, University of California, San Francisco, in July 2012. As an Assistant Professor at UNC–Chapel Hill, Qian is currently exploring the reprogramming approaches for cardiac regeneration and disease modeling.



FINALIST

Yosef Buganim

Yosef Buganim received undergraduate degrees from Bar-Ilan University and a Ph.D. from the Weizmann Institute of Science. As a postdoctoral fellow at the Whitehead Institute for Biomedical Research at MIT, he used single-cell technologies and bioinformatic approaches to shed light on the molecular mechanisms that underlie the reprogramming of somatic cells to iPSCs. Currently the leader of his own laboratory at The Hebrew University of Jerusalem, Buganim uses somatic cell conversion models to identify and investigate the elements that facilitate safe and complete nuclear reprogramming.

http://science.sciencemag.org/content/352/6292/1401

ESSAY

REGENERATIVE MEDICINE

Back to basics

Refined nuclear reprogramming techniques yield higher-quality stem cells

by **Yosef Buganim**

The "hyperdynamic" chromatin state, which characterizes embryonic stem cell (ESC) epigenetic state, facilitates a rapid and efficient reaction to external and internal cues. These cues lead to the activation of key master regulators that drive the cells into their developmental fate. Thus, ESCs hold great promise as a source of diverse differentiated cell types for cell therapy and regenerative medicine.

Two major bottlenecks to realizing such potential are allogenic immune rejection of ESC-derived cells and ethical concerns related to the use of fetal tissues. In 2006, two Japanese scientists, Kazutoshi Takahashi and Shinya Yamanaka, made progress toward addressing both of these hurdles when they showed that introducing four transcription factors—Oct4, Sox2, Klf4, and Myc (OSKM)—can reprogram fibroblasts derived from adult mice into functional embryonic stem-like cells, known as induced pluripotent stem cells (iPSCs). (Resetting the epigenome of a somatic cell to a pluripotent state has already been achieved with somatic cell nuclear transfer, but this type of transformation is mediated by the actions of a large number of genes expressed within the host oocyte.)

The notion that as few as four factors are sufficient to reset the epigenome of a cell has inspired scientists to attempt to convert different adult cells into other somatic cell types by avoiding the pluripotent state, using a specific combination of key master regulators. Several subsets of cell types—including hematopoietic cells, various neuronal cells, cardiomyocytes, hepatocytes, embryonic Sertoli cells, endothelial cells, and retinal pigment epithelial cells—have been generated from somatic cells by using the direct conversion approach pioneered by Takahashi and Yamanaka.

Unfortunately, the majority of directly converted cells are not stable and represent mostly incomplete reprogramming states, and the vast majority of iPSCs exhibit poor potential for incorporation into all developing tissues in mice (*1*). This suggests that the current prevailing reprogramming methods are not ideal and must be improved before we consider using these cells in the clinic.

During my postdoctoral training in the laboratory of Rudolf Jaenisch at the Massachusetts Institute of Technology (MIT), I explored the reprogramming process at the single-cell level. I analyzed the expression of 48 genes in single mouse embryonic fibroblasts (MEFs) at various stages during the reprogramming process (*2*). Analysis of early stages revealed considerable variation in gene expression between cells, in contrast to late stages, which exhibited more homogeneous expression between reprogrammable cells. We found that cells that expressed *Esrrb*, *Utf1*, *Lin28*, and *Dppa2* were more likely to become iPSCs than were those that expressed previously suggested reprogramming markers such as *Fbxo15*, *Fgf4*, and *Oct4*. We also noticed that stochastic gene expression early in reprogramming is followed by a late deterministic phase, in which *Sox2* is the upstream factor in the gene expression network. Exploiting this network allowed us to reprogram MEFs to iPSCs in the absence of "Yamanaka" key pluripotency genes and Nanog, a transcription factor that had previously been considered crucial for reprogramming, demonstrating for the first time that the core pluripotency circuitry can be activated via different entries.

In a follow-up study, we reasoned that a combination of key factors derived from the later phase of reprogramming would convert cells in a more controlled way and therefore might uniformly yield iPSCs of high quality. We chose *Sall4*, *Esrrb*, and *Lin28*—key downstream players during the late reprogramming phase—and *Nanog* for our initial experiment. We observed that ectopic expression of *Sall4*, *Nanog*, *Esrrb*, and *Lin28* (SNEL) in MEFs generated iPSCs of superior quality to iPSCs derived from OSKM (*3*). SNEL-iPSCs contributed to high-grade chimeras and produced "all-iPSC" mice by means of 4n complementation at a significantly higher frequency than did OSKM-derived iPSCs. These results demonstrate that bioinformatic models derived from in vitro single-cell data can aid in improving the quality of iPSCs.

In January 2014, I became an independent faculty member at the Hebrew University of Jerusalem. My first goal was to determine whether a high degree of nuclear reprogramming can be achieved even in cells undergoing conversion to nonpluripotent cells. Indeed, in a paper that we published in *Cell Stem Cell* (*4*), we showed for the first time that a high nuclear resetting state can be attained in cells undergoing conversion to induced trophoblast stem cells (iTSCs) and that this complete nuclear reprogramming process is independent from the pluripotent state. The iTSCs resemble blastocyst-derived trophoblast stem cells in all examined parameters, including transcription, methylation, epigenome, and function. As such, we believe that these cells hold great promise for modeling and treating placental dysfunction diseases and for mitigating recurrent miscarriage.

Overall, the main goal of my laboratory is to identify and investigate the components that regulate cell plasticity and epigenome resetting. To that end, we study several con-

FINALIST

Yosef Buganim

Yosef Buganim received undergraduate degrees from Bar-Ilan University and a Ph.D. from the Weizmann Institute of Science. As a postdoctoral fellow at the Whitehead Institute for Biomedical Research at MIT, he used single-cell technologies and bioinformatic approaches to shed light on the molecular mechanisms that underlie the reprogramming of somatic cells to iPSCs. Currently the leader of his own laboratory at The Hebrew University of Jerusalem, Buganim uses somatic cell conversion models to identify and investigate the elements that facilitate safe and complete nuclear reprogramming.

PHOTO: SHAY HERMAN

version models in order to identify common and more global elements that facilitate nuclear reprogramming and improve the quality of converted cells. We investigate both nearly complete conversion models, such as conversion of fibroblasts into iPSCs and iTSCs, together with less stable, partial-reprogramming models, such as the conversion of somatic cells into induced embryonic Sertoli-like cells, induced neurons, and induced hepatocytes. To capture rare reprogrammable cells, we have established triple/quadruple fluorescent knock-in reporter systems, using the CRISPR/Cas9 technique, to mark the cells that are destined to become converted. We have applied cutting-edge single-cell technologies such as RNA-sequencing, Fluidigm-BioMark, and single-molecule-mRNA–fluorescence in situ hybridization to probe the transcriptome of multiple individual reprogrammable cells from the various conversion models. To understand how different key master regulators define a new epigenome and reshape the chromatin, we have performed chromatin immunoprecipitation–sequencing on various transcription factors and histone marks from different models. These approaches allow us to dissect the most global and fundamental events that facilitate complete nuclear reprogramming and will ultimately improve the quality of iPSCs and converted cells available for clinical use.

REFERENCES

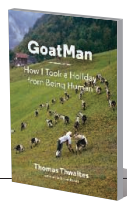
1. S. Sebban, Y. Buganim, *Trends Cell Biol.* **26**, 65–75 (2016).
2. Y. Buganim *et al.*, *Cell* **150**, 1209–1022 (2012).
3. Y. Buganim, *Cell Stem Cell* **15**, 295–309 (2014).
4. H. Benchetrit *et al.*, *Cell Stem Cell* **17**, 543–556 (2015).

10.1126/science.aag1215

Thomas Thwaites does his best impression of a goat in the Swiss Alps.

Being a Beast
Adventures Across the Species Divide
Charles Foster
Metropolitan Books, 2016. 251 pp.

GoatMan
How I Took a Holiday from Being Human
Thomas Thwaites
Princeton Architectural Press, 2016. 208 pp.



ANIMAL BEHAVIOR

Wild things

Two authors embrace unusual methods to shed light on the lives of other species

Reviewed by **Carolyn A. Ristau**

"When you bite into a worm for the first time … nip it genteelly with your incisors," instructs Charles Foster in *Being a Beast: Adventures Across the Species Divide*. Foster is a fellow at Green Templeton College, Oxford, teaching medical law and ethics. He is also a veterinarian, a barrister, and a philosopher. Sometimes, he also lives like a badger. In the book, he describes his attempts to live not only as a badger but also as an otter, an urban London fox, a red deer, and a swift. "I want to know what it is like to be a wild thing," he writes. Would an animal who lived in this wood "recognize our description of the wood?" Is our world his world?

Such inquiries have a long tradition in the natural sciences, having been raised by Charles Darwin in *The Expression of the Emotions in Man and Animals* and emphasized by Jakob von Uexküll, whose theory of "Umwelt" described how species sense, prioritize, and interact with the environment differently. These ideas were later dismissed by Thomas Nagel in "What is it like to be a bat?" (we can't ever know) and reclaimed as an avenue of scientific inquiry in Donald R. Griffin's groundbreaking, though controversial, *The Question of Animal Awareness*.

In *Being a Beast*, informed by scientific research on animal physiology and behavior, Foster and his son live, badger-like, in a trench in the woods for 6 weeks, scratching in the dirt, crawling close to the earth, and eating worms that they find there. The swifts inspire his awe, although his attempts to fly with them fail awkwardly. To sense like the bird, he stands naked in the wind atop the moor. The swift's minute filoplumes would convey essential information on feather orientation to its brain; Foster's body hair ruffles empathetically. The urban fox engenders his deep respect: choosing to hunt, even though it could survive on pizza scraps. In contrast, humans seem "sclerosed superspecialists."

Foster has created an enthralling and deeply reflective book. His prose is not merely apt but charming, witty, and poetic, capturing the imagined state and being of animals and of us humans as well. This reader was anxious, reading of the otter living on the metabolic edge. The hypertensive red deer, alert for a wolf, aroused similar emotions.

We are drawn into a deeper cultural discussion through Foster's allusions to history, philosophy, political systems, music, and literature and references to Descartes, Messiaen, Jung, C. S. Lewis, Heraclitus, Darwin, and the Emperor Augustus, among others. Intertwined with his riveting first-hand observations of nature are issues of "other minds," of animal cognitive and emotional capacities, of suffering, and of conservation.

Thomas Thwaites's *GoatMan* is a shorter book, lighter and more humorous, that delves into similar issues. Thwaites's impetus for embarking on a quest to become a goat is "to have a holiday from *being human*," to live without past regrets or worries for the future.

Thwaites, with a background in biology and a degree in design, originally conceived of the project as one in which he would transform himself into an elephant and traverse the Alps. A Wellcome Trust grant finally materialized, the award committee's confidence based, no doubt, on Thwaites's previously well-received project in which he created a toaster from scratch (*1*). After visiting real elephants and consulting a shaman, he realized that becoming a goat was more practicable.

Thwaites begins by recounting his first attempt to design prosthetic goat legs, which he hopes will enable four-legged movement and sufficient flexibility to eat grass. He visits a motion study laboratory and offers an engrossing discussion of the mechanics and evolution of human and animal movement.

At a gut ecosystems laboratory, he searches for information that will help him create a synthetic stomach that will allow him to digest grass. (The lab is researching the possibility of creating modified livestock stomachs to reduce climate-changing methane burps.) Further exploits, all with fascinating scientific renderings, take him to a prosthetics facility, a goat sanctuary, and several farms.

Clothed in his mom's homemade "goat coat," he eventually heads for the Alps. Does he traverse them? Does he "hang out" with the alpine goats? No spoiler, I.

Being a Beast, particularly, held me mesmerized. Foster has, for decades, been an observer and participant in animal lives. His immersion into nature reveals itself in the gripping details of animals and their environments. The goat project, in contrast, may reflect the beginning of an artist's similar journey. Unlike Foster's emphasis on the mentation of his beasts, Thwaites is focusing on achieving the physicality of the goat.

Still, both books have important messages. Of Thwaites's: Be more playful; take some risks and explore the (real and imagined) world. Of Foster's: Use your capacities, expand them, engage with the wondrous natural world. Do appreciate the rich and complex experiences of the other beings that live alongside us.

The reviewer is affiliated with the Columbia University Seminars, Columbia University, New York, NY 10027, USA. Email: carolynaristau@aol.com

REFERENCES

1. T. Thwaites, *The Toaster Project: Or a Heroic Attempt to Build a Simple Electric Appliance from Scratch* (Princeton Architectural Press, New York, 2011).

10.1126/science.aaf8235

PHOTO: TIM BOWDITCH

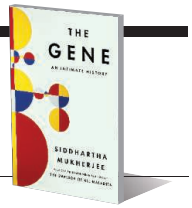
GENETICS

Form. Function. Fate.

A rich history of genetics probes the social and ethical implications of human heredity

Reviewed by **Michael A. Goldman**

The Gene
An Intimate History
Siddhartha Mukherjee
Scribner, 2016. 611 pp.



"This missing science of heredity … is, in simple truth, ten times more important to humanity than all the chemistry and physics, all the technical and industrial science that ever has been or ever will be discovered," wrote H. G. Wells in 1903 (*1*). Having dedicated more than 500 pages to the topic in his new book, *The Gene*, Siddartha Mukherjee would seem to agree. The book examines human genetics, from its philosophical roots in the work of Aristotle, through Gregor Mendel's garden and Thomas Hunt Morgan's work on fruitflies, and including active work in genome editing.

"The study that founded modern biology was buried in the pages of an obscure journal of an obscure scientific society, read mostly by plant breeders in a declining Central European town," writes Mukherjee, referencing Mendel's groundbreaking paper, "Experiments on Plant Hybridization," published in *Proceedings of the Natural History Society of Brünn* in 1866. As a scientist, I can't help but wonder: Would I have been one of those people who overlooked that landmark paper and the significance of DNA as the genetic material?

The eugenics movement is, arguably, one of the darkest chapters in the history of genetics and of America, an embarrassment from which Mukherjee does not shy away. It would lead to the sterilization of Carrie Buck, and many others, and the infamous turn of phrase written by Justice Oliver Wendell Holmes Jr. in the Supreme Court's decision to approve compulsory sterilization in 1927: "Three generations of imbeciles is enough."

Sterilization in Europe would give way to euthanasia and, ultimately, human experimentation and genocide. Yet, the Nazi regime would ultimately contribute to the advancement of genetics, albeit unwittingly. "The exodus from Germany," Mukherjee notes, "allowed scientists to travel not just between nations, but also between disciplines." The intertwining of physics and biology would lead, more or less, to the very birth of molecular genetics, he argues.

Mukherjee, ever the scholar, reaches back to the work of the British embryologist C. H. Waddington, who coined the term "epigenetics" to introduce the broad strokes of how genotypes give rise to phenotypes. As epigenetics came to refer to the effects of DNA modifications and chromatin structure on gene expression, the differences in phenotype between genetically identical cells or organisms (e.g., monozygotic twins) became confounded with genuine molecular differences. Epigenetics has since become shorthand for the effect of environment on the genome, a simplification that neglects the rich and varied ways that the environment influences our identity.

Fourteen children wait to be examined in a "Better Babies" contest in 1931.

PHOTO: EVERETT COLLECTION/NEWSCOM

Switching gears to the topic of gene therapy, Mukherjee recounts a harrowing incident that brought the burgeoning field "to the brink of disaster." In 1997, 18-year-old Jesse Gelsinger, mildly affected by a deficiency of ornithine transcarbamylase (OTC), opted into a gene therapy trial that, if successful, would free him from his rigorous diet and onerous medication regime. Four days after the experiment began, Gelsinger was dead. Subsequent investigations would reveal that the trial's consent forms had failed to reveal pertinent information, that serious side effects had been minimized or ignored, and that one of the clinicians involved in the trial had a financial stake in the company that stood to benefit from a successful trial. "The pattern of neglect was so damning that it nearly obscured the most important scientific lessons of the trial," writes Mukherjee.

In 2011, Jennifer Doudna's work on CRISPR/Cas-9 brought renewed enthusiasm and hope for gene therapy, or at least sent "a restless ping of fantasy through the mind of any human geneticist." Mukherjee describes efforts to control this technology, but the signs of wayward progress in human germline genetic engineering are already clear. "Can we make our genomes a 'little better' without risking the possibility of making ourselves substantially worse?" he asks.

Some of my favorite chapters in human genetics are conspicuously absent from *The Gene*. There isn't a word about the subject of genomic imprinting, in which genes inherited from the mother are expressed differently than the same genes inherited from the father. Bioinformatics, "big data," precision, and personalized medicine might as well be just passing fads. Mukherjee overlooks the vast contribution of chromosome studies in the "pregenome" phase of human genetics and barely scratches the surface of work on the regulation of gene expression in development. Although his coverage of some of the controversies around recombinant DNA technology is thorough and even personal, the book's history of industry is confined to just a few of the first employees and founders at Genentech.

In the concluding pages, Mukherjee distills the story of genetics to 13 overarching principles. They wouldn't get my students through the final exam; they don't even contain the simplest rules of Mendelian inheritance. Still, *The Gene* will have relevance to a wide swath of readers and has the potential to make a working knowledge of genetics a fundamental part of literacy in the 21st century.

The reviewer is at the Department of Biology, San Francisco State University, San Francisco, CA 94132, USA. Email: goldman@sfsu.edu

REFERENCES

1. H.G. Wells, *Mankind in the Making* (Chapman & Hall, London, 1903).

10.1126/science.aaf6085

LETTERS

Edited by **Jennifer Sills**

Time to return blue skies to Iran

BETWEEN OCTOBER 2015 and February 2016, Iran's Meteorological Organization issued multiple "red status" warnings for the local air quality index. In Tehran, schools, construction sites, and polluting factories were closed, smoke-belching cars were fined or towed away, and vulnerable populations were advised to stay indoors. Other Iranian cities have also frequently suffered from extreme air pollution, with lethal consequences (*1*, *2*). Apart from dust storms in some western cities of Iran, many specialists blame Iran's poisonous air on the quantity and quality of the gasoline consumed in the country (*2*, *3*).

Tehran, Iran has suffered multiple "red status" warnings for air quality.

Iranian gasoline consumption is skyrocketing as vehicle numbers soar (*4*). It hit 25 billion liters in 2013 (*5*)—or about 323 liters per capita, which is double that of the average EU citizen (*6*). Not surprisingly, in 2013 Iran's national greenhouse gas emissions reached the ninth highest in the world (*7*).

Gasoline quality is also problematic, and well below global standards. During a decade of sanctions, import restrictions forced the pursuit of self-sufficiency. Iran ramped up domestic gasoline production by converting petrochemical plants to refineries. Substitute—or ersatz—fuel produced in such plants has been shown to be underrefined and heavier in pollutants than gasoline from refineries built for fuel production. Although regulations require ersatz gasoline to meet Euro-4 levels and the government is now seeking to reduce local production with imports, some politicians allege that local fuel meets only Euro-2 standards (*8*).

Consequently, Iranian air quality is well below that of its Persian Gulf neighbors. Qatar consumes twice as much gasoline, but air pollution is only 20 μg/m³, half that of Iran (*9*). Sanctions also crippled Iran's ability to equip new vehicles with catalytic converters for pollution control (*2*). Old domestically made "gas guzzlers" dominate the roads, often belching heavy smoke. Energy efficiency is low in industry, too, because generous government fuel subsidies encourage wasteful practices (*10*).

With sanctions now lifted and government recognition of air pollution mounting, the time is ripe for international cooperation to clean up Iranian skies. This requires a comprehensive approach in energy production and use. Oil consumption must be curbed by ramping up vehicle fuel efficiency standards and lowering fuel subsidies. Enforcement of gasoline quality standards is also crucial. Lastly, Iran's government must invest in clean public transportation to decrease growing vehicle and gasoline dependency.

M. Mozafari,[1,2,3]* G. Trencher,[4] A. Beynaghi[5]

[1]Nanotechnology and Advanced Materials Department, Materials and Energy Research Center (MERC), Tehran, Iran. [2]Cellular and Molecular Research Center, Iran University of Medical Sciences, Tehran, Iran. [3]Department of Tissue Engineering & Regenerative Medicine, Faculty of Advanced Technologies in Medicine, Iran University of Medical Sciences, Tehran, Iran. [4]Department of International Development, Community, and Environment, Clark University, MA 01610, USA. [5]Office of Sustainability, Amirkabir University of Technology, Tehran, Iran.

*Corresponding author. Email: mozafari.masoud@gmail.com

REFERENCES

1. WHO, "WHO Global Urban Ambient Air Pollution Database" (2016); www.who.int/phe/health_topics/outdoorair/databases/cities/en/.
2. C. Mawer, *BMJ* **348**, g1586 (2014).
3. H. H. Jafari, A. Baratimalayeri, *Energy Pol.* **36**, 2536 (2008).
4. "Iran's government steps up efforts to tackle pollution," *The Guardian* (2014); www.theguardian.com/world/iran-blog/2014/mar/10/irans-government-steps-up-efforts-to-tackle-pollution.
5. W. B. Group, *World Development Indicators 2015* (World Bank Publications, 2015).
6. Eurostat, Final Energy Consumption by Product (http://ec.europa.eu/eurostat/tgm/refreshTableAction.do?tab=table&plugin=1&pcode=ten00095&language=en).
7. International Energy Agency, "CO_2 emissions from fuel combustion highlights 2015" (IEA, 2015); www.iea.org/publications/freepublications/publication/co2-emissions-from-fuel-combustion-highlights-2015.html.
8. M. Amiri, "Quality of Iran petrol below global standards–report," Reuters (2010); www.reuters.com/article/iran-gasoline-quality-idAFHOS45321820101214.
9. A. Waked, C. Afif, *Atmos. Env.* **61**, 446 (2012).
10. A. Karbassi, M. Abduli, E. M. Abdollahzadeh, *Energy Pol.* **35**, 5171 (2007).

10.1126/science.aag1449

The age of the Twitter conference

CONFERENCES ALLOW RESEARCHERS to present findings, exchange ideas, and develop networks. Unfortunately, they also incur high financial costs for participants and high environmental costs in the form of carbon emissions associated with air and other travel. Twitter—already popular with many academics (*1*)—can reach large audiences in a short period of time. We propose that global Twitter conferences could be a cost-effective and low-carbon complement to traditional conferences.

The World Seabird Union hosted Twitter Conferences in 2015 and 2016. The 2016 conference featured 72 presenters from 11 countries. Each presenter had 15 minutes to present six tweets (at 140 characters each). The constraints forced presenters to be succinct and use photographs, infographics, or animations to explain their work. The 2016 World Seabird Twitter Conference hashtag (#WSTC2) was used by 633 people in 3931 posts, compared with the 540 people who wrote 2700 posts tagged with the 2015 hashtag (#WSTC1). The success and growth of these events demonstrate the effectiveness of Twitter as a peer-to-peer platform for scientific interaction and science communication.

Stephanie Avery-Gomm,[1]* Sjúrður Hammer,[2] Grant Humphries[3]

[1]Centre for Biodiversity and Conservation Science, University of Queensland, St. Lucia, Queensland,

4067, Australia. [2]Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, UK. [3]Department of Ecology and Evolution, Stony Brook University, Stony Brook, NY 11794, USA.

*Corresponding author. Email: s.averygomm@uq.net.au

REFERENCE

1. G. Orizaola, A. E. Valdés, *Science* **350**, 170 (2015).

Subsidizing truly open access

IN HIS NEWS In Depth story "Dutch push for a quantum leap in open access" (15 April, p. 279), M. Enserink details the European Union initiative to make all EU-published papers open access (OA) by 2020, echoing previous European ideas (*1*). We agree with the initiative's goals but not with its emphasis on a Gold OA model, in which authors pay for both editorial production and the publisher's profit margin. Mesmerized by a quick fix for wealthier European nations and institutions, Berlin 12 Open Access (*2*) and the Max Planck Digital Library (*3*) support Gold OA. However, the Association of Research Libraries and others have pointed out that Gold OA publishing fees constitute a barrier to full global participation in scholarly communication and exacerbate economic and political marginalization (*2*, *4*, *5*).

We advocate a fundamental shift to a Platinum OA system, in which external subsidies, instead of authors, cover publication fees, allowing free access to scientists submitting papers as well as those accessing published content. Many projects and models are exploring the Platinum OA landscape (*6*, *7*), and about 70% of OA journals already have this system (*8*, *9*). However, Platinum OA could be expanded if noncommercial publishers (e.g., scholarly societies) prioritized open communication over short-term profit [e.g., from contracts with commercial publishers (*10*)]; in tandem, academic institutions can incrementally reinvest institutional resources currently dedicated to paying commercial access charges in platinum OA. This plan would open access to the entire scholarly community, avoid drainage of funding out of academia, and realign scholarly communication with academic and socially equitable goals.

A. Townsend Peterson,* Ada Emmett, Josh Bolick, Marc Greenberg, Brian Rosenblum

University of Kansas, Lawrence, KS 66044, USA.

*Corresponding author. Email: town@ku.edu

REFERENCES

1. J. Finch, "Accessibility, sustainability, excellence: How to Expand access to research publications" (2012); www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf.
2. Association of Research Libraries, "Report on Berlin 12 Open Access Conference" (2015); www.arl.org/storage/documents/publications/2015.12.18-Berlin12Report.pdf.
3. Max Planck Digital Library, Open Access 2020 (http://oa2020.org).
4. A. T. Peterson, A. Emmett, M. L. Greenberg, *J. Librarianship Scholarly Commun.* **1**, eP1064 (2013).
5. E. Bonaccorso *et al.*, *J. Librarianship Scholarly Commun.* **2**, eP1126 (2014).
6. R. Kennison, L. Norberg, "A scalable and sustainable approach to open access publishing and archiving for humanities and social sciences" (2014); http://knconsultants.org/wp-content/uploads/2014/01/OA_Proposal_White_Paper_Final.pdf.
7. J. Willinsky, The open access publishing cooperative study (http://oa-cooperative.org/participants.html).
8. W. Crawford, "72% and 41%: A Gold OA 2011–2014 preview" (2015); http://walt.lishost.org/2015/08/72-and-41-a-gold-oa-2011-2014-preview/.
9. P. Suber, "How many peer-reviewed OA journals charge publication fees?" (2015); https://plus.google.com/+PeterSuber/posts/Cqv4oq3LuFr.
10. S. Morris, *Learned Publishing* **20**, 299 (2007).

10.1126/science.aag0946

TECHNICAL COMMENT ABSTRACTS

Comment on "Sensitivity of seafloor bathymetry to climate-driven fluctuations in mid-ocean ridge magma supply"

Peter Huybers, Charles Langmuir, Richard F. Katz, David Ferguson, Cristian Proistosescu, Suzanne Carbotte

Olive *et al.* (Reports, 16 October 2015, p. 310) argue that ~10% fluctuations in melt supply do not produce appreciable changes in ocean ridge bathymetry on time scales less than 100,000 years and thus cannot reflect sea level forcing. Spectral analysis of bathymetry in a region they highlight as being fault controlled, however, shows strong evidence for a signal from sea level variation.

Full text at http://dx.doi.org/10.1126/science.aae0451

Response to Comment on "Sensitivity of seafloor bathymetry to climate-driven fluctuations in mid-ocean ridge magma supply"

J.-A. Olive, M. D. Behn, G. Ito, W. R. Buck, J. Escartín, S. Howell

Huybers *et al.* present new bathymetric spectra from an intermediate-spreading ridge as evidence for a primary contribution of sea-level cycles to the morphology of the seafloor. Although we acknowledge the possibility that sea level–modulated magmatic constructions may be superimposed on a first-order tectonic fabric, we emphasize the difficulty of deciphering these different contributions in the frequency domain alone.

Full text at http://dx.doi.org/10.1126/science.aaf2021

TECHNICAL COMMENT

OCEANOGRAPHY

Comment on "Sensitivity of seafloor bathymetry to climate-driven fluctuations in mid-ocean ridge magma supply"

Peter Huybers,[1]* Charles Langmuir,[1] Richard F. Katz,[3] David Ferguson,[1] Cristian Proistosescu,[1] Suzanne Carbotte[2]

Olive *et al*. (Reports, 16 October 2015, p. 310) argue that ~10% fluctuations in melt supply do not produce appreciable changes in ocean ridge bathymetry on time scales less than 100,000 years and thus cannot reflect sea level forcing. Spectral analysis of bathymetry in a region they highlight as being fault controlled, however, shows strong evidence for a signal from sea level variation.

Olive *et al*. (*1*) argue that abyssal hills do not contain appreciable contributions from the influence of sea level change on magma production at ocean ridges. This hypothesis was originally suggested by Huybers and Langmuir (*2*) and further elaborated and supported by mantle melt modeling and observations on the Australian-Antarctic ridge by Crowley *et al*. (*3*).

Olive *et al*. consider separately intrusive and extrusive magma additions but claim that bathymetric variations from both of these can only result in topographic variations of less than 85-m amplitude for magma volumes that fluctuate at Milankovitch periods [see figure 2 in (*1*)]. Constructional topography of far greater extent, however, is common at slow and intermediate-spreading ridges. A split seamount and the pairs of symmetric ridges at the intermediate-spreading Endeavor segment in the northeast Pacific are both demonstrably constructional features created on-axis and preserved off-axis (*4*). The common axial volcanic ridges at slowly spreading ridges (*5*), where elastic thickness is greatest, are also constructional features, tens of km long with vertical relief of several hundred meters and a cross-sectional scale that can be less than 2 km. Lucky Strike seamount rises 1 km above the axis of the slow-spreading Mid-Atlantic ridge (*6*). The Southeast Indian Ridge (*7*, *8*) and Galapagos Spreading Center (*9*, *10*), which have near constant spreading rates but variable magma supply along axis, show marked differences in axial topography in response to melt flux variations of similar magnitudes to those proposed by Crowley *et al*. Despite the inferences Olive *et al*. draw from their models, seafloor topography of many hundreds of meters that reflects changes in magmatic budget is a common feature of ocean ridges.

A closer examination of the Chile Ridge segment used by Olive *et al*. offers evidence for the conclusions of Crowley *et al*. Large offset faults are apparent in the bathymetry, but there are also smaller linear ridges and numerous volcanic cones with relief of about 100 m, particularly at the edges of the segment [see figure 1 in (*1*)]. Accordingly, we analyze a transect near 39° South from cruise PANR04MV near the northern edge of the Chile Ridge segment. Crustal age is estimated using a Brunhes-Matuyama magnetic reversal of 700 thousand years (ky). Spectral analysis is carried out following the multitaper methodology used by Crowley *et al*., wherein statistical significance is judged at the 95% confidence level relative to a chi-square null distribution (*11*). Significant spectral peaks occur at the Milankovitch periods of 100 ky, 41 ky, and near the 23-ky period, matching those found in sea level variability (Fig. 1), and in agreement with the independent analysis of the Australian-Antarctica Ridge by Crowley *et al*. Other transects from the Chile Ridge show similar behavior, especially near the edges of the segment, but not as clearly. Topography associated with short-term fluctuations in magma supply may be better preserved near the ends of ridge segments on the Chile Ridge, where crustal magma reservoirs may be present only during periods of magma excess.

Olive *et al*. argue that the appearance of 100-ky variability could be an emergent time scale associated with extensional faulting and that spectral peaks at higher frequencies could result from overtones of 100-ky fault spacing. Two further observations, however, indicate that the Chile Ridge bathymetry is not faulting masquerading as Milankovitch. First, bathymetry variations before 700 thousand years ago are characterized by smaller amplitude oscillations and a 41-ky time scale, consistent with the dominant 41-ky obliquity period variability found in Early Pleistocene sea level. Second, examination of the temporal variability shows high-frequency variations in bathymetry similar in structure to sea level estimates and distinct from the abrupt changes that give rise to strong overtones in Olive *et al*.'s simulations (see Fig. 1).

Numerous processes missing from Olive *et al*.'s models might explain their inability to produce the significant sea level–induced bathymetry emerging from observations. For example, dike-induced faulting (*12*) and eruption dynamics, including effusion rates that govern how lavas accumulate on the seafloor (*13*), are not considered. Rather than new pulses of magma effectively mixed across a 1-km-wide magma storage zone as modeled by Olive *et al*., local crustal magma sills could wax and wane as magma supply varies (*14*), and large magma pulses could migrate vertically without mixing in a broader crustal reservoir. Although we agree that Crowley *et al*.'s approach of predicting bathymetric variations assuming local isostasy is probably too simple, it appears that Olive *et al*.'s models also omit important processes.

Olive *et al*. dismiss the importance of volcanic construction on the basis that it is "unlikely to strongly overprint the tectonic fabric of the seafloor, which typically consists of fault scarps greater than 200 m at slow and intermediate-spreading ridges." Normal faults do, of course, contribute importantly to the fabric of the sea floor but do not preclude the presence of additional structure. We suggest the need for a more resolved approach than the historical characterization of abyssal hills using a single wavelength (*15*, *16*). The approach in Crowley *et al*. of prewhitening the spectral estimate has the effect of emphasizing higher-frequency variability and leads to the identification of multiple relevant time scales, namely those associated with 100-ky glacial, 41-ky obliquity, and 23-ky precession variations. This more-detailed seafloor relief is an opportunity to better understand the relative contributions of volcanic and tectonic processes at mid-ocean ridges.

[1]Harvard University, Cambridge, MA, USA. [2]Lamont-Doherty Earth Observatory, Palisades, NY, USA. [3]University of Oxford, Oxford, UK.
*Corresponding author. Email: phuybers@fas.harvard.edu

REFERENCES AND NOTES

1. J. A. Olive *et al*., *Science* **350**, 310–313 (2015).
2. P. Huybers, C. H. Langmuir, *Earth Planet. Sci. Lett.* **286**, 479–491 (2009).
3. J. W. Crowley, R. F. Katz, P. Huybers, C. H. Langmuir, S. H. Park, *Science* **347**, 1237–1240 (2015).
4. D. A. Clague *et al*., *Geochem. Geophys. Geosyst.* **15**, 3364–3391 (2014).
5. R. C. Searle *et al*., *Earth Planet. Sci. Lett.* **299**, 228–241 (2010).
6. R. S. Detrick, H. D. Needham, V. Renard, *J. Geophys. Res.* **100**, 3767–3787 (1995).
7. J. C. Sempéré, J. R. Cochran, *J. Geophys. Res.* **102**, 15489– 15505 (1997).
8. R. C. Holmes, M. Tolstoy, J. R. Cochran, J. S. Floyd, *Geochem. Geophys. Geosyst.* **9**, Q12020 (2008).
9. J. Sinton, R. Detrick, J. P. Canales, G. Ito, M. Behn, *Geochem. Geophys. Geosyst.* **4**, 8515 (2003).
10. J. P. Canales, G. Ito, R. S. Detrick, J. Sinton, *Earth Planet. Sci. Lett.* **203**, 311–327 (2002).
11. D. B. Percival, A. T. Walden, *Spectral Analysis for Physical Applications* (Cambridge Univ. Press, 1993)
12. S. M. Carbotte *et al*., *Geology* **34**, 209–212 (2006).
13. M. R. Perfit, W. W. Chadwick, *Geophys. Monogr.* **106**, 59–116 (1998).

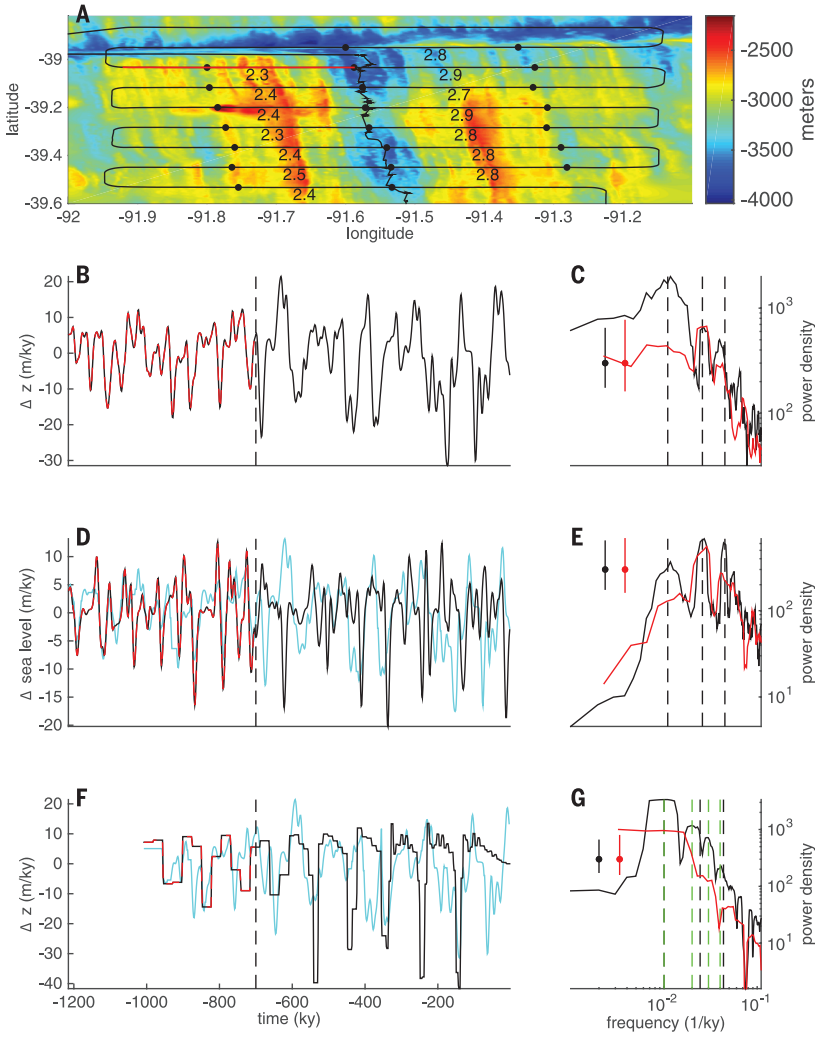


Fig. 1. Bathymetry from the Chile Ridge. (**A**) Map of bathymetry (coloring), the track from cruise PANR04MV (black line), Brunhes-Matuyama reversals from magnetics and visually identified ridge center (dots), inferred spreading rates (numbers in cm/year), and the transect that we focus on (thicker red line). (**B**) Rate of change of the bolded bathymetry section and (**C**) the associated spectral estimate (black), indicating peaks near the 1/100, 1/41, and 1/23 ky^{-1} Milankovitch bands (marked with vertical dashed lines). Spectral peaks are statistically significant when they rise above the background continuum by more than the 95th percentile (i.e., after aligning the dot on the black confidence bar with the spectral peak, the lower vertical bar does not reach the level of the background continuum). Note the use of logarithmic axes. Also shown is a spectral estimate for the time period between 1.2 and 0.7 million years ago during the "41 ky" world (red), where there is significant spectral energy at the 41-ky obliquity band (judged using the red vertical bar) but the other Milankovitch bands are diminished. (**D** and **E**) Analysis of sea level changes (*17*) (note reversal of the *y* axis) shows spectral peaks matching (C). A version of the bathymetry rates of change are also shown (cyan) after alignment with the sea level variations using a dynamic time warping algorithm and scaling to match variance. (**F** and **G**) The analysis is repeated using the faulting simulations from Olive *et al*. with a 100-ky time scale, but which obviously cannot reproduce the transition to 41-ky variability. Overtones of the 100-ky time scale are indicated at 2/100, 3/100, and 4/100 ky^{-1} (green vertical dashed lines). An aligned version of the bathymetry data is also presented (cyan) that illustrates the difference between the continuous high-frequency variability recorded in bathymetry and abrupt transitions in the fault simulation.

14. J. M. Sinton, R. S. Detrick, *J. Geophys. Res.* **97**, 197–288 (1992).
15. J. A. Goff, *J. Geophys. Res.* **96**, 21713–21737 (1991).
16. J. A. Goff, *Science* **349**, 1065 (2015).
17. M. Siddall, B. Hönisch, C. Waelbroeck, P. Huybers, *Quat. Sci. Rev.* **29**, 170–181 (2010).

ACKNOWLEDGMENTS

Research leading to these results received funding through U.S. National Science Foundation award 1338832 and European Research Council grant 279925.

9 December 2015; accepted 17 May 2016
10.1126/science.aae0451

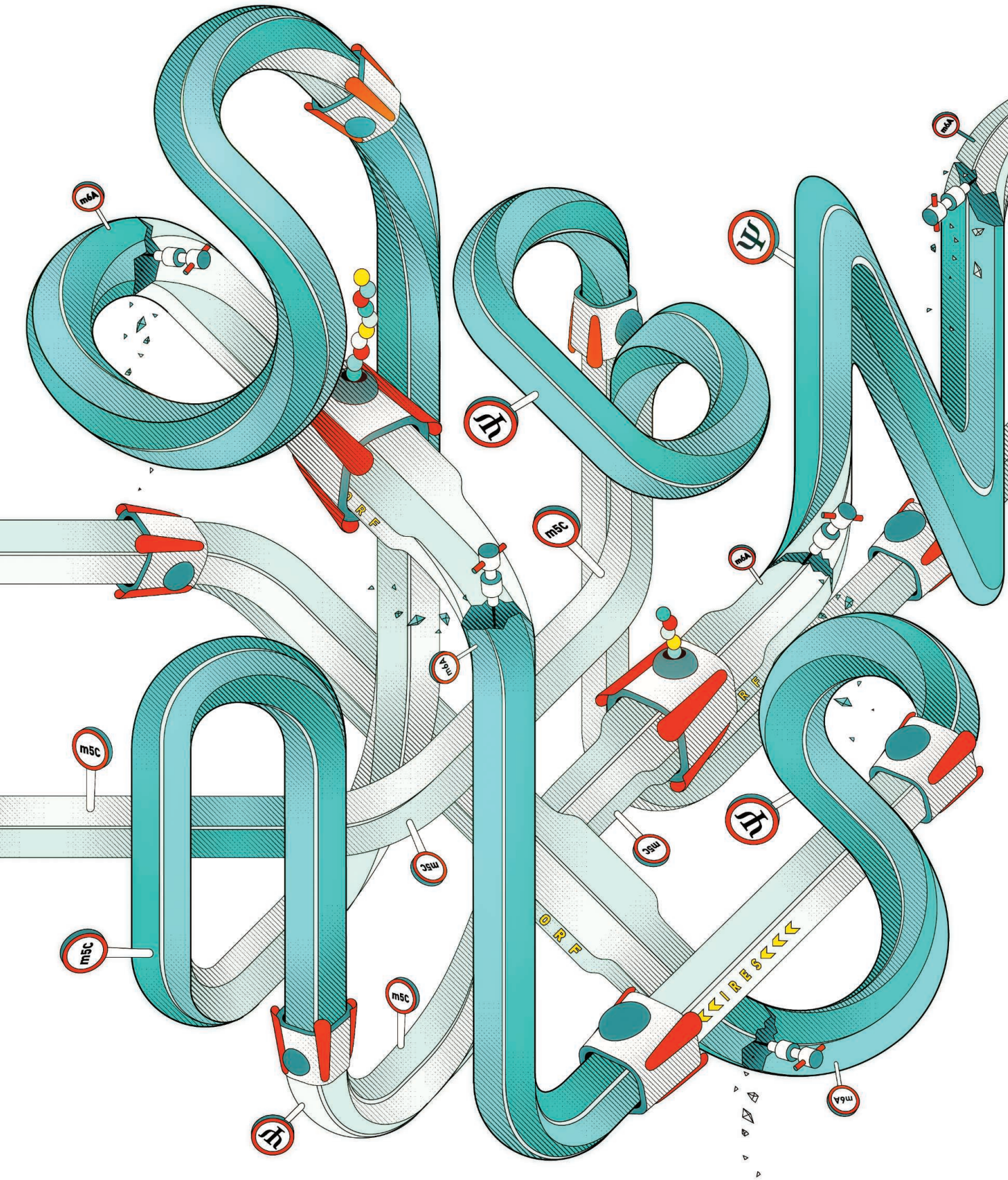
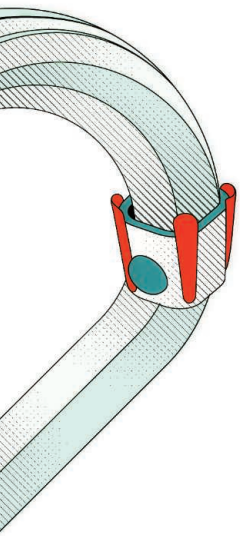


ILLUSTRATION: CHARLES WILLIAMS/MADEUP.ORG



SIGNALS IN RNA

By **Guy Riddihough**

RNA is deeply entwined in the molecular biology of gene expression: in a structural capacity; as an information carrier; and as a regulator in its own right. In this special issue, for example, small noncoding microRNAs, which are known gene expression regulators, are linked to gene expression variability in mammalian stem cells, and this variability is, in turn, linked to cell fate determination during embryonic development.

DNA methylation is critical for the epigenetic regulation of gene expression. RNA, like DNA, can also be covalently modified, and we are discovering that covalent marks on mRNAs and long noncoding RNAs are much more extensive than previously thought. That enzymes can both add and, in some cases, remove these modifications implies that the marks have regulatory functions, and initial evidence is consistent with the idea of an "RNA epigenetics" potentially analogous to that found in DNA.

The 5′ untranslated region of mRNA—the stretch of sequence that the ribosome must traverse to find the start of the protein-coding region—is often rich in regulatory features. For example, very short upstream open reading frames, found in many messages, can modulate how downstream proteins are made. Importantly, the ability to analyze how individual mRNAs are translated in vivo is further revealing details of their kinetics and subcellular localization, and thus the potential to regulate protein synthesis.

Given RNA's critical role in gene expression, it is unsurprising that it has attracted attention as a therapeutic reagent. Indeed, there are a number of ongoing clinical trials that explore different RNA drug modalities and their potential to translate to use in the clinic.

All in all, RNA is continuing to transform the way we think about genes, gene expression, and gene regulation.

INSIDE

REVIEWS

RELATED ITEMS

▸ PERSPECTIVES PP. 1390 & 1391

▸ RESEARCH ARTICLES PP. 1425 & 1430

An artist's impression of some of the sequence features and base modifications that modulate the many and diverse functions of RNA.

REVIEW

Messenger RNA modifications: Form, distribution, and function

Wendy V. Gilbert,[1]* Tristan A. Bell,[1,2]† Cassandra Schaening[1,3]†

RNA contains more than 100 distinct modifications that promote the functions of stable noncoding RNAs in translation and splicing. Recent technical advances have revealed widespread and sparse modification of messenger RNAs with N^6-methyladenosine (m^6A), 5-methylcytosine (m^5C), and pseudouridine (Ψ). Here we discuss the rapidly evolving understanding of the location, regulation, and function of these dynamic mRNA marks, collectively termed the epitranscriptome. We highlight differences among modifications and between species that could instruct ongoing efforts to understand how specific mRNA target sites are selected and how their modification is regulated. Diverse molecular consequences of individual m^6A modifications are beginning to be revealed, but the effects of m^5C and Ψ remain largely unknown. Future work linking molecular effects to organismal phenotypes will broaden our understanding of mRNA modifications as cell and developmental regulators.

The first modified RNA nucleoside was identified almost 60 years ago by analyzing salt-soluble RNA from yeast (*1*). Since then, more than 100 chemically distinct modified nucleotides have been characterized, most of which were identified in tRNAs and other abundant noncoding RNAs from diverse organisms (*2*). N^6-methyladenosine (m^6A) was the first internal mRNA modification discovered (*3*, *4*), and at ~1 to 3 m^6A residues per message, it is abundant enough to be readily detected by bulk mRNA analysis. Next-generation sequencing approaches have allowed mapping of the locations of m^6A and less-abundant modified nucleosides. Here we summarize these methods and the mRNA modification landscape they reveal. We then discuss the enzymes responsible for installing m^6A, m^5C, and pseudouridine (Ψ) at specific sites and emphasize questions regarding the basis for target specificity and regulation. Finally, we describe the remaining challenges in determining the functions of mRNA modifications, which have the potential to regulate genes with widespread consequences for development and disease.

Transcriptome-wide mapping of m^6A, m^5C, and Ψ

All high-throughput methods for locating modified nucleosides rely on one of two approaches: Either antibodies are used to isolate modified RNA fragments for sequencing or some modification-selective RNA chemistry is exploited (Fig. 1, A to C). The first genome-wide RNA modification maps were generated using antibodies against m^6A to identify thousands of ~100 nucleotide RNA fragments containing the modification in mammalian cells (*5*, *6*). This approach has been adapted to give single-nucleotide–resolution m^6A maps by cross-linking the antibody-RNA complexes and determining the sites of cross-link–induced mutations within enriched RNA fragments (*7*, *8*). Antibody-based modification profiling has the advantage of concentrating sequencing efforts on sites of interest.

> *"...mRNA modifications have the potential to affect most posttranscriptional steps in gene expression."*

There is some evidence for artifactual enrichment of mRNA fragments lacking m^6A. In budding yeast, where the only known m^6A-generating enzyme is not essential for cell viability, it was possible to rigorously determine the background association of unmethylated mRNA fragments with anti-m^6A antibodies. Notably, almost half of the putative m^6A peaks were found to be methyltransferase independent (*9*), suggesting a potentially widespread problem with the current m^6A maps in many systems. Higher-confidence m^6A sites can be identified by performing methylome mapping after genetic manipulation of methyltransferases and/or demethylases or by overlaying m^6A peak profiles with maps of methyltransferase interaction sites obtained by cross-linking and immunoprecipitation (CLIP) approaches.

Pseudouridine mapping relies on chemical strategies to selectively derivatize Ψ nucleosides with *N*-cyclohexyl-*N*′-beta-(4-methylmorpholinium) ethylcarbodiimide p-tosylate (CMCT), a bulky covalent adduct that creates a block to reverse transcriptase (RT) (*10*). The use of a "click" chemistry–compatible CMC coupled to biotin allows pre-enrichment of Ψ-containing RNA fragments (*11*). Different computational strategies have been used to identify Ψ sites from analysis of CMC-dependent RT stops in yeast and human cells (*11*–*14*), which may explain some discrepancies between Ψ annotations.

The potential effect of an mRNA modification depends on both the molecular consequences and the percentage of transcripts that are modified. For example, a modification that leads to accelerated mRNA decay is unlikely to have much biological effect if only 1% of transcripts are modified, whereas a modification that causes an alternative protein variant to be produced could be functionally important, even at very low levels. A limitation of current m^6A and Ψ profiling methods is the lack of quantitative information about the extent of modification. Changes in the relative enrichment of a particular sequence in m^6A pull-downs from different growth states have been used to infer regulation of modification, but the absolute fraction of mRNA that is modified cannot be determined from these data. Similarly, differences in the abundance of CMC-dependent reads can indicate relative changes in pseudouridylation when comparing the same mRNA site in different conditions but cannot be compared between different sites due to sequence-dependent capture biases in library preparation. A method has been developed to enable absolute quantitation of modified nucleosides at specific mRNA sites (*15*), but this technique does not scale to allow parallel measurements at many sites. High-throughput methods to quantify site-specific m^6A and Ψ would considerably advance the field.

In contrast to m^6A and Ψ, the level of m^5C at specific sites in RNA or DNA can be quantitatively determined by bisulfite sequencing. For transcriptome-wide analysis of m^5C, isolated RNA is treated with bisulfite to convert unmodified cytosines to uracils before cDNA synthesis. The extent of modification at each cytosine residue is then determined by observing the rate of nonconversion—the fraction of reads that do not show the expected C to T sequence change—assuming complete conversion of unmodified C nucleosides. The first map of m^5C claimed more than 8000 candidate sites in mRNAs from human (HeLa) cells (*16*); however, these sites could include other cytosine modifications known or suspected to interfere with bisulfite conversion (*17*) and may include false positives from stochastic nonconversion events. More targeted high-throughput methods exploit the catalytic mechanism of m^5C methyltransferases to trap covalent intermediates formed between these enzymes and their target sites, which are then identified by RNA sequencing after immunoprecipitation of the methyltransferase (*18*). Targeted bisulfite sequencing at candidate m^5C sites allows verification and quantitation of modification and could be scaled to monitor hundreds

[1]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Graduate Program in Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Graduate Program in Computational and Systems Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*Corresponding author. Email: wgilbert@mit.edu †These authors contributed equally to this work.

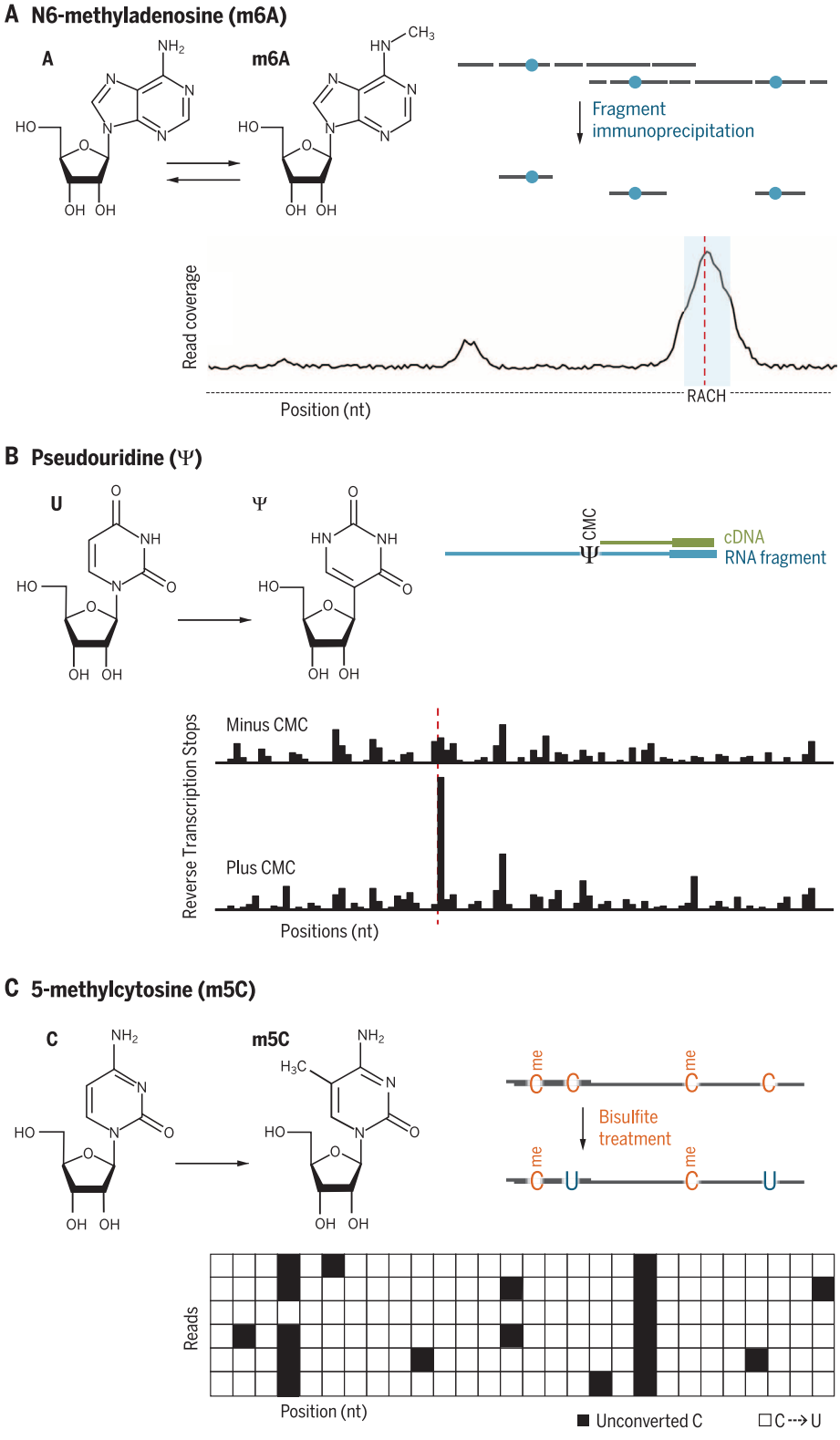


Fig. 1. Nucleotide modifications and detection strategies. For each modification, the chemical structure (left), detection strategy (right), and sample output of mapped sequencing reads (bottom) are shown. (**A**) For detection of m^6A, antibodies are used to select methylated RNA fragments. A typical broad peak in read coverage overlaps an m^6A site. nt, nucleotide. (**B**) Ψ are detected as CMC-dependent reverse transcriptase stops 3′ of U sites. (**C**) m^5C is protected from conversion into U during bisulfite treatment, and putative sites are identified by their high rates of nonconversion. Each row represents one sequencing read, and filled squares indicate unconverted Cs. me, methylated site.

of sites using microfluidics-based multiplex polymerase chain reaction and deep sequencing (*19*).

Although noncoding RNAs are extensively modified in all organisms, the modification landscape of prokaryotic mRNAs has barely begun to be explored. m^6A has been reported in mRNA from *Escherichia coli* and *Pseudomonas aeruginosa* (*20*), and m^5C has been mapped in *Sulfolobus solfataricus* mRNAs (*21*). Because there are substantial differences between prokaryotic and eukaryotic mRNA metabolism, the functional consequences of mRNA modifications are likely to differ as well. It will be interesting to see how these ancient and conserved RNA modifications are harnessed for posttranscriptional gene regulation in organisms with widely differing lifestyles.

Reproducibility and orthogonal validation give confidence in mark annotations. The m^6A methylomes of different human cell lines are highly overlapping (*5*, *22*), genetic manipulation of methyltransferases and demethylases produces the expected changes in m^6A signals (*9*, *22*–*24*), and thousands of m^6A target RNAs have been shown to cross-link to the modifying enzymes in vivo (*24*, *25*). In contrast, for m^5C and Ψ, there are still few transcriptome-wide profiles and almost no independent analysis of similar cell types or growth states by different groups. Because functionally important sites are likely to be identified in multiple studies, a critical next phase will be the determination of m^5C and Ψ sites whose detection is robust to technical variation.

mRNA-modifying enzymes and target specificity

In mammalian mRNA, m^6A is primarily produced by the methyltransferases METTL14 and METTL3, orthologs of yeast *IME4*. METTL3 and METTL14 associate with the regulatory subunit WTAP (Wilms' tumor 1–associating protein) to form a 200-kDa methyltransferase complex (*24*, *25*). Knockout mouse embryonic stem cell lines lacking METTL3 or METTL14 display up to 99% reductions in bulk mRNA methylation and reduced m^6A signal at thousands of sites (*22*, *26*). Most mammalian m^6A sites are found within the consensus sequence Rm^6ACH (R = G or A, H = A, C, or U), which is consistent with the enriched binding motifs observed in CLIP studies of METTL14, METTL3, and WTAP (GGAC, GGAC, and GACU, respectively) (*24*).

Despite the strong consensus, only a small fraction of RACH sites are detectably methylated in vivo, arguing that the sequence motif is not sufficient to determine the distribution of m^6A. Furthermore, m^6A sites are strongly biased toward the 3′ ends of transcripts in organisms as diverse as yeast and humans (*5*, *6*, *9*). This distribution suggests a functional coupling between modification and other RNA processing steps. There is also substantial enrichment of m^6A sites in the vicinity of stop codons in a variety of mammalian cell types (*5*, *6*, *22*), which hints at a role for the ribosome. A simple model can predict which RGAC sites will be methylated in yeast using

just three features: the sequence flanking the site, the proximity of the site to the 3′ end, and the predicted secondary structure at the site (unstructured favored). Expanding the consensus sequence to ANRGACNNU yielded the greatest predictive power; nevertheless, only ~10% of sites that matched this consensus were observed to be methylated (*9*). It remains unclear how much of the observed site specificity of m^6A modifications is due to the intrinsic substrate preferences of these enzymes, and it would be useful to systematically and quantitatively test modification of diverse methyltransferase substrates.

Additional methyltransferase enzymes may also modify mRNAs with m^6A. METTL4 is closely related to METTL14 and METTL3 and is predicted to have catalytic activity (*27*). Although knockdown of METTL4 had no detectable effect on bulk m^6A levels in mRNA from HeLa cells (*24*), the sensitivity of this assay is limited and would not detect changes in modification of a small subpopulation of mRNA substrates. Human cells express at least one active m^6A methyltransferase in addition to METTL14 and METTL3, as suggested by a study of an m^6A site on the human U6 small nuclear RNA (snRNA). The sequence context of this site does not match the RACH consensus, and methylation of this site cannot be outcompeted by an mRNA target of METTL3-METTL14 (*28*).

Pseudouridine sites in mRNA are much less abundant than m^6A, yet the enzymology is more complex (Table 1). There is evidence for mRNA pseudouridylation by 8 out of 10 yeast Ψ synthases, whose canonical targets are mostly sites in tRNA but also include snRNA and ribosomal RNA sites (*12*–*14*). A plurality of identified mRNA substrates in yeast have been genetically assigned to two tRNA-modifying enzymes: Pus1 (*12*), whose human ortholog also modifies mRNAs (*11*), and Pus7 (*13*). The basis for Pus1's target site specificity has long puzzled those studying tRNA modification. Pus7 modifies mRNA sites found in the specific sequence context UGΨAR, in agreement with its known tRNA target site preferences. However, the structural context of this motif is critical for pseudouridylation of noncoding RNAs (*29*), which may explain why only a tiny fraction of RNAs with UGUAR motifs are modified. Pseudouridylation by all Pus enzymes likely involves recognition of RNA secondary structures, the details of which remain to be discovered. Transcriptome-wide RNA structure probing techniques should facilitate identification of the structural determinants for mRNA pseudouridylation, similar to work on m^6A (*30*). mRNA pseudouridylation has been determined in only a few eukaryotic cell types thus far. Because the Ψ synthases that modify mRNAs in yeast and human cells have homologs in all domains of life, mRNA pseudouridylation is likely to be widely distributed.

Comparatively little is known about the enzymology of m^5C deposition in mRNA. Although m^5C is common in noncoding RNAs from all domains of life, there are marked species differences in the reported prevalence of m^5C in mRNA: ~8000 m^5C sites in human mRNAs compared with a single mRNA m^5C site in budding yeast (*16*, *21*). *Saccharomyces cerevisiae* expresses three known m^5C methyltransferases: Ncl1, which modifies multiple tRNAs at several positions, and Nop2 and Rcm1, which target the ribosomal RNA. In human cells, the methyltransferases Dnmt2 and Nsun2 have been shown to modify certain mRNAs (*16*, *18*), but their verified targets do not include most reported m^5C sites, suggesting that additional enzymes may be active toward mRNA substrates. Humans have seven known proteins in the Nsun methyltransferase family (Nop2 and Nsun2 to Nsun7), of which five have been shown to have catalytic activity, and all seven have the predicted active-site cysteine residues. No sequence motifs are common among reported m^5C sites in human mRNAs, which is expected if multiple enzymes are responsible. In contrast to humans, all known m^5C sites in *S. solfataricus* exactly match the sequence of a ribosomal m^5C site, suggesting a common enzyme (*21*).

Regulation of mRNA modifications

m^6A is a truly dynamic mRNA modification that can be enzymatically removed by demethylase "eraser" enzymes. Fat mass and obesity-associated protein, FTO, was the first mammalian RNA demethylase implicated in the dynamics of m^6A modifications (*23*). ALKBH5 is a second, conserved eraser of m^6A marks that is highly expressed in the testes and is required for spermatogenesis and fertility in mice (*31*). The sequence or structural preferences of the demethylases may account for some or all of the observed specificity in the distribution of m^6A among individual RACH sites.

Alternatively, the m^6A "writing" and "erasing" enzymes may have only minimal sequence requirements (for RACH motifs), allowing the accessibility of potential sites and/or the availability of competing RNA-binding proteins to shape the landscape of m^6A modification. An example of this mode of regulating m^6A comes from studies of the heat shock response in mammalian cells (MEF and HeLa) (*32*). It is plausible that m^6A modification is somehow coupled to cotranscriptional RNA processing events, given the pronounced 3′ bias of modified sites. Moreover, shifts in the relative abundance of m^6A in 5′ compared with 3′UTRs (3′ untranslated regions) [e.g., upon heat shock (*32*, *33*)] could arise from a global change in recruitment of methyltransferases or demethylases to promoters.

Changes in enzyme or substrate localization may also play an important role in regulating mRNA modifications. In yeast, heat shock induces relocalization of the Ψ synthase Pus7 from the nucleus to the cytoplasm. This shift correlates with a 10-fold increase in the number of Pus7-dependent pseudouridylated mRNA target sites (*13*), which could reflect an increased window of opportunity for Pus7 to interact with potential mRNA substrates. Similarly, conditions that decrease the rate of mRNA export, either globally or for specific messages, could increase mRNA modification by nuclear enzymes. In *S. cerevisiae*, the Ime4-Mum2-Slz1 m^6A methyltransferase complex localizes to nucleoli at the stage of meiosis in which m^6A mRNA levels peak, and mutations in *SLZ1* that prevent nucleolar accumulation of Ime4 also reduce m^6A levels by about a factor of 3 (*9*). Another apparently localization-dependent regulatory mechanism involves heat shock–induced nuclear accumulation of the mammalian m^6A "reader" protein, YTHDF2, which competes with the FTO demethylase for binding to m^6A sites (*32*). The observation that heat shock causes a substantial shift in the distribution of m^6A from 5′ to 3′ sites implies that 5′ sites are more efficiently targeted by demethylases during normal growth and may be more inherently dynamic.

Pseudouridylation is thought to be irreversible, so its observed dynamics are likely mediated by the production or degradation of pseudouridylated messages. It is also possible that the

Table 1. mRNA-modifying enzymes and modification readers. m, mammalian; y, yeast.

Factors	m^6A	Ψ	m^5C
Writers	METTL14 (m) (*22*, *26*), METTL3 (m) (*22*, *26*), METTL4? (m) (*27*)	Pus1 (m, y) (*11*–*13*), Pus2 (y) (*12*, *13*), Pus3 (y) (*12*, *13*), Pus4 (y) (*11*–*13*), Pus6 (y) (*12*, *13*), Pus7 (m, y) (*11*–*13*), Pus9 (y) (*12*, *13*), Cbf5 (y) (*12*, *13*)	Dnmt2 (m) (*16*, *18*), Nsun2 (m) (*16*, *18*)
Erasers	FTO (m) (*23*), ALKBH5 (m) (*31*)	None identified	None identified
Readers	YTH family (m) (*32*, *36*, *40*), YTHDF2 (m) (*5*, *32*)	None identified	None identified

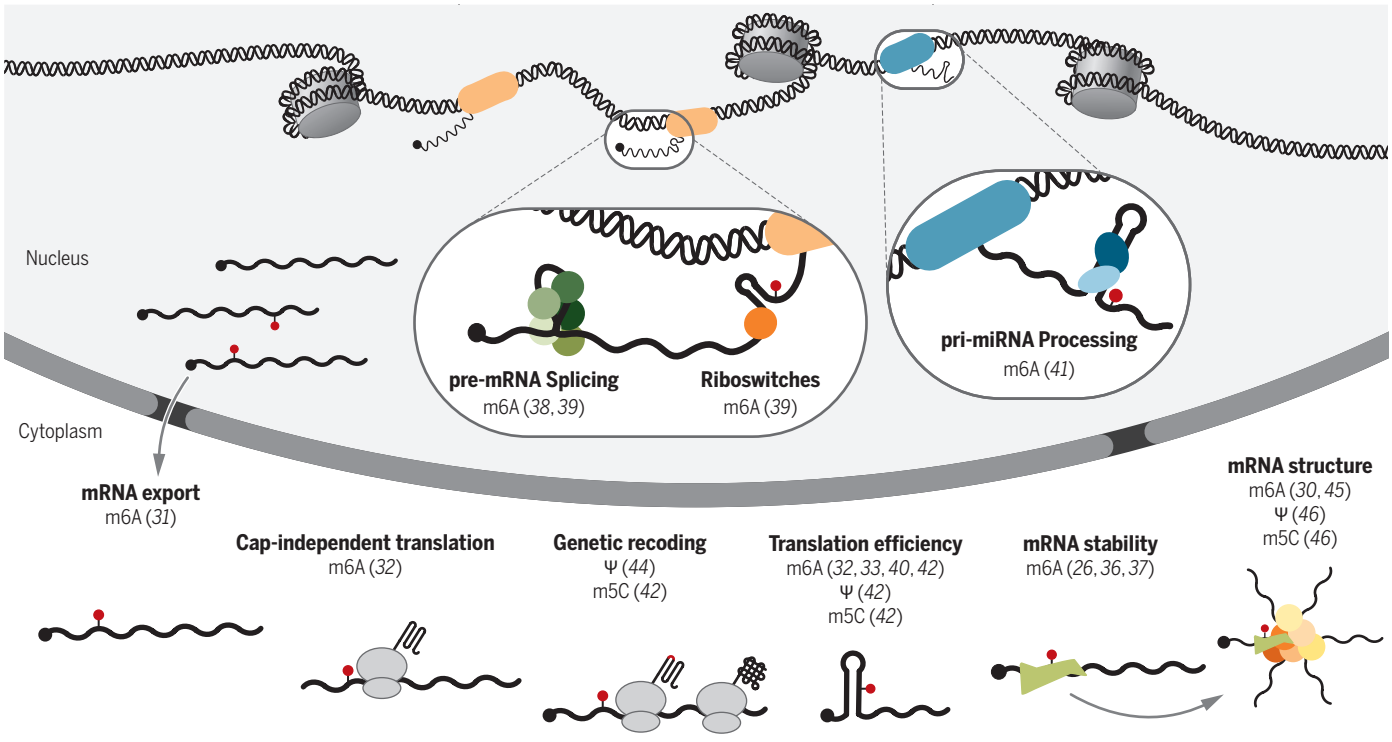


Fig. 2. Diverse molecular functions of m^6A, Ψ, and m^5C in coding RNAs. Nascent RNA transcripts in eukaryotic cells are chemically modified (red dot) by m^6A, Ψ, and m^5C "writer" enzymes. In the nucleus, m^6A, and potentially other mRNA modifications, alters processing of pre-mRNA and pri-miRNA, through both direct recognition and induced changes in RNA secondary structure (*38*, *39*, *41*). After export to the cytoplasm, which is enhanced by m^6A (*31*), mRNA modifications alter the efficiency and fidelity of translation (*32*, *33*, *40*, *42*, *44*) and turnover of transcripts in the actively translating mRNA pool (*26*, *36*, *37*).

molecular consequences of an irreversible modification such as Ψ could be functionally mitigated by additional chemical transformations of the modified nucleoside (i.e., by affecting RNA-protein interactions and/or RNA structure). Ψ can be further modified by N1 methylation (*34*), raising the possibility that the effects of mRNA pseudouridylation can be modulated if not reversed. Similarly, m^5C in RNA can be further modified to 5-hydroxymethylcytidine, 5-formylcytidine, and 5-carboxylcytidine by ten-eleven translocation (Tet) family enzymes (*35*). The direct reversibility of m^5C in RNA is an open question, but there are multiple pathways for demethylation of m^5C in DNA. However, given that all of the known mRNA demethylases act in the nucleus and that mRNAs can be turned over rapidly, reversible and irreversible mRNA modifications may have similar regulatory potential. Consistent with this possibility, m^6A shows highly dynamic changes during meiosis in *S. cerevisiae*, despite the apparent lack of an m^6A demethylase (*9*).

It is notable that most known mRNA-modifying enzymes are nuclear during normal growth. This localization allows m^6A, and likely other modifications, to influence mRNA biogenesis from the earliest stages and may enable coupling between transcriptional control circuits and mRNA modification state.

Functional consequences

At the molecular level, mRNA modifications have the potential to affect most posttranscriptional steps in gene expression. Some broad regulatory themes have emerged from mechanistic studies of mRNA regulation by m^6A, which has been linked to control of mRNA stability (*26*, *36*, *37*), splicing (*38*, *39*), and translational efficiency (*32*, *33*, *40*) and to pri–microRNA (miRNA) processing (*41*) (Fig. 2). m^6A, and likely other modifications, can mediate diverse effects on mRNA metabolism by affecting interactions with RNA binding proteins. YTHDF2—the first modification-specific RNA binding protein, or "reader," identified (*5*)—increases turnover of m^6A-modified mRNA by promoting colocalization with decay factors (*36*). This destabilizing function of m^6A plays an important biological role in stem cell differentiation by regulating key pluripotency factors (*26*, *37*). Individual YTH proteins interact with distinct subsets of m^6A sites and produce different effects on gene expression when perturbed in different cellular contexts (*32*, *36*, *40*), suggesting combinatorial control by uncharacterized co-regulators.

Dedicated reader proteins have not yet been identified for m^5C or Ψ, but the ribosome has the potential to function as a universal reader of mRNA modifications. Incorporation of single m^6A, m^5C, or Ψ modifications in specific codon contexts reduced protein production in *E. coli* by 20 to 70% (*42*), a level of repression exceeding that of many conserved microRNA target sites (*43*). There is also the possibility of encoding truncated proteins through site-specific ribosome stalling at modified codons, depending on how the stall is resolved. Perhaps the most exciting potential for mRNA modifications to effect gene function is through regulated rewiring of the genetic code, and there are limited but intriguing observations supporting this possibility. Insertion of m^5C led to 4% recoding of proline as leucine in *E. coli* (*42*), and Ψ-containing stop codons were efficiently mistranslated as specific amino acids in budding yeast (*44*). The effects of modified nucleosides on translation have been tested in very limited contexts, and the mechanisms responsible for noncanonical decoding events are unknown. If coding-sequence modifications lead to the production of alternative protein variants, even low-occupancy modifications could have substantial biological effects.

RNA modifications may broadly influence mRNA metabolism through their effects on RNA structures. m^6A destabilizes RNA duplexes in vitro (*45*), and m^6A sites in mRNA tend to be unstructured in vivo (*30*). Both m^5C and Ψ affect tRNA folding and are likely to affect mRNA structures as well (*46*). By changing the accessibility of binding sites for regulatory

factors, the effect of RNA structure may be direct (e.g., inhibiting splicing or translation initiation by blocking access to functional sites) or indirect, as was recently shown for m^6A structural switches affecting splicing (*39*). Given the stabilizing effects of Ψ on RNA structure and the central role of RNA structure in prokaryotic gene regulation, it will be particularly interesting to see if eubacterial and archaeal mRNAs are also pseudouridylated. The effect of individual modifications is likely to be highly dependent on context. Therefore, it will be important to balance unifying descriptions of the effects of particular modifications against the need to understand sites that deviate from global trends. The incorporation of context effects has been critical for progress toward elucidating the "splicing code" (*47*), which offers an instructive model for investigating the regulatory effects encoded in mRNA modifications.

As the mRNA modification field develops further, the most promising future work will directly relate individual modifications to cellular and organismal phenotypes. To date, most mRNA modification studies have employed sequencing-based assays to explore the epitranscriptomic landscape in vivo. Although this approach has rapidly expanded our knowledge of the frequency and diversity of cellular mRNA modifications, the importance of most modification sites remains enigmatic. Detailed investigations of individual modification sites are sorely needed. A key step is determining the molecular effects of preventing or introducing individual modifications at physiologically relevant sites, as recently determined for m^6 methylation of A103 of the 5′UTR of Hsp70. This modification is necessary and sufficient to promote noncanonical cap-independent translation (*32*), though the cellular importance of this methylation event in the heat shock response remains unclear.

"Given the enormous diversity of RNA modifications found in tRNA ...the mRNA-modification landscape is likely to be very rich indeed."

The most substantial barrier to detailed studies of molecular function is identifying which modification sites are the most biologically relevant. Conservation analysis is an underutilized and cost-effective tool for prioritizing specific mRNA modification sites for in-depth characterization. Hundreds of m^6A sites are conserved between human and mouse embryonic stem cells (*22*). Likewise, dozens of orthologous sites become modified with m^6A during meiosis in yeast species that last shared a common ancestor more than 5 million years ago (*9*). There is substantially more conservation of modified genes than sites in both yeast and mammals. This might be expected if the relevant (conserved) functional consequence of m^6A addition is recruitment of a trans-acting factor that works similarly from anywhere within the 3′UTR. On the other hand, modification sites that control alternative splicing or protein recoding should be conserved at the nucleotide level. Thus, examining the patterns of conservation could illuminate the likely molecular function of a modification, as well as its biological importance.

With the use of CRISPR-based genome engineering, it is currently possible to directly test the influence of altering any modification site in many organisms. CRISPR multiplexing strategies could also potentially permit interrogation of many sites in parallel and hasten functional discoveries. As a complementary approach, it would be valuable to promote de novo modification at specific sites by engineering fusions between catalytic domains (e.g., methyltransferase, Ψ synthase) and a programmable RNA binding domain scaffold such as Pumilio.

The expanding epitranscriptome

Just how complex is the epitranscriptome? High-throughput sequencing methods were required to reveal the presence of sparse Ψ modifications in mRNA, suggesting that new approaches will discover still more previously unseen mRNA modifications. In fact, while this Review was in preparation, the first maps of 5-hydroxymethylcytidine and N^1-methyladenosine sites were reported using new sequencing techniques to examine mRNA from flies, yeast, mice, and human cells (*48*–*50*). Given the enormous diversity of RNA modifications found in tRNA and the demonstrations that many tRNA-modifying enzymes also have mRNA substrates (*11*–*14*, *16*, *18*), the mRNA-modification landscape is likely to be very rich indeed.

The past 5 years have witnessed a revolution in our understanding of the extent and diversity of mRNA modifications. Many clever enrichment and chemical modification strategies coupled to high-throughput sequencing have produced maps of m^6A, m^5C, and Ψ across the transcriptomes of many organisms and demonstrated the dynamics of these modifications across different cellular growth states. Now that diverse molecular functions of mRNA modifications are beginning to emerge, the field is poised for breakthroughs in understanding the most important cellular and organismal functions of mRNA modification.

REFERENCES AND NOTES

1. F. F. Davis, F. W. Allen, *J. Biol. Chem.* **227**, 907–915 (1957).
2. M. A. Machnicka *et al.*, *Nucleic Acids Res.* **41**, D262–D267 (2013).
3. R. P. Perry, D. E. Kelley, *Cell* **1**, 37–42 (1974).
4. R. Desrosiers, K. Friderici, F. Rottman, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3971–3975 (1974).
5. D. Dominissini *et al.*, *Nature* **485**, 201–206 (2012).
6. K. D. Meyer *et al.*, *Cell* **149**, 1635–1646 (2012).
7. K. Chen *et al.*, *Angew. Chem. Int. Ed. Engl.* **54**, 1587–1590 (2015).
8. B. Linder *et al.*, *Nat. Methods* **12**, 767–772 (2015).
9. S. Schwartz *et al.*, *Cell* **155**, 1409–1421 (2013).
10. A. Bakin, J. Ofengand, *Biochemistry* **32**, 9754–9762 (1993).
11. X. Li *et al.*, *Nat. Chem. Biol.* **11**, 592–597 (2015).
12. T. M. Carlile *et al.*, *Nature* **515**, 143–146 (2014).
13. S. Schwartz *et al.*, *Cell* **159**, 148–162 (2014).
14. A. F. Lovejoy, D. P. Riordan, P. O. Brown, *PLOS ONE* **9**, e110799 (2014).
15. N. Liu, T. Pan, *Methods Enzymol.* **560**, 149–159 (2015).
16. J. E. Squires *et al.*, *Nucleic Acids Res.* **40**, 5023–5033 (2012).
17. M. Schaefer, T. Pollex, K. Hanna, F. Lyko, *Nucleic Acids Res.* **37**, e12 (2009).
18. V. Khoddami, B. R. Cairns, *Nat. Biotechnol.* **31**, 458–464 (2013).
19. R. Zhang *et al.*, *Nat. Methods* **11**, 51–54 (2014).
20. X. Deng *et al.*, *Nucleic Acids Res.* **43**, 6557–6567 (2015).
21. S. Edelheit, S. Schwartz, M. R. Mumbach, O. Wurtzel, R. Sorek, *PLOS Genet.* **9**, e1003602 (2013).
22. P. J. Batista *et al.*, *Cell Stem Cell* **15**, 707–719 (2014).
23. G. Jia *et al.*, *Nat. Chem. Biol.* **7**, 885–887 (2011).
24. J. Liu *et al.*, *Nat. Chem. Biol.* **10**, 93–95 (2014).
25. X.-L. Ping *et al.*, *Cell Res.* **24**, 177–189 (2014).
26. S. Geula *et al.*, *Science* **347**, 1002–1006 (2015).
27. J. M. Bujnicki, M. Feder, M. Radlinska, R. M. Blumenthal, *J. Mol. Evol.* **55**, 431–444 (2002).
28. S. Shimba, J. A. Bokar, F. Rottman, R. Reddy, *Nucleic Acids Res.* **23**, 2421–2426 (1995).
29. A. Urban, I. Behm-Ansmant, C. Branlant, Y. Motorin, *J. Biol. Chem.* **284**, 5845–5858 (2009).
30. R.C. Spitale *et al.*, *Nature* **519**, 486–490 (2015).
31. G. Zheng *et al.*, *Mol. Cell* **49**, 18–29 (2013).
32. J. Zhou *et al.*, *Nature* **526**, 591–594 (2015).
33. K. D. Meyer *et al.*, *Cell* **163**, 999–1010 (2015).
34. B. Meyer *et al.*, *Nucleic Acids Res.* **39**, 1526–1537 (2011).
35. L. Shen, C.-X. Song, C. He, Y. Zhang, *Annu. Rev. Biochem.* **83**, 585–614 (2014).
36. X. Wang *et al.*, *Nature* **505**, 117–120 (2014).
37. Y. Wang *et al.*, *Nat. Cell Biol.* **16**, 191–198 (2014).
38. X. Zhao *et al.*, *Cell Res.* **24**, 1403–1419 (2014).
39. N. Liu *et al.*, *Nature* **518**, 560–564 (2015).
40. X. Wang *et al.*, *Cell* **161**, 1388–1399 (2015).
41. C. R. Alarcón, H. Lee, H. Goodarzi, N. Halberg, S. F. Tavazoie, *Nature* **519**, 482–485 (2015).
42. T. P. Hoernes *et al.*, *Nucleic Acids Res.* **44**, 852–862 (2016).
43. S. W. Eichhorn *et al.*, *Mol. Cell* **56**, 104–115 (2014).
44. J. Karijolich, Y.-T. Yu, *Nature* **474**, 395–398 (2011).
45. C. Roost *et al.*, *J. Am. Chem. Soc.* **137**, 2107–2115 (2015).
46. M. Helm, *Nucleic Acids Res.* **34**, 721–733 (2006).
47. Z. Wang, C. B. Burge, *RNA* **14**, 802–813 (2008).
48. B. Delatte *et al.*, *Science* **351**, 282–285 (2016).
49. D. Dominissini *et al.*, *Nature* **530**, 441–446 (2016).
50. X. Li *et al.*, *Nat. Chem. Biol.* **12**, 311–316 (2016).

ACKNOWLEDGMENTS

We thank members of the Gilbert lab for discussion. This work was supported by grants from the NIH (GM101316 and CA187236) and the American Cancer Society (RSG-13-396-01-RMC) to W.V.G., a NIH Pre-Doctoral Training grant (T32GM007287) to T.A.B., and an NSF Graduate Research Fellowship to C.S.

10.1126/science.aad8711

REVIEW

Translational control by 5′-untranslated regions of eukaryotic mRNAs

Alan G. Hinnebusch,[1] Ivaylo P. Ivanov,[1] Nahum Sonenberg[2]*

The eukaryotic 5′ untranslated region (UTR) is critical for ribosome recruitment to the messenger RNA (mRNA) and start codon choice and plays a major role in the control of translation efficiency and shaping the cellular proteome. The ribosomal initiation complex is assembled on the mRNA via a cap-dependent or cap-independent mechanism. We describe various mechanisms controlling ribosome scanning and initiation codon selection by 5′ upstream open reading frames, translation initiation factors, and primary and secondary structures of the 5′UTR, including particular sequence motifs. We also discuss translational control via phosphorylation of eukaryotic initiation factor 2, which is implicated in learning and memory, neurodegenerative diseases, and cancer.

Most eukaryotic mRNAs are translated by the scanning mechanism, which begins with assembly of a 43*S* preinitiation complex (PIC), containing methionyl-initiator tRNA (Met-tRNA$_i$) in a ternary complex (TC) with guanosine triphosphate (GTP)–bound eukaryotic initiation factor 2 (eIF2). The 43*S* PIC assembly is stimulated by eIFs 1, 1A, 3, and 5 (Fig. 1). Its subsequent attachment to mRNA at the m^{7}G-capped 5′ end is facilitated by the eIF4F complex—composed of cap-binding protein eIF4E, eIF4G, and RNA helicase eIF4A—and by poly(A)-binding protein (PABP). The PIC scans the mRNA 5′ untranslated region (UTR) for an AUG nucleotide triplet start codon using complementarity with the anticodon of Met-tRNA$_i$. AUG recognition evokes hydrolysis of the GTP bound to eIF2 to produce a stable 48*S* PIC. Release of eIF2-GDP is followed by joining of the large (60*S*) ribosome subunit, catalyzed by eIF5B, to produce an 80*S* initiation complex ready to begin protein synthesis (Fig. 1) (*1*). There are exceptions to the scanning mechanism in which PICs are recruited by specialized sequences in the 5′UTR, called internal ribosome entry sites (IRESs).

The scanning mechanism of translation initiation

The nature of scanning, its 5′ to 3′ directionality, dictates that the initiation codon is frequently the AUG triplet closest to the 5′ end, encountered first by the scanning PIC. The first AUG can be skipped when it is flanked by an unfavorable sequence—a process termed "leaky scanning"—to use a downstream AUG. A favorable sequence context in mammals is the "Kozak consensus," 5′ (A/G)CCAUGG 3′ (*2*). Although not always the same in plants and fungi, a purine at the −3 position from the AUG both is conserved and functionally predominates over other positions in all organisms (*1*). When an upstream AUG (uAUG) is in-frame with a downstream AUG without an intervening stop codon, leaky scanning may occur at some frequency to allow production of two protein isomers differing only by an N-terminal extension, with the longer form often targeted to a particular cellular compartment. If the uAUG is followed by a stop codon in the same open reading frame (ORF), then translation of the upstream ORF (uORF) will attenuate translation of the downstream main ORF, because reinitiation is generally inefficient (Fig. 2A). Some uORFs inhibit downstream translation primarily because ribosomes stall during their translation and create a roadblock to scanning PICs that bypass the uORF start codon (Fig. 2B). These principles account for the fact that polycistronic mRNAs, common in bacteria, are rare in eukaryotes (*1*).

"Near-cognate" triplets, differing from AUG by a single base, can be selected by the scanning PIC but with lower frequencies, owing to the mismatch with the anticodon of tRNA$_i$ and attendant destabilization of the 48*S* PIC. NUG (N is any nucleotide) triplets generally function the best, whereas A(A/G)G triplets are the worst, and the use of near-cognates relies more heavily than AUG on optimal context (*1*). Although CUG codons are usually decoded by Met-tRNA$_i$, leucyl-tRNA^{Leu} can be engaged by a scanning PIC in a manner requiring the noncanonical initiation factor eIF2A but not eIF2. This occurs in the synthesis of antigenic precursors for loading on major histocompatibility complex molecules (*3*).

Multiple eIFs, structural elements of tRNA$_i$, and both rRNA and protein components of the 40*S* subunit participate in discrimination between AUGs and non-AUG triplets as start codons, and good versus poor Kozak context, by the scanning PIC (*1*). eIF1 and eIF5 have opposing effects, with

[1]Group on Cell Regulation and Development, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA. [2]Department of Biochemistry and Goodman Cancer Centre, McGill University, Montreal, Quebec H3A 1A3, Canada.
*Corresponding author. Email: nahum.sonenberg@mcgill.ca

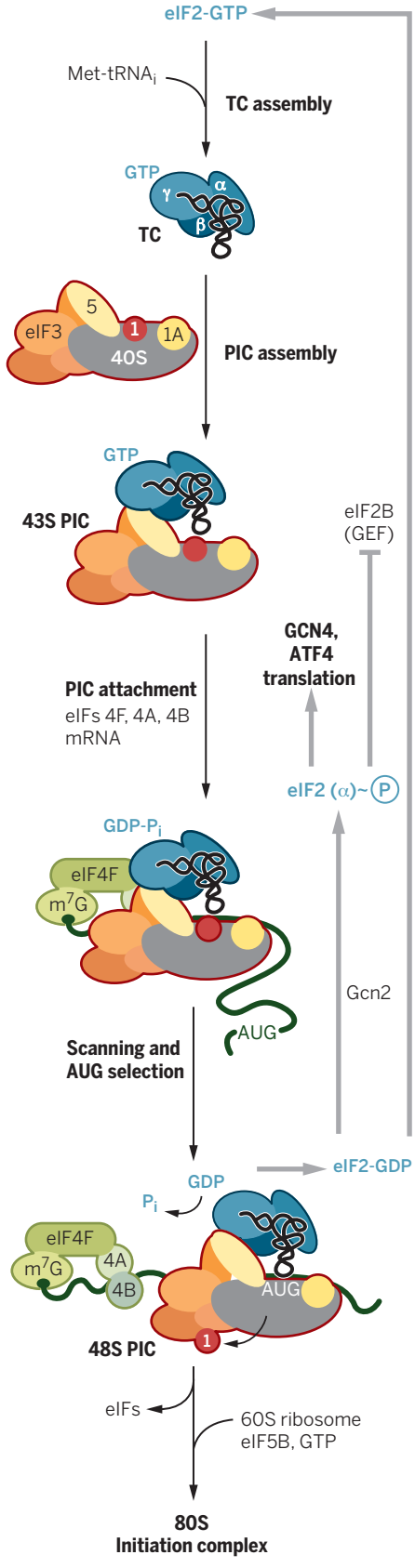


Fig. 1. The scanning mechanism of translation initiation. The simpler 5-subunit version of budding yeast eIF3 is depicted.

eIF1 promoting scanning and blocking recognition of non-AUGs and AUGs in poor context and eIF5 antagonizing eIF1 function. These activities are exploited for autoregulation and cross-regulation of eIF1/eIF5 expression in most eukaryotes (*4*), because the eIF1 AUG is in weak context and poorly recognized unless eIF1 levels drop below a certain threshold (*5*). Initiation at the eIF5 start codon is impaired by an uORF whose AUG is in weak context and hence efficiently bypassed by leaky scanning only at low eIF5 (or high eIF1) levels (Fig. 3A) (*4*). Yeast mutants of these and other initiation factors (eIF1A, eIF2 subunits, and eIF3c) can increase or decrease discrimination against suboptimal start codons (*1*, *6*). There is a potential for regulating such initiation events through posttranslational modifications of these factors or with small molecules.

Based on studies in yeast, in which dramatically lengthening the 5′UTR conferred no reduction in translational efficiency, it appears that the scanning PIC is highly processive (*7*). However, an excessively short 5′UTR [≤20 nucleotides (nt)] is generally detrimental and can evoke leaky scanning (*1*, *8*). Indeed, genome-wide mapping of yeast 5′UTRs identified many mRNAs with short 5′UTRs exhibiting lower-than-average translational efficiency (TEs) (*9*). Leaky scanning induced by short 5′UTRs allows production of different isoforms differing at the N termini from the same mRNA (*1*, *10*). In contrast, the mammalian 5′UTR element translation initiator of short 5′UTRs (TISU) allows cap-dependent but scanning-independent initiation on mRNAs with 5′UTRs as short as 5 nt. Although not requiring eIF4A, TISU's function paradoxically depends on eIF1 (*11*), which normally blocks selection of AUGs too close to the cap (*8*). mRNAs encoding mitochondrial proteins are enriched in TISU, which appears to confer maintenance of translation at low energy levels (*11*).

Translational control by 5′UTR structure

Secondary structures in the 5′UTR can also influence the initiation efficiency of suboptimal start codons. A strong stem-loop (SL) structure just downstream of the start codon will stall the scanning 40*S* subunit, increasing its "dwell time" and thus decreasing the probability of leaky scanning through near-cognates or AUG triplets in poor context (*12*).

Although a precise SL-AUG spacing is required for the SL stimulatory effect, mRNA structures of sufficient stability inhibit all scanning-dependent initiation downstream (*1*). DEAD-box adenosine triphosphate (ATP)–dependent RNA helicases can overcome SL structural impediments, and they might be specialized for certain types or locations of mRNA structures. The fact that eIF4A is recruited to the mRNA 5′ end and activated as a component of eIF4F positions eIF4A to facilitate PIC attachment near the cap to initiate scanning at the mRNA 5′ end (*1*). eIF4E overexpression preferentially stimulates translation of mRNAs containing excessive secondary structure, implying that eIF4F is limiting for translation of mRNAs with structured leaders (*13*). Mammalian DHX29 and the yeast homolog of DDX3 (Ded1) appear to be crucial for resolving stable structures distal from the cap that impede scanning (*14*). Indeed, genome-wide analysis of TEs in Ded1 and eIF4A yeast mutants revealed that Ded1-hyperdependent genes tend to have atypically long and structured 5′UTRs, whereas eIF4A contributes more equally to translation of all mRNAs (*15*). This differs in mammals, where eIF4A-dependence is conferred by long 5′ UTR sequences capable of forming stable secondary structure (*16*) or G-quadruplex structures (*17*). Moreover, mammalian mRNAs containing 5′UTR secondary structure are hyperdependent on eIF4A for translation in vitro (*18*). Cap-proximal structures can also impede eIF4F binding to the cap (*19*), and DDX3 was implicated in resolving cap-proximal SLs to enhance eIF4F recruitment (*20*).

Analogous to the inhibitory effects of cap-proximal SL elements, a paradigm of translational control in mammals involves formation of an mRNA-protein complex composed of the iron regulatory protein (IRP) and a cap-proximal SL known as the iron response element, which blocks 43*S* attachment to mRNAs encoding ferritin or other iron metabolism proteins in iron-deprived cells (*21*).

Translational control by uORFs

Genome-wide sequencing of 5′UTRs reveals that uORFs are pervasive, occurring in ~50% of mammalian mRNAs, and there is evidence from ribosome footprint profiling that a sizable fraction of uORFs are translated (*22*–*24*), although only a small fraction produces peptides sufficiently abundant and stable for detection (*25*). It is likely that ribosome occupancies of uORFs detected in certain profiling experiments overestimate their TEs in cells (*26*, *27*). This is especially true for uORFs initiated by near-cognates under conditions where bulk protein synthesis is diminished, where their representation is substantially elevated compared with their use as start sites for main ORFs. However, the facts that the occurrence of AUG-initiated uORFs is below the frequency predicted by chance; that, when present, their start codons tend to be in poor initiation context; that their occurrence and translation is associated with below-average TEs for the downstream ORFs genome-wide; and that they show evidence of evolutionary sequence conservation are good indicators that AUG-initiated uORFs function broadly to throttle down translation initiation, whereas the same evidence is lacking for most non-AUG–initiated uORFs (*9*, *24*, *28*, *29*). Regulation via uORFs is likely coupled to transcriptional control in yeast meiosis, where the transcription start sites of certain genes shift upstream to include one or more AUG-initiated uORFs, which is accompanied by diminished TE of the downstream ORFs (*29*). Termination at an uORF stop codon can elicit the same mRNA destabilization evoked by the nonsense-mediated decay pathway at premature termination codons in ORFs, magnifying the inhibitory effects of uORFs (*9*, *30*).

Despite their widespread occurrence, direct evidence that particular uORFs inhibit translation of downstream ORFs exists only for a relatively small number of genes, with two primary control mechanisms at play. For one class of regulatory uORFs, the encoded peptide acts to stall the elongating 80*S* ribosome engaged in its synthesis at or near the uORF stop codon, creating a "roadblock" to scanning PICs that leaky-scanned the uORF AUG codon (Fig. 2B). This roadblock can be modulated by ligands to

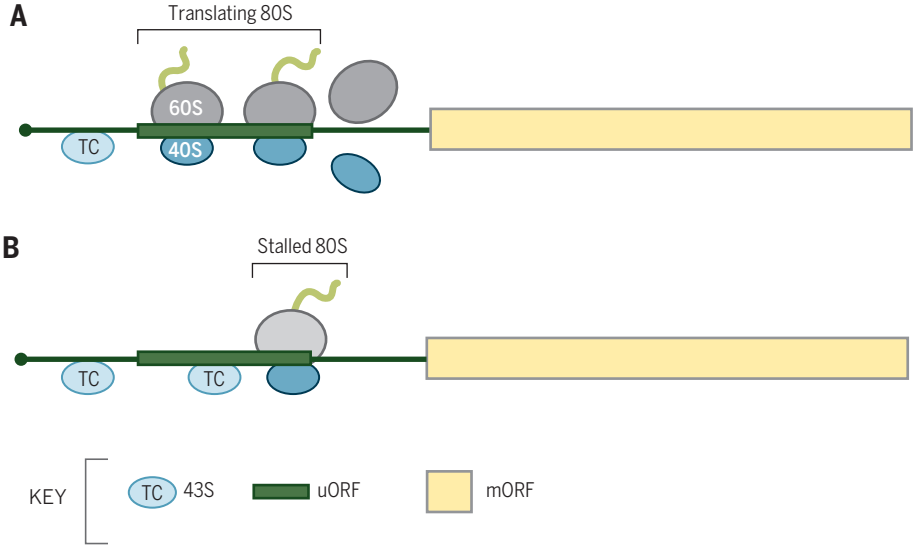


Fig. 2. Mechanisms of translational control by short uORFs. (**A**) The 43S PICs scanning from the mRNA 5′ end translate the uORF (as 80S ribosomes), and free subunits dissociate from the mRNA after termination, preventing translation of the main ORF (mORF). (**B**) The 80S ribosomes are stalled during elongation or termination by the uORF-encoded attenuator peptide and impose a barrier to scanning 43S PICs that leaky-scan the uORF start codon, preventing translation of the mORF. Stalling is modulated by small molecules.

achieve translational control—e.g., arginine for yeast *CPA1* and *Neurospora crassa arg-2* attenuator peptides (*31*, *32*) or spermidine for *ADM1* (*33*, *34*).

An encoded peptide sequence is irrelevant for a second class of regulatory uORFs that function only to waylay scanning PICs from the downstream ORF start codon (Fig. 2A). That the barrier imposed by such uORFs is generally overcome by leaky scanning is suggested by genome-wide data indicating that uORFs whose AUG codons better conform to the Kozak consensus are more inhibitory (*23*, *28*). Also, that upstream start codons tend to be near-cognates or AUGs in poor context should favor leaky scanning (*23*). Leaky scanning of an inhibitory uORF, through an unknown mechanism, is increased under stress conditions that evoke phosphorylation of eIF2 on serine-51 of its α-subunit eIF2(αP) (Fig. 3A) (*35*). This applies to GADD34, a targeting subunit for protein phosphatase-1 that dephosphorylates eIF2α, enabling autoregulation of eIF2(αP) accumulation (*36*). Phosphorylation of eIF2α converts eIF2-GDP into a competitive inhibitor of the guanine nucleotide exchange factor eIF2B and thereby decreases TC assembly (Fig. 1) (*37*). This might allow a fraction of PICs scanning from the cap to reach the uORF without harboring the TC, bypass the uAUG (owing to the absence of Met-tRNA$_i$), and bind TC while scanning the remainder of the 5′UTR, and initiate at the main ORF. Alternatively, phosphorylation might also alter eIF2 function in start codon recognition to allow leaky scanning even with TC bound to the PIC.

The presence of multiple uORFs can greatly amplify the effect of eIF2α phosphorylation on leaky scanning, as demonstrated first for yeast *GCN4* (*37*) and subsequently mammalian *ATF4* and *ATF5* (*38*, *39*), which encode transcription factors instrumental in responding to stresses that activate eIF2α kinases, such as amino acid deprivation for kinase Gcn2 (Fig. 1) and endoplasmic reticulum stress for protein kinase R–like endoplasmic reticulum kinase (PERK). The first (uORF1) is translated by most scanning PICs and optimized to allow a fraction of 40*S* ribosomes to remain attached to the 5′UTR and reinitiate downstream (*37*). For *GCN4* uORF1, sequences/structures upstream of the uORF functionally interact with the a-subunit of eIF3 and AU-rich sequences 3′ of the uORF stop codon to allow scanning to resume (*40*). With abundant TC in nonstressed cells, "rescanning" PICs rebind TC rapidly and efficiently reinitiate at the downstream uORF(s), optimized to evict postermination 40*S* subunits from the mRNA and prevent translation of the downstream main ORF. The decreased TC levels evoked by eIF2α phosphorylation allow a fraction of rescanning PICs to rebind TC only after leaky scanning the inhibitory uORFs, and initiate downstream at the ORF instead (Fig. 3B). Because of minimal leaky scanning of the inhibitory uORFs in nonstressed cells, owing to their optimum context, only a modest reduction in their recognition engenders large increases in main ORF translation

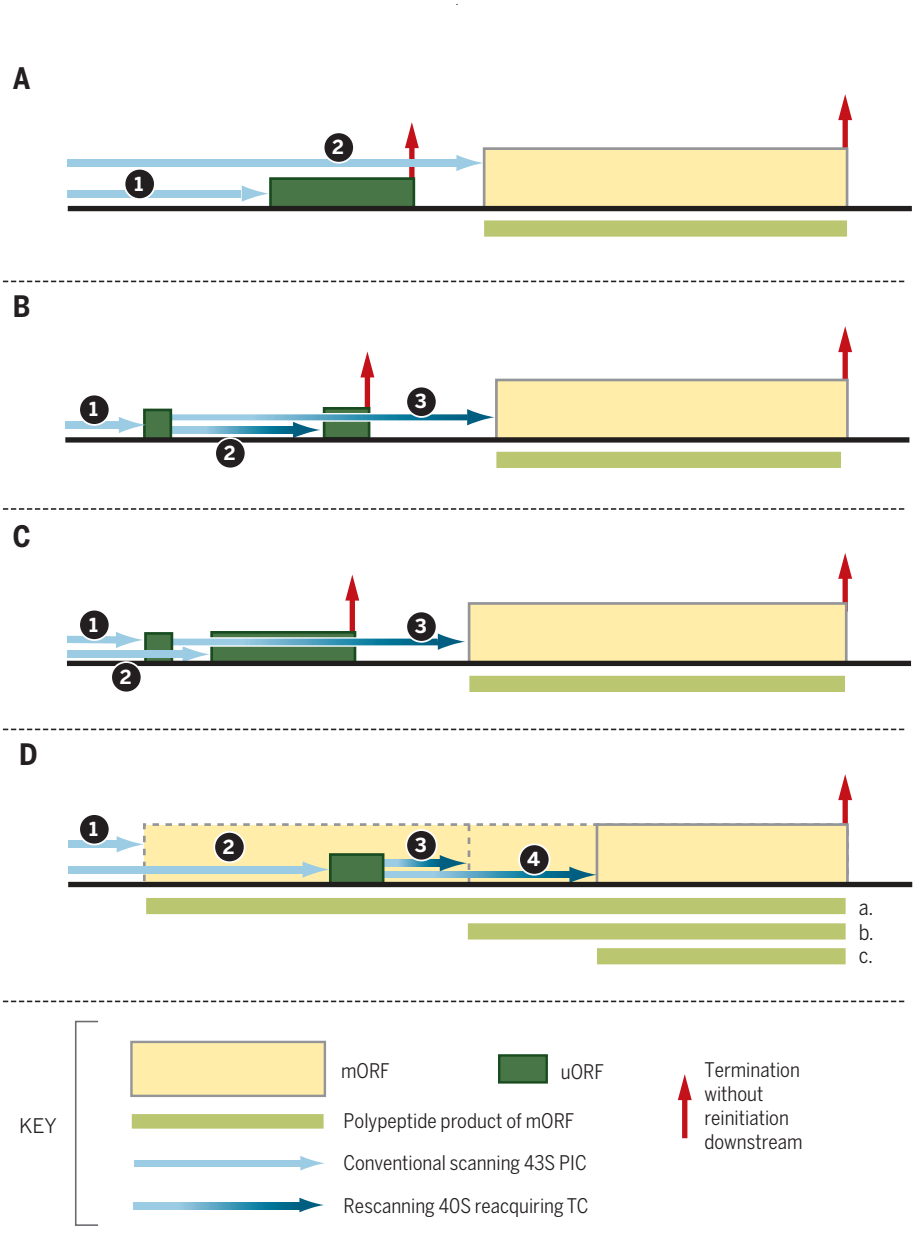


Fig. 3. Different gene architectures conferring translational control by short uORFs. (**A**) 1. Scanning PICs that translate the uORF fail to reinitiate at the mORF, as depicted in Fig. 2. 2. A fraction of scanning PICs leaky-scan the uORF start codon, enhanced by its suboptimal context, and initiate at the mORF instead. Leaky scanning can be inhibited by elevated eIF5 levels, lowering translation of the eIF5 gene itself (*4*); by eIF2(αP), e.g., for *GADD34* (*36*) and *IFRD* (*35*); and by polyamines for *AMD1* (encoding S-adenosylmethionine decarboxylase (SAMDC)) and *AZIN1* (*34*). (**B**) 1. Scanning ribosomes initiate translation of a short uORF whose translation does not preclude reinitiation. 2. Resumed scanning followed by quick reacquisition of TC enables translation of an inhibitory downstream uORF that precludes further reinitiation. 3. Slow reacquisition of TC at reduced TC concentrations evoked by eIF2(αP) allows reinitiation further downstream at the mORF. Examples include *GCN4* (*37*) and *ATF4* (*38*, *39*). (**C**) 1. Scanning ribosomes initiate translation of a short uORF permissive for reinitiation. 2. Ribosomes that leaky-scan the first uORF translate a second inhibitory uORF that precludes reinitiation. 3. Ribosomes that translate the first uORF resume scanning and reacquire TC only after bypassing the second uORF, avoiding its inhibitory effect, exemplified by polyamine regulation of SAMDC synthesis in plants (*34*). (**D**) 1. Ribosomes initiate at an upstream start codon in-frame with the mORF and bypass an inhibitory uORF during elongation while producing protein isoform "a" with specific properties. 2. Scanning subunits bypass the in-frame start site, owing to its suboptimal context, and initiate downstream at the uORF. 3. Rescanning followed by quick reacquisition of TC leads to reinitiation at a proximal start codon to produce protein isoform "b". 4. Slow reacquisition of TC allows initiation further downstream producing the shortest isoform, "c," with activities opposite those of "a" and "b." Examples include C/EBP-α and C/EBP-β (*49*).

in stressed cells (*37*). This mechanism enables the rapid, strategic induction of key transcription factors, while the reduced TC levels dampen bulk protein synthesis, for a two-pronged stress response.

The short length of uORF1 is crucial for reinitiation and might facilitate retention of eIF3 during its translation (*37*). Reinitiation after longer uORFs requires additional cis-acting sequences, such as the termination upstream ribosomal binding site (TURBS) element of polycistronic calicivirus mRNA that base pairs with 18*S* rRNA sequences of the 40*S* subunit (*41*, *42*), similar to Shine-Dalgarno sequences in bacterial mRNAs and eukaryotic viral IRESs such as in HCV. Some cellular IRESs might also function in this manner (*43*, *44*). The role of eIF3 subunits in promoting reinitiation appears widespread (*45*, *46*). Accessory factors, including ligatin/eIF2D or related proteins MCT-1 and DENR, can promote reinitiation, possibly through an alternative pathway for recovering Met-tRNA$_i$ by rescanning PICs (*42*, *47*).

The role of uORFs in regulating reinitiation in response to eIF2(αP) was implicated in learning and memory, neurodegenerative diseases, and cancer. For example, eliminating Gcn2 improved memory in mice owing to decreased eIF2(αP), leading to reduced *ATF4* translation (*48*). Alternative outcomes of leaky scanning versus reinitiation imposed by a uORF in CCAAT/enhancer-binding protein (C/EBP) mRNAs determine the balance of isoforms differing at their N termini, one activating and the other inhibiting transcription, important for mouse liver differentiation and regeneration (Fig. 3D) (*49*). An unusual role for a uORF in facilitating translational repression of *Drosophila msl-2* mRNA in conjunction with 5′UTR binding by sex-lethal (SXL) protein regulates dosage compensation (*50*). There is a growing list of mutations associated with human disease that increase or decrease the influence of uAUGs/uORFs on translation of the main ORF (*28*, *51*).

Other 5′UTR regulatory elements

The 5′ terminal oligopyrimidine (5′TOP) motif plays a role in mammalian target of rapamycin (mTOR)–dependent stimulation of the expression of proteins of the translation machinery to promote cell growth. mTOR complex 1 (mTORC1) activates the La-related protein 1 (LARP1) that binds to the TOP sequence (*52*). Many less-abundant mRNAs lacking 5′TOP exhibit mTOR dependence, encoding mitochondrial and growth/survival-promoting proteins (*53*). Additionally, the m^6-adenine methylation of 5′UTR sequences seems to have stimulatory effects (*54*).

That uORFs appear to influence translation genome-wide suggests that scanning operates widely in the eukaryotic translatome. However, scanning can be circumvented by specialized elements that enable PICs to enter the 5′UTR internally. mRNAs with a stretch of unstructured nucleotides in the 5′UTR can bypass the requirement for the m^{7}G cap and eIF4F, as shown for poly(A) sequences in poxvirus mRNA 5′UTRs (*55*) and CAA nucleotide triplet repeats in the Ω leader of tobacco mosaic virus mRNAs. Although dispensable for 48*S* assembly per se, eIF4A accelerated the process on a Ω reporter (*56*). eIF4F independence for 48*S* PIC assembly was also observed for mRNAs with synthetic unstructured 5′UTRs, which, like the poly(A) 5′UTRs, still require eIF1, eIF1A, TC, and (in mammals) eIF3 (*8*, *57*). Unstructured nucleotides might bind directly in the mRNA binding cleft of the 40*S* subunit, with ATP hydrolysis by eIF4A enabling subsequent 5′-to-3′ directional scanning. A group of mRNAs in yeast are refractory to widespread translational repression in carbon-starved cells and contain a poly(A) stretch in their 5′UTRs that might recruit eIF4F to the 5′UTR via PABP-eIF4G association, augmenting cap-binding by eIF4F (*58*).

Many viral mRNAs circumvent the scanning process with highly structured IRES elements that interact with the 40*S* subunit or particular eIFs to recruit the PIC to internal sites in the 5′UTR (*59*). These mechanisms persist when cap/eIF4F-dependent initiation is impaired, such as in virus infections where eIF4G is cleaved. Whereas viral IRESs have been extensively characterized biochemically and structurally (*59*), such studies have not been accomplished for potential IRESs in cellular mRNAs (*60*). A genome-wide search yielded a large number of mammalian cellular IRESs (10% of randomly selected 5′UTRs) (*44*), which, if validated for individual mRNAs, would be highly important for understanding gene regulation in humans.

Important progress has been made in elucidating mechanisms by which the 5′UTR regulates translation initiation. This includes molecular and structural understanding of the assembly and recruitment of the PIC to the 5′UTR, scanning, and start codon selection. Acutely lacking is a precise kinetic analysis of the pathway. Single-molecule approaches can be expected to fill this gap and identify intermediate states too transient for detection by ensemble kinetics. Ribosome profiling (*22*) can be adapted to analyze the kinetics and regulation of scanning on all 5′UTRs. Advanced cryo–electron microscopy will continue to yield high-resolution structures of the PIC in different stages of initiation. The new information is bound to aid efforts to discover new drugs to treat diseases whose etiology is associated with dysregulated translation.

REFERENCES

1. A. G. Hinnebusch, *Microbiol. Mol. Biol. Rev.* **75**, 434–467 (2011).
2. M. Kozak, *Cell* **44**, 283–292 (1986).
3. S. R. Starck *et al.*, *Science* **336**, 1719–1723 (2012).
4. G. Loughran, M. S. Sachs, J. F. Atkins, I. P. Ivanov, *Nucleic Acids Res.* **40**, 2898–2906 (2012).
5. I. P. Ivanov, G. Loughran, M. S. Sachs, J. F. Atkins, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18056–18060 (2010).
6. T. F. Donahue, A. M. Cigan, E. K. Pabich, B. Castilho Valavicius, *Cell* **54**, 621–632 (1988).
7. K. Berthelot, M. Muldoon, L. Rajkowitsch, J. Hughes, J. E. McCarthy, *Mol. Microbiol.* **51**, 987–1001 (2004).
8. T. V. Pestova, V. G. Kolupaeva, *Genes Dev.* **16**, 2906–2922 (2002).
9. J. A. Arribere, W. V. Gilbert, *Genome Res.* **23**, 977–987 (2013).
10. S. A. Sedman, G. W. Gelembiuk, J. E. Mertz, *J. Virol.* **64**, 453–457 (1990).
11. H. Sinvani *et al.*, *Cell Metab.* **21**, 479–492 (2015).
12. M. Kozak, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8301–8305 (1990).
13. A. E. Koromilas, A. Lazaris-Karatzas, N. Sonenberg, *EMBO J.* **11**, 4153–4158 (1992).
14. I. S. Abaeva, A. Marintchev, V. P. Pisareva, C. U. Hellen, T. V. Pestova, *EMBO J.* **30**, 115–129 (2011).
15. N. D. Sen, F. Zhou, N. T. Ingolia, A. G. Hinnebusch, *Genome Res.* **25**, 1196–1205 (2015).
16. C. A. Rubio *et al.*, *Genome Biol.* **15**, 476 (2014).
17. A. L. Wolfe *et al.*, *Nature* **513**, 65–70 (2014).
18. Y. V. Svitkin *et al.*, *RNA* **7**, 382–394 (2001).
19. J. Pelletier, N. Sonenberg, *Mol. Cell. Biol.* **5**, 3222–3230 (1985).
20. R. Soto-Rifo *et al.*, *EMBO J.* **31**, 3745–3756 (2012).
21. M. W. Hentze, F. Gebauer, T. Preiss, in *Translational Control in Biology and Medicine*, M. B. Mathews, N. Sonenberg, J. W. B. Hershey, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2007), pp. 269–295.
22. N. T. Ingolia, S. Ghaemmaghami, J. R. Newman, J. S. Weissman, *Science* **324**, 218–223 (2009).
23. S. Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2424–E2432 (2012).
24. T. G. Johnstone, A. A. Bazzini, A. J. Giraldez, *EMBO J.* **35**, 706–723 (2016).
25. B. Vanderperre *et al.*, *PLOS ONE* **8**, e70698 (2013).
26. F. Zhang, A. G. Hinnebusch, *Nucleic Acids Res.* **39**, 3128–3140 (2011).
27. M. V. Gerashchenko, V. N. Gladyshev, *Nucleic Acids Res.* **42**, e134 (2014).
28. S. E. Calvo, D. J. Pagliarini, V. K. Mootha, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7507–7512 (2009).
29. G. A. Brar *et al.*, *Science* **335**, 552–557 (2012).
30. A. Gaba, A. Jacobson, M. S. Sachs, *Mol. Cell* **20**, 449–460 (2005).
31. M. Werner, A. Feller, F. Messenguy, A. Piérard, *Cell* **49**, 805–813 (1987).
32. J. Wei, C. Wu, M. S. Sachs, *Mol. Cell. Biol.* **32**, 2396–2406 (2012).
33. A. Raney, G. L. Law, G. J. Mize, D. R. Morris, *J. Biol. Chem.* **277**, 5988–5994 (2002).
34. I. P. Ivanov, J. F. Atkins, A. J. Michael, *Nucleic Acids Res.* **38**, 353–359 (2010).
35. D. E. Andreev *et al.*, *eLife* **4**, e03971 (2015).
36. S. K. Young, J. A. Willy, C. Wu, M. S. Sachs, R. C. Wek, *J. Biol. Chem.* **290**, 28257–28271 (2015).
37. A. G. Hinnebusch, *Annu. Rev. Microbiol.* **59**, 407–450 (2005).
38. P. D. Lu, H. P. Harding, D. Ron, *J. Cell Biol.* **167**, 27–33 (2004).
39. K. M. Vattem, R. C. Wek, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11269–11274 (2004).
40. B. Szamecz *et al.*, *Genes Dev.* **22**, 2414–2425 (2008).
41. C. Luttermann, G. Meyers, *Genes Dev.* **23**, 331–344 (2009).
42. A. Zinoviev, C. U. Hellen, T. V. Pestova, *Mol. Cell* **57**, 1059–1073 (2015).
43. S. A. Chappell, J. Dresios, G. M. Edelman, V. P. Mauro, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 9488–9493 (2006).
44. S. Weingarten-Gabbay *et al.*, *Science* **351**, aad4939 (2016).
45. H. S. Park, A. Himmelbach, K. S. Browning, T. Hohn, L. A. Ryabova, *Cell* **106**, 723–733 (2001).
46. B. Roy *et al.*, *RNA* **16**, 748–761 (2010).
47. S. Schleich *et al.*, *Nature* **512**, 208–212 (2014).
48. M. Costa-Mattioli *et al.*, *Nature* **436**, 1166–1173 (2005).
49. K. Wethmar *et al.*, *Genes Dev.* **24**, 15–20 (2010).
50. J. Medenbach, M. Seiler, M. W. Hentze, *Cell* **145**, 902–913 (2011).
51. J. Somers *et al.*, *Genes Dev.* **29**, 1891–1896 (2015).
52. J. Tcherkezian *et al.*, *Genes Dev.* **28**, 357–371 (2014).
53. V. Gandin *et al.*, *Genome Res.* **26**, 636–648 (2016).
54. K. D. Meyer *et al.*, *Cell* **163**, 999–1010 (2015).
55. N. E. Shirokikh, A. S. Spirin, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10738–10743 (2008).
56. S. Ch. Agalarov, P. A. Sakharov, D. Kh. Fattakhova, E. A. Sogorin, A. S. Spirin, *Sci. Rep.* **4**, 4438 (2014).
57. M. A. Algire *et al.*, *RNA* **8**, 382–397 (2002).
58. W. V. Gilbert, K. Zhou, T. K. Butler, J. A. Doudna, *Science* **317**, 1224–1227 (2007).
59. J. S. Kieft, *Trends Biochem. Sci.* **33**, 274–283 (2008).
60. R. J. Jackson, *Cold Spring Harb. Perspect. Biol.* **5**, a011569 (2013).

10.1126/science.aad9868

REVIEW

From the RNA world to the clinic

Bruce A. Sullenger* and Smita Nair

The study of RNA has continually emphasized the structural and functional versatility of RNA molecules. This versatility has inspired translational and clinical researchers to explore the utility of RNA-based therapeutic agents for a wide variety of medical applications. Several RNA therapeutics, with diverse modes of action, are being evaluated in large late-stage clinical trials, and many more are in early clinical development. Hundreds of patients are enrolled in large trials testing messenger RNAs to combat cancer, small interfering RNAs to treat renal and hepatic disorders, and aptamers to combat ocular and cardiovascular disease. Results from these studies are generating considerable interest among the biomedical community and the public and will be important for the future development of this emerging class of therapeutic agents.

During the past 25 years, the idea of using RNA molecules as therapeutic agents has grown from a concept into a clinical reality (*1*, *2*). Initially, RNA was thought to be a poor choice for a therapeutic agent, given its relatively short half-life in vivo. However, with improvements in stabilization chemistry, and by making informed decisions about where a relatively short-lived molecule may be useful clinically, much of this skepticism has been overcome. RNA molecules possess many properties that potentially make them useful therapeutics. They can fold into complex conformations that allow them to specifically bind proteins, small molecules, or other nucleic acids or even to form catalytic centers (*3*). Coding mRNAs are the exclusive carriers of genetic instructions from DNA to the ribosome. Many types of noncoding RNA (ncRNA) orchestrate the mobilization and utilization of genetic information through a variety of mechanisms whose discovery has dramatically changed our understanding of RNA biology (*3*). The four classes of therapeutic RNAs that have received the most attention can be classified by their modes of action: (i) encoding therapeutic proteins or vaccine antigens (mRNAs), (ii) inhibiting pathogenic RNA activity [small interfering RNAs (siRNAs), microRNAs (miRNAs), and antisense RNAs], (iii) modulating protein activity (RNA aptamers), and (iv) reprogramming genetic information [trans-splicing ribozymes and CRISPR guide RNAs (gRNAs)]. Here we briefly review progress in and challenges for the clinical development of such coding and noncoding RNA therapeutics, focusing on those that have reached large phase 3 clinical studies (table S1).

Therapeutic coding RNAs: mRNA-based immunotherapy

Preclinical studies suggest that direct delivery of therapeutic mRNAs into cells ex vivo or in vivo may result in substantial health benefits in a number of clinical settings where transient generation of a therapeutic protein is effective and often preferred. mRNAs that encode tumor antigens have been used to induce tumor-specific immune responses. Such mRNA-based vaccination has largely been performed by either loading dendritic cells (DCs) with mRNA ex vivo and reintroducing these DCs back into the body to translate, present, and elicit an immune response against the encoded tumor antigen (Fig. 1) (*4*, *5*), or by direct injection of tumor antigen–encoding mRNA in vivo for uptake, translation, and presentation by antigen-presenting cells such as DCs (*6*, *7*). In animal studies, such transient antigen presentation has led to long-lasting antitumor immune responses. Phase 1 clinical trials have been carried out in patients with several types of cancer [reviewed in (*4*, *5*)], and an open-label phase 2 clinical study suggests that mRNA-based DC immunotherapy can improve median overall survival rates in renal cancer patients (*8*). Other phase 2 clinical studies in HIV-infected individuals have demonstrated that introducing mRNAs that encode viral antigens into DCs can induce long-lasting HIV-specific T cell responses (*9*, *10*). Thus, even transient expression of tumor or viral antigens via mRNA transfection can result in prolonged effects. An ongoing phase 3 clinical study is using patient-derived DCs loaded with mRNA amplified from metastatic renal cell carcinoma tumors (Fig. 1 and table S1); this will help determine whether such transient mRNA-based immunotherapy improves long-term survival in patients.

While this initial ex vivo strategy of using mRNA-based immunotherapy has been moving through clinical development, several improvements have been established to enhance mRNA-induced immunity. Electroporation has been used to more effectively load DCs with mRNA ex vivo (*11*). To further enhance the immune response, mRNAs encoding CD40L, CD70, and a constitutively active form of TLR4 have been delivered in conjunction with tumor antigens to generate optimized human DCs for use as cancer vaccines (*12*). Also, preconditioning the skin site before intradermal injection of a DC-loaded mRNA vaccine leads to increased DC migration to lymph nodes and improved survival in patients with glioblastoma (*13*).

Ex vivo DC vaccine strategies require complicated personalized cell therapy. Direct injection in vivo of mRNA encoding tumor antigens has the potential to develop into an off-the-shelf therapeutic agent. Direct injection of naked or protamine-protected mRNA leads to the induction of T and B cell responses (*14*), and intradermal administration of mRNA promotes the induction of antitumor immunity (*15*). However, such unmodified mRNA is expected to have a very short half-life in vivo. To optimize mRNA pharmacokinetics and limit unwanted inflammatory responses, investigators have modified elements of mRNA structure such as the 5′ and 3′ untranslated region; 5′ cap; coding

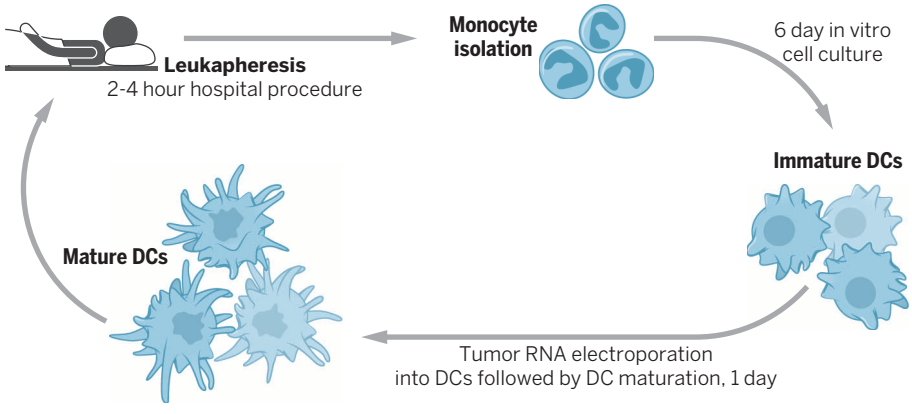


Fig. 1. RNA-based tumor vaccines. A DC-based vaccine is made using tumor antigen, either RNA isolated from a patient's tumor biopsy or mRNA encoding a known tumor antigen. Patient peripheral blood mononuclear cells are used to isolate $CD14^+$ monocytes, which in turn are used to generate immature DCs. RNA is transfected into the patient's immature DCs, followed by DC activation and maturation ex vivo. These RNA-transfected mature DCs are administered to patients to elicit an antitumor immune response.

Duke Translational Research Institute and Department of Surgery, Duke University Medical Center, Durham, NC 27710, USA.
*Corresponding author. Email: bruce.sullenger@duke.edu

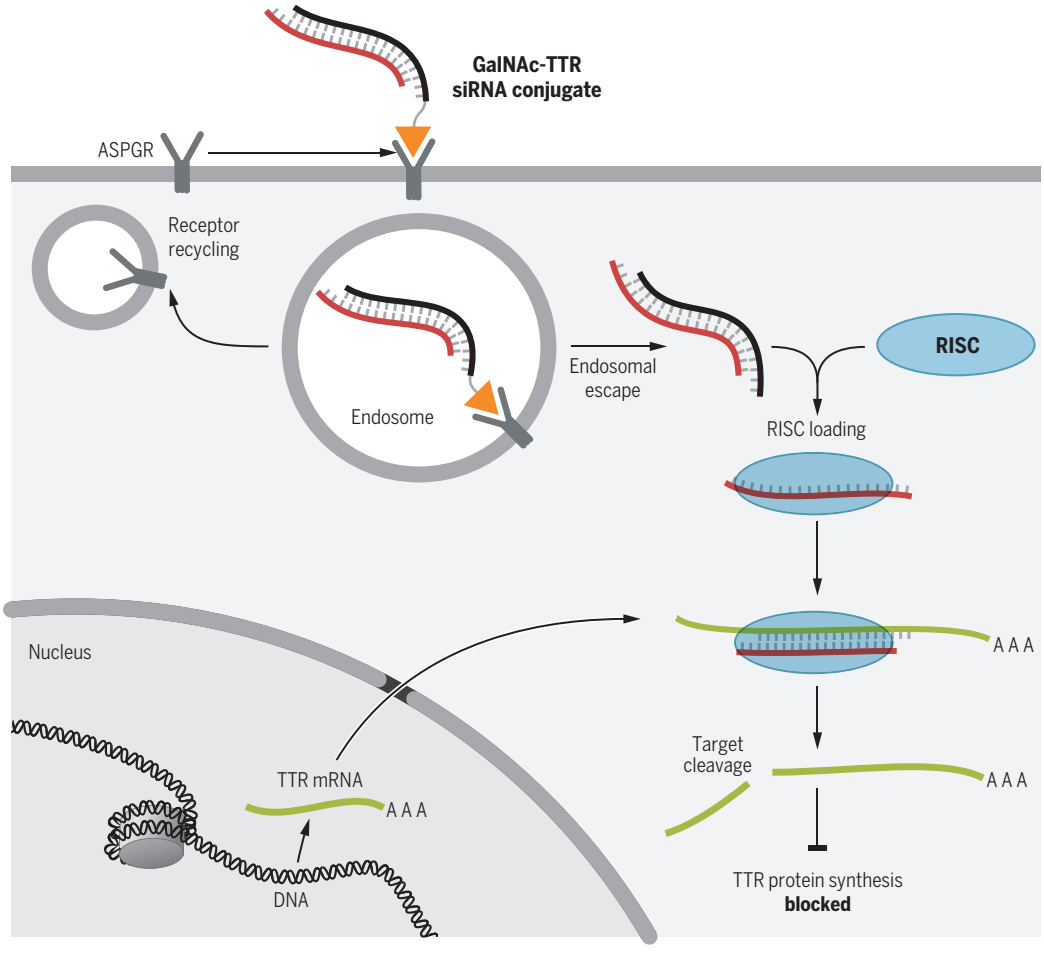


Fig. 2. siRNA-mediated treatment of TTR amyloidosis. siRNA-mediated inhibition of mutant TTR synthesis in the liver is used to limit TTR protein generation and subsequent protein misfolding, oligomerization, and systemic pathology. GalNAc (orange triangle) is appended to the siRNA (ALN-TTRSC, black line) to facilitate delivery to hepatocytes through the asialoglycoprotein receptor (ASGPR) and release from the endosomal compartment into the cytoplasm, where the targeting strand (red line) is taken up by the RISC complex. The RISC complex then targets the TTR mRNA (green line; AAA, poly(A) tail) for destruction, preventing synthesis of the TTR protein.

region, poly(A) tail, and nucleotide chemistry [reviewed in-depth in (*16*)]. These optimized mRNAs are now being evaluated in ongoing clinical studies to determine whether they can effectively replace the more complicated and costly cell-based approaches.

In addition to active immunization, mRNA can be used to transiently generate T cells (or natural killer cells) that express tumor antigen–specific T cell receptors or chimeric antigen receptors for therapeutic applications (*17*, *18*). Cell therapy using T cells electroporated with mRNA encoding mesothelin chimeric antigen receptors is in clinical trials in patients with pancreatic cancer and other solid malignancies (*19*). Lastly, in addition to being used as cancer vaccines, tumor mRNA–loaded DCs have been used to expand autologous T cells in vitro for adoptive cell therapy. Thus, coding RNA is central to a number of potential applications in immunotherapy settings where even transient production of a therapeutic agent can elicit a prolonged therapeutic effect.

Therapeutic proteins generated by mRNA

mRNA can be used to generate a variety of therapeutic proteins for applications that include protein replacement therapy, regenerative medicine, and therapeutic genome editing. Methods have been developed to reduce innate immune activation caused by the mRNA itself and to increase mRNA translation by incorporating modified nucleotides (*16*), because many therapeutic protein replacement applications will probably require repeated administration.

mRNAs can be used to reprogram cells. Electroporation into B cells of a cocktail of four different mRNAs encoding immune stimulatory proteins produces reprogrammed cells that are similar to natural DCs and as effective at stimulating T cells (*20*). A cocktail of mRNAs that encode four transcription factors was shown to efficiently reprogram cells to pluripotency, even though these mRNAs are expected to have a relatively short half-life in cells (*21*). mRNA-based approaches will probably be evaluated in the clinic in the near future, given that using mRNA for regenerative medicine applications poses less of a safety concern than DNA-based reprogramming. Another potential application of mRNA-based therapeutics is to transiently generate ZFNs (zinc-finger nucleases), TALENs (transcription activator-like effector nucleases), and Cas9 for therapeutic genome editing. For example, genetically modified cynomolgus monkeys have been generated by injecting Cas9 mRNA and CRISPR gRNA into one-cell embryos (*22*). Given ongoing concerns about off-target effects, mRNA-based transient expression of such nucleases is being actively pursued as a strategy to reduce risk. A number of clinical studies slated to begin in 2016 or 2017 will explore these new applications. Thus, though coding RNAs only transiently induce protein production, an increasing number of potential clinical applications appear to be well suited for these relatively short-lived therapeutic agents.

The three translational applications of therapeutic ncRNAs that have received the most clinical attention to date are (i) inhibition of mRNA translation, (ii) protein modulation, and (iii) genetic reprogramming. As shown in table S1, 10 ncRNA-based therapeutics have already entered large randomized phase 3 clinical studies.

ncRNA-mediated inhibition of mRNA translation

Antisense RNA binding to and regulating the expression of a complementary mRNA is a well-characterized phenomenon in prokaryotes, and expression of antisense RNA in mammalian cells has been used to inhibit a wide variety of target mRNAs. These observations have led to the development of several antisense oligonucleotide–based therapeutic agents that have been making their way through clinical trials. Because many of these antisense oligonucleotides contain DNA, we do not discuss them here in detail; a recent review considers the clinical development of this class of therapeutic agents (*23*).

The discovery of RNA interference (RNAi) (*24*) greatly increased interest in the potential utility of ncRNAs that can inhibit therapeutically relevant mRNAs through posttranscriptional gene silencing. Gene silencing can be induced by synthetic siRNAs (*25*) or miRNA mimics (*1*) or by

expression of short hairpin RNAs (shRNAs) or primary miRNAs that are processed into therapeutic siRNAs or miRNAs. After loading into the RNA-induced silencing complex (RISC), targeted RNAs have their translation repressed (miRNA) or are cleaved and degraded (siRNA). In addition, once associated with silencing complexes, RNAi can induce the silencing of multiple target RNAs. Moreover, related RNA complexes have recently been shown to induce transcriptional gene silencing, which may elicit an even longer-lasting effect in cells (*26*). More than 20 clinical trials are underway evaluating the potential utility of RNAi-based therapeutics for a range of human diseases. RNAi-based therapies that have incorporated modified nucleotides to limit innate immune system activation seem to be well tolerated in patients (*26*), and six RNAi-based therapeutic agents have progressed into phase 3 clinical trials (table S1). Of the six, five use synthetic siRNAs and the sixth uses an expressed bifunctional shRNA in cancer cells. The shRNA approach involves two shRNAs targeted to furin mRNA (*27*) and is designed to augment immune responses against patient-derived tumor cells that are transfected with the shRNA expression cassette, lethally irradiated, and intradermally re-administered to the patient. Currently, this ex vivo shRNA-based cancer immunotherapy approach is being evaluated in a phase 3 clinical study to determine whether such transient gene silencing can improve long-term survival in cancer patients (table S1). Several other RNAi-based therapeutic agents are being evaluated for treatment of a variety of cancers, and several that do not involve ex vivo cell therapy are progressing through early clinical development [reviewed in (*26*)].

Two major pharmacokinetic challenges exist for developing synthetic oligonucleotide therapeutics: their limited oral bioavailability and the rapid rate at which short nuclease-resistant RNAs are cleared from circulation. All five of the current synthetic siRNA phase 3 clinical trials either use local delivery of siRNA to the eye or target delivery of siRNA to the liver or kidney, organs that are involved in oligonucleotide clearance after intravenous systemic administration (table S1). The siRNA bevasiranib was designed to silence vascular endothelial growth factor (VEGF) mRNA after direct intraocular administration in patients with age-related macular degeneration. Despite encouraging phase 2 results, this first-generation siRNA-based therapy did not reach the intended clinical endpoint in a phase 3 trial and was discontinued; this is probably because the chemistry used was not particularly potent, and because other VEGF inhibitors (aptamer and antibody) made it to market first and proved superior (*28*). Subsequent early-phase clinical studies with siRNAs targeting different mRNAs in the eye have been more promising (*26*). The ocular siRNA QPI-1007 targets Cas2 (*29*) to reduce retinal ganglion apoptosis in patients with optic neuropathy (table S1). This chemically modified siRNA drug has been granted an orphan drug designation by the U.S. Food and Drug Administration (FDA) and is moving into a phase 3 clinical trial. A large clinical trial is under way of an intravenously administered siRNA (QPI-1002) that targets p53 mRNA, localizes to the kidney, and is designed to transiently limit p53-induced apoptosis and thereby mitigate programmed cell death and kidney injury after transplantation (table S1) (*30*). Thus, siRNA appears to be particularly well suited for local or transient therapeutic applications.

Advances in siRNA formulation allow efficient delivery to the liver (*26*). The utility of siRNA-mediated silencing of transthyretin (TTR) mRNA in the liver of patients with familial TTR amyloidosis (*31*, *32*) is being studied in two separate phase 3 clinical trials. The first trial uses a siRNA formulated with lipid nanoparticle technology (ALN-TTR02) and intravenous administration to attempt to improve neuropathy in patients. This is the therapeutic siRNA that is closest to potential approval. In contrast, the second trial uses a TTR-targeted siRNA (ALN-TTRSC) that is formulated with *N*-acetylgalactosamine (GalNAc) to facilitate uptake by and release into the cytoplasm of hepatocytes, as well as to allow for subcutaneous administration (Fig. 2 and table S1). Many other receptor-targeting agents are being evaluated preclinically to determine whether siRNA can be selectively and effectively delivered to the cytoplasm of other cell types in other tissues.

ncRNAs as protein antagonists

Certain viruses express short, structured ncRNAs (aptamers) that bind and antagonize cellular proteins involved in antiviral responses. That these aptamers can have therapeutic applications has been demonstrated by the expression of the short, stem-loop HIV-derived TAR ncRNA in $CD4^+$ T cells, where it bound and repressed the viral protein Tat and thereby inhibited HIV replication (*33*). RNA aptamers can be generated against novel targets in vitro by using a combinatorial chemistry method termed SELEX (systematic evolution of ligands by exponential enrichment); well over 1000 proteins can now be targeted by high-affinity and -specificity aptamers, including many proteins of therapeutic interest (*34*). SELEX allows the incorporation of modified nucleotides so that RNA aptamers can be generated that are highly nuclease-resistant and therefore suitable for animal and clinical studies. Because nuclease-resistant RNA aptamers can target extracellular proteins and do not have to cross the cell membrane to elicit a therapeutic effect, their clinical development has focused largely on targeting soluble extracellular proteins or extracellular domains of cell-surface receptors. Several aptamers have now been evaluated in the clinic (*1*), with four moving into large phase 3 clinical studies (table S1).

Ocular diseases present a natural fit for the clinical development of therapeutic aptamers, because local delivery to the eye (*2*) would theoretically eliminate pharmacokinetic challenges associated with systemic delivery of short oligonucleotides. Aptamers have been created against several proteins that are thought to be important therapeutic targets for ocular indications, including VEGF, platelet-derived growth factor (PDGF), and complement component 5 (C5) (*2*); aptamers targeting each of these are in or have completed phase 3 clinical trials. Patients suffering from neovascular age-related macular degeneration and treated with the aptamer that targets VEGF (Macugen) demonstrated a 45% improvement in limiting disease progression compared with the standard of care (*2*). Macugen received FDA approval for ophthalmic use in 2004 and reached sales of $185 million in the United States in 2005. Protein-based inhibitors of VEGF in the eye have since largely replaced Macugen in the market because of their improved efficacy (*2*). Nonetheless, Macugen demonstrated that an RNA therapeutic was marketable and addressed a then-unmet medical need. Two additional aptamers targeting PDGF and C5 in combination with VEGF protein inhibitors (e.g., Lucentis) are now progressing through phase 3 clinical studies (table S1) after encouraging phase 2 studies, which indicate that such aptamer-based combination therapy further improves outcomes for patients with age-related macular degeneration (*2*).

A drug-antidote approach has been used to develop a rapid-onset and rapidly reversible RNA aptamer–based anticoagulant agent for use in acute care and clinical settings requiring short-term yet potent anticoagulation, such as percutaneous coronary intervention in cardiovascular disease patients. The aptamer RB006, which was developed for systemic delivery [formulated with a 40-kDa polyethylene glycol (PEG)], targets coagulation factor IXa (*35*), and the matched RNA oligonucleotide (RB007) is the "antidote" that forms base pairs with the aptamer (Fig. 3). This drug-antidote approach highlights a major advantage of aptamers over antibody-based drugs: It is easy to generate antidotes to aptamers to provide physicians with additional control over aptamer-based medications. Unfortunately, after treating only 25% of an anticipated 6600 patients, the phase 3 trial was halted because of serious allergic reactions in a very small number of recipients that received the PEGylated aptamer (*36*). These rare allergic reactions are associated with the presence of high amounts of preexisting antibodies to PEG in allergic individuals (*37*). However, it is clear that the factor IXa RNA aptamer can serve as a potent rapid-onset anticoagulant, controlling coagulation and limiting thrombosis as effectively as the standard of care in this common clinical setting (Fig. 3). The study also demonstrated that RNA aptamers are rapidly controllable therapeutic agents with the potential to improve safety and efficacy and that alternative formulations to PEG should be considered to improve aptamer pharmacokinetic properties. The collective clinical experience with the ocular aptamers in phase 3 clinical studies and the anticoagulant activity achieved with the factor IXa aptamer in over 1600 cardiovascular disease patients bode well for the future development of RNA aptamers as a class of therapeutic agents. The major challenge for the widespread commercialization of RNA aptamer–based therapeutics remains the need to distinguish them

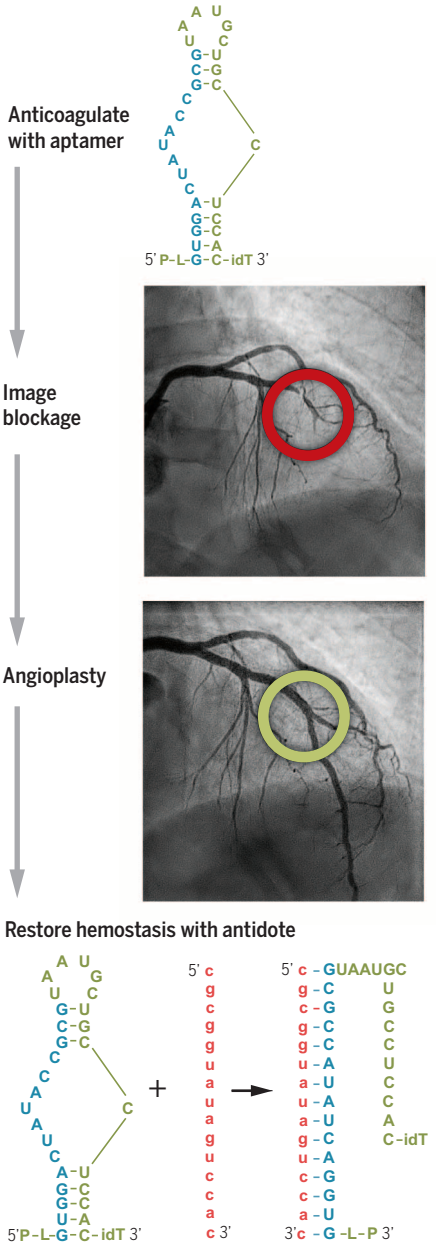


Fig. 3. The factor IXa aptamer anticoagulation system. An anticoagulant aptamer (RB006, blue and green) targeting blood coagulation factor IXa is used to control blood coagulation during percutaneous coronary intervention in patients with acute coronary syndrome (blue bases are the region of aptamer that is recognized by the antidote). Substantial anticoagulation is required to limit clotting during this procedure, in which a catheter is inserted into a peripheral artery and guided to the heart to administer a contrast agent that will indicate sites of blockage in a coronary angiography. If blockage is found in the arteries of a patient's heart (red circle), angioplasty is often performed to restore blood flow (green circle). To limit bleeding after treatment, the antidote oligonucleotide (RB007, pink) is administered to rapidly unfold the aptamer and restore normal blood hemostasis.

from monoclonal antibody competitors, which naturally have longer-lasting pharmacokinetic properties after systemic administration.

ncRNA-mediated genetic reprogramming

The observation that ncRNAs can serve as guides to target endonucleases—not simply to cleave and destroy genetic instructions but rather to reprogram them—has received considerable attention. ncRNAs created by trans-splicing ribozymes and CRISPR gRNAs can, in principle, be used to reengineer the transcriptome or genome for therapeutic applications (*38*, *39*). Though this RNA-based genetic reprogramming technology is not nearly as clinically advanced as other RNA therapeutics, the first therapeutic ncRNA designed to reprogram genetic instructions entered a phase 1 clinical trial in 2015. After encouraging preclinical animal studies, a trans-splicing ribozyme designed to recognize telomerase reverse transcriptase mRNA and reprogram it to encode herpes simplex virus–thymidine kinase only in cancer cells has now been tested in five cancer patients with hepatic metastases (*40*). In that study, the trans-splicing ribozyme was expressed after in vivo gene transfer, and the safety and potential efficacy of this ncRNA-mediated transcriptome engineering was assessed. Numerous RNA-guided endonuclease-based therapies are currently in preclinical development. If it can be adequately developed during the next decade, ncRNA-mediated genetic engineering may prove useful for treating a variety of inherited diseases (including genetic disorders) through genome engineering and acquired diseases (such as cancer) through transcriptome engineering, where transient reprogramming is expected to be sufficiently effective and also safer than permanently reprogramming the genome.

Conclusion

The field of RNA therapeutics is currently undergoing a major expansion and a continued renaissance. Eleven RNA-based therapeutics have now reached late-stage clinical development, with one on the market and several others close behind (table S1). The continued discovery of activities performed by RNAs, taken together with the increasing recognition that transient induction of a therapeutic effect can have long-lasting health benefits, promises to keep the clinical development pipeline full of therapeutic RNAs well into the future.

REFERENCES AND NOTES

1. J. C. Burnett, J. J. Rossi, *Chem. Biol.* **19**, 60–71 (2012).
2. D. W. Drolet, L. S. Green, L. Gold, N. Janjic, *Nucleic Acid Ther.* **26**, 127–146 (2016).
3. T. R. Cech, J. A. Steitz, *Cell* **157**, 77–94 (2014).
4. D. Benteyn, C. Heirman, A. Bonehill, K. Thielemans, K. Breckpot, *Expert Rev. Vaccines* **14**, 161–176 (2015).
5. S. Nair, D. Boczkowski, S. Pruitt, J. Urban, in *Cancer Vaccines: From Research to Clinical Practice*, A. Bot, M. Obrocea, F. M. Marincola, Eds. (CRC Press, ed. 1, 2011), pp. 217–231.
6. A. Bringmann, S. A. Held, A. Heine, P. Brossart, *J. Biomed. Biotechnol.* **2010**, 623687 (2010).
7. S. Van Lint *et al.*, *Expert Rev. Vaccines* **14**, 235–251 (2015).
8. A. Amin *et al.*, *J. Immunother. Cancer* **3**, 14 (2015).
9. S. D. Allard *et al.*, *Clin. Immunol.* **142**, 252–268 (2012).
10. E. Van Gulck *et al.*, *AIDS* **26**, F1–F12 (2012).
11. V. F. Van Tendeloo *et al.*, *Blood* **98**, 49–56 (2001).
12. A. M. Van Nuffel *et al.*, *Cancer Immunol. Immunother.* **61**, 1033–1043 (2012).
13. D. A. Mitchell *et al.*, *Nature* **519**, 366–369 (2015).
14. I. Hoerr, R. Obst, H. G. Rammensee, G. Jung, *Eur. J. Immunol.* **30**, 1–7 (2000).
15. R. D. Granstein, W. Ding, H. Ozawa, *J. Invest. Dermatol.* **114**, 632–636 (2000).
16. U. Sahin, K. Karikó, Ö. Türeci, *Nat. Rev. Drug Discov.* **13**, 759–780 (2014).
17. Y. Zhao *et al.*, *Cancer Res.* **70**, 9053–9061 (2010).
18. S. H. Yoon *et al.*, *Cancer Gene Ther.* **16**, 489–497 (2009).
19. G. L. Beatty *et al.*, *Cancer Immunol. Res.* **2**, 112–120 (2014).
20. J. Lee, C. M. Dollins, D. Boczkowski, B. A. Sullenger, S. Nair, *Immunology* **125**, 229–240 (2008).
21. L. Warren *et al.*, *Cell Stem Cell* **7**, 618–630 (2010).
22. Y. Niu *et al.*, *Cell* **156**, 836–843 (2014).
23. G. McClorey, M. J. Wood, *Curr. Opin. Pharmacol.* **24**, 52–58 (2015).
24. A. Fire *et al.*, *Nature* **391**, 806–811 (1998).
25. S. M. Elbashir *et al.*, *Nature* **411**, 494–498 (2001).
26. M. L. Bobbin, J. J. Rossi, *Annu. Rev. Pharmacol. Toxicol.* **56**, 103–122 (2016).
27. J. Nemunaitis *et al.*, *Oncology* **87**, 21–29 (2014).
28. A. O. Garba, S. A. Mousa, *Ophthalmol. Eye Dis.* **2**, 75–83 (2010).
29. V. Vigneswara *et al.*, *Brain* **137**, 1656–1675 (2014).
30. A. M. Ramos *et al.*, *Expert Opin. Drug Discov.* **10**, 541–556 (2015).
31. O. B. Suhr *et al.*, *Orphanet J. Rare Dis.* **10**, 109 (2015).
32. M. Hanna, *Curr. Heart Fail. Rep.* **11**, 50–57 (2014).
33. B. A. Sullenger, H. F. Gallardo, G. E. Ungers, E. Gilboa, *Cell* **63**, 601–608 (1990).
34. M. R. Mehan *et al.*, *Adv. Exp. Med. Biol.* **735**, 283–300 (2013).
35. C. P. Rusconi *et al.*, *Nature* **419**, 90–94 (2002).
36. A. M. Lincoff *et al.*, *Lancet* **387**, 349–356 (2016).
37. N. J. Ganson *et al.*, *J. Allergy Clin. Immunol.* **137**, 1610–1613.e7 (2016).
38. N. Lan, R. P. Howrey, S. W. Lee, C. A. Smith, B. A. Sullenger, *Science* **280**, 1593–1596 (1998).
39. J. A. Doudna, E. Charpentier, *Science* **346**, 1258096 (2014).
40. Y. H. Kim *et al.*, *Theranostics* **6**, 357–368 (2016).

ACKNOWLEDGMENTS

We would like to thank R. Becker, T. Povsic, S. Nimjee, and D. Boczkowski for useful discussions. This work was supported in part by grants from the NIH (to B.A.S.) and the Department of Defense (to S.N. and B.A.S.). B.A.S. and Duke University have filed U.S. patent applications (numbers 7,300,922 and 7,312,325) that relate to anticoagulant aptamers and antidote oligonucleotides. S.N. is a co-inventor on the patent describing the use of DCs transfected with tumor antigen–encoding RNA that has been licensed by Argos Therapeutics (Durham, NC) through Duke University. S.N. has no financial interest in and is not compensated by Argos Therapeutics.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1417/suppl/DC1
Table S1

10.1126/science.aad8709

RESEARCH ARTICLE SUMMARY

IMMUNE REGULATION

T helper 1 immunity requires complement-driven NLRP3 inflammasome activity in CD4$^+$ T cells

Giuseppina Arbore,* Erin E. West,* Rosanne Spolski, Avril A. B. Robertson, Andreas Klos, Claudia Rheinheimer, Pavel Dutow, Trent M. Woodruff, Zu Xi Yu, Luke A. O'Neill, Rebecca C. Coll, Alan Sher, Warren J. Leonard, Jörg Köhl, Pete Monk, Matthew A. Cooper, Matthew Arno, Behdad Afzali, Helen J. Lachmann, Andrew P. Cope, Katrin D. Mayer-Barber, Claudia Kemper†

INTRODUCTION: The inflammasomes and the complement system are traditionally viewed as quintessential components of innate immunity required for the detection and elimination of pathogens. Assembly of the NLRP3 inflammasome in innate immune cells controls the maturation of interleukin (IL)–1β, a proinflammatory cytokine critical to host defense, whereas activation of the liver-derived complement key components C3 and C5 in serum leads to opsonization and removal of microbes and induction of the inflammatory reaction. Recent studies, however, have highlighted an unanticipated direct role for complement C3 also in human T cell immunity: The anaphylatoxin C3a receptor (C3aR) and the complement regulator CD46 (which binds C3b) are critical checkpoints in human T cell lineage commitment, and they control initiation and resolution of T helper 1 (T_H1) responses in an autocrine fashion via T cell–derived and intracellularly activated C3. We explored a novel functional cross-talk of complement with the NLRP3 inflammasome within CD4$^+$ T cells and determined how the cooperation between these two "classically" innate systems directly affects interferon-γ (IFN-γ) production by adaptive immune cells.

RATIONALE: Given the critical role of intracellular C3 activation in human T_H1 responses and the importance of C5 activation products in inflammation, we investigated whether human CD4$^+$ T cells also harbor an "intracellular C5 activation system" and by what means this system may contribute to effector responses by using C5aR1 and C5aR2 agonists and antagonists, T cells from patients with cryopyrin-associated periodic syndromes (CAPS), and mouse models of infection and autoimmunity.

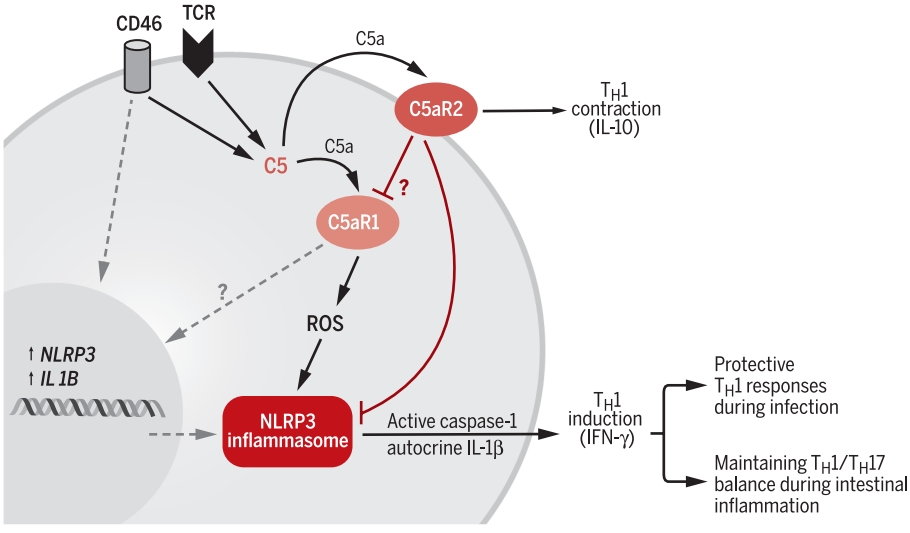
RESULTS: Human CD4$^+$ T cells expressed C5 and generated increased intracellular C5a upon T cell receptor activation and CD46 autocrine costimulation. Subsequent engagement of the intracellular C5aR1 by C5a induced the generation of reactive oxygen species (ROS) and the unexpected assembly of a functional NLRP3 inflammasome in CD4$^+$ T cells, whereas the surface-expressed C5aR2 negatively controlled this process.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aad1210

NLRP3 inflammasome–dependent autocrine IL-1β secretion and activity were required for optimal IFN-γ production by T cells; consequently, dysregulation of NLRP3 function in these cells affected their normal effector responses. For example, mutated, constitutively active NLRP3 in T cells from patients with CAPS induced hyperactive T_H1 responses that could be normalized with a NLRP3 inhibitor. The in vivo importance of a T cell–intrinsic NLRP3 inflammasome was further supported by the finding that IFN-γ production by *Nlrp3*$^{-/-}$ CD4$^+$ T cells was significantly reduced during viral infections in mice and that diminished T_H1 induction due to lack of NLRP3 function in a CD4$^+$ T cell transfer model of colitis led to uncontrolled T_H17 infiltration and/or expansion in the intestine and aggravated disease.

CONCLUSION: Our results demonstrate that the regulated cross-talk between intracellularly activated complement components (the "complosome") and the NLRP3 inflammasome is fundamental to human T_H1 induction and regulation. The finding that established innate immune pathways are also operative in adaptive immune cells and orchestrate immunological responses contributes to our understanding of immunobiology and immune system evolution. In addition, the results suggest that the complement-NLRP3 axis in T cells represents a novel therapeutic target for the modulation of T_H1 activity in autoimmunity and infection. ■



An intrinsic complement-NLRP3 axis regulates human T_H1 responses. T cell receptor activation and CD46 costimulation trigger *NLRP3* expression and intracellular C5a generation. Subsequent intracellular C5aR1 engagement induces ROS production (and possibly *IL1B* gene transcription) and NLRP3 assembly, which in turn mediates IL-1β maturation. Autocrine IL-1β promotes T_H1 induction (IFN-γ production) but restricts T_H1 contraction (IL-10 coexpression). C5aR2 cell surface activation by secreted C5a negatively controls these events via undefined mechanisms. Dysfunction of this system contributes to impaired T_H1 responses in infection or increased T_H17 responses during intestinal inflammation.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: claudia.kemper@kcl.ac.uk
Cite this article as G. Arbore *et al.*, *Science* 352, aad1210 (2016). DOI: 10.1126/science.aad1210

RESEARCH ARTICLE

IMMUNE REGULATION

T helper 1 immunity requires complement-driven NLRP3 inflammasome activity in CD4$^+$ T cells

Giuseppina Arbore,[1]* Erin E. West,[2]* Rosanne Spolski,[2] Avril A. B. Robertson,[3] Andreas Klos,[4] Claudia Rheinheimer,[4] Pavel Dutow,[4] Trent M. Woodruff,[3] Zu Xi Yu,[5] Luke A. O'Neill,[6] Rebecca C. Coll,[3] Alan Sher,[7] Warren J. Leonard,[2] Jörg Köhl,[8,9] Pete Monk,[10] Matthew A. Cooper,[3] Matthew Arno,[11] Behdad Afzali,[1,12] Helen J. Lachmann,[13] Andrew P. Cope,[14] Katrin D. Mayer-Barber,[15] Claudia Kemper[1,2]†

The NLRP3 inflammasome controls interleukin-1β maturation in antigen-presenting cells, but a direct role for NLRP3 in human adaptive immune cells has not been described. We found that the NLRP3 inflammasome assembles in human CD4$^+$ T cells and initiates caspase-1–dependent interleukin-1β secretion, thereby promoting interferon-γ production and T helper 1 (T_H1) differentiation in an autocrine fashion. NLRP3 assembly requires intracellular C5 activation and stimulation of C5a receptor 1 (C5aR1), which is negatively regulated by surface-expressed C5aR2. Aberrant NLRP3 activity in T cells affects inflammatory responses in human autoinflammatory disease and in mouse models of inflammation and infection. Our results demonstrate that NLRP3 inflammasome activity is not confined to "innate immune cells" but is an integral component of normal adaptive T_H1 responses.

The complement system is an ancient innate immune sensor system that is essential for elimination of pathogens by the host. Processing in serum of liver-derived C3 into C3a and C3b and of C5 into C5a and C5b activation fragments leads to opsonization and removal of invading microbes, mobilization of innate immune cells, and induction of inflammatory reactions (*1*). However, complement also profoundly regulates adaptive immunity: In addition to T cell receptor (TCR) activation, costimulation, and the presence of interleukin (IL)–12 (*2*), human CD4$^+$ T cells also depend on the activation of T cell–expressed complement receptors binding C3 activation fragments for normal T helper 1 (T_H1) induction (*3*). Unexpectedly, the engagement of complement receptors on T cells is independent of systemic complement but instead is mediated in an autocrine manner by complement activation fragments produced by the T cell itself. In particular, C3a and C3b are generated intracellularly via cathepsin L–mediated cleavage of C3 in T cells upon TCR activation (*4*). These engage their respective receptors—a G protein–coupled receptor (GPCR) C3a receptor (C3aR) and the complement regulator CD46 (which binds C3b)—and induce autocrine interferon-γ (IFN-γ) (*5*, *6*). Mechanistically, C3aR- and CD46-mediated signals (i) regulate IL-2R assembly, (ii) up-regulate the glucose transporter GLUT1 and the amino acid transporter LAT1, and (iii) up-regulate mTORC1 activation, which is required for the metabolic programming essential for IFN-γ induction (*7*).

However, CD46 costimulation is not only essential for IFN-γ production and human T_H1 induction; it also contributes to the negative control of T_H1 responses. Together with IL-2, CD46-mediated signals drive the coexpression of immunosuppressive IL-10 in T_H1 cells and initiate their switch into a (self-)regulatory and contracting phase (*3*). Accordingly, C3- and CD46-deficient patients suffer from recurrent infections and have severely reduced T_H1 responses in vitro and in vivo, whereas T_H2 responses remain intact (*5*, *8*). Conversely, uncontrolled intracellular C3 activation (or dysregulated CD46 engagement) in T cells contributes to hyperactive T_H1 responses observed in autoimmunity (*3*, *4*, *9*) that can be normalized pharmacologically by targeting intracellular cathepsin L function (*4*). Of note, CD46 is not expressed on somatic tissue in rodents and a functional homolog has not yet been identified. This indicates the existence of substantial differences in the complement receptor–driven pathways regulating T cell responses between species [reviewed in (*6*)].

Given the critical role of intracellular C3 processing in human T_H1 induction and contraction and the importance of C5a generation in inflammation, we investigated whether human CD4$^+$ T cells also harbor an "intracellular C5 activation" system contributing to effector responses.

Autocrine activation of C5a receptors regulates IFN-γ production by human CD4$^+$ T cells

Human CD4$^+$ T lymphocytes isolated from healthy donors contained intracellular stores of C5 and produced low levels of C5a in the resting state. TCR activation, in particular TCR + CD46 costimulation, increased the amounts of intracellular C5a, and this was associated with the secretion of C5a to the cell surface (Fig. 1, A and B). C5a, as well as the C5a "des-arginized" form of C5a (C5adesArg) generated by carboxypeptidase processing, can bind two distinct GPCR receptors, C5aR1 (CD88) and C5aR2 (GPR77, C5L2) (*10*, *11*). Binding of C5a to C5aR1 preferentially mediates proinflammatory responses. The function of C5aR2 varies with cell type; C5aR2 can act either as a nonsignaling decoy receptor antagonizing C5aR1 or as an active transducer of pro- or anti-inflammatory signals (*11*–*14*).

Both extra- and intracellular localization of C5aR1 and C5aR2 on human monocytes have been reported (*14*, *15*), but expression patterns in human CD4$^+$ T cells have not been described in detail. We detected expression of both *C5AR1* and *C5AR2* mRNA in human CD4$^+$ T cells (Fig. 1C) and protein by immunoblotting (fig. S1A), confocal microscopy (Fig. 1D), and flow cytometry (Fig. 1, E and F). Although mRNA amounts for *C5aR1* and *C5aR2* vary in T cells (Fig. 1C) (*16*), the protein levels for these receptors are comparable among donors (Fig. 1E). In resting and activated CD4$^+$ T cells, C5aR1 is expressed exclusively intracellularly and in low amounts, whereas the C5aR2 receptor is abundantly present inside and to a lesser degree on the cell surface (Fig. 1F). Using human embryonic kidney (HEK) 293 cells that had been stably transfected to express C5aR1, C5aR2, or no receptor, we corroborated the specificity of reagents used for C5a receptor detection (fig. S1, B and C). Competitive binding studies of C5a labeled with radioactive ^{125}I (Fig. 1G and fig. S1D) confirmed the ability of resting and activated human CD4$^+$ T cells to bind C5a.

To determine whether autocrine engagement of the C5a receptors on T cells regulates T_H1 induction, we activated human CD4$^+$ T cells with

[1]MRC Centre for Transplantation, Division of Transplant Immunology and Mucosal Biology, King's College London, London SE1 9RT, UK. [2]Laboratory of Molecular Immunology and Immunology Center, National Heart, Lung, and Blood Institute, Bethesda, MD 20892, USA. [3]Institute for Molecular Bioscience and School of Biomedical Sciences, University of Queensland, QLD 4072, Australia. [4]Institute for Medical Microbiology and Hospital Epidemiology, Medizinische Hochschule Hannover, 30625 Hannover, Germany. [5]Pathology Core, National Heart, Lung, and Blood Institute, Bethesda, MD 20892, USA. [6]School of Biochemistry and Immunology, Trinity College Dublin, Dublin, Ireland. [7]Laboratory of Parasitic Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA. [8]Institute for Systemic Inflammation Research, University of Lübeck, Lübeck, Germany. [9]Division of Immunobiology, Cincinnati Children's Hospital Medical Center and University of Cincinnati College of Medicine, Cincinnati, OH, USA. [10]Department of Infection and Immunity, University of Sheffield, Sheffield S10 2RX, UK. [11]Genomics Centre, Faculty of Life Sciences and Medicine, King's College London, London SE1 9NH, UK. [12]Lymphocyte Cell Biology Section, Molecular Immunology and Inflammation Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, MD 20892, USA. [13]UK National Amyloidosis Centre, Division of Medicine, University College London, Royal Free Campus, London NW3 2PF, UK. [14]Academic Department of Rheumatology, Division of Immunology, Infection and Inflammatory Diseases, King's College London, London SE1 1UL, UK. [15]Laboratory of Clinical Infectious Diseases, Inflammation and Innate Immunity Unit, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA.
*These authors contributed equally to this work. †Corresponding author. Email: claudia.kemper@kcl.ac.uk

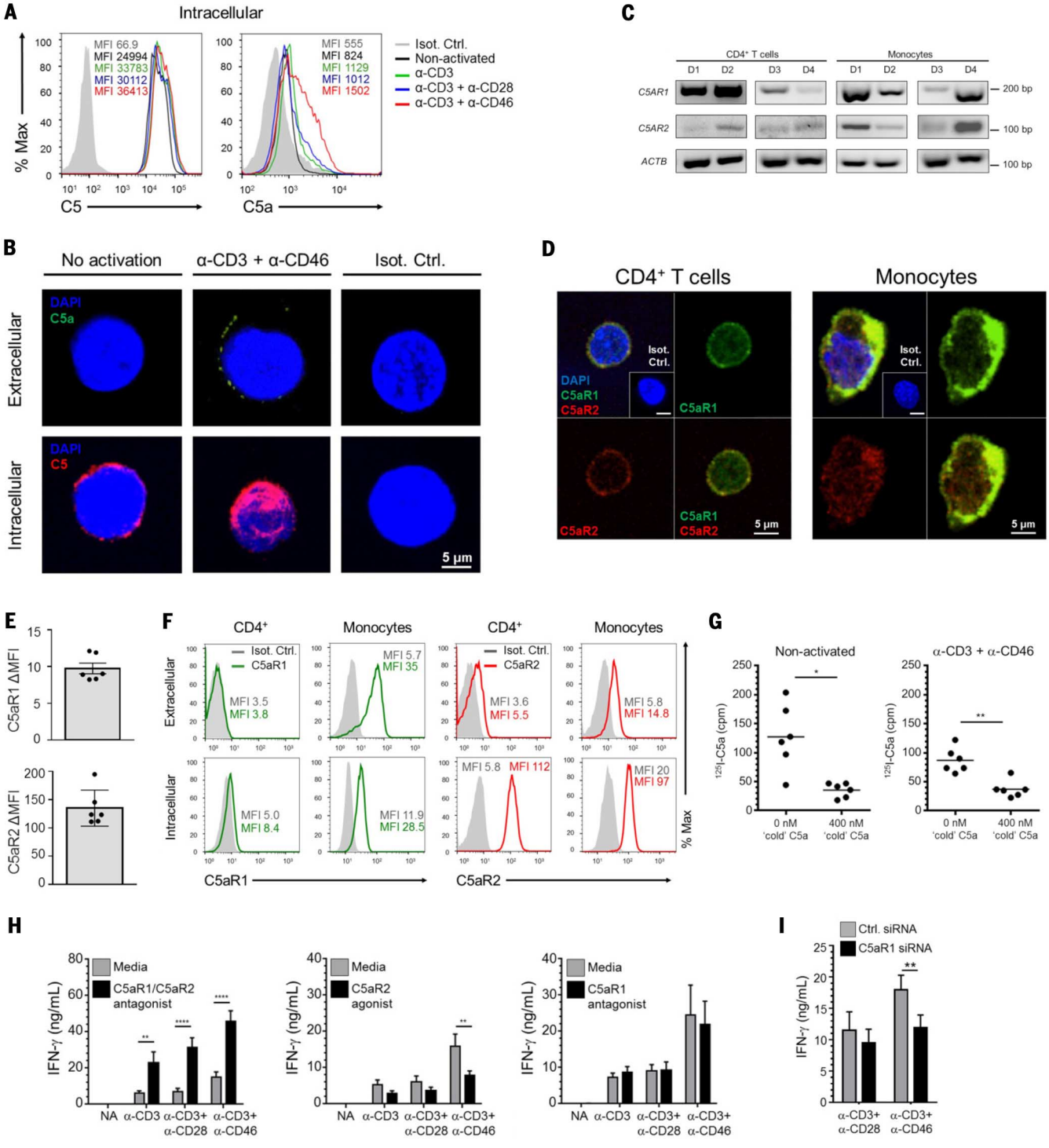


Fig. 1. Autocrine activation of C5a receptors regulates IFN-γ production by human CD4⁺ T cells. (**A** and **B**) Intracellular C5 and C5a generation in CD4⁺ T lymphocytes, left nonactivated or activated (36 hours) with anti-CD3 (α-CD3), α-CD3 + α-CD28, or α-CD3 + α-CD46 by flow cytometry (A) and confocal microscopy (B) (data representative of $n = 3$). (**C**) RT-PCR analysis for *C5AR1* and *C5AR2* mRNA in resting human CD4⁺ cells and monocytes ($n = 4$, donors D1 to D4, endogenous control *ACTB*). (**D**) Intracellular immunofluorescence on resting T cells and monocytes with antibodies to C5aR1 (green) and C5aR2 (red) (data are representative of $n = 3$). (**E**) C5aR1 and C5aR2 protein amounts in T cells with expression normalized to respective isotype control staining for each donor (change in mean fluorescence intensity ΔMFI ± SEM, $n = 6$). (**F**) Flow cytometry for C5aR1 and C5aR2 on resting T cells and monocytes, with representative histogram plots shown ($n = 6$). (**G**) Binding of radioactively labeled ^{125}I-C5a in absence or presence of nonlabeled "cold" C5a as competitor to resting or α-CD3 + α-CD46 activated (4 hours) T cells ($n = 6$). (**H**) IFN-γ secretion in nonactivated (NA) and activated (36 hours) CD4⁺ T cells in the absence or presence of a C5aR1/C5aR2 double receptor antagonist ($n = 9$), a C5aR2 agonist ($n = 8$), or a C5aR1 antagonist ($n = 7$). (**I**) IFN-γ production by T cells transfected with C5aR1-specific siRNA or a scrambled control siRNA (Ctrl. siRNA) 36 hours after activation ($n = 7$). Data are means ± SEM. *$P < 0.05$, **$P < 0.01$, ****$P < 0.0001$. (G), paired t test; (H) and (I), two-way ANOVA with Bonferroni multiple comparison test.

immobilized antibodies to CD3, CD3 and CD28, or CD3 and CD46 in the presence or absence of (i) a specific antagonist to C5aR1 [PMX53 (*17*)]; (ii) the C5aR1/C5aR2 receptor double antagonist $A8^{\Delta 71\text{-}73}$ [dRA (*18*)], targeting only C5aR2 (as the C5aR1 is expressed intracellularly); or (iii) a specific C5aR2 agonist (*19*). All reagents were cell-impermeable. Blocking C5aR2 activity significantly increased T_H1 induction (Fig. 1H, left), and activating C5aR2 with the agonist or with C5a or C5adesArg reduced T_H1 responses (Fig. 1H, middle, and fig. S1E). Blockade of C5aR2 also led to increased T_H17 (IL-17) but not T_H2 (IL-4) responses (fig. S1F) without altering cell viability (fig. S1G). Consistent with the solely intracellular localization of C5aR1, the C5aR1-specific antagonist had no effect on IFN-γ production because it could not "reach" and block intracellular C5aR1 (Fig. 1H, right). However, reduction of intracellular C5aR1 by small interfering RNA (siRNA) gene targeting led to a commensurate decrease in IFN-γ production (Fig. 1I and fig. S1H). Together, these data show that intracellular C5 activation contributes to induction of IFN-γ production in $CD4^+$ T cells via intracellular C5aR1 engagement, and that the surface-expressed C5aR2 exerts negative control of IFN-γ, possibly via suppression of intracellular C5aR1 signals.

Canonical NLRP3 inflammasome activation in $CD4^+$ T cells enhances IFN-γ production

To delineate the autocrine C5-driven pathways contributing to regulation of IFN-γ in $CD4^+$ T cells, we performed a transcriptome analysis using T cells from three healthy donors activated, or not, with anti-CD3 and anti-CD46 in the presence or absence of the C5aR1/C5aR2 antagonist. Surprisingly, we observed enrichment of transcripts associated with inflammasome activation, including *NLRP3* and *IL1B* (Fig. 2, A and B, and table S1), in cells activated with anti-CD3 and anti-CD46. Inhibition of C5aR2 during these activation conditions further increased some of these transcripts, notably *IL1A* and *IL1B* (fig. S2A and table S2); this finding offers further support for the idea that blocking C5aR2 leads to unrestrained or increased engagement of intracellular C5aR1 driven by the anti-CD3– and anti-CD46–induced increase in intracellular C5a generation.

IL-1α and IL-1β are prototypical proinflammatory cytokines involved in innate immune responses and contributing to the development of several pathogenic autoimmune diseases, including

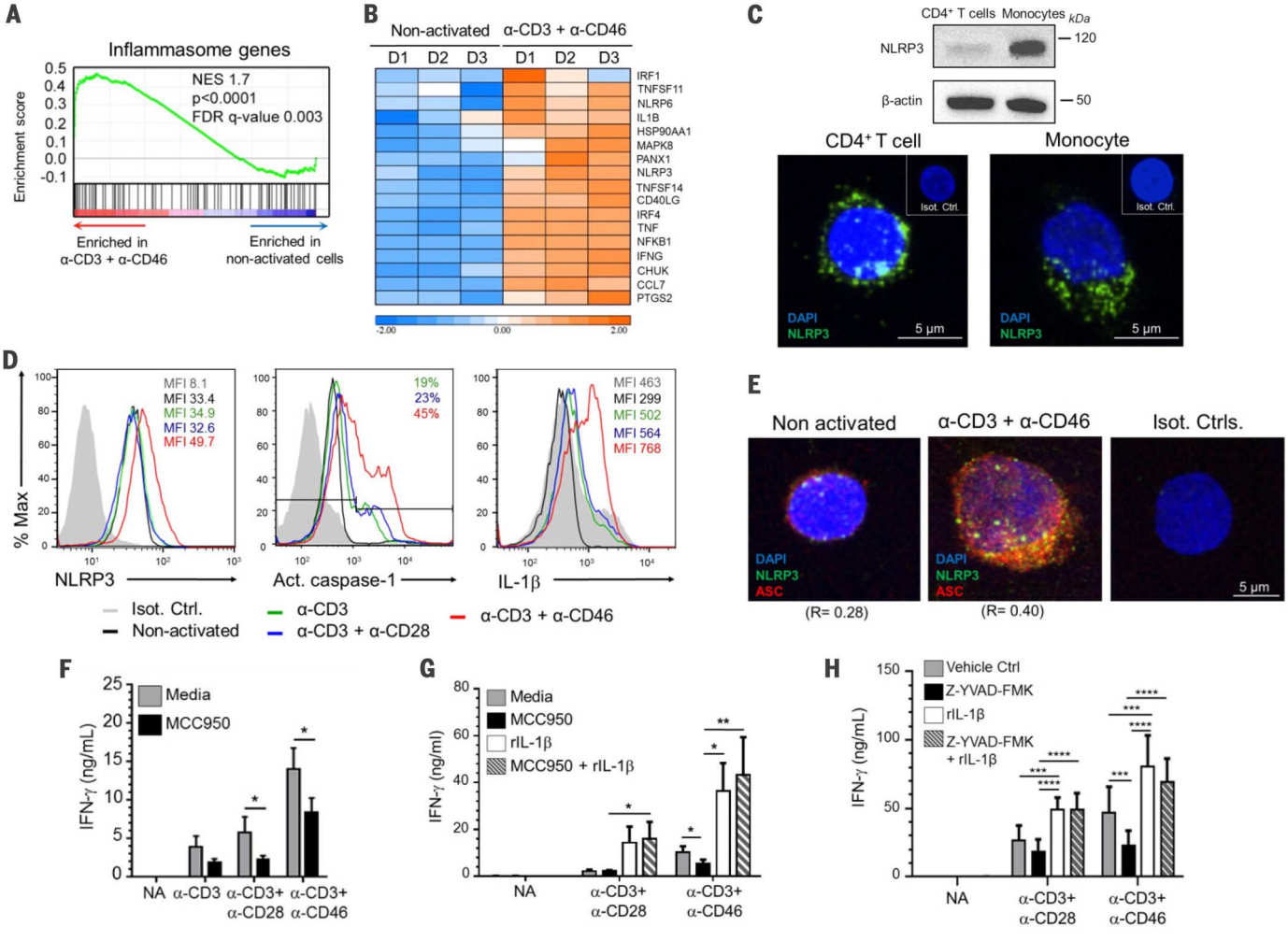


Fig. 2. NLRP3 inflammasome activation occurs in $CD4^+$ T cells and enhances IFN-γ production. (**A**) Gene set enrichment analysis (GSEA) for inflammasome-related genes in $CD4^+$ T cells after α-CD3 + α-CD46 activation (2 hours) compared to resting cells (donors D1 to D3). (**B**) Heat map depicting leading edge analysis (the core enriched genes) of the data in (A). (**C**) NLRP3 immunoblot (upper panel) and immunofluorescence (lower panel) on $CD4^+$ lymphocytes and monocytes (data representative of $n = 3$). (**D**) NLRP3, activated caspase-1 and total IL-1β protein expression in activated $CD4^+$ cells (data representative of $n = 3$). (**E**) Representative immunofluorescence costaining for NLRP3 (green) and ASC (red) on resting and α-CD3 + α-CD46 activated T cells (r = Pearson correlation coefficient between NLRP3 and ASC fluorescence, $n = 3$). (**F** and **G**) IFN-γ production by resting (NA) and activated $CD4^+$ T cells with or without MCC950 addition ($n = 7$) (F) and with or without rhIL-1β supplementation ($n = 3$) (G). (**H**) IFN-γ production in presence of the specific caspase-1 inhibitor Z-YVAD-FMK with or without rhIL-1β addition. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. (F) to (H), two-way ANOVA with Bonferroni multiple comparison test.

type 1 diabetes and arthritis (20–22). Both IL-1α and IL-1β bind to IL-1 receptor 1 (IL-1R1). Antigen-presenting cell (APC)–derived IL-1β supports T cell priming and imprinting of T helper effector function (23), including enhancement of IFN-γ and IL-17 production from $CD4^+$ T cells (24–26). Further, mice with deletion of the IL-1β signal transducer MyD88 in T lymphocytes cannot generate memory T cells (27). Pro–IL-1β is synthesized as a 31-kDa precursor and converted to mature 17-kDa IL-1β via caspase-1 cleavage (28). Caspase-1 is regulated by proteolytic activation during oligomerization with NLRP3 and the adaptor ASC (apoptosis-associated speck-like protein containing a caspase recruitment domain), which is triggered in response to danger signals (29, 30). NLRP3 inflammasome function requires a priming signal 1 (which induces *NLRP3* and *IL1B* gene transcription) and a signal 2 that induces functional inflammasome assembly (30) and has been described in myeloid innate immune cells, with monocytes as the main source of IL-1β (25, 31), and in several nonimmune cell types (such as microglia, endothelial cells, and retinal pigment epithelial cells) (32–34). However, canonical NLRP3 inflammasome activity has not been demonstrated in lymphoid adaptive immune cells.

We confirmed the presence of an "NLRP3 signature" in T cells by demonstrating *NLRP3* and *IL1B* gene (fig. S2B) and protein expression, as well as generation of activated caspase-1 and mature IL-1β, in activated human $CD4^+$ T cells (Fig. 2, C and D, and fig. S2, C to F). Consistent with our gene array data, anti-CD3 and anti-CD46 activation led to robust NLRP3 activation and IL-1β generation (Fig. 2D) and increased colocalization of NLRP3 and ASC (Fig. 2E). Notably, both resting naïve and memory $CD4^+$ T cells expressed NLRP3 protein (fig. S2, C and D).

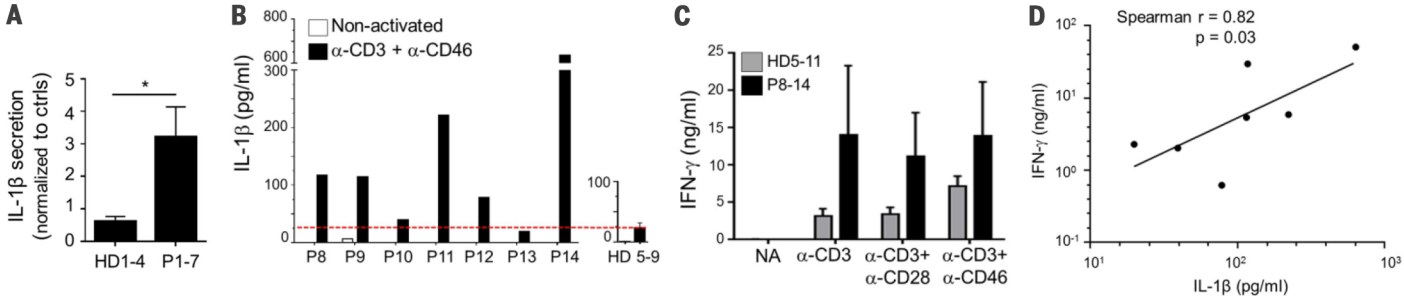
Because IL-1β supports T_H1 induction (35) and is most strongly induced by the T_H1 driver CD46, we next assessed whether inhibition of NLRP3 activity in $CD4^+$ T cells perturbs IFN-γ production. To this end, $CD4^+$ T cells were activated in the presence of MCC950, a specific NLRP3 inhibitor (36), and T_H1, T_H2, and T_H17 cytokine production was measured 36 hours after activation. NLRP3 inhibition during T cell activation specifically attenuated IFN-γ (Fig. 2F), whereas differences in IL-4 and IL-17 production did not reach significance (fig. S2G) and cell viability was unaffected (fig. S2H). The effects of the NLRP3 inhibitor could be fully reversed by the addition of recombinant human IL-1β (rhIL-1β) to cultures (Fig. 2G). Similarly, reduction of active caspase-1 activity by the specific inhibitor Z-YV AD-FMK repressed IL-1β and IFN-γ secretion (Fig. 2H and fig. S2I), and rhIL-1β provision normalized T_H1 induction in these cultures. The role for IL-1β as critical autocrine "T_H1 supporter" is reinforced by our observation that no IL-18 [which also depends on NLRP3 activation and can support T_H1 responses (37)] was measurable in our cultures and that addition of IL-18 binding protein had no effect on cytokine production (fig. S2J).

The hyperactive in vitro T_H1 response in CAPS patients is normalized by NLRP3 inhibition

To further explore this pathway, we measured the effects of NLRP3 hyperactivity in $CD4^+$ T cells isolated from the blood of patients with distinct gain-of-function mutations in NLRP3 (patient characteristics are summarized in table S3). This class of NLRP3 mutations is associated with a group of heritable monogenic syndromes known as cryopyrin-associated periodic syndromes (CAPS), characterized by excessive production of IL-1β from APCs with recurrent fevers, skin rashes, joint and ocular inflammation, and amyloidosis (38). Therapeutic suppression of the inflammatory responses can be achieved by IL-1R blockade with anakinra, an IL-1R antagonist, or canakinumab, a monoclonal antibody (mAb) targeting IL-1β (38, 39). Despite their medication regimen and the fact that cytokine production by immune cells from CAPS patients can vary with their respective "flare status" (40), T cells from a first cohort of CAPS patients that we assessed had significantly increased IL-1β secretion relative to sex- and age-matched healthy donors (Fig. 3A), indicating that increased NLRP3 activity in $CD4^+$ T cells indeed induces heightened IL-1β secretion.

We next performed a more in-depth analysis of T cell in vitro responses from another cohort of seven CAPS patients (table S4). All patients had a naïve versus memory T cell distribution comparable to those of healthy donors (fig. S3A), and T cells from five patients of this second cohort also showed significantly increased IL-1β secretion upon activation (Fig. 3B). Furthermore, $CD4^+$ T cells from these patients trended toward substantially increased IFN-γ relative to T cells from sex- and age-matched healthy donors, and we observed a statistically significant correlation between increased IL-1β and IFN-γ secretion (Fig. 3, C and D).

T cells from CAPS patients displayed significantly reduced in vitro IL-17 responses (Fig. 3E).



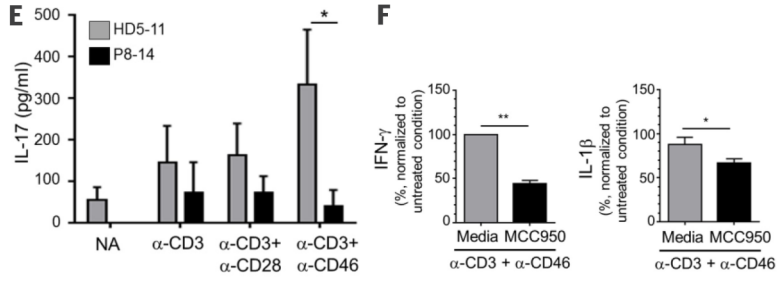


Fig. 3. T cells from CAPS patients have increased NLRP3 inflammasome activity and hyperactive T_H1 responses. (**A**) IL-1β production from $CD4^+$ T cells activated with α-CD3 + α-CD46 for 36 hours from four healthy donors (HD1 to HD4) and seven patients with CAPS (cohort 1, P1 to P7). (**B**) IL-1β secretion from resting and α-CD3 + α-CD46 activated $CD4^+$ cells from seven patients with CAPS (P8 to P14, individual values) and five healthy sex- and age-matched donors (HD5 to HD9, combined values). (**C**) IFN-γ secretion from resting and activated $CD4^+$ cells from seven patients with CAPS (P8 to P14) and seven healthy sex- and age-matched donors (HD5 to HD11). (**D**) Correlation between IL-1β and IFN-γ production in T cells from patients P8 to P14 upon α-CD3 + α-CD46 activation (Spearman correlation analysis). (**E**) IL-17 production by resting and activated T cells from CAPS patients P8 to P14 and healthy donors H5 to H11. (**F**) IFN-γ and IL-1β secretion by $CD4^+$ T cells from P8, P11, and P14 after α-CD3 + α-CD46 activation with or without MCC950 treatment (% normalized to nontreated). Analyses on (A) to (F) were performed at 36 hours after activation. Values correspond to two technical replicates for every patient and healthy control sample in each experiment. Data are means ± SEM. $^*P < 0.05$, $^{**}P < 0.01$. (A), unpaired *t* test; (C) and (E), two-way ANOVA with Bonferroni multiple comparison test; (D), Spearman correlation test; (F), paired *t* test.

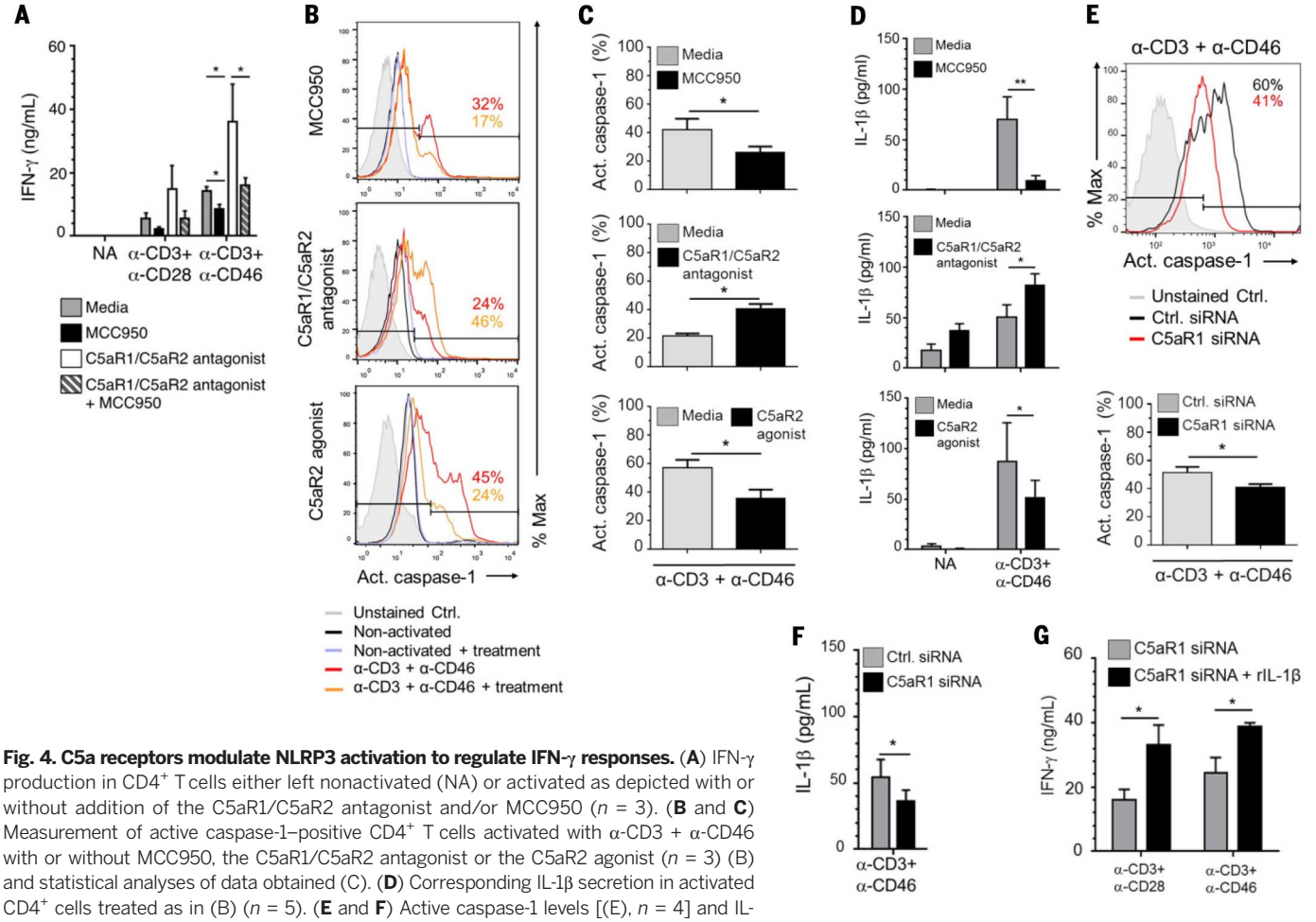


Fig. 4. C5a receptors modulate NLRP3 activation to regulate IFN-γ responses. (**A**) IFN-γ production in $CD4^+$ T cells either left nonactivated (NA) or activated as depicted with or without addition of the C5aR1/C5aR2 antagonist and/or MCC950 (n = 3). (**B** and **C**) Measurement of active caspase-1–positive $CD4^+$ T cells activated with α-CD3 + α-CD46 with or without MCC950, the C5aR1/C5aR2 antagonist or the C5aR2 agonist (n = 3) (B) and statistical analyses of data obtained (C). (**D**) Corresponding IL-1β secretion in activated $CD4^+$ cells treated as in (B) (n = 5). (**E** and **F**) Active caspase-1 levels [(E), n = 4] and IL-1β secretion [(F), n = 7] in T cells after transfection with either C5aR1-specific siRNA or scrambled control (Ctrl.) siRNA. (**G**) IFN-γ production in activated $CD4^+$ T cells after transfection with C5aR1-specific siRNA or a scrambled control siRNA (Ctrl. siRNA) with or without addition of rhIL-1β (n = 3). Analyses were performed at 36 hours after activation. Data are means ± SEM. *P < 0.05, **P < 0.01. (A), (D), and (G), two-way ANOVA with Bonferroni multiple comparison test; (C), (E), and (F), paired t test.

Although caspase-1 activity was not significantly increased in the patients' T cells at the time point assessed (36 hours), the patients with highest IL-1β secretion also had the highest active caspase-1 levels (fig. S3B). Activation of $CD4^+$ T cells from CAPS patients in the presence of the NLRP3 inhibitor MCC950 led to a reduction of both IL-1β and IFN-γ secretion (Fig. 3F). Together, these data demonstrate that human $CD4^+$ T cells produce IL-1β in an NLRP3-dependent manner, that autocrine IL-1β generation supports IFN-γ secretion, and that dysregulation of this pathway occurs in human autoinflammatory disease.

C5a receptors modulate NLRP3 inflammasome activity to regulate IFN-γ production

We next asked whether C5aR signaling could directly regulate NLRP3 activity in human $CD4^+$ T cells. C5aR2 blockade in CD3 + CD46–activated T cells further increased *IL1B* but not *NLRP3* mRNA (fig. S4, A and B). Enhanced IFN-γ secretion driven by C5aR2 blockage could be reversed by inhibition of NLRP3 with MCC950 (Fig. 4A) without affecting IL-17 or IL-4 production (fig. S4C). Pharmacological targeting of C5aR2 via either dRA (blockage of C5aR2 signaling) or a C5aR2 agonist (activation of C5aR2 signaling) revealed that C5aR2 negatively regulates active caspase-1 and mature IL-1β expression in T cells (Fig. 4, B to D) but does not affect NLRP3 protein levels per se (fig. S4D). Silencing of *C5AR1* expression had also no effect on NLRP3 protein levels (fig. S4E) but reduced active caspase-1 (Fig. 4E) and IL-1β expression (Fig. 4F). Moreover, the reduction of IFN-γ secretion after *C5AR1* gene silencing was "rescued" by addition of rhIL-1β (Fig. 4G). Together, these data suggest that CD46-mediated signals increase *NLRP3* mRNA expression in T cells, whereas C5aR1 supports subsequent NLRP3 assembly and C5aR2 is a negative regulator of this process.

Reactive oxygen species (ROS) are "classical" upstream stimulators (signal 2) of NLRP3 assembly (*41*) and are strongly induced by C5aR1 in monocytes and neutrophils (*42*). Furthermore, generation of ROS within $CD4^+$ T cells is required for T cell activation and induction of IL-2, a key cytokine for T_H1 biology (*43*). We therefore assessed whether autocrine C5aR1 engagement by intracellular C5a generation in T cells induces NLRP3 inflammasome assembly via ROS generation. We observed potent generation of ROS in anti-CD3– and anti-CD46–induced T_H1 cells (Fig. 5A) and poor T_H1 induction in the presence of a ROS inhibitor (Fig. 5B). Further, reduction of C5aR1 protein expression by gene silencing decreased ROS production (Fig. 5C, left), whereas inhibition of C5aR2 surface activation significantly increased ROS generation in T cells (Fig. 4C, right).

Enhanced IFN-γ production by T cells induced by C5aR2 blockade could be entirely reversed by the presence of a ROS inhibitor (Fig. 5B and fig. S5). This finding suggests that NLRP3 activation in human T cells involves intracellular C5-driven ROS production.

NLRP3 activity in $CD4^+$ T cells is required for optimal IFN-γ responses during viral infection

To address the biological importance of NLRP3-driven autocrine IL-1β production by $CD4^+$ T cells, we analyzed $CD4^+$ T cell responses of *Nlrp3*$^{-/-}$,

Il1a$^{-/-}$/*Il1b*$^{-/-}$, and *Il1r1*$^{-/-}$ mice initially in vitro and subsequently in an established in vivo viral infection model. Similar to human CD4$^+$ lymphocytes, CD4$^+$ cells from wild-type mice expressed NLRP3 and IL-1β; neither NLRP3 nor IL-1β mRNA (Fig. 6A) and protein (fig. S6A) were detectable in T cells from respective gene-deficient animals. We observed no difference in the proportion of naïve versus memory T cells or in T cell survival between wild-type and knockout strains (fig. S6, B and C). However, upon in vitro CD3 + CD28 activation, CD4$^+$ T cells from *Nlrp3*$^{-/-}$, *Il1a*$^{-/-}$/*Il1b*$^{-/-}$, and *Il1r1*$^{-/-}$ mice had a reduction of ~75% in IFN-γ production when compared to T cells from wild-type animals (Fig. 6B); in contrast, IL-10, IL-4, and IL-17 production were unaffected in all three mouse mutant lines (fig. S6D). Further, although activation of T cells from wild-type mice in the presence of the NLRP3 inhibitor MCC950 had no effect on cell viability (fig. S6E), only IFN-γ production was significantly reduced (Fig. 6C and fig. S6F). These results indicate that diminished IFN-γ secretion in the "knockout T cells" was not due to a developmental defect, whereas NLRP3 activity is required for normal IFN-γ induction. Moreover, both naïve and memory mouse CD4$^+$ T cells displayed a requirement for NLRP3-driven IL-1β activity for optimal IFN-γ secretion (fig. S6, G and H).

Using a lymphocytic choriomeningitis virus (LCMV) model (Fig. 6D), we next demonstrated an in vivo role for NLRP3-driven IL-1β generation in T_H1 responses during infection. Irradiated mice were reconstituted with equal parts bone marrow cells isolated from wild-type mice mixed with bone marrow cells from *Nlrp3*$^{-/-}$, *Il1a*$^{-/-}$/*Il1b*$^{-/-}$, or *Il1r1*$^{-/-}$ mice before infection with LCMV. Analysis of splenic CD4$^+$ T cells 12 days after infection revealed comparable numbers of GP66-77$^+$Ki67$^+$ LCMV tetramer$^+$ cells generated by all animals (Fig. 6, E and F), indicating that *Nlrp3*$^{-/-}$, *Il1a*$^{-/-}$/*Il1b*$^{-/-}$, and *Il1r1*$^{-/-}$ CD4$^+$ T cells survived normally. However, T cells deficient in any of these components displayed substantially reduced ability to generate IFN-γ$^+$ virus-specific cells in vivo (with an average decrease of ~50%) (Fig. 6, G and H). Together, these data demonstrate that autocrine canonical NLRP3 inflammasome activity is required for optimal protective IFN-γ production by CD4$^+$ T cells during viral infection.

Autocrine NLRP3 activity in T cells controls the T_H1-T_H17 balance during intestinal inflammation

To further substantiate the in vivo importance of NLRP3 inflammasome activity in CD4$^+$ T cells, we also measured the effects of NLRP3 deficiency in an autoimmune disease setting by assessing its influence on disease outcome in a CD4$^+$ T cell transfer model of colitis, where IL-1β and both T_H1 and T_H17 responses in the intestine have been shown to be involved (*44*, *45*). To this end, sorted CD4$^+$CD25$^-$CD45RB^{hi} T cells isolated from C57BL/6 wild-type or *Nlrp3*$^{-/-}$ mice were injected intraperitoneally (i.p.) into age- and sex-matched C57BL/10 *Rag2*$^{-/-}$ mice. Body weight and disease score were monitored and cytokine production by lamina propria CD4$^+$ T cells measured after animals displayed disease symptoms and were killed. Unexpectedly, relative to mice injected with wild-type CD4$^+$ T cells, mice that had received *Nlrp3*$^{-/-}$ CD4$^+$ T cells developed more severe disease with significantly increased weight loss, reduction in colon length, and higher disease scores (Fig. 7, A to C). Similar to our observation in the LCMV model, colonic *Nlrp3*$^{-/-}$ T cells displayed a substantial reduction in IFN-γ production (average decrease ~45%); however, we also observed a concurrent significant increase in T_H17 responses in these animals (Fig. 7, D and E). These observations were confirmed using a CD4$^+$ T cell–driven model of graft-versus-host disease (GvHD) where mice receiving *Nlrp3*$^{-/-}$ T cells also displayed more severe illness with reduced T_H1 but concurrently increased T_H17 induction (fig. S7, A to D). Together, these data demonstrate that the NLRP3 inflammasome mediates functionally important CD4$^+$ T cell intrinsic effects that not only are required for normal IFN-γ production but also control the T_H1-T_H17 balance during (at minimum) intestinal inflammation. These latter findings align with our observation that T cells from CAPS patients indeed have increased T_H1 but also decreased in vitro T_H17 responses (Fig. 3, C and E).

Discussion

Our results show that canonical NLRP3 inflammasome function is not confined to innate immune cells but is operative in adaptive CD4$^+$ T cells and, via autocrine IL-1β activity, is required for the optimal production of the key host defense factor IFN-γ. Further, and unexpectedly, NLRP3 assembly in human T cells requires TCR-induced intracellular C5 activation and stimulation of intracellular C5aR1, which initiates the generation of ROS and thereby provides a critical signal 2 for inflammasome assembly (*30*). Secretion of intracellularly generated C5a/C5ades-Arg engages the surface-expressed "alternative" C5aR2, which negatively controls NLRP3 activation either through inhibition of the C5aR1 or via a yet undefined mechanism. Given that APCs provide generally ample amounts of IL-1β during the cognate APC/T cell interaction, our observation that normal IFN-γ production requires also T cell autocrine IL-1β production is initially somewhat surprising. However, we envisage that, whereas APC-derived NLRP3-activated IL-1β supports initial T_H1 priming, proper "imprinting" or maintenance of the T_H1 phenotype during differentiation and migration into the periphery may rely on autocrine NLRP3 activity. IL-1β production by T cells, relative to myeloid cells, is comparatively low and, as we have shown, tightly regulated by an autocrine C5aR1 versus C5aR2 activation balance. The likely reason for this is that rapid control of local IL-1β is critical to normal termination of T_H1 responses: Human T_H1 cells co-induce IL-10 secretion in a CD46-dependent fashion during their contraction phase, and failure of this "IL-10 switch" underlies hyperactive T_H1 responses observed in rheumatoid arthritis and multiple sclerosis (*3*, *9*). IL-1β is a strong suppressor of IL-10 production (*23*) and, accordingly, we found that blockade of C5aR2 increased the IFN-γ/IL-10 ratio in CD4$^+$ T cells (fig. S8A), whereas

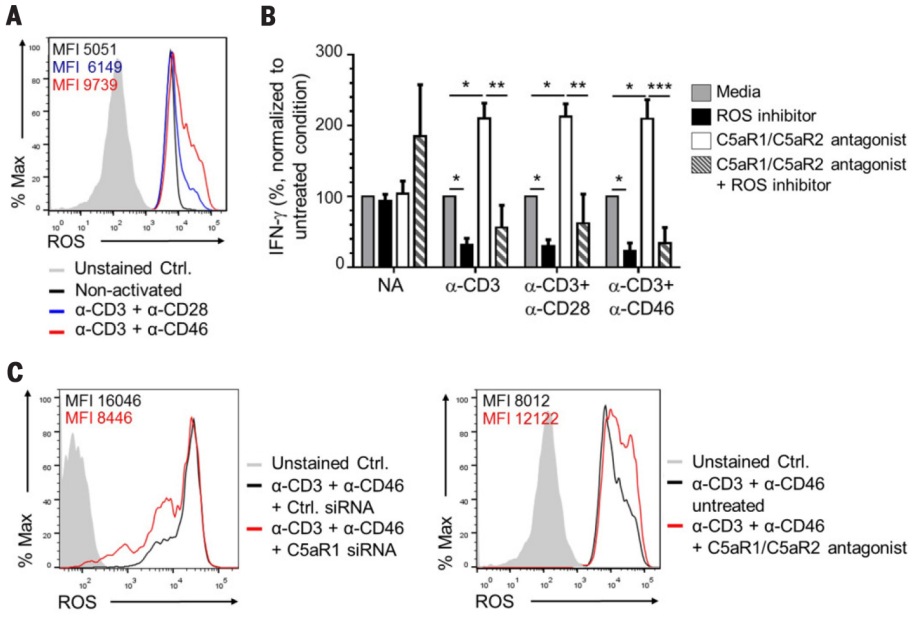


Fig. 5. Intracellular C5aR1 activation induces ROS generation in CD4$^+$ T cells. (**A**) ROS production in CD4$^+$ T cells activated under the depicted conditions (data shown are representative of $n = 3$). (**B**) IFN-γ production from CD4$^+$ T cells left nonactivated or activated as indicated with and without a specific ROS inhibitor and/or the C5aR1/C5aR2 antagonist ($n = 3$). Data are from a two-way ANOVA with Bonferroni multiple comparison test. (**C**) ROS production in α-CD3 + α-CD46 activated CD4$^+$ cells after transfection with C5aR1-specific siRNA (left panel) or with or without the C5aR1/C5aR2 double antagonist (right panel) (data shown are representative of $n = 3$). Analyses were performed 36 hours after activation. Data are means ± SEM. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

IL-1β addition to cultures increased IFN-γ (Fig. 2G) but blocked proportional IL-10 secretion (fig. S8B). Moreover, IFN-γ to IL-10 switching was significantly reduced in T cells from CAPS patients (fig. S8C).

Further supporting the notion that autocrine T cell cytokine production needs to be carefully controlled in the microenvironment is our observation that the reduction in IFN-γ secretion by T cells from *Nlrp3*$^{-/-}$ mice led to a concurrent increase in colonic lamina propria T_H17 cells and increased intestinal inflammation. These findings demonstrate that T_H1 cells negatively control the influx and/or expansion of T_H17 cells during colitis and that T_H17 induction at this location (and in this model) is independent of intrinsic T cell NLRP3 activity. These observations may also help to explain why some groups observed protection of *Nlrp3*$^{-/-}$ animals in models of inflammatory bowel disease while others observed aggravated disease (*46*, *47*); these earlier studies had not controlled for a T cell–intrinsic function of the NLRP3 inflammasome. Also, because IL-1β also boosts the production of other cytokines including IL-4 and IL-17 (*35*), the activation of the NLRP3 inflammasome in T cells and its functional outcome could be context-dependent. For example, Bruchard *et al.* recently observed a noncanonical function for NLRP3 in mouse CD4$^+$ T cells (they did not assesses human T cells) independent of inflammasome formation and IL-1β secretion, during T_H2 induction and tumor growth (*48*).

In addition, there are clear species-specific differences in the relative contributions of complement receptor activities to IL-1β and/or IFN-γ induction in CD4$^+$ T cells. Although we found that CD4$^+$ T cells from *C5ar2*$^{-/-}$ mice have increased in vitro IFN-γ production, which was reduced to normal levels by MCC950 treatment (fig. S9), and *C5ar1*$^{-/-}$ mice have impaired in vitro and in vivo T_H1 responses (*49*), the role and expression of anaphylatoxin receptors on mouse T cells remains a matter of controversy (*49*–*51*). But more important, mice lack expression of CD46 on all immune cells (*6*), whereas in humans, CD46 costimulation is required for IL-2R assembly and the metabolic reprogramming in CD4$^+$ T cells that drives IFN-γ secretion (*5*, *7*, *52*). Further, we show here that CD46 engagement, aside from amplifying intracellular C5a generation and ROS production (signal 2), also delivers an important signal 1 for NLRP3 inflammasome activation by mediating increased transcription of the *NLRP3* gene [this likely occurs via CD46-induced increased NF-κB nuclear translocation during T cell activation (*6*)]. Because a functional homolog for CD46 has not been identified in rodents, the exact upstream signals controlling NLRP3 inflammasome activation and IL-1β production in murine T cells remain to be defined.

In summary, the regulated cross-talk between intracellularly activated complement components ("complosome") and the NLRP3 inflammasome emerges as fundamental to human T_H1 induction and regulation. That established innate immune pathways previously not thought to be operative

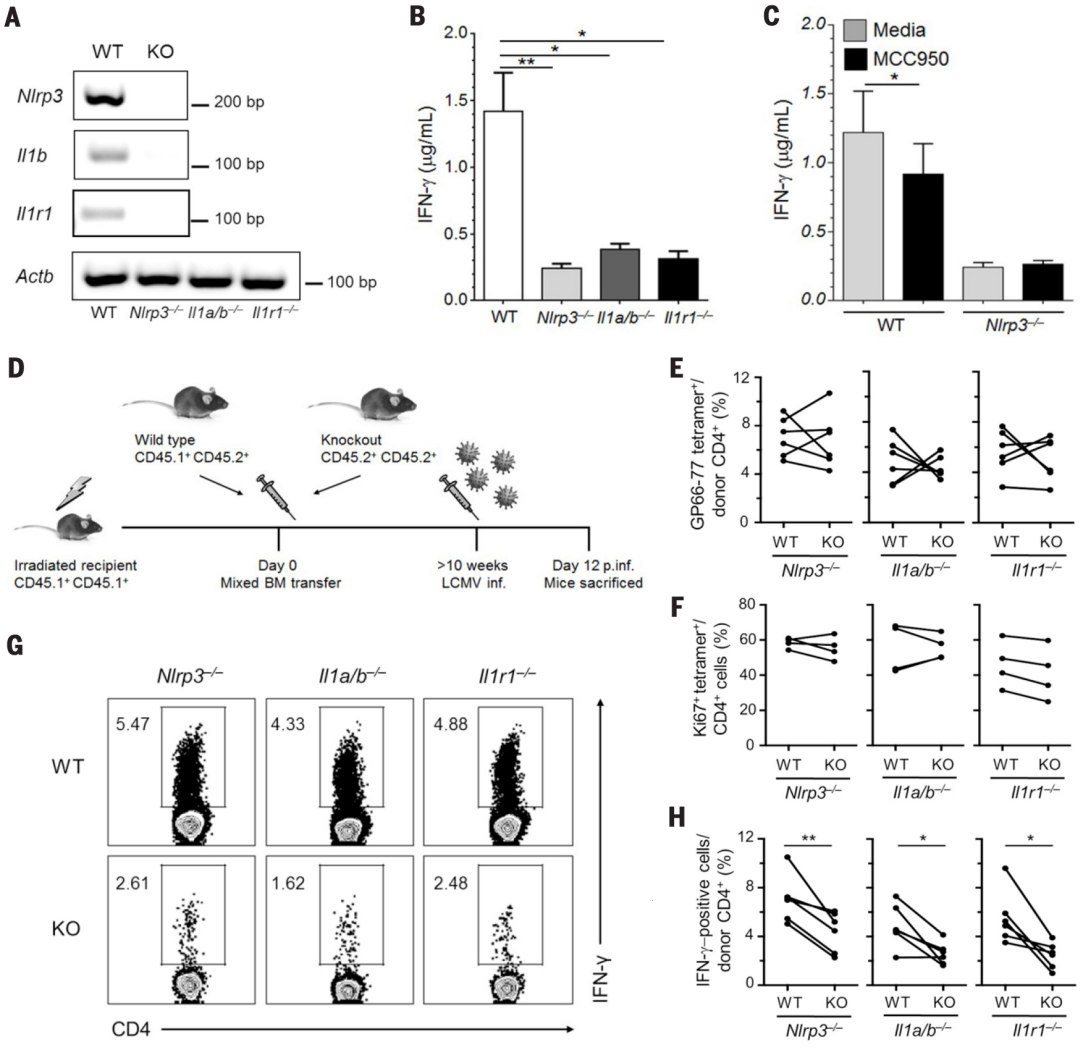


Fig. 6. NLRP3 function in CD4$^+$ T cells drives optimal IFN-γ production during viral infection. (**A**) RT-PCR analysis on CD4$^+$ T cells isolated from wild type (WT), *Nlrp3*$^{-}$, combined *Il1a*$^{-/-}$ and *Il1b*$^{-/-}$ (*Il1a/b*$^{-/-}$) and *Il1r1*$^{-/-}$ mice for corresponding gene mRNA expression. (**B**) Cytokine secretion from CD4$^+$ T cells isolated from wild-type and knockout mice at 96 hours after α-CD3 + α-CD28 activation (n = 3). (**C**) Cytokine production from CD4$^+$ T cells from wild-type and *Nlrp3*$^{-/-}$ mice after α-CD3 + α-CD28 activation (96 hours) with or without addition of MCC950 (n = 4). (**D**) Schematic of the acute lymphocytic choriomeningitis virus (LCMV) infection model used in this study. (**E** and **F**) Percentage of LCMV tetramer-positive CD4$^+$ T cells isolated from the spleens of the three bone marrow chimeric mice groups used 12 days after infection (E) and percentages of Ki67$^+$GP66-77$^+$/tetramer-positive cells (F). (**G** and **H**) Representative intracellular IFN-γ staining in splenic CD4$^+$ T cells of one mouse from each group after LCMV peptide restimulation (5 hours) [(G), n = 6] with corresponding statistical analyses [(H), n = 6]. Data are means ± SEM. *P < 0.05, **P < 0.01. (B), one-way ANOVA with Tukey multiple comparison test; (C), two-way ANOVA with Bonferroni multiple comparison test; (E) to (H), paired t test. Data in (G) are representative of two independent experiments; data in (E), (F), and (H) are pooled from two independent experiments.

in adaptive immune cells are not only present but also are key in directing immunological responses is of substantial importance to our understanding of immunobiology and immune system evolution. Further, the possibility that this normal functional cross-talk may also be target of viral immune evasion strategies (*53*) suggests that the complement-NLRP3 axis in T cells could represent a novel therapeutic target for the modulation of IFN-γ responses in autoimmunity and infection. In this regard, it will be valuable to explore whether optimal production of IFN-γ by $CD8^+$ T cells (*54*), natural killer T (NKT) cells, and/or innate lymphoid type 1 (ILC1) cells also relies on autocrine complement-NLRP3 inflammasome activity.

Materials and methods

Healthy donors and patients

Blood samples were obtained with ethical and institutional approvals (Wandsworth Research Ethics Committee, REC number 09/H0803/154). T cells were purified blood samples from healthy volunteers after informed consent. Fourteen adult patients with CAPS were recruited at the National Amyloidosis Centre, University College London (ethical approval REC reference number 06/Q0501/42) with key information on the patients summarized in tables S3 and S4. In all experiments that involved T cells from CAPS patients, T cells from age- and sex-matched healthy volunteers were used as controls.

Mice used in the study

All mice used in this study are on a C57BL/6 background (with the exception of the GvHD experiment, where Balb/c mice were used). Wild-type and *Il1r1*$^{-/-}$mice were purchased from Jackson Laboratories and subsequently backcrossed to B6 for 10 generations at NIH (Bar Harbor, ME). The *C5ar2*$^{-/-}$ (human gene symbol GPR77) mice were previously described (*12*), *Nlrp3*$^{-/-}$ animals were provided by V. Dixit of Genentech, and mice deficient in *Il1a* and *Il1b* (*Il1a/Il1b*$^{-/-}$ animals) were kindly provided by Y. Iwakura (Tokyo University) (*55*). The C57BL/10 $RAG2^{-/-}$ mice were obtained from Taconic. All animals were maintained in AALAC-accredited BSL2 or BSL3 facilities at the NIH or FDA and experiments performed in compliance with an animal study proposal approved by the NIAID or FDA Animal Care and Use Committee.

Cell isolation and activation

Human cells: $CD4^+$ T cells and monocytes were isolated from blood as previously published using the MACS Human $CD4^+$ Positive T cell Isolation Kit or the MACS Human $CD14^+$ Cell Positive Isolation Kit (both Miltenyi Biotech, Bisley, UK), respectively (*7*). Purity of bead-isolated T lymphocyte fractions was typically >98% and for monocytes >95%. For naïve and memory $CD4^+$ T cell sorting, cells were stained with appropriate antibodies (naïve cells, $CD4^+$, $CD45RA^+$, $CD45RO^-$, and $CD25^-$; memory cells, $CD4^+$, $CD45RA^-$, $CD45RO^+$, and $CD25^-$) and sorted with a BD FACSAria II Cell Sorter (KCL Flow Core facility). $CD4^+$ T cells were activated in 48-well culture plates (2.5×10^5 to 3.0×10^5 cells per well) coated with mAbs to CD3, CD28, or CD46 (2.0 μg/ml PBS each) and addition of rhIL-2 (25 U/ml), thus, under nonskewing conditions. Monocytes were activated in 24-well plates (2.5×10^5 to 5.0×10^5 cells per well) by addition of LPS (50 ng/ml). Cell viability was monitored by either propidium iodide (BD Biosciences) or the LIVE/DEAD Cell Viability Assay (Life Technologies).

Mouse cells: Single cell suspensions of spleen cells were generated and red blood cells lysed using ACK lysis buffer (Life Technologies). $CD4^+$ T cells were isolated by negative selection using the Stem Cell Technologies EasySep Mouse $CD4^+$ T Cell Isolation Kit (Tukwila, WA). To obtain pure $CD4^+$ T cell populations, $CD4^+$ cells were sorted using a FACS Aria (BD Biosciences) based on $CD4^+$ $CD45.2^+$ staining and to separate naïve versus memory $CD4^+$ T lymphocytes. T cells were sort-separated based on $CD4^+$ $CD44^+$ (memory) and $CD4^+$ $CD44^-$ (naïve) stainings. For in vitro T cell activation, 48- or 96-well plates were coated with anti-CD3 (2 μg/ml) overnight at 4°C, and $CD4^+$ T cells (0.5×10^6 to 1.0×10^6 per well of 48-well plates or 0.2×10^6 per well of 96-well plates) were added to the appropriate wells. Anti-CD28 (1 μg/ml) was added to the media to provide costimulation.

Lymphocytic choriomeningitis virus (LCMV) infection in mice

Preparation of mixed bone marrow (BM) chimeric mice: B6.SJL (CD45.1,1) mice were lethally irradiated (950 rad) and reconstituted with a total of 10^7 donor BM cells from C57BL/6 CD45.1,2 wild-type mice mixed at equal parts with BM cells from CD45.2,2 mice deficient (KO) in *Nlrp3*$^{-/-}$, *Il1r1*$^{-/-}$,

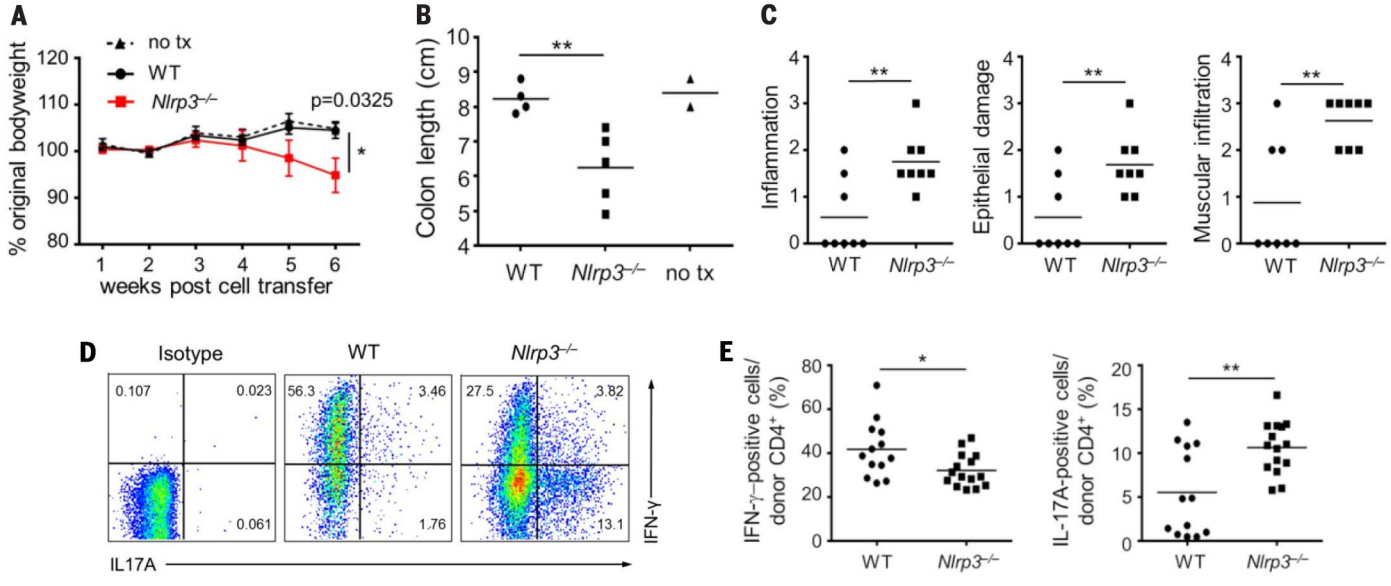


Fig. 7. T cell intrinsic NLRP3 activity regulates the T_H1-T_H17 balance in intestinal inflammation. (**A** to **E**) Naïve splenic $CD25^-CD45RB^{hi}$ $CD4^+$ T cells from wild-type or *Nlpr3*$^{-/-}$ mice were transferred into C57BL/10 *Rag2*$^{-/-}$ mice. (**A**) Weight change over the course of colitis induction. (**B**) Colon length at the study endpoint. (**C**) Inflammation score of the colons according to blinded histological analysis with assessment of inflammation (left panel), epithelial damage (middle panel) and muscular immune cell infiltration (right panel). (**D** and **E**) Intracellular IFN-γ and IL-17A staining of colonic $CD4^+$ T cells at the study endpoint after overnight α-CD3 + α-CD28 stimulation and brefeldin A and monensin addition for 5 hours (gated on live $CD4^+$ $Thy1.2^+$ T cells). Representative flow cytometric plots (D) with corresponding statistical analysis shown from two combined independent experiments [(E), $n = 13$ wild-type, $n = 15$ KO]. Data are means ± SEM. $^*P < 0.05$, $^{**}P < 0.01$. (A) and (B), one-way ANOVA with Sidak multiple-comparisons test; (C) and (E), unpaired t test. Data are representative of two experiments [(A) to (C)] or are combined from two experiments (E).

or *Il1a/Il1b*$^{-/-}$. Mice were allowed to reconstitute for 10 weeks before infection with LCMV.

LCMV infection and assessment of antigen-specific CD4$^+$ T cell response: Ten weeks after reconstitution, the mice were infected i.p. with 10^5 pfu of LCMV-Armstrong. On day 12 after infection, the mice were killed and the spleens removed for processing. For ex vivo cytokine staining of mouse cells after LCMV infection, cells were incubated with LCMV GP61-80 peptide (1 μg/ml) in the presence of monensin and brefeldin for 5 hours at 37°C. Staining for LCMV-specific CD4$^+$ T cells was performed using an APC-labeled 1A^b LCMV GP66-77 tetramer (NIH tetramer core facility) as described (*56*). Data were acquired with a FACS Calibur, Fortessa LSRIII, or FACS Aria cytometer (BD Biosciences) and analyzed with FlowJo 10.0.8 software (Ashland, OR).

Induction of colitis and colon cell isolation

Splenic CD4$^+$ T cells were isolated from C57BL/6 or *Nlpr3*$^{-/-}$ mice using a negative selection CD4 T cell enrichment kit (Stemcell tech), were stained with anti-CD45RB FITC, anti-CD25 APC, and anti-CD4 BV421, and sorted on a FACS Aria (BD biosciences) for CD4$^+$ CD25$^-$CD45RB^{hi} (brightest 35%) cells. Wild-type or *Nlpr3*$^{-/-}$ cells (2×10^5 each) were injected i.p. into age- and sex-matched C57BL/10 RAG2$^{-/-}$ mice. The mice were killed when symptoms of clinical disease (5 to 10% weight loss of original body weight and/or diarrhea) were observed in at least one group, approximately 6 to 11 weeks after adoptive transfer. Colon lamina propria cells were isolated as described (*57*), with the additional step of further purifying the cells over a 44 and 67% Percoll gradient to enrich for the mononuclear cells.

Scoring of intestinal inflammation

Samples of the proximal, mid-, and distal colon were excised after feces were flushed from the colons, placed into 3.7% formaldehyde solution, and then paraffin-embedded. Cross-sectional sections were cut and stained with hematoxylin and eosin (H&E). Colon pathology scores were based on severity of mononuclear cell inflammation, intestinal wall thickening, including infiltration to the muscularis, and epithelial damage, including edema, degeneration, and necrosis on a graded scale where 0 = normal, 0.5 = very mild, 1 = mild, 2 = moderate, 3 = severe. Samples were scored blinded by a pathologist from the NIH Pathology Score.

Induction of graft versus host disease (GvHD)

Balb/c mice were lethally irradiated with 900 cGy (two doses of 450 cGy 3 hours apart) on day –1. C57BL/6 wild-type bone marrow was depleted of T cells with the use of a CD90.2 Positive Selection Kit (Stemcell tech), and 5×10^6 cells were transferred on the following day (day 0) alone (control), or in addition to 1×10^6 wild-type B6 or *Nlpr3*$^{-/-}$ naïve CD4$^+$ T cells isolated with the Negative Selection Naïve CD4 T Cell Kit (Stemcell tech). Mice were killed upon clinical symptoms of disease (diarrhea and weight loss) on day 12 after cell transfer.

Detection of active caspase-1 and reactive oxygen species (ROS)

Generation of cleaved and active caspase-1 in cells was monitored by Western blotting for appropriate active fragment generation and by using the Green FLICA Caspase-1 Assay Kit (ImmunoChemistry Technologies, Bloomington, MN) according to the manufacturer's protocol with subsequent FACS analysis. ROS staining was performed by incubating cells to be assayed in dihydrorhodamine 123 (17 μg/ml) diluted in Hank's balanced salt solution with 10 mM HEPES (all from Sigma Aldrich) for 15 min at 37°C. Data were acquired on a FACS Calibur or Fortessa LSRIII cytometer (BD Biosciences) and analyzed with FlowJo software.

Confocal microscopy

Cells were fixed and permeabilized using the Cytofix/Cytoperm Kit (BD Biosciences) and stained with the indicated primary antibodies overnight and with secondary antibodies for 30 min at 4°C. Cells were mounted using VECTASHIELD media with DAPI (Vector Laboratories, Burlingame, CA) and images were acquired with a Nikon A1R confocal microscope (Nikon Imaging Centre, King's College London) and analyzed using NIS Elements (Nikon) and ImageJ software (National Institutes of Health).

Binding studies with recombinant human ^{125}I-labeled C5a

CD4$^+$ T cells from healthy donors were left nonactivated or activated for 4 hours with immobilized antibodies to CD3 and CD46 and then incubated for 2 hours at 4°C (1×10^7 cells/ml) with 10 μl of 0.1 nM ^{125}I-rhC5a (Perkin Elmer) and either 400 nM nonlabeled rhC5a in HAG-CM buffer (1 mM $CaCl_2$, 1 mM $MgCl_2$, 0.25% bovine serum albumin, 0.5 mM glucose, pH 7.4) or buffer without rhC5a addition. Cells were vacuum-transferred onto 96-well MultiScreen-HV filter plates (MAHVN4510; Millipore/Merck), nonbound ^{125}I-rhC5a removed by washing and cell-bound ^{125}I-rhC5a detected on the filter membranes by ^{125}I using a Packard Cobra II Gamma Counter (Perkin Elmer). For binding controls, HEK 293 cells (ATCC CRL 1573) were stably transfected with the pQCXIN vector expressing hC5aR1 or hC5aR2 (leading to expression of >1 Mio. of the respective C5aR/cell) or with the "empty" vector as control (*58*) (these cell lines also served as specificity controls for the anti-C5a receptor antibodies used in this study). In order to get comparable low CPM values as observed with purified T cells, only 5×10^4 cells/ml of C5aR1- or C5aR2-expressing HEK cells were applied. They were diluted in buffer containing "no-C5aR-expressing control" cells. The constant higher number of cells (5×10^5 HEK cells/ml in the 30-μl volume later used in the binding assay) permitted repetitive washing without cell loss and ensured identical nonspecific binding in all samples containing the same cell type. C5aR1-, C5aR2-expressing or control HEK 293 cells were incubated for 1 hour with or without 100 nM of nonlabeled rhC5a, washed thoroughly and then incubated for an additional 2 hours with 10 μl of 0.1 nM ^{125}I-rhC5a. After removal of nonbound rhC5a, binding of ^{125}I-C5a to the respective HEK 293 cell lines was determined by measuring gamma radioactivity. To exclude C5a-induced C5aR-internalization during all binding studies all steps in the binding experiments were performed at 4°C and HEK 293 cells were additionally preincubated 15 min at 37°C with 0.1% NaAcid and Cytochalasin B (21 μg/ml) and then cooled on ice for 5 min before their incubation with rhC5a.

Cytokine measurements

Cytokine production by cells in culture was quantified from cell supernatants using either the human or mouse $T_H1/T_H2/T_H17$ Cytometric Bead Array (BD Bioscience) or via intracellular cytokine staining after treated for 4 hours with PMA (50 ng/ml), ionomycin (1 μg/ml) (both Sigma Aldrich), and 1× Golgi Plug (BD Biosciences). Secreted human IL-1β and IL-18 were measured using the Human IL-1β/IL-1F2 DuoSet Kit or the Human IL-18 Platinum ELISA kit (R&D Systems and eBiosciences, respectively) in combination with SIGMAFAST OPD tablets (Sigma Aldrich) as substrate for detection.

RNA extraction, reverse transcription polymerase chain reaction (RT-PCR), and quantitative RT-PCR

RNA was extracted with the RNeasy Mini Kit including DNase digestion and DNA cleanup (Qiagen), and reverse transcription was performed with the One Step RT-PCR (Qiagen). For quantitative PCR, RNA was reverse-transcribed with 2.5 μM random hexamers, 1 mM dNTPs, 40 U of RiboLock RNase inhibitor, and 400 U of RevertAid H Minus Reverse Transcriptase (Thermo Scientific). Quantitative PCR was performed using KI-Q Hot Start Sybr Green Mix (Sigma Aldrich), with 150 pmol of forward and reverse primers and data acquired on the CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Primer sequences are listed in table S5.

RNA silencing

SiRNA targeting human C5AR1 mRNA and control scrambled siRNA were purchased from Origene (Rockville, MD) and delivered at a final concentration of 15 nM (mixture of three different C5aR1 siRNA used at 5 nM each or scramble control at 15 nM) into primary human CD4$^+$ T cells by transfection with Lipofectamine RNAiMAX (Life Technologies) following the manufacturer's instructions. *C5AR1* mRNA level reduction was consistently about 30%.

Microarray data generation and analysis

Transcriptome profiling was performed by the KCL Genomic Centre (London) using human exon 1.0 ST arrays (Affymetrix) on CD4$^+$ T cells isolated from three different healthy donors that were left either nonactivated or were activated with antibodies to CD3 and CD46 for 2 hours in the absence or presence of the C5aR1/C5aR2 antagonist. Expression data were analyzed using

Partek Genomics Suite (Partek Inc., St. Louis, MO) version 6.6 and Gene Set Enrichment Analysis, GSEA (*59*) (Broad Institute of MIT and Harvard) with a normalized enrichment score of 1.8 to derive normalized enrichment score (NES), nominal *P* value, and FDR *q* value. Microarray data sets were used in conjunction with the Qiagen-generated inflammasome gene set (84 members). Heat maps for the leading edge subset were drawn with Partek genomics suite. Table S1 shows the normalized read values from microarrays for Fig. 2, A and B. The list of annotated genes differentially regulated by the C5aR1/C5aR2 double antagonist (fig. S2) is given in table S2.

Statistical analysis

Analyses were performed with GraphPad Prism (La Jolla, CA). Data are presented as mean ± SEM and compared using either paired *t* tests with Bonferroni correction for multiple comparisons, one- or two-way analysis of variance (ANOVA) with a Tukey multiple comparison post hoc test or with Sidak multiple-comparisons test, as appropriate. Correlation analysis (Fig. 3D and fig. S3B) was performed with Spearman's correlation test. *P* values < 0.05 denote statistical significance.

REFERENCES AND NOTES

1. D. Ricklin, G. Hajishengallis, K. Yang, J. D. Lambris, Complement: A key system for immune surveillance and homeostasis. *Nat. Immunol.* **11**, 785–797 (2010). doi: 10.1038/ni.1923; pmid: 20720586
2. K. M. Murphy, B. Stockinger, Effector T cell plasticity: Flexibility in the face of changing circumstances. *Nat. Immunol.* **11**, 674–680 (2010). doi: 10.1038/ni.1899; pmid: 20644573
3. J. Cardone *et al*., Complement regulator CD46 temporally regulates cytokine production by conventional and unconventional T cells. *Nat. Immunol.* **11**, 862–871 (2010). doi: 10.1038/ni.1917; pmid: 20694009
4. M. K. Liszewski *et al*., Intracellular complement activation sustains T cell homeostasis and mediates effector differentiation. *Immunity* **39**, 1143–1157 (2013). doi: 10.1038/ni.1917; pmid: 20694009
5. G. Le Friec *et al*., The CD46-Jagged1 interaction is critical for human TH1 immunity. *Nat. Immunol.* **13**, 1213–1221 (2012). doi: 10.1038/ni.2454; pmid: 23086448
6. H. Yamamoto, A. F. Fara, P. Dasgupta, C. Kemper, CD46: The 'multitasker' of complement proteins. *Int. J. Biochem. Cell Biol.* **45**, 2808–2820 (2013). doi: 10.1016/j.biocel.2013.09.016; pmid: 24120647
7. M. Kolev *et al*., Complement regulates nutrient influx and metabolic reprogramming during Th1 cell responses. *Immunity* **42**, 1033–1047 (2015). doi: 10.1016/j.immuni.2015.05.024; pmid: 26084023
8. A. Ghannam *et al*., Human C3 deficiency associated with impairments in dendritic cell differentiation, memory B cells, and regulatory T cells. *J. Immunol.* **181**, 5158–5166 (2008). doi: 10.4049/jimmunol.181.7.5158; pmid: 18802120
9. A. L. Astier, G. Meiffren, S. Freeman, D. A. Hafler, Alterations in CD46-mediated Tr1 regulatory T cells in patients with multiple sclerosis. *J. Clin. Invest.* **116**, 3252–3257 (2006). doi: 10.1172/JCI29251; pmid: 17099776
10. J. V. Sarma, P. A. Ward, New developments in C5a receptor signaling. *Cell Health Cytoskelet.* **4**, 73–82 (2012). pmid: 23576881
11. A. M. Scola, K. O. Johswich, B. P. Morgan, A. Klos, P. N. Monk, The human complement fragment receptor, C5L2, is a recycling decoy receptor. *Mol. Immunol.* **46**, 1149–1162 (2009). doi: 10.1016/j.molimm.2008.11.001; pmid: 19100624
12. N. P. Gerard *et al*., An anti-inflammatory function for the complement anaphylatoxin C5a-binding protein, C5L2. *J. Biol. Chem.* **280**, 39677–39680 (2005). doi: 10.1074/jbc.C500287200; pmid: 16204243
13. C. E. Bamberg *et al*., The C5a receptor (C5aR) C5L2 is a modulator of C5aR-mediated signal transduction. *J. Biol. Chem.* **285**, 7633–7644 (2010). doi: 10.1074/jbc.M109.092106; pmid: 20044484
14. R. Li, L. G. Coulthard, M. C. Wu, S. M. Taylor, T. M. Woodruff, C5L2: A controversial receptor of complement anaphylatoxin, C5a. *FASEB J.* **27**, 855–864 (2013). doi: 10.1096/fj.12-220509; pmid: 23239822
15. D. E. Croker, R. Halai, D. P. Fairlie, M. A. Cooper, C5a, but not C5a-des Arg, induces upregulation of heteromer formation between complement C5a receptors C5aR and C5L2. *Immunol. Cell Biol.* **91**, 625–633 (2013). doi: 10.1038/icb.2013.48; pmid: 24060963
16. C. Vogel, E. M. Marcotte, Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* **13**, 227–232 (2012). pmid: 22411467
17. T. M. Woodruff *et al*., Therapeutic activity of C5a receptor antagonists in a rat model of neurodegeneration. *FASEB J.* **20**, 1407–1417 (2006). doi: 10.1096/fj.05-5814com; pmid: 16816116
18. M. Otto *et al*., C5a mutants are potent antagonists of the C5a receptor (CD88) and of C5L2: Position 69 is the locus that determines agonism or antagonism. *J. Biol. Chem.* **279**, 142–151 (2004). doi: 10.1074/jbc.M310078200; pmid: 14570896
19. P. N. Monk *et al*., De novo protein design of agonists and antagonists of C5a receptors. *Immunobiology* **217**, 1162–1163 (2012). doi: 10.1016/j.imbio.2012.08.097
20. A. Kastbom *et al*., Genetic variation in proteins of the cryopyrin inflammasome influences susceptibility and severity of rheumatoid arthritis (the Swedish TIRA project). *Rheumatology* **47**, 415–417 (2008). doi: 10.1093/rheumatology/kem372; pmid: 18263599
21. A. Pontillo *et al*., Two SNPs in NLRP3 gene are involved in the predisposition to type-1 diabetes and celiac disease in a pediatric population from northeast Brazil. *Autoimmunity* **43**, 583–589 (2010). doi: 10.3109/08916930903540432; pmid: 20370570
22. P. J. Shaw, M. F. McDermott, T. D. Kanneganti, Inflammasomes and autoimmunity. *Trends Mol. Med.* **17**, 57–64 (2011). doi: 10.1016/j.molmed.2010.11.001; pmid: 21163704
23. C. E. Zielinski *et al*., Pathogen-induced human TH17 cells produce IFN-γ or IL-10 and are regulated by IL-1β. *Nature* **484**, 514–518 (2012). doi: 10.1038/nature10957; pmid: 22466287
24. C. Sutton, C. Brereton, B. Keogh, K. H. Mills, E. C. Lavelle, A crucial role for interleukin (IL)-1 in the induction of IL-17-producing T cells that mediate autoimmune encephalomyelitis. *J. Exp. Med.* **203**, 1685–1691 (2006). doi: 10.1084/jem.20060285; pmid: 16818675
25. D. A. Rao, K. J. Tracey, J. S. Pober, IL-1α and IL-1β are endogenous mediators linking cell injury to the adaptive alloimmune response. *J. Immunol.* **179**, 6536–6546 (2007). doi: 10.4049/jimmunol.179.10.6536; pmid: 17982042
26. C. E. Sutton *et al*., Interleukin-1 and IL-23 induce innate IL-17 production from γδ T cells, amplifying Th17 responses and autoimmunity. *Immunity* **31**, 331–341 (2009). doi: 10.1016/j.immuni.2009.08.001; pmid: 19682929
27. D. Schenten *et al*., Signaling through the adaptor molecule MyD88 in $CD4^+$ T cells is required to overcome suppression by regulatory T cells. *Immunity* **40**, 78–90 (2014). doi: 10.1016/j.immuni.2013.10.023; pmid: 24439266
28. W. P. Arend, G. Palmer, C. Gabay, IL-1, IL-18, and IL-33 families of cytokines. *Immunol. Rev.* **223**, 20–38 (2008). doi: 10.1111/j.1600-065X.2008.00624.x; pmid: 18613828
29. F. Martinon, J. Tschopp, NLRs join TLRs as innate sensors of pathogens. *Trends Immunol.* **26**, 447–454 (2005). doi: 10.1016/j.it.2005.06.004; pmid: 15967716
30. J. K. Dowling, L. A. O'Neill, Biochemical regulation of the inflammasome. *Crit. Rev. Biochem. Mol. Biol.* **47**, 424–443 (2012). doi: 10.3109/10409238.2012.694844; pmid: 22681257
31. C. A. Dinarello, Biologic basis for interleukin-1 in disease. *Blood* **87**, 2095–2147 (1996). doi: 10.1111/j.1600-065X.2008.00624.x; pmid: 18613828
32. M. T. Heneka *et al*., NLRP3 is activated in Alzheimer's disease and contributes to pathology in APP/PS1 mice. *Nature* **493**, 674–678 (2013). doi: 10.1038/nature11729; pmid: 23254930
33. K. Shahzad *et al*., Nlrp3-inflammasome activation in non-myeloid-derived cells aggravates diabetic nephropathy. *Kidney Int.* **87**, 74–84 (2015). doi: 10.1038/ki.2014.271; pmid: 25075770
34. W. A. Tseng *et al*., NLRP3 inflammasome activation in retinal pigment epithelial cells by lysosomal destabilization: Implications for age-related macular degeneration. *Invest. Ophthalmol. Vis. Sci.* **54**, 110–120 (2013). doi: 10.1167/iovs.12-10655; pmid: 23221073
35. S. Z. Ben-Sasson *et al*., IL-1 acts directly on CD4 T cells to enhance their antigen-driven expansion and differentiation. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7119–7124 (2009). doi: 10.1073/pnas.0902745106; pmid: 25686105
36. R. C. Coll *et al*., A small-molecule inhibitor of the NLRP3 inflammasome for the treatment of inflammatory diseases. *Nat. Med.* **21**, 248–255 (2015). pmid: 25686105
37. H. Okamura *et al*., Cloning of a new cytokine that induces IFN-γ production by T cells. *Nature* **378**, 88–91 (1995). doi: 10.1038/378088a0; pmid: 7477296
38. P. Menu, J. E. Vince, The NLRP3 inflammasome in health and disease: The good, the bad and the ugly. *Clin. Exp. Immunol.* **166**, 1–15 (2011). doi: 10.1038/378088a0; pmid: 7477296
39. H. J. Lachmann *et al*., In vivo regulation of interleukin 1β in patients with cryopyrin-associated periodic syndromes. *J. Exp. Med.* **206**, 1029–1036 (2009). doi: 10.1084/jem.20082481; pmid: 19364880
40. S. Carta *et al*., Cell stress increases ATP release in NLRP3 inflammasome-mediated autoinflammatory diseases, resulting in cytokine imbalance. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2835–2840 (2015). doi: 10.1073/pnas.1424741112; pmid: 25730877
41. K. Schroder, R. Zhou, J. Tschopp, The NLRP3 inflammasome: A sensor for metabolic danger? *Science* **327**, 296–300 (2010). doi: 10.1126/science.1184003; pmid: 20075245
42. O. Samstad *et al*., Cholesterol crystals induce complement-dependent inflammasome activation and cytokine release. *J. Immunol.* **192**, 2837–2845 (2014). doi: 10.1126/science.1184003; pmid: 20075245
43. L. A. Sena *et al*., Mitochondria are required for antigen-specific T cell activation through reactive oxygen species signaling. *Immunity* **38**, 225–236 (2013). doi: 10.1016/j.immuni.2012.10.020; pmid: 23415911
44. M. Coccia *et al*., IL-1β mediates chronic intestinal inflammation by promoting the accumulation of IL-17A secreting innate lymphoid cells and $CD4^+$ Th17 cells. *J. Exp. Med.* **209**, 1595–1609 (2012). doi: 10.1084/jem.20111453; pmid: 22891275
45. M. F. Neurath, Cytokines in inflammatory bowel disease. *Nat. Rev. Immunol.* **14**, 329–342 (2014). doi: 10.1038/nri3661; pmid: 24751956
46. C. Bauer, P. Duewell, H. A. Lehr, S. Endres, M. Schnurr, Protective and aggravating effects of Nlrp3 inflammasome activation in IBD models: Influence of genetic and environmental factors. *Dig. Dis.* **30** (suppl. 1), 82–90 (2012). doi: 10.1159/000341681; pmid: 23075874
47. N. Powell *et al*., The transcription factor T-bet regulates intestinal inflammation mediated by interleukin-7 $receptor^+$ innate lymphoid cells. *Immunity* **37**, 674–684 (2012). doi: 10.1016/j.immuni.2012.09.008; pmid: 23063332
48. M. Bruchard *et al*., The receptor NLRP3 is a transcriptional regulator of T_H2 differentiation. *Nat. Immunol.* **16**, 859–870 (2015). doi: 10.1038/ni.3202; pmid: 26098997
49. M. G. Strainic *et al*., Locally produced complement fragments C5a and C3a provide both costimulatory and survival signals to naive $CD4^+$ T cells. *Immunity* **28**, 425–435 (2008). doi: 10.1016/j.immuni.2008.02.001; pmid: 18328742
50. J. Dunkelberger, L. Zhou, T. Miwa, W. C. Song, C5aR expression in a novel GFP reporter gene knockin mouse: Implications for the mechanism of action of C5aR signaling in T cell immunity. *J. Immunol.* **188**, 4032–4042 (2012). doi: 10.4049/jimmunol.1103141; pmid: 22430734
51. C. M. Karsten *et al*., Monitoring and cell-specific deletion of C5aR1 using a novel floxed GFP-C5aR1 reporter knock-in mouse. *J. Immunol.* **194**, 1841–1855 (2015). pmid: 25589074
52. M. Kolev, G. Le Friec, C. Kemper, Complement—tapping into new sites and effector systems. *Nat. Rev. Immunol.* **14**, 811–820 (2014). doi: 10.1038/nri3761; pmid: 25394942
53. G. Doitsh *et al*., Cell death by pyroptosis drives CD4 T-cell depletion in HIV-1 infection. *Nature* **505**, 509–514 (2014). doi: 10.1038/nature12940; pmid: 24356306
54. S. Z. Ben-Sasson *et al*., IL-1 enhances expansion, effector function, tissue localization, and memory response of antigen-specific CD8 T cells. *J. Exp. Med.* **210**, 491–502 (2013). doi: 10.1084/jem.20122006; pmid: 23460726
55. K. D. Mayer-Barber *et al*., Host-directed therapy of tuberculosis based on interleukin-1 and type I interferon crosstalk. *Nature* **511**, 99–103 (2014). doi: 10.1038/nature13489; pmid: 24990750
56. M. A. Rasheed *et al*., Interleukin-21 is a critical cytokine for the generation of virus-specific long-lived plasma cells. *J. Virol.* **87**, 7737–7746 (2013). doi: 10.1128/JVI.00063-13; pmid: 23637417
57. V. Valatas *et al*., Host-dependent control of early regulatory and effector T-cell differentiation underlies the genetic susceptibility of RAG2-deficient mouse strains to transfer

colitis. *Mucosal Immunol.* **6**, 601–611 (2013). doi: 10.1038/mi.2012.102; pmid: 23149660
58. D. Bock *et al.*, The C terminus of the human C5a receptor (CD88) is required for normal ligand-dependent receptor internalization. *Eur. J. Immunol.* **27**, 1522–1529 (1997). doi: 10.1002/eji.1830270631; pmid: 9209506
59. A. Subramanian *et al.*, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi: 10.1073/pnas.0506580102; pmid: 16199517

ACKNOWLEDGMENTS

We thank the healthy volunteers and the patients for their support, A. Fara and J. Sumner (King's College London) for initial contributions to the C5aR2 antagonism studies, and Y. Iwakura (Tokyo University) for providing the combined *Il1a/Il1b*$^{-/-}$ mice. The MHC class II tetramer to LCMV (GP66-77) was kindly provided by the NIH Tetramer Core. We thank the Nikon Imaging Centre and the Genomics Centre at King's College London. The data presented in this manuscript are tabulated in the main paper and in the supplementary materials. The detailed gene array data have been deposited in the GEO database with accession number GSE69090. Supported by MRC Centre grant MR/J006742/1, an EU-funded Innovative Medicines Initiative BTCURE (C.K. and A.P.C.); a Wellcome Trust Investigator Award (C.K.); a Wellcome Trust Intermediate Clinical Fellowship award (B.A.); British Heart Foundation grant PG/09/018/25279 (P.M.); Science Foundation Ireland grant G20598 and Australia National Health and Medical Research Council project grant APP1086786 (R.C.C.); the King's Bioscience Institute at King's College London (G.A.); the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St. Thomas' NHS Foundation Trust and King's College London; and the Division of Intramural Research, National Heart, Lung, and Blood Institute, and the intramural research program of NIAID. T.M.W., P.M., and M.A.C. are co-inventors on a U.S. provisional patent application (62/242632) related to the C5aR2 agonist used this manuscript. J.K. is inventor on U.S. patent 8/542,862, which was filed by Cincinnati Children's Hospital related to the C5aR1/C5aR2 antagonist A8D71-73. The authors declare no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/aad1210/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S5

29 July 2015; accepted 20 April 2016
10.1126/science.aad1210

RESEARCH ARTICLES

TRANSLATION

Real-time quantification of single RNA translation dynamics in living cells

Tatsuya Morisaki,[1] Kenneth Lyon,[1] Keith F. DeLuca,[1] Jennifer G. DeLuca,[1] Brian P. English,[2] Zhengjian Zhang,[2] Luke D. Lavis,[2] Jonathan B. Grimm,[2] Sarada Viswanathan,[2] Loren L. Looger,[2] Timothee Lionnet,[2] Timothy J. Stasevich[1]*

Although messenger RNA (mRNA) translation is a fundamental biological process, it has never been imaged in real time in vivo with single-molecule precision. To achieve this, we developed nascent chain tracking (NCT), a technique that uses multi-epitope tags and antibody-based fluorescent probes to quantify protein synthesis dynamics at the single-mRNA level. NCT reveals an elongation rate of ~10 amino acids per second, with initiation occurring stochastically every ~30 seconds. Polysomes contain ~1 ribosome every 200 to 900 nucleotides and are globular rather than elongated in shape. By developing multicolor probes, we showed that most polysomes act independently; however, a small fraction (~5%) form complexes in which two distinct mRNAs can be translated simultaneously. The sensitivity and versatility of NCT make it a powerful new tool for quantifying mRNA translation kinetics.

At the core of all gene regulatory networks are the processes of DNA transcription and RNA translation. Although transcription is now regularly quantified in real time in vivo with single-gene resolution (*1*, *2*), the same cannot be said for translation. In principle, fluorescent tags such as green fluorescent protein (GFP) could be used to do this, but in practice, these tags take too long to fluoresce or their signals are too weak to visualize translation of a single RNA in real time (*3*). Recently, some of these problems were overcome by using a RNA biosensor that is dislodged by translating ribosomes (*4*). However, fluorescence is lost on the first round of translation with this assay, so ongoing translation dynamics cannot be visualized.

Developing a method to visualize the translation of single mRNAs in living cells

To visualize translation of a single RNA in real time, we developed a system based on bright photostable small-molecule dyes, antibody enhancement, and multi-epitope protein tags. We constructed a plasmid encoding the large nuclear protein KDM5B, N-terminally tagged with a 10× FLAG tag (which we refer to as the spaghetti monster, SM) (*5*) and containing a 24× MS2 tag (*6*) in the 3′ untranslated region (UTR) (Fig. 1A). The FLAG SM tag created a highly avid site for binding of fluorescently labeled fragments of antibody to FLAG (anti-FLAG Fab), whereas the MS2 stem-loop repeat allowed visualization with labeled MS2 coat protein (MCP) (*2*, *6*) (Fig. 1B). We transiently transfected this plasmid into U2OS cells that were subsequently bead-loaded with Cy3-labeled anti-FLAG Fab and Halo-tagged MCP [labeled with the far-red JF646 fluorophore (*7*)]. Twenty-four hours after transfection, MCP marked mRNA in the cytoplasm and Fab marked KDM5B in the nucleus, suggesting that neither the FLAG SM tag nor the presence of Fab interfered with mRNA and KDM5B production and localization (Fig. 1C). To see how soon Fab could mark protein, we imaged ~6 hours after transfection. At this early time, Fab only lightly marked the nucleus, suggesting that very little KDM5B had been synthesized (Fig. 1D). Fab also colocalized and co-moved with many MCP-labeled mRNAs in the cytoplasm (Fig. 1, D and E, and movie S1). These bright co-moving spots displayed RNA-like diffusivity, were very stable—lasting for 2 hours or more (movie S2)—and could only be seen in the cytoplasm of cells that were both transfected with FLAG SM–tagged KDM5B and bead-loaded with Fab.

We wondered whether the co-moving protein-mRNA spots were bona fide translation sites. To test this, we treated cells with 50 μg/ml of puromycin, an inhibitor of translation that releases nascent chains from ribosomes (*8*). Within minutes of drug addition, the number of co-moving spots dropped exponentially (Fig. 1, F and G, and movie S3). By tracking co-moving spots, we could see the disappearance of Fab-labeled protein, despite the persistence of the mRNA (Fig. 1F). The time of complete protein disappearance varied among different mRNAs, but there was a well-defined exponential decay in the number of protein spots (Fig. 1G). To further confirm that protein-mRNA spots were translation sites, we treated cells with 4 μg/ml cycloheximide to slow elongation and load more ribosomes per transcript, and the spots became brighter (fig. S1). Together, these data suggest that the co-moving spots were indeed translation sites.

KDM5B is a large protein (1544 amino acids), so its translation should be relatively easy to detect. To see whether we could also detect translation of smaller proteins, we constructed two plasmids (Fig. 2A) encoding either β-actin (374 amino acids) or the core histone H2B (125 amino acids). As with KDM5B, neither the FLAG SM tag nor Fab disturbed the localization of these proteins (Fig. 2, B and C), and Western blots confirmed that tagged proteins were of the expected length (fig. S2). Furthermore, we could again observe translation sites ~6 hours after transient transfection (Fig. 2D and movies S4 and S5), indicating that our system is useful for imaging the translation of protein-coding genes of varying sizes.

Nascent chain tracking to quantify polysome mobility and size

We tracked translation of single molecules by using a technique that we call nascent chain tracking (NCT). With NCT, we followed individual H2B, β-actin, and KDM5B translation sites for hundreds to thousands of seconds in full cell volumes (300 time points × 2 colors × 13 *z* planes = 7800 images per movie). We adjusted laser powers to focus exclusively on translation sites rather than on fully translated single protein products (which could interfere with tracking). As shown in the inset of Fig. 2D, this allowed us to accurately compare (i) the appearance frequency and brightness, (ii) the mobility, and (iii) the size of translation sites. Of these parameters, frequency and brightness varied the most, tending to increase with construct length. We detected the translation of 86 ± 2% of KDM5B mRNA but just 19 ± 4% of β-actin mRNA and 4 ± 1% of H2B mRNA (error, SEM) (Fig. 2E). Furthermore, KDM5B translation sites, as marked by Fab, were over 1.5 times as bright as β-actin sites, which themselves were nearly 1.5 times as bright as H2B sites (Fig. 2F).

One explanation for the difference in brightness in translation sites could be a difference in the number of nascent peptide chains per mRNA, as would be the case in polysomes (*9*). To determine precisely how many nascent chains exist per site, we calibrated fluorescence by imaging a new β-actin plasmid containing a single 1× FLAG tag rather than the 10× FLAG SM tag (fig. S3A). With this plasmid, only one Fab can be bound per peptide chain, allowing a direct comparison between translation site fluorescence and single Fab fluorescence. By imaging cells transfected with this plasmid at high laser powers, both single Fabs and translation sites could be detected and tracked (fig. S3, B and C), revealing translation sites to be on average 3.1 ± 0.5 times as bright as single Fabs (fig. S3D). We therefore estimate that there are 3.1 ± 0.5 nascent peptide chains per β-actin translation site, 2.1 ± 0.4 per H2B site, and 5.1 ± 0.9 per KDM5B site (Fig. 2F, right axis). Combining these data and assuming one ribosome per

[1]Department of Biochemistry and Molecular Biology and Institute for Genome Architecture and Function, Colorado State University, Fort Collins, CO, USA. [2]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA.
*Corresponding author. Email: tim.stasevich@colostate.edu

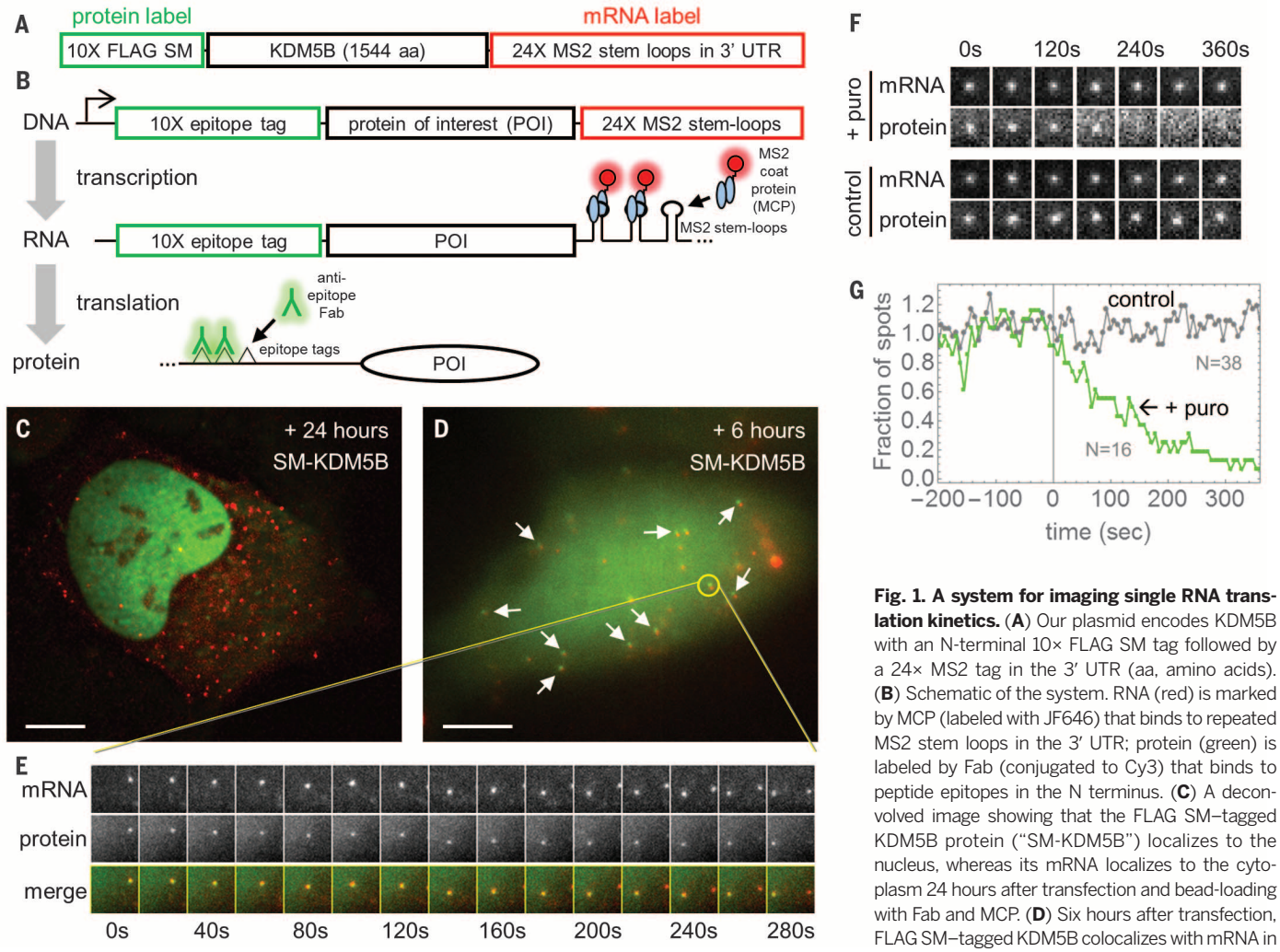


Fig. 1. A system for imaging single RNA translation kinetics. (**A**) Our plasmid encodes KDM5B with an N-terminal 10× FLAG SM tag followed by a 24× MS2 tag in the 3′ UTR (aa, amino acids). (**B**) Schematic of the system. RNA (red) is marked by MCP (labeled with JF646) that binds to repeated MS2 stem loops in the 3′ UTR; protein (green) is labeled by Fab (conjugated to Cy3) that binds to peptide epitopes in the N terminus. (**C**) A deconvolved image showing that the FLAG SM–tagged KDM5B protein ("SM-KDM5B") localizes to the nucleus, whereas its mRNA localizes to the cytoplasm 24 hours after transfection and bead-loading with Fab and MCP. (**D**) Six hours after transfection, FLAG SM–tagged KDM5B colocalizes with mRNA in punctae (arrows). (**E**) Example co-movement of FLAG SM–tagged KDM5B and mRNA punctae circled in yellow in (D). (**F**) Addition of puromycin (50 μg/ml) leads to a loss of the FLAG SM–tagged KDM5B signal in punctae; this does not occur in control cells loaded with vehicle. (**G**) Quantification of the loss of the FLAG SM–tagged KDM5B signal from punctae (lower green curve) in a single cell as a function of time after addition of puromycin (which takes effect at 0 s), compared with the signal from a control cell (upper gray curve). Curves are normalized to pre-puromycin levels. Scale bars, 10 μm.

nascent chain, we conclude that detected translation sites are polysomes that can contain as few as one ribosome every 900 mRNA bases or as many as one ribosome every 200 mRNA bases.

In addition to differences in their brightness, NCT also exposed differences in the mobility of polysomes. We quantified this by measuring the mean squared displacement of tracked polysomes as a function of time. For the nuclear proteins H2B and KDM5B, mobility was modeled well by diffusion. Not only did displacement increase linearly with time for at least 20 s, the less massive H2B polysomes also moved significantly faster than KDM5B polysomes (Fig. 2G). In contrast, β-actin polysomes displayed constrained diffusion, with jump sizes that were initially between those of KDM5B and H2B (consistent with diffusion) but that ultimately lagged behind both at longer times. This constrained movement of β-actin could be due to interactions with cytoplasmic binding partners. Despite these trends, there was substantial variability in mobility between mRNA, so that we sometimes saw rapidly moving KDM5B polysomes (up to 6 $\mu m^2/s$) and nearly immobile H2B polysomes (~0.01 $\mu m^2/s$; fig. S4). This made it nearly impossible to identify a translated mRNA based on mobility alone, implying that the translation machinery only weakly alters mRNA movement in our system.

Unlike their brightness and mobility, the size of polysomes was less variable between constructs. To quantify sizes, we measured the distance between the 3′ UTR of polysomal mRNA (labeled with MCP) and the nascent peptide chains (labeled with Fab). The fluorescence from polysomes was within diffraction-limited spots, so we determined their mean positions with superresolution by using Gaussian fitting (fig. S5). According to hairpin models of polysome organization (*10*), this distance should grow as the length of the mRNA grows. Instead, we found that the distance was shortest in KDM5B polysomes, typically around 65 nm, compared with 85 nm for H2B and 105 nm for β-actin (Fig. 2H). This suggests that the polysomes that we imaged are organized in a globular shape rather than an elongated shape, consistent with recent atomic force microscopy images (*11*).

Extracting translation kinetics from NCT data

Having measured the basic physical properties of polysomes, we next focused on translation kinetics. In particular, we wondered how the number of ribosomes in polysomes is controlled. This number reflects a balance between incoming and outgoing ribosomes and therefore depends on the ribosome elongation rate. One way to noninvasively estimate this rate is to examine the correlation of fluctuations in NCT data by means of fluorescence correlation spectroscopy (FCS), similar to how transcription elongation rates have been measured by using MS2 fluorescence fluctuations (*12*). Before doing so, however, we needed to ensure that Fabs would bind polysomes quickly and irreversibly on the time scale of translation. If not, their dynamics would contribute to the

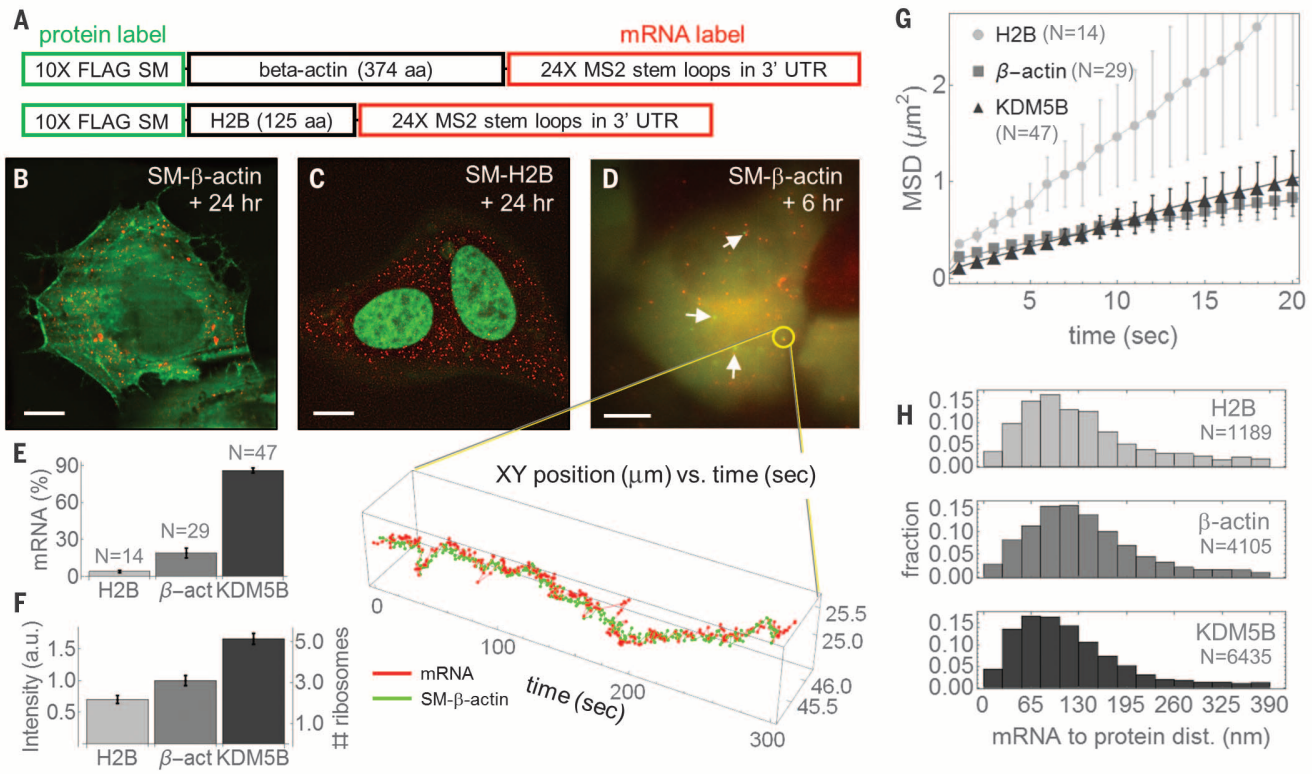


Fig. 2. Quantifying the mobility, ribosomal content, and structure of translation sites. (**A**) Plasmids encoding β-actin (SM-β-actin) and H2B (SM-H2B), analogous to the FLAG SM–tagged KDM5B construct in Fig. 1A. (**B** and **C**) Deconvolved images showing that SM-β-actin [(B), green] and SM-H2B [(C), green] localize to the cytoplasm and nucleus, respectively, whereas their mRNAs (red) both localize to the cytoplasm 24 hours after transfection and bead-loading with Fab and MCP. (**D**) Six hours after transfection, SM-β-actin translation sites can be seen (arrows) and tracked (yellow circle and inset). (**E** and **F**) Quantification of the frequency (E) and intensity (F) of translation sites. Normalization to the number of ribosomes is shown on the right axis in (F) (a.u., arbitrary units). (**G**) Measured mean squared displacements (MSD) of tracked polysomes as a function of time. (**H**) Histograms of the Gaussian fit distances between mRNA and protein in tracked polysomes. Error bars show SEM. Scale bars, 10 μm.

fluctuations, and this would distort our FCS measurements of elongation times.

To measure how quickly Fabs bind polysomes, we microinjected them into cells transfected 6 hours earlier with our KDM5B construct and preloaded with MCP. Just 3 s after microinjection (as soon as we could image), many polysomes were labeled by Fab, implying that the binding time is less than 3 s (Fig. 3, A and B). To measure the lifetime of Fab binding, we performed fluorescence recovery after photobleaching (FRAP) experiments in cells transfected with the H2B construct and bead-loaded with Fab and MCP 24 hours earlier. We chose H2B because it is known to remain bound for hours (*13*, *14*), so any fluorescence recovery on the time scale of minutes would be exclusively due to Fab turnover. As Fig. 3C shows, there was little FRAP recovery in 4 min, implying that most Fabs are bound much longer. These binding kinetics (Fig. 3D) make Fabs ideal tools for measuring translation elongation times on time scales ranging from ~10 s to ~5 min.

Knowing the limits of Fab, we analyzed the fluorescence fluctuations of polysomes by extracting their intensity time series from our tracking data. We began with KDM5B polysomes because these were the brightest and most numerous. The intensity of polysomes fluctuated with time (Fig. 3E), reflecting changes in the number of elongating ribosomes. From each intensity time series, we computed the correlation curve, and we averaged these together. There was high variability among mRNAs (fig. S6), but the average correlation curve revealed a clear linear drop in correlation that went to zero at 149 ± 20 s (Fig. 3F, upper panel). In direct analogy to transcription correlation analyses (*12*), the time at which the correlation drops to zero marks the total elongation dwell time. To confirm this, we treated cells with 100 μg/ml of cycloheximide to stall translation. As expected, the correlation disappeared (fig. S7 and movies S6 and S7).

To corroborate these measurements, we performed FRAP on KDM5B polysomes. We photobleached a large section of the cytoplasm where many polysomes were present (fig. S8). By tuning the powers of the photobleaching laser, we could selectively photobleach just the Fab, leaving the mRNA bright. This allowed us to monitor the fluorescence recovery of the relatively slower-moving polysomes. On average, polysomes recovered 80 to 90% of their fluorescence in 125 to 180 s, although there was again high variability among mRNAs, just as with FCS. Nevertheless, the average recovery time was on the same time scale as our FCS measurements.

Given this consistency, we next performed FCS on the shorter β-actin and H2B constructs. Again, the correlation curves were linear and went to zero at distinct elongation dwell times, whereas random background spots showed no correlations from frame to frame (fig. S9). As expected, the dwell times decreased with mRNA length, being 32 ± 9 s for β-actin and just 16 ± 7 s for H2B (Fig. 3F, lower panels). Importantly, for all constructs, the correlation vanished at times greater than the dwell time. This implies that initiation is random, so there is no memory between initiation events, similar to transcriptional initiation (*12*) and in contrast to bursting (*15*, *16*).

To calculate the elongation rate, we divided the length of the encoded protein by the elongation dwell time of each construct. The calculated rates were all within error (Fig. 3G), yielding a single consistent elongation rate of 10 ± 2.3 amino acids/s, which is fairly close to what has been measured using genome-wide ribosomal profiling (5.6 amino acids/s) (*17*, *18*). The difference is probably due to single-molecule variability (as shown in figs. S4 and S6) or differences in mRNA sequence and codon usage (*19*).

With a consistent elongation rate, we can unify our observations. First, assuming KDM5B elongation occurs at 10 ± 2.3 amino acids/s, a new

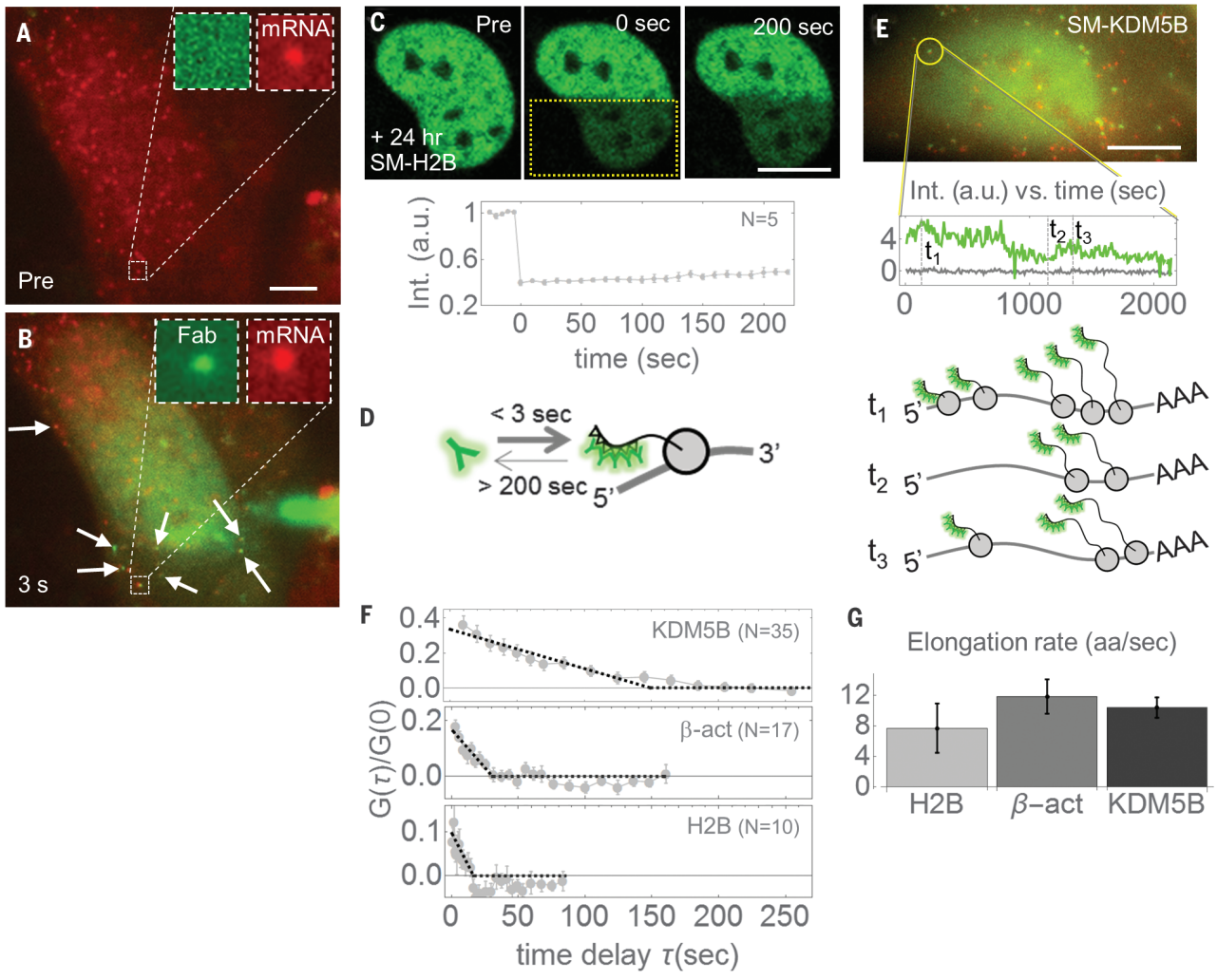


Fig. 3. Quantifying the translational kinetics of tracked polysomes. (**A**) Sample cell transfected with FLAG SM–tagged KDM5B and loaded with MCP before Fab microinjection. Many mRNAs (red) can be seen. Their fluorescence does not bleed into the green channel (inset). (**B**) Three seconds after microinjection, Fabs (green) co-localize with mRNAs (arrows and inset). The site of microinjection can be seen on the right (bright green smear). (**C**) Sample FRAP experiment on a cell transfected with SM-H2B and bead-loaded with MCP and Fab 24 hours earlier. There is little recovery in 200 s (lower curve; int., intensity). Error bars show SEM. (**D**) A cartoon of results from (A) to (C) showing fast on rates and slow off rates for Fab (green Y shapes) binding to SM epitopes (triangles) as they emerge from a ribosome (circle) translating mRNA (thick line). (**E**) The intensity of a tracked FLAG SM–tagged KDM5B polysome (yellow circle and inset) can be measured as a function of time. The cartoon below shows how movement of ribosomes along mRNA and the emergence of elongating peptide chains can produce intensity fluctuations at the indicated times t_1, t_2, and t_3 (AAA, poly-A tail). (**F**) The average correlation curves calculated from intensity fluctuation data for each construct (error bars show SEM; *G*, autocorrelation function). The time at which the correlation hits zero can be obtained from fits (dashed lines) to estimate the elongation dwell time. (**G**) Calculated elongation rates (amino acids per second) from fits in (F) (error bars show 95% confidence intervals). Scale bars, 10 μm.

ribosome would have to initiate on average every 30 ± 9 s to maintain the measured 5.1 ± 0.9 ribosomes per polysome. From this we can predict that 96 ± 3% of KDM5B mRNA will be translated by polysomes, 3.1 ± 2.5% of KDM5B mRNA will be translated by a single ribosome, and 0.6 ± 0.6% will be untranslated. Moreover, using the same initiation and elongation rates for the other constructs (because they have the same 5′ and 3′ UTRs), we can predict that 35 ± 9% of β-actin mRNA will be translated by polysomes containing 2.5 ± 0.1 ribosomes on average (36 ± 2% translated by single ribosomes and 29 ± 7% untranslated), whereas just 6.5 ± 2.5% of H2B mRNA will be translated by polysomes containing 2.2 ± 0.1 ribosomes on average (27 ± 3% translated by single ribosomes and 66 ± 5% untranslated).

These predictions (detailed in the supplementary materials and summarized in figs. S10 and S11) are consistent with our earlier measurements of polysomes (19 ± 4% of β-actin mRNA containing 3.1 ± 0.5 ribosomes and 4 ± 1% of H2B mRNA containing 2.1 ± 0.4 ribosomes). They are also in agreement with independent measurements of the number of ribosomes per polysome obtained by polysome profiling (figs. S10 and S11). Although we found a lower density of ribosomes within polysomes than others have (*20*, *21*), the consistency of our live-cell and biochemical data suggests that the difference is due to mRNA variability rather than to experimental stress that might artificially lower the density. In particular, other mRNAs with different 5′ and 3′ architectures (*22*, *23*) will probably have different polysome occupancies and dynamics, depending not only on the balance of elongation and initiation but also on the metabolic status of the cell and the local environment.

Simultaneous multicolor imaging of distinct mRNAs being translated in a single cell

One advantage of using Fab to mark translation sites is the large number of high-affinity antibodies for multicolor applications. To demonstrate this,

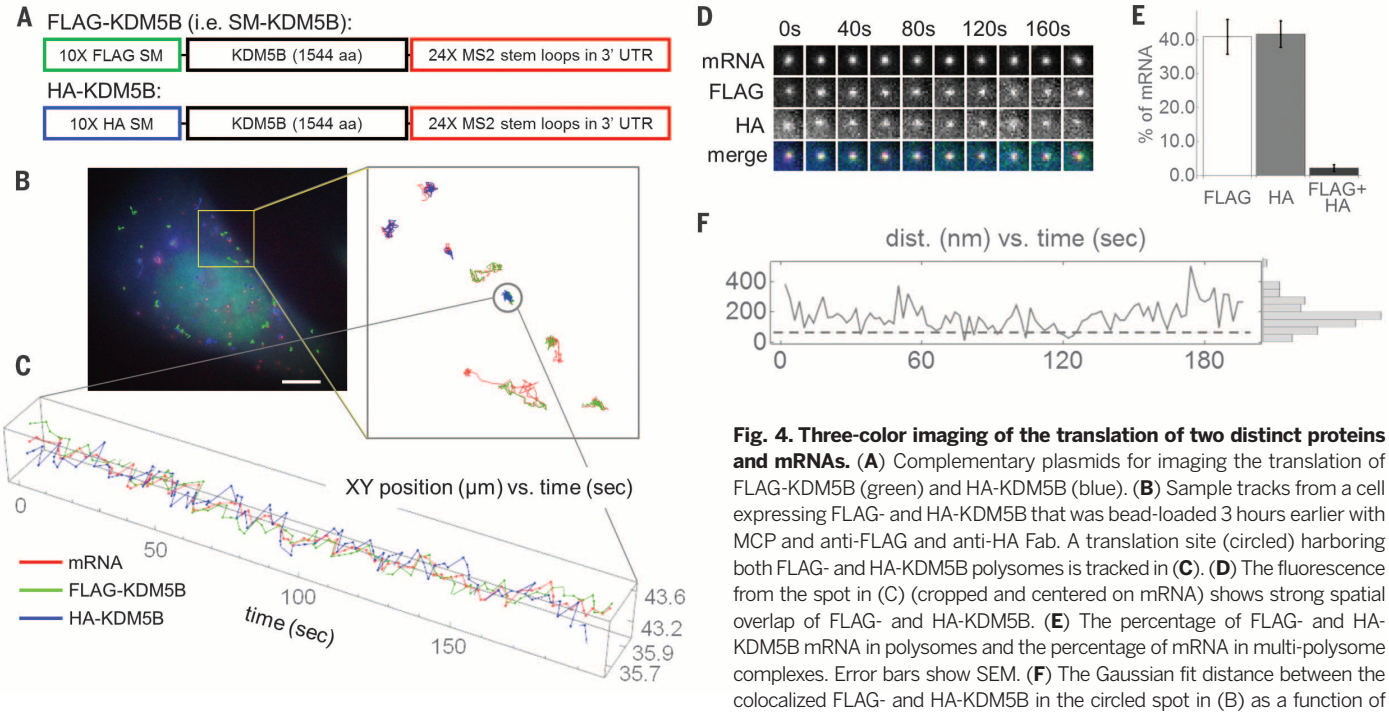


Fig. 4. Three-color imaging of the translation of two distinct proteins and mRNAs. (**A**) Complementary plasmids for imaging the translation of FLAG-KDM5B (green) and HA-KDM5B (blue). (**B**) Sample tracks from a cell expressing FLAG- and HA-KDM5B that was bead-loaded 3 hours earlier with MCP and anti-FLAG and anti-HA Fab. A translation site (circled) harboring both FLAG- and HA-KDM5B polysomes is tracked in (**C**). (**D**) The fluorescence from the spot in (C) (cropped and centered on mRNA) shows strong spatial overlap of FLAG- and HA-KDM5B. (**E**) The percentage of FLAG- and HA-KDM5B mRNA in polysomes and the percentage of mRNA in multi-polysome complexes. Error bars show SEM. (**F**) The Gaussian fit distance between the colocalized FLAG- and HA-KDM5B in the circled spot in (B) as a function of time. The distribution of these distances is shown on the right. It peaks at ~130 nm, twice the distance reported in Fig. 2H between KDM5B nascent chains and mRNA in a single polysome (~65 nm, dashed line).

we generated new Fabs from hemagglutinin (HA) antibodies and labeled these with Alexa488 dye. In parallel, we engineered a new KDM5B construct with a 10× HA SM tag (HA-KDM5B) (*5*) to complement FLAG SM–tagged KDM5B (hereafter, FLAG-KDM5B), as shown in Fig. 4A. As a first application of this technology, we wanted to test whether polysomes interact with each other to form higher-order structures that can translate two distinct mRNAs at the same time. For this, we cotransfected cells with HA- and FLAG-KDM5B and bead-loaded them with MCP and anti-HA and anti-FLAG Fab. As anticipated, cotransfected cells contained two types of polysomes in equal fractions (Fig. 4, B to E), one type labeled by anti-HA Fab (HA-KDM5B) and the other labeled by anti-FLAG Fab (FLAG-KDM5B) (Fig. 4B and movie S8). For the most part, there was little interaction between the two, providing direct evidence that the vast majority of KDM5B polysomes act independently of one another. However, a small fraction (~5%) of KDM5B polysomes formed complexes that co-moved for hundreds of seconds (Fig. 4, C and D, and fig. S12) and that produced both HA- and FLAG-tagged nascent peptide chains. By measuring the distance between the nascent HA and FLAG chains, we found the complexes to be roughly twice the size of a single polysome (Fig. 4F), suggesting that the component polysomes remain compartmentalized. These complexes could reflect a more general strategy to either assemble higher-order complexes cotranslationally (*24*) or co-regulate the expression of two genes.

This work is similar to a companion study by Wu *et al.* (*25*) that combined the SunTag (*26*) with MS2 to image single mRNA translation kinetics in live cells. Their measurements and our measurements of translation elongation and initiation rates are within a factor of 2, indicating that the general approach is reproducible between laboratories. Although the techniques are similar, the combination of anti-FLAG Fab and anti-HA Fab enables multicolor experiments that are not possible with the SunTag system. Also, the FLAG and HA epitopes (8 and 9 amino acids, respectively) are just over one-third the size of the SunTag epitope (22 amino acids), so spatio-temporal resolution can be up to three times as good and imaging is potentially less invasive. These advantages will make NCT a powerful new tool for detecting, tracking, and quantifying translation dynamics and for dissecting gene regulatory networks in vivo.

REFERENCES AND NOTES

1. G. L. Hager, J. G. McNally, T. Misteli, *Mol. Cell* **35**, 741–753 (2009).
2. X. Darzacq *et al.*, *Annu. Rev. Biophys.* **38**, 173–196 (2009).
3. J. A. Chao, Y. J. Yoon, R. H. Singer, *Cold Spring Harb. Perspect. Biol.* **4**, a012310–a012310 (2012).
4. J. M. Halstead *et al.*, *Science* **347**, 1367–1671 (2015).
5. S. Viswanathan *et al.*, *Nat. Methods* **12**, 568–576 (2015).
6. E. Bertrand *et al.*, *Mol. Cell* **2**, 437–445 (1998).
7. J. B. Grimm *et al.*, *Nat. Methods* **12**, 244–250 (2015).
8. G. Blobel, D. Sabatini, *Proc. Natl. Acad. Sci. U.S.A.* **68**, 390–394 (1971).
9. J. R. Warner, A. Rich, C. E. Hall, *Science* **138**, 1399–1403 (1962).
10. A. K. Christensen, C. M. Bourne, *Anat. Rec.* **255**, 116–129 (1999).
11. G. Viero *et al.*, *J. Cell Biol.* **208**, 581–596 (2015).
12. D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, R. H. Singer, *Science* **332**, 475–478 (2011).
13. H. Kimura, P. R. Cook, *J. Cell Biol.* **153**, 1341–1354 (2001).
14. F. Mueller, T. Morisaki, D. Mazza, J. G. McNally, *Biophys. J.* **102**, 1656–1665 (2012).
15. D. R. Larson, R. H. Singer, D. Zenklusen, *Trends Cell Biol.* **19**, 630–637 (2009).
16. J. Yu, J. Xiao, X. Ren, K. Lao, X. S. Xie, *Science* **311**, 1600–1603 (2006).
17. N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, *Science* **324**, 218–223 (2009).
18. N. T. Ingolia, L. F. Lareau, J. S. Weissman, *Cell* **147**, 789–802 (2011).
19. J.-R. Yang, X. Chen, J. Zhang, *PLOS Biol.* **12**, e1001910 (2014).
20. R. D. Palmiter, *Cell* **4**, 189–197 (1975).
21. R. Jackson, N. Standart, *RNA* **21**, 652–654 (2015).
22. G. S. Wilkie, K. S. Dickson, N. K. Gray, *Trends Biochem. Sci.* **28**, 182–188 (2003).
23. N. Sonenberg, A. G. Hinnebusch, *Cell* **136**, 731–745 (2009).
24. Y.-W. Shieh *et al.*, *Science* **350**, 678–680 (2015).
25. B. Wu, C. Eliscovich, Y. J. Yoon, R. H. Singer, *Science* **352**, 1430–1435 (2016).
26. M. E. Tanenbaum, L. A. Gilbert, L. S. Qi, J. S. Weissman, R. D. Vale, *Cell* **159**, 635–646 (2014).

ACKNOWLEDGMENTS

This work was supported by start-up funds from Colorado State University. We thank R. Cohen, F. Scavone, H. Kimura, S. DiPietro, A. Ambrosio, L. Stargell, A. Thurston, and J. Chao for help with experiments. We thank Wu *et al.* (*25*) for sharing their preliminary data and the Transcription Imaging Consortium at the Janelia Research Campus of the Howard Hughes Medical Institute for creating the environment that made our interactions and discussions possible. T.M. and T.J.S. planned, analyzed, and carried out all experiments except for microinjections. T.J.S., T.M., and K.L. wrote the manuscript. K.L. helped with plasmid construction and transfections, cell culture, analysis, Western blotting, and figures. K.F.D. and J.G.D. performed microinjections. B.P.E. provided guidance for the multicolor single-molecule microscopy. Z.Z. purified Halo-MCP. L.L.L. and S.V. provided HA and FLAG SM constructs. L.D.L. and J.B.G. provided fluorophores. T.L. provided help with experimental design.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1425/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
References (*27*–*33*)
Movies S1 to S8

17 December 2015; accepted 28 April 2016
Published online 5 May 2016
10.1126/science.aaf0899

TRANSLATION

Translation dynamics of single mRNAs in live cells and neurons

Bin Wu,[1,2] Carolina Eliscovich,[1] Young J. Yoon,[1,3] Robert H. Singer[1,2,3,4]*

Translation is the fundamental biological process converting mRNA information into proteins. Single-molecule imaging in live cells has illuminated the dynamics of RNA transcription; however, it is not yet applicable to translation. Here, we report single-molecule imaging of nascent peptides (SINAPS) to assess translation in live cells. The approach provides direct readout of initiation, elongation, and location of translation. We show that mRNAs coding for endoplasmic reticulum (ER) proteins are translated when they encounter the ER membrane. Single-molecule fluorescence recovery after photobleaching provides direct measurement of elongation speed (5 amino acids per second). In primary neurons, mRNAs are translated in proximal dendrites but repressed in distal dendrites and display "bursting" translation. This technology provides a tool with which to address the spatiotemporal translation mechanism of single mRNAs in living cells.

Genome-wide studies have shown that protein abundance is predominantly controlled at the level of translation (*1*). Translational regulation allows cells to respond rapidly to environmental cues and synthesize proteins with precise timing and at specific subcellular locations (*2*, *3*). Numerous studies have concentrated on RNA localization and its underlying mechanism (*4*). However, the translation of localized mRNA in living cells remains poorly understood because unlike transcription, a single-molecule method to directly image the process is lacking (*5*). Ensemble biochemical measurements such as ribosome profiling can provide a genome-wide measurement of translation (*6*, *7*). Approaches that measure the association of ribosomes with mRNAs are not direct readouts of translation (*8*). Pulse-labeling in cell culture [stable isotope labeling with amino acids in cell culture (SILAC)] quantifies newly synthesized proteins by means of mass spectrometry (*1*). Cellular approaches, such as fluorescent noncanonical amino acid tagging (FUNCAT) (*9*) and puromycylation (*10*), measure overall protein synthesis but are not gene-specific. Fluorescent protein–based translation assays (*11*–*14*) require lengthy maturation of fluorophores. Recently, translating RNA imaging by coat protein knock off (TRICK) (*15*) distinguished previously translated from untranslated mRNAs. A single-molecule fluorescence imaging approach would complement these cellular and biochemical methods by providing precise spatiotemporal molecular information, including variance among mRNAs. Here, we describe a strategy for directly visualizing the nascent peptide (NAP) on translating mRNA with single-molecule imaging of nascent peptides (SINAPS). It provides precise parameters and reveals unpredicted phenomena of mRNA regulation in the endoplasmic reticulum (ER) and neuronal dendrites.

[1]Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [2]Gruss-Lipper Biophotonics Center, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [3]Department of Neuroscience, Albert Einstein College of Medicine, Bronx, NY 10461, USA. [4]Janelia Research Campus, Howard Hughes Medical Institute (HHMI), Ashburn, VA 20147, USA.
*Corresponding author. Email: robert.singer@einstein.yu.edu; singerr@janelia.hhmi.org

Development of SINAPS, a single-molecule translation assay in live cells

There are two obstacles to visualize the translation site (TLS): (i) the weak signal of the nascent chain and (ii) the background due to preexisting labeled proteins. To address the weak signal, we used the recently developed SunTag technology for signal amplification (*16*), which uses a genetically encoded antibody fragment, the single-chain variable fragment (scFV) of GCN4 antibody fused to a super folder green fluorescent protein (sfGFP) (scFV-sfGFP) (*17*). When the epitope peptide for the scFV is multimerized, multiple scFV-sfGFPs bind to a single reporter protein, rendering it visible above the fluorescence background, similar to MS2 stem loops that make the mRNA visible (Fig. 1, A and B) (*18*). Because the sfGFP is already mature and fluorescent, the NAP could be detected as soon as it emerges from the ribosome (Fig. 1B). In a polysome, multiple NAPs accumulate at the site of mRNA and form a bright TLS. To address the background fluorescence, we removed completed proteins by fusing the auxin-induced degron (AID) to the C terminus of the reporter in order to regulate its stability (*19*, *20*). We inserted the protein of interest [blue fluorescent protein (oxBFP) (*21*)] between the SunTag and the AID (*20*) with a flag tag in the N terminus. The reporter mRNA included the 3′ untranslated region (3′UTR) of β-actin mRNA for cytoplasmic localization and the MS2 binding sites (MBS) for its visualization (*22*). The reporter (flag-SINAPS) was cloned into a lentiviral backbone for stable integration into the target cells (Fig. 1A).

The SINAPS reporter was expressed by viral infection in a U2OS human osteosarcoma cell line stably expressing two accessory proteins: the antibody scFV-sfGFP for labeling and the thermally stable rice Oryza Sativa transport inhibitor response 1 (OsTIR1) for degrading AID-containing protein (*19*, *23*). A protein of the correct size was synthesized but became undetectable with Western blot in presence of 500 μM auxin Indole-3-acetic acid (IAA) (fig. S1). We performed single-molecule fluorescence in situ hybridization (smFISH) to detect mRNA and immunofluorescence (IF) against GFP to measure the protein (Fig. 1, C and D). The bright spots in the IF channel (green) colocalized with mRNA spots (red). The numerous dim green spots were single proteins labeled with SunTag (Fig. 1D, white arrow heads). They were used to convert the TLS intensity into the number of NAPs (maximum, 14; mean, 4.5) (Fig. 1E). We used puromycin, a tRNA analog that dissociates NAPs from the ribosomes, to verify that TLS disappeared (Fig. 1F and fig. S2, A and B); its removal (Puro/Wash) restored the TLS to steady-state levels (Fig. 1F and fig. S2, C to E). Low concentrations of the elongation inhibitor cycloheximide (CHX) increased NAPs on mRNAs (Fig. 1F and fig. S2F). We inserted varying lengths of the protein domains between SunTag and the AID to determine whether the number of NAPs scaled linearly with the length of the coding region (Fig. 1G and fig. S2G). A linear fit described the data with a slope of 0.0044 per amino acid (Fig. 1H). With the translation elongation speed determined as v = 4.7 amino acids/s, this reveals an average initiation rate of 1.3 per min (supplementary materials, Eq. 4).

Diffusion dynamics of translating RNA

To describe the dynamics of translation, we performed live cell imaging. We labeled the mRNA with MS2 coat protein (MCP) fused with Halotag (*24*) and applied Halo-JF549 dyes for its superior brightness and photostability (*23*, *25*). We imaged mRNAs and TLS simultaneously using their distinct fluorescent tags (stdMCP-Halo-JF549 and scFV-sfGFP) (*26*) on two precisely aligned cameras (*27*). We observed bright TLS moving together with mRNAs (Fig. 2, A to D, and movies S1 and S2). As a control, in puromycin-treated cells there were no TLSs detected (movie S3). The brightness of the mRNA and TLS allowed accurate determination of their positions with Matlab software (*28*). We tracked the positions of mRNA and TLS (*29*) and identified TLS moving together with mRNAs (Fig. 2E) (*23*). We classified the tracks of mRNAs into two categories: confined and freely diffusing. For each category, we determined the fraction of mRNA tracks with TLS in each cell and defined them as translating mRNAs. Translation was independent of whether the mRNAs were confined or freely diffusing (Fig. 2F). For freely diffusing mRNAs, the diffusion coefficients (D) were measured through mean-square displacements. The translating mRNAs diffused slower on average but overlapped in their distribution with untranslating ones (Fig. 2G). The D of translating mRNA was weakly anti-correlated with TLS intensity, which is consistent with larger polysomes slowing mRNA mobility (Fig. 2H). Therefore, the mRNA mobility correlates with its translation level on average but is not an exact predictor of translation status of individual

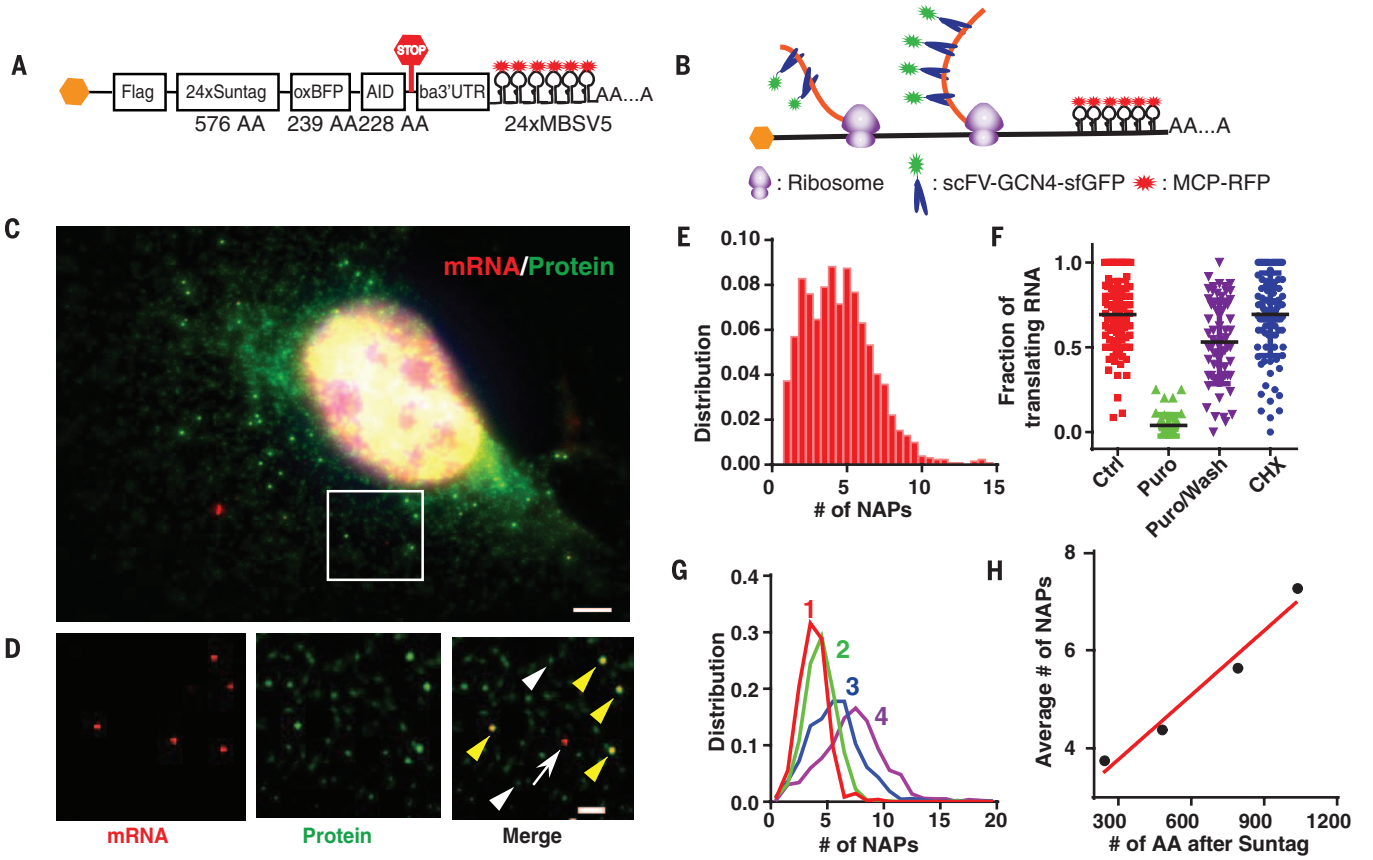


Fig. 1. System for single-molecule imaging of nascent peptides in live cells. (**A**) Schematic of the SINAPS construct. Flag, flag tag; AID, auxin induced degron; ba3′UTR, β-actin 3′ untranslated region; AA, amino acid. (**B**) Schematic of SINAPS. scFV-sfGFP binds to and labels NAPs containing SunTag epitope emerging from ribosome. Polysome assembling on mRNA results in multiple NAPs. (**C**) smFISH and IF experiments on flag-SINAPS constructs. Green, IF against GFP; red, smFISH against mRNA. The bright green spots colocalizing with red mRNAs are TLSs. Scale bar, 5 μm. (**D**) The deconvolved image of the box in (C). (Left) smFISH. (Middle) IF. (Right) Merged. Yellow arrowheads indicate TLS; white arrowheads indicate single flag-SINAPS protein; and the white arrow indicates untranslating mRNA. Scale bar, 2 μm. (**E**) The integrated intensity of TLS is normalized with that of single proteins, giving the number of NAPs (full-length equivalent). (**F**) The fraction of translating mRNA under treatments. Each symbol represents a single cell. Ctrl, control; Puro, 100 μg/mL puromycin for 10 min; Puro/Wash, 100 μg/mL puromycin treatment for 10 min, then incubation in normal medium for 20 min; CHX, 2 μg/mL CHX treatment for 30 min. (**G**) The number of NAPs for different constructs at steady state. The longer the mRNA, the more NAPs at the TLS. (**H**) The mean number of NAPs scales linearly with the length of the coding region. Red line describes a linear fit with slope 0.0044 per amino acid.

mRNAs coding for cytosolic proteins, which is different from membrane-targeted mRNAs.

Local translation on the ER

Local protein synthesis on the ER and other membrane-bound cellular compartments plays a critical role in protein trafficking to establish and maintain membrane function (*30*). Introducing an ER-targeting signal in the N terminus of the SINAPS reporter will target the reporter protein to the ER (Fig. 3A). Because the scFV antibody is cytoplasmic, the SunTag motif needs to be exposed to cytosol to be labeled (Fig. 3B). We fused the first 29 amino acids of cytochrome p450 to the N terminus of the reporter [cytoplasmic end of an ER signal-anchor membrane protein (CytERM)] (*21*). CytERM-SINAPS labeled the tubular ER structure when expressed in U2OS cells stably infected with scFV-sfGFP and OsTIR1 (fig. S3, A and B). At low expression, individual CytERM-SINAPS proteins were visible owing to their slower mobility on the membrane (fig. S3, C to E, and movie S4). Unlike mRNAs coding for cytoplasmic proteins, the majority of CytERM-SINAPS mRNAs showed confined motion (Fig. 3, C to G, and movies S5 to S7). Consistent with this, all mRNAs colocalizing with TLSs were confined and showed low mobility (Fig. 3, H to J). Confined mRNAs without bright translation sites might have recently initiated translation. In contrast, none of the freely diffusing mRNAs moved together with TLS (Fig. 3, C to J). This confirmed that the reporter proteins were inserted into the ER cotranslationally, and the mRNAs were tethered to the ER by NAPs. To further support this interpretation, when treated with puromycin the mRNAs immediately dissociated from the ER (within 2 min) and became freely diffusing particles without TLS (Fig. 3, G and K, and movies S5 and S8). These experiments demonstrate that the mRNAs are tethered to the ER only when translated.

Translation kinetics in live cells

Because the translating CytERM-SINAPS mRNA were confined and could be tracked for many minutes, we performed fluorescence recovery after photobleaching (FRAP) experiments on single TLSs (Fig. 4A). We applied a two-second pulse of diffraction-limited 491-nm laser light to bleach the selected single TLS (*23*). The integrated fluorescence intensity was monitored while tracking the TLS (movies S9 and S10). The fluorescence recovered over 90% of the pre-bleach value within 4 min (Fig. 4, B and E). As a control, neighboring TLS not bleached showed consistent signal throughout the experiment (Fig. 4C). In contrast, the fluorescence only recovered to ~20% of pre-bleach value when ribosomes were stalled by high concentrations of CHX (Fig. 4, D and E). To derive the recovery curve, we assumed that the scFV binds to each epitope segment with equal probability at the TLS or on mature proteins. We modeled the translation by ribosomes as a stochastic process, with constant initiation and elongation rates. The normalized recovery curve depended only on one parameter: the elongation rate of ribosomes [the gene length is a constant for a given reporter (supplementary text)]. The

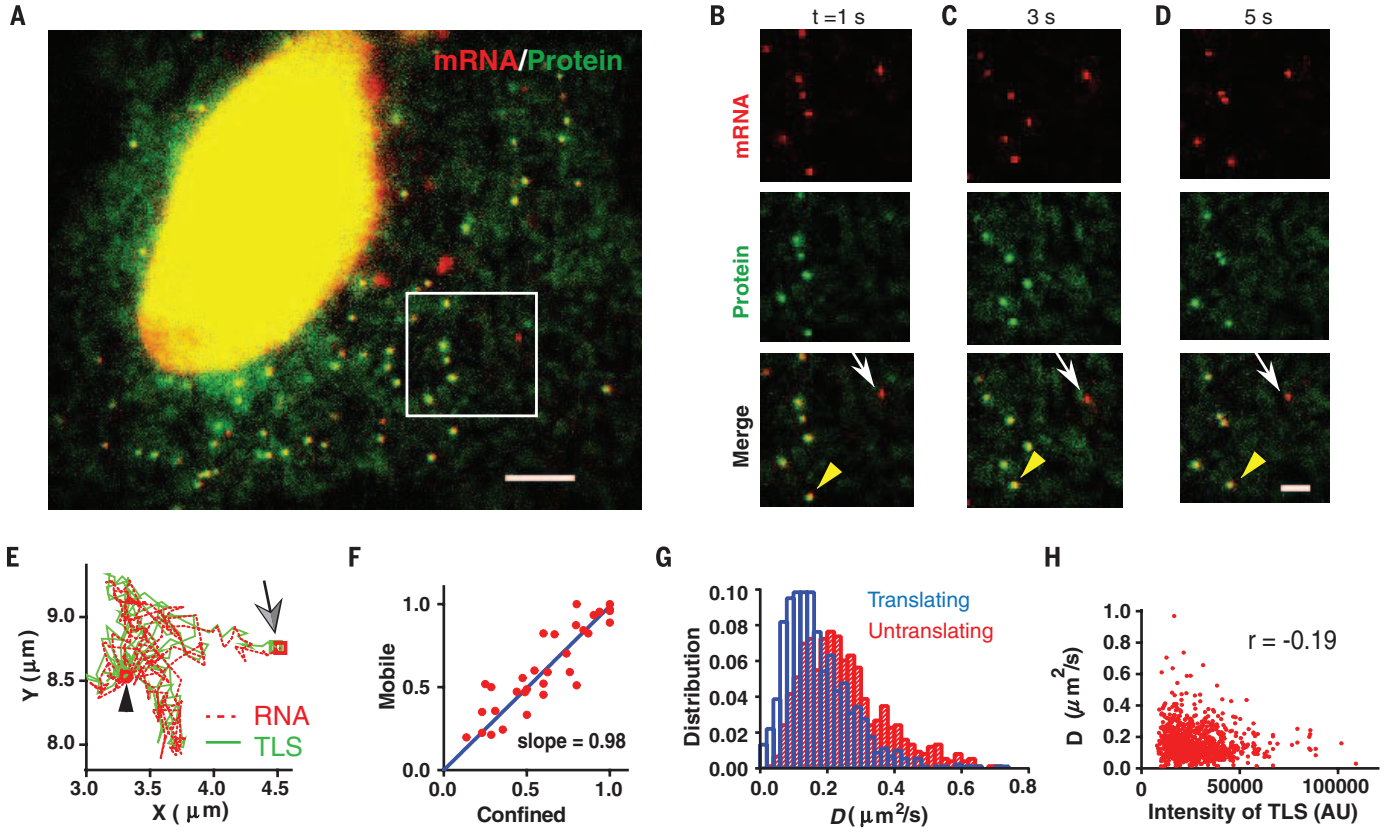


Fig. 2. Dynamics of mRNAs in translation. (**A**) Snapshot of live cell imaging of cells stably expressing flag-SINAPS (movie S1). Green, scFV-sfGFP; red, stdMCP-Halotag-JF549. Scale bar, 5 μm. (**B** to **D**) Montage of movie S2 at time (*t*) = 1, 3 and 5 s. (Top) mRNA. (Middle) TLS and free proteins. (Bottom) Merged. Yellow arrowheads inidcate TLS, and white arrows indicate untranslating mRNAs. Scale bar, 2 μm. (**E**) The mRNA (red) and TLS (green) move together. Arrow (square) indicates the start of the track. The black arrowhead (circle) indicates the end of the track. (**F**) The motion of mRNA is classified either as confined or mobile. The fraction of translating mRNA in each category is almost identical in each cell. Each symbol represents a cell. (**G**) The histogram of diffusion coefficients (*D*) of freely diffusing mRNAs. Blue, translating mRNA; red, untranslating mRNA. (**H**) The diffusion coefficient of the translating mRNA is only weakly anticorrelated with the integrated intensity of TLS; correlation coefficient (*r*) = −0.19.

modeling correctly predicted the FRAP curve with Monte Carlo simulation (fig. S4). Fitting the theoretical curve to the experimentally observed data yielded the average translation elongation rate v = 4.7 ± 0.6 amino acids/s (Fig. 4E). The translation rate was in agreement with previous measurements by using ribosome profiling (v = 5.6 amino acids/s) (*31*). Experiments on mRNAs with different lengths of coding regions yielded similar translation rates (v = 5.6 ± 1.4 amino acids/s and v = 5.0 ± 1.1 amino acids/s) (fig. S5). These results indicate that the SINAPS reporter mRNAs undergo constitutive translation in U2OS cells.

Translation in neurons

We applied SINAPS to study local translation in neurons, where local regulation of translation is particularly important owing to polarized neuronal processes (*32*). We combined the OsTIR1 and scFV-sfGFP into a polycistronic plasmid with an internal ribosome entry site (IRES) (*23*). We expressed flag-SINAPS and OsTIR1-IRES-scFV-sfGFP in cultured primary hippocampal neurons by dual lentiviral infection. We performed smFISH and IF experiments to investigate the translation of individual mRNAs. Bright puncta in the SunTag channel colocalizing with the mRNA indicated that the mRNA was being translated (Fig. 5, A to C, and fig. S6). Although the numbers of NAPs on mRNAs in primary neurons and glia were similar to that in U2OS (Fig. 1E and fig. S6, B and E), the fraction of translating mRNAs was less (Figs. 1G and 5D), suggesting a more stringent regulation of mRNA translatability in primary cells. In addition, translation in neurons was spatially modulated. In proximal dendrites (<30 μm from soma), the percentage of translating mRNAs (~40%) was similar to that of the glial cells cultured in the same dish. However, the percentage gradually decreased as a function of distance to the soma (Fig. 5E), as low as ~10% for distal dendrites (>100 μm) (Fig. 5D). This suggests translation repression in distal dendrites (*33*).

To investigate the dynamics of translation, we imaged the SINAPS reporter in live neurons. We observed bright green particles, mostly in proximal dendrites, but some also distributed in distal dendrites (movie S11). They were much brighter than the fast diffusing single proteins (visible at high power). These were bona fide TLSs because puromycin treatment abolished them within 2 min (fig. S7). We focused on RNA in the distal dendrites because they showed dynamic behavior and were sparsely distributed. We tracked the TLS for 2 hours. The position and integrated intensity of TLS were determined with a tracking algorithm in order to generate the spatial trajectory and intensity trace of the sites (*23*). TLSs were mobile in a time scale of minutes (Fig. 6A and movies S11 to S13). The mRNA sampled several anchoring sites in dendrites while undergoing active translation. This is in contrast to the hypothesis that the RNA was translationally repressed during transport (*34*). The TLS intensity over time encoded the translation dynamics of the mRNA (Fig. 6B). To extract the kinetic parameters from these observations, it was necessary to deconvolve the effect of multiple ribosomes generating NAPs. We adopted a fluctuation analysis approach that was developed for transcription (*5*). The translation dynamics was described by an autocorrelation function $G(\tau)$ (Fig. 6C and supplementary text). We observed that many translation events occurred in a bursting fashion (movies S11 to S13): The translation was "on" or "off" with interspersed long periods of no translation at all. We modeled translation as the random telegraph

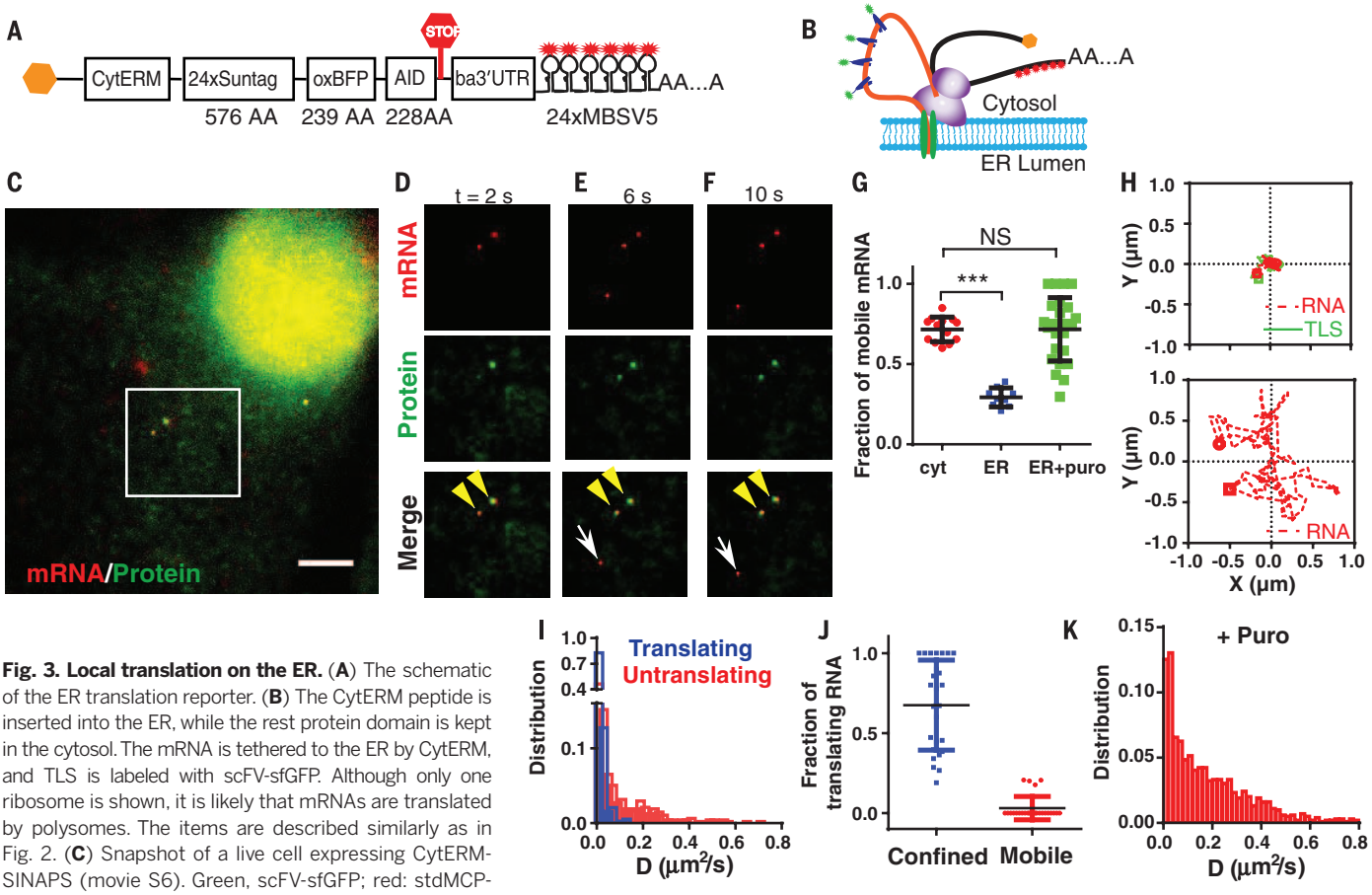


Fig. 3. Local translation on the ER. (**A**) The schematic of the ER translation reporter. (**B**) The CytERM peptide is inserted into the ER, while the rest protein domain is kept in the cytosol. The mRNA is tethered to the ER by CytERM, and TLS is labeled with scFV-sfGFP. Although only one ribosome is shown, it is likely that mRNAs are translated by polysomes. The items are described similarly as in Fig. 2. (**C**) Snapshot of a live cell expressing CytERM-SINAPS (movie S6). Green, scFV-sfGFP; red: stdMCP-Halotag-JF549. Scale bar, 5 μm. (**D** to **F**) Montage of movie S7 at t = 2, 6 and 10 s. (Top) mRNAs. (Middle) TLS and free proteins. (Bottom) Merged. Yellow arrowheads indicate TLS; white arrows indicate untranslating mRNAs. Scale bar, 2 μm. (**G**) The fraction of mobile mRNAs for cytoplasmic mRNAs (cyt), Flag-SINAPS; the ER-targeted mRNAs (ER), CytERM-SINAPS; and CytERM-SINAPS mRNAs in the presence of 100 μg/mL puromycin (ER+Puro). Unpaired t test, ***$P < 0.001$. NS, not significant. See also movie S5. (**H**) (Top) A translating mRNA in (C) to (F) shows a confined motion. (Bottom) An untranslating mRNA is freely diffusing. (**I**) Most translating CytERM-SINAPS mRNAs were confined with very small diffusion coefficient (blue). The untranslating mRNA had higher diffusion coefficient (red). (**J**) The confined mRNAs had a significantly higher fraction in translation (average 69%) than that of freely diffusing ones (average 3%; all fast moving mRNAs are not translating). (**K**) The diffusion coefficient of CytERM-SINAPS mRNA when treated with 100 μg/mL puromycin. mRNAs were released from the ER and freely diffuse.

model used for describing transcription kinetics. When the RNA is on, the translation is described by a stochastic process, with constant initiation and elongation rate. The autocorrelation function for this scenario has been worked out previously for modeling transcription (*5*). The autocorrelation function depends on the initiation rate c and a total dwell time T of NAP at TLS determined by elongation speed and the length of the coding region. When the mRNA is in the off state, there is no translation. The two states interchange randomly and have lifetimes τ_{on} and τ_{off}, respectively. We further assumed that translations were correlated in the same on state but statistically independent between successive on states. Fitting the theoretical function to the experimental data yielded a dwell time of an individual NAP (T = 170 ± 50 s, which is consistent with the length of the reporter) and an average translation initiation rate of c = 2.1/min. The average lifetime of the on state was τ_{on} = 13 min (Fig. 6C). Alternatively, we measured the width of the translation bursts directly from intensity traces (only from the full off-on-off intensity traces), which yielded the average lifetime of bursts τ_{on} = 17 min, agreeing with the fluctuation analysis (Fig. 6D). About 20% of mRNAs underwent constitutive translation during the 2 hours of imaging (Fig. 6, E to G, and movie S14). By singling out those TLSs and analyzing them separately, we obtained a similar dwell time (T = 164 ± 24 s) but a much longer τ_{on} > 120 min, which is consistent with a constitutive translation kinetics for these mRNAs. Therefore, the elongation speed was independent of whether translation was bursting or constitutive.

Discussion

We studied translation dynamics of single mRNAs in live cells by directly imaging the "newly born" nascent peptide emerging from the ribosome. This was possible because the scFV-GCN4 had high affinity to its epitope [the dissociation constant K_d is less than 4.4×10^{-11} M for wild-type scFV-GCN4 (*35*)]. FRAP experiments of scFV-GCN4 bound to mitochondria showed that the off time was in the 10-min range (*16*). Therefore, the on rate (k_{on}) for the scFV-GCN4 is fast (~10^7/M/s). The concentration of scFV-GCN4 in the cytoplasm is typically >100 nM. Therefore, the time between the NAP segment appearing from the ribosome and being bound by scFV is in the sub-second range. Each epitope peptide has 24 amino acids and takes ~5 s to synthesize. Therefore, the binding of the scFV is fast enough to track translation kinetics. The measured translation speed is consistent with the literature (*31*), further demonstrating the validity of the technology and that the binding of the scFV does not interfere with the translation kinetics.

Our results using single-particle tracking of mRNA and TLS support the widespread view of translation dynamics on the ER. For secretory and membrane proteins, their mRNAs show the following characteristics: (i) Translating mRNAs were tethered to the ER and show confined motion (Fig. 3, G to J). The translation-dependent tethering is universal for all signal peptides we have tested (CytERM, prolactin, and immunoglobulin heavy chain). (ii) mRNAs were translated

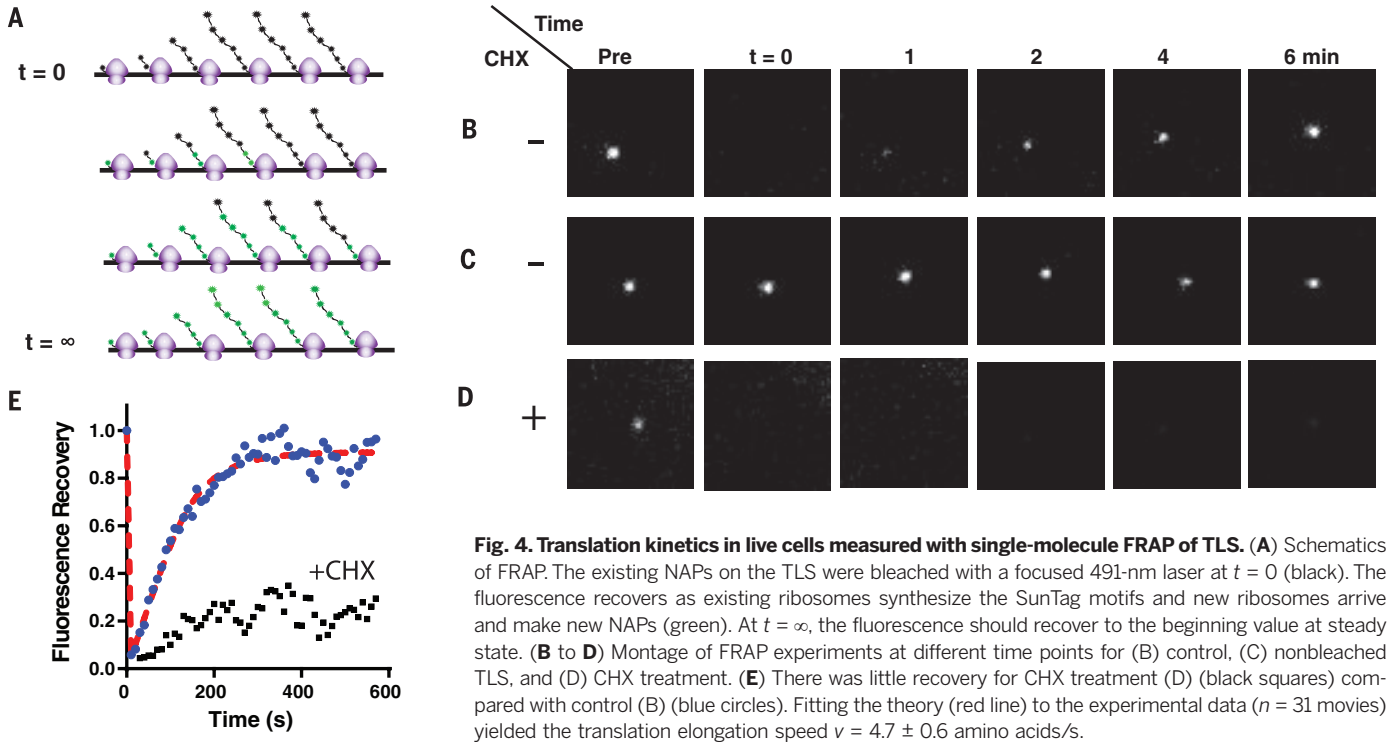


Fig. 4. Translation kinetics in live cells measured with single-molecule FRAP of TLS. (**A**) Schematics of FRAP. The existing NAPs on the TLS were bleached with a focused 491-nm laser at $t = 0$ (black). The fluorescence recovers as existing ribosomes synthesize the SunTag motifs and new ribosomes arrive and make new NAPs (green). At $t = \infty$, the fluorescence should recover to the beginning value at steady state. (**B** to **D**) Montage of FRAP experiments at different time points for (B) control, (C) nonbleached TLS, and (D) CHX treatment. (**E**) There was little recovery for CHX treatment (D) (black squares) compared with control (B) (blue circles). Fitting the theory (red line) to the experimental data ($n = 31$ movies) yielded the translation elongation speed $v = 4.7 \pm 0.6$ amino acids/s.

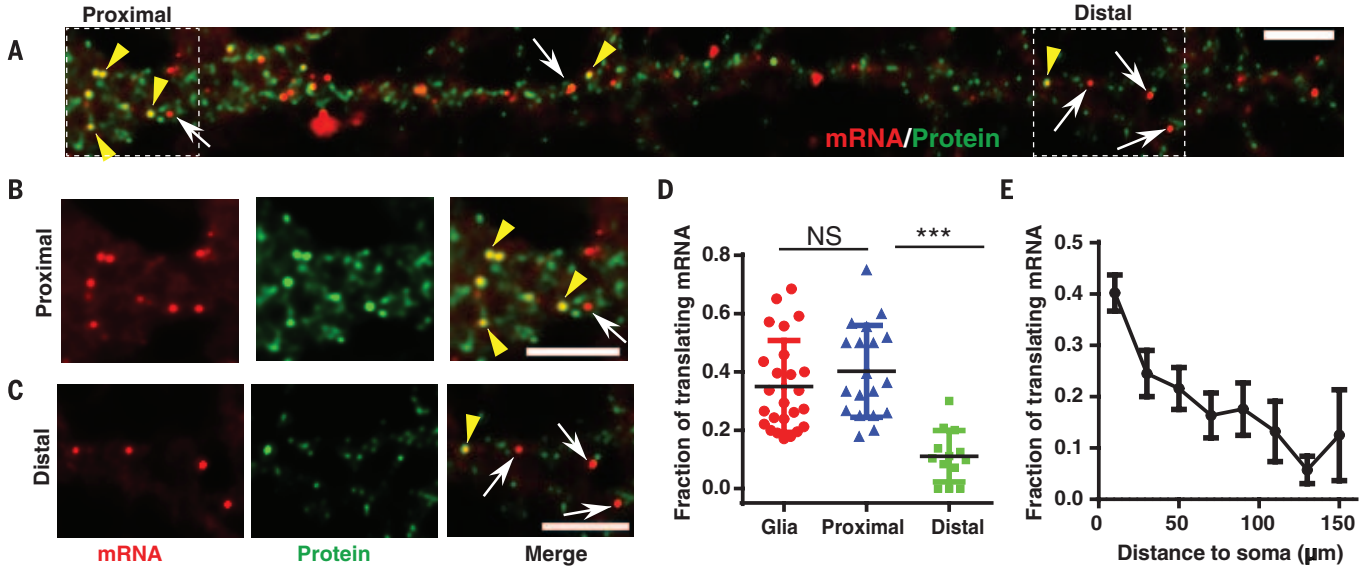


Fig. 5. Spatial distribution of translation sites in neurons. (**A**) smFISH and IF experiment. Flag-SINAPS was coexpressed with OsTIR1-IRES-scFV-sfGFP in dissociated primary hippocampal neuron with lentiviral infection at 7 days in vitro. The neurons were fixed at 14 to 21 days in vitro, and smFISH and IF experiments were performed. The dendrites of the neuron were straightened with ImageJ. The whole neuron is shown in fig. S6. Red, mRNA; green, TLS and free protein. Scale bar, 5 μm. (**B** and **C**) Segments in the (B) proximal and (C) distal dendrite were enlarged. (Left) mRNA. (Middle) TLS and free protein. (Right) Merged. Yellow arrowheads indicate TLS; white arrows indicate untranslating mRNAs. Scale bar, 5 μm. (**D**) The fraction of translating mRNA was similar for proximal dendrites (<30 μm to soma) and glial cells in the same culture dish (fig. S6), but significantly less in distal dendrites (>100 μm from soma). Unpaired t test, $^{***}P < 0.001$. (**E**) The fraction of translating mRNA in dendrite as a function of distance to soma (53 dendrites, 19 neurons).

multiple rounds on the ER because mRNA were tethered to the same location much longer than the time needed for a single round of translation. (iii) Multiple ribosomes translated the same mRNA tethered to the ER at the same time because TLS was much brighter than a single protein. And (iv) the tethering of the mRNA to the ER depended on the NAP because puromycin treatment releases mRNA from restricted motion to freely diffusing motion (Fig. 3, G and K). Therefore, our data are consistent with a "cotranslational translocation" model (*7*, *36*) that the mRNA that contains ER-targeting signal peptide was tethered to the ER by the NAP during the initial translation. The mRNA remained on the ER because ribosomes continually initiated translation. For cytosolic proteins, their mRNAs were mainly freely diffusing. Therefore, they were mostly translated by cytosolic ribosomes, not by the ER-bound ribosomes (*37*, *38*).

In neurons, it has been proposed that localizing mRNAs are packaged into granules and transported

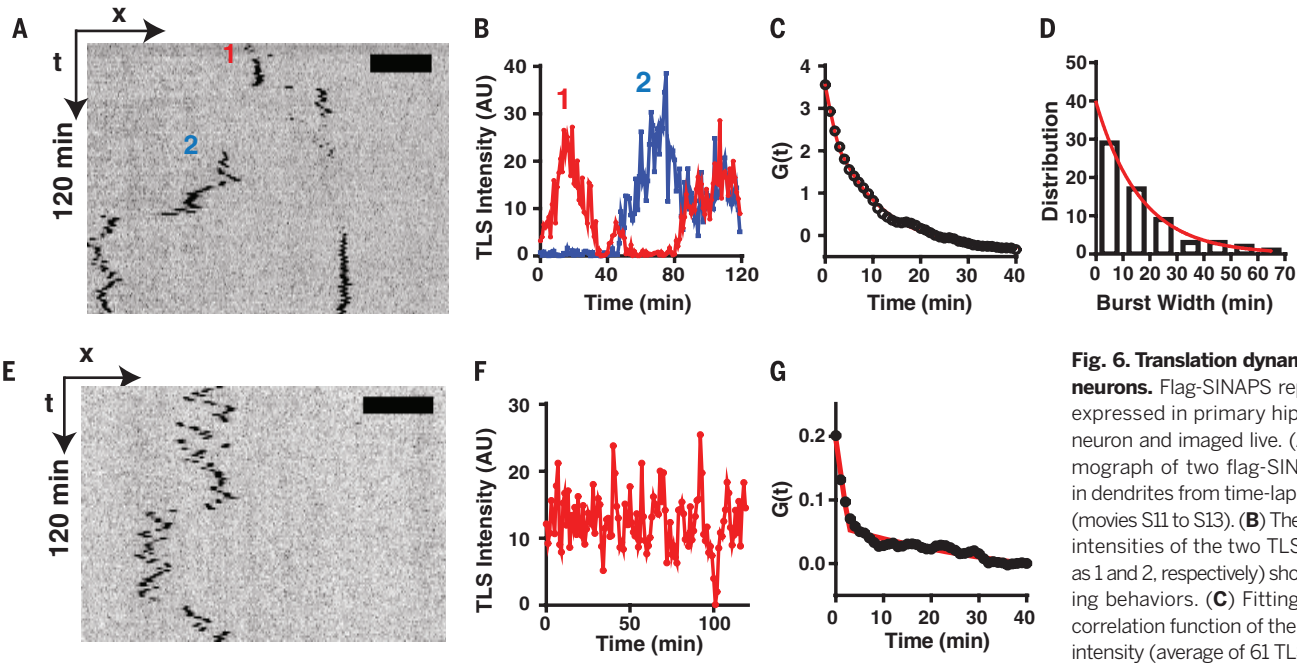


Fig. 6. Translation dynamics in live neurons. Flag-SINAPS reporter was expressed in primary hippocampal neuron and imaged live. (**A**) The kymograph of two flag-SINAPS TLSs in dendrites from time-lapse imaging (movies S11 to S13). (**B**) The integrated intensities of the two TLSs (labeled as 1 and 2, respectively) showing bursting behaviors. (**C**) Fitting the autocorrelation function of the integrated intensity (average of 61 TLSs) yielded the residence time $T = 170 \pm 50$ s, initiation rate = 2.1/min, and the average length of translation bursts $\tau_{on} = 13$ min. (**D**) The translation burst size was directly measured for tracks showing complete off-on-off cycles. Exponential fit of the histogram yielded the average length of the burst $\tau_{on} = 17$ min. (**E**) Kymograph, (**F**) integrated intensity trace, and (**G**) autocorrelation function of constitutive TLS (defined as translating more than 90% of the time during the 2-hour imaging window, average of 13 TLSs). Fitting of correlation function yielded $T = 164 \pm 24$ s, initiation rate = 2.9/min. The length of translation burst was τ_{on} >120 min, which is consistent with a constitutive translation.

long distances (*33*). It is hypothesized that mRNAs are suppressed for translation during this transport (*34*). By visualizing translation of single mRNAs, we demonstrated that mRNAs with the β-actin 3′UTR are actively translated in proximal dendrites within 30 μm of the soma. Many fewer mRNAs in distal dendrites are actively translating, possibly in a repressed granule state (*33*). These mRNAs show a bursting translation behavior, similar to that seen for nascent mRNA during transcription (*5*). In addition, we found that mRNA can scan while actively translating and possibly provide newly synthesized proteins immediately for visited spines. This single-molecule methodology provides a powerful tool with which to determine the mechanism of local translation of mRNA and particularly its relationship to local neuronal activity.

The SINAPS technology provides a general approach to measure translation kinetics of single mRNAs. Because all components of the technology are genetically encoded, it can elucidate translation regulation and dynamics with unprecedented spatiotemporal precision without perturbation directly in live cells or animals. For low-abundance proteins, the degron in the reporter is not needed. Further development will be needed to study translation of endogenous genes. This work is complementary to a companion manuscript (*39*) that demonstrates single-molecule translation by the use of a multi-epitope reporter paired with exogenous fragment antigen-binding (FAB) antibody to detect nascent chains. The measurements for initiation and elongation agree within a factor of two, indicating that the general approach is reproducible between laboratories and different cell types.

REFERENCES AND NOTES

1. B. Schwanhäusser *et al.*, *Nature* **473**, 337–342 (2011).
2. N. Sonenberg, A. G. Hinnebusch, *Cell* **136**, 731–745 (2009).
3. C. E. Holt, E. M. Schuman, *Neuron* **80**, 648–657 (2013).
4. A. R. Buxbaum, G. Haimovich, R. H. Singer, *Nat. Rev. Mol. Cell Biol.* **16**, 95–109 (2015).
5. D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, R. H. Singer, *Science* **332**, 475–478 (2011).
6. N. T. Ingolia, S. Ghaemmaghami, J. R. Newman, J. S. Weissman, *Science* **324**, 218–223 (2009).
7. C. H. Jan, C. C. Williams, J. S. Weissman, *Science* **346**, 1257521 (2014).
8. B. Wu, A. R. Buxbaum, Z. B. Katz, Y. J. Yoon, R. H. Singer, *Cell* **162**, 211–220 (2015).
9. D. C. Dieterich *et al.*, *Nat. Neurosci.* **13**, 897–905 (2010).
10. S. R. Starck, H. M. Green, J. Alberola-Ila, R. W. Roberts, *Chem. Biol.* **11**, 999–1008 (2004).
11. G. Aakalu, W. B. Smith, N. Nguyen, C. Jiang, E. M. Schuman, *Neuron* **30**, 489–502 (2001).
12. D. O. Wang *et al.*, *Science* **324**, 1536–1540 (2009).
13. M. T. Butko *et al.*, *Nat. Neurosci.* **15**, 1742–1751 (2012).
14. Y. Taniguchi *et al.*, *Science* **329**, 533–538 (2010).
15. J. M. Halstead *et al.*, *Science* **347**, 1367–1671 (2015).
16. M. E. Tanenbaum, L. A. Gilbert, L. S. Qi, J. S. Weissman, R. D. Vale, *Cell* **159**, 635–646 (2014).
17. J. D. Pédelacq, S. Cabantous, T. Tran, T. C. Terwilliger, G. S. Waldo, *Nat. Biotechnol.* **24**, 79–88 (2006).
18. E. Bertrand *et al.*, *Mol. Cell* **2**, 437–445 (1998).
19. A. J. Holland, D. Fachinetti, J. S. Han, D. W. Cleveland, *Proc. Natl. Acad. Sci. U.S.A.* **109**, E3350–E3357 (2012).
20. K. Nishimura, T. Fukagawa, H. Takisawa, T. Kakimoto, M. Kanemaki, *Nat. Methods* **6**, 917–922 (2009).
21. L. M. Costantini, M. Fossati, M. Francolini, E. L. Snapp, *Traffic* **13**, 643–649 (2012).
22. B. Wu *et al.*, *Genes Dev.* **29**, 876–886 (2015).
23. Materials and methods are available as supplementary materials on *Science* Online.
24. G. V. Los *et al.*, *ACS Chem. Biol.* **3**, 373–382 (2008).
25. J. B. Grimm *et al.*, *Nat. Methods* **12**, 244–250, 3, 250 (2015).
26. B. Wu, J. A. Chao, R. H. Singer, *Biophys. J.* **102**, 2936–2944 (2012).
27. D. Grünwald, R. H. Singer, *Nature* **467**, 604–607 (2010).
28. T. Lionnet *et al.*, *Nat. Methods* **8**, 165–170 (2011).
29. K. Jaqaman *et al.*, *Nat. Methods* **5**, 695–702 (2008).
30. T. A. Rapoport, *Nature* **450**, 663–669 (2007).
31. N. T. Ingolia, L. F. Lareau, J. S. Weissman, *Cell* **147**, 789–802 (2011).
32. H. Jung, C. G. Gkogkas, N. Sonenberg, C. E. Holt, *Cell* **157**, 26–40 (2014).
33. A. R. Buxbaum, B. Wu, R. H. Singer, *Science* **343**, 419–422 (2014).
34. M. Doyle, M. A. Kiebler, *EMBO J.* **30**, 3540–3552 (2011).
35. A. Wörn *et al.*, *J. Biol. Chem.* **275**, 2795–2803 (2000).
36. C. H. Jan, C. C. Williams, J. S. Weissman, *Science* **348**, 1217 (2015).
37. D. W. Reid, C. V. Nicchitta, *Nat. Rev. Mol. Cell Biol.* **16**, 221–231 (2015).
38. D. W. Reid, C. V. Nicchitta, *Science* **348**, 1217 (2015).
39. T. Morisaki *et al.*, *Science* **352**, 1425–1429 (2016).

ACKNOWLEDGMENTS

We thank X. Meng for cloning some of the constructs and Western blotting, E. L. Snapp for providing the oxBFP plasmid and the sequence of CytERM, E. Tutucci for constructive discussion and suggestions, and L. Lavis for providing JF dyes Halotag. The authors thank T. Morisaki *et al.* for sharing their preliminary data and the Transcription Imaging Consortium at the Janelia Research Campus of the HHMI for creating the environment for our interactions and discussions. This work was supported by NIH grant NS083085 to R.H.S.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1430/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
References (*40–42*)
Movies S1 to S14

17 December 2015; accepted 28 April 2016
10.1126/science.aaf1084

AGING

NAD+ repletion improves mitochondrial and stem cell function and enhances life span in mice

Hongbo Zhang,[1] Dongryeol Ryu,[1] Yibo Wu,[2] Karim Gariani,[1] Xu Wang,[1] Peiling Luan,[1] Davide D'Amico,[1] Eduardo R. Ropelle,[1,3] Matthias P. Lutolf,[4] Ruedi Aebersold,[2,5] Kristina Schoonjans,[6] Keir J. Menzies,[1,7]* Johan Auwerx[1]*

Adult stem cells (SCs) are essential for tissue maintenance and regeneration yet are susceptible to senescence during aging. We demonstrate the importance of the amount of the oxidized form of cellular nicotinamide adenine dinucleotide (NAD^+) and its effect on mitochondrial activity as a pivotal switch to modulate muscle SC (MuSC) senescence. Treatment with the NAD^+ precursor nicotinamide riboside (NR) induced the mitochondrial unfolded protein response and synthesis of prohibitin proteins, and this rejuvenated MuSCs in aged mice. NR also prevented MuSC senescence in the *mdx* (C57BL/10ScSn-*Dmd*^{mdx}/J) mouse model of muscular dystrophy. We furthermore demonstrate that NR delays senescence of neural SCs and melanocyte SCs and increases mouse life span. Strategies that conserve cellular NAD^+ may reprogram dysfunctional SCs and improve life span in mammals.

In adults, tissue homeostasis is highly dependent on stem cell (SC) function. Adult SCs are not only essential in continuously proliferating tissues (like the hematopoietic, intestinal, and skin systems) but also in normally quiescent tissues (such as skeletal muscle and the brain) that require regeneration after damage or exposure to disease (*1*). Aging is accompanied by a decline in adult SC function, termed SC senescence, which leads to the loss of tissue homeostasis and regenerative capacity (*2*, *3*).

Homeostasis and regeneration of skeletal muscle depend on normally quiescent muscle SCs (MuSCs), which are activated upon muscle damage to expand and give rise to differentiated myogenic cells that regenerate damaged muscle fibers (*4*, *5*). These responses are blunted in aged muscle, probably because of the reduced number and function of MuSCs (*6*–*8*). In aging, MuSC dysfunction may be caused by extrinsic signals (*9*, *10*), intrinsic cellular senescence signaling pathways (*11*), or both. One general regulator of cellular senescence, cyclin-dependent kinase inhibitor 2A (CDKN2A, also known as $p16^{INK4A}$), shows increased expression in geriatric MuSCs, which causes permanent cell cycle withdrawal and senescence of MuSCs in very old mice (28 to 32 months of age) (*11*). However, reductions in MuSC number and function can already be observed before this stage (*6*, *11*), indicating that MuSC senescence may be initiated at an earlier time point. Pre-geriatric mice, approximately 2 years old, can exhibit features of MuSC senescence (*8*, *12*–*15*). However, the early mechanisms that instigate MuSC senescence are still largely unknown.

One of the hallmarks of organismal aging is the appearance of mitochondrial dysfunction (*2*, *3*). Induced by calorie-dense diets or aging, mitochondrial dysfunction can result from depletion of the oxidized form of nicotinamide adenine dinucleotide (NAD^+), whereas NAD^+ repletion, with precursors such as nicotinamide riboside (NR), can reverse this process (*16*–*20*). SCs are thought to rely predominantly on glycolysis for energy, a process that would reduce cellular concentrations of NAD^+ (*21*). Mitochondrial function is linked to SC maintenance and activation (*22*–*25*), yet its role in senescence is unknown.

Mitochondrial dysfunction is a biomarker of MuSC senescence

To identify the role of mitochondrial function in SCs, we compared MuSCs from young and aged mice. For the purpose of identifying the principal mechanisms that initiate MuSC senescence, we examined publically available MuSC gene expression data sets from young (~3 months) and aged (~24 months) mice using gene set enrichment analysis [GSEA; Gene Expression Omnibus (GEO) accession numbers GSE47177 (*14*), GSE47104 (*12*), and GSE47401 (*8*)]. Enrichment scores of data sets from young versus aged mice revealed up-regulation of senescence pathways and down-regulation of cell cycle pathways with age (Fig. 1A, tables S1 to S3, and fig. S1, A and B). This finding is consistent with the idea that irreversible cell cycle arrest is a primary marker of cellular senescence (*2*, *3*). In all three data sets, tricarboxylic acid (TCA) cycle and oxidative phosphorylation (OXPHOS) pathways were among the most down-regulated pathways in aged MuSCs (Fig. 1A, tables S1 to S3, and fig. S1, A and B). Analysis of gene ontology terms that were significantly (GSEA; family-wise error rate $P < 0.05$) down-regulated in aged MuSCs further demonstrated links to mitochondrial function (fig. S1C). Genes that were commonly down-regulated during aging showed a substantial overlap (113 genes; 11.59%) with mitochondrial genes (*26*) (Fig. 1B and table S4), in contrast to the minimal overlap (11 genes; 1.92%) among commonly up-regulated genes and mitochondrial genes (fig. S1D and table S4). Among the 113 down-regulated mitochondrial genes in aged MuSCs, 41.6% were related to the TCA cycle and OXPHOS (fig. S1E), which is higher than their percent composition of the whole mitochondrial proteome (~14%) (*27*, *28*). This indicates a dominant decline in expression of mitochondrial respiratory genes in aged MuSCs. The reduction in mitochondrial OXPHOS and TCA cycle genes was consistent for all independent data sets (fig. S1, F and G). We isolated primary aged and young MuSCs and confirmed reduced abundance of OXPHOS and TCA cycle transcripts (Fig. 1C), as well as reduced oxidative respiration rates in both freshly isolated (Fig. 1D) and cultured MuSCs (fig. S1H). MuSC mitochondrial dysfunction in aged mice was further confirmed by the loss of mitochondrial membrane potential (Fig. 1E) and a reduction in cellular adenosine triphosphate (ATP) concentrations (Fig. 1F). Several important markers and regulators of the mitochondrial unfolded protein response (UPR^{mt}), a stress-response pathway that mediates adaptations in mitochondrial content and function, were down-regulated in aged MuSCs (fig. S1F and Fig. 1G). Despite the absence of consistent changes in the CDKN2A (fig. S1I and table S4) or mitogen-activated protein kinase 14 (MAPK14, also known as p38) pathways (table S4), previously reported to regulate MuSC senescence, expression of cell cycle genes was decreased and expression of genes encoding the senescent proinflammatory secretome was increased (fig. S1, I to K). The reduction in cell cycle signaling was accompanied by increased expression of genes in the cyclin-dependent kinase inhibitor 1 (CDKN1A; also known as $p21^{CIP1}$)–mediated pathway (fig. S1K and table S4), which suggests that early senescence in MuSCs may involve CDKN1A.

NAD^+ repletion improves MuSC function in aged mice

Given the importance of NAD^+ concentrations in the control of mitochondrial function (*16*, *29*), we examined the potential of NAD^+ repletion to improve MuSC numbers and muscle function in aged mice. Amounts of NAD^+ in freshly isolated MuSCs were lower in those isolated from aged mice, and 6 weeks of NR treatment increased NAD^+ concentration in MuSCs from young and old mice (Fig. 2A). Amounts of NADH (the reduced form of NAD^+) were relatively stable

[1]Laboratory of Integrative and Systems Physiology, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. [2]Department of Biology, Institute of Molecular Systems Biology, Eidgenössische Technische Hochschule Zürich (ETHZ), 8093 Zurich, Switzerland. [3]Laboratory of Molecular Biology of Exercise, School of Applied Science, University of Campinas, CEP 13484-350 Limeira, São Paulo, Brazil. [4]Laboratory of Stem Cell Bioengineering, EPFL, 1015 Lausanne, Switzerland. [5]Faculty of Science, University of Zurich, 8057 Zurich, Switzerland. [6]Metabolic Signaling, EPFL, 1015 Lausanne, Switzerland. [7]Interdisciplinary School of Health Sciences, University of Ottawa Brain and Mind Research Institute, 451 Smyth Road, K1H 8M5 Ottawa, Ontario, Canada.
*Corresponding author. Email: kmenzies@uottawa.ca (K.J.M.); admin.auwerx@epfl.ch (J.A.)

(fig. S2A). Muscle from aged mice contained fewer MuSCs compared with that of young animals (Fig. 2, B and C, and fig. S2, B and C). However, NR treatment increased MuSC numbers in young and old mice (Fig. 2, B and C, and fig. S2, B and C). The increase in MuSC numbers was confirmed by staining with paired box protein Pax-7 (PAX7), a known MuSC marker (*4*) (Fig. 2D and fig. S2, D and E). In both young and aged mice, the effect of NR did not appear to result from changes in muscle or body mass, as these measures remained comparable among all groups over the short treatment period (fig. S2, F to I). NR treatment also enhances muscle function in aged animals through an independent mechanism acting directly on the muscle fibers (*16*), as was apparent from improvements in maximal running times and distances, along with limb grip strength in aged mice (Fig. 2, E to G). Young animals showed no such changes (fig. S2, J to L).

Impairments in muscle regeneration efficiency have been linked to the decline in MuSC function in aged mice (*6*). We therefore examined the benefits of NR on muscle regeneration with cardiotoxin (CTX)–induced muscle damage (*4*). NR treatment accelerated muscle regeneration in aged and young mice (Fig. 2H and fig. S2M). NR-induced improvements in regeneration were paralleled by increases in PAX7-stained MuSCs in aged mice (Fig. 2I and fig. S2N) but not in young mice (fig. S2, O and P). NAD^+ repletion also improved the stemness of the aged MuSCs, as demonstrated by a reduction in the ratio of myoblast determination protein 1 (MYOD1) and PAX7–double-positive to PAX7-positive myofibers (Fig. 2J and fig. S2Q), as MYOD1 is a transcriptional factor that activates MuSC differentiation. Seven days after CTX-induced damage, aged mice treated with NR exhibited elevated levels of fibers positive for embryonic myosin heavy chain, a protein expressed in fetal and newly regenerating adult muscle fibers (*30*) (Fig. 2K). MuSCs were then transplanted from NR-treated or control aged mice into *mdx* (C57BL/10ScSn-*Dmd*^{mdx}/J) mice (fig. S2R) (a mouse model of Duchenne muscular dystrophy that gradually loses MuSC function in aging due to the strain of continual muscle regeneration). MuSCs isolated from NR-treated donors more effectively replenished the MuSC compartment and stimulated myogenesis of dystrophin-positive myofibers when transplanted into either young or aged *mdx* recipients (fig. S2S and Fig. 2L, respectively). Thus, NR treatment can improve both muscle regeneration and MuSC transplantation efficiency.

The inappropriate accumulation of nonmyogenic fibro-adipogenic progenitors (FAPs) and inflammatory cells has been reported to impair MuSC function and muscle regeneration, especially in aged or chronically damaged muscle, as found in *mdx* mice (*31*). NR treatment largely attenuated increases in FAP numbers 7 days after CTX-induced damage in aged mice but had no effect on FAPs under basal conditions (fig. S2, T to V). This finding is consistent with benefits to FAP clearance in later periods of muscle regeneration (*31*). NR also alleviated macrophage infiltration 7 days after CTX-induced regeneration in aged mice (fig. S2, W and X).

NR prevents MuSC senescence by improving mitochondrial function

To explain the improvements in MuSCs from aged animals after NAD^+ repletion, we examined effects on MuSC senescence. Freshly isolated MuSCs from NR-treated young and aged mice were immunostained for phosphorylation of histone 2A.X at Ser^{139} (γH2AX), a marker of DNA damage (*2*). γH2AX-stained nuclei were more abundant in MuSCs from aged animals, and staining was reduced with NR treatment (Fig. 3, A and B; results for young controls can be found in fig. S3, A and B). The reduction of the nuclear damage response in MuSCs was confirmed by a single-cell gel electrophoresis (comet) assay, a sensitive measure of DNA strand breaks as an indicator of senescence (Fig. 3C), as well as by staining for β-galactosidase (β-Gal), another classical

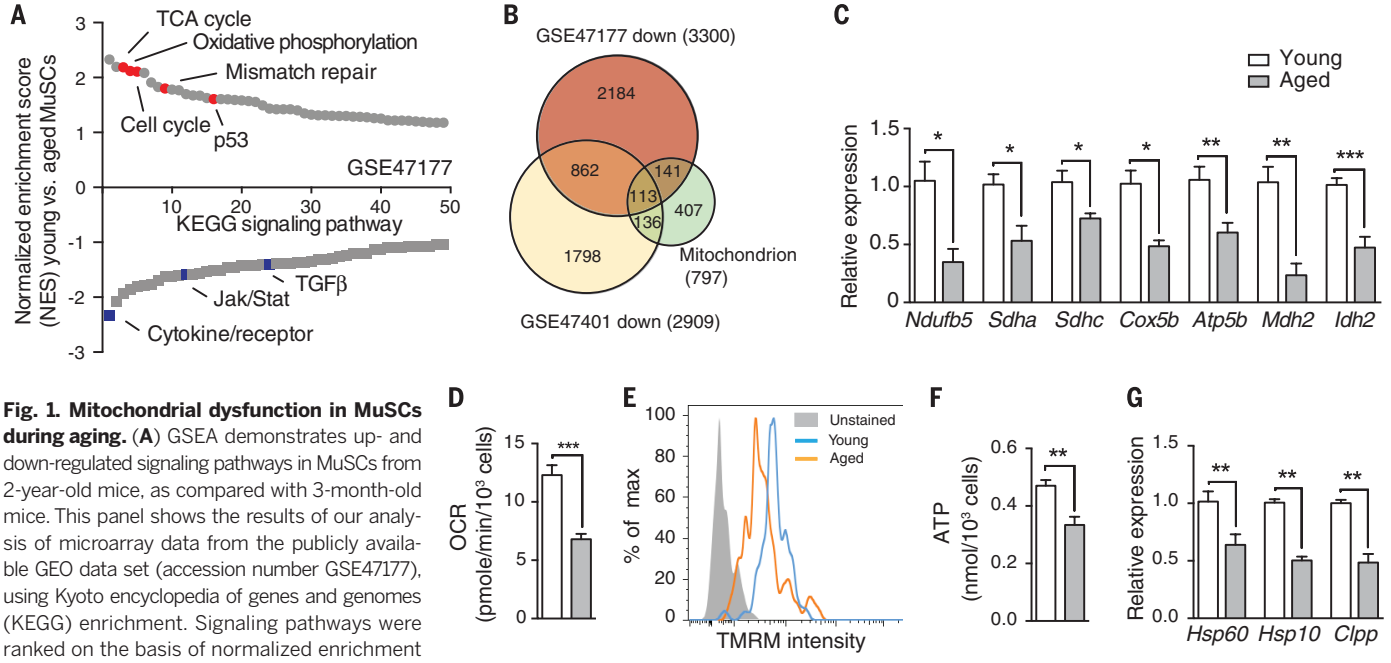


Fig. 1. Mitochondrial dysfunction in MuSCs during aging. (**A**) GSEA demonstrates up- and down-regulated signaling pathways in MuSCs from 2-year-old mice, as compared with 3-month-old mice. This panel shows the results of our analysis of microarray data from the publicly available GEO data set (accession number GSE47177), using Kyoto encyclopedia of genes and genomes (KEGG) enrichment. Signaling pathways were ranked on the basis of normalized enrichment scores (NESs); positive and negative NESs indicate down- or up-regulation, respectively, in aged MuSCs. Specific pathways related to MuSC function are highlighted in red and blue. TGFβ, transforming growth factor–β. (**B**) Area-proportional Venn diagram representing 113 common genes between the significantly down-regulated genes ($P < 0.05$) in MuSC transcriptomes originating from aged mice [GSE47177 and GSE47401 (*12*)] and genes from the human mitochondrial transcriptome (*26*). (**C** to **G**) Young (1 month old) and aged (22 to 24 months old) C57BL/6J mice received a dietary supplement with NR for 6 weeks. (C) Quantitative real-time fluorescence polymerase chain reaction validation of transcriptional changes in mitochondrial genes of freshly sorted MuSCs. (D) Oxygen consumption rate (OCR) in freshly isolated MuSCs after 16 hours of recovery at 37°C. Pmole, picomoles. (E and F) Mitochondrial membrane potential, as measured by a tetramethylrhodamine, methyl ester (TMRM) assay (E) and cellular ATP levels (F) in freshly isolated MuSCs. (G) Relative gene expression for UPR^{mt} genes and cell senescence markers in freshly sorted MuSCs. Data are normalized to *36b4* mRNA transcript levels. All statistical significance was calculated by Student's *t* test. All data are shown as mean ± SEM (error bars). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. In (C), (D), (F), and (G), $n = 6$ mice per group; in (E), $n = 3$ mice per group.

senescence marker (*2*) (fig. S3C). A 6-hour NR treatment of late-passage C2C12 myoblasts also reduced the expression of cell senescence and apoptosis markers (*32*) (Fig. 3D). MuSCs isolated from NR-treated aged mice showed enhanced potential to form myogenic colonies (Fig. 3E and fig. S3D). Thus, NR exerts a protective effect against intrinsic MuSC senescence.

Nicotinamide riboside treatment of MuSCs from aged mice reduced abundance of mRNAs encoding CDKN1A and related proinflammatory proteins and increased the expression of cell cycle genes (Fig. 3F). These effects were not observed in nonsenescent MuSCs from young animals (fig. S3E). NR treatment of MuSCs from aged animals increased expression of genes whose products function in the TCA cycle and OXPHOS (Fig. 3G), an effect that was not evident in young animals (fig. S3F). To quantify protein expression levels under different conditions, we applied a new mass spectrometry–based proteomics technique, the sequential windowed acquisition of all theoretical fragment ion mass spectra (SWATH-MS) (*33*), which allows accurate and reproducible protein quantification across sample cohorts. Using this technique, we were able to quantify the expression changes of more than 1100 proteins in MuSCs across the various conditions. The SWATH-MS results show that a significant amount of proteins that function in OXPHOS and in the UPR^{mt} were decreased in MuSCs from aged animals (Fig. 3H and table S6). The overall amount of these same proteins was increased after NR supplementation [two-way analysis of variance (ANOVA) test, $P < 0.05$ (Fig. 3H and table S6)]. Protein immunoblotting of freshly

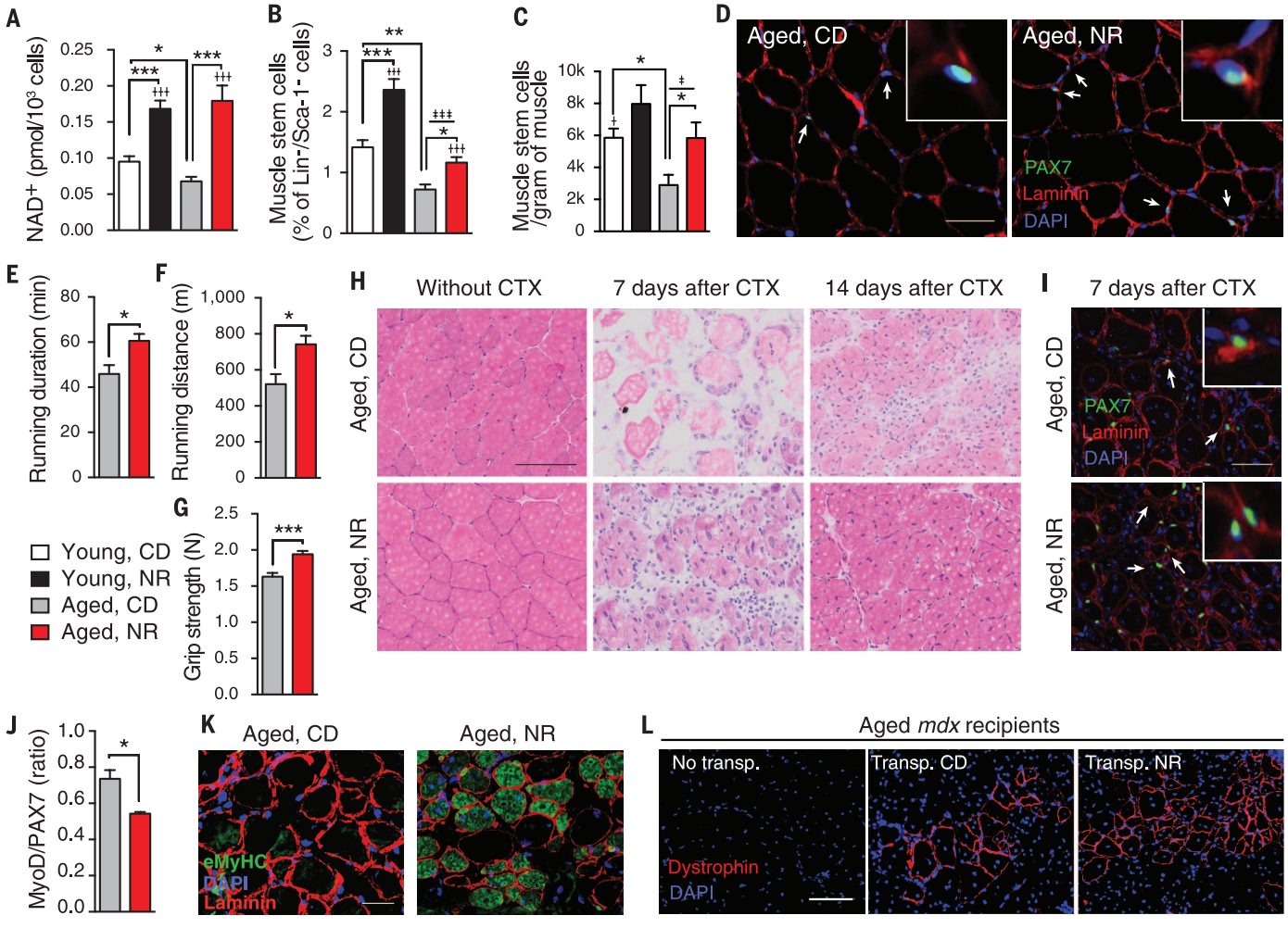


Fig. 2. Improved MuSC numbers and muscle function in NR-treated aged mice. Young (1 month old) and aged (22 to 24 months old) C57BL/6J mice received a chow diet (CD) or a CD supplemented with NR for 6 weeks. All results are compared with those of age-matched mice given a control diet. (**A**) NAD$^+$ concentrations in freshly isolated MuSCs. (**B** and **C**) Percentage of fluorescence-activated cell sorting (FACS)–quantified CD34$^+$/integrin $\alpha7^+$/Lin$^-$/Sca-1$^-$ MuSCs relative to the total Lin$^-$/Sca-1$^-$ cell population (B) or to muscle weight (C). (**D**) Representative images of PAX7- and laminin-immunostained tibialis anterior (TA) muscle. Arrows point to PAX7-positive SCs. Insets (20 μm by 20 μm) show a single MuSC. Scale bar, 50 μm. DAPI, 4′,6-diamidino-2-phenylindole. (**E** to **G**) Comparison of maximal running duration (E), running distance (F), and grip strength (G) in NR-treated aged mice. (**H**) Hematoxylin and eosin (H&E)–stained TA tissue sections from NR-treated aged mice 7 and 14 days after CTX-induced muscle damage. Scale bar, 100 μm. (**I**) Images of PAX7- and laminin-immunostained TA muscle cross sections taken from NR-treated aged mice 7 days after CTX-induced muscle damage. Arrows point to PAX7-positive MuSCs. Insets (20 μm by 20 μm) show a single MuSC. Scale bar, 50 μm. (**J**) Quantification of MYOD1 and PAX7–double positive to PAX7-positive myofibers, performed on sections isolated 7 days after muscle damage in aged mice. (**K**) Newly regenerated muscle fibers stained by embryonic myosin heavy chain (eMyHC) 7 days after muscle damage in aged mice. Scale bar, 50 μm. (**L**) Dystrophin immunostaining of TA muscle sections in aged (16 months old) *mdx* mice 4 weeks after receiving transplantations of MuSCs isolated from control or NR-treated aged C56BL/6J donors. Scale bar, 100 μm. All statistical significance was calculated by Student's *t* test or two-way ANOVA. All data are shown as mean ± SEM (error bars). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Main effects for treatment or age are denoted with † or ‡ symbols, respectively. † or ‡, $P < 0.05$; ††† or ‡‡‡, $P < 0.001$. In (A), $n = 6$ mice per group; in (B) to (D) and (H) to (K), $n = 3$ to 6 mice per group; in (E) to (G), $n = 10$ mice fed the control diet and 7 NR-treated mice; in (L), $n = 12$ donor mice and 3 recipient mice for each treatment. For (E) to (J), corresponding young control data can be found in fig. S2, J to O, respectively.

isolated MuSCs from aged animals confirmed increased expression of proteins related to cell cycle and senescence that could not be detected by SWATH-MS (Fig. 3I).

Muscle SCs from aged mice treated with NR exhibited increases in oxidative respiration (Fig. 3, J and K). NR-treated MuSCs from aged animals also showed increased mitochondrial membrane potential (Fig. 3L and fig. S3G) and increased abundance of ATP (Fig. 3M). To test whether this protective effect of NR on MuSC senescence relies on mitochondrial function, we created a tamoxifen-inducible sirtuin-1 (SIRT1) MuSC-specific knockout mouse (SIRT1$^{MuSC-/-}$) by crossing SIRT1$^{flox/flox}$ mice with the Pax7^{creER} strain. SIRT1 is a NAD$^+$-dependent deacetylase that increases mitochondrial biogenesis (*16*). The beneficial effect of NR on muscle regeneration after CTX injection appeared to be attenuated in SIRT1$^{MuSC-/-}$ mice (Fig. 3N). Supporting this qualitative observation, the act of knocking out SIRT1 in MuSCs blocked the beneficial effects of NR on MuSC activation (Fig. 3, O to Q) and senescence (Fig. 3R and fig. S3H) 7 days after regeneration. These data indicate that NR inhibits MuSC senescence by improving mitochondrial function in a SIRT1-dependent manner. This finding is consistent with a report linking activation of FOXO3, a SIRT1 target, to improved mitochondrial metabolism in hematopoietic SCs (*34*).

Rejuvenating MuSCs by activating the UPR^{mt} and prohibitin pathways

We further explored how UPR^{mt} might regulate senescence by examining the role of prohibitins, a family of stress-response proteins. Prohibitins sense mitochondrial stress and modulate senescence in fibroblasts in mammals (*35*), maintain replicative life span in yeast (*36*), and promote longevity in worms (*37*), animals that lack adult SCs. Expression of the prohibitins *Phb* and *Phb2* was reduced in the bioinformatics analysis (fig. S1F),

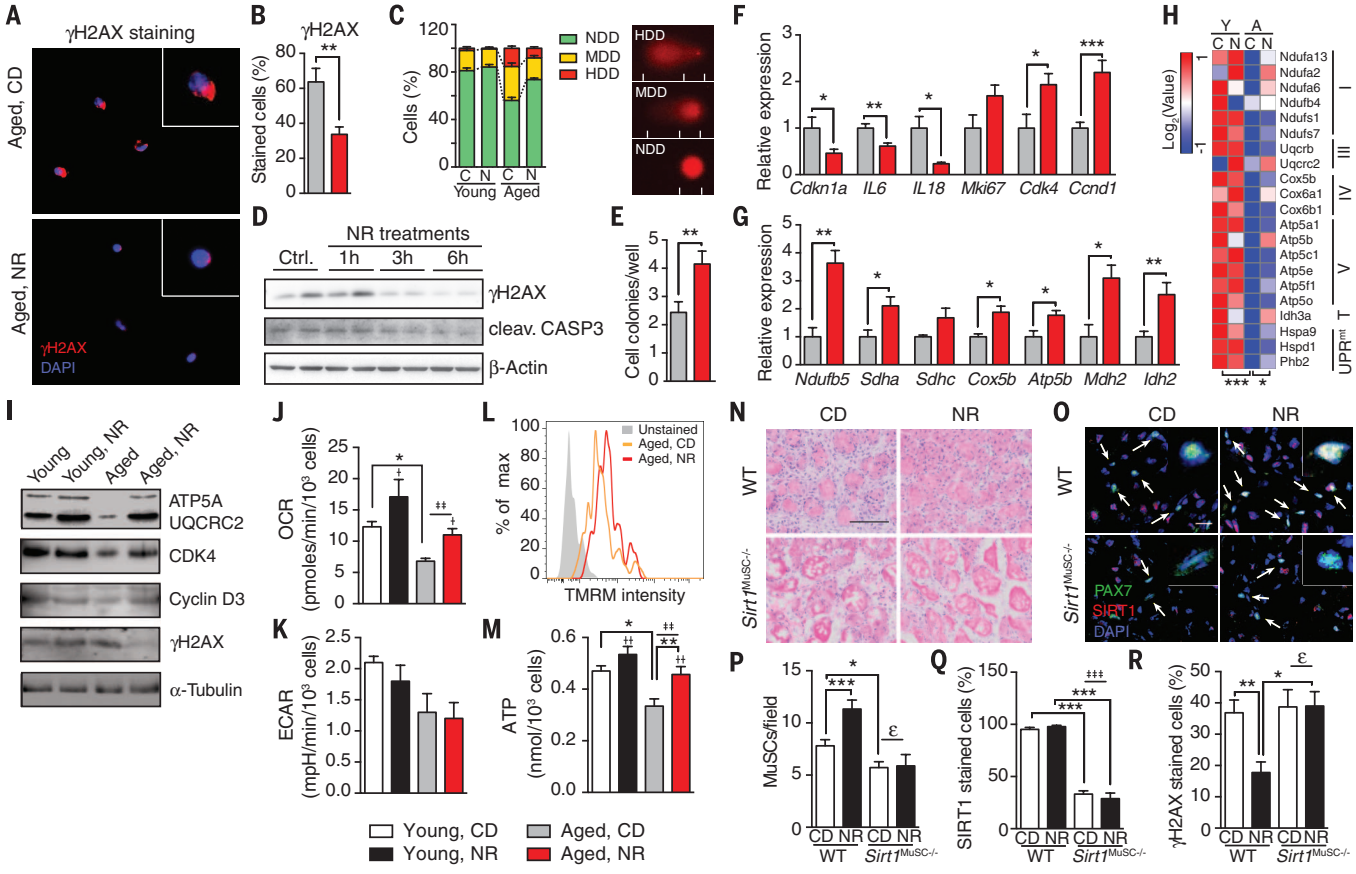


Fig. 3. NR treatment prevents MuSC senescence and improves mitochondrial function. Aged (22 to 24 months old) C57BL/6J mice or 8-month-old SIRT1$^{MuSC-/-}$ mice received a dietary supplement with NR for 6 weeks. All isolated MuSCs were freshly FACS sorted for assay. Most comparative data from young mice (1 month old) are presented in fig. S3. (**A** and **B**) Immunostaining (A) and quantification (B) of γH2AX staining in freshly sorted MuSCs from aged mice. Insets (20 μm by 20 μm) show single MuSCs. (**C**) Single-cell gel electrophoresis (comet) assay of MuSCs from aged mice. C, chow diet; N, NR treated; NDD, nondamaged DNA; MDD, moderately damaged DNA; HDD, heavily damaged DNA. (**D**) Proteins levels in C2C12 myoblasts after NR treatment for 1, 3, or 6 hours. (**E**) Colony-formation ability assay in isolated MuSCs. (**F** and **G**) Quantification of transcript expression for cell cycle and inflammatory secretome genes (F) or OXPHOS and TCA cycle genes (G) in MuSCs. (**H**) Abundance of proteins from MuSCs of young (Y) and aged (A) mice fed with a chow (C) or NR (N) diet. Protein abundances were calculated using the intensities of peptides detected in the SWATH-MS maps. Roman numerals indicate corresponding OXPHOS complexes. T, TCA cycle. (**I**) Protein levels in MuSCs. (**J** and **K**) OCR (J) and extracellular acidification rate (ECAR) (K), in MuSCs after 16 hours of recovery at 37°C. (**L**) Mitochondrial membrane potential, as measured by a TMRM assay in MuSCs. (**M**) Cellular ATP concentration in MuSCs. (**N**) H&E stained TA muscle from wild-type (WT) or SIRT1$^{MuSC-/-}$ mice 7 days after CTX-induced muscle damage. Scale bar, 100 μm. (**O** to **Q**) Representative images (O) and quantification of PAX7-positive MuSCs in random fields of view (160 μm by 160 μm) (P) and the percentage of SIRT1-positive MuSCs (Q) in immunostained TA 7 days after CTX-induced muscle damage. Arrows in (O) point to PAX7-positive MuSCs. Insets (20 μm by 20 μm) show a single MuSC. Scale bar, 50 μm. (**R**) Quantification of γH2AX-positive MuSCs in immunostained TA 7 days after CTX-induced muscle damage. All statistical significance was calculated by Student's *t* test or two-way ANOVA. All data are represented as mean ± SEM (error bars). **P* < 0.05; ***P* < 0.01; ****P* < 0.001. Main effects for treatment or age are denoted as † or ‡, respectively, with interactions denoted as ε. † or ‡, *P* < 0.05; †† or ‡‡, *P* < 0.01; ††† or ‡‡‡, *P* < 0.001. In (A) to (C) and (N) to (R), *n* = 3 to 6 mice per group; in (E), *n* = 24 repeats per group; in (F) and (G) and (J) to (M), *n* = 6 mice per group; in (H), protein was extracted and pooled from 6 mice in each group. For (A), (B), (E), (F), (G), and (L), corresponding young control data can be found in fig. S3, A, B, D, E, F, and G, respectively.

as well as in freshly isolated aged MuSCs (fig. S4A). The addition of NR increased the expression of prohibitin proteins in C2C12 myoblasts (Fig. 4A) and transcripts in MuSCs of young and aged mice (Fig. 4B and fig. S4B). NR treatment was also shown to increase the expression of prohibitins concurrent to markers of UPR^{mt} and the cell cycle (fig. S4C). Moreover, the overexpression of prohibitins, in the absence of NR, likewise increased UPR^{mt} and cell cycle protein expression (Fig. 4C). Demonstrating the dependency of the NR effect on prohibitins, improvements in UPR^{mt} and cell cycle protein expression were not observed with NR treatment after the knockdown of prohibitins (Fig. 4D and fig. S4D). To confirm the regulation of prohibitins on cell cycle proteins and to explore the effect of prohibitins on MuSC function, *Phb* was depleted in vivo through an intramuscular injection of sh*Phb* lentivirus (PHB and PHB2 are functional only as a heterozygous protein complex) (*38*). Impairment of muscle regeneration and a reduction in MuSC numbers were observed in sh*Phb* lentivirus–injected mice 7 days after CTX-induced muscle regeneration (Fig. 4, E and F). Quantifying these results, *Phb* knockdown is shown to block the NR-induced increase of MuSCs upon regeneration (Fig. 4, G and H). *Phb* knockdown does not induce more MuSC senescence in aged mice but does prevent the beneficial effect of NR on MuSC senescence (Fig. 4I and fig. S4E). Initiation of the UPR^{mt} by thiamphenicol also induced expression of prohibitins and cell cycle genes in C2C12 cells (fig. S4F). These results indicate that NR activates the UPR^{mt} and the prohibitin signaling pathway as it inhibits MuSC senescence.

NR reprograms senescence-prone MuSCs in *mdx* mice

With continuous muscle regeneration, MuSCs in *mdx* mice are abnormally active at a young age, leading to MuSC depletion and dysfunction later in life. As a result, primary MuSCs isolated from 14-week-old *mdx* mice were more intensively and frequently stained with β-Gal and had a larger cell size than those of control mice (Fig. 5A and fig. S5, A and B). Similar to the effect in aged animals, NR treatment of *mdx* mice increased

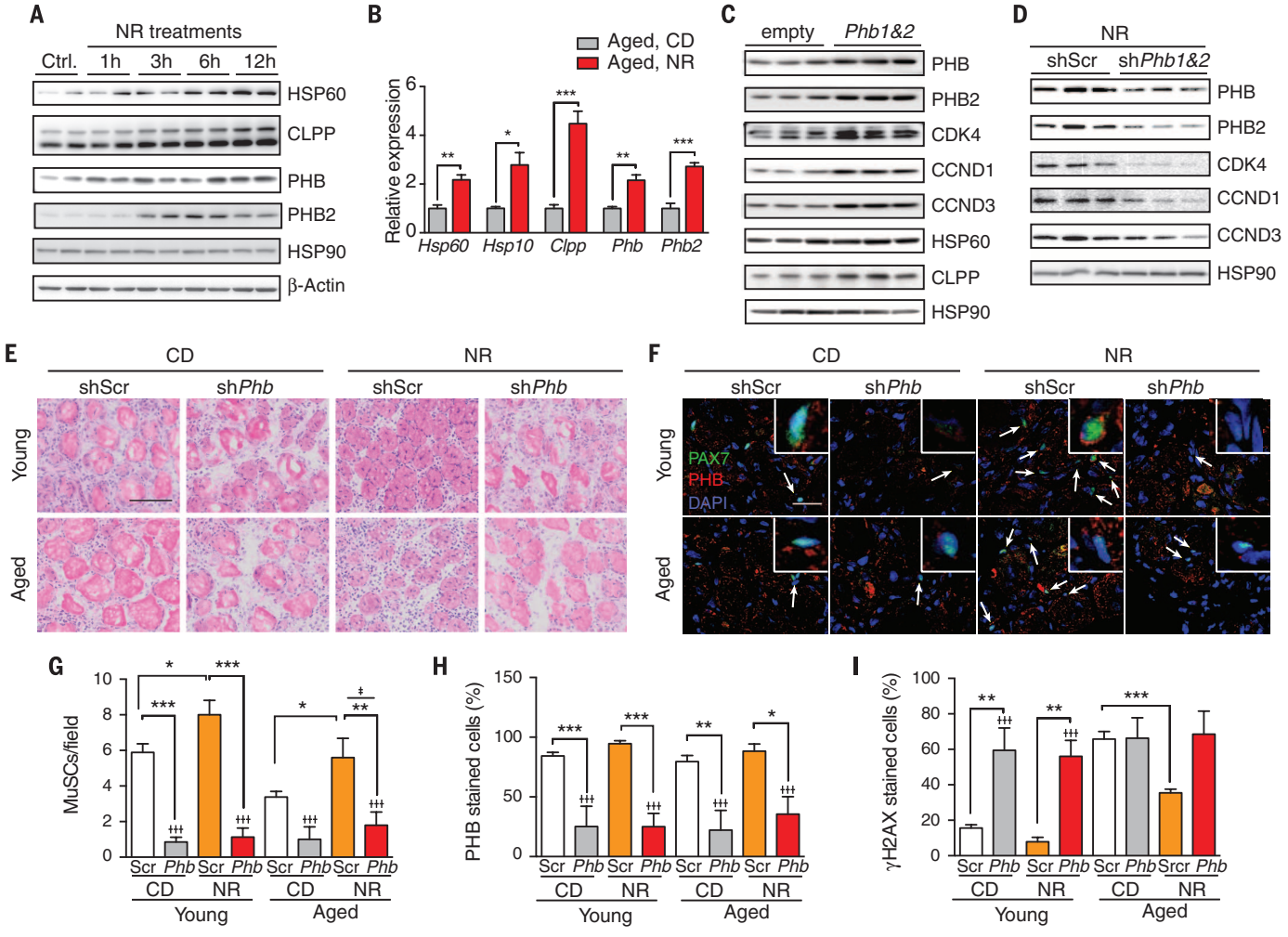


Fig. 4. Effects of NR on prohibitins, UPR^{mt}, and MuSC senescence. (**A**) Expression of HSP60, CLPP, and prohibitins in C2C12 myoblasts upon NR treatment at the indicated time points. (**B**) Quantification of transcript expression for UPR^{mt} and prohibitin genes in MuSCs from aged (22 to 24 months old) C57BL/6J mice after 6 weeks of chow or NR diets. (**C**) Expression of prohibitins and cell cycle proteins in C2C12 myoblasts with the combined overexpression of *Phb* and *Phb2*. (**D**) Expression of prohibitins and cell cycle genes with a 6-hour NR treatment in C2C12 myoblasts after a combined *Phb* and *Phb2* short hairpin RNA knockdown. (**E**) H&E staining of TA muscle in NR-treated or intramuscular sh*Phb* lentivirus–injected C57BL/6J mice 7 days after CTX-induced muscle damage. Scale bar, 100 μm. (**F** to **H**) Representative images (F) and quantification of PAX7-positive MuSCs in randomly chosen fields of view (160 μm by 160 μm) (G) and the percentage of PHB-positive MuSCs (H) in immunostained TA muscle 7 days after CTX-induced muscle damage. Arrows in (F) point to PAX7-positive MuSCs. Insets (20 μm by 20 μm) show a single MuSC. Scale bar, 50 μm. (**I**) Quantification of γH2AX-positive MuSCs in immunostained TA muscle cross sections taken from control and NR-treated mice 7 days after CTX-induced muscle damage. All statistical significance was calculated by Student's *t* test or two-way ANOVA. All data are represented as mean ± SEM (error bars). *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. Main effects for treatment or *Phb* knockdown are denoted as † or ‡, respectively. † or ‡, $P < 0.05$; ††† or ‡‡‡, $P < 0.001$. In (B), $n = 6$ mice per group; in (E) to (I), $n = 3$ mice per group. For (B), corresponding young control data can be found in fig. S4B.

MuSC numbers by a factor of ~1.8 in vivo (Fig. 5, B to D, and fig. S5, C and D), as also confirmed by PAX7 immunostaining (fig. S5E). Along with the increase in MuSCs, there was an increase in regenerated muscle fibers after NR treatment (Fig. 5E and fig. S5F). We extended this analysis by examining the self-renewal capacity of MuSCs in *mdx* mice. The cellular redox ratio decreases as MuSCs differentiate (*39*), which can be detected by an increase in 450-nm autofluorescence (*40*). In line with NR increasing the number of MuSCs in *mdx* mice, we found reduced autofluorescence from MuSCs isolated from these animals (Fig. 5, F to H). We performed β-Gal staining on primary MuSCs that had been isolated from *mdx* mice with or without NR treatment in vivo and then further cultured with or without NR in vitro. MuSCs isolated from NR-treated mice were less prone to senescence (Fig. 5I and fig. S5G). When MuSCs isolated from control *mdx* mice were treated with NR in vitro, there was also a reduction in senescence (Fig. 5I and fig. S5G). The inhibition of MuSC senescence in NR-treated *mdx* mice was confirmed by the attenuation of γH2AX and cleaved caspase-3 immunostaining (Fig. 5J). To evaluate MuSC function, CTX-induced muscle regeneration was examined in NR-treated *mdx* mice. Consistent with the prevention of MuSC senescence, muscle regeneration was improved with NR in both aged (Fig. 5K and fig. S5H) and young *mdx* mice (fig. S5I). We also examined the effect of NR on the FAP population and muscle regeneration in *mdx* mice. NR treatment increased MuSCs and reduced FAP numbers in basal conditions and 7 days after CTX-induced damage (fig. S5, J to L). Abnormal activation of FAPs in *mdx* mice contributes to fibrosis (*31*). *mdx* mice treated with CTX and then exposed to NR showed lower levels of macrophage infiltration 7 days after damage (fig. S5, M and N). Hence, our results indicate a beneficial effect of NR on MuSC function and regeneration in *mdx* mice.

NR attenuates senescence of neural and melanocyte SCs and increases mouse life span

Aging is accompanied by a decline in the number and function of neural SCs (NSCs) (*23*) and melanocyte SCs (McSCs) (*41*). Therefore, to examine the generalized importance of NAD^+ homeostasis in somatic SCs, we assessed the effect of NR in NSCs from aged mice. NR increased proliferation, as shown by 5-ethynyl-2-deoxyuridine (EdU) and

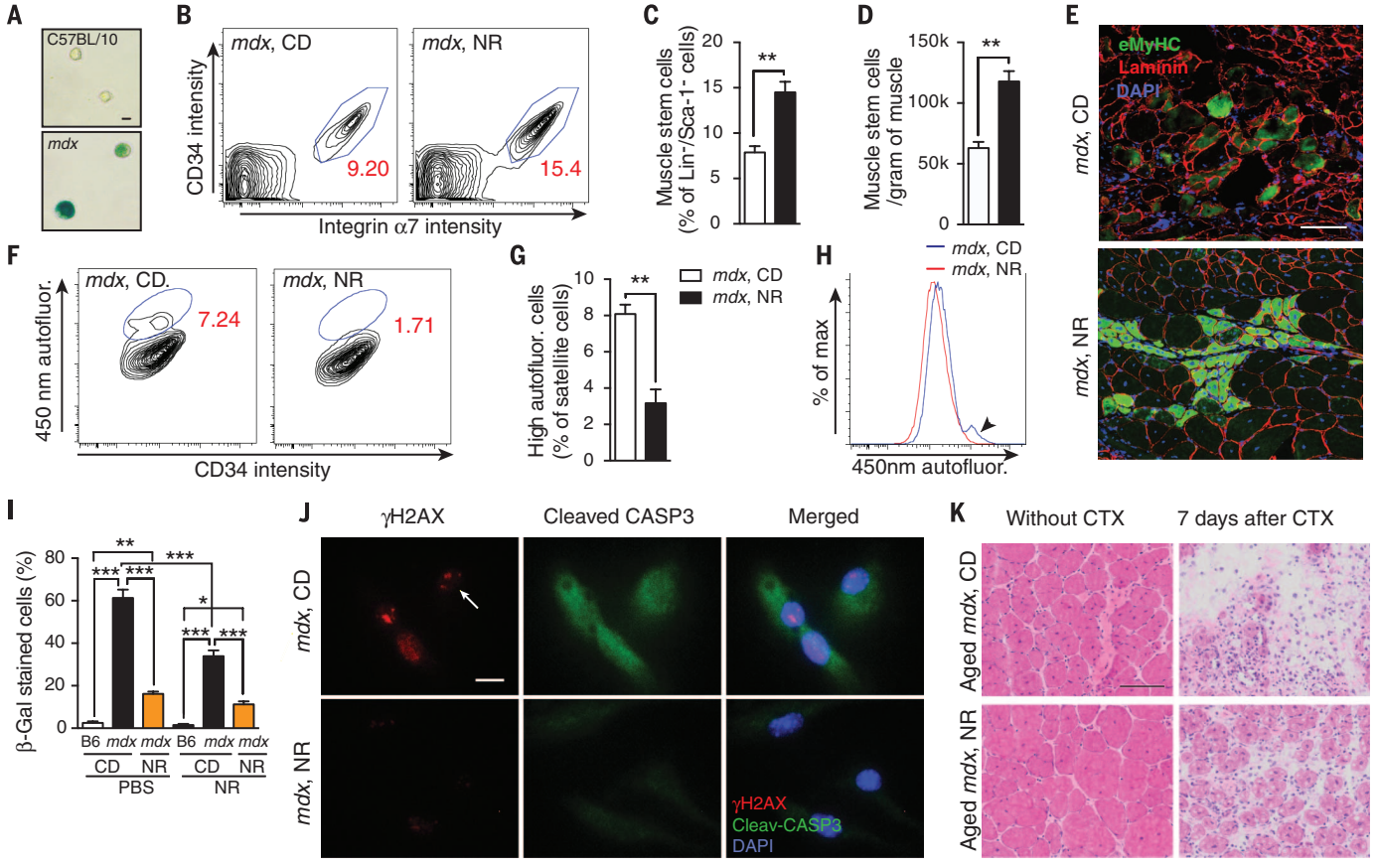


Fig. 5. Increased SC number and function in NR-treated *mdx* mice. *mdx* mice received a dietary supplement with NR for 10 weeks. All results were compared with those of *mdx* mice given a control diet. (**A**) β-Gal staining of MuSCs isolated from C57BL/10SnJ or *mdx* mice and cultured in vitro for three generations. Scale bar, 10 μm. (**B** to **D**) FACS contour plots of Sca-1^-/Lin^- cells isolated from muscle. Percentages of the $CD34^+$/integrin $\alpha7^+$/Lin^-/Sca-1^- MuSC populations are noted in red in the contour plots (B) and are quantified relative to the total Lin^-/Sca-1^- cell population (C) or muscle weight (D). (**E**) Immunostaining of $eMyHC^+$ fibers in tissue sections of NR-treated *mdx* mice 7 days after CTX-induced muscle damage. Scale bar, 100 μm. (**F** to **H**) FACS contour plots (F), quantification (G), and distribution (H) of MuSC autofluorescence as a measure of the relative NAD(P)H concentration (where NADPH is the reduced form of nicotinamide adenine dinucleotide phosphate) upon ultraviolet light excitation. Autofluorescence emission was detected with a wavelength of 450 nm. The arrowhead in (H) points to the highly autofluorescent SC population. (**I**) Quantification of β-Gal staining of FACS-sorted MuSCs from C57BL/6J (B6 with CD), untreated (*mdx* with CD), or NR-treated *mdx* (*mdx* with NR) mice challenged with phosphate-buffered saline or NR for 6 hours in vitro. (**J**) Immunostaining showing γH2AX and cleaved caspase-3 in MuSCs cultured in vitro for three generations. The arrow points to a γH2AX-positive nucleus. Scale bar, 10 μm. (**K**) H&E staining of tissue sections from NR-treated aged *mdx* mice (16 months old) with 7 days of recovery after CTX-induced muscle damage. Scale bar, 100 μm. All statistical significance was calculated by Student's *t* test or one-way ANOVA. All data are represented as mean ± SEM (error bars). $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. In (A) to (H), (J), and (K), $n = 3$ to 5 mice per treated group; in (I), $n = 3$ mice and $n = 6$ in vitro treatments. More than 500 cells were quantified for each condition.

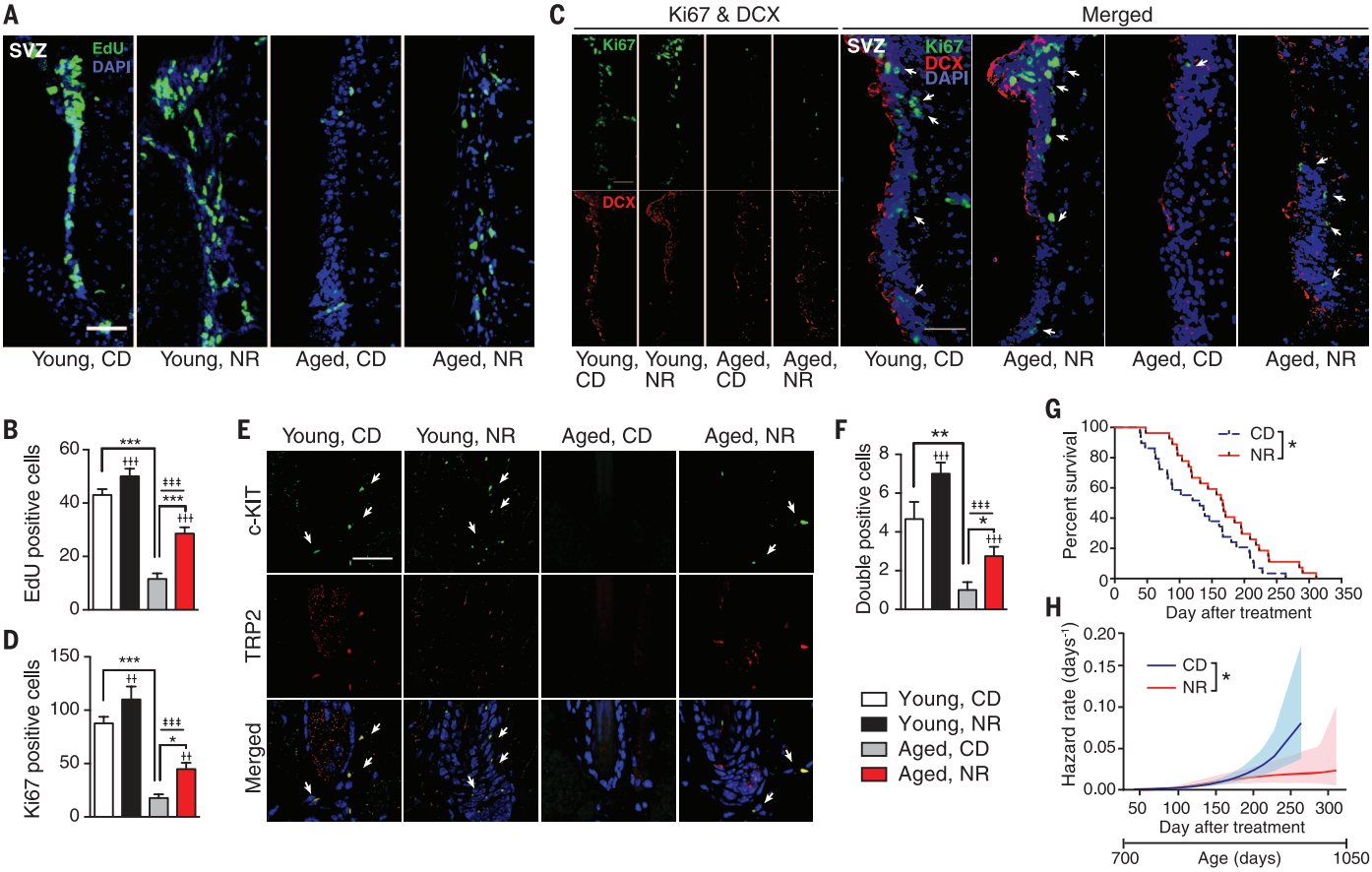


Fig. 6. NR improves NSC and McSC function and increases the life span of aged C57BL/6J mice. Aged (22 to 24 months old) C57BL/6J mice received a dietary supplement with NR for 6 weeks. (**A** and **B**) Representative images (A) and quantification (B) of EdU-positive NSCs in the SVZ from young and aged mice after NR treatment. Scale bar, 50 μm. (**C** and **D**) Representative images (C) and quantification (D) of Ki67- and DCX-positive NSCs in the SVZ, harvested from young and aged mice treated with or without NR. Arrows in (C) point to Ki67-positive NSCs. Scale bar, 50 μm. (**E** and **F**) Representative images (E) and quantification (F) of c-KIT and TRP2 double-positive McSCs in the bulge of hair follicles from dorsal skin, harvested from young and aged mice treated with or without NR. Arrows in (E) point to double-positive McSCs. Scale bar, 50 μm. (**G**) Kaplan-Meier survival curves of control and NR-treated aged mice, with the NR treatment beginning at 2 years (700 days) of age. (**H**) Hazard rate decreased under NR treatment. Individual life spans were collected and used to estimate the hazard function of each population, using numerical differentiation of the Kaplan-Meier survival estimator (solid lines). The shaded areas represent the 95% confidence bands of the true hazard. *P* values were calculated with the use of a Cox proportional hazards model. All statistical significance was calculated by Student's *t* test or two-way ANOVA, except in (G) and (H). All data are represented as mean ± SEM (error bars). $*P < 0.05$; $**P < 0.01$; $***P < 0.001$. Main effects for treatment or age are denoted as † or ‡, respectively. †† or ‡‡, $P < 0.01$; ††† or ‡‡‡, $P < 0.001$. In (A) to (F), $n = 6$ mice per group; in (H), $n = 30$ mice per treated group.

antigen K_i-67 (Ki67) staining, and induced neurogenesis, as indicated by doublecortin (DCX) staining, in both the subventricular zone (SVZ) (Fig. 6, A to D) and the dentate gyrus of the hippocampus (fig. S6, A to D) in aged mice. Nicotinamide mononucleotide, another NAD^+ precursor, also has beneficial effects in aged NSCs (*23*). Similarly, NR rescued the decline of McSCs in hair follicles of aged mice, as reflected by increases in mast/stem cell growth factor receptor Kit (c-KIT) and short transient receptor potential channel 2 (TRP2), known McSC markers, in NR-treated aged mice (Fig. 6, E and F). NR treatment of C57BL/6J mice slightly increased life span (chow diet: mean 829 ± 12.0 days; NR treatment: mean 868 ± 12.4 days; $P = 0.034$) (Fig. 6G). The beneficial effect of NR on survival was further confirmed by Cox proportional hazards analysis (Fig. 6H). Although the life span benefit is small, it was obtained with the NR treatment commencing late in life at 24 months, which indicates that aging may partially stem from the dysregulation of general SC NAD^+ homeostasis.

Conclusions

Oxidative stress, potentially introduced by mitochondrial respiration, is thought to be circumvented in SCs by their reliance on glycolysis as a primary energy resource (*42*). However, our study demonstrates that mitochondrial oxidative respiration is important for the functional maintenance of multiple types of adult SCs during aging. In fact, the reduction in cellular NAD^+ pools blunts the adaptive UPR^{mt} pathway (*18*), ultimately leading to a loss of mitochondrial homeostasis with a concurrent reduction in the number and self-renewal capacity of MuSCs. Accordingly, by boosting the MuSC concentration of NAD^+, proteotoxic stress resistance may be restored due to the activation of the UPR^{mt} pathway, stimulating the prohibitin family of mitochondrial stress sensors and effectors. In turn, this effect will improve mitochondrial homeostasis, protecting MuSCs from senescence and safeguarding muscle function in aged mice (fig. S6E). With the use of a MuSC-specific loss-of-function model for *Sirt1*, an essential regulator governing mitochondrial homeostasis (*43*), the importance and vital nature of the relationship between the NAD^+-SIRT1 pathway, mitochondrial activity, and MuSC function were unequivocally established in vivo. Furthermore, maintaining healthy mitochondria by replenishing NAD^+ stores seems to have beneficial effects beyond MuSCs and also appears to protect NSC and McSC populations from aging.

Our results demonstrate that the depression of prohibitin signaling, leading to mitochondrial

dysfunction, can be reversed in aging by using a nutritional intervention to boost NAD^+ concentrations in SCs. Additionally, our findings suggest that NAD^+ repletion may be revealed as an attractive strategy for lengthening mammalian life span.

REFERENCES AND NOTES

1. A. J. Wagers, I. L. Weissman, *Cell* **116**, 639–648 (2004).
2. T. Kuilman, C. Michaloglou, W. J. Mooi, D. S. Peeper, *Genes Dev.* **24**, 2463–2479 (2010).
3. C. López-Otín, M. A. Blasco, L. Partridge, M. Serrano, G. Kroemer, *Cell* **153**, 1194–1217 (2013).
4. H. Yin, F. Price, M. A. Rudnicki, *Physiol. Rev.* **93**, 23–67 (2013).
5. M. Tabebordbar, E. T. Wang, A. J. Wagers, *Annu. Rev. Pathol.* **8**, 441–475 (2013).
6. Y. C. Jang, M. Sinha, M. Cerletti, C. Dall'Osso, A. J. Wagers, *Cold Spring Harb. Symp. Quant. Biol.* **76**, 101–111 (2011).
7. C. S. Fry *et al.*, *Nat. Med.* **21**, 76–80 (2015).
8. F. D. Price *et al.*, *Nat. Med.* **20**, 1174–1181 (2014).
9. I. M. Conboy *et al.*, *Nature* **433**, 760–764 (2005).
10. J. V. Chakkalakal, K. M. Jones, M. A. Basson, A. S. Brack, *Nature* **490**, 355–360 (2012).
11. P. Sousa-Victor *et al.*, *Nature* **506**, 316–321 (2014).
12. J. D. Bernet *et al.*, *Nat. Med.* **20**, 265–271 (2014).
13. B. D. Cosgrove *et al.*, *Nat. Med.* **20**, 255–264 (2014).
14. L. Liu *et al.*, *Cell Reports* **4**, 189–204 (2013).
15. M. T. Tierney *et al.*, *Nat. Med.* **20**, 1182–1186 (2014).
16. C. Cantó *et al.*, *Cell Metab.* **15**, 838–847 (2012).
17. E. Pirinen *et al.*, *Cell Metab.* **19**, 1034–1041 (2014).
18. L. Mouchiroud *et al.*, *Cell* **154**, 430–441 (2013).
19. J. Yoshino, K. F. Mills, M. J. Yoon, S. Imai, *Cell Metab.* **14**, 528–536 (2011).
20. A. P. Gomes *et al.*, *Cell* **155**, 1624–1638 (2013).
21. K. Ito, T. Suda, *Nat. Rev. Mol. Cell Biol.* **15**, 243–256 (2014).
22. M. Cerletti, Y. C. Jang, L. W. Finley, M. C. Haigis, A. J. Wagers, *Cell Stem Cell* **10**, 515–519 (2012).
23. L. R. Stein, S. Imai, *EMBO J.* **33**, 1321–1340 (2014).
24. P. Katajisto *et al.*, *Science* **348**, 340–343 (2015).
25. J. G. Ryall *et al.*, *Cell Stem Cell* **16**, 171–183 (2015).
26. T. R. Mercer *et al.*, *Cell* **146**, 645–658 (2011).
27. A. Sickmann *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13207–13212 (2003).
28. D. J. Pagliarini *et al.*, *Cell* **134**, 112–123 (2008).
29. R. H. Houtkooper, C. Cantó, R. J. Wanders, J. Auwerx, *Endocr. Rev.* **31**, 194–223 (2010).
30. S. Sartore, L. Gorza, S. Schiaffino, *Nature* **298**, 294–296 (1982).
31. D. R. Lemos *et al.*, *Nat. Med.* **21**, 786–794 (2015).
32. E. Hara *et al.*, *Mol. Cell. Biol.* **16**, 859–867 (1996).
33. L. C. Gillet *et al.*, *Mol. Cell. Proteomics* **11**, O111.016717 (2012).
34. P. Rimmelé *et al.*, *EMBO Rep.* **16**, 1164–1176 (2015).
35. P. J. Coates *et al.*, *Exp. Cell Res.* **265**, 262–273 (2001).
36. P. J. Coates, D. J. Jamieson, K. Smart, A. R. Prescott, P. A. Hall, *Curr. Biol.* **7**, 607–610 (1997).
37. M. Artal-Sanz, N. Tavernarakis, *Nature* **461**, 793–797 (2009).
38. C. Osman, C. Merkwirth, T. Langer, *J. Cell Sci.* **122**, 3823–3830 (2009).
39. M. Fulco *et al.*, *Mol. Cell* **12**, 51–62 (2003).
40. K. P. Quinn *et al.*, *Sci. Rep.* **3**, 3432 (2013).
41. E. K. Nishimura, S. R. Granter, D. E. Fisher, *Science* **307**, 720–724 (2005).
42. C. D. Folmes, P. P. Dzeja, T. J. Nelson, A. Terzic, *Cell Stem Cell* **11**, 596–606 (2012).
43. C. Cantó, K. J. Menzies, J. Auwerx, *Cell Metab.* **22**, 31–53 (2015).

ACKNOWLEDGMENTS

H.Z., D.R., K.J.M., J.A., and the EPFL have filed a provisional patent application on the use of NAD boosting to enhance SC function.

We thank T. Langer for sharing the *Phb* plasmids; S. Wang and M. Knobloch for technical help in McSC and NSC experiments; H. Li, L. Mouchiroud, P. Moral Quiros, and all members of the Auwerx and Schoonjans groups for helpful discussions; and the EPFL histology and flow cytometry core facilities for technical assistance. H.Z. is the recipient of a doctoral scholarship from the China Scholarship Council and a fellowship from CARIGEST SA. D.D. was supported by a fellowship from Associazione Italiana per la Ricerca sul Cancro. K.J.M. is supported by the University of Ottawa and the Heart and Stroke Foundation of Canada. J.A. is the Nestlé Chair in Energy Metabolism, and his research is supported by EPFL, the NIH (grant R01AG043930), Krebsforschung Schweiz/SwissCancerLeague (grant KFS-3082-02-2013), Systems X (grant SySX.ch 2013/153), and the Swiss National Science Foundation (grant 31003A-140780).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1436/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S6
References (*44*–*54*)

16 January 2016; accepted 13 April 2016
Published online 28 April 2016
10.1126/science.aaf2693

REPORTS

MOLECULAR JUNCTIONS

Covalently bonded single-molecule junctions with stable and reversible photoswitched conductivity

Chuancheng Jia,[1]* Agostino Migliore,[2]* Na Xin,[1]* Shaoyun Huang,[3]* Jinying Wang,[1] Qi Yang,[3] Shuopei Wang,[4] Hongliang Chen,[1] Duoming Wang,[4] Boyong Feng,[3] Zhirong Liu,[1] Guangyu Zhang,[4] Da-Hui Qu,[5] He Tian,[5] Mark A. Ratner,[6] H. Q. Xu,[3]† Abraham Nitzan,[7,8]† Xuefeng Guo[1,9]†

Through molecular engineering, single diarylethenes were covalently sandwiched between graphene electrodes to form stable molecular conduction junctions. Our experimental and theoretical studies of these junctions consistently show and interpret reversible conductance photoswitching at room temperature and stochastic switching between different conductive states at low temperature at a single-molecule level. We demonstrate a fully reversible, two-mode, single-molecule electrical switch with unprecedented levels of accuracy (on/off ratio of ~100), stability (over a year), and reproducibility (46 devices with more than 100 cycles for photoswitching and ~10^5 to 10^6 cycles for stochastic switching).

Rapidly growing research in nanoscience has implications for the development of computing devices, solar energy harvesting, chemical sensing (*1*–*4*), photonics and optoelectronics, biomedical electronics [such as cell-chip connections, cyborg cells (*5*), electroceuticals (*6*), and prosthetics], and biofuel cells (*7*). The development of electronic devices based on controllable molecular conduction aims to meet the urgent demand for further device miniaturization on one hand, and the need to effectively interface organic and inorganic materials for biomedical and nanoelectronic applications on the other. To this end, diverse approaches to molecular nanodevices have been proposed and have faced important issues of reproducibility and stability (*8*).

Switches are basic components of almost any electronic device. The manufacturing of reliable molecular switches is a crucial test of the possibility to use molecules as pivotal components of electronic devices. Molecular switches have been intensively investigated for two decades (*9*–*11*), but only a few studies have demonstrated unidirectional switching (namely, irreversible change) in molecular conduction (*12*–*14*), while failing to produce actual (bidirectional) conductance switching. In this work, we demonstrate a reversible molecular electrical switch that consists of a single diarylethene molecule in a junction comprising graphene electrodes.

[1]Beijing National Laboratory for Molecular Sciences, State Key Laboratory for Structural Chemistry of Unstable and Stable Species, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, P. R. China. [2]Department of Chemistry, Duke University, Durham, NC 27708, USA. [3]Department of Electronics and Key Laboratory for the Physics and Chemistry of Nanodevices, Peking University, Beijing 100871, P. R. China. [4]Institute of Physics, Chinese Academy of Sciences, Beijing 100190, P. R. China. [5]Key Laboratory for Advanced Materials and Institute of Fine Chemicals, East China University of Science and Technology, Shanghai 200237, P. R. China. [6]Department of Chemistry, Northwestern University, Evanston, IL 60208, USA. [7]Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104-6323, USA. [8]School of Chemistry, Tel Aviv University, Tel Aviv, 69978, Israel. [9]Department of Materials Science and Engineering, College of Engineering, Peking University, Beijing 100871, P. R. China.
*These authors contributed equally to this work. **†Corresponding author. Email: guoxf@pku.edu.cn (X.G.); anitzan@sas.upenn.edu (A.N.); hqxu@pku.edu.cn (H.Q.X.)**

One of the most challenging issues in fabricating reliable (namely, stable and reproducible) molecular switches is the lack of control of the properties of the molecule-electrode interface (*15*). In particular, only unidirectional optoelectronic switching (from closed and conducting to open and nonconducting diarylethene) was observed for single diarylethenes sandwiched between gold electrodes via Au–S bonds (*12*). This response was attributed to quenching of the excited state of the open molecular form in the presence of the Au electrode. We have reported the opposite unidirectional photoswitching of single diarylethenes (from the insulating open form to the conducting closed form) when they were used to bridge nanogaps between carbon electrodes [either single-walled carbon nanotubes (*13*) or graphene sheets (*14*)] because the molecule remained locked in its closed form as a consequence of the energy transfer from the photoexcited molecule to the extended π-electron system in the electrodes. This quenching effect resulted from the strong molecule-electrode couplings produced by the covalent amide linkages, which we originally used to enhance the device stability (*16*, *17*). These results have stressed the crucial importance of the molecule-electrode coupling strength in determining the device performance.

To overcome this issue, in the present study we have incorporated three methylene (CH_2) groups into each side of the molecular backbone (Fig. 1, scheme S1, and fig. S1) (*18*) in order to decrease the effective molecule-electrode coupling. Theoretical analysis suggests that this structural modification reduces the electrode-molecule coupling to a level that may avoid the quenching of the molecular excited state by the electrodes and moves the system from the Landauer regime without the CH_2 groups closer to the Coulomb blockade regime, in which the molecule is locked in either the closed or open conformation. Our calculations show narrow resonance half-widths of the transmission peaks [in correspondence to the perturbed highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies (fig. S2)] (*18*), which amount to interfacial graphene-molecule coupling on the order of 1 meV in both open and closed forms—that is, considerably weaker than those in previous cases (*14*).

It was found that diarylethene-reconnected molecular junctions with graphene electrodes reversibly switched at a single-molecule level in a robust and reproducible way (Fig. 2); 46 devices were successfully implemented and studied. The on/off ratio was 107.1 ± 56.3 when light with different wavelengths was toggled back and forth (Fig. 2, A and B; figs. S3 to S7; and table S1) (*18*). The assignment of the open and closed molecular forms was further confirmed by means of low-temperature inelastic electron tunneling spectroscopic measurements and related theoretical analysis (figs. S9 to S12) (*18*).

Our theoretical analysis (supplementary materials, section S9.2) (*18*) shows that the HOMO of diarylethene is the only molecular orbital with energy close enough to the unbiased electrode chemical potentials to enter the electrode Fermi window

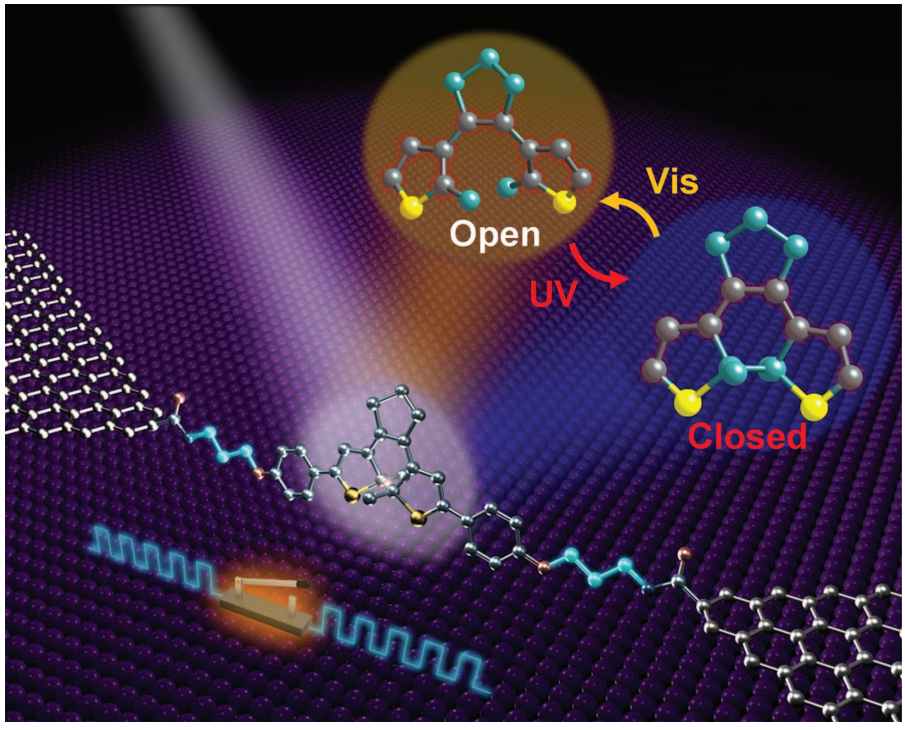


Fig. 1. Schematic of a graphene-diarylethene-graphene junction that highlights the expansion of the molecular bridge by methylene groups.

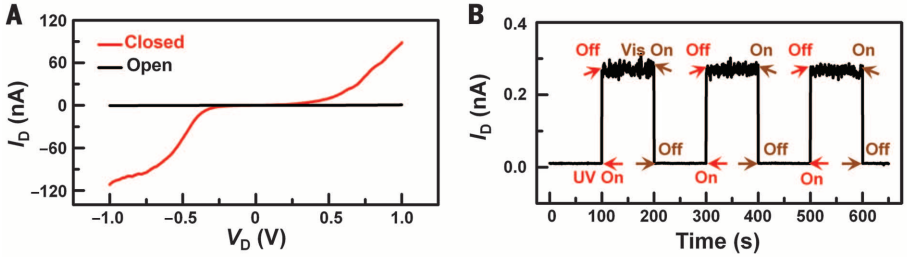


Fig. 2. Reversible photoswitching of graphene-diarylethene-graphene junctions. (**A**) *I-V* characteristics of individual diarylethenes in open (black line) and closed (red line) form at gate voltage V_G = 0 V. V_D, drain voltage; I_D, drain current. (**B**) Real-time measurement of the current through a diarylethene molecule that reversibly switches between the closed and open forms, upon exposure to ultraviolet (UV) and visible (Vis) radiation, respectively. V_D = 100 mV and V_G = 0 V.

in the bias range explored, thus enabling conduction. Assuming a charge-hopping transport mechanism, for the closed-ring molecular bridge at room temperature, our analysis establishes that the threshold bias voltage for the onset of appreciable current is in the range 0 to 0.38 V [as shown in supplementary materials, section S9.2 (*18*), this range is narrowed to 0 to 0.23 V if the computational uncertainties in the HOMO and reorganization energy are limited to the respective standard deviations]. Moreover, our theoretical analysis predicts a minimum threshold bias voltage of 1.2 V for observing appreciable conduction through the open molecular form. The theoretical predictions for both closed and open molecular forms are in agreement with the experimental current-voltage characteristics reported in Fig. 2A, where the onset of the current through the closed molecule occurs at ~0.3 V, whereas no appreciable conduction through the open molecule is observed up to 1 V. Assuming that tunneling mediated by a molecular level is at play and using the elegant theoretical approach by Bâldea *et al.* (*19*), we again find that the offset of the HOMO from the electrode Fermi level is significantly larger for the open form than for the closed form of the diarylethene bridge (fig. S8) (*18*).

If resonant tunneling (as is described by the Landauer-Büttiker formalism) is assumed as the dominant conduction mechanism, as was recently done for other diarylethene junctions (*20*), the threshold bias voltage for the onset of appreciable current is located in the range of 0.3 to 1.0 V, which is still compatible with the characteristics

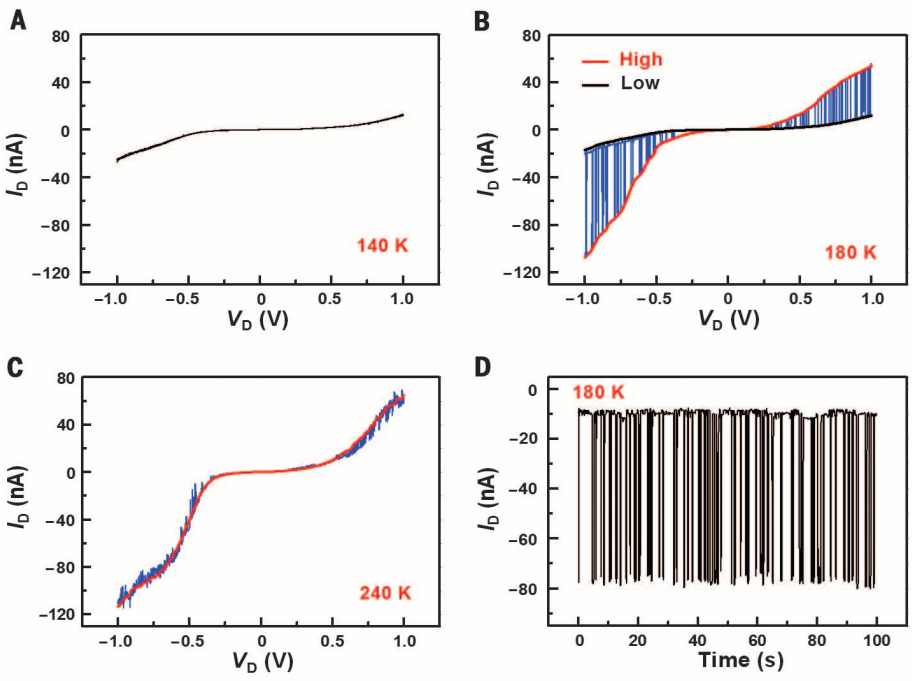


Fig. 3. Temperature-dependent stochastic switching of graphene-diarylethene-graphene junctions. (**A**) *I-V* characteristics of a single diarylethene junction in the closed form screened at 140 K for at least three times, which do not show stochastic switching. (**B**) Stochastic switching between two conductive states at 180 K. (**C**) *I-V* characteristics at 240 K. The stochastic conductance switching starts to disappear at this temperature. (**D**) Real-time recording of stochastic conductance switching at 180 K, with a source-drain bias of −0.8 V and a gate voltage of 0 V. The full details of the temperature-dependent stochastic conductance switching are provided in figs. S13 to S21 (*18*).

in Fig. 2A. A reliable discrimination between the hopping and resonant tunneling mechanisms for conduction requires further experimental and theoretical analysis (table S4 and its discussion) (*18*). In either mechanism, the energy gap between the molecular HOMO and the graphene Fermi level plays a crucial role in determining the threshold bias voltage for conduction, also because of the relatively small reorganization energy, which we have calculated to be $\lambda = (0.27 \pm 0.02)$ eV.

In addition to the photoinduced switching (Fig. 2B), systematic measurements over a wide temperature range revealed stochastic switching of the closed junction between two conductive modes (Fig. 3). This phenomenon was observed only for the closed form in the temperature range of 160 to 240 K (figs. S13 to S21 and tables S2 and S3) (*18*). Although the detailed mechanism of this behavior cannot be decided yet, some conclusions can be reached from our observations and theoretical analysis. This analysis indicates that neither the SOMO (singly occupied molecular orbital) of the molecular cation nor the LUMO of the neutral molecule are involved in this stochastic switching because these levels are much lower and higher in energy, respectively, than the unbiased graphene Fermi level (*21*, *22*). In either the hopping or the resonant tunneling model (*23*), the stochastic switching may be caused by a diarylethene conformational change, possibly driven by transient molecular charging—that is, changing occupation of the molecular HOMO—which is similar to that recently observed in organometallic single-molecule junctions with a redox center that is weakly coupled to the electrodes (*24*). Such a conformational change can lead to a bistable free-energy surface, with relative stability determined by entropy (number of microstates involved) and therefore sensitive to temperature. The current-voltage (*I-V*) characteristics in figs. S14 and S20 (*18*) are consistent with this picture if the two locally stable conformations have different conductions. In a simple model for such behavior—the polaron model of Galperin *et al.* (*23*)—the entropic term results in a temperature-dependent molecular electronic (free) energy, and the relative stability of the two conductive states predicted by the model is inverted as the temperature increases (fig. S27) (*18*). The current measurements do not rule out the possibility that stochastic conductance switching similar to that in Fig. 3 may be observed for the open-form molecular bridge at higher voltages for which the open-form HOMO enters the electrode Fermi window.

Both the photoswitching (Fig. 2) and the stochastic conductance switching (Fig. 3) were consistently observed in many devices for many switching cycles (more than 100 cycles for photoswitching and ~10^5 to 10^6 cycles for stochastic switching). Furthermore, the photoswitches produced in this work have been stable for over a year.

This work shows that molecules with suitable conductive properties and amenable to anchoring to solid-state electrodes can be key components for future nanoelectronics. Molecular engineering aimed to modulate the molecule-electrode coupling strengths has been the key to our success in constructing fully reversible, two-mode, single-molecule electrical switches.

REFERENCES AND NOTES

1. R. Naaman, *Phys. Chem. Chem. Phys.* **13**, 13153–13161 (2011).
2. A. K. Tatikonda, M. Tkachev, R. Naaman, *Biosens. Bioelectron.* **45**, 201–205 (2013).
3. S. Rochat, T. M. Swager, *ACS Appl. Mater. Interfaces* **5**, 4488–4502 (2013).
4. A. Migliore, R. Naaman, D. N. Beratan, *Proc. Natl. Acad. Sci. U.S.A.* **112**, E2419–E2428 (2015).
5. B. G. Mathapa, V. N. Paunov, *Biomater. Sci.* **2**, 212–219 (2014).
6. K. Famm, B. Litt, K. J. Tracey, E. S. Boyden, M. Slaoui, *Nature* **496**, 159–161 (2013).
7. M. T. Meredith, S. D. Minteer, *Annu. Rev. Anal. Chem.* **5**, 157–179 (2012).
8. *Nat. Nanotechnol.* **8**, 377–467 (2013).
9. M. Irie, *Chem. Rev.* **100**, 1685–1716 (2000).
10. B. K. Pathem, S. A. Claridge, Y. B. Zheng, P. S. Weiss, *Annu. Rev. Phys. Chem.* **64**, 605–630 (2013).
11. J. L. Zhang *et al.*, *Chem. Soc. Rev.* **44**, 2998–3022 (2015).
12. D. Dulić *et al.*, *Phys. Rev. Lett.* **91**, 207402 (2003).
13. A. C. Whalley, M. L. Steigerwald, X. Guo, C. Nuckolls, *J. Am. Chem. Soc.* **129**, 12590–12591 (2007).
14. C. Jia *et al.*, *Angew. Chem. Int. Ed.* **52**, 8666–8670 (2013).
15. C. Jia, X. Guo, *Chem. Soc. Rev.* **42**, 5642–5660 (2013).
16. X. Guo *et al.*, *Science* **311**, 356–359 (2006).
17. Y. Cao *et al.*, *Angew. Chem. Int. Ed.* **51**, 12228–12232 (2012).
18. Materials and methods are available as supplementary materials on *Science* Online.
19. I. Bâldea, Z. Xie, C. D. Frisbie, *Nanoscale* **7**, 10465–10471 (2015).
20. T. Sendler *et al.*, *Adv. Sci.* **2**, 1500017 (2015).
21. A. Migliore, A. Nitzan, *ACS Nano* **5**, 6669–6685 (2011).
22. A. Migliore, A. Nitzan, *J. Am. Chem. Soc.* **135**, 9420–9432 (2013).
23. M. Galperin, M. A. Ratner, A. Nitzan, *Nano Lett.* **5**, 125–130 (2005).
24. F. Schwarz *et al.*, *Nat. Nanotechnol.* **11**, 170–176 (2016).

ACKNOWLEDGMENTS

We acknowledge Z. Liu, N. Kang, J. Bergfield, and C. Van Dyck for insightful discussions. This work was supported by the National Basic Research Program of China (2012CB921404 and 2012CB932703) (X.G. and H.Q.X.), the National Natural Science Foundation of China (21225311, 91333102, 21373014, 21190033, 91221202, and 61321001) (X.G., D.-H.Q., and H.Q.X.), the Israel Science Foundation (A.N. and A.M.), the U.S.-Israel Bi-national Science Foundation (A.N.), and the University of Pennsylvania (A.N.).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1443/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S27
Databases S1 to S6
References (*25*–*51*)

5 March 2016; accepted 3 May 2016
10.1126/science.aaf6298

APPLIED PHYSICS

Band structure engineering in organic semiconductors

Martin Schwarze,[1] Wolfgang Tress,[2*] Beatrice Beyer,[3] Feng Gao,[2] Reinhard Scholz,[1,4] Carl Poelking,[5] Katrin Ortstein,[1] Alrun A. Günther,[1] Daniel Kasemann,[1] Denis Andrienko,[5] Karl Leo[1]†

A key breakthrough in modern electronics was the introduction of band structure engineering, the design of almost arbitrary electronic potential structures by alloying different semiconductors to continuously tune the band gap and band-edge energies. Implementation of this approach in organic semiconductors has been hindered by strong localization of the electronic states in these materials. We show that the influence of so far largely ignored long-range Coulomb interactions provides a workaround. Photoelectron spectroscopy confirms that the ionization energies of crystalline organic semiconductors can be continuously tuned over a wide range by blending them with their halogenated derivatives. Correspondingly, the photovoltaic gap and open-circuit voltage of organic solar cells can be continuously tuned by the blending ratio of these donors.

The working principles of semiconductor devices crucially depend on the electronic energy levels of conduction and valence states. This also holds for organic semiconductors, which have recently gained much attention for their application in flexible large-area devices not realizable with traditional inorganic semiconductors. In organic semiconductor devices such as organic solar cells (OSCs), light-emitting diodes (OLEDs), and field-effect transistors (OFETs), an accurate energy-level tuning of electron- and hole-transporting electronic states—represented by the frontier orbitals, the highest occupied (HOMO) and lowest unoccupied molecular orbital (LUMO)—is necessary to optimize device efficiency (*1*–*4*). For instance, the relative energy position of the HOMO of the donor and the LUMO of the acceptor material is paramount in OSCs, as it controls the open-circuit voltage (*1*, *5*). Also, the tuning of the energy level alignment at the electrode/semiconductor interface is crucial for efficient charge injection in OLEDs and OFETs, and extraction in OSCs (*4*, *6*–*8*).

In inorganic semiconductors, band-gap engineering (or more generally, band structure engineering) has opened up a new dimension in device design: By blending materials with different energy levels, it became possible to continuously tune band energies by simply varying the composition of binary, ternary, or quaternary alloys (*9*–*11*).

So far, this design principle has not been available in organic semiconductors. Here, the established means of tuning the electronic levels is by varying the molecular design—for example, by halogenation of organic compounds (*12*–*15*). However, such changes in molecular design require a new synthesis and do not provide the same degree of fine-scale tunability obtained by the blending of inorganic semiconductors. Blending of organic molecules has been applied to the active layer in organic solar cells (*16*, *17*) and for interface layers in OLEDs (*18*, *19*). However, in those experiments, only the electronic states of individual molecules were relevant. Accordingly, blend properties were not used to achieve continuous tuning.

In this study, we show that band structure engineering is possible in organic semiconductors. Despite the strongly localized electronic states of the individual molecules, long-range Coulomb interactions enable continuous tuning in blends (*20*, *21*). As a model material system, we chose zinc phthalocyanine and boron subphthalocyanine chloride (ZnPc and SubPc; Fig. 1A and fig. S3) and their halogenated derivatives (F_4ZnPc, F_8ZnPc, F_{16}ZnPc, and Cl_6SubPc; structural formulas in figs. S1 and S3) (*22*). The energetic, optical, and structural properties of phthalocyanines have already been well investigated in the literature, as also motivated by their good device performance (*23*).

We first performed ultraviolet photoelectron spectroscopy (UPS) measurements on ZnPc, F_4ZnPc, F_8ZnPc, and F_{16}ZnPc thin films to obtain measurements of the ionization energy (IE) of neat layers (fig. S1). The IE increases with higher degrees of fluorination from 5.1 eV for ZnPc to 6.75 eV for F_{16}ZnPc, whereas the shapes of the spectra remain very similar, as reported previously (*24*). This increase of IE in fluorinated compounds has been explained by a stabilization of the HOMO level in compounds with electron-withdrawing ligands (*13*). In ordered thin films, large orientation dependencies of the IE were seen for molecules containing polar bonds (*25*). For highly ordered thin films of pentacene and perfluoropentacene, this was explained by an orientation dependence of the polarization energy originating from different charge-quadrupole interactions (*26*, *27*).

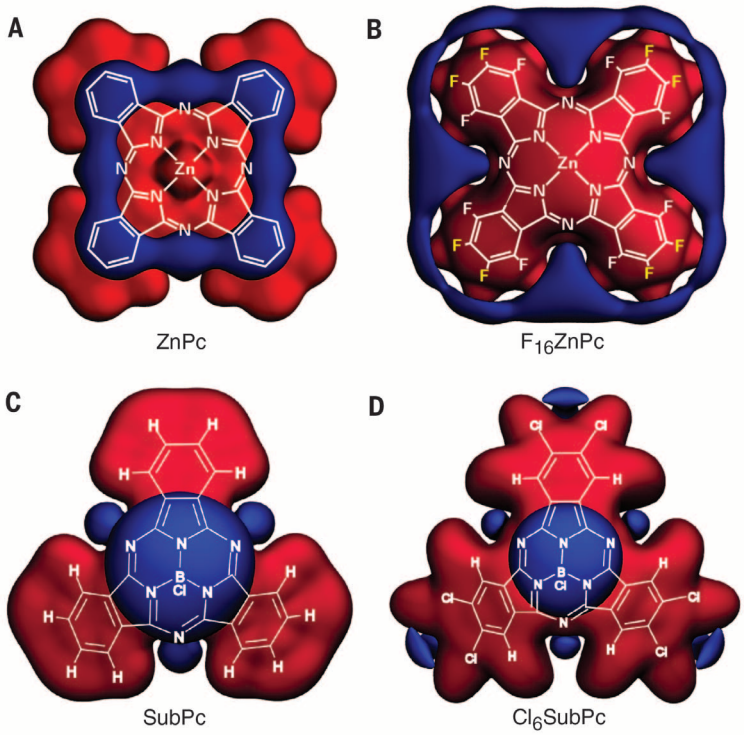


Fig. 1. Molecular structural properties of (halogenated) ZnPc and SubPc. Isopotential surfaces at −0.3 V (blue) and +0.3 V (red) for (**A**) ZnPc, (**B**) F_{16}ZnPc, (**C**) SubPc, and (**D**) Cl_6SubPc calculated via the DFT method B3LYP/6-311+g(d,p).

[1]Institut für Angewandte Photophysik, Technische Universität Dresden, 01069 Dresden, Germany. [2]Biomolecular and Organic Electronics, IFM, Linköping University, 58183 Linköping, Sweden. [3]Fraunhofer Institute for Electron Beam, Plasma Technology and COMEDD, 01109 Dresden, Germany. [4]Dresden Center for Computational Materials Science, Technische Universität Dresden, 01062 Dresden, Germany. [5]Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany.
*Present address: Laboratory of Photonics and Interfaces, Swiss Federal Institute of Technology (EPFL), Station 6, CH 1015 Lausanne, Switzerland. †Corresponding author. Email: leo@iapp.de

By comparing literature values for the IE of ZnPc (6.3 eV) and F_{16}ZnPc (7.2 eV) molecules measured by gas-phase UPS with the IE obtained in this study, we can obtain a first estimation of the relaxation energy contribution to the IE of our solid thin-film samples (*14*). The large difference of ~0.75 eV is rationalized by their opposing quadrupole moments (Fig. 1 and table S3).

Salzmann *et al.* (*28*) showed large shifts in the IE by intermixing of pentacene and perfluoropentacene. However, the IE did not shift linearly with mixing ratio, which was explained by an undesired phase separation in the blend films. In contrast, good intermixing on a molecular scale was shown by x-ray diffraction (XRD) and absorption measurements for a blend of CuPc:F_{16}CuPc (*29*). As shown by comparison of our XRD results on a ZnPc, an F_4ZnPc, and a ZnPc:F_4ZnPc (1:1) blend layer (fig. S2) to previous results, the α-phase crystal structure with an edge-on orientation and fine molecular intermixing in blend layers can be expected (*24*, *29*, *30*). Crystallinity of mixed systems and large differences in quadrupole moments of single molecules are key preconditions for tuning energy levels over a wide range (see below).

UPS results on mixed ZnPc:F_nZnPc layers show a monotonous shift of the HOMO distribution to lower binding energies with increasing ZnPc content, corresponding to a decrease of the IE (Fig. 2, A to C). Whereas the HOMO distributions of neat layers can be fitted by single Gaussian functions, the broader HOMO distributions of the blend layers can only be described by a superposition of two Gaussian peaks (see supplementary materials). The molar ratio of ZnPc and F_nZnPc in the blends, obtained from the intensity ratio of fluorine and zinc emission in x-ray photoelectron spectra, was used as the intensity ratio of the peaks.

A detailed analysis of the HOMO peaks in the blends yields several remarkable results. First, the difference of the IE of ZnPc and F_nZnPc in the blends is considerably smaller than the difference in the neat films. Second, this difference is constant over a broad range of mixing ratios: The IE of ZnPc and F_nZnPc linearly shift to higher (lower) binding energies with increasing content of F_nZnPc (ZnPc) in the blends. Hence, the IE in blends of ZnPc with F_4ZnPc, F_8ZnPc, and F_{16}ZnPc can be continuously shifted over a range of 0.5 eV, 0.86 eV, and 1.17 eV, respectively (Fig. 2, D to F, maximum position of Gaussian fits is used), by simply blending these materials. As the linear IE change with mixing ratio is shared by all systems, with larger degrees of fluorination enhancing the slope, a common physical mechanism can already be anticipated.

Density functional theory (DFT) allows estimation of intermolecular interactions along the stacking direction of molecular crystals (*31*–*33*). Here, we applied DFT calculations to stacks of F_nZnPc molecules with the geometry of the β phase. For model geometries consisting of five stacked molecules, interactions along the stacking direction produce substantial shifts of the HOMO energies. In ZnPc stacks, the inner sites suffer a HOMO destabilization of 0.24 eV on average, whereas in F_8ZnPc stacks the respective states are stabilized by –0.22 eV, and the HOMO energies in F_4ZnPc stacks remain nearly unaffected relative to the gas phase. HOMO shifts on surface molecules of the finite stacks are about half as large.

DFT calculations reproduce an essential finding of the UPS measurements (Fig. 3): In mixed stacks of ZnPc and F_nZnPc, intermolecular interactions result in two distinct groups of HOMO and LUMO energies, one arising from the fluorinated and one from the nonfluorinated sites, but the splitting between these two groups of orbitals is significantly reduced with respect to the pure phases. The energy difference between the two subgroups reaches a value of 0.20 eV for ZnPc/F_4ZnPc and of 0.34 eV for ZnPc/F_8ZnPc, in both cases in excellent agreement with the splitting found in UPS.

The reduced splitting between the orbital energies of fluorinated and nonfluorinated phthalocyanines seems counterintuitive, because any off-diagonal interaction between nondegenerate levels should increase the splitting. Instead, DFT calculations on mixed stacks demonstrate that electrostatic interactions of a molecular orbital with neighboring molecules move the on-site energies of the HOMOs of fluorinated and nonfluorinated ZnPc toward each other. This effect can be rationalized by accounting for electrostatic interactions of a net charge with quadrupole moments of surrounding molecules.

A deficit of the DFT calculations concerns the overall shift of the HOMO energy arising from fluorination: The UPS measurements reveal a shift of 0.7 eV from ZnPc to F_4ZnPc, whereas DFT predicts only 0.46 eV. For a more quantitative description of the ionization energies, an explicit account of the excess charge (hole) is required, achieved here via a perturbative embedding procedure applied to large atomistic models (fig. S8; see supplementary materials for simulation details). Such a coarse-grained approach allows the investigation of much larger molecular assemblies.

Our simulations reveal that the microscopic mechanism responsible for the linear behavior is the superposition of quadrupolar molecular fields of two compounds that act on the excess hole. Indeed, in the absence of intermolecular interactions, the density of states (DOS) of the binary mixture would consist only of two peaks corresponding to the ionization energies of ZnPc and F_nZnPc in the gas phase. Changes in composition would then only shift the relative height of these

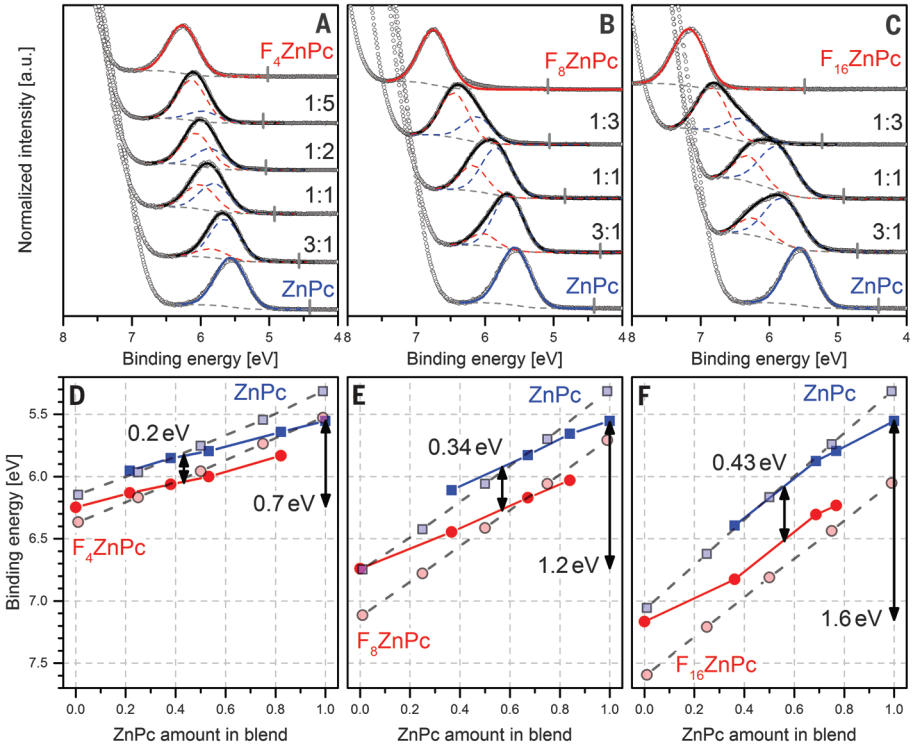


Fig. 2. Thin-film ionization energies at different mixing ratios. UPS spectra (gray circles) and fits (solid lines) for mixed blends of (**A**) ZnPc:F_4ZnPc, (**B**) ZnPc:F_8ZnPc, and (**C**) ZnPc:F_{16}ZnPc. In blends, parts of spectra belonging to ZnPc (blue) and F_nZnPc (red) are obtained from Gaussian fits with Shirley backgrounds (dashed lines). Mixing ratios of the two compounds in the blends are given by the mass ratios obtained from quartz crystals. Fermi level positions are depicted as short gray solid lines. (**D** to **F**) Maximum positions of HOMO distributions of ZnPc (blue squares) and F_nZnPc (red circles) over ZnPc content are shown together with simulation results (dashed lines) of the IE of ZnPc (light blue squares) and F_nZnPc (light red circles). Simulated onset values are shifted by 500 meV, which is the typical difference between the maximum and the onset of the HOMO peak in the spectra.

peaks, but not their positions. The linear dependence on mixing ratio is thus conditioned solely by the external, solid-state contribution determined by the molecular packing, degree of molecular ordering, and molecular architecture. In fact, ZnPc carries a net-negative out-of-π-plane quadrupole component, as opposed to the positive out-of-π-plane quadrupole component that characterizes the F_nZnPc family (see again the isopotential maps of ZnPc and F_{16}ZnPc in Fig. 1, A and B, and table S3). The long-range character of the charge-quadrupole interaction effectively results in a concentration-weighted average over both contributions, which thus serves as the mechanism for level engineering. Note that the long-range electrostatic contribution is virtually homogeneous across the film. Polarization additionally reduces the level offset between ZnPc and F_nZnPc, which accounts for the anomalously small IE offset in the thin film relative to the gas phase. This microscopic view is fostered by the very good agreement between experimentally measured and simulated, composition-dependent IEs (see Fig. 2, D to F, and fig. S9).

To prove the generality of the observations, we performed UPS experiments on blends of SubPc and Cl_6SubPc, a material system with a different symmetry (fig. S3). This system also shows a tuning effect of the ionization energies with mixing ratio and the difference of the IE in the blend is smaller than in neat films. Notably, the IE difference of around 300 meV is similar to the value in ZnPc:F_8ZnPc blends, where the degree of halogenation (i.e., the amount of electron-withdrawing ligands) is comparable. However, the shift of the ionization energies is much smaller than the shift in blends of ZnPc:F_8ZnPc, which already is partly explained by the smaller difference of quadrupole moments of the single molecules (table S3). An orientation deviating from edge-on stacks would decrease the tuning effect even further.

We next address the transport properties of such tunable systems. Previously, a substantial decrease of conductivity in p-doped CBP was seen by admixing molecules with lower ionization energies—that is, by the introduction of traps in the energy gap of CPB (*34*). Because of the energy distance of the HOMO and LUMO levels between the single molecules in the blends, a negative impact on the charge transport properties might be expected. We obtained mobility data for undoped and temperature-dependent conductivity data for n-doped blends (see supplementary materials). The mobilities slightly decrease for the mixed systems but still allow device application (fig. S4). The room-temperature conductivity increases with mixing ratio from the value of n-doped ZnPc to n-doped F_4ZnPc (figs. S5 and S6). Accordingly, the activation energy does not show an increase in the blend layers. Obviously, the transport level shifts with the mixing ratio without significantly impairing the transport properties.

To prove that the continuous shift of energy levels can be exploited in devices, we applied the approach in organic solar cells. In previous studies, a correlation of the open-circuit voltage (V_{oc}) and the effective gap was measured for several donor-acceptor systems. Here, the effective gap is defined as the difference between the IE of the donor and the electron affinity (EA) of the acceptor (*1*, *5*). As tuning of the IE of the donor by intermixing of ZnPc and F_4ZnPc is possible, this should equally apply to the effective gap. We chose to study a ternary bulk heterojunction of the two donors with C_{60} as acceptor. The heterojunction was embedded in a p-i-n–type solar cell (see supplementary materials). The mixing ratio of F_4ZnPc and ZnPc was varied from pure F_4ZnPc to pure ZnPc while the C_{60} content was fixed at 60 weight percent.

Tuning the IE of the donor indeed changes the current-voltage characteristics (Fig. 4A and table S1). The V_{oc} shifts almost linearly to higher values with increasing F_4ZnPc content, as observed in earlier studies on other ternary blends (*35*, *36*). The 300-meV shift of V_{oc} in our solar cells is, however, smaller than expected from UPS measurements on binary ZnPc:F_4ZnPc blends. For that reason, we also performed UPS measurements on ternary blends with varying donor:donor ratio (fig. S9). From pure ZnPc to pure F_4ZnPc, the IE of the donor shifts to higher binding energies by 700 meV (i.e., the same magnitude as in binary

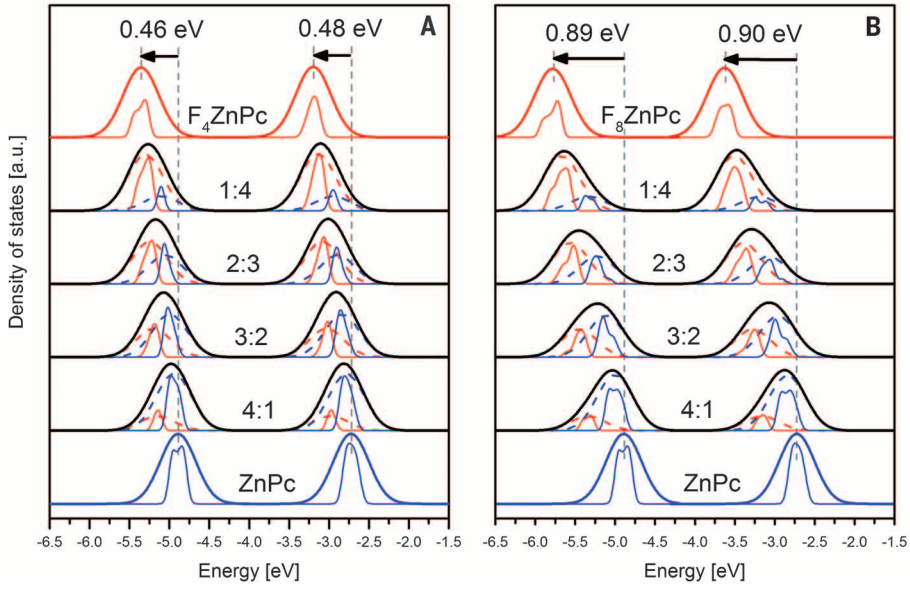


Fig. 3. Calculated DOS of the HOMO and LUMO levels. Mixed stacks of (**A**) ZnPc:F_4ZnPc and (**B**) ZnPc:F_8ZnPc consisting of five molecules according to the geometry of β-ZnPc. The results are averaged over all possible orderings of the stacks, convoluted with a Gaussian broadening of 100 meV (thin solid lines) and 500 meV (dashed lines) for ZnPc (blue) and F_nZnPc (red). Thick solid lines depict the sum of ZnPc and F_nZnPc levels with a broadening of 500 meV.

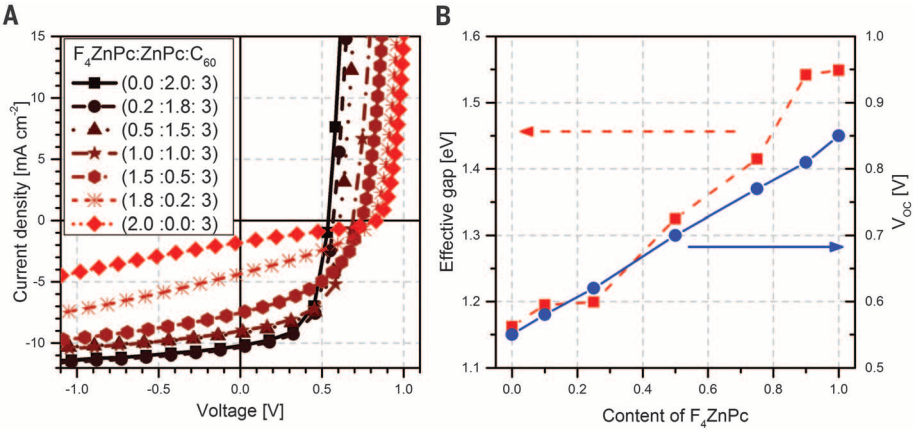


Fig. 4. Solar cell performance of ternary blends. (**A**) Current density–voltage curves of ternary blends with varied donor composition. (**B**) Open-circuit voltage (V_{oc}, blue circles) and effective gap (red squares) obtained by UPS as a function of F_4ZnPc content.

blends). The IE of the acceptor C_{60} also shifts to higher binding energies by 300 meV, mirroring the increase in IE that is driven by the change in long-range quadrupolar fields. Assuming a constant distance between the IE and the EA of 2.4 eV for C_{60} in all ternary blends, the effective band gap for all samples can be calculated (*37*). The relation of V_{oc} and the effective gap deviates from a linear correlation only at higher F_4ZnPc contents (Fig. 4B and table S1).

Tuning of energy levels by superimposing quadrupole fields is expected to work in a variety of semiconducting small molecules and polymers, but a few preconditions appear necessary. Besides the large difference in the magnitude of the quadrupole moments along the thin-film normal, the superposition of their fields must be coherent facilitated by a systematic orientation of both species in the films. A tuning effect can even be realized with constituents of the same polarity, but different orientation in the film, which thus effect a different out-of-plane quadrupole moment. Although the spatial range of this tuning effect and the required degree of molecular intermixing need further investigation, not only bulk tuning but also tuning the energy levels spatially by gradients of the mixing ratios is possible. This may motivate entirely new designs of device architectures for organic semiconductors.

REFERENCES AND NOTES

1. J. Widmer, M. Tietze, K. Leo, M. Riede, *Adv. Funct. Mater.* **23**, 5814–5821 (2013).
2. L. E. Polander *et al.*, *APL Mater.* **2**, 081503 (2014).
3. M. Schober *et al.*, *Phys. Rev. B* **84**, 165326 (2011).
4. M. Barret, S. Sanaur, P. Collot, *Org. Electron.* **9**, 1093–1100 (2008).
5. D. Veldman, S. C. J. Meskers, R. A. J. Janssen, *Adv. Funct. Mater.* **19**, 1939–1948 (2009).
6. J. Hwang, A. Wan, A. Kahn, *Mater. Sci. Eng. Rep.* **64**, 1–31 (2009).
7. N. Koch, A. Elschner, J. Schwartz, A. Kahn, *Appl. Phys. Lett.* **82**, 2281 (2003).
8. W. Tress, K. Leo, M. Riede, *Adv. Funct. Mater.* **21**, 2140–2149 (2011).
9. F. Capasso, *Science* **235**, 172–176 (1987).
10. J. A. Van Vechten, T. K. Bergstresser, *Phys. Rev. B* **1**, 3351–3358 (1970).
11. R. Hill, *J. Phys. C* **7**, 521–526 (1974).
12. M. L. Tang, Z. Bao, *Chem. Mater.* **23**, 446–455 (2011).
13. A. Ghosh, P. G. Gassman, J. Almloef, *J. Am. Chem. Soc.* **116**, 1932–1940 (1994).
14. D. Schlettwein *et al.*, *J. Phys. Chem. B* **105**, 4791–4800 (2001).
15. F. Anger *et al.*, *J. Chem. Phys.* **136**, 054701 (2012).
16. M. Hiramoto, H. Fujiwara, M. Yokoyama, *Appl. Phys. Lett.* **58**, 1062 (1991).
17. G. Yu, J. Gao, J. C. Hummelen, F. Wudl, A. J. Heeger, *Science* **270**, 1789–1791 (1995).
18. J. Kido, K. Hongawa, K. Okuyama, K. Nagai, *Appl. Phys. Lett.* **64**, 815 (1994).
19. S. Reineke *et al.*, *Nature* **459**, 234–238 (2009).
20. C. Poelking *et al.*, *Nat. Mater.* **14**, 434–439 (2015).
21. C. Poelking, D. Andrienko, *J. Am. Chem. Soc.* **137**, 6320–6326 (2015).
22. H. Brinkmann *et al.*, *Phys. Status Solidi* **205**, 409–420 (2008).
23. J. Meiss *et al.*, *Adv. Funct. Mater.* **22**, 405–414 (2012).
24. M. Brendel *et al.*, *Adv. Funct. Mater.* **25**, 1565–1573 (2015).
25. S. Duhm *et al.*, *Nat. Mater.* **7**, 326–332 (2008).
26. H. Yoshida, K. Yamada, J. Tsutsumi, N. Sato, *Phys. Rev. B* **92**, 075145 (2015).
27. B. J. Topham, Z. G. Soos, *Phys. Rev. B* **84**, 165405 (2011).
28. I. Salzmann *et al.*, *J. Am. Chem. Soc.* **130**, 12870–12871 (2008).
29. A. Opitz *et al.*, *Org. Electron.* **10**, 1259–1267 (2009).
30. C. Schünemann *et al.*, *Thin Solid Films* **519**, 3939–3945 (2011).
31. Y. C. Cheng *et al.*, *J. Chem. Phys.* **118**, 3764 (2003).
32. L. Gisslén, R. Scholz, *Phys. Rev. B* **80**, 115309 (2009).
33. N. Marom, A. Tkatchenko, M. Scheffler, L. Kronik, *J. Chem. Theory Comput.* **6**, 81–90 (2010).
34. J.-H. Lee *et al.*, *Org. Electron.* **15**, 16–21 (2014).
35. P. P. Khlyabich, B. Burkhart, B. C. Thompson, *J. Am. Chem. Soc.* **133**, 14534–14537 (2011).
36. P. P. Khlyabich, B. Burkhart, B. C. Thompson, *J. Am. Chem. Soc.* **134**, 9074–9077 (2012).
37. M. L. Tietze *et al.*, *Phys. Rev. B* **88**, 085119 (2013).

ACKNOWLEDGMENTS

We thank O. Inganäs for his support and insightful discussions, D. Wöhrle for supplying F_8ZnPc, O. Kaveh for conductivity measurements, D. Schütze for building OFETs, L. Wilde for performing GIXRD measurements, and F. Holzmüller for evaluating and discussing GIXRD measurements. The research in Dresden was funded by the DFG project MatWorldNet LE-747/44-1, as well as the European Community's Seventh Framework Programme under grant agreement FP7-267995 (NUDEV). The research at Linköping was supported by the Knut and Alice Wallenberg Foundation through a Wallenberg Scholar grant to O. Inganäs, the Swedish Research Council (VR, 330-2014-6433), and the European Commission Marie Skłodowska-Curie Actions (INCA 600398). Also supported by Bundesministerium für Bildung und Forschung project MEDOS grant FKZ 03EK3503B (C.P. and D.A.) and by the Dr. Isolde-Dietrich-Stiftung (A.A.G.). K.L. is a fellow of the Canadian Institute for Advanced Research (CIFAR). Author contributions: M.S. and K.O. acquired UPS data; W.T., B.B., and F.G. acquired and evaluated solar cell data; M.S. evaluated UPS and conductivity data; R.S. performed DFT calculations; C.P. performed molecular simulations; A.A.G. obtained mobility values; M.S., K.L., C.P., R.S., and D.A. wrote the manuscript; all authors contributed to discussions and finalizing the manuscript. There are no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1446/suppl/DC1
Materials and Methods
Figs. S1 to S11
Tables S1 to S3
References (*38*–*49*)

11 December 2015; accepted 13 April 2016
10.1126/science.aaf0590

ASTROCHEMISTRY

Discovery of the interstellar chiral molecule propylene oxide (CH_3CHCH_2O)

Brett A. McGuire,[1,2]*† P. Brandon Carroll,[2]*† Ryan A. Loomis,[3] Ian A. Finneran,[2] Philip R. Jewell,[1] Anthony J. Remijan,[1] Geoffrey A. Blake[2,4]

Life on Earth relies on chiral molecules—that is, species not superimposable on their mirror images. This manifests itself in the selection of a single molecular handedness, or homochirality, across the biosphere. We present the astronomical detection of a chiral molecule, propylene oxide (CH_3CHCH_2O), in absorption toward the Galactic center. Propylene oxide is detected in the gas phase in a cold, extended molecular shell around the embedded, massive protostellar clusters in the Sagittarius B2 star-forming region. This material is representative of the earliest stage of solar system evolution in which a chiral molecule has been found.

The origin of homochirality is a key mystery in the study of our cosmic origins (*1*). Although homochirality is itself evolutionarily advantageous (*2*), the mechanism for the selection of one iso-energetic enantiomer over another is uncertain. Many routes to homochirality have been proposed through the amplification and subsequent transfer of a small primordial enantiomeric excess (e.e.). Disentangling these possible mechanisms requires that we understand the potential sources from which an e.e. may arise. The oldest material on which e.e. data have been taken in the laboratory are meteoritic samples (*3*), yet the provenance of this e.e. remains a matter of considerable debate (*4*). Material in molecular clouds from which planetary systems form is processed through circumstellar disks (*5*) and can subsequently be incorporated into planet(esimal)s (*6*). Thus, a primordial e.e. found in the parent molecular cloud may be inherited by the fledgling system. Constraining the origin of e.e.s found in meteorites therefore requires the determination of the possible contributions of primordial e.e.s and thus the detection of a chiral molecule in these environments.

For the past 50 years, radio astronomy has been the primary method for studying the gaseous, complex molecular content of interstellar clouds. In this regime, observed spectral features correspond to fine-structure transitions of atoms, or pure rotational transitions of polar molecules, that can uniquely identify their carrier. The observations presented here were taken toward the Sagittarius B2 North [Sgr B2(N)] molecular cloud, the preeminent source for new complex-molecular detections in the interstellar medium (ISM).

Propylene oxide (Fig. 1) was initially detected using data from the publicly available Prebiotic

[1]National Radio Astronomy Observatory, Charlottesville, VA 22903, USA. [2]Division of Chemistry and Chemical Engineering, California Institute of Technology Pasadena, CA 91125, USA. [3]Department of Astronomy, Harvard University, Cambridge, MA 02138, USA. [4]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA.
*These authors contributed equally to this work. **†Corresponding author. Email: bmcguire@nrao.edu (B.A.M.); pcarroll@caltech.edu (P.B.C.)**

Interstellar Molecular Survey (PRIMOS) project at the Green Bank Telescope (GBT), which provides nearly frequency-continuous, high-resolution, high-sensitivity spectral survey data toward Sgr B2(N) from 1 to 50 GHz (*7*). Based on our model of rotationally cold propylene oxide absorbing against the Sgr B2(N) continuum (*8*), only three transitions are predicted to have appreciable intensity above the survey noise floor: the b-type Q-branch $1_{1,0} - 1_{0,1}$, $2_{1,1} - 2_{0,2}$, and $3_{1,2} - 3_{0,3}$ transitions at 12.1, 12.8, and 14.0 GHz (λ = 2.478, 2.342, and 2.141 cm), respectively (*8*). The $1_{1,0} - 1_{0,1}$ line at 12.1 GHz is obscured by radio frequency interference (RFI) at the GBT; however, clear absorption signatures are observed from the $2_{1,1} - 2_{0,2}$ and $3_{1,2} - 3_{0,3}$ transitions (Fig. 2).

These features may be sufficient for a detection on their own at these wavelengths; however, we endeavored to confirm the detection by observing the $1_{1,0} - 1_{0,1}$ line at 12.1 GHz using the Parkes Radio Telescope (see the supplementary materials for a detailed description of the Parkes observations), which does not suffer from RFI in the region of the line. The data confirm the presence of a feature at the same velocity (~64 km s^{-1}) as the transitions from PRIMOS, as well as fortuitously detecting a nearby feature of propanal, a structural isomer of propylene oxide (Fig. 2). The far larger Parkes beam (~115″ versus 60″) encompasses a much larger sample of environments, inhomogeneously broadening the observed transition and incorporating a second, distinct ~46 km s^{-1} component not seen by the GBT beam but previously observed in the material surrounding Sgr B2 (Fig. 3) (*9*).

A fit to the observations using a single excitation temperature for propylene oxide finds a column density of $N_T = 1 \times 10^{13}$ cm^{-2} and a rotational excitation temperature of $T_{ex} = 5$ K (*8*). Although an excitation temperature of 5 K is indeed the best fit to the data, we note that the most rigorous constraint on T_{ex} is from the nondetection at higher frequencies in PRIMOS, giving an upper limit of ~35 K. Changes in T_{ex} greatly affect N_T, and model parameters that fit the data nearly as well are possible for excitation conditions between $T_{ex} = 5$ and 35 K. These models all reproduce the observed features from the GBT and Parkes and are consistent with the nondetection of propylene oxide at 3 mm; under these conditions, no transitions of propylene oxide would be detectable in the reported observations (*10*).

A search of spectral line catalogs reveals no reasonable interfering transitions from other molecular species. Propylene oxide is an asymmetric rotor with modest rotational constants and therefore has numerous (~450) transitions that fall within the PRIMOS data. For lower excitation temperatures (T_{ex} ~ 10 to 35 K), at most 80 have an intensity ≥1% of the strongest predicted line. Of these, ~13% are unobservable due to a lack of available receivers at the GBT. Inspection of the entire PRIMOS data set showed no absorption or emission features attributable to propylene oxide at any of these frequencies except the three listed above, in good agreement with the model and the sensitivity of the survey.

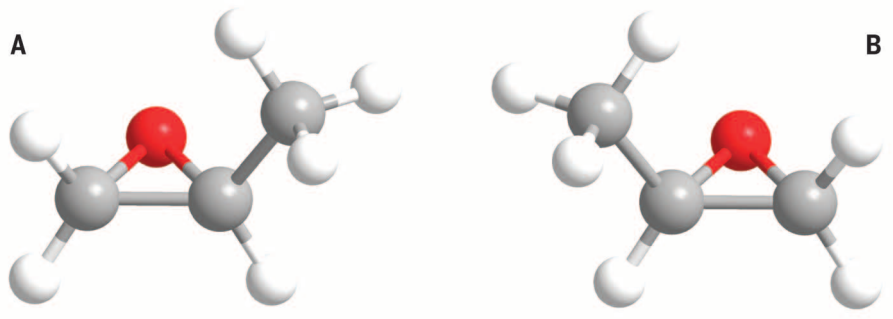


Fig. 1. The molecular structure of *S*-propylene oxide and *R*-propylene oxide. (**A**) *S*-propylene oxide. (**B**) *R*-propylene oxide. Carbon, hydrogen, and oxygen atoms are indicated by gray, small white, and red spheres, respectively.

This detection is complementary to the upper limit placed by (*10*) on the nondetection of warm, compact propylene oxide at T_{ex} = 200 K toward Sgr B2(N) at mm wavelengths using the Mopra Telescope. This search was sensitive only to a warm population of propylene oxide, however, and resulted in a nondetection, with an upper limit column density of 6.7×10^{14} cm^{-2} for an excitation temperature of T_{ex} = 200 K and compact source size (5″), such as that expected for gas associated with the embedded protostellar clusters/hot cores in this cloud (*10*).

In sources with strong background continuum, of which Sgr B2(N) is a prominent example, many rotationally cold, high-dipole moment species are observed almost exclusively in absorption against the continuum source, as shown in Fig. 3. Because of the exceptionally low line densities, only two to five well-measured centimeter-wavelength spectral features are needed to securely claim a detection [see, e.g., (*11*–*13*)]. This stands in stark contrast to mm-wave detections, particularly toward Sgr B2(N), where dozens of lines are typically required. Based on a statistical analysis of the line density in our observations of Sgr B2(N), we find the likelihood of three random features falling within three resolution elements of the propylene oxide transitions to be ≤6×10^{-8} (*8*).

Taken together, the GBT and Parkes observations provide strong evidence of cold, low-abundance propylene oxide toward the Sgr B2 cloud complex, in excellent agreement with previously established upper limits, as well as with previous observations of complex organic molecules. Indeed, many of the complex organics seen toward Sgr B2(N) are found not in or near the hot cores, but, like propylene oxide, in a cold, extended shell around the source. In these regions, molecules are often liberated into the gas phase via nonthermal, shock-driven desorption, resulting in colder, spatially extended gas-phase populations that are often more abundant than predicted by standard warm-up models (*14*). This is consistent with the observation that the structurally similar ethylene oxide is consistently found to have low excitation temperatures (11 to 35 K), well below the temperature of the surrounding grains (*15*), with the detections of glycolaldehyde (*11*), ethanimine (*16*), and propylene oxide's structural isomers propanal (*13*) and acetone (*8*), in this region, and with the general pattern of shock-driven liberation

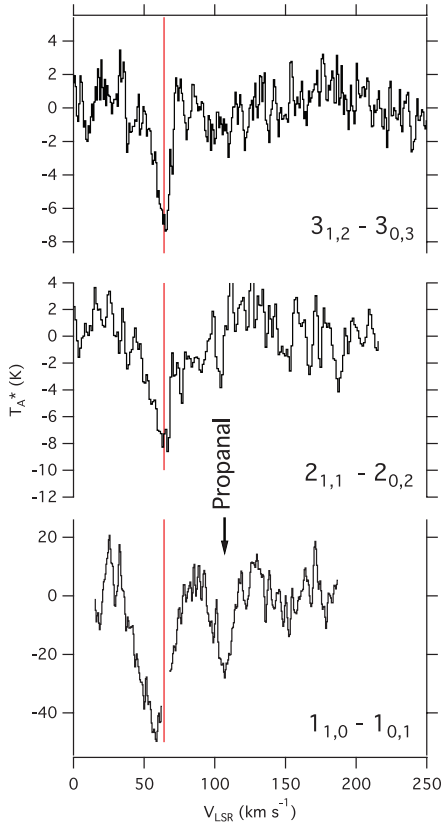


Fig. 2. Observations of the $1_{1,0} - 1_{0,1}$ (Parkes), $2_{1,1} - 2_{0,2}$ (GBT), and $3_{1,2} - 3_{0,3}$ (GBT) transitions of propylene oxide, in absorption, toward the Galactic center. The 64 km s^{-1} systematic velocity characteristic of Sgr B2(N) is indicated by a vertical red line. The $1_{1,0} - 1_{0,1}$ transition of propanal is also seen in the Parkes data. Data are given as antenna temperature (T_A^*) as a function of shift from local standard of rest velocity (V_{LSR}), where 0 km s^{-1} is the measured laboratory frequency of the transition (*8*), and have been Hanning smoothed.

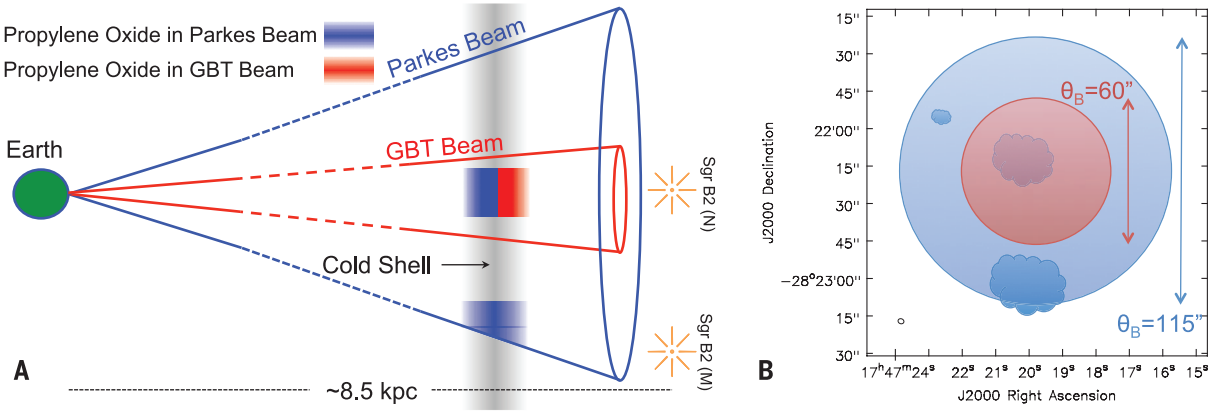


Fig. 3. Illustration of source structure within the Sgr B2 region. (**A**) The GBT and Parkes beams probe different portions of the cold molecular shell in front of the bright continuum sources/hot cores within Sgr B2. Molecules in the shell that are not backlit by continuum sources are not seen in absorption. (**B**) As the schematic of the sky view shows, the GBT (red) and Parkes (blue) beams probe different continuum sources, with the GBT beam probing only Sgr B2(N), whereas the Parkes beam also includes most of Sgr B2(M) to the south.

of complex molecules in the so-called central molecular zone (*14*).

From a chemical perspective, the presence of propylene oxide in Sgr B2(N) is not surprising. Propylene oxide is the third species of the C_3H_6O family detected toward this source. Its structural isomers, propanal (CH_3CH_2CHO) (*13*) and acetone [$(CH_3)_2CO$] (*17*), are both seen toward Sgr B2(N), and propylene oxide is not the first epoxide found in the ISM. Ethylene oxide (CH_2OCH_2) is structurally similar to propylene oxide, differing by only a methyl group, and has been detected toward numerous massive star-forming regions, including Sgr B2(N) (*15*, *18*). In the case of acetone, Belloche *et al.* (*19*) report a column density of $N_T = 1.49 \times 10^{17}$ cm^{-2}, but for a warm population with T_{ex} = 100 K that peaks at the position of the hot core. In the detection of propanal, column densities were not determined (*13*).

To determine the relative populations of these molecules in the cold shell around Sgr B2(N), we have used the same procedure as for propylene oxide (*8*). We find that a column density of $N_T = 6 \times 10^{13}$ cm^{-2} with T_{ex} = 6.2 K reproduces the 11 propanal transitions observed in the full PRIMOS data set (see the supplementary materials) to within a factor of ~2. Similarly, using 18 detected lines of acetone in PRIMOS (see the supplementary materials), we find that a column density of $N_T = 2.1 \times 10^{14}$ cm^{-2} with T_{ex} = 6.2 K reproduces these features within a factor of ~2. The best-fit T_{ex} = 5 K for propylene oxide is in remarkably good agreement with these values, which, due to the larger number of observed transitions over a wider frequency range for propanal and acetone, are much more rigorously constrained. T_{ex} up to 35 K for propylene oxide is formally allowed in the propylene oxide fit, due to loose constraints stemming from the narrow range of energy levels covered in a narrow frequency window. However, the best-fit, T_{ex} = 5 K, is greatly bolstered by the similar conditions exhibited by the acetone and propanal populations.

All three members of the C_3H_6O family are then detected in absorption in the PRIMOS data at remarkably similar excitation conditions, suggesting that they likely occupy the same cold, shocked region surrounding Sgr B2(N). Propanal and acetone are thermodynamically favored over propylene oxide, residing 22.7 and 30.8 kcal mol^{-1} (0.98 and 1.33 eV) lower in energy, respectively (*20*). However, although the relative column densities derived here do roughly follow the pattern of increasing abundance with increasing stability, chemistry in molecular clouds is largely kinetically controlled, rather than thermodynamically, and relative abundances do not regularly follow thermodynamic patterns (*21*, *22*). The recent detection of acetone and propanal at an abundance ratio of three to one in comet 67P/Churyumov-Gerasimenko shows that members of the C_3H_6O family also feature prominently in the volatile organic content of comet nuclei, and the remarkably similar ratios to those observed toward Sgr B2(N) suggest that such kinetically controlled routes to both species are widespread and not isolated to extraordinary interstellar sources (*23*).

The leading models for the production and enhancement of an e.e. in the interstellar medium likely act over time scales far longer than the delivery of complex organic material to the planet-forming region of disks (*24*–*26*). A number of mechanisms have been proposed for gas-phase routes in the ISM to create such a primordial e.e. Although beta decay–related chemistry has been proven to generate slight chiral asymmetries (*25*) that would be universal in nature, perhaps the most intriguing route, astronomically, is enantiomerically selective photochemistry induced by circularly polarized light (CPL) (*24*). Here, the chirally sensitive chemical reaction networks would be stochastically driven on the spatial scales of giant molecular cloud complexes. Toward the Orion Nebula cluster, for example, significant CPL patterns capable of producing e.e. do not extend over the entire protostellar cluster but have been detected over regions that are large compared with individual protoplanetary disks (*26*). We have rigorously examined the possible mechanisms for determining an e.e. (see the supplementary materials) and concluded that the standard, total power observations shown here cannot determine whether such an e.e. exists in the case of propylene oxide but that high-precision, full-polarization-state measurements can, in principle. Critically, the detection of gas-phase propylene oxide toward the Galactic center provides a molecular target for such observations and demonstrates that interstellar chemistry can reach sufficient levels of complexity to form chiral species in environments with the physical conditions required to produce an enantiomeric excess.

REFERENCES AND NOTES

1. S. Pizzarello, T. L. Groy, *Geochim. Cosmochim. Acta* **75**, 645–656 (2011).
2. J. Lunine, *Astrobiology: A Multi-Disciplinary Approach* (Pearson Education Inc., San Francisco, CA, 2005).
3. M. H. Engel, S. A. Macko, *Nature* **389**, 265–268 (1997).
4. D. P. Glavin, J. P. Dworkin, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5487–5492 (2009).
5. L. I. Cleeves *et al.*, *Science* **345**, 1590–1593 (2014).
6. C. F. Chyba, P. J. Thomas, L. Brookshaw, C. Sagan, *Science* **249**, 366–373 (1990).
7. J. L. Neill *et al.*, *Astrophys. J.* **755**, 153 (2012).
8. Materials and methods are available as supplementary materials on *Science* Online.
9. P. A. Jones *et al.*, *Mon. Not. R. Astron. Soc.* **386**, 117–137 (2008).
10. M. R. Cunningham *et al.*, *Mon. Not. R. Astron. Soc.* **376**, 1201–1210 (2007).
11. J. M. Hollis, P. R. Jewell, F. J. Lovas, A. Remijan, *Astrophys. J.* **613**, L45–L48 (2004).
12. B. A. McGuire *et al.*, *Astrophys. J.* **758**, L33 (2012).
13. J. M. Hollis, P. R. Jewell, F. J. Lovas, A. Remijan, H. Møllendal, *Astrophys. J.* **610**, L21–L24 (2004).
14. M. A. Requena-Torres *et al.*, *Astron. Astrophys.* **455**, 971–985 (2006).
15. M. Ikeda *et al.*, *Astrophys. J.* **560**, 792–805 (2001).
16. R. A. Loomis *et al.*, *Astrophys. J.* **765**, L9 (2013).
17. L. E. Snyder *et al.*, *Astrophys. J.* **578**, 245–255 (2002).
18. J. E. Dickens *et al.*, *Astrophys. J.* **489**, 753–757 (1997).
19. A. Belloche, H. S. P. Müller, K. M. Menten, P. Schilke, C. Comito, *Astron. Astrophys.* **559**, A47 (2013).
20. A. Lifshitz, C. Tamburu, *J. Phys. Chem.* **98**, 1161–1170 (1994).
21. E. Herbst, E. F. van Dishoeck, *Annu. Rev. Astron. Astrophys.* **47**, 427–480 (2009).

22. R. A. Loomis *et al.*, *Astrophys. J.* **799**, 34 (2015).
23. F. Goesmann *et al.*, *Science* **349**, aab0689 (2015).
24. P. Modica *et al.*, *Astrophys. J.* **788**, 79 (2014).
25. J. M. Dreiling, T. J. Gay, *Phys. Rev. Lett.* **113**, 118103 (2014).
26. J. Bailey *et al.*, *Science* **281**, 672–674 (1998).

ACKNOWLEDGMENTS

We thank the PRIMOS team, GBT, and Parkes staff for their ongoing support in acquiring the GBT and Parkes data and S. Breen, S. Mader, and J. Reynolds for assistance with Parkes data reduction. We acknowledge the support of L. Snyder, J. M. Hollis, and F. Lovas. B.A.M. thanks J. Mangum and J. Corby for helpful discussions. P.B.C. and B.A.M. acknowledge the support of a NASA Astrobiology Institute Early Career Collaboration Award. B.A.M. is funded by a National Radio Astronomy Observatory Jansky Postdoctoral Fellowship. R.A.L. and I.A.F. are funded by a National Science Foundation Graduate Research Fellowship. P.B.C., I.A.F., and G.A.B. acknowledge support from the NASA Astrobiology Institute through the Goddard Team (M. J. Mumma, PI) under Cooperative Research Agreements NNX09AH63A and NNX15AT33A (NNX09AH63A), and the NSF Astronomy and Astrophysics (AST-1109857) grant program. Access to the entire PRIMOS data set, specifics on the observing strategy, and overall frequency coverage information is available at www.cv.nrao.edu/PRIMOS/. The spectra obtained with Parkes are available through this website as well. Data from project AGBT06B-006 are available in the NRAO Archive at https://science.nrao.edu/observing/data-archive. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. The Australia Telescope Compact Array (Parkes Radio Telescope/Mopra Radio Telescope/Long Baseline Array) is part of the Australia Telescope National Facility, which is funded by the Australian government for operation as a National Facility managed by the Commonwealth Scientific and Industrial Research Organisation (CSIRO).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1449/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S4
References (*27*–*47*)

31 December 2015; accepted 11 May 2016
10.1126/science.aae0328

FOREST ECOLOGY

Northeastern North America as a potential refugium for boreal forests in a warming climate

L. D'Orangeville,[1,2]* L. Duchesne,[3] D. Houle,[3,4] D. Kneeshaw,[1] B. Côté,[5] N. Pederson[6]

High precipitation in boreal northeastern North America could help forests withstand the expected temperature-driven increase in evaporative demand, but definitive evidence is lacking. Using a network of tree-ring collections from 16,450 stands across 583,000 km^2 of boreal forests in Québec, Canada, we observe a latitudinal shift in the correlation of black spruce growth with temperature and reduced precipitation, from negative south of 49°N to largely positive to the north of that latitude. Our results suggest that the positive effect of a warmer climate on growth rates and growing season length north of 49°N outweighs the potential negative effect of lower water availability. Unlike the central and western portions of the continent's boreal forest, northeastern North America may act as a climatic refugium in a warmer climate.

The boreal forest biome is responsible for ~20% of the total carbon (C) sequestrated annually by forest ecosystems (*1*) and contains a large fraction of the planet's remaining unmanaged forests (*2*). Over the current century, this ecosystem is expected to undergo one of the largest increases in temperatures (*3*). Low mean annual temperatures (MATs) are a major constraint on boreal forest productivity, and increases in temperature and growing season length in recent decades are reported to have benefited tree growth over large areas of Fennoscandia (*4*) and Russia (*5*). In central and western areas of North American boreal forests, however, analysis of recent tree-ring width data (*6*, *7*), in combination with stable carbon isotope analysis (*8*, *9*) and satellite images (*10*), suggests that changes in soil water balance have canceled potential gains in forest growth from recent warming. An increase in water deficit in these regions is suggested to have increased tree mortality in recent decades (*11*).

The boreal forest of northeastern North America (NENA) receives more than twice the mean annual precipitation (MAP) as its central and western counterparts, typically exceeding the mean annual evapotranspiration demand (MAE). This pattern results in a strong west-east gradient in water availability (Fig. 1A), which should continue in the future, because climate projections suggest that NENA could be the only area across the circumboreal forest with sufficient precipitation to fully compensate for increasing evaporative stress induced by warmer temperatures (*2*). Future increases in temperature combined with an earlier snowmelt, leading to an earlier start to the growing season, could thus allow local tree species of NENA to better withstand and even thrive in a warmer climate. Although satellite-driven measures generally support this hypothesis for NENA (*12*, *13*), empirical studies and climate-growth models have not yet reached a consensus (*14*–*18*), and definitive evidence from well-replicated large-scale tree growth data is lacking. Should water availability become a chronic factor limiting tree growth and survival across much of the boreal forest, NENA could act as a climatic refugium in the near future for black spruce [*Picea mariana* (Mill.) B.S.P.], the most abundant tree species throughout the North American boreal forest (*19*), including our study region (fig. S1).

Using a network of tree-ring collections from 16,450 stands representing much of the boreal forest east of 80°W (Fig. 1B), our objective was to determine how water availability and air temperature control interannual variations in the growth of black spruce (*20*). To do so, we used bootstrapped correlations to measure the influence of seasonal climate on radial growth for the period 1960–2004. Should the dominant climatic constraint on growth be low water availability, growth should be positively correlated with precipitation and soil moisture and, because evaporative demand increases with temperature, negatively correlated with temperature. In contrast, opposite results would support the hypothesis that low temperatures are the dominant climatic growth constraint. By determining the climatic sensitivity of the annual growth of 26,697 trees across a 583,000-km^2 area in boreal NENA, we had the potential to resolve the discrepancy between empirical studies and model forecasts on the fate of these forests under anthropogenic climate change. To this end, individual tree-ring series were standardized to emphasize interannual variations due to climatic variables and were averaged according to landscape unit and soil type (Fig. 1C).

Two strikingly different types of tree growth response to seasonal climate were detected by bootstrapped correlations (Fig. 2). The first response, dominant across the northern part of the study area, indicates that an increase in temperature with a concomitant reduction in available water has positive effects on growth. The second response, more frequent in forests south of 49°N, indicates that an increase in water availability rather than temperature has a positive effect on growth.

Support for the low-temperature constraint hypothesis is observed in summer (July and August), winter (November to April), and to a lesser extent prior-year fall (September and October) (Fig. 2). Out of the 109 chronologies, hereafter called forests, which display significant growth correlations

[1]Centre d'Étude de la Forêt, Université du Québec à Montréal, Case Postale 8888, Succursale Centre-Ville, Montreal, Quebec H3C 3P8, Canada. [2]Department of Biology, Indiana University, 1001 East 3rd Street, Jordan Hall 142, Bloomington, IN 47405-7005, United States. [3]Direction de la Recherche Forestière, Ministère des Forêts, de la Faune et des Parcs du Québec, 2700 Einstein, Québec City, Quebec G1P 3W8, Canada. [4]Consortium sur la Climatologie Régionale et l'Adaptation aux Changements Climatiques (Ouranos), 550 Sherbrooke W, Montreal, Quebec H3A 1B9, Canada. [5]Department of Natural Resource Sciences, McGill University, 21,111 Lakeshore Road, Sainte-Anne-de-Bellevue, Quebec H9X 3V9, Canada. [6]Harvard Forest, 324 North Main Street, Petersham, MA 10366, USA.
***Corresponding author. Email: loicdorangeville@gmail.com**

($P < 0.05$) with summer climate, a majority (92%), distributed at all latitudes, correlate either positively with temperature, negatively with precipitation, or negatively with a soil moisture index (SMI). Exceptions are concentrated around the Gaspé Peninsula on the south shore of the Saint Lawrence River and in southwestern Québec, where these regions have higher summer temperatures or lower precipitation than other sites at a similar latitude (fig. S2). Significant growth correlations with winter temperatures are positive in 54 out of 57 forests (Fig. 2). Finally, 18 out of 19 significant growth correlations with prior-fall temperatures are positive.

On the other hand, low water availability during the prior-year summer constrains growth, because 107 of the 109 forests with significant growth correlations with climate displayed either negative correlations with prior-summer temperature or positive correlations with prior-summer precipitation and SMI (Fig. 2). It should be noted that the contradictory patterns reported for current-year and prior-year summer are only observed in 8% of the same forests.

Current-summer temperature has the highest number of significant correlations with growth (67), 97% of which are positive. This largely positive trend is intuitive, given that the temperature of high-latitude forests typically limits growth rates. Across seasons, however, a majority of significant correlations with growth (76%) are related to spring, summer, fall, or winter climate before wood formation (Fig. 2). These correlations consist in lagged effects, which are common in boreal forests of North America and Europe (*16*, *21*). High winter snowfall has been reported to delay the spring onset of growth in Eurasia (*22*), which is a primary driver of the productivity of boreal forests (*23*, *24*). In turn, spring snowmelt controls soil water saturation and temperature (*25*). That is, an early snowmelt hastens the drainage of excess water and allows soils to warm up faster at the beginning of the growing season. These processes have positive effects on black spruce growth (*26*, *27*). The large-scale positive growth response to drier springs and low soil moisture for the areas north of 49°N supports these mechanisms. Reduced spring precipitation could also have benefited tree growth indirectly by being associated with higher solar radiation, which has been observed to limit photosynthesis in mature black spruce forests (*24*). At high latitudes, spring precipitation may still include some snowfall, because spring snowmelt averaged over 1960–2004 was completed by May 29 at the northern margin of the study area. Thus, the winter mechanisms mentioned above may also occur in the spring.

The positive growth response to high winter temperatures observed over the entire study area has been previously reported for black spruce (*15*) and other conifers at the northern margin of their distribution in the northeastern United States (*28*). It is probably associated with a reduction in winter desiccation injuries (*29*). Prior-year climate also influences growth in various ways: Significant amounts of nonstructural carbohydrates (NSCs) created in prior years are used for growth (*30*); weather conditions toward the end of the growing season can stimulate water recharge in trees, a prerequisite to the following spring growth onset (*31*, *32*); and black spruce buds formed during the previous growing season already contain a determined number of needle primordia, which will have a determinant effect on leaf area, transpiration, photosynthetic capacity, and ultimately growth. Our results support prior work underscoring the importance of all seasons when forecasting the boreal forest response to climate change (*17*).

From our broadly distributed network, clear spatial patterns in climatic sensitivity indicate that high temperatures and low water availability primarily constrain growth south of 49°N. Below 49°N, we find 95% (n = 18 forests) of all significant positive correlations with precipitation or SMI during prior-year spring (May to June), 91% (n = 33 forests) of all significant negative correlations with temperature over the current spring, and 87 to 93% (n = 27 to 47 forests) of all significant positive correlations with soil moisture during prior-year summer and fall (Fig. 2). Many of these correlations are observed in the southwestern region, which combines the lowest MAP with the highest MAT of the study area (fig. S2). The latitudinal thermal gradient (Fig. 3A) and its effect on water availability thus exert a strong control on the direction of the correlation of growth with climatic variables. Combined across seasons, the majority of significant growth correlations with temperature become positive at 48° to 51°N (Fig. 3B). The majority of significant

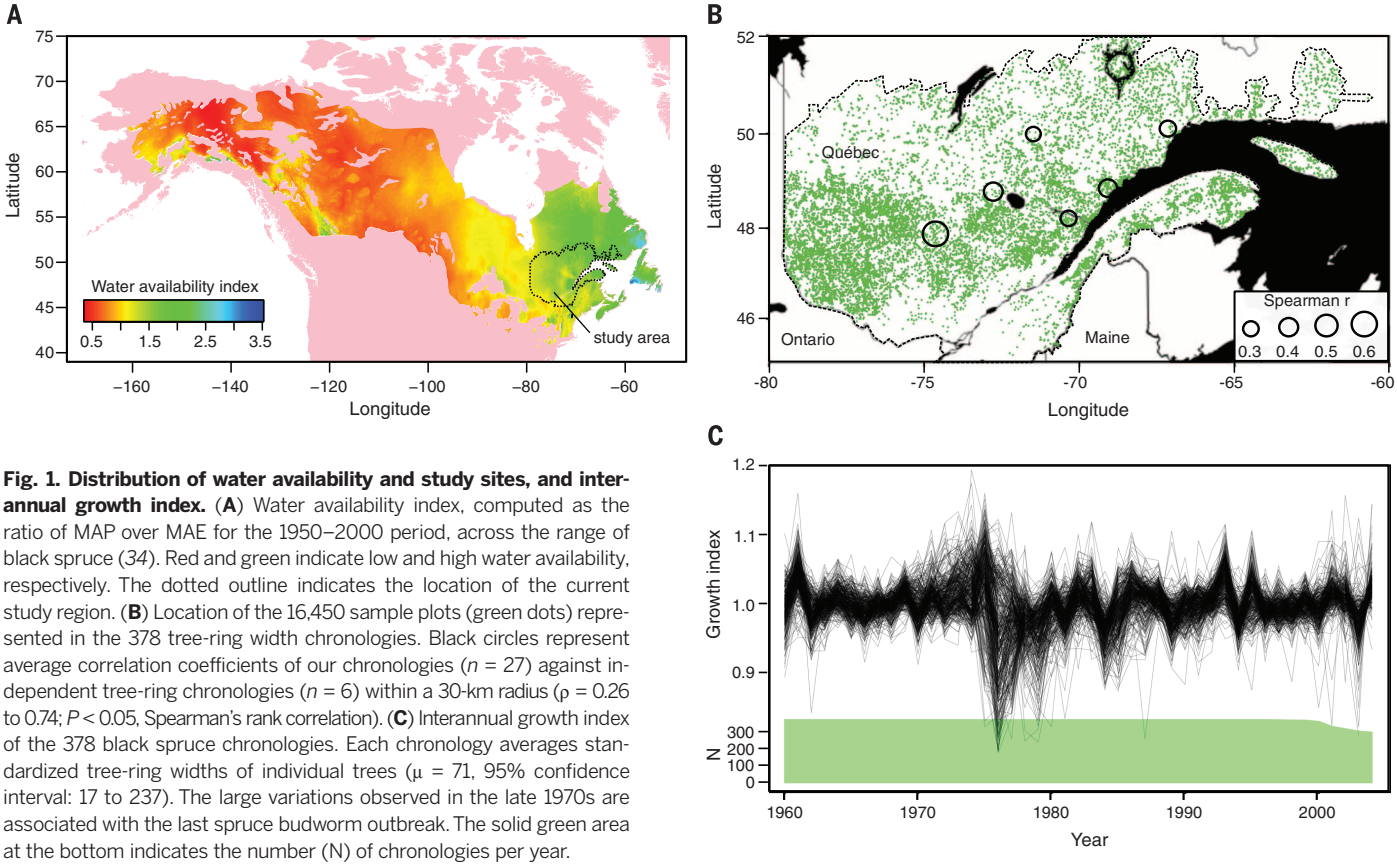


Fig. 1. Distribution of water availability and study sites, and interannual growth index. (**A**) Water availability index, computed as the ratio of MAP over MAE for the 1950–2000 period, across the range of black spruce (*34*). Red and green indicate low and high water availability, respectively. The dotted outline indicates the location of the current study region. (**B**) Location of the 16,450 sample plots (green dots) represented in the 378 tree-ring width chronologies. Black circles represent average correlation coefficients of our chronologies (n = 27) against independent tree-ring chronologies (n = 6) within a 30-km radius (ρ = 0.26 to 0.74; $P < 0.05$, Spearman's rank correlation). (**C**) Interannual growth index of the 378 black spruce chronologies. Each chronology averages standardized tree-ring widths of individual trees (μ = 71, 95% confidence interval: 17 to 237). The large variations observed in the late 1970s are associated with the last spruce budworm outbreak. The solid green area at the bottom indicates the number (N) of chronologies per year.

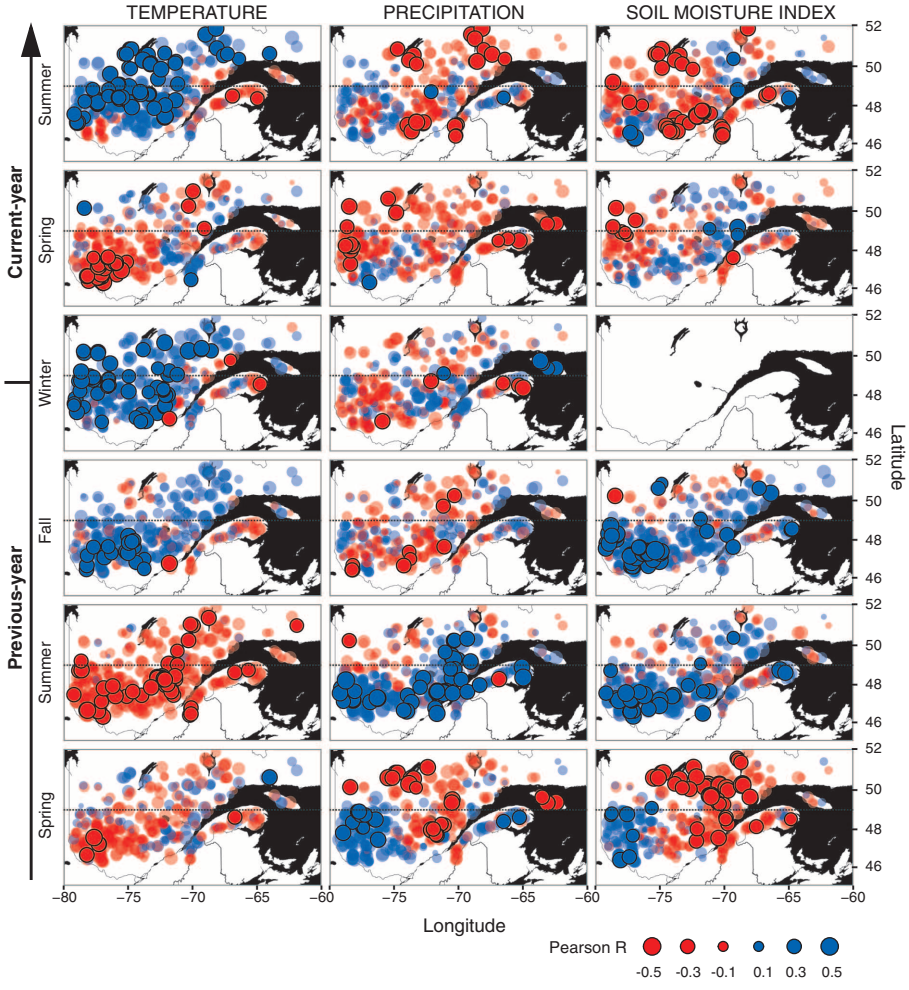


Fig. 2. Seasonal correlations of annual black spruce growth (1960–2004) with temperature, precipitation, and SMI. Seasons include prior-year spring (May and June), summer (July and August), fall (September and October), and winter (November to April) and the current-year spring and summer. Black contours around each circle indicate significant correlations ($P < 0.05$) from 1000 bootstrap replicates. Of the 378 tree-ring width chronologies, 290 (77%) display at least one significant correlation with one of the three seasonal climate variables. The 49th parallel north is indicated by a dotted line. Because the soil-drying model used to estimate the SMI is designed and calibrated for the snow-free season, we did not calculate the correlation of winter SMI with growth.

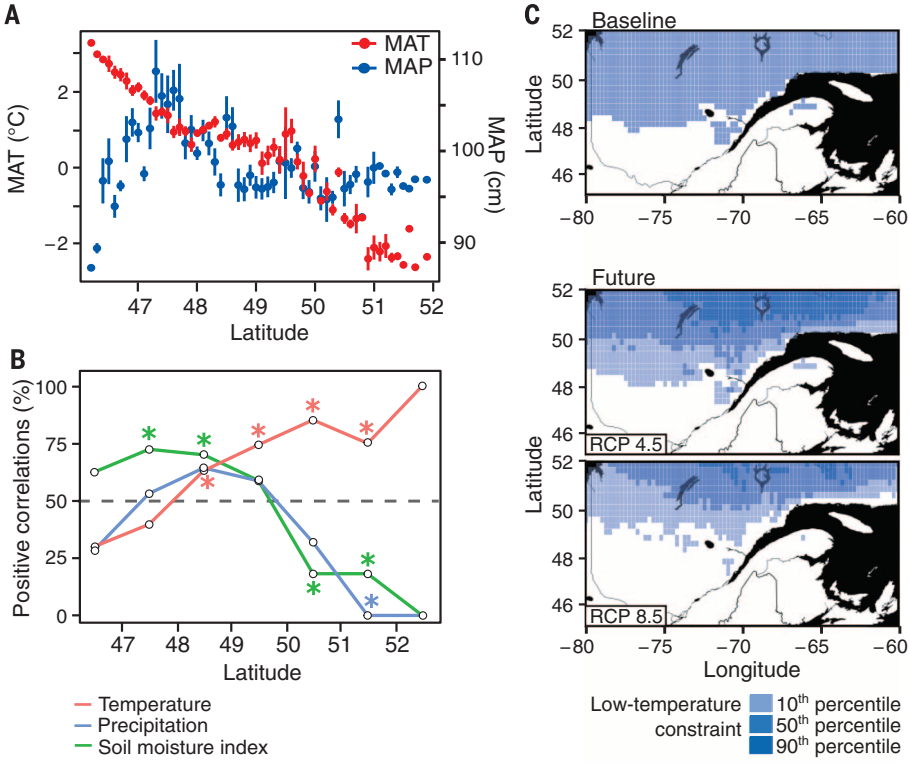


Fig. 3. Latitudinal and temporal shift in temperature, water availability, and growth response. (**A**) Average MAT and MAP of study forests (1960–2004) binned by 0.1° latitude. Data are means ± SEM. Temperature decreases linearly with latitude ($P < 0.001$, two-tailed t test), unlike precipitation ($P > 0.05$). (**B**) Proportion of positive correlations with growth ($P < 0.05$, stationary bootstrap) across seasons, binned by degree of latitude. Asterisks indicate values differing significantly from 0.5 ($P < 0.05$, two-tailed chi-squared tests). Linear regression model P values for slopes of temperature, precipitation, and SMI are <0.01, 0.11, and <0.01, respectively (one-way analysis of variance models). Forests above 49°N display significant patterns of low-temperature constraint. (**C**) Areas with baseline (1971–2000) and future (2041–2070) MAT corresponding to the forests with low-temperature constraints from (B). Colors indicate the 10th, 50th, and 90th percentiles of projected MAT from 21 climate simulations and two greenhouse gas emission scenarios [representative concentration pathways (RCPs)] of 4.5 and 8.5 W m^{-2} (*35*).

correlations with SMI are positive between 47° and 49°N and become negative above 50°N. As for precipitation, the majority of the significant correlations are negative above 51° to 52°N (Fig. 3B). To summarize, all three climate variables point toward a low-temperature constraint in black spruce forests north of approximately 49°N. The average MAT of forests sampled at this latitude is 1.1 ± 0.7°C (SD) and may be a threshold of MAT below which the growth of black spruce trees is constrained by low temperatures (Fig. 3C). In contrast to typical climate envelope models, which use species distribution data to estimate their climatic niche, our approach uses the climate sensitivity of thousands of black spruce trees.

According to median temperature projections for a low- and a high-emission scenario (4.5 and 8.5 W m^{-2}) for 2041–2070, 63 to 80% of the territory from 49° to 52°N should still be subject to MAT associated with positive temperature responses (Fig. 3C). Considering that (i) increasing growth rates are being reported at the species treeline [55° to 58°N (*18*)] and (ii) the species is already dominant at these latitudes although at lower density, we see no major constraint against a shift of the refugium into the open-crown forests located north of the study area, despite the presence of less fertile soils. We acknowledge that there is a potential warming threshold when the region would lose its capacity to favor black spruce growth.

The essentially monotypic black spruce boreal forest dominating at latitudes from 49°to 52°N has a largely positive growth response to the combined increase in temperature and decrease in precipitation, thus supporting the hypothesis that low temperatures are the dominant climatic growth constraint. Conversely, growth reductions associated with increases in temperature and decreases in precipitation and SMI are mostly found south of 49°N. This conclusion agrees well with (i) satellite-derived observations of recent increases in photosynthetic activity in high-latitude forests of NENA (*12*, *13*, *33*), (ii) ground-based reports of a recent increase in black spruce growth in the northern forest-tundra of NENA (*18*), and (iii) predictive growth models for boreal tree species of NENA (*17*). The poor adaptation of black spruce to warm temperatures (*6*) that is responsible for its lower relative abundance south of 49°N (fig. S1), coupled with the higher water requirements of the denser, taller, and more productive forest stands found at these latitudes, may contribute to the observed response gradient. Being mainly driven by temperature, this gradient is likely to also affect other boreal species of NENA, although species-specific adaptations at the scale of this study are unknown.

In contrast to the moisture-sensitive boreal forests of central and western North America, results from this heavily replicated network indicate that eastern black spruce populations north of 49°N show no sign of a negative response to climate warming and instead respond positively to increased temperature and reduced precipitation. Although these conclusions do not take into account the predicted changes in biotic and abiotic disturbances (*2*), they do suggest that the higher NENA water availability could allow boreal tree species such as black spruce to better withstand a warmer climate in NENA than in the central and western portions of North America. Outside of the potential for extreme disturbance events, NENA may act as a refugium for the boreal forest.

REFERENCES AND NOTES

1. Y. Pan *et al.*, *Science* **333**, 988–993 (2011).
2. S. Gauthier, P. Bernier, T. Kuuluvainen, A. Z. Shvidenko, D. G. Schepaschenko, *Science* **349**, 819–822 (2015).
3. Intergovernmental Panel on Climate Change (IPCC), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, New York, 2013).
4. P. E. Kauppi, M. Posch, P. Pirinen, *PLOS ONE* **9**, e111340 (2014).
5. S. Schaphoff, C. P. O. Reyer, D. Schepaschenko, D. Gerten, A. Shvidenko, *For. Ecol. Manage.* **361**, 432–444 (2016).
6. A. H. Lloyd, A. G. Bunn, *Environ. Res. Lett.* **2**, 045013 (2007).
7. M.-P. Girardin, J. Tardif, *Can. J. For. Res.* **35**, 48–64 (2005).
8. X. J. Walker, M. C. Mack, J. F. Johnstone, *Glob. Change Biol.* **21**, 3102–3113 (2015).
9. V. A. Barber, G. P. Juday, B. P. Finney, *Nature* **405**, 668–673 (2000).
10. P. S. A. Beck *et al.*, *Ecol. Lett.* **14**, 373–379 (2011).
11. C. Peng *et al.*, *Nature Clim. Change* **1**, 467–471 (2011).
12. C. Boisvenue, S. W. Running, *Glob. Change Biol.* **12**, 862–882 (2006).
13. R. de Jong, J. Verbesselt, M. E. Schaepman, S. de Bruin, *Glob. Change Biol.* **18**, 642–655 (2012).
14. F. Girard, S. Payette, R. Gagnon, *Ecoscience* **18**, 279–294 (2011).
15. J. Huang *et al.*, *Glob. Change Biol.* **16**, 711–731 (2010).
16. M. P. Girardin *et al.*, *Glob. Change Biol.* **22**, 627–643 (2016).
17. J.-G. Huang *et al.*, *PLOS ONE* **8**, e56758 (2013).
18. I. Gamache, S. Payette, *J. Ecol.* **92**, 835–845 (2004).
19. A. Beaudoin *et al.*, *Can. J. For. Res.* **44**, 521–532 (2014).
20. See the supplementary materials on *Science* Online.
21. F. Babst *et al.*, *Environ. Res. Lett.* **7**, 045705 (2012).
22. M. K. Hughes, E. A. Vaganov, A. V. Kirdyanov, F. H. Schweingruber, P. P. Silkin, *Nature* **400**, 149–151 (1999).
23. A. D. Richardson *et al.*, *Agric. For. Meteorol.* **169**, 156–173 (2013).
24. O. Bergeron *et al.*, *Glob. Change Biol.* **13**, 89–107 (2007).
25. A. Turcotte, H. Morin, C. Krause, A. Deslauriers, M. Thibeault-Martel, *Agric. For. Meteorol.* **149**, 1403–1409 (2009).
26. S. J. Steele, S. T. Gower, J. G. Vogel, J. M. Norman, *Tree Physiol.* **17**, 577–587 (1997).
27. J. M. Wolken, S. M. Landhäusser, V. J. Lieffers, U. Silins, *Can. J. For. Res.* **41**, 2292–2300 (2011).
28. N. Pederson, E. R. Cook, G. C. Jacoby, D. M. Peteet, K. L. Griffin, *Dendrochronologia* **22**, 7–29 (2004).
29. S. Sevanto *et al.*, *Tree Physiol.* **26**, 749–757 (2006).
30. M. C. Dietze *et al.*, *Annu. Rev. Plant Biol.* **65**, 667–687 (2014).
31. W. M. Havranek, W. Tranquillini, in *Ecophysiology of Coniferous Forests* (Academic Press, 1995), chap. 5, pp. 95–124.
32. L. D'Orangeville, B. Côté, D. Houle, H. Morin, *Tree Physiol.* **33**, 516–526 (2013).
33. M. Zhao, S. W. Running, *Science* **329**, 940–943 (2010).
34. A. Trabucco, R. J. Zomer, CGIAR Consortium for Spatial Information, Global Soil Water Balance Geospatial Database, available from the CGIAR-CSI GeoPortal at www.cgiar-csi.org (2009).
35. B. Thrasher, E. P. Maurer, C. McKellar, P. B. Duffy, *Hydrol. Earth Syst. Sci.* **16**, 3309–3314 (2012).

ACKNOWLEDGMENTS

We gratefully acknowledge the staff of the Ministère des Forêts, de la Faune et des Parcs du Québec (MFFP) for the meticulous work related to tree core sampling, preparation, and measurements; M.-C. Lambert, who generated meteorological data with the BioSim software; R. Ouimet for graciously sharing tree-ring chronologies from the Réseau d'Etude et de Surveillance des Ecosystèmes Forestiers; J. Noël for the vegetation and climate maps; and the anonymous referees for improving the manuscript with their thoughtful comments. Climate scenarios used were from the NEX-GDDP data set, prepared by the Climate Analytics Group and the NASA Ames Research Center, using the NASA Earth Exchange, distributed by the NASA Center for Climate Simulation and transformed by T. Logan at Ouranos. Data are available at the Dryad Digital Repository, at http://dx.doi.org/10.5061/dryad.785cv. This work was made possible by a NSERC scholarship to L.D'O. as well as funding provided by the MFFP and Le Fond Vert du Ministère du Développement Durable, de l'Environnement et de la Lutte contre les Changements Climatiques du Québec within the framework of Action Plan 2013-2018 on climate change. The authors declare that they have no conflicts of interest.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1452/suppl/DC1
Materials and Methods
Figs. S1 and S2
References (*36*–*53*)

18 February 2016; accepted 19 May 2016
10.1126/science.aaf4951

ECOSYSTEM SERVICES

Improvements in ecosystem services from investments in natural capital

Zhiyun Ouyang,[1]* Hua Zheng,[1] Yi Xiao,[1] Stephen Polasky,[2] Jianguo Liu,[3] Weihua Xu,[1] Qiao Wang,[4] Lu Zhang,[1] Yang Xiao,[1] Enming Rao,[1] Ling Jiang,[1] Fei Lu,[1] Xiaoke Wang,[1] Guangbin Yang,[5] Shihan Gong,[1] Bingfang Wu,[6] Yuan Zeng,[6] Wu Yang,[7] Gretchen C. Daily[8]*

In response to ecosystem degradation from rapid economic development, China began investing heavily in protecting and restoring natural capital starting in 2000. We report on China's first national ecosystem assessment (2000–2010), designed to quantify and help manage change in ecosystem services, including food production, carbon sequestration, soil retention, sandstorm prevention, water retention, flood mitigation, and provision of habitat for biodiversity. Overall, ecosystem services improved from 2000 to 2010, apart from habitat provision. China's national conservation policies contributed significantly to the increases in those ecosystem services.

Through pursuit of rapid economic development, China has become the second largest economy in the world and has lifted hundreds of millions of people out of poverty since the "reform and opening up," begun in the 1970s. Yet the costs of this success are reflected in high levels of environmental degradation. In 1998, massive deforestation and erosion contributed to severe flooding along the Yangtze River, killing thousands of people, rendering 13.2 million

homeless, and causing about U.S. $36 billion in property damage (*1*). This crisis prompted creation of the world's largest government-financed payment for ecosystem services programs: the Natural Forest Conservation Program (NFCP) and the Sloping Land Conversion Program (SLCP) (*2*, *3*). By 2009, the cumulative total investment through the NFCP and SLCP exceeded U.S. $50 billion and directly involved more than 120 million farmers in 32 million households in the SLCP alone (*4*). These programs aim to reduce natural disaster risk by restoring forest and grassland, while improving livelihood options and alleviating poverty.

[1]State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China. [2]Department of Applied Economics and Natural Capital Project, Institute on the Environment, University of Minnesota, St. Paul, MN 55108, USA. [3]Center for Systems Integration and Sustainability, Michigan State University, East Lansing, MI 48823-5243, USA. [4]Satellite Environment Center, Ministry of Environmental Protection, Beijing 100094, China. [5]School of Geographic and Environmental Sciences, Guizhou Normal University, Guiyang 550001, China. [6]Key Laboratory of Digital Earth Sciences, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, Beijing 100101, China. [7]College of Environmental and Resource Sciences, Zhejiang University, Hangzhou 310058, China. [8]Department of Biology, Natural Capital Project, and Woods Institute for the Environment, Stanford University, Stanford, CA 94305, USA.
***Corresponding author. Email: zyouyang@rcees.ac.cn (Z.O.); gdaily@stanford.edu (G.C.D.)**

Although there are scattered case studies [e.g., (*5*)], systematic, comprehensive, and rigorous assessments of the ecosystem services and people affected by these conservation policies at the national level have been lacking. To address this knowledge gap, in 2012, China's Ministry of Environmental Protection and Chinese Academy of Sciences launched a national ecosystem assessment to quantify ecosystem status and trends, and ecosystem service provision between 2000 and 2010. The China ecosystem assessment (CEA) was designed to address central questions of how ecosystem services are changing, where important services originate, and what should be protected and restored to increase ecosystem services.

Here, we report on results of the first CEA, which covered all of mainland China from 2000 to 2010. The assessment used data from a variety of sources, including >20,000 multisource satellite images, recorded biophysical data [such as soil, digital elevation models (DEMs), hydrology, and meteorology], >100,000 field surveys; historical records of biodiversity; and special assessments from several government ministries (e.g., surveys of desertification, soil erosion). All lands were classified using a newly established ecosystem classification system for China (*6*). The CEA collected data on food production by crop converted to kilocalories (kcal) and modeled the level of provision for six other important ecosystem services [carbon sequestration (metric tons), soil retention (metric tons), sandstorm prevention (metric tons), water retention (metric tons), flood mitigation (m^3), and habitat provision for biodiversity (total habitat area of endemic, endangered, and nationally protected species per county)] using InVEST (a suite of free, open-source software models designed for Integrated Valuation of Ecosystem Services and Tradeoffs) (*7*, *8*) and other biophysical models (*6*).

We translated biophysical supply of ecosystem services into importance of service provision by weighting supply by the number of people affected. For example, sandstorm prevention is weighted by the population downwind. The importance of food supply and carbon sequestration has the same value in all locations, which reflects the uniform atmospheric mixing of CO_2 and integrated markets for crops (*6*).

All ecosystem services evaluated increased between 2000 and 2010, with the exception of habitat provision for biodiversity. Food production had the largest increase (38.5%), followed by carbon sequestration (23.4%), soil retention (12.9%), flood mitigation (12.7%), sandstorm prevention (6.1%), and water retention (3.6%), whereas habitat provision decreased slightly (–3.1%) (Fig. 1A).

Not all regions had a positive trend. Ecosystem services increased, in aggregate, in the Loess Plateau in western China (the most severe soil erosion area in the world); the Sanjiangyuan area in the center of the Tibetan Plateau (the headwater region of the Yellow River, Yangtze River, and Mekong River); and the Taihang Mountains in north China (the water provision area for the North China Plain) (Fig. 1B). Ecosystem services decreased, in aggregate, in the southwestern Tibetan Plateau, the western Hunshandake Sandy Area in northern China, and the northern Tianshan Mountains in western China (Fig. 1C). There are also tradeoffs between services (e.g., food production and soil retention). However, we see many synergistic increases or decreases among services (e.g., carbon sequestration, soil retention, and sandstorm prevention).

Food production in China is concentrated in the eastern plains (including the Northeast Plain, North China Plain, and Middle and Lower Yangtze Plain) and the Sichuan Basin (Fig. 2A and fig. S2). Important areas providing the other ecosystem services (carbon sequestration, soil retention, sandstorm prevention, water retention, flood mitigation, and provision of habitat for biodiversity) occur throughout the country. The government's priority areas for securing these ecosystem services are in the Great Khingan and Changbai Mountains in northeastern China, the Hunshandake Sandy Area in northern China, the Tianshan Mountains, and Loess Plateau in northwestern China, the Sanjiangyuan Area and Hengduan Mountains in southwestern China, the adjacent Mountains of Zhejiang Province and Fujian Province, the Nanling Mountains in southeastern China, and the Qinling-Daba Mountains in central China (Fig. 2, B to G, and fig. S5). These priority areas provide 83.4% of carbon-sequestration services, 77.7% of soil retention services, 59.1% of sandstorm prevention services,

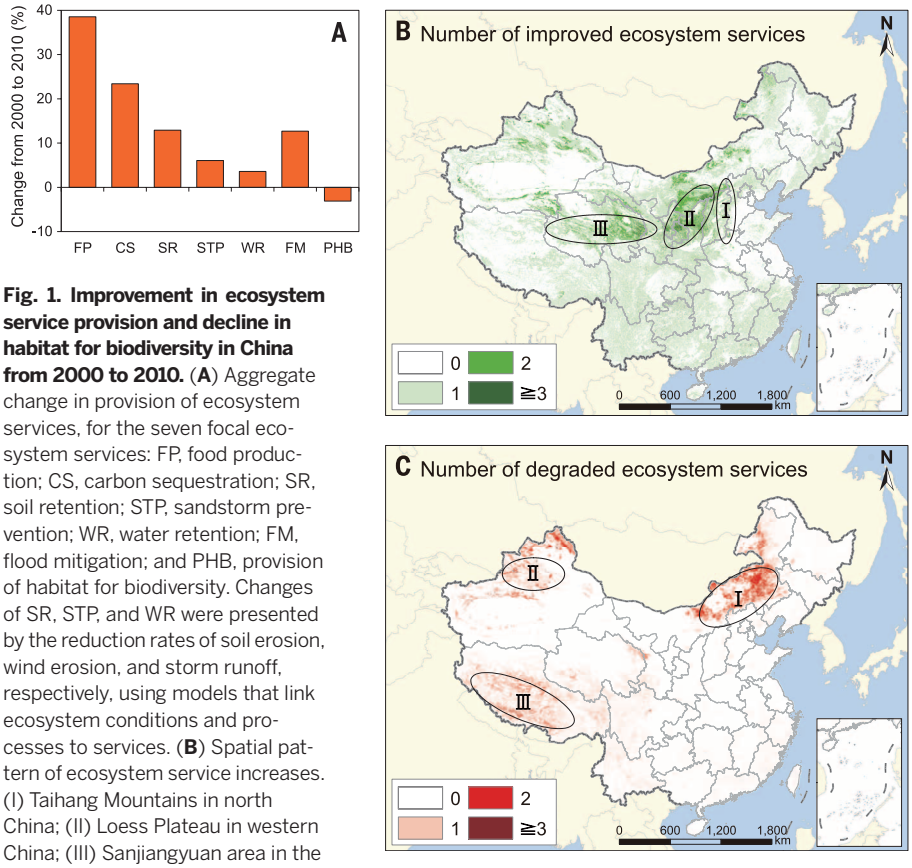


Fig. 1. Improvement in ecosystem service provision and decline in habitat for biodiversity in China from 2000 to 2010. (**A**) Aggregate change in provision of ecosystem services, for the seven focal ecosystem services: FP, food production; CS, carbon sequestration; SR, soil retention; STP, sandstorm prevention; WR, water retention; FM, flood mitigation; and PHB, provision of habitat for biodiversity. Changes of SR, STP, and WR were presented by the reduction rates of soil erosion, wind erosion, and storm runoff, respectively, using models that link ecosystem conditions and processes to services. (**B**) Spatial pattern of ecosystem service increases. (I) Taihang Mountains in north China; (II) Loess Plateau in western China; (III) Sanjiangyuan area in the center of the Tibetan Plateau. (**C**) Spatial pattern of ecosystem service decreases. (I) Western Hunshandake Sandy Area in northern China; (II) Northern Tianshan Mountains in western China; (III) Southwestern Tibetan Plateau.

80.4% of water retention services, and 56.3% of natural habitats, although they make up only 37.0% of the area of China (Fig. 2H).

The changes in the provision of ecosystem services from 2000 to 2010 are the result of natural capital investment policies, changes in biophysical factors, and socioeconomic development (Table 1) (*6*). Overall, our results suggest that China's national conservation policies contributed significantly to the increases in four key ecosystem services. For carbon sequestration and soil retention, coefficients for the SLCP targeting forest restoration (SLCP_F) and NFCP are positive and statistically significant. For sand fixation, SLCP targeting grassland restoration (SLCP_G) is positive and statistically significant ($P < 0.05$) whereas SLCP_F is not. For water retention, NFCP is positive and statistically significant ($P < 0.001$) but SLCP_F is not.

The results of the CEA show that improving ecosystem services and economic growth can coexist. Analyses using model simulations in the United Kingdom (*9*), the United States (*10*), and Australia (*11*) also show that it is possible to increase the provision of key ecosystem services with economic growth through intelligent policy design, although ecosystem services can decline without proper policies in place. Further developing the CEA can help inform future efforts to sustain and enhance ecosystem services and human well-being, not only in China but worldwide (*12*–*15*).

The results generated by the CEA have already been applied by policy-makers in China at national, provincial, and local levels, by several parts of government (e.g., Ministry of Environmental Protection and the National Development Reform Commission). For example, 49.4% of China's land area (4.74 million km^2 over 63 locations) has been newly incorporated into Ecosystem Function Conservation Areas (EFCAs), designed to secure the nation's most vital natural capital, on the basis of CEA's characterization of important source areas for ecosystem service provision (Fig. 2H) (*16*, *17*). EFCAs include areas that provide 77.7% of carbon-sequestration services, 75.3% of soil-retention services, 60.7% of sandstorm-prevention services, 76.8% of water-retention services, 60.2% of flood-mitigation services, and 67.6% of natural habitats. The CEA also informed the national-level policy of ecological protection red-lining (EPR) that designates lands for strict protection to ensure sustainable provision of ecosystem services (*18*). The national EPR, as well as EPR planning, in provinces and localities was based on priority sources of ecosystem services and covers 34.4% of the area of China (Fig. 2H) (*18*, *19*). The results of the CEA have also been applied in national transportation network planning to identify sensitive areas for protection when designing road projects (*20*).

Although the CEA and some other studies [e.g., (*21*)] have documented improvement in ecosystem services, there remain serious environmental challenges, including deteriorating air and water quality, increasing greenhouse gas emissions, and an

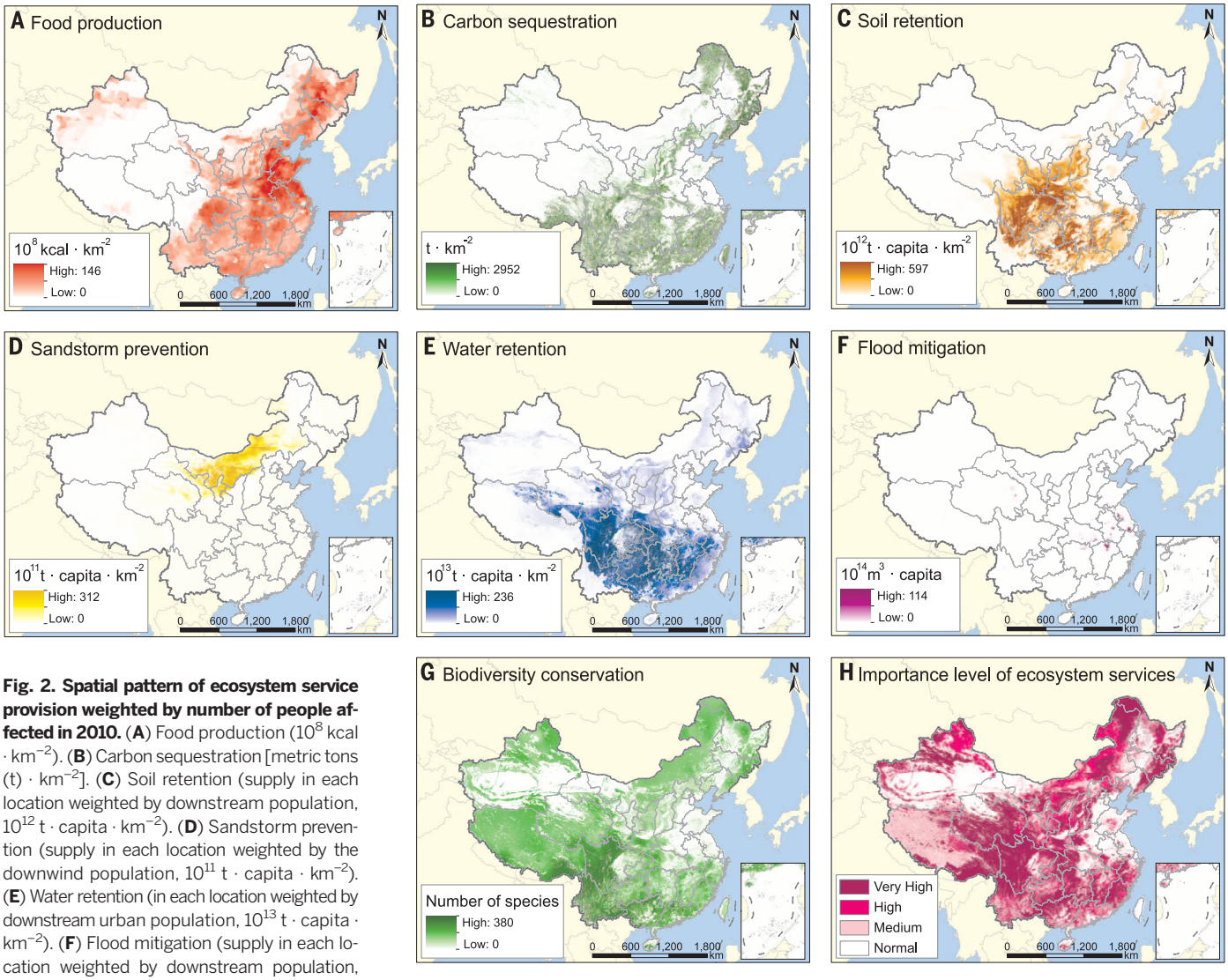


Fig. 2. Spatial pattern of ecosystem service provision weighted by number of people affected in 2010. (**A**) Food production (10^8 kcal · km^{-2}). (**B**) Carbon sequestration [metric tons (t) · km^{-2}]. (**C**) Soil retention (supply in each location weighted by downstream population, 10^{12} t · capita · km^{-2}). (**D**) Sandstorm prevention (supply in each location weighted by the downwind population, 10^{11} t · capita · km^{-2}). (**E**) Water retention (in each location weighted by downstream urban population, 10^{13} t · capita · km^{-2}). (**F**) Flood mitigation (supply in each location weighted by downstream population, 10^{14} m^3 · capita). (**G**) Provision of habitat for biodiversity (total species richness of endemic, endangered, and nationally protected species per county). (**H**) Index of relative importance of ecosystem services (*6*).

Table 1. Factors associated with increases in four key ecosystem services. Unit of analysis is the county. Dependent variables are increases in per-unit-area carbon sequestration, soil retention, sandstorm prevention, and water retention, respectively. Standardized coefficients and robust standard errors are reported outside and inside parentheses, respectively. Model results passed standard regression diagnostics. Variance inflation factors (VIFs) were tested to be <5.

Category	Independent variable	Carbon sequestration	Soil retention	Sandstorm prevention	Water retention
Policy	SLCP targeting forest restoration (1: yes; 0: no)	0.029** (0.006)	0.069* (0.067)	0.060 (0.152)	0.005 (0.107)
	SLCP targeting grassland restoration (1: yes; 0: no)	-	-	0.125* (0.138)	-
	NFCP (1: yes; 0: no)	0.062*** (0.007)	0.227*** (0.051)	-	0.094*** (0.387)
Biophysical variables	Initial service amount in 2000 (10^3 ton · km^{-2})	0.873*** (0.004)	0.676*** (0.002)	0.729*** (0.042)	0.454*** (0.002)
	Above-ground forest biomass per unit area in 2000 (10^3 t · km^{-2})	-	−0.586*** (0.042)	-	-
	Above-ground grass biomass per unit area in 2000 (10^3 t · km^{-2})	-	-	−0.135** (0.612)	-
	Proportion of forest area in 2000	-	-	-	−0.255*** (1.797)
	Change in proportion of forest area 2000 to 2010	-	-	-	0.799*** (54.831)
	Proportion of shrub area in 2000	0.074*** (0.037)	−0.060* (0.235)	-	−0.034 (1.617)
	Change in proportion of shrub area 2000 to 2010	0.021* (0.146)	0.006 (1.448)	-	0.580*** (50.793)
	Proportion of grassland area in 2000	−0.059*** (0.017)	-	-	−0.063* (1.245)
	Change in proportion of grassland area 2000 to 2010	0.036** (0.259)	-	-	0.192*** (18.980)
	Proportion of cropland area in 2000	-	0.045 (0.136)	−0.083 (0.401)	-
	Change in proportion of cropland area 2000 to 2010	-	−0.285** (2.780)	−0.1112* (1.632)	-
Socioeconomic variables	Human population density in 2000 (10^3 individual · km^{-2})	−0.024† (0.012)	−0.172** (0.168)	-	0.059† (0.992)
	Change in human population density 2000 to 2010 (10^3 individual · km^{-2})	0.025* (0.060)	0.074** (0.616)	-	−0.085** (5.942)
	Proportion of urban population in 2000	−0.008 (0.022)	−0.007 (0.136)	-	−0.021 (1.237)
	Change in proportion of urban population 2000 to 2010	−0.008 (0.016)	0.045 (0.171)	-	−0.093*** (1.274)
	Livestock inventory in 2000 (10^3 sheep unit · km^{-2})	−0.013 (0.009)	−0.116*** (1.12E-04)	-	−0.125** (0.001)
	Change in livestock inventory 2000 to 2010 (10^3 sheep unit · km^{-2})	−0.039*** (0.014)	0.007 (1.23E-04)	-	−0.064* (0.001)
R^2	-	0.859	0.285	0.596	0.545
N	-	1296	1136	186	871

†P < 0.1; *P < 0.05; **P < 0.01; ***P < 0.001.

expanding global ecological footprint from importing raw materials (*22*, *23*). Solutions will require interventions beyond the ecosystem restoration that is the focus of the CEA.

Although the CEA has already had notable success, providing improved policy guidance in the future depends on making progress in several aspects. First, rapid technological advances can enable more frequent data collection at finer resolution. Second, direct measurement of variables that are more directly linked to service provision can improve accuracy of results (e.g., wide

spatial coverage of measures of soil loss). Third, an expanded set of ecosystem services could be quantified, including ecosystem contributions to securing water and air quality, both of which have deteriorated in China in recent decades, and mental health benefits of exposure to nature (*24*). Fourth, improved measures can be used that more directly link ecosystem services to human well-being, such as economic measures of value and direct measures of impact on health, livelihoods, happiness, or other aspects of well-being (*25*, *26*). Finally, better understanding of human behavioral responses to changes in policy or market conditions could improve policy effectiveness. Regularly repeating the CEA can provide insight into future national development pathways (*27*).

REFERENCES AND NOTES

1. Q. Ye, M. H. Glantz, *Mitig. Adapt. Strategies Glob. Change* **10**, 159–182 (2005).
2. J. Liu, S. Li, Z. Ouyang, C. Tam, X. Chen, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 9477–9482 (2008).
3. P. Zhang *et al.*, *Science* **288**, 2135–2136 (2000).
4. J. Liu, Z. Ouyang, W. Yang, W. Xu, S. Li, in *Encyclopedia of Biodiversity*, S. A. Levin, Ed. (Academic Press, Waltham, MA, ed. 2, 2013), pp. 372–384.
5. M.-N. Tuanmu *et al.*, *Conserv. Biol.*, 10.1111/cobi.12669 (2016).
6. Materials and methods are available as supplementary materials on *Science* Online.
7. R. Sharp *et al.*, *InVEST +VERSION+ User's Guide* (The Natural Capital Project, Stanford University, University of Minnesota, The Nature Conservancy, and World Wildlife Fund, 2015).
8. P. Kareiva, H. Tallis, T. H. Ricketts, G. C. Daily, S. Polasky, Eds., *Natural Capital: Theory and Practice of Mapping Ecosystem Services* (Oxford Univ Press, New York, 2011).
9. I. J. Bateman *et al.*, *Science* **341**, 45–50 (2013).
10. J. J. Lawler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7492–7497 (2014).
11. S. Hatfield-Dodds *et al.*, *Nature* **527**, 49–53 (2015).
12. Millennium Ecosystem Assessment, *Ecosystems and Human Well-Being: Synthesis* (Island Press, Washington, DC, 2005).
13. UKNEA (UK National Ecosystem Assessment), *The UK National Ecosystem Assessment Technical Report* [United Nations Environment Programme (UNEP)–World Conservation Monitoring Centre, Cambridge, 2011].
14. UNEP, Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Decision IPBES-2/4: Conceptual framework for the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services: Report of the second session of the plenary of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (UNEP, 2014).
15. WAVES (Wealth Accounting and the Valuation of Ecosystem Services), *WAVES Annual Report 2015* (World Bank, Washington, DC, 2015).
16. P. R. Ehrlich, P. M. Kareiva, G. C. Daily, *Nature* **486**, 68–73 (2012).
17. Ministry of Environmental Protection of China and Chinese Academy of Sciences, *National Ecosystem Service Zoning in China* (Ministry of Environmental Protection and CAS, Beijing, 2015).
18. Ministry of Environmental Protection, *National Ecological Protection Redlining* (Ministry of Environmental Protection, Beijing, 2015).
19. China Council for International Cooperation on Environment and Development, *Report on Institutional Innovation of Ecological Protection Redlining* (CCICED, Beijing, 2014).
20. Ministry of Transport of China, *National Road Development Planning (2014–2030)* (MTC, Beijing, 2013).
21. A. Viña, W. McConnell, H. B. Yang, Z. C. Xu, J. G. Liu, *Sci. Adv.* **2**, e1500965 (2016).
22. B. Fu, *Science* **321**, 611 (2008).
23. J. G. Liu, P. H. Raven, *Crit. Rev. Environ. Sci. Technol.* **40**, 823–851 (2010).
24. G. N. Bratman, J. P. Hamilton, G. C. Daily, *Ann. N.Y. Acad. Sci.* **1249**, 118–136 (2012).
25. S. R. Carpenter *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 1305–1312 (2009).
26. C. Folke, S. R. Carpenter, B. Walker, M. Scheffer, *Ecol. Soc.* **15**, 20 (2010).
27. National Development and Reform Commission of China, *Opinions on Accelerating the Construction of Ecological Civilization* (NDRCC, Beijing, 2013).

ACKNOWLEDGMENTS

We thank the Ministry of Environmental Protection (MEP) of China and Chinese Academy of Sciences (CAS) for organizing the Project, and experts in 31 provinces for collecting the field data for statistics analysis on this manuscript. This work was supported by the Ministry of Finance of China through the MEP/CAS project "Survey and Assessment of National Ecosystem Changes Between 2000 and 2010, China" and by the international Natural Capital Project. All the data are available at www.sciencedb.cn/dataSet/handle/73.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1455/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S5
References (*28*–*98*)

10 January 2016; accepted 18 May 2016
10.1126/science.aaf2295

HUMAN BEHAVIOR

New online ecology of adversarial aggregates: ISIS and beyond

N. F. Johnson,[1] M. Zheng,[1] Y. Vorobyeva,[2] A. Gabriel,[1] H. Qi,[1] N. Velasquez,[2] P. Manrique,[1] D. Johnson,[3] E. Restrepo,[4] C. Song,[1] S. Wuchty[5,6]*

Support for an extremist entity such as Islamic State (ISIS) somehow manages to survive globally online despite considerable external pressure and may ultimately inspire acts by individuals having no history of extremism, membership in a terrorist faction, or direct links to leadership. Examining longitudinal records of online activity, we uncovered an ecology evolving on a daily time scale that drives online support, and we provide a mathematical theory that describes it. The ecology features self-organized aggregates (ad hoc groups formed via linkage to a Facebook page or analog) that proliferate preceding the onset of recent real-world campaigns and adopt novel adaptive mechanisms to enhance their survival. One of the predictions is that development of large, potentially potent pro-ISIS aggregates can be thwarted by targeting smaller ones.

Extremist entities such as ISIS (known as Islamic State) stand to benefit from the global reach and speed of the Internet for propaganda and recruiting purposes in ways that were unthinkable for their predecessors (*1*–*10*). This increased connectivity not only may facilitate the formation of real-world organized groups that subsequently carry out violent attacks (e.g., the ISIS-directed attacks in Paris in November 2015) but also may inspire self-radicalized actors with no known history of extremism or links to extremist leadership to operate without actually belonging to a group (e.g., the ISIS-inspired attack in San Bernardino in December 2015) (*11*). Recent research has used records of attacks to help elucidate group structure in past organizations for which the Internet was not a key component (*3*, *6*, *12*), the nature of attacks by lone-wolf actors (*13*), and the relationship between general online buzz and real-world events (*14*–*16*). Online buzz created by individuals that casually mention ISIS or protests is insufficient to identify any long-term buildup ahead of sudden real-world events (see, for example, fig. S1). This leaves open the question of how support for an entity like ISIS develops online—possibly before any real-world group has been formed or any real-world attack has been perpetrated—whether by "recruits" or by those simply "inspired."

Our data sets consist of detailed second-by-second longitudinal records of online support activity for ISIS from its 2014 development onward and, for comparison, online civil protestors across multiple countries within the past 3 years, following the U.S. Open Source Indicator (OSI) project (*14*–*16*). The supplementary materials (SM) provide a roadmap for the paper, data descriptions, and downloads. The data show that operational pro-ISIS and protest narratives develop through self-organized online aggregates, each of which is an ad hoc group of followers of an online page created through Facebook or its global equivalents, such as ВКонтакте (VKontakte) at http://vk.com/ (Fig. 1). These generic web-based interfaces allow such aggregates to form in a language-agnostic way and with freely chosen names that help attract followers without publicizing their members' identities. Because the focus in this paper is on the ecosystem rather than the behavior of any individual aggregate, the names are not being

[1]Department of Physics, University of Miami, Coral Gables, FL 33126, USA. [2]Department of International Studies, University of Miami, Coral Gables, FL 33126, USA. [3]Department of Government, Harvard University, Cambridge, MA 02138, USA. [4]Department of Geography and Regional Studies, University of Miami, Coral Gables, FL 33126, USA. [5]Department of Computer Science, University of Miami, Coral Gables, FL 33126, USA. [6]Center for Computational Science, University of Miami, Coral Gables, FL 33126, USA.
***Corresponding author. Email: wuchtys@cs.miami.edu**

released. They are available on request from the authors. Pro-ISIS aggregates inhabit an online environment in which predatory entities such as police cybergroups, individual hackers, and website moderators seek to shut down pro-ISIS activity and narratives (*17*, *18*). In contrast to the largely mundane chatter that may casually mention ISIS on Twitter and in aggregates focused on sport, for example, pro-ISIS aggregates frequently discuss operational details such as routes for financing, technological know-how, and avoiding drone strikes. We chose VKontakte for our pro-ISIS analysis because (i) pro-ISIS aggregates are shut down essentially immediately on Facebook, but not on VKontakte; (ii) it is the largest European online social networking service, with more than 350 million users; (iii) it allows multiple languages and is used worldwide; (iv) being based in Russia, it has a high concentration of users of Chechen origin focused in the Caucasus region near ISIS's main area of influence in the Levant; and (v) ISIS used it to spread propaganda among the Russian-speaking population (*2*).

Our methodology for identifying these pro-ISIS aggregates was as follows. We manually identified relevant narratives using hashtags in multiple languages—e.g., #isn, #khilafah, #fisyria, #игиш (i.e., ISIS), #дауля (i.e., *dawla*, meaning "state"), and #халифат (i.e., "Caliphate")—and traced these to underlying aggregates. The specific criterion for inclusion in the list was that the group explicitly expressed its support for ISIS, publishing ISIS-related news or propaganda and/or calling for jihad in the name of ISIS. This list was fed into software application programming interfaces that expanded it by means of automated search snowballing (fig. S2). The expanded aggregate list was then cross-checked to eliminate false identifications. New embedded links were manually searched to identify more aggregates and hashtags. We then iterated this process until closure of the aggregate list (i.e., the search led back to aggregates that were already in the list). Although this process was labor intensive, we were able to find closure on a daily basis in real time. A similar process was followed for the civil protest data.

We uncovered 196 pro-ISIS aggregates involving 108,086 individual followers between 1 January and 31 August 2015. On any given day, the total number of follows in the follower-aggregate network [i.e., the total number of links that existed on that day from followers (blue nodes) into the various aggregates (red nodes), as shown in the inset to Fig. 1) ranged up to 134,857. The data provided us with bipartite graphs in which individual members belong to aggregates but aggregates are not linked to each other except through people. This two-mode network has a highly complex temporal evolution—with strong heterogeneity in both the number of follows per individual follower (i.e., the number of links emanating from a given blue node) and the number of follows per aggregate (i.e., the number of links entering a given red node, which we define as the aggregate's size)—and no obvious hierarchical structure. This suggests that the follower-aggregate dynamics are driven by self-organization. Such online support is likely a necessary but not sufficient condition for any real-world actions to take place subsequently, because many additional factors can hinder real-world execution. However, Fig. 2 suggests that the online proliferation of pro-ISIS or protest aggregates can indeed act as an indicator of conditions becoming right for the onset of a real-world attack campaign or mass protests, respectively. We fit the trend in the creation dates of new online aggregates (Fig. 2, A and B) to a well-known organizational development curve (*19*). The escalation parameter b diverges at these real-world onsets (Fig. 2, C and D) and follows the same mathematical dependence $(T_c - t)^{-1}$ as a wide class of physical phase transitions (*20*), with the divergence date T_c matching the actual onset almost exactly (SM). The connection to physical phase transitions again suggests that self-organization is a driving factor (*20*). Although such a divergence will not necessarily preempt attacks involving only a few individuals, such as in San Bernardino or Paris, it can help indicate an alignment of favorable conditions and has the advantage that it does not rely on any real-world events having yet occurred or likely dates having been circulated through social media in advance (*14*–*16*). The far longer lifetimes for online aggregates of protestors in Fig. 2B, as compared with pro-ISIS aggregates in Fig. 2A, makes sense because predatory online shutdown pressure was far less for the civil protestors; in particular, we found no evidence of any shutdowns in Fig. 2B, in stark contrast to Fig. 2A. Figure 2D is likely smoother than Fig. 2C for the same reason. More aggressive antigovernment protests, such as the sudden outburst in Venezuela in February 2014 (fig. S6),

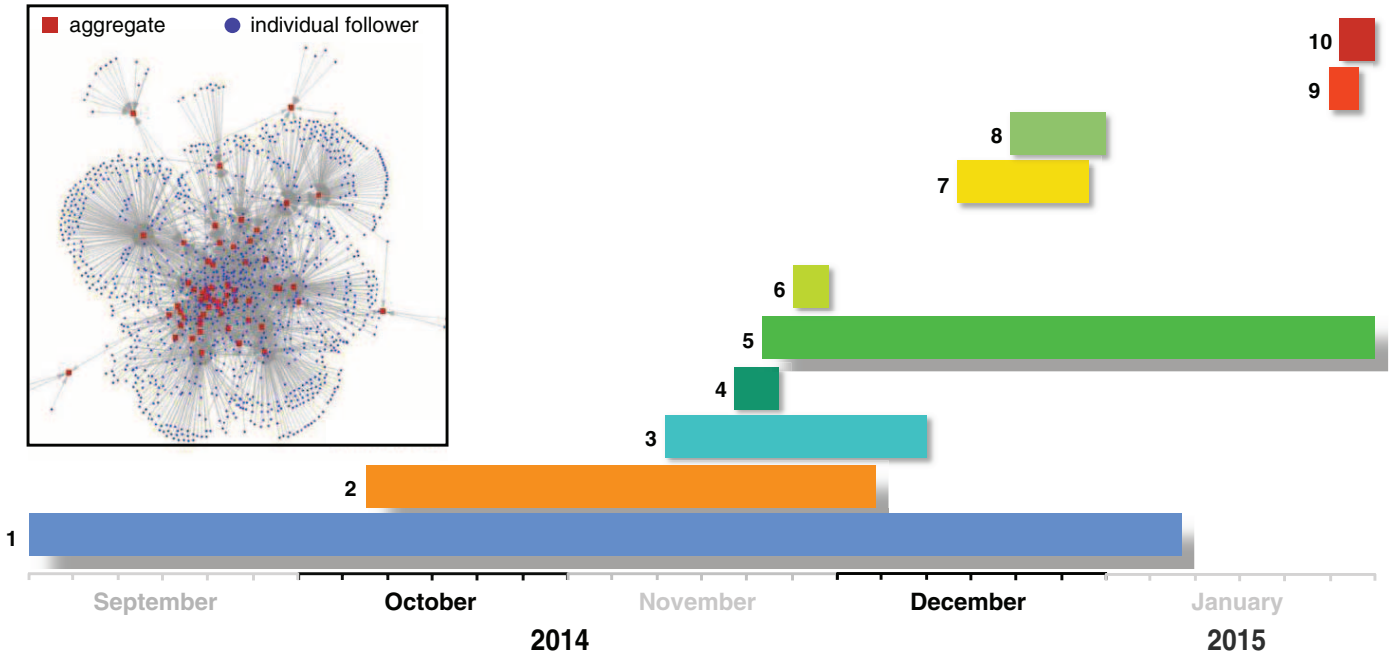


Fig. 1. Pro-ISIS aggregates. Horizontal bars illustrate timelines of some typical pro-ISIS aggregates. Their names are available from the authors. Each timeline starts when the aggregate appears and ends when it disappears. (Inset) Snapshot of part of an aggregate-follower network on 1 January 2015 showing individual followers (blue nodes) linking to pro-ISIS aggregates (red nodes). Followers can link into as many aggregates as they wish. Aggregates emerge of all sizes, where an aggregate's size is the number of follows linking into it.

generate an intermediate case between Fig. 2, A and B.

We now develop a systems-level theory of this online aggregate ecology. The aggregate size variations observed empirically were characterized by distinctive shark-fin shapes (Fig. 3A), with each shutdown of a pro-ISIS aggregate severing the links into that particular aggregate—hence the abrupt drop. This fragmentation coexists with self-organized coalescence by which individual followers sporadically link into existing aggregates while existing aggregates sporadically link into each other. Although each aggregate's precise shark-fin shape will depend on its content and noticeability to external predators, Fig. 3 shows that the system-level features are captured using only this minimal coalescence-fragmentation process. At each time step, a phenomenological probability v_{coal} describes the sporadic addition of 1,2,3,... etc. follows to an aggregate (coalescence of followers), whereas v_{frag} describes the sporadic sudden shutdown of an aggregate (fragmentation of followers). Such stochastic shutdown is realistic because the predators (e.g., government monitors or individual hackers) are largely independent entities and can only shut down aggregates that they happen to find. Larger aggregates should be more noticeable; hence, we can take the probability of a particular aggregate being picked for coalescence or targeted for shutdown as proportional to the aggregate's size (*21*), although this is generalizable to other algebraic forms without affecting our main findings (SM). The total number of potential follows N in the system is a sum, over all potential followers, of the maximum number of aggregates that each follower is prepared to follow. The number of follows per individual can be heterogeneous, and at any time step, not all N follows are necessarily used. Computer simulations of this coalescence-fragmentation process reproduce the ecology of shark-fin shapes of all sizes (Fig. 3B) with a power-law distribution $s^{-\alpha}$ for the time-average number of aggregates of size s where $\alpha = 2.5$. This is similar to the empirical value of $\alpha = 2.33$ that had high goodness-of-fit ($P = 0.86$) (Fig. 3C). These shark-fin dynamics are robust in that they emerge irrespective of when we examine the model's evolution (Fig. 3B) and for any value of N as a result of the model's self-similarity—i.e., the coalescence-fragmentation process generates the same dynamics across all aggregate sizes (*22*). Connecting to real-world ISIS activity, we note that the severity of ISIS attacks is also approximately power-law distributed with exponent $\alpha = 2.44$ and goodness-of-fit $P > 0.1$. The model can be represented mathematically by the following coupled, nonlinear differential equations describing the number n_s of pro-ISIS aggregates of size s ($s > 1$) over time.

$$\frac{\partial n_s(t)}{\partial t} = \frac{v_{coal}}{N^2}\sum_{k=1}^{s-1} k(s-k)n_k(t)n_{s-k}(t) - \frac{2v_{coal}sn_s(t)}{N^2}\sum_{k=1}^{\infty} kn_k(t) - \frac{v_{frag}sn_s(t)}{N} \quad (1)$$

A detailed discussion of Eq. 1 is given in the SM. Like the data and computer simulation, solving Eq. 1 mathematically yields a power-law $s^{-\alpha}$ for the time-averaged aggregate size, with an exact exponent $\alpha = 2.5$ [see (*23*, *24*) and SM for the mathematical proof]. The spatial independence of Eq. 1 is consistent with online interactions being largely independent of followers' separation across the globe. The first term on the right describes the formation of an aggregate of size s (i.e., s follows) from a smaller one through the addition of 1,2,3,... etc. new follows; the second describes the loss of an aggregate that coalesced with another aggregate; and the third describes the fragmentation of an aggregate of size s. $n_{s=1}(t)$ is the pool of isolated (i.e., unused) follows at time t—i.e., potential "recruits", with $\Sigma_{s=1}^{\infty} sn_s(t) = N$. We take

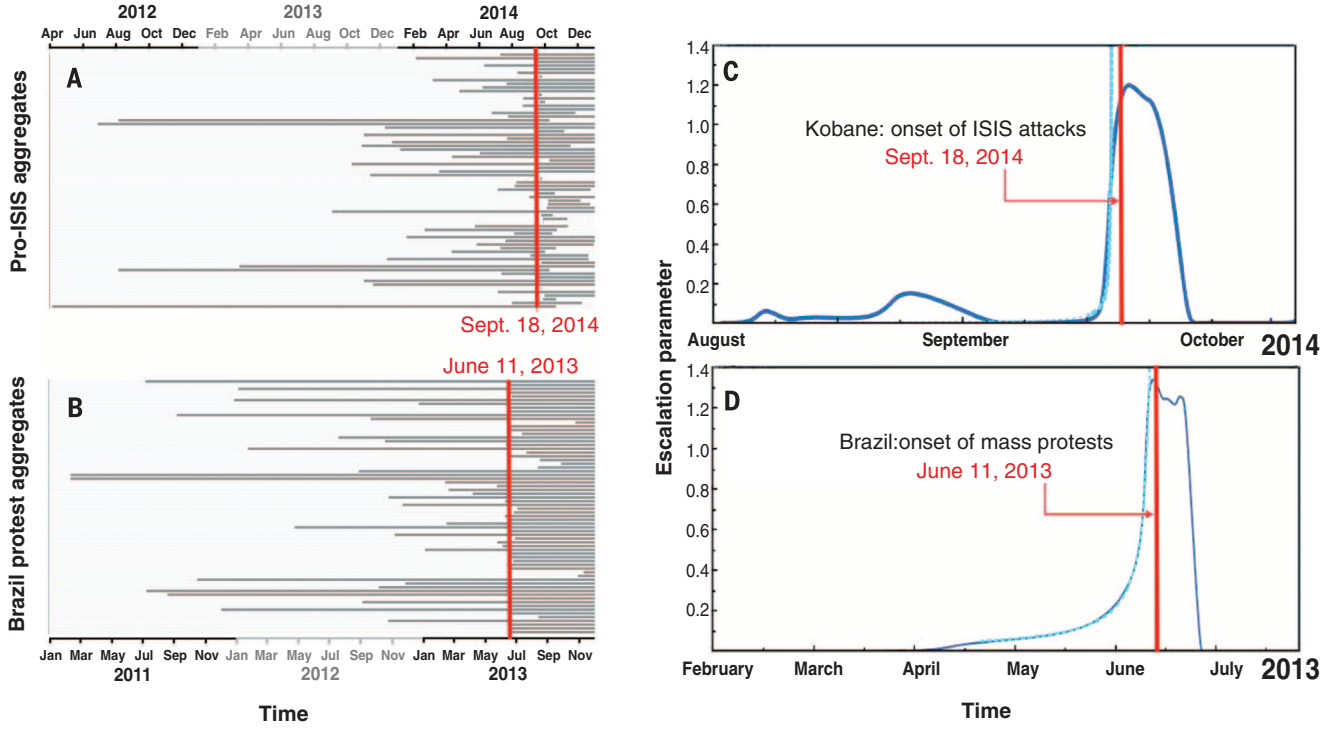


Fig. 2. Proliferation in online aggregate creation before the onsets of recent real-world campaigns (red vertical lines). (**A** and **C**) concern the unexpected assault by ISIS on Kobane in September 2014. (**B** and **D**) concern the unexpected outburst of protests in Brazil in June 2013, commonly termed the "Brazil Winter," which involved some violence and for which we were able to collect accurate information following the Intelligence Advanced Research Projects Activity (IARPA) OSI program (*14*, *15*). Horizontal bars in (A) and (B) show timelines for (A) pro-ISIS aggregates on VKontakte and (B) protestor aggregates on Facebook in Brazil. Each horizontal bar represents one aggregate. The aggregates are stacked separately along the vertical axis. [(C) and (D)] Divergence of escalation parameter b for aggregate creation (dark blue solid line) coincides with real-world onset at time T_c (vertical red line). The light blue dashed line shows theoretical form $(T_c - t)^{-1}$. The subsequent decrease in both curves likely occurs for system-specific reasons associated with coalition bombings starting in (C) and loss of public interest in (D).

N to be reasonably slowly varying, although this can be generalized (SM). Adding heterogeneity to the aggregate formation process (e.g., preference for similar or diverse follows) leaves the exponent $\alpha = 2.5$ unchanged, as do a variety of other generalizations (table S2) (*23*, *24*).

Our theoretical model generates various mathematically rigorous yet operationally relevant predictions. First, anti-ISIS agencies can thwart development of large aggregates that are potentially far more potent (*21*) by breaking up smaller ones. As shown in Fig. 3D, adding a simple cost into the model for shutting down an aggregate makes this strategy actually more effective than targeting the largest aggregates (SM). Second, if anti-ISIS agencies are insufficiently active in countermeasures and hence the overall rate at which they fragment pro-ISIS clusters becomes too small—specifically, if the aggregate fragmentation rate $v_{frag} < (N \ln N)^{-1}$—then pro-ISIS support will grow exponentially fast into one super-aggregate (fig. S11). Third, when fragmentation rates drop below a critical value $v_{frag}^{critical}$, the system enters a regime in which any piece of pro-ISIS material can spread globally across the pro-ISIS support network through contagion: $v_{frag}^{critical} = v_{coal} p/q$, with p and q representing the probabilities of follower-to-follower transmission and follower recovery, respectively (*25*). To prevent diffusion of potentially dangerous material and ideas, the fragmentation rate should be greater than $v_{coal} p/q$. Fourth, any online "lone wolf" actor will only truly be alone for short periods of time (on the order of weeks in Fig. 3A, for example) before being attracted into one aggregate or another through coalescence. Fifth, a systems-level tool emerges for detecting the future online emergence of new ISIS-like entities, which is to employ our methodology to determine whether a crude power-law distribution with α near 2.5 begins to emerge for aggregate support surrounding a particular theme.

At a more microscopic level, the data reveal that pro-ISIS aggregates exhibit the ability to collectively adapt in a way that can extend their lifetime and increase their maximal size (Fig. 4), despite the fact that each aggregate is an ad hoc group of followers who likely have never met, do not know each other, and do not live in the same city or country. For the civil protests, by contrast, we detected no such adaptations and no online predatory shutdowns, adding support to the notion that the pro-ISIS adaptations are a response to their high-pressured online environment. Figure 4, A to C, illustrates the remarkable speed, variety, and novelty of these adaptations, with 15% of aggregates exhibiting name changes; 7% exhibiting flips between online visibility (i.e., content open to any VKontakte user) and invisibility (i.e., content open only to current followers of the aggregate); and 4% exhibiting reincarnation in which an aggregate disappears completely and then reemerges at a later time with another identity but with most (e.g., >60%) of the same followers. Such reincarnation is not known to occur in real-world ecologies of living organisms. Figure 4D confirms that these adaptations tend to increase not only the maximum number of followers attracted into the aggregate (maximum size) but also its lifetime. The 0.9 value for the reincarnation lifetime can be understood as follows: Reincarnation involves the aggregate temporarily disappearing; therefore, an aggregate that uses reincarnation runs a high risk of losing followers because they do not know when, and with what identity, the core follower group will reemerge. Reincarnation hence tends to be used by aggregates that are attracting unusually high predation and would otherwise have had a much shorter lifetime. Reincarnation extends this lifetime beyond its otherwise much shorter value, but not enough to reach the value of 1 corresponding to aggregates that experience less intense shutdown pressure and hence do not employ adaptations. These observations open up the possibility to add evolutionary game theoretic features into our systems-level theory to explain the multiple use of particular adaptations by particular aggregates and their decision of when to adapt. A future generalized theory could prove possible, employing game theoretic ideas from (*26*), for example.

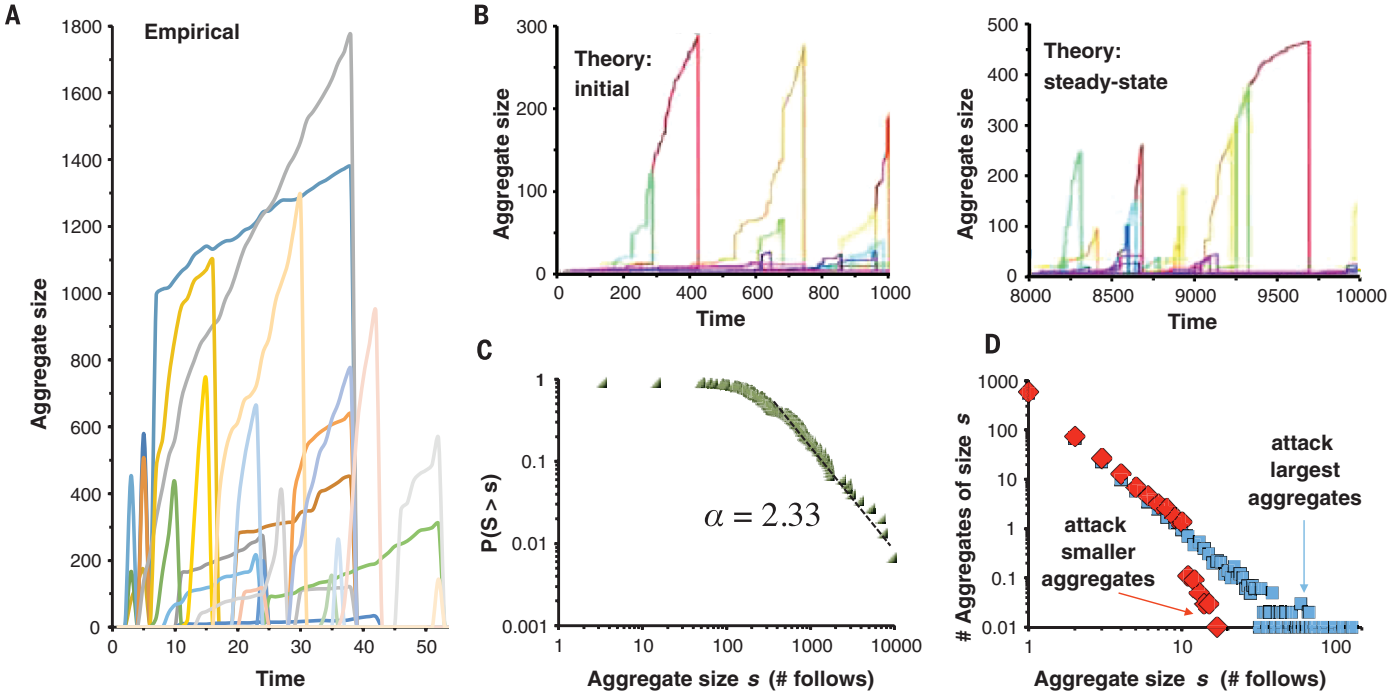
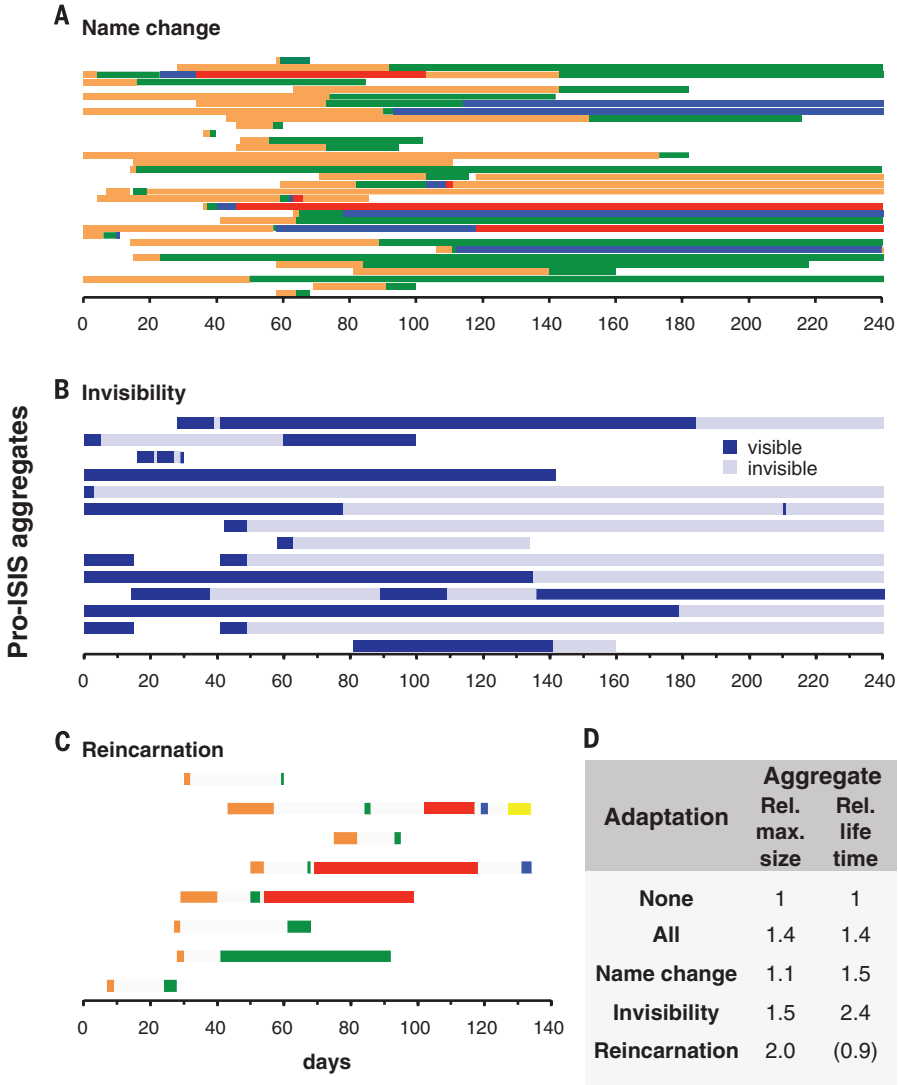


Fig. 3. Size dynamics of pro-ISIS aggregates. (**A**) Empirical size variation of pro-ISIS aggregates. Shark-fin shapes of all sizes emerge with shutdowns that are not strongly correlated. (**B**) Similar results are predicted by our theoretical model, irrespective of whether we consider the model's initial (left) or steady-state (right) dynamics. Here, the total number of potential follows is $N = 500$; however, the model's self-similar dynamics generate the same picture for any N with shark-fin shapes of all sizes. An aggregate grows by an individual linking in (i.e., size increases by 1) or by an existing aggregate linking in (i.e., size increases by >1), as shown by the color change. In (B), knowledge of the theory's microscopic dynamics allows us to denote each coalescence of a large aggregate by a color change, whereas in the empirical data (A), we maintain a constant color for each aggregate. (**C**) Complementary distribution function for the observed aggregate sizes. (**D**) Effect of intervention strategy involving dismantling smaller aggregates (SM). Using a larger N increases the vertical and horizontal scales without changing the main results (see fig. S10). Red diamonds: $s_{min} = 10$ and $s_{max} = 50$. Blue squares: $s_{min} = 200$ and $s_{max} = 1000$.



D

Adaptation	Aggregate Rel. max. size	Aggregate Rel. life time
None	1	1
All	1.4	1.4
Name change	1.1	1.5
Invisibility	1.5	2.4
Reincarnation	2.0	(0.9)

Fig. 4. Evolutionary adaptations. (**A** to **C**) Sample of pro-ISIS aggregate timelines showing evolutionary adaptations (shown by switches in colors) that tend to increase an aggregate's maximum attained size and extend its lifetime (D). Time is measured in days from 1 January 2015. In (A), the switch in colors within a given timeline indicates a switch in aggregate name. (B) Dark blue means the aggregate is visible (i.e., content open to any VKontakte user), while light blue means it is invisible (i.e., content open only to current followers of the aggregate). (C) Aggregate has a specific initial identity (orange), then disappears from the Internet for an extended time (white), then reappears with another identity shown by a switch in color. (**D**) Relative maximum aggregate size and relative lifetime for particular adaptations and their combinations, given as average values relative to the values for aggregates employing no adaptation. "All" corresponds to aggregates that use name change, invisibility, and reincarnation. See the text for explanation of the "(0.9)" entry.

More generally, our findings suggest that instead of having to analyze the online activities of many millions of individual potential actors worldwide (*27*), interested parties can shift their focus to aggregates, of which there will typically be only a few hundred. Our approach, combining automated data-mining with subject-matter expert analysis and generative model-building drawn from the physical and mathematical sciences, goes beyond existing approaches to mining such online data (*28*–*30*). Although recent reports (*31*) suggest that the amount of explicit pro-ISIS material online may have declined since summer 2015, it is possible that there is lower detection due to novel adaptations being employed—as in Fig. 4, but now likely more sophisticated.

REFERENCES AND NOTES

1. M. Weiss, H. Hassan, *ISIS: Inside the Army of Terror* (Regan Arts, New York, 2015).
2. BBC News, "Russian students targeted as recruits by Islamic State"; http://linkis.com/dmFgu (24 July 2015).
3. P. Gill, J. Lee, K. Rethemeyer, J. Horgan, V. Asal, *Int. Interact.: Empir. Theor. Res. Int. Relat.* **40**, 52–78 (2014).
4. R. Wrangham, L. Glowacki, *Hum. Nat.* **23**, 5–29 (2012).
5. J. E. Smith *et al.*, *Trends Ecol. Evol.* **31**, 54–66 (2016).
6. M. Kenney *et al.*, *Appl. Ergon.* **44**, 739–747 (2013).
7. J. N. Shapiro, *The Terrorist's Dilemma: Managing Violent Covert Organizations* (Princeton Univ. Press, 2013).
8. A. M. Morgenstern *et al.*, *Am. J. Phys.* **81**, 805 (2013).
9. G. King, J. Pan, M. E. Roberts, *Science* **345**, 1251722 (2014).
10. D. J. Watts, P. S. Dodds, *J. Consum. Res.* **34**, 441–458 (2007).
11. F. Karimi, C. E. Shoichet, D. Ford, "San Bernardino shooters were radicalized 'for quite some time,' FBI says." CNN News, 7 December 2015; www.cnn.com/2015/12/07/us/san-bernardino-shooting/index.html.
12. V. S. Subrahmanian, A. Mannes, A. Sliva, J. Shakarian, J. P. Dickerson, *Computational Analysis of Terrorist Groups: Lashkar-e-Taiba* (Springer, New York, 2013).
13. P. Gill, *Lone Actor Terrorists: A Behavioural Analysis* (Routledge, London, 2015).
14. P. Manrique *et al.*, Context matters: Improving the uses of big data for forecasting civil unrest: Emerging phenomena and big data. *IEEE Intelligence and Security Informatics* (IEEE, 2013), pp. 169–172.
15. J. Matheny, Test and evaluation in ACE and OSI IARPA (2013); available online at http://semanticommunity.info/@api/deki/files/21696/3-ACE_and_OSI_NIST_Brief.pdf.
16. J. Cadena *et al.*, *PLOS ONE* **10**, e0128879 (2015).
17. BBC News, "Anonymous targets IS sympathisers on Twitter"; www.bbc.com/news/technology-33608369 (21 July 2015).
18. BBC News, "Islamic State web accounts to be blocked by new police team"; www.bbc.com/news/world-europe-33220037 (22 June 2015).
19. N. Johnson *et al.*, *Science* **333**, 81–84 (2011).
20. A.-L. Barabasi, H. E. Stanley, *Fractal Concepts in Surface Growth* (Cambridge University Press, 1995).
21. V. Asal, S. M. Chermak, S. Fitzgerald, J. D. Freilich, *Crim. Justice Rev.* **2016**, 10.1177/0734016815626970 (2016).
22. N. F. Johnson, P. Jefferies, P. M. Hui, *Financial Market Complexity* (Oxford Univ. Press, 2003), chap. 5.
23. J. C. Bohorquez, S. Gourley, A. R. Dixon, M. Spagat, N. F. Johnson, *Nature* **462**, 911–914 (2009).
24. B. Ruszczycki, B. Burnett, Z. Zhao, N. F. Johnson, *Eur. Phys. J.* **72**, 289–302 (2009).
25. Z. Zhao *et al.*, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **81**, 056107 (2010).
26. M. A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Belknap Press, New York, 2006).
27. N. Perra *et al.*, *Phys. Rev. Lett.* **109**, 238701 (2012).
28. D. Koutra, D. Jin, Y. Ning, C. Faloutsos, Perseus: An interactive large-scale graph mining and visualization tool (DEMO), VLDB 2015, Kohala Coast, Hawaii, 31 August to 4 September 2015.
29. Q. Zhao, M. Erdogdu, H. He, A. Rajaraman, J. Leskovec, SEISMIC: A self-exciting point process model for predicting tweet popularity, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Sydney, 10 to 13 August 2015.
30. A. Stanton, A. Thart, A. Jain, P. Vyas, A. Chatterjee, P. Shakarian, arXiv:1508.01192v1 (2015).
31. J. M. Berger, H. Perez, The Islamic State's diminishing returns on Twitter. GW Program on Extremism (2016); available at https://cchs.gwu.edu/sites/cchs.gwu.edu/files/downloads/Berger_Occasional%20Paper.pdf.

ACKNOWLEDGMENTS

N.F.J. gratefully acknowledges partial support for preliminary work from Intelligence Advanced Research Projects Activity (IARPA) under grant D12PC00285 and recent funding under National Science Foundation (NSF) grant CNS1500250 and Air Force (AFOSR) grant 16RT0367. The views and conclusions contained herein are solely those of the authors and do not represent official policies or endorsements by any of the entities named in this paper. Data described are presented in an Excel file available in the supplementary materials, and code is provided in the SM document.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1459/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 to S3
Database S1
References

13 December 2015; accepted 12 May 2016
10.1126/science.aaf0675

BRAIN RESEARCH

Organizing conceptual knowledge in humans with a gridlike code

Alexandra O. Constantinescu,[1*†] Jill X. O'Reilly,[1,2,3†] Timothy E. J. Behrens[1,4*]

It has been hypothesized that the brain organizes concepts into a mental map, allowing conceptual relationships to be navigated in a manner similar to that of space. Grid cells use a hexagonally symmetric code to organize spatial representations and are the likely source of a precise hexagonal symmetry in the functional magnetic resonance imaging signal. Humans navigating conceptual two-dimensional knowledge showed the same hexagonal signal in a set of brain regions markedly similar to those activated during spatial navigation. This gridlike signal is consistent across sessions acquired within an hour and more than a week apart. Our findings suggest that global relational codes may be used to organize nonspatial conceptual representations and that these codes may have a hexagonal gridlike pattern when conceptual knowledge is laid out in two continuous dimensions.

Humans have a remarkable capacity for generalizing experiences to novel situations (*1*, *2*). It has been hypothesized that this capacity relies on a "cognitive map," allowing conceptual relationships to be navigated in a manner similar to that of space (*3*–*6*). Grid cells use a hexagonally symmetric code to organize spatial representations (*7*). Here we ask whether conceptual knowledge may also be organized by gridlike codes.

Human grid cells have been identified during intraoperative recordings (*8*) and are the likely source of a precise sixfold (hexagonal) symmetry in the functional magnetic resonance imaging (fMRI) signal, as a function of movement direction during virtual navigation (*9*–*11*). This hexagonal signal varies depending on whether the direction of moving in space is aligned or misaligned with the orientation of the grid (Fig. 1, D and E). Such a signal is a precise and unusual prediction for fMRI (supplementary text): It is predicted in the bulk activity because grid cells share a common grid axis (*12*–*14*), and conjunctive grid cells fire faster on average when movement is aligned to this axis (*9*). It pertains to the moving direction and cannot be explained by any characteristics of the currently experienced visual scene. It is hexagonal and therefore does not align on average to the cardinal directions. Its temporal waveform is different in every scan and every subject (fig. S4) and thus cannot be easily predicted by imaging artefacts (supplementary text and table S1).

This gridlike signal is not unique to the entorhinal cortex, but can be measured during spatial navigation in prescribed parts of the medial frontal, medial parietal, and lateral temporal cortices (*9*). Despite no report in rodents of grid cells outside the hippocampal formation, direct recordings during brain surgery in humans have confirmed gridlike firing patterns in some of these areas (*8*). This same network of brain regions, often referred to as the "default mode network" (*15*), is also regularly activated in nonspatial tasks that involve the manipulation of conceptual knowledge, such as memory (*16*), imagination (*17*), scene construction (*18*), valuation (*19*), and theory of mind (*20*), and in situations when subjects must generalize learned concepts to novel situations (*1*, *2*, *21*).

The ability to interact with knowledge in this flexible and generalizable way is the central advantage of maintaining an explicit cognitive map (*3*). Together with the regions' established role in nonspatial conceptual generalization, the finding of gridlike activity in these brain regions during spatial navigation therefore raises the possibility of common neural coding mechanisms for storing spatial and conceptual representations. Indeed, this hypothesis is strengthened by the findings that hippocampal cells (analogous to rodent place cells) encode individual concepts in humans (*22*) and sound frequency in rodents (*23*), and that rodent grid cell coding may not be restricted to spatial dimensions, but also represent time (*24*).

We used fMRI to test if humans use a hexagonally symmetric code when navigating through abstract conceptual representations. We designed a task analogous to the one used for navigation in physical space (*9*), with the notable difference that our dimensions were organized in an abstract, rather than physical, space.

Twenty-eight healthy subjects performed a stimulus-outcome (S-O) learning task in which they learned that bird stimuli (*25*) were associated with different Christmas symbols (Fig. 1A). The study therefore resembles other S-O learning tasks except that here, the bird stimuli were not independent fractals or symbols but, instead, they varied according to two continuous dimensions: the lengths of the neck and legs. Each stimulus could therefore be described within a two-dimensional conceptual "bird space" (Fig. 1B). Even though the features of this bird space were lengths as in physical maps, here they had to be extracted from a one-dimensional (vertical) visual space where different stimuli did not form any angles (Fig. 1A) and transformed into a two-dimensional conceptual map (Fig. 1B).

Participants had extensive experience of the bird space the day before scanning (*26*). Briefly, we first trained them to morph birds with specific neck:legs ratios, using a nonspatial controller (fig. S1). Next, the participants learned which bird stimuli were paired with outcomes by freely morphing the neck and legs dimensions. The outcome symbols would appear on screen whenever the morph matched the associated bird (fig. S2). This ensured that subjects became familiar with the entire bird space and not just with the S-O pairings. Indeed, participants progressively refined the locations of the outcomes through training (Fig. 1F and fig. S6). We periodically tested their knowledge of the bird space by asking them to find specific outcomes from arbitrary start positions, by correctly choosing the appropriate neck:legs ratio. When asked to make such ballistic movements, subjects significantly increased their precision through training, and therefore learned to visualize the target bird (Fig. 1G and fig. S7). In the scanner, participants continued to improve performance each day, possibly because they received fresh training before each day's scanning (Fig. 1G and fig. S8).

Each subject participated in two to four separate fMRI sessions, spanning two separate days, at least 1 week apart. In each trial during scanning, subjects watched a video of a bird morphing according to a predefined neck:legs ratio (Fig. 1C and movie S1). They were then instructed to imagine the outcome if the bird continued to morph with the same neck:legs ratio. In some trials, they had to choose one of three offered outcomes: two outcomes they were trained with and a "no outcome" option (black square). Participants reached a performance of 72.8 ± 1.0% accuracy in predicting outcomes, and 0 out of 28 reported conceiving of the relationships between birds or outcomes as lying in a spatial map (*26*).

To test whether the fMRI signal had hexagonal symmetry, as a proxy for grid cells, we ensured that the orientation of the trajectories for movement in bird space was dissociated from the properties of the visual scene. That is, trajectories with the same orientation were formed by different stimuli, and trajectories with different orientations could pass through the same stimulus. Indeed, the critical hexagonal symmetry regressors described below did not share more than 5% variance with any tested basic visual property of the stimuli, outcomes, or behavioral accuracy in any subject (Fig. 1I). The trajectories were sampled evenly across directions both in sum and when separated according to outcome (Fig. 1H and fig. S5).

Despite the absence of any hexagonal modulation of these confounding factors, we found a hexagonal modulation effect in the neural activity. We first identified hexagonally symmetric signals across the whole brain and then we focused on those regions where the effect was

[1]Oxford Centre for Functional MRI of the Brain, University of Oxford, John Radcliffe Hospital, Headington, Oxford OX3 9DU, UK. [2]Department of Experimental Psychology, University of Oxford, 9 South Parks Road, Oxford OX1 3UD, UK. [3]Donders Institute, Radboud University, Nijmegen, Netherlands. [4]Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London WC1N 3BG, UK.
***Corresponding author. Email: alexandra.constantinescu@magd.ox.ac.uk (A.O.C.); behrens@fmrib.ox.ac.uk (T.E.J.B.)**
†These authors contributed equally to this work.

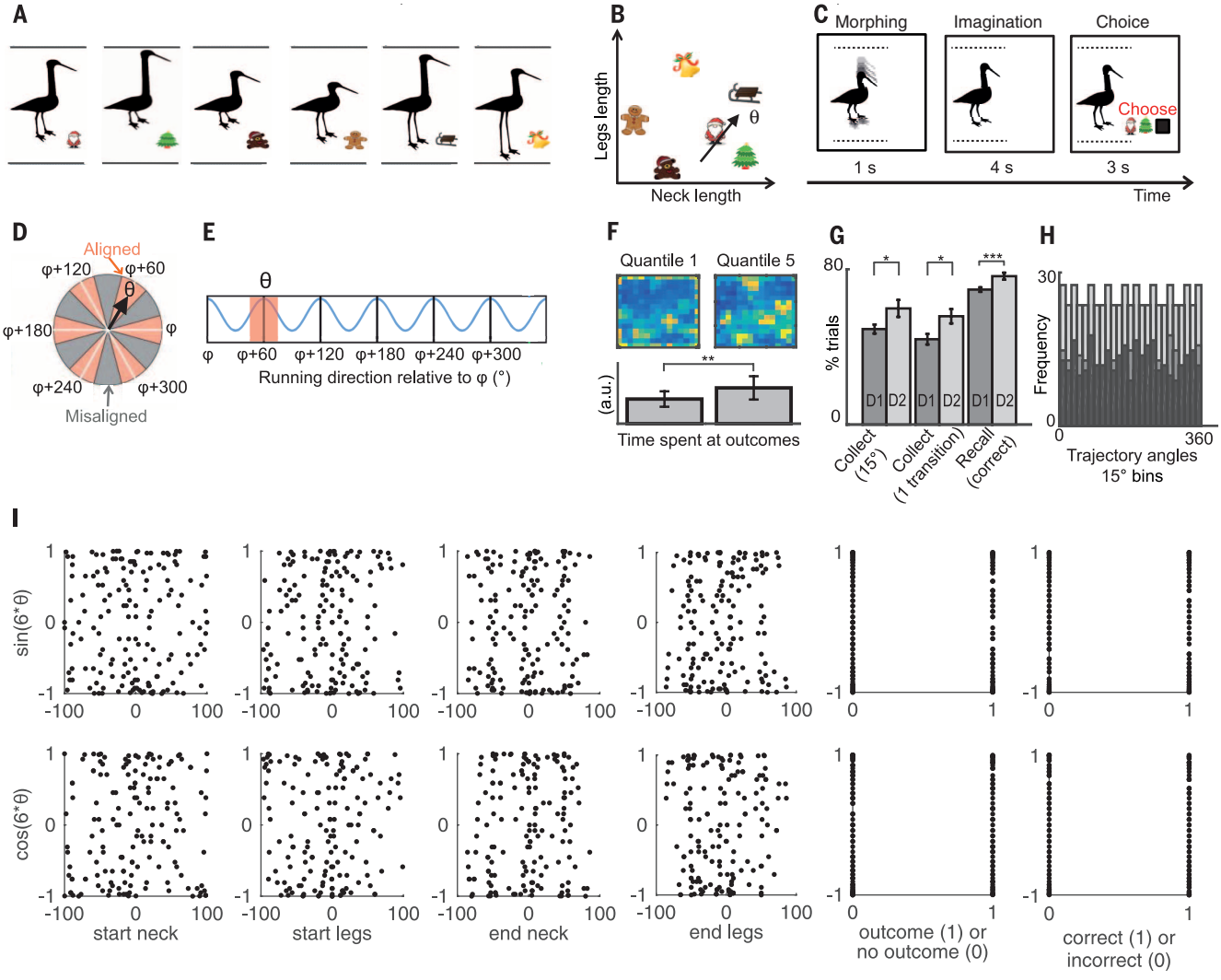
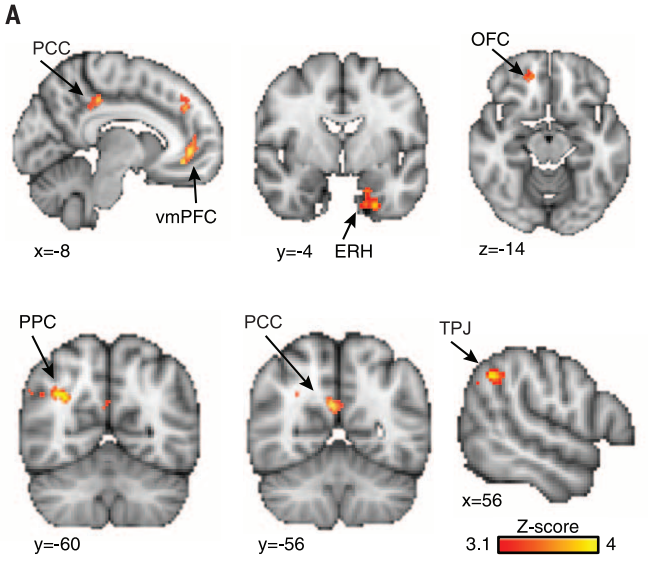


Fig. 1. Experimental design for navigation in abstract space. (**A**) Subjects were trained to associate stimuli (birds) with outcomes (Christmas symbols). (**B**) Example trajectory in abstract space. A location in this abstract space was represented by a bird stimulus. A trajectory was equivalent to visually morphing one bird into another (Fig. 1C). The direction θ of the trajectory depended on the ratio of the rates of change of the legs and the neck (movie S1). Subjects were not consciously aware that these associations could be organized in a continuous "bird space." (**C**) Example trial corresponding to the trajectory with direction θ. (**D**) Trajectories can be categorized as aligned (red sectors) or misaligned (gray sectors) with the mean orientation φ of the hexagonal grid. Note that φ is different for each participant [see (*26*) for details on how φ was calculated]. Here, the direction θ is aligned with the grid. (**E**) fMRI markers of grid cells showing hexagonal symmetry: The signal is bigger for trajectories aligned versus those misaligned with the grid. (**F**) (Top) Color-coded trajectory maps illustrating time spent in each part of the environment during the "explore" task in the first (quantile1) and last parts of training (quantile5). Yellow is maximum and dark blue is 0. (Bottom) Barplots showing the amount of time spent at the locations/stimuli paired with outcomes in each epoch relative to the total time spent navigating ("time at outcomes" quantile1 versus quantile5, $t_{22} = -3.17$, $^{**}P < 0.01$). (**G**) In the "collect" task, participants made significant improvements in training day 2 compared to training day 1: The percentage of trials with an angle error < 15° ($t_{33} = 2.37$, $^{*}P < 0.05$) and with only one transition increased ($t_{33} = 2.55$, $^{*}P < 0.05$). In the "recall" task, participants made significantly more correct responses in day 2 compared to day 1 ($t_{41} = 3.89$, $^{***}P < 0.001$). (**H**) Example data from the most commonly used schedule: even distribution of trajectory angles across all trials (light gray), outcome trials (medium gray) and nonoutcome trials (dark gray). (**I**) Example data from the most commonly used schedule: We tested if the sin(6θ) and cos(6θ) regressors correlated with multiple confounding factors. These regressors did not correlate with the start neck, start legs, end neck, and end legs lengths, whether the subject responded accurately or whether the morph passed through an outcome (all coefficients of determination R^2 averaged across all subjects < 0.02). [(D) and (E) are adapted by permission from *Nature* (*9*)]

strongest. This approach allowed us to test in an unbiased way if this hexagonal symmetry had a consistent grid angle across two sessions acquired on the same day and more than a week apart.

To identify brain regions sensitive to hexagonal symmetry, we used a *Z*-transformed F-statistic to test for a significant modulation of the fMRI signal by any linear combination of sin(6θ) and cos(6θ), where θ is the trajectory angle in bird space (*26*) (fig. S3). We found hexagonal symmetry in a network of brain regions that overlapped anatomically with the network found during navigation in physical space (*8*, *9*) and with the default-mode network (*15*) (Fig. 2A and fig. S10). However, although these brain regions all survived whole-brain cluster correction, this quadrature test could overestimate the *Z* scores (*26*). Thus, we used this test not for statistical inference per se, but rather to create orthogonal regions of interest that allowed us to test in an unbiased way if the grid angle was consistent across separate experimental sessions. This was possible because the quadrature test was independent from the phase of the periodic signal, that is, the grid



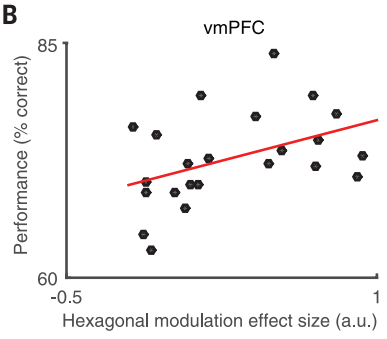
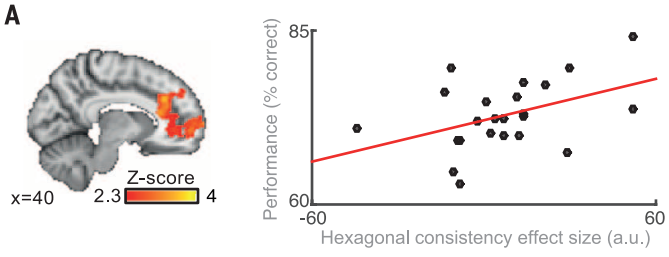
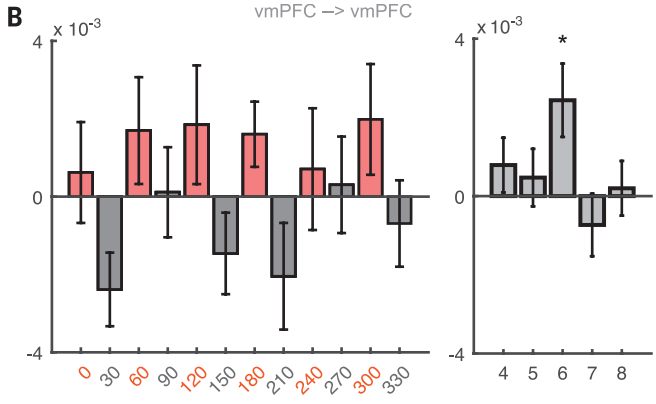


Fig. 2. Identifying hexagonally symmetric signals across the whole brain. (**A**) Hexagonal modulation in a network of brain regions including the medial prefrontal cortex, with a peak in its ventral region (vmPFC; peak Montreal Neurological Institute coordinates –8/42/0, peak *Z* score = 4.09), the medial entorhinal (ERH; –18/0/–38; *Z* = 4.41), the orbitofrontal (OFC; 6/44/–10; *Z* = 4.27), the posterior cingulate (PCC; 0/–32/28; *Z* = 4.3), retrosplenial (RSC; 6/–52/24; *Z* = 4.73), and lateral parietal cortices (LPC; 30/–62/28; *Z* = 4.96) and the temporoparietal junction (TPJ; 52/–42/40; *Z* = 4.13). For visualization purposes, the maps are cluster corrected at a cluster threshold *Z* = 3.1 and *P* < 0.05 for all brain regions apart from the ERH, where we used a more lenient threshold of *Z* = 2.3 and *P* < 0.05. (**B**) Subjects who performed better at the task had significantly more hexagonal signal modulation in the vmPFC (correlation coefficient *r* = 0.432, *P* = 0.039).





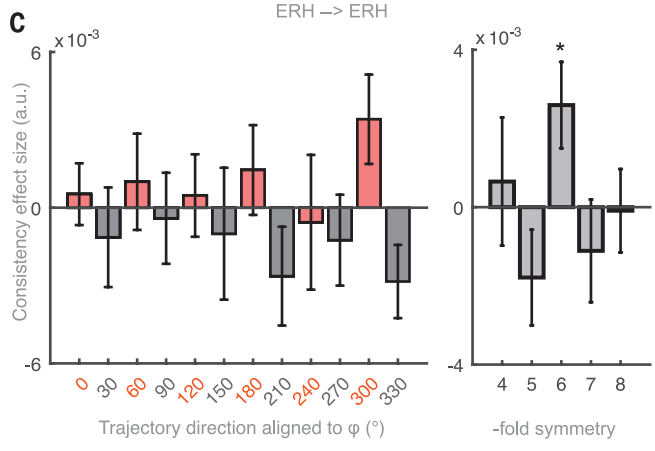


Fig. 3. Grid angle consistency between separate sessions acquired within the same day. (**A**) Left: Whole-brain level grid angle consistency in vmPFC (cluster corrected *Z* = 2.3 and *P* < 0.05; 16/54/–2; *Z* = 3.76, *P* < 0.0001). Right: participants with higher hexagonal consistency performed more accurately on the task (*r* = 0.431, *P* = 0.039). (**B** and **C**) Left panels: Sixfold modulation signals aligned to the same grid angle in the vmPFC (t_{26} = 2.61, **P* < 0.05) and ERH (t_{27} = 2.36, **P* < 0.05). The effect is plotted separately for all aligned (red) and misaligned (gray) trajectories. Right panels: This effect was specific for sixfold, but not any other control periodicities between four- and eightfold (all *P* > 0.15).

angle. We focused on brain regions where grid cells have been recorded in humans during spatial tasks (*8*): that is, the anterior cingulate/medial prefrontal cortex (mPFC) and entorhinal cortex (ERH).

Using the peak coordinate of the hexagonal modulation signal in the mPFC located ventrally (vmPFC), we found that subjects with greater hexagonal modulation had a more accurate performance at the task (Fig. 2B). This region has also been shown to correlate with the performance in memory and conceptual knowledge tests (*1*, *21*). As previously described for spatial hexagonal symmetries (*9*, *10*), we next asked whether the grid angle to which this hexagonal modulation was aligned was consistent between separate experimental sessions. We thus used the data from one session to estimate the grid angle for a given participant, using the beta coefficients for the sin(6θ) and cos(6θ) regressors (*26*) (fig. S3). We then took the data from a separate session and looked for differences in activation between trials in which the trajectories were aligned versus misaligned to this hexagonal grid. This was achieved with the regressor cos(6[θ(*t*) – φ]), where θ(*t*) is the trajectory orientation in trial *t* and φ is the mean grid orientation across the region of interest (*9*, *10*). This "cross-validation" procedure was counterbalanced

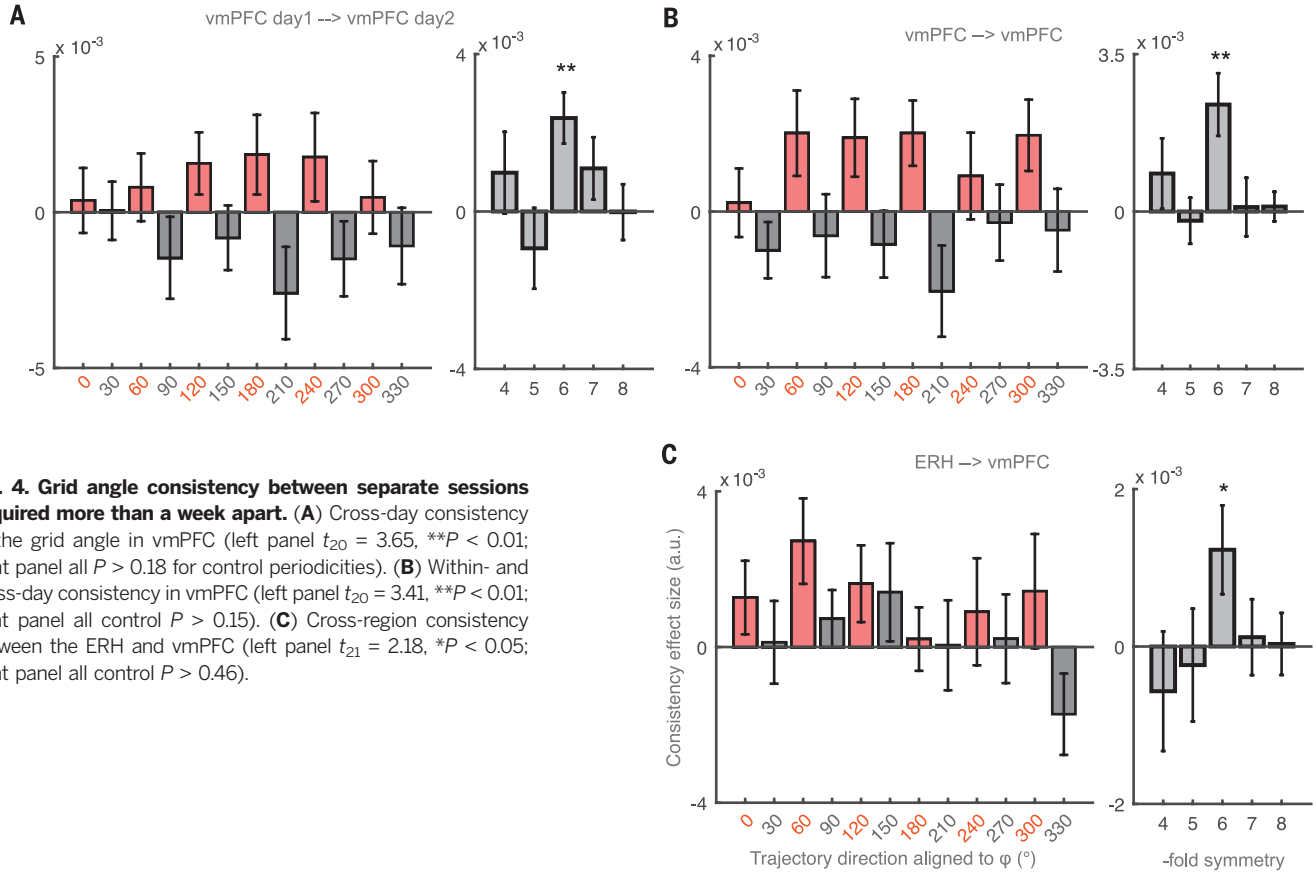


Fig. 4. Grid angle consistency between separate sessions acquired more than a week apart. (**A**) Cross-day consistency of the grid angle in vmPFC (left panel t_{20} = 3.65, $^{**}P < 0.01$; right panel all $P > 0.18$ for control periodicities). (**B**) Within- and cross-day consistency in vmPFC (left panel t_{20} = 3.41, $^{**}P < 0.01$; right panel all control $P > 0.15$). (**C**) Cross-region consistency between the ERH and vmPFC (left panel t_{21} = 2.18, $^{*}P < 0.05$; right panel all control $P > 0.46$).

across sessions. We then performed a one-sample *t* test across the group on the resulting regression coefficients.

Using this approach, we tested for hexagonal consistency between separate sessions acquired half an hour apart. We found such an effect at the whole-brain level in the vmPFC (Fig. 3A, left). To test if this effect was a function of the presence of outcomes, we included in the design matrix a confound regressor that modeled out the effect of outcomes. We found that the map did not change (fig. S11). Again, the consistency effect correlated with behavior in subjects who performed better at the task (Fig. 3A, right). Thus, we replicated the finding that participants with more gridlike representations performed better at the task, using two independent analyses. The same neural signal was not predicted by the speed of learning of the task during training (*26*) (Fig. 1, F and G). Gridlike activity was therefore related to the current performance rather than the trajectory of learning.

Next, to examine the pattern underlying this hexagonal effect, we separated all aligned (red) and misaligned (gray) trajectories (*26*) (Fig. 1D). The signal in the vmPFC was significantly higher for aligned than misaligned trajectories (Fig. 3B). The same pattern appeared in ERH (Fig. 3C). This effect was significant only for sixfold but not control four-, five-, seven- and eightfold symmetries (Fig. 3, B and C, right panels).

We also tested for hexagonal consistency between separate sessions acquired more than a week apart. We found a significant effect in the vmPFC (Fig. 4A). When grouping together within- and between-day data, the consistency effect was strongest in the vmPFC (Fig. 4B). Moreover, we found hexagonal consistency also between the ERH and vmPFC, suggesting that different brain regions may contain gridlike activity that is aligned to the same angle (Fig. 4C). Again, all these effects were significant only in hexagonal but not control symmetries (Fig. 4, right panels).

Although the coarse nature of the fMRI signal urges caution in making conclusions at the level of neuronal codes, we have reported an unusually precise hexagonal modulation of the fMRI signal during nonspatial cognition. When subjects perform this abstract cognitive task, this signal exists in a set of brain regions markedly similar to those observed when subjects run in a virtual reality spatial environment (*9*) (fig. S9), despite profound differences in the cognitive and perceptual demands of the two tasks. The hexagonal grid is consistently oriented across sessions that are acquired both half an hour and more than a week apart. Together, this evidence suggests that gridlike codes that are known to underlie spatial navigation, and recently discovered in the temporal dimension (*24*), can also be used to organize abstract knowledge of the type that is difficult to study in nonhuman species. In the event that such conceptual grid cells can be recorded directly, it will be of interest to know whether they share relationships (such as relative phases) across spatial and conceptual tasks, suggesting that conceptual tasks can be solved by subconsciously mapping abstract dimensions onto preexisting spatial maps, or whether new organizations can emerge to represent conceptual problems. It will also be informative to study how such cells behave in conceptual problems that are not easily mapped onto continuous two-dimensional spaces (*27*, *28*).

REFERENCES AND NOTES

1. D. Kumaran, J. J. Summerfield, D. Hassabis, E. A. Maguire, *Neuron* **63**, 889–901 (2009).
2. H. C. Barron, R. J. Dolan, T. E. J. Behrens, *Nat. Neurosci.* **16**, 1492–1498 (2013).
3. E. C. Tolman, *Psychol. Rev.* **55**, 189–208 (1948).
4. J. O'Keefe, L. Nadel, *The Hippocampus as a Cognitive Map* (Oxford Univ. Press, 1978).
5. G. Buzsáki, E. I. Moser, *Nat. Neurosci.* **16**, 130–138 (2013).
6. H. Eichenbaum, N. J. Cohen, *Neuron* **83**, 764–770 (2014).
7. T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, E. I. Moser, *Nature* **436**, 801–806 (2005).
8. J. Jacobs *et al.*, *Nat. Neurosci.* **16**, 1188–1190 (2013).
9. C. F. Doeller, C. Barry, N. Burgess, *Nature* **463**, 657–661 (2010).
10. L. Kunz *et al.*, *Science* **350**, 430–433 (2015).
11. A. J. Horner, J. A. Bisby, E. Zotow, D. Bush, N. Burgess, *Curr. Biol.* **26**, 842–847 (2016).
12. C. Barry, R. Hayman, N. Burgess, K. J. Jeffery, *Nat. Neurosci.* **10**, 682–684 (2007).
13. H. Stensola *et al.*, *Nature* **492**, 72–78 (2012).
14. T. Stensola, H. Stensola, M.-B. Moser, E. I. Moser, *Nature* **518**, 207–212 (2015).
15. M. D. Fox, M. E. Raichle, *Nat. Rev. Neurosci.* **8**, 700–711 (2007).
16. J. R. Binder, R. H. Desai, W. W. Graves, L. L. Conant, *Cereb. Cortex* **19**, 2767–2796 (2009).
17. D. L. Schacter *et al.*, *Neuron* **76**, 677–694 (2012).

18. D. Hassabis, E. A. Maguire, *Trends Cogn. Sci. (Regul. Ed.)*. **11**, 299–306 (2007).
19. J. A. Clithero, A. Rangel, *Soc. Cogn. Affect. Neurosci.* **9**, 1289–1302 (2014).
20. R. Saxe, S. Carey, N. Kanwisher, *Annu. Rev. Psychol.* **55**, 87–124 (2004).
21. R. G. Benoit, K. K. Szpunar, D. L. Schacter, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 16550–16555 (2014).
22. R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, I. Fried, *Nature* **435**, 1102–1107 (2005).
23. D. Aronov, R. Nevers, D. W. Tank, *Proceedings of the Computational and Systems Neurosciscience Meeting* **T-29**, 41 (2016).
24. B. J. Kraus *et al*., *Neuron* **88**, 578–589 (2015).
25. T. Davis, R. A. Poldrack, *Cereb. Cortex* **24**, 1720–1737 (2014).
26. Materials and methods are available as supplementary materials on *Science* Online.
27. Y. Dordek, D. Soudry, R. Meir, D. Derdikman, *eLife* **5**, 682 (2016).
28. K. L. Stachenfeld, M. M. Botvinick, S. J. Gershman, *NIPS Proceedings, Advances in Neural Information Processing Systems*, 2528–2536 (2014).

ACKNOWLEDGMENTS

We thank N. Burgess, Z. Kurth-Nelson, H. Barron, L. Hunt, M. Garvert, and T. Muller for helpful discussions and comments on this manuscript, as well as the anonymous peer reviewers who greatly improved the manuscript. We thank S. Jbabdi for help with data collection and analysis, as well as A. Winkler and L. Griffanti for guidance. We also thank all participants for volunteering in this study. The study was supported by a Wellcome Trust 4-year Ph.D. studentship (099715/Z/12/Z) to A.O.C., a Medical Research Council Career Development Award (MR/L019639/1) to J.X.O.R., and a Wellcome Trust Senior Research Fellowship (WT104765MA) together with a James S. McDonnell Foundation Award (JSMF220020372) to T.E.J.B. The authors declare no conflicts of interest. All raw data are archived at the Oxford Centre for Functional MRI of the Brain.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1464/suppl/DC1
Materials and Methods
Supplementary text
Figs. S1 to S11
Table S1
Movie S1
References (*29*–*35*)

16 December 2015; accepted 17 May 2016
10.1126/science.aaf0941

ADDICTION RESEARCH

Carrots and sticks fail to change behavior in cocaine addiction

Karen D. Ersche,[1]* Claire M. Gillan,[1,2] P. Simon Jones,[1] Guy B. Williams,[1] Laetitia H. E. Ward,[1] Maartje Luijten,[3] Sanne de Wit,[4] Barbara J. Sahakian,[1] Edward T. Bullmore,[1,5,6] Trevor W. Robbins[1]

Cocaine addiction is a major public health problem that is particularly difficult to treat. Without medically proven pharmacological treatments, interventions to change the maladaptive behavior of addicted individuals mainly rely on psychosocial approaches. Here we report on impairments in cocaine-addicted patients to act purposefully toward a given goal and on the influence of extended training on their behavior. When patients were rewarded for their behavior, prolonged training improved their response rate toward the goal but simultaneously rendered them insensitive to the consequences of their actions. By contrast, overtraining of avoidance behavior had no effect on patient performance. Our findings illustrate the ineffectiveness of punitive approaches and highlight the potential for interventions that focus on improving goal-directed behavior and implementing more desirable habits to replace habitual drug-taking.

Why do some people take drugs by any possible means, seemingly without regard for the consequences? Actions normally constrained by their outcome become "out of control" in drug-addicted individuals, who fail to stop taking drugs despite being aware that continuing drug use provides little pleasure while inflicting considerable damage on their lives. Even the prospect of contracting an infectious disease fails to deter these individuals from sharing drug paraphernalia. Such maladaptive and ill-judged behaviors may be explained in terms of aberrant learning processes (*1*), where drug-taking is a learned behavior initially directed toward a conscious desire to enjoy a rush or avoid feelings of discomfort. Such goal-directed actions, whether appetitive or avoidant, are modulated by their outcomes. Following extended practice, however, drug-taking may deteriorate into a stimulus-driven habit that is elicited by antecedent stimuli and is thus performed regardless of any goals (*2*). This proposal is consistent with the notion of behavior being jointly regulated by goal-directed and habitual brain systems (*3*, *4*) and the disruption of this balance during the course of addiction (*1*).

Maladaptive behavior in drug-addicted individuals may thus result from impairments in goal-directed control, an enhanced propensity to develop stimulus-driven habits, or a combination of these factors. Preclinical evidence supports both accounts. Exposure to either cocaine or stress amplifies the transition from goal-directed to stimulus-driven behavior (*5*, *6*). Cocaine administration also diminishes information processing about consequences, leading to failures to adjust behavior during goal reevaluation (*7*).

We studied 125 participants to determine whether a newly learned behavior is under voluntary (goal-directed) or habitual (stimulus-driven) control using both positive and negative reinforcement. Seventy-two individuals met the DSM-IV-TR criteria for cocaine dependence and were actively using cocaine, as verified by urine screen (*8*), whereas 53 healthy control volunteers had no history of chronic drug or alcohol abuse (table S1). Participants learned by trial and error that an action was associated with a particular outcome, such as earning points toward a monetary reward (Fig. 1A) or avoiding an unpleasant electrical shock (Fig. 2, A and B). We then reduced the value of previously reinforcing outcomes by discontinuing point allocation for certain outcomes in the appetitive task (Fig. 1B) and physically disconnecting participants from the electrical stimulator in the avoidance task (Fig. 2C). We then tested whether participants made fewer responses to obtain or avoid the (now) devalued outcome, reflecting a goal-directed strategy, or whether they maintained their previously learned behavior despite outcome devaluation, as an index of habit.

In participants with cocaine use disorder (CUD), instrumental learning performance fell significantly short of that of control volunteers, irrespective of whether the goal was to make responses to obtain symbolic rewards or to avoid electrical shocks (Figs. 1A and 2B). However, depending on the type of reinforcement, prolonged training had a differential effect on the behavior of these individuals. For appetitive behavior, extensive training rendered CUD patients less sensitive to outcome devaluation (Fig. 1B). They persistently responded to stimuli previously associated with reward, irrespective of whether their behavior was actually rewarded or not (Fig. 1C). In fact, the shift toward habitual responding improved their response rate to the valued outcome (Fig. 1C). The strong habit bias in the slip-of-action test was not due to executive impairments (*9*, *10*), which were assessed separately in a control task (Fig. 1D) and included as a covariate in the statistical model.

By contrast, overtraining avoidance behavior had no effect on task performance in individuals with CUD. Despite intact fear conditioning (Fig. 2B), CUD patients continued to show attenuated avoidance responses to the conditioned stimulus

[1]Departments of Psychiatry, Psychology, and Clinical Neurosciences and the Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, UK. [2]Department of Psychology, New York University, 6 Washington Place, New York, NY 10003, USA. [3]Behavioural Science Institute, Radboud University, Nijmegen, Netherlands. [4]Department of Clinical Psychology, University of Amsterdam, Amsterdam, Netherlands. [5]Cambridgeshire and Peterborough National Health Service Foundation Trust, Cambridge, UK. [6]Immunopsychiatry Discovery Performance Unit, Alternative Discovery and Development Division, GlaxoSmithKline R&D, Stevenage SG1 2NY, UK.
*Corresponding author. Email: ke220@cam.ac.uk

(CS) associated with a shock, even after extended training (Fig. 2D). Such impairments in the initiation of goal-directed avoidance behavior have previously been reported in animals after dopamine receptor blockade (*11*) or experimental lesions of dopamine neurons (*12*). Although CUD patients undervalued the aversive outcome, overtraining did not change their sensitivity to outcome devaluation, either in terms of behavior or skin conductivity. As shown in Fig. 2D, CUD patients' responses were comparable to controls when the CS was no longer associated with a shock.

In light of the high prevalence of comorbid addictions in CUD, we sought to determine the extent to which the increased formation of appetitive habits and the persistent deficiencies in avoiding aversive outcomes resulted from cocaine addiction specifically or from addiction to other drugs. We also assessed the influence of vulnerability factors such as impulsivity-compulsivity traits, stress, and poor instrumental learning performance (*8*). Addiction to cocaine, but not to other drugs, explained ~13% of the variance of appetitive habits in the slip-of-action test (coefficient of determination $R^2 = 0.13$; $F_{4,117} = 4.48$, $P = 0.002$). However, reduced performance accuracy during training ($\beta = -0.410$, $P < 0.001$) and higher numbers of stressful life events ($\beta = 0.30$, $P = 0.015$) were factors of even greater weight in the model, accounting for one-third of the variance ($R^2 = 0.31$; $F_{8,113} = 6.32$, $P < 0.001$). Hence, our results suggest that, in individuals with prior exposure to cocaine and stress, impairments in instrumental learning lead to a shift from goal-directed to goal-independent habitual behavior.

We also applied a similar model to examine attenuated avoidance responses to the valued CS in extinction (table S2), revealing that addiction to cocaine (but not to other drugs) accounted for only 9% of the variance ($R^2 = 0.09$; $F_{4,119} = 2.82$, $P = 0.028$). High levels of impulsivity ($\beta = 0.18$, $P = 0.047$) and low avoidance accuracy during overtraining ($\beta = -0.67$, $P < 0.001$)—both associated with reduced striatal dopaminergic neurotransmission (*12*, *13*)—were the strongest predictors in this model, accounting for more than half of the variance of attenuated avoidance ($R^2 = 0.52$; $F_{8,115} = 15.85$, $P < 0.001$). These results are consistent with preclinical evidence for impulsivity predicting compulsive cocaine-seeking, even in the face of aversive consequences (*14*).

Our data provide compelling evidence for impairments in instrumental learning in CUD, regardless of affective valence and whether rewards were primary (shock) or secondary (monetary). In the case of appetitive learning, increased habitual responding may either be an indirect consequence of poor goal-directed action (*7*) or result from stronger habit learning. Both explanations would be consistent with disruptions of the balance between goal-directed and habitual control hypothesized to underlie compulsive cocaine-seeking (*1*). By contrast, impaired performance for instrumental avoidance in CUD patients

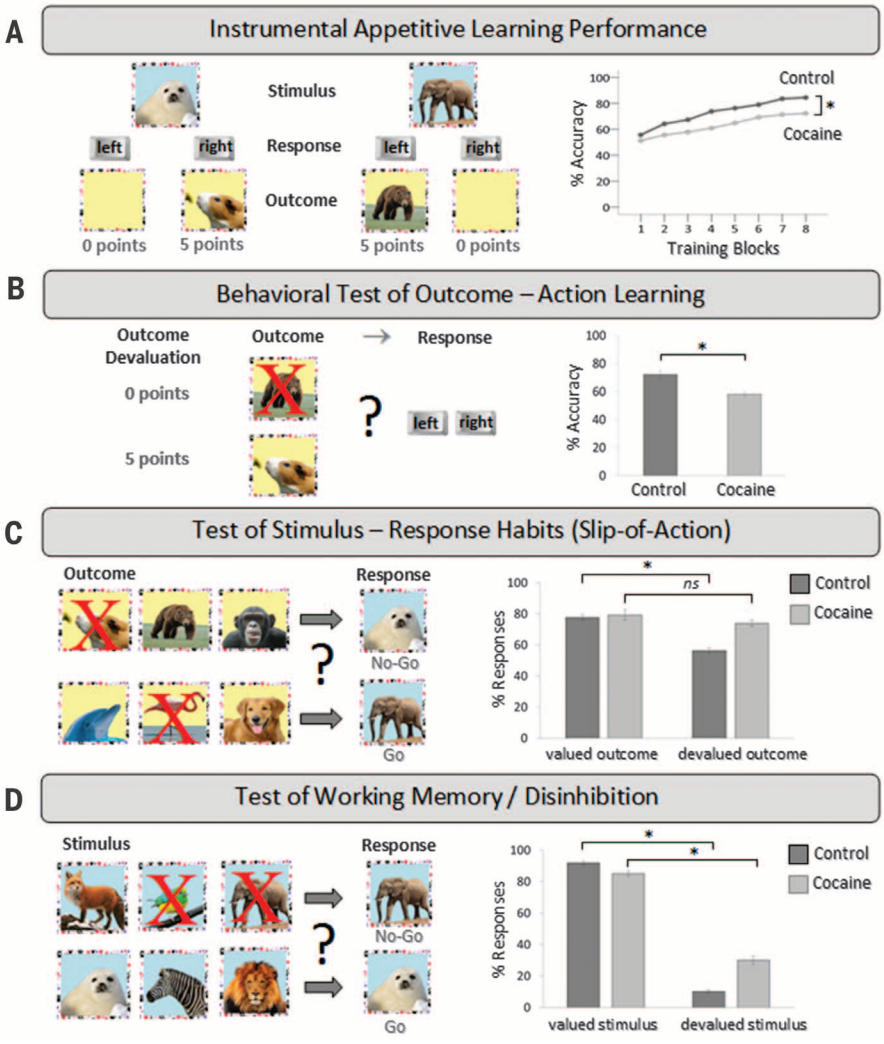


Fig. 1. Appetitive instrumental learning task. (**A**) Participants learned by trial and error which response associated with an animal picture gained them points. Feedback was provided by a picture of another animal, coupled with a number of points, or an empty box with no points. Goal-directed discrimination learning performance improved steadily in all participants over eight training blocks ($F_{6,684} = 43.98$, $P < 0.001$), but performance accuracy in individuals with CUD was reduced compared with that in control volunteers ($F_{1,121} = 20.19$, $P < 0.001$). (**B**) Participants were instructed that some of the pictures that were previously associated with points would no longer lead to point increases. Sensitivity to outcome devaluation was tested by simultaneous presentation of two outcome-related pictures and the instruction to select the response leading to a valued outcome without providing performance feedback. CUD patients showed significant impairments when outcome-action knowledge was tested behaviorally ($t_{88.2} = 3.83$, $P < 0.001$). (**C**) Slip-of-action test to determine the balance between goal-directed and habitual responses: Participants were asked to selectively respond to those stimuli still associated with reward and to withhold responding to stimuli that had been devalued. (For demonstration only, we indicated "go" and "no-go" below the pictures to denote the correct response.) Habitual behavior is reflected by continued responses to devalued outcomes, implying reduced sensitivity to outcome value. We observed a highly significant group–by–outcome-value interaction. CUD patients responded significantly more often than controls to the stimuli associated with the devalued outcome ($t_{121} = -4.72$, $P < 0.001$), whereas the level of responding toward valued outcomes did not differ between the groups ($t_{121} = -0.65$, $P = 0.520$). (**D**) Immediately after the slip-of-action test, a control task was introduced: Participants were instructed to respond only to those stimuli still associated with reward and to withhold responding to devalued stimuli. All participants responded more frequently to stimuli associated with the valued rather than the devalued outcome ($F_{1,121} = 111$, $P < 0.001$), but this difference was significantly smaller in CUD patients ($F_{1,121} = 42.10$, $P < 0.001$). (In all panels, error bars denote SEM, *ns* indicates $P > 0.05$, and asterisks indicate $P < 0.05$.) Analysis of covariance showed that executive impairments in the control task were not sufficient to account for the impaired "slip-of-action" performance. The significant group–by–outcome-value interactions in (C) survived statistical correction ($F_{1,120} = 8.79$, $P = 0.004$), indicating enhanced habitual control (see text).

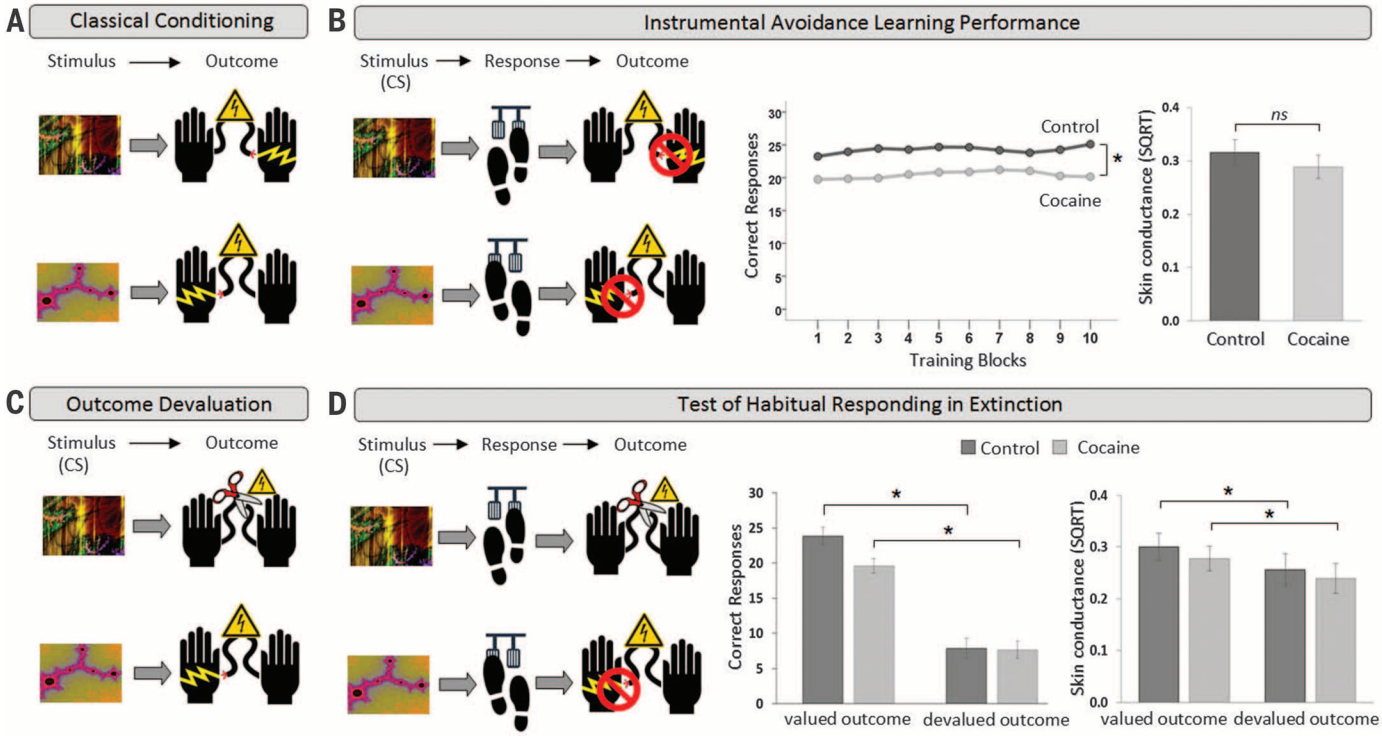


Fig. 2. Avoidance instrumental learning task. (**A**) Participants were trained to associate distinctive visual stimuli with an electrical shock to one wrist or the other. (**B**) Participants were instructed to avoid receiving shocks by pressing a foot-pedal on the side corresponding to the wrist where they were expecting to receive an electrical shock in response to the appearance of the CS. Individuals with CUD made significantly fewer successful avoidance responses compared with controls ($F_{1,121}$ = 11.28, P = 0.001). No group differences in skin conductance responses to the CS were observed ($F_{1,89}$ = 0.71, P = 0.401). (**C**) In the outcome devaluation procedure, we disconnected one wrist from the electrical stimulator (devalued) while leaving the other wrist connected (valued). Participants were made explicitly aware that one wrist previously associated with an electrical shock was now safe. (**D**) During the extinction procedure, the number of unnecessary foot-pedal presses to avoid shocks from the now disconnected electrical stimulator was measured. The events discussed in (C) and (D) were conducted twice: once after a short period of training and again after overtraining to promote habit formation. All participants made a greater number of successful avoidance responses to the CS associated with the valued outcome compared with the devalued outcome ($F_{1,121}$ = 20.05, P < 0.001). This difference was marginally smaller in CUD patients compared with controls ($F_{1,121}$ = 3.23, P = 0.075). Consistent with their poor performance during the training phases, individuals with CUD remained less successful than controls in avoiding shocks. Skin conductance increased in all participants in response to the CS associated with the valued outcome compared with the devalued outcome ($F_{1,88}$ = 8.23, P = 0.005), but this did not differ between the groups ($F_{1,88}$ = 0.29, P = 0.592). [Results were statistically corrected for group differences in subjective shock intensity. In (B) and (D), error bars denote SEM, SQRT signifies square-root transformation, *ns* indicates P > 0.05, and asterisks denote P < 0.05.]

occurred in the context of intact fear conditioning and was not accompanied by habit learning. This could be interpreted as a motivational impairment that is consistent with theories of the role of dopamine in motivational processes (*11*, *12*) and with reports of reduced striatal dopamine function in CUD (*15*, *16*). Our findings are also in line with evidence indicating that manipulations of dopamine neurotransmission alter instrumental learning (*17*) and shift the balance between goal-directed and habitual responding (*18*, *19*).

Although the observed appetitive habit bias was specific to cocaine addiction, the main contributory factors were impaired goal-directed learning and accumulated life stress. We also report evidence of additional executive impairments consistent with previous findings (*9*); however, these were insufficient to explain the increased goal-to-habit shift in appetitive behavior. Nonetheless, impulsivity and instrumental learning impairments are critical factors in explaining the reduced propensity to avoid aversive outcomes.

How can these findings be applied to other addictive and compulsive behaviors? Emerging evidence in alcoholism has already shown disruptions in the balance of action control for appetitive behavior (*20*, *21*). Avoidance habits might be more relevant for opiate addiction, given that the avoidance of unpleasant withdrawal symptoms is thought to play an important role in its development. Although we did not find supportive evidence in our comorbid sample, this hypothesis should be tested in opiate-addicted patients without such comorbidity. The performance profile of CUD patients in the appetitive condition may reflect a transdiagnostic risk factor for developing compulsive habits, as was recently shown to explain common deficits seen in obsessive-compulsive disorder (OCD), alcohol addiction, and eating disorders (*22*, *23*). Notably, however, our data show that this pattern may not hold in the context for avoidance behavior, where, for example, OCD patients (unlike our CUD sample) exhibit greater habitual learning (*24*).

Our findings illustrate the particular difficulty of treating CUD: The persistent deficits in avoiding aversive consequences highlight the ineffectiveness of punitive interventions for cocaine addiction. Moreover, the tendency of patients to perform a rewarded behavior in an automatic fashion, irrespective of its consequences, is unlikely to be affected by cognitive interventions that target the enhancement of alternative outcomes. Treatment of cocaine addiction should thus focus on training desirable habits that replace habitual drug-taking while protecting CUD patients from aversive consequences that they may fail to avoid.

REFERENCES AND NOTES

1. B. J. Everitt, T. W. Robbins, *Nat. Neurosci.* **8**, 1481–1489 (2005).
2. F. J. Miles, B. J. Everitt, A. Dickinson, *Behav. Neurosci.* **117**, 927–938 (2003).

3. A. Dickinson, *Philos. Trans. R. Soc. London Ser. B* **308**, 67–78 (1985).
4. B. W. Balleine, J. P. O'Doherty, *Neuropsychopharmacology* **35**, 48–69 (2010).
5. L. H. Corbit, B. C. Chieng, B. W. Balleine, *Neuropsychopharmacology* **39**, 1893–1901 (2014).
6. E. Dias-Ferreira *et al.*, *Science* **325**, 621–625 (2009).
7. G. Schoenbaum, B. Setlow, *Cereb. Cortex* **15**, 1162–1169 (2005).
8. Supplementary materials are available on *Science* Online.
9. K. D. Ersche *et al.*, *Science* **335**, 601–604 (2012).
10. R. Z. Goldstein, N. D. Volkow, *Nat. Rev. Neurosci.* **12**, 652–669 (2011).
11. R. J. Beninger, S. T. Mason, A. G. Phillips, H. C. Fibiger, *J. Pharmacol. Exp. Ther.* **213**, 623–627 (1980).
12. J. D. Salamone, M. Correa, *Behav. Brain Res.* **137**, 3–25 (2002).
13. J. W. Dalley *et al.*, *Science* **315**, 1267–1270 (2007).
14. D. Belin, A. C. Mar, J. W. Dalley, T. W. Robbins, B. J. Everitt, *Science* **320**, 1352–1355 (2008).
15. N. D. Volkow *et al.*, *Nature* **386**, 830–833 (1997).
16. D. Martinez *et al.*, *Am. J. Psychiatry* **166**, 1170–1177 (2009).
17. M. J. Frank, L. C. Seeberger, R. C. O'Reilly, *Science* **306**, 1940–1943 (2004).
18. S. de Wit *et al.*, *Psychopharmacology* **219**, 621–631 (2012).
19. K. Wunderlich, P. Smittenaar, R. J. Dolan, *Neuron* **75**, 418–424 (2012).
20. Z. Sjoerds *et al.*, *Transl. Psychiatry* **3**, e337 (2013).
21. J. M. Barker, J. R. Taylor, *Neurosci. Biobehav. Rev.* **47**, 281–294 (2014).
22. C. M. Gillan *et al.*, *Am. J. Psychiatry* **168**, 718–726 (2011).
23. C. M. Gillan, M. Kosinski, R. Whelan, E. A. Phelps, N. D. Daw, *eLife* **5**, e11305 (2016).
24. C. M. Gillan *et al.*, *Biol. Psychiatry* **75**, 631–638 (2014).

ACKNOWLEDGMENTS

We thank all volunteers for their participation in this study, as well as the staff at the Mental Health Research Network and the Cambridge BioResource for their assistance with volunteer recruitment. We are especially grateful to N. Flake and S. Whittle for their exceptional commitment in this regard. We also thank the staff at the National Institute for Health Research (NIHR) Clinical Research Facility at Addenbrooke's Hospital for their support throughout this study. We are grateful to S. Abbott, R. Lumsden, J. Arlt, C. Whitelock, I. Lee, and M. Pollard for their assistance. C.M.G. is supported by a Sir Henry Wellcome Postdoctoral Fellowship (101521/Z/12/Z). This work was funded by a grant from the Medical Research Council (MR/J012084/1) and was conducted within the NIHR Cambridge Biomedical Research Centre and the Behavioral and Clinical Neuroscience Institute, which is jointly funded by the Medical Research Council and the Wellcome Trust. The data described in this paper are stored at the University of Cambridge's institutional repository, Apollo (http://dx.doi.org/10.17863/CAM.122).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1468/suppl/DC1
Materials and Methods
Supplementary Text
Tables S1 and S2
References (*25*–*30*)

3 February 2016; accepted 19 May 2016
10.1126/science.aaf3700

STEM CELLS

Spatiotemporal coordination of stem cell commitment during epidermal homeostasis

Panteleimon Rompolas,[1]*† Kailin R. Mesa,[1]† Kyogo Kawaguchi,[2] Sangbum Park,[1] David Gonzalez,[1] Samara Brown,[1] Jonathan Boucher,[1] Allon M. Klein,[2]‡ Valentina Greco[1,3]‡

Adult tissues replace lost cells via pools of stem cells. However, the mechanisms of cell self-renewal, commitment, and functional integration into the tissue remain unsolved. Using imaging techniques in live mice, we captured the lifetime of individual cells in the ear and paw epidermis. Our data suggest that epidermal stem cells have equal potential to either divide or directly differentiate. Tracking stem cells over multiple generations reveals that cell behavior is not coordinated between generations. However, sibling cell fate and lifetimes are coupled. We did not observe regulated asymmetric cell divisions. Lastly, we demonstrated that differentiating stem cells integrate into preexisting ordered spatial units of the epidermis. This study elucidates how a tissue is maintained by both temporal and spatial coordination of stem cell behaviors.

Tissue homeostasis requires the ability to replace damaged or lost cells while maintaining tissue structure and function. A model for studying this process is the mouse adult interfollicular epidermis (IFE), where organized layers of progressively differentiated epithelial cells form a barrier from which suprabasal cells are continuously shed and replenished by an underlying proliferative basal layer (*1*–*3*). Understanding how basal stem cell proliferation and terminal differentiation remain balanced in homeostasis is a central question in both epithelial and stem cell biology.

Initial models of epidermal maintenance recognized the three-dimensional organization of discrete columns, called epidermal proliferative units (EPUs), which are defined by the perimeter of the most external, terminally differentiated cells (*4*–*6*). An important implication of the EPU model is that each unit is autonomously maintained by an asymmetrically dividing, basally located stem cell, with slow-cycling characteristics (*7*–*9*). Recent studies support the presence of slow-cycling stem cells in mouse epidermis (*10*, *11*). However, long-term lineage-tracing studies show that basal clones do not strictly adhere to the columnar borders of EPUs and support a model based on a single stem cell population that makes stochastic fate choices, while still relying on mostly (60 to 84%) asymmetric divisions to generate one stem cell and one terminally differentiated cell (*12*–*16*). These studies provide critical insights into epidermal homeostasis but remain disconnected and don't explain how individual stem cells and their progeny are integrated into the existing structure of a tissue.

A major challenge in elucidating cell fate has been the inability to resolve individual cell fate choices within clones. Individual cell behaviors have been indirectly inferred from time series of fixed clonal samples (*17*). Therefore, we developed an in vivo pulse/chase system for single-cell genetic label retention to continuously track entire lineages across multiple generations and capture the fate of individual basal cells within them (*18*) (Fig. 1A and fig. S1A). For that, we acquired serial optical sections of the epidermis from the same live adult mice at successive time points and captured the differentiation state of single labeled cells by position and cellular morphology within the entire volume of the IFE (fig. S1B) (*19*–*22*). To distinguish between region-specific characteristics and more general epidermal principles, we performed our lineage tracing in both ear and plantar epidermis. Cells that committed to differentiation were scored by their departure from the basal layer and their gradual movement toward the surface of the skin, which was irreversible in all cases (fig. S1C). Cell divisions in the basal layer generated two daughter cells that remained within the basal layer upon division (fig. S1D).

Analysis of division and differentiation events in clonal lineage trees provided direct access to lifetimes and fate choices of individual basal cells, and revealed fate correlations that could not be addressed from static clonal analysis (Fig. 1, B and C, and fig. S2A). We tested two key hypotheses: First, we asked whether the basal layer is maintained through a proliferative hierarchy by a small population of stem cells (*10*, *11*); if so, mother and daughter cell fates should be correlated, because only stem cells should give rise to daughter stem cells. We performed this multigenerational analysis in the ear epidermis and detected no mother-daughter bias in fate choice [supplementary theory (ST) S5] or in their lifetimes (Pearson correlation $R = -0.11$, $P = 0.2$).

[1]Department of Genetics, Yale School of Medicine, New Haven, CT 06510, USA. [2]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. [3]Departments of Dermatology and Cell Biology, Yale Stem Cell Center, Yale Cancer Center, Yale School of Medicine, New Haven, CT 06510, USA.
*Present address: Department of Dermatology, Institute for Regenerative Medicine, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104, USA.
†These authors contributed equally to this work. **‡Corresponding author. Email: valentina.greco@yale.edu (V.G.); allon_klein@hms.harvard.edu (A.M.K.)**

and thereby competes for free NAD^+ in the absence of substrate. We observed a rapid recovery of fluorescence upon GAPDH addition (Fig. 1F).

To determine sensor specificity, we evaluated fluorescence in the presence of related nucleotides and NAD^+ precursors (Fig. 1G). Only NAD^+ appreciably decreased sensor fluorescence. The absolute fluorescence intensities of the sensor and cpVenus displayed similar sensitivities to pH, and the NAD^+-dependent responses of the sensor were minimally affected from pH 6.5 to 8.0 (fig. S7). Thus, pH effects can be accounted for by normalizing to cpVenus. Fluorescence intensity was slightly affected by temperature, but the K_d of the unbound pool was not appreciably affected between 20° and 37°C (fig. S8).

Localization sequences directing the sensor to the nucleus, cytoplasm, and mitochondria (Fig. 2A) did not affect responses to NAD^+ in vitro (fig. S9). We generated clonal human embryonic kidney (HEK) 293T lines stably expressing the localized sensors or their corresponding cpVenus controls and verified that expression of the sensors did not affect NAD^+ levels (fig. S10). To estimate concentrations of free intracellular NAD^+, we permeabilized cells with digitonin to allow internal NAD^+ to equilibrate with externally determined concentrations and monitored fluorescence by flow cytometry. Equilibration was assessed with propidium iodide (PI), whose molecular size is similar to that of NAD^+ (fig. S11). NAD^+ decreased fluorescence of the cytoplasmic sensor in a dose-dependent manner (apparent K_d, ~300 μM; dynamic range, 30 μM to 1 mM) (Fig. 2B and fig. S12) and minimally affected cpVenus. The mean of the fluorescence ratio (488 nm/405 nm) for the cytoplasmic sensor in nonpermeabilized HEK293T cells relative to that of cpVenus was interpolated to reveal a free NAD^+ value of 106 μM [95% confidence interval (CI), 92 to 122 μM]. Using the same strategy, we determined that the concentration of free NAD^+ was 109 μM in the nucleus (95% CI, 87 to 136 μM) (fig. S13) and 230 μM in mitochondria (95% CI, 191 to 275 μM). Mitochondrial measurements were fit to the curve obtained with the cytoplasmic sensor. We confirmed the cytoplasmic calculations with live microscopy, using adherent human cervical cancer HeLa cells permeabilized with saponin and equilibrated with varying external NAD^+ concentrations (Fig. 2C). Equilibration with 100 μM NAD^+ minimally changed sensor fluorescence compared with either higher (1 mM) or lower (0 μM) NAD^+ concentration. We observed similar fluorescence changes in populations of partially permeabilized HeLa cells that were analyzed by flow cytometry (fig. S14). Many nuclear and cytoplasmic NAD^+-consuming enzymes have Michaelis constant (K_m) values around 100 μM for NAD^+, and the K_m of

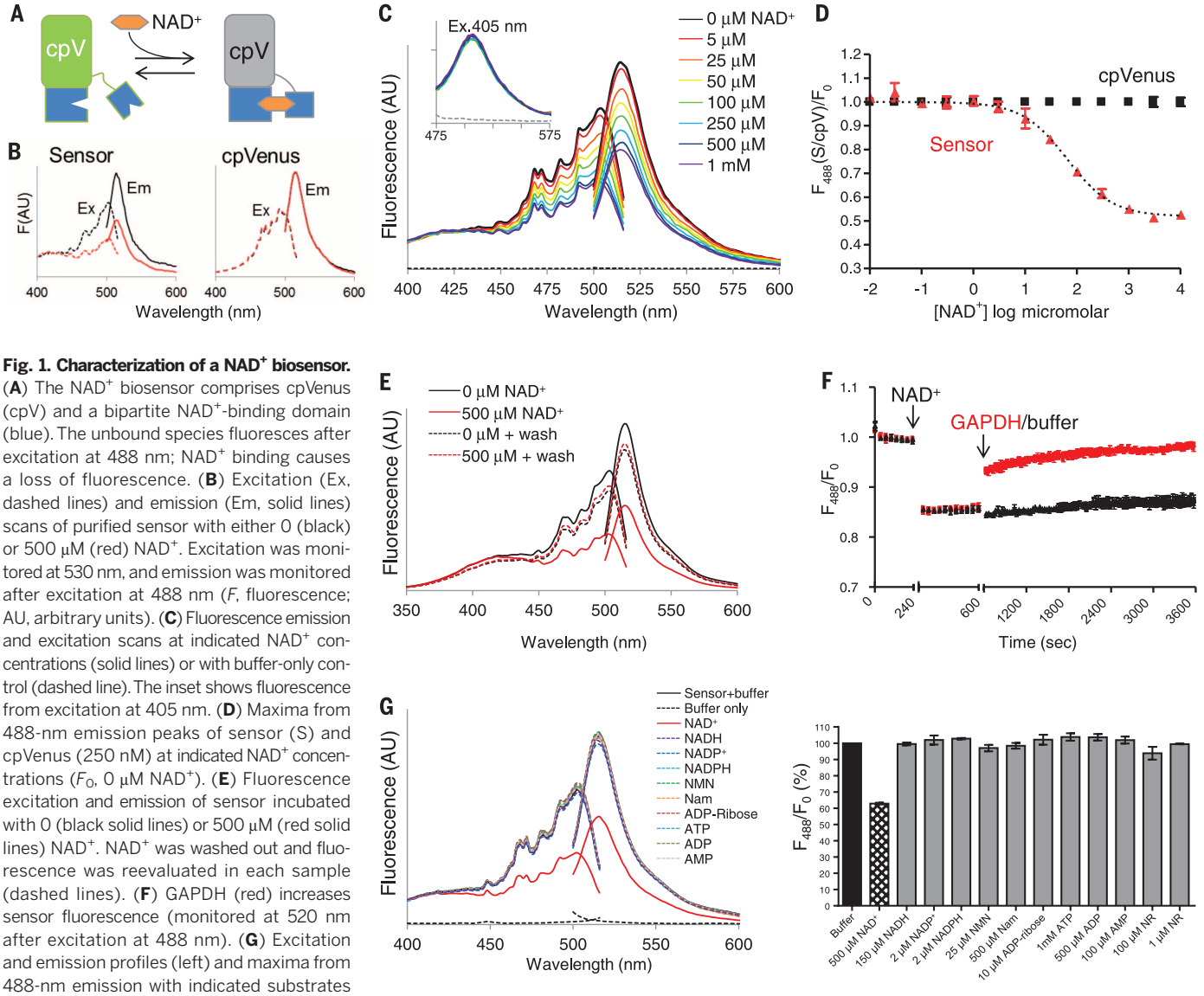


Fig. 1. Characterization of a NAD^+ biosensor. (**A**) The NAD^+ biosensor comprises cpVenus (cpV) and a bipartite NAD^+-binding domain (blue). The unbound species fluoresces after excitation at 488 nm; NAD^+ binding causes a loss of fluorescence. (**B**) Excitation (Ex, dashed lines) and emission (Em, solid lines) scans of purified sensor with either 0 (black) or 500 μM (red) NAD^+. Excitation was monitored at 530 nm, and emission was monitored after excitation at 488 nm (*F*, fluorescence; AU, arbitrary units). (**C**) Fluorescence emission and excitation scans at indicated NAD^+ concentrations (solid lines) or with buffer-only control (dashed line). The inset shows fluorescence from excitation at 405 nm. (**D**) Maxima from 488-nm emission peaks of sensor (S) and cpVenus (250 nM) at indicated NAD^+ concentrations (F_0, 0 μM NAD^+). (**E**) Fluorescence excitation and emission of sensor incubated with 0 (black solid lines) or 500 μM (red solid lines) NAD^+. NAD^+ was washed out and fluorescence was reevaluated in each sample (dashed lines). (**F**) GAPDH (red) increases sensor fluorescence (monitored at 520 nm after excitation at 488 nm). (**G**) Excitation and emission profiles (left) and maxima from 488-nm emission with indicated substrates (right). NADH, reduced NAD^+; $NADP^+$, nicotinamide adenine dinucleotide phosphate; NADPH, reduced $NADP^+$; Nam, nicotinamide; ATP, adenosine triphosphate; AMP, adenosine monophosphate. In (D), (F), and the right panel of (G), data are shown as means ± SD (n = 3).

mitochondrial SIRT3 is around 250 μM (*1*), supporting the idea that these enzymes are poised to be regulated by local NAD^+ fluctuations. The similarity in nuclear and cytoplasmic NAD^+ levels suggests that NAD^+ is readily exchangeable between the nucleus and cytoplasm.

Mammalian cells predominantly rely on the NAMPT-dependent salvage pathway for NAD^+ biosynthesis (*4*), so presumably NAD^+ concentrations in all subcellular compartments would be affected by NAMPT inhibition. To test this idea, cells were treated with FK866 {N-[4-(1-benzoyl-4-piperidinyl)butyl]-3-(3-pyridinyl)-2E-propenamide}, an NAMPT inhibitor, and NAD^+ concentrations in various compartments were monitored with live flow cytometry (Fig. 3A) or microscopy (fig. S15). In the absence of FK866, NAD^+ concentrations remained constant (fig. S16). The rate of free NAD^+ depletion was similar in the nucleus and cytoplasm (half-life, ~2 hours) and largely depended on PARP activity, as demonstrated by treatment with the inhibitor Tiq-A [thieno(2,3-*c*)isoquinolin-5-one] (fig. S16). Depletion of mitochondrial free NAD^+ occurred at a slower rate (half-life, ~8 hours) and was minimally affected by Tiq-A. FK866 treatment for 16 hours significantly decreased free NAD^+ by more than 85% in all compartments—nuclear (8 μM; 95% CI, 6 to 11 μM), cytoplasmic (3 μM; 95% CI, 0.1 to 6 μM), and mitochondria (32 μM; 95% CI, 27 to 38 μM); this decrease was not prevented by Tiq-A treatment (fig. S16).

We confirmed our pharmacological studies with NAMPT by using small interfering RNAs (siRNAs), which depleted free NAD^+ in all compartments (fig. S17). Nicotinamide riboside (NR)

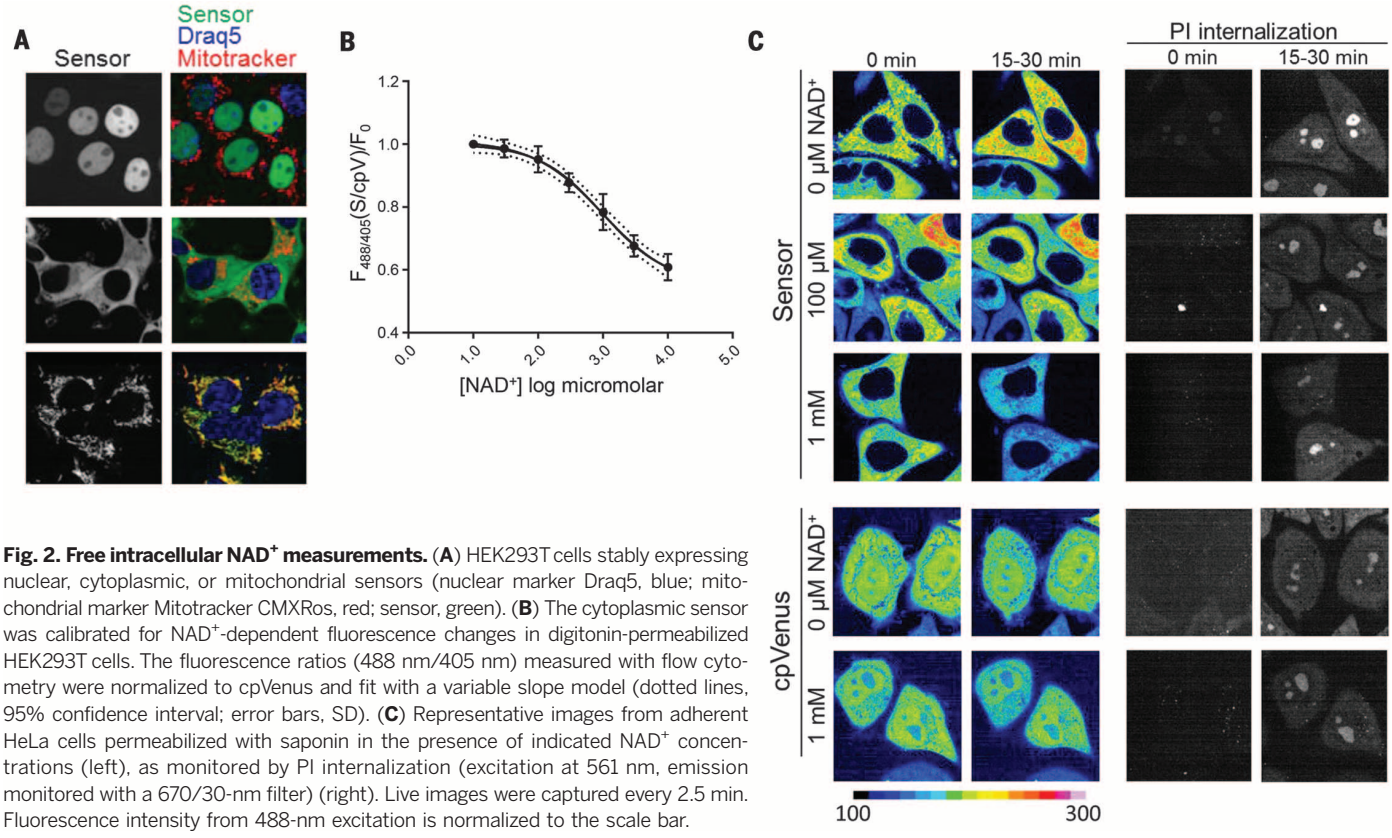


Fig. 2. Free intracellular NAD^+ measurements. (**A**) HEK293T cells stably expressing nuclear, cytoplasmic, or mitochondrial sensors (nuclear marker Draq5, blue; mitochondrial marker Mitotracker CMXRos, red; sensor, green). (**B**) The cytoplasmic sensor was calibrated for NAD^+-dependent fluorescence changes in digitonin-permeabilized HEK293T cells. The fluorescence ratios (488 nm/405 nm) measured with flow cytometry were normalized to cpVenus and fit with a variable slope model (dotted lines, 95% confidence interval; error bars, SD). (**C**) Representative images from adherent HeLa cells permeabilized with saponin in the presence of indicated NAD^+ concentrations (left), as monitored by PI internalization (excitation at 561 nm, emission monitored with a 670/30-nm filter) (right). Live images were captured every 2.5 min. Fluorescence intensity from 488-nm excitation is normalized to the scale bar.

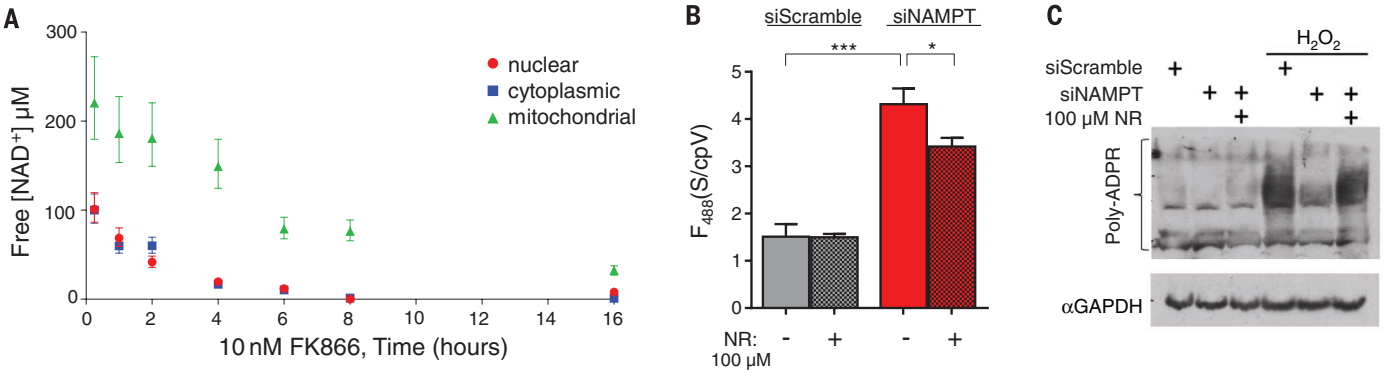


Fig. 3. NAD^+ fluctuations in cells. (**A**) Free NAD^+ concentrations in HEK293T cells after treatment with FK866 (10 nM). Sensor/cpVenus (488 nm/405 nm) fluorescence ratios were measured by flow cytometry, and the fold change compared with untreated controls was interpolated onto an in vitro standard curve. (**B**) Cytoplasmic free NAD^+ was decreased by NAMPT depletion (siNAMPT) and partially restored by NR (100 μM) treatment. Imaging measurements are from excitation at 488 nm. Shown are means ± SEM (n = 3) from residual maximum likelihood (REML) analysis. ***P < 0.001; *P < 0.05. (**C**) Effect of free NAD^+ on PARP activation (poly-ADPR). siScramble, scrambled RNA control; αGAPDH, anti-GAPDH antibody.

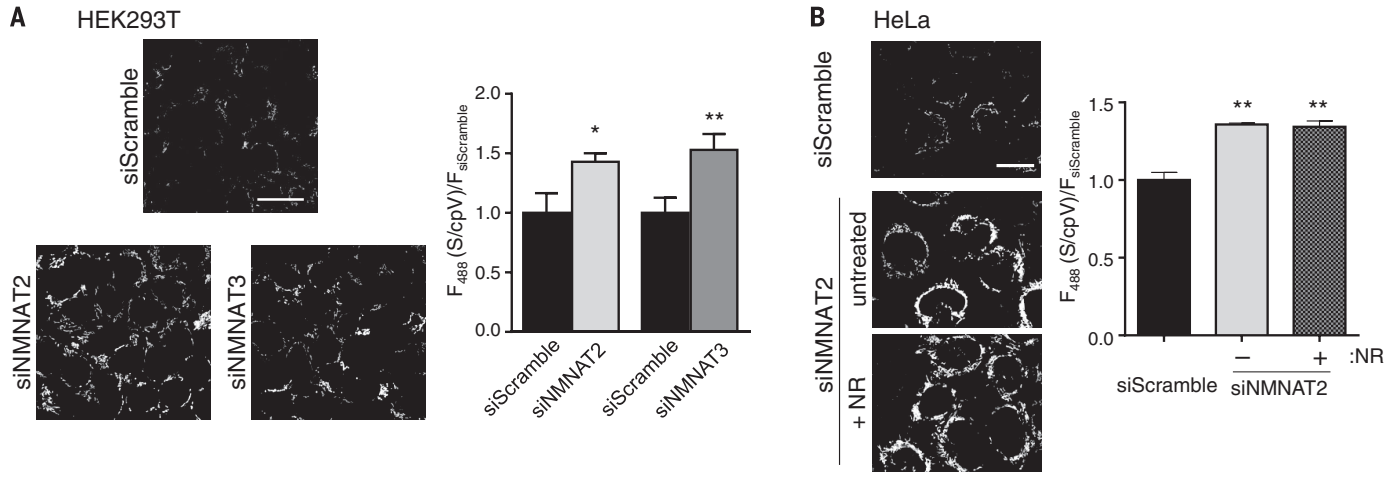


Fig. 4. Multiple sources for mitochondrial NAD$^+$. (**A**) Representative images are shown on the left of the mitochondrial NAD$^+$ sensor after depletion of either NMNAT2 (siNMNAT2) or NMNAT3 (siNMNAT3) in HEK293T cells. Scale bar, 25 μm. Excitation at 488 nm. Fluorescence intensities were normalized to scramble RNA (siScramble) and mitochondrial cpVenus (cpV) controls. Plotted are means ± SEM (n = 3) from REML analysis. *P = 0.03; **P < 0.005. (**B**) Effect of NR treatment (250 μM, 24 hours) on NAD$^+$ concentrations in the mitochondria of HeLa cells depleted of NMNAT2 (siNMNAT2). Representative images are shown on the left (scale bar, 25 μm). Fluorescence intensities were normalized to scramble RNA (siScramble) and mitochondrial cpVenus (cpV) controls. Plotted are means ± SEM (n = 3) from REML analysis. **P < 0.01.

increased free NAD$^+$ to 83 μM (95% CI, 72 to 98 μM) through a parallel pathway that bypasses NAMPT (*11*) (Fig. 3B and fig. S18). NR itself was not recognized by the sensor (Fig. 1G) and did not alter sensor fluorescence (fig. S19). The levels of free NAD$^+$ correlated with PARP activity, induced by H_2O_2 genotoxic stress (Fig. 3C).

A heretofore unanswered question is how pools of NAD$^+$ in the nucleus, cytoplasm, and mitochondria are established and maintained. To address the specific roles of NMNAT isoforms in the creation and maintenance of compartmentalized NAD$^+$ pools, we used validated siRNAs (fig. S20) to systematically deplete the NMNAT enzymes that catalyze the final step of NAD$^+$ biosynthesis in each of these subcellular compartments (fig. S21) (*5*). Depletion of NMNAT2 decreased cytoplasmic NAD$^+$ to 58 μM (95% CI, 50 to 67 μM), consistent with its subcellular expression pattern. Nuclear NAD$^+$ concentrations, however, were not significantly affected (fig. S22), indicating that NMNAT1 is sufficient to meet the nuclear NAD$^+$ demand but cannot fully compensate for depletion of NAD$^+$ in the cytoplasm. This finding is consistent with the lethality of individual knockout animal models (*12*, *13*). Depletion of NMNAT1 did not significantly change nucleocytoplasmic NAD$^+$ concentrations, but levels were lower (fig. S23A). We wondered whether the relatively similar amounts of NMNAT1 and NMNAT2 in HEK293T cells masked the NMNAT1 contribution to the cytoplasmic compartment, so we examined HeLa cells, which express relatively less NMNAT2 (fig. S24). In HeLa cells, depleting NMNAT1 significantly decreased cytoplasmic NAD$^+$ to 38 μM (95% CI, 32 to 44 μM) (fig. S23B). Together, these data demonstrate that NMNAT1 can contribute to the cytoplasmic NAD$^+$ pool and highlight cell type–dependent differences in NAD$^+$ regulation.

Our data indicate that free NAD$^+$ in mitochondria fluctuates distinctly from that in the nucleocytoplasm (Fig. 3A and fig. S25). The mitochondrial NMNAT isoform, NMNAT3, is thought to generate mitochondrial NAD$^+$; consistent with this idea, depleting NMNAT3 in HEK293T cells significantly decreased mitochondrial NAD$^+$ concentrations to 103 μM (95% CI, 88 to 123 μM) (Fig. 4A and fig. S26). Depleting NMNAT2 also decreased mitochondrial NAD$^+$ concentrations (Fig. 4A and fig. S26). This suggests that NAD$^+$ produced in the cytoplasm can influence mitochondrial stores and that NMN is not the sole source of mitochondrial NAD$^+$.

Supporting this idea, NMNAT3 depletion in HeLa cells, which contain low amounts of this enzyme, did not affect mitochondrial NAD$^+$ concentrations (fig. S27). In contrast, NMNAT2 depletion both decreased cytoplasmic NAD$^+$ to 38 μM (95% CI, 33 to 45 μM) and decreased mitochondrial NAD$^+$ to 134 μM (95% CI, 113 to 161 μM) (Fig. 4B and fig. S28). Addition of NR did not restore mitochondrial NAD$^+$ concentrations, implying that cytoplasmic NAD$^+$, and not NMN, maintains mitochondrial NAD$^+$ in HeLa cells (Fig. 4B). Thus, there appear to be at least two mechanisms for maintaining mitochondrial NAD$^+$ in various cell types: conversion of NMN by NMNAT3 and transport of cytoplasmic NAD$^+$. A NAD$^+$ transporter has been identified in bacteria (*14*), yeast (*15*), and plants (*16*), but a mammalian homolog has not yet been identified.

REFERENCES AND NOTES

1. C. Cantó, K. J. Menzies, J. Auwerx, *Cell Metab.* **22**, 31–53 (2015).
2. A. A. Sauve *et al.*, *Biochemistry* **40**, 15456–15463 (2001).
3. P. O. Hassa, S. S. Haenni, M. Elser, M. O. Hottiger, *Microbiol. Mol. Biol. Rev.* **70**, 789–829 (2006).
4. J. R. Revollo, A. A. Grimm, S. Imai, *J. Biol. Chem.* **279**, 50754–50763 (2004).
5. F. Berger, C. Lau, M. Dahlmann, M. Ziegler, *J. Biol. Chem.* **280**, 36334–36341 (2005).
6. F. Koch-Nolte, S. Fischer, F. Haag, M. Ziegler, *FEBS Lett.* **585**, 1651–1656 (2011).
7. S. Imai, L. Guarente, *Trends Cell Biol.* **24**, 464–471 (2014).
8. R. H. Houtkooper, C. Cantó, R. J. Wanders, J. Auwerx, *Endocr. Rev.* **31**, 194–223 (2010).
9. M. Tantama, Y. P. Hung, G. Yellen, *Prog. Brain Res.* **196**, 235–263 (2012).
10. K. S. Gajiwala, C. Pinko, *Structure* **12**, 1449–1459 (2004).
11. P. Bieganowski, C. Brenner, *Cell* **117**, 495–502 (2004).
12. L. Conforti *et al.*, *FEBS J.* **278**, 2666–2679 (2011).
13. A. N. Hicks, L. Campeau, D. Burmeister, C. E. Bishop, K. E. Andersson, *Neurourol. Urodyn.* **32**, 1130–1136 (2013).
14. I. Haferkamp *et al.*, *Nature* **432**, 622–625 (2004).
15. S. Todisco, G. Agrimi, A. Castegna, F. Palmieri, *J. Biol. Chem.* **281**, 1524–1531 (2006).
16. F. Palmieri *et al.*, *J. Biol. Chem.* **284**, 31249–31259 (2009).

ACKNOWLEDGMENTS

NAD$^+$ sensors are available from Oregon Health & Science University (OHSU) under a material transfer agreement with the authors. We thank D. Piston for guidance and the cpVenus construct, C. Brenner and M. Michaud for NR, E. Cambronne for the *Enterococcus fecalis* isolate, J. Wiedrick and S. Markwardt for statistical help, G. Mandel for critical reading of the manuscript, and P. Brehm, J. and C. Galbraith, H. Zhong, and R. Hanson for helpful discussions. Imaging, flow cytometry, and high-performance liquid chromatography were supported by OHSU core facilities. This work was supported by the Pew Charitable Trust and the NIH (grants NS088629, UL1TR000128, P30NS061800, and P30CA069533). X.A.C., M.L.S., M.S.C., and R.H.G. are inventors on patent PCT/US15/62003 for the NAD$^+$ sensor. Materials and methods are described in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/352/6292/1474/suppl/DC1
Materials and Methods
Figs. S1 to S28
References (*17*–*20*)

23 September 2015; accepted 3 May 2016
10.1126/science.aad5168

CALL FOR ENTRIES

Apply for the 2017 Award now!

http://www.sciencemag.org/prizes/boyalife
Questions/Inquiries:BoyalifeSciencePrize@aaas.org

Stem cells and regenerative medicine is the new frontier in life sciences. Boyalife,*Science*, and *Science Translational Medicine* jointly establish a global annual prize to recognize and reward scientists in the fields of stem cells and regenerative medicine with a focus on developing cell-based treatments for cancer, degenerative disorders, immunotherapy and stem cells transplantation.

2016 WINNER

Li Qian, Ph.D.
(UNC-Chapel Hill)

Exploring the reprogramming approaches for cardiac regeneration and disease modeling

★ The winners will receive: A Grand Prize of $25,000 and a Runner-Up Prize of $5,000 will be awarded.

★The Grand Prize Winning Essay and an abstract of the Runner-Up Essay will be published in *Science*.

The 2016 winners are Dr. Li Qian and Dr.Yossi Buganim, You can read their winning essays in the June 17 issue of *Science*.

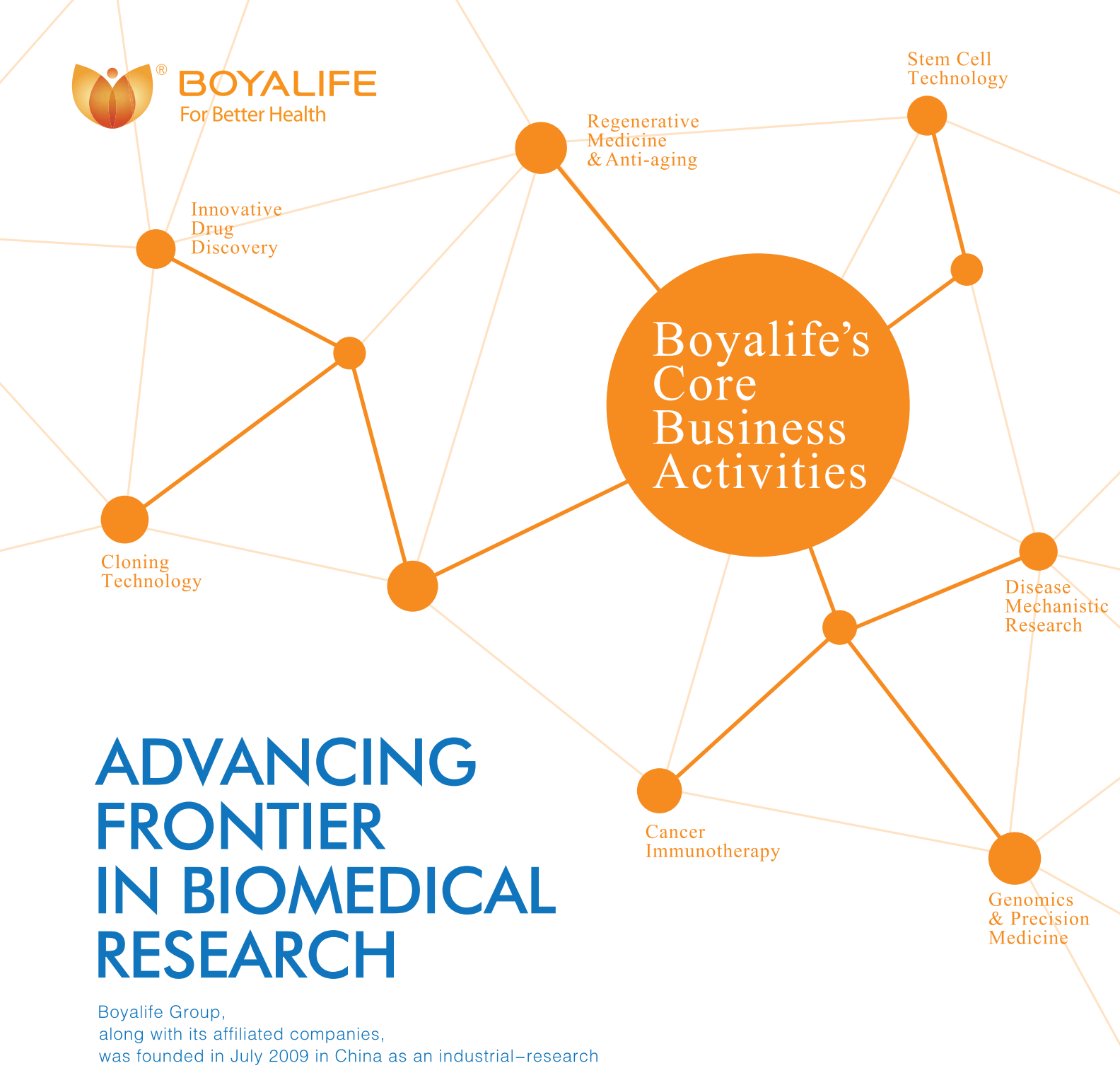
BOYALIFE Science & Science Translational Medicine Award | in Stem Cells and Regenerative Medicine

BOYALIFE For Better Health

Science AAAS

Science Translational Medicine AAAS

BOYALIFE
For Better Health



ADVANCING FRONTIER IN BIOMEDICAL RESEARCH

Boyalife Group,
along with its affiliated companies,
was founded in July 2009 in China as an industrial–research alliance among top research institutes for stem cell and regenerative medicine.
In 2012, Boyalife Group won the prestigious "2012 CCTV China Brand Award",
emerging as a leading brand in healthcare in China.

Start your career at Boyalife!
Contact:career@boyalife.com
www.boyalifegroup.com

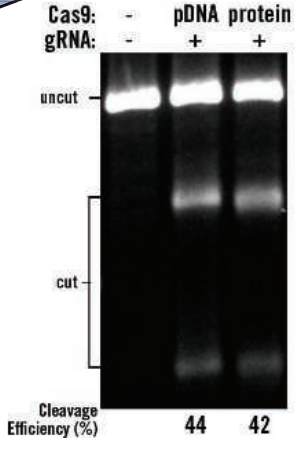
Mirus | The Transfection Experts

X2 for CRISPR

Optimized DNA, RNA and RNP Delivery Using CRISPR/Cas9

Download the white paper
www.mirusbio.com/CRISPR

*Trans*IT-X2® Dynamic Delivery System



T7E1 cleavage efficiency in 293T/17 cells transfected with Cas9 pDNA/gRNA or Cas9 protein/gRNA (RNP) using *Trans*IT-X2®.

Now you can use the same transfection reagent for your knockout, knockdown or expression experiments. The *Trans*IT-X2® Dynamic Delivery System takes CRISPR genome editing a step further with an advanced polymeric technology that efficiently delivers plasmid DNA, small RNAs such as siRNA and CRISPR guide RNA or RNP complexes.

The *Trans*IT-X2® Dynamic Delivery System delivers CRISPR/Cas components in multiple formats:

- **DNA**–deliver plasmid DNA expressing Cas9 or guide RNA
- **RNA**–deliver sgRNA or crRNA:tracrRNA
- **Protein**–deliver Cas9:gRNA RNP complexes

ADVANCE YOUR TRANSFECTIONS.
Request a FREE SAMPLE of
*Trans*IT-X2® Dynamic Delivery System.
Visit www.mirusbio.com, call 888.530.0801 (U.S. only) or +1.608.441.2852

PASSION VISION INSPIRATION
Celebrating 20 YEARS

mirusbio.com
Providing gene delivery expertise since 1995

AAAS 2017 ANNUAL MEETING

FEBRUARY 16-20 • BOSTON

SERVING SOCIETY THROUGH

SCIENCE POLICY

Join Us in Boston

The AAAS Annual Meeting is interdisciplinary and inclusive. Thousands of leading scientists, engineers, educators, policymakers, and journalists from around the world will gather to discuss recent developments in science and technology.

Registration opens August 2016.

aaas.org/meetings

Genotyping Software

GeneMarker software version 2.6.7 incorporates several enhanced and automated quality alerts for trisomy, multiplex ligation-dependent probe amplification (MLPA), and microsatellite instability analysis. The automated alerts in the form of pop-up windows, which include those recommended by MLPA kit manufacturer MRC-Holland, inform users when sequencing failures such as saturation or quality control fragment failure have been detected by the software. In the event the user overrides the alert, the particular failure will still be indicated on the final analysis report. Additionally, the latest version includes the ability to automatically export completed analysis projects to a common folder on the user's intranet, as well as new .pdf file electronic-reporting capabilities to provide paperless saving of analysis results. GeneMarker software is compatible with Windows OS 7–10, commercial and custom chemistries, and data files from all major capillary electrophoresis sequencing platforms, including Thermo Fisher Scientific, Beckman Coulter, and Gel Company's MegaBACE.

SoftGenetics
For info: 888-791-1270
www.softgenetics.com

DNA Synthesis Platform

A novel multiplex DNA assembly approach scales the manufacture of synthetic DNA by nearly two orders of magnitude. Gen9 has applied this technology to long-length DNA construction, building 50 gene-length DNA constructs simultaneously in a single reaction. Comprising advancements in software, chemistry, and biology, multiplexing technology will be applied to Gen9's BioFab next-generation DNA synthesis platform over the next 12 months. The BioFab platform is the first and only industrialized, chip-based manufacturing technology for DNA synthesis and assembly. This first-of-its-kind advancement will also narrow the DNA read-write gap, allowing scientists to leverage DNA sequence information to discover the next generation of breakthroughs in therapeutics, biofuels, chemicals, and more. The BioFab platform makes manual, time-consuming PCR cloning methods for working with DNA obsolete. BioFab enables scientists to get to their answers faster by designing, building, and testing many sequences at once, rather than serially.

Gen9
For info: 617-250-8433
www.gen9bio.com

PCR System

The CE-IVD marked QX200 Droplet Digital PCR (ddPCR) System can be used for highly accurate diagnostic detection and quantification of nucleic acids, aiding clinical decision-making in the treatment of diseases ranging from cancer to transplant rejection and viral infection. The system includes a QX200 Droplet Generator, QX200 Droplet Reader, and QuantaSoft software. The droplet generator and reader are CE-IVD (European Conformity-in vitro diagnostic) marked for use in the European Union for in vitro diagnostics, and are the only digital PCR instruments available for use in a clinical laboratory. The QX200 Droplet Generator is used to partition a ddPCR reaction mixture into thousands of nanoliter-sized droplets. After endpoint PCR on a thermal cycler, droplets are read individually as they are streamed single file past a two-color optical detection system on the QX200 Droplet Reader. The PCR-positive and PCR-negative droplets are counted to provide absolute quantification of target DNA.

Bio-Rad
For info: 800-424-6723
www.bio-rad.com

Microinjection-Ready Cas9

Microinjection-ready Cas9 Accelerate Ribonucleoprotein (RNP) adds to Horizon's existing collection of reagents for gene editing of animal models. Cas9 Accelerate RNP maximizes gene editing efficiency by enabling CRISPR reagents to sidestep the transcription and translation steps that traditional CRISPR-Cas9 plasmid and messenger RNA formats must complete before being activated, thereby maximizing gene editing efficiency and lowering the risk of off-target effects. Horizon's collection of microinjection-ready CRISPR reagents represent expertly designed, rigorously validated, fast-acting, and efficient gene editing tools. Horizon Discovery uses the same optimized single guide RNA and Cas9 Accelerate RNP premixed formulation for its in-house, custom animal-model generation service.

Horizon Discovery
For info: +44-(0)-1223-655580
www.horizondiscovery.com

CGH Labeling Kit

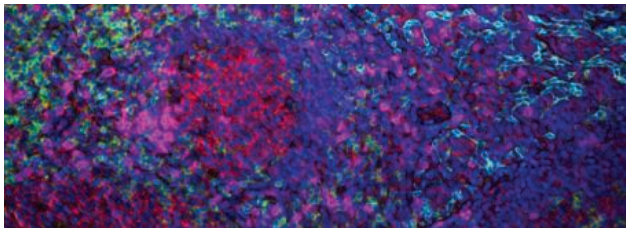
The CYTAG TotalCGH Labeling Kit contains a complete set of reagents to prepare and label DNA for hybridization to a comparative genomic hybridization and single nucleotide polymorphism (CGH+SNP) array. The kit is compatible with both CGH arrays and CGH+SNP arrays and includes the CYTAG CGH Labeling Kit, restriction enzymes AluI and RsaI, and PCR and gel cleanup columns. The CGH+SNP array enables detection of genomic aberrations associated with copy number variations, but it can also detect copy-neutral aberrations, such as loss of heterozygosity and uniparental disomy. The array is a robust method to detect chromosomal abnormalities and enables rapid screening of the entire genome, as well as detection of variation at a single nucleotide. The CYTAG TotalCGH Labeling Kit delivers low-derivative log ratio scores, assuring more confident aberrant calls and fewer failed runs than other methods can deliver. The kit works with most common dye-based arrays and has been validated on multiple scanning platforms.

Enzo Life Sciences
For info: 800-942-0430
www.enzolifesciences.com

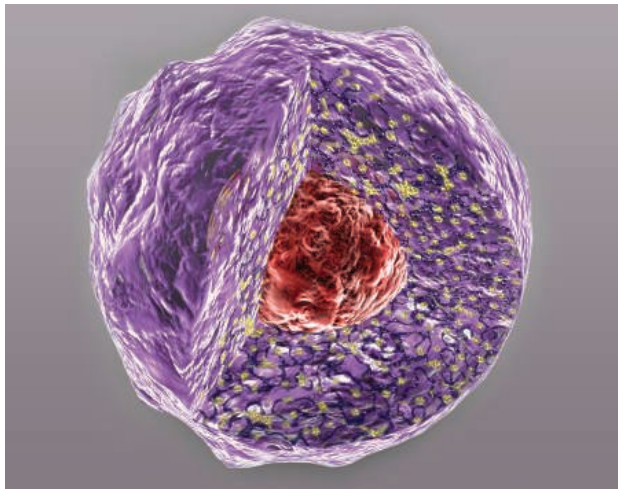
Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

want new technologies?



antibodies
apoptosis
biomarkers
cancer
cytometry
data
diseases
DNA
epigenetics
genomics
immunotherapies
medicine
microbiomics
microfluidics
microscopy
neuroscience
proteomics
sequencing
toxicology
transcriptomics



watch our webinars

Learn about the latest break-throughs, new technologies, and ground-breaking research in a variety of fields. Our expert speakers explain their quality research to you and answer questions submitted by live viewers.

VIEW NOW!
webinar.sciencemag.org

Science
AAAS

Brought to you by the *Science*/AAAS Custom Publishing Office

@SciMagWebinars

NEW ENGLAND BioLabs® Inc.

PASSION IN SCIENCE AWARDS™

What is your PASSION?

New England Biolabs® is pleased to announce our 2016 Passion In Science Awards™, recognizing those within the scientific community working to make a difference in the world through their scientific mentorship, humanitarianism, environmental stewardship, or love of the arts.

Through July 31st, 2016, we invite you to nominate yourself or a colleague for one of these awards. Award winners will be our guests at a celebration at the NEB® campus in Ipswich MA, USA and will also receive a donation to help further their cause. Remember, no act is too big or too small.

Help us celebrate those individuals who demonstrate true passion in science.

Tell us your story.

Visit NEBPassionInScience.com

Science and art come together in the artwork displayed on the front of the NEB laboratory facility in Ipswich, MA, USA. Artwork created by Nicole Gordon.

C | A | U Kiel University
Christian-Albrechts-Universität zu Kiel

Kiel University wants to attract more qualified women to professorships.

The Kiel University Medical Faculty invites applications for the following position to be filled as soon as possible:

W 3 Professorship in Human Genetics

The future occupant of this position is expected to represent the subject in research, teaching and in patient care, as well as to manage the Department of Human Genetics as its Director.

Kiel's human genetics has gained national and international recognition in research into lymphoma and leukaemia in particular. The Faculty would expressly like the Department to continue with a comparable profile (successor to Prof. R. Siebert). His successor should therefore have an outstanding level of scientific expertise, e.g. in haemato-oncological tumour genome research or in a different field with an excellent subject-specific outlook. Human genetics in Kiel has access to a university environment which is very strong from both a technical and research perspective. Interdisciplinary platforms are available in the field of high-throughput techniques for molecular genetic research and patient care in academia, to which the occupant of the position will have appropriate access.

The scientific focus of the professorship should be geared towards one of the Faculty of Medicine's profile areas. The occupant of the position is expected to become actively involved in the faculty's research activities, which are under the theme of "Age-related mechanisms in the manifestation of diseases". The close collaboration with the three research profiles of oncology, neuroscience and inflammation medicine should be manifested in local research groups. Further development of joint projects and an active role across all locations is expected in the institution.
Teaching duties include organising and holding human genetic lectures and examinations for the medicine and dentistry degree programmes as well as on the "Medical Life Sciences" Master's degree programme.

The occupant of the position is expected to cover the full range of clinical duties in diagnostics and patient care at the UKSH on the Kiel Campus. This means that the most eligible applicants will be doctors, as a licence for advanced training in human genetics is also expected. The clinical focus is derived from the specific applicant profile; close collaboration with other established diagnostic and clinical institutes and clinics is anticipated on the basis of the cooperative model. Outpatient care in the field of human genetics is currently being provided by a Medical Care Centre (MVZ). Remuneration for services provided in patient care will be regulated in a separate contract of employment with the UKSH. Applicants without any recognition as a consultant for human genetics are expected to submit a concept as to how the clinical responsibilities in diagnostics and patient care can be satisfied in full.

The requirements for appointment provided for under Article 61 apply, and a limited contract according to Article 63 (1) of the Higher Education Act (Hochschulgesetz) of the State of Schleswig-Holstein is possible. Further information is available at www.berufungen.uni-kiel.de/de (German only).

Kiel University seeks to increase the proportion of female scientists in teaching and research and urges qualified female candidates to apply. Women with equal qualifications, competence and professional performance will be given priority.

The university is committed to employing individuals with disabilities. For this reason, disabled applicants with equal qualifications, competence and professional performance will be given priority.

We explicitly encourage candidates with a migration background to apply.
Candidates are not required to submit a photograph with their application and are requested not to do so.

The Higher Education Act requires the Kiel University Medical Faculty and University of Lübeck Department of Medicine to work closely together as well as in close cooperation with the University Medical Center Schleswig-Holstein to determine and coordinate focus areas. The state further expects the clinics and institutes to cooperate with each other accordingly.

Applications should include the usual documents (curriculum vitae, publication list, list of lectures, copies of academic certificates), private and business addresses, telephone number and email address and should be submitted by **July 15th, 2016** to: Dekan der Medizinischen Fakultät der Christian-Albrechts-Universität zu Kiel, Olshausenstr. 40, 24098 Kiel

For more information on structuring the application, see www.medizin.uni-kiel.de > Faculty > Appointment process

暨南大學
JINAN UNIVERSITY

Job Vacancy

Jinan University is one of China's "One Hundred Key Universities of 21st Century"(the "211 Project") and is operated under the leadership of the Overseas Chinese Affairs Office of State Council. As the first university established by the State for overseas Chinese students, JNU currently has the largest number of overseas and foreign students and is honored as the "top university for overseas Chinese". Abiding by the motto of "loyalty, sincerity, integrity and respect", the university is committed to cultivating talents with the excellent traditional Chinese moralities and culture. In June 2015, JNU was selected into the "High-level University Construction Program" by Guangdong provincial government. The University has 27 colleges, comprising 62 departments, 188 research institutions, 77 laboratories and offering 89 undergraduate majors, 189 master degree programs in 38 Level-I and 5 Level-II disciplines as well as 74 doctoral degree programs in 15 Level-I and 4 Level-II disciplines. Besides, we have 4 national key disciplines (industrial economics, aquatic biology, finance and literary theory), 8 key disciplines at the level of the Overseas Chinese Affairs Office of the State Council, 20 provincial Level-I key disciplines, and 4 provincial Level-II key disciplines. The University has the following teaching and research bases: a key research base of national humanity & social sciences, a teaching & research base for Chinese language & literature of the Education Ministry, a base for national university cultural quality education, a national base for teaching Chinese as a foreign language, an educational base for overseas Chinese education of the Overseas Chinese Affairs Office of the State Council and a key research base of humanity & social science of Guangdong Province. It also has one national engineering center, 14 ministerial and provincial engineering centers and 13 key ministerial and provincial laboratories.

Institute of Rail Transit Jinan University (the former is Institute of Electrical Automation Jinan University) is founded in December 2013. The head of the institute is the Academician of Chinese Academy of Engineering Qian Qingquan who is the academic leading in the field of Chinese rail transit electrification and automation. 80% researchers of the institute have the PhD degree. The research directions of the institute are the rail transit safety monitoring and detection and the power regeneration and utilization in the rail transit. The research funds has been tens of millions Yuan. Especially, in July 2015 Collaborative Innovation Center of Rail Transit Healthy Operation Zhuhai was approved and the financial support for the center construction is 500 million Yuan.

We are seeking experienced researchers with a passion for exploring the scientific research of rail and transit. Join us and help shape the future of rail transit.

Requirements: The researcher candidates should have a PhD degree in rail transit, or a relevant engineering discipline.

Contact: Mr. Li
Email: liweihua@jnu.edu.cn
Tel: 00-86-13411373512

山东理工大学

Shandong University of Technology Is Seeking for High-level Talents at Home and Abroad

I. Recruiting subjects

Mechanical engineering, agricultural engineering, chemical engineering and technology, electrical engineering, computer science and technology, civil engineering, surveying and mapping science and technology, transportation engineering, biology, economics, ect.

II. Recruitment Scope

Professor, Associate Professor, Assistant Professor, Post doctor, Doctor

III. Welfare and Benefits

(1) Annual salary for assistant professor or higher-level talents is RMB 400,000-1,200,000 with setting-in allowance of RMB 500,000-1,000,000;
(2) Annual salary for post doctor and doctor is RMB 300,000-400,000 and other relevant benefits.

IV. Talents who are eligible for the State's Thousand Talents Plan and Shandong Taishan Scholars as well as other talent projects enjoy corresponding policies and treatments after successfully recommended.

V. Contact information

Address: Recruitment Office of Shandong University of Technology, No. 266, Xincun West Road, Zhangdian District, Zibo City, Shandong Province
Zip code: 255049
Contact: Professor Zhang
E-mail: rshch@sdut.edu.cn

第三軍醫大學

Recruiting High-level Medical Talents in the Third Military Medical University, Chongqing, China

The Third Military Medical University (TMMU), with a history of over 70 years, is located in the mountain city of Chongqing, one of the four municipalities directly under the jurisdiction of the Central Government of China. This university is a national key university, one of the military "2110 Project" key universities , and one of the first universities allowed to grant doctor's degree and to offer eight-year medical education program in China. Now, the TMMU is seeking **high-level medical** talents from home and abroad.

1. Positions and Remuneration Packages

1) Members of the national "One Thousand Talents Long-term Program": A settling-down allowance of 3 million yuan and 8-million-yuan research fund with an annual salary of 1 million yuan.
2) Members of the national "One Thousand Yong Talents Program": A settling-down allowance of 1 million and 500 thousand yuan and 5-million-yuan research fund with an annual salary of 400 thousand yuan.

2. Qualifications for the Position

1) Having the nationality of the People's Republic of China;
2) Be patriotic and willing to be devoted to national defense and health;
3) Having a strong academic and research ability, and be capable of undertaking major scientific research programs and engaging in original academic researches.

3.Salary and Benefits and Further More Information please check this link
www.edu.cn/tmmu

Contact Person: Ms. Fan
Tel: +86-23-68752105 13883596876
Email: sydrcb@sina.com
Address: Cadre Office of Politics Department, Third Military Medical University, Gaotanyan No.30 Center Street, Shapingba District, Chongqing 400038, P.R.China.

10 ways that *Science* Careers can help advance your career

1. Register for a free online account on **ScienceCareers.org**.
2. Search thousands of job postings and find your perfect job.
3. Sign up to receive e-mail alerts about job postings that match your criteria.
4. Upload your resume into our database and connect with employers.
5. Watch one of our many webinars on different career topics such as job searching, networking, and more.
6. Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
7. Complete an interactive, personalized career plan at "my IDP."
8. Visit our Career Forum and get advice from career experts and your peers.
9. Research graduate program information and find a program right for you.
10. Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

HARVARD MEDICAL SCHOOL | DEPARTMENT OF Biomedical Informatics

Master of Biomedical Informatics

Program Description:

The program provides the intellectual framework for clinicians and biomedical scientists in the systematic and sound use of quantitative methods to increase agility with such methods in their respective domains. The program includes an intensive, hands-on quantitative boot camp, a range of foundational courses, and courses in emerging areas such as precision medicine, data science, and data visualization. All students are expected to complete a research project and to participate in a longitudinal seminar series.

Who is this Program for?

- Postdoctoral students who recognize the relevance of informatics to their research
- MD's who are interested in qualifying for the subspecialty in clinical informatics
- Medical students who would like to take a research year during their training to explore the importance of informatics in the practice of medicine

Contact Information:

To learn more about the program, please visit our website and email us with any questions through our 'Contact Us' page: **https://informaticstraining.hms.harvard.edu/**

****Applications for Fall 2016 are due no later than July 8, 2016****

EGL Charitable Foundation

The EGL Charitable Foundation invites you to apply to the

Gruss Lipper Post-Doctoral Fellowship Program

Eligibility

- Israeli citizenship
- Candidates must have completed PhD and/or MD/PhD degrees in the Biomedical Sciences at an accredited Israeli University/Medical School or be in their final year of study
- Candidates must have been awarded a postdoctoral position in the U.S. host research institution.

Details regarding the fellowship are available at www.eglcf.org

NW210508R

UPPSALA UNIVERSITET

Seven Postdoc Positions Open at Uppsala University, Sweden to Study Evolution of New Genes and Proteins

We are seeking 7 highly motivated postdocs to work in a multi-disciplinary project to study in eubacteria and eukaryotes how new genes and proteins emerge and acquire novel capabilities. This 5-year project is funded by the Knut and Alice Wallenberg Foundation and includes six principal investigators that are specialists in high-resolution structures of molecules, biochemistry, molecular evolution and genetics. By combining controlled evolution experiments and genomic analysis with detailed studies of protein structure and function, we aim to describe in detail the birth and evolution of new protein functions.

- **Please send your application to the respective principal investigator.**
- **For further information and application details please visit http://jobs.sciencecareers.org**
- **The positions are for two years with the possibility of extension.**
- **Please note that you can ONLY apply for one of the advertised positions.**

Short description of the 7 postdoc positions.

Prof. Dan Andersson, Dept. of Medical Biochemistry and Microbiology
Postdoc 1: Experimental evolution of new genes from non-functional DNA (de novo origination).
Postdoc 2: Experimental evolution of new genes from pre-existing genes via duplication-divergence mechanisms.

Prof. Leif Andersson, Dept. of Medical Biochemistry and Microbiology
Postdoc 1: Bioinformatics of the genetic basis of ecological adaptation in Atlantic herring
Postdoc 2: Functional genomics of the genetic basis of ecological adaptation in Atlantic herring

Prof. Siv Andersson, Dept. of Cell and Molecular Biology
Postdoc 1: Bioinformatic studies of how new genes evolve from non-functional DNA (de novo origination)
Postdoc 2: Experimental studies of how new genes evolve from non-functional DNA (de novo origination).

Prof. Johan Åqvist, Dept. of Cell and Molecular Biology
Postdoc 1: Computational Biochemistry to study enzyme evolution

KEYSTONE SYMPOSIA on Molecular and Cellular Biology

Connecting the Scientific Community

The Keystone Symposia on Molecular and Cellular Biology is conducting a retained search for a Chief Scientific Officer.

The Keystone Symposia is a nonprofit 501(c)(3) educational organization, and is the recognized leader in conducting international biomedical and life science conferences. Each year more than 12,000 scientists gather at locations around the world for approximately 60 Keystone Symposia covering topics such as genomics, cancer, infectious diseases, immunology, metabolic diseases, biochemistry and cell biology. The Keystone Symposia is based in Summit County, Colorado, and convenes meetings in Colorado and other destinations around the globe.

The Chief Scientific Officer drives the scientific agenda of the organization, both internally and externally, and manages relationships across a wide range of non-profit, governmental and industry stakeholders to advance the mission of the organization. The CSO also plays a significant role in soliciting, cultivating, and stewarding relationships with academia, life science companies, advisory board members, as well as other collaborators and contributors.

The CSO will be a visionary leader who will enhance and expand the scientific content offerings and continue to build on a vision centered around the Keystone Symposia's mission to impact life sciences.

Inquiries and letters of application, along with a CV and letters of nominations should be submitted to: ***jeff@continuum-search.com***.

UNIVERSITY OF SOUTH CAROLINA
School of Medicine

FACULTY POSITION
PHARMACOLOGY, PHYSIOLOGY and NEUROSCIENCE

The Department of Pharmacology, Physiology and Neuroscience at the University of South Carolina (USC) School of Medicine in Columbia, SC invites applications for a faculty position at the Assistant Professor level. New faculty members will join a collegial and collaborative department at a university in the midst of an ambitious program to achieve national prominence in research and education. Candidates with research interests that complement the departmental research programs that focus on studying the molecular or cellular mechanisms underlying physiological processes, complex behaviors, or drug action are desirable, with priority given to USC's expanding neuroscience based research focus. Successful candidates will be expected to participate in medical and graduate teaching, with an emphasis on neuroanatomical and pharmacological topics. We are especially interested in research on novel compounds aimed at molecular or cellular targets that play a key role in neurological and neuropsychiatric disorders, since suitable candidates are eligible for appointment as target faculty in the NIH funded COBRE Center in Targeted Therapeutics (http://sccp.sc.edu/ctt). Applicants must have a doctoral degree and postdoctoral experience. Preference will be given to individuals with experience in medical education, funding success and potential, and research interests that enhance departmental programs targeting neurobiological disorders. Substantial start up funds and laboratory facilities will be provided. Candidates with necessary credentials will be considered for appointment on the tenure-track as an Assistant or Associate Professor.

Qualified applicants may apply by submitting a single electronic file (PDF or Word) that includes a cover letter summarizing qualifications, *curriculum vitae* and publication list, a statement of research plans and professional goals, and contact information for four references. The file should be attached to an e-mail message sent to **Dr. Jim Fadel** at **ppn.search@uscmed.sc.edu** with **"PPN Faculty Search"** as the subject. Review of applications will begin **July 1, 2016** and continue until the position is filled. For more information about the department including our research programs please visit: http://ppn.med.sc.edu/.

The University of South Carolina is an AA/EOE.

PENNSTATE HERSHEY College of Medicine

A **tenure-track faculty position** (Assistant or Associate Professor level) is available in the **Department of Anesthesiology and Perioperative Medicine** to strengthen and complement ongoing research. The candidate must have an outstanding record of research accomplishments, as documented by publications in leading peer-reviewed journals. Specific area of emphasis includes but is not limited to, addiction, inflammation, opioid management, aging and neurocognitive dysfunction. The successful applicant will be provided with a generous package including competitive salary, start-up funds and laboratory space. The candidate will be expected to establish an externally funded research program, actively collaborate with other basic and clinical investigators in the department, and mentor junior faculty and residents. To learn more about the Department of Anesthesiology please visit http://www.pennstatehershey.org/web/anesthesia/home.

Please send a cover letter, curriculum vitae, a 2 to 4 page research plan, and arrange for submission of 3 letters of recommendation. Application materials can be sent electronically to **vruizvelasco@psu.edu** or by mail to:

Victor Ruiz-Velasco, Ph.D.
Department of Anesthesiology and Perioperative Medicine
Penn State Hershey Medical Center
P.O. Box 850, H187
Hershey, PA 17033

By **Andy Tay**

Show us the money

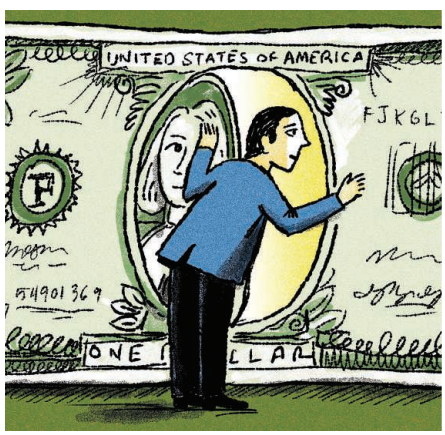
Nine months into working in the lab as a Ph.D. student, my professor called me into his office. I expected that he wanted my typical weekly research update, so I was dumbfounded when he told me that the grant supporting my research would expire in 3 months. He went on to say that I needed to start working on one of the other, funded projects in the lab or take on teaching assistantships to cover my stipend. I remained calm on the outside, but inside I was overwhelmed. I worried that changing my research direction would delay my graduation and hurt my publication record, making me less competitive for postdocs down the road. I was also kicking myself for not asking for details about how he would fund me and my work before I joined the lab.

There were good reasons why I hadn't. When I started grad school, I knew too little about how academic research is paid for to even know which questions to raise. And I worried that asking potential advisers about their grants would come across as rude or disrespectful. So I had kept quiet, opting to remain in the dark rather than risk alienating powerful senior scientists. But this approach left me utterly unprepared for my adviser's pronouncement, and I wished that I had been more proactive earlier.

I couldn't go back in time, though, so I spent most of that night coming up with possible solutions. "I can give up on my current project, even though it's very intellectually engaging, and start working on a topic I'm less excited about," I thought. "Or maybe I should become a teaching assistant, even though I want to focus on research, or even transfer to a different lab."

The next morning, I mustered the courage to ask my professor for a follow-up meeting. I told him my concerns about changing projects and shared some of the options I had come up with. Together, we brainstormed about how I could integrate my existing project in neural engineering with one that had more stable funding, on magnetotactic bacteria. Now I work on two separate but interconnected projects that together offer secure funding and intellectual fulfillment. This arrangement has increased my workload, but it has also significantly expanded my areas of expertise, offering me flexibility for my future research.

I wouldn't say that I'm glad I hit that funding snag, but a year later, I'm happy with my research direction, and I think the experience of working through adversity has helped improve my confidence and my communication skills. I have also found that I'm not the only student whose training has faced potential disruption because of an adviser's changing funding situation.



"None of us had known much about our advisers' funding."

A few weeks after my moment of crisis, I shared my experience with some other students, and many similar stories came out. I learned that one friend had to switch projects midway through her Ph.D. for similar reasons, and that some of my colleagues were earning their stipends by taking on extra teaching duties because their professors couldn't obtain enough funding. A common thread was that none of us had known much about our advisers' funding before we found ourselves in uncomfortable situations.

Realizing that many of us feared repercussions if we inquired about our professors' finances and that we felt ashamed about our predicaments, my friends and I set up departmental focus group discussions between administrative staff and graduate students to help us better understand how each lab is funded. It may seem like a small step, but I hope that it will help empower more graduate students to learn about their funding sources and take control of their own financial situation.

For myself, I am glad in hindsight that this issue arose early in my research career. It helped me learn how to come up with creative solutions to a challenge, and it taught me that it's always better to ask questions, even if it feels awkward or risky, than to stay quiet and hope for the best. ■

Andy Tay is a graduate student at the University of California, Los Angeles. Send your story to SciCareerEditor@aaas.org. For more on life and careers, visit sciencecareers.org.