A suicide spurs a bioethicist's crusade *p. 793*

Coral responses to climate change *pp. 798 & 895*

Simulation of a black hole evaporating *pp. 806 & 882*

Science

$10
23 MAY 2014
sciencemag.org

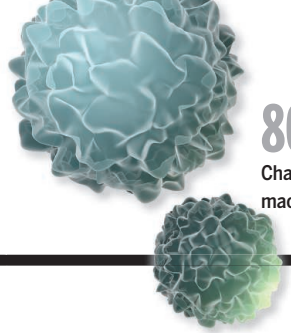
AAAS

SPECIAL ISSUE

Haves and have-nots

The science of inequality *p. 818*

CONTENTS



801 & 921
Characterizing tumor-fighting macrophages

23 MAY 2014 • VOLUME 344 • ISSUE 6186

836

SPECIAL SECTION

The science of inequality

INTRODUCTION

818 What the numbers tell us

NEWS

822 The ancient roots of the 1%
By H. Pringle

824 Our egalitarian Eden
By E. Pennisi

826 Tax man's gloomy message: the rich will get richer
By E. Marshall

828 Physicists say it's simple
By A. Cho

829 Can disparities be deadly?
By E. Underwood

832 While emerging economies boom, equality goes bust
By M. Hvistendahl

836 Tracking who climbs up—and who falls down—the ladder
By J. Mervis

RESEARCH REVIEWS

838 Inequality in the long run
By T. Piketty and E. Saez

843 Skills, education, and the rise of earnings inequality among the "other 99 percent"
By D. H. Autor

851 Income inequality in the developing world
By M. Ravallion

832

856 The intergenerational transmission of inequality: Maternal disadvantage and health at birth
By A. Aizer and J. Currie

862 On the psychology of poverty
By J. Haushofer and E. Fehr

SEE ALSO

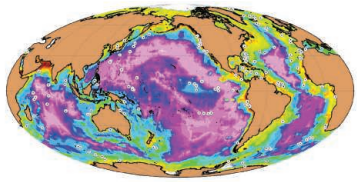
- EDITORIAL P. 783
- PERSPECTIVE P. 809
- BOOK REVIEWS PP. 811 & 812
- ONLINE PODCAST, VIDEO, AND BLOGS
- SCIENCEMAG.ORG/SPECIAL/INEQUALITY

ON THE COVER

A woman carries her son as she begs for money at a red light in Shanghai, China. Researchers are struggling to understand the causes and consequences of a rapid rise in inequality in many countries. *See page 818. Photo: REUTERS/Carlos Barria*

Published by AAAS

CONTENTS



889
Global distribution of sulfate-reduction rates

23 MAY 2014 • VOLUME 344 • ISSUE 6186

NEWS

IN BRIEF

784 Roundup of the week's news

IN DEPTH

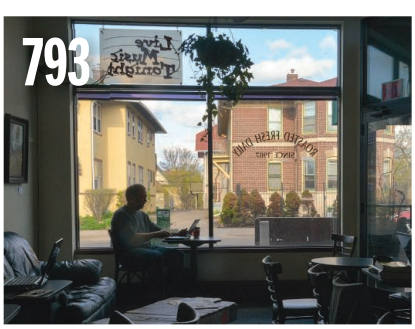
787 NASA PLANNERS GEAR UP FOR MARTIAN SAMPLE RETURN
Engineers designing the Mars 2020 rover keep an eye on geologic diversity—and the clock *By E. Hand*

788 REPLICATION EFFORT PROVOKES PRAISE—AND 'BULLYING' CHARGES
Global network fails to confirm 10 of 27 psychology findings, but some call project an inquisition *By J. Bohannon*

790 BLOCKBUSTER CLAIM COULD COLLAPSE IN A CLOUD OF DUST
Smoking-gun evidence for cosmic inflation may actually be radiation from within our galaxy *By A. Cho*

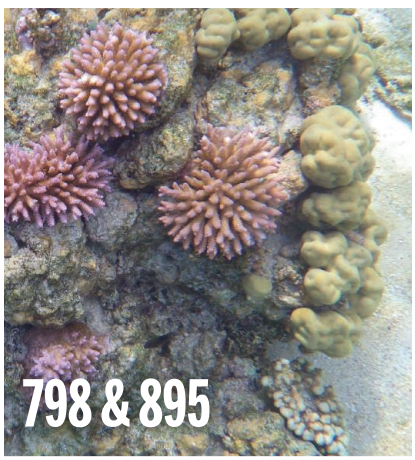
791 DOING THE MATH IN BERLIN
To keep Germany's science from stalling, research minister Johanna Wanka must break a federal-state impasse
By K. Kupferschmidt and G. Vogel

792 SOUTHERN SKY DIMS
Astronomy is at risk as Australia cuts most areas of science
By L. Dayton



FEATURES

793 A LONELY CRUSADE
A death during a clinical trial galvanized bioethicist Carl Elliott
By J. Couzin-Frankel



INSIGHTS

PERSPECTIVES

798 LAMARCK WAS PARTIALLY RIGHT—AND THAT IS GOOD FOR CORALS
Tabletop corals can adapt to changing temperature conditions on shorter time scales than thought *By C. M. Eakin*
▸ REPORT P. 895

800 DEEP MANTLE MATTERS
Experiments reveal how some deep seismic anomalies near the core-mantle boundary might be generated
By Q. Williams
▸ RESEARCH ARTICLE P. 877; REPORT P. 892

801 IDENTIFYING THE INFILTRATORS
Molecular analysis of macrophages reveals distinct types during tumorigenesis
By E. Gomez Perdiguero and F. Geissmann
▸ REPORT P. 921

803 RECORD-BREAKING WINTERS AND GLOBAL CLIMATE CHANGE
Rising greenhouse gas emissions may have played a role in the severe 2013/2014 winter in the U.S. Midwest *By T. Palmer*

804 A UNIFIED CAUSE FOR ADRENAL CUSHING'S SYNDROME
Mutations in a signaling protein underlie a class of functional adrenal tumors
By L. S. Kirschner
▸ REPORTS PP. 913 & 917

806 TESTING GAUGE/GRAVITY DUALITY ON A QUANTUM BLACK HOLE
A numerical test shows that string theory can provide a self-consistent quantization of gravity *By J. Maldacena*
▸ REPORT P. 882

807 TARGETING THE HOST IMMUNE RESPONSE TO FIGHT INFECTION
Strategies to modify immune responses to infection can be found in our genome
By J. K. Baillie

809 "UNDEMOCRACY": INEQUALITIES IN SCIENCE
Inequality, an intrinsic feature of science, has increased *By Y. Xie*
▸ THE SCIENCE OF INEQUALITY SECTION P. 818

BOOKS *ET AL.*

811 GDP
By D. Coyle, reviewed by N. Oulton
▸ THE SCIENCE OF INEQUALITY SECTION P. 818

812 THE SON ALSO RISES
By G. Clark with others, reviewed by M. Corak
▸ THE SCIENCE OF INEQUALITY SECTION P. 818

LETTERS

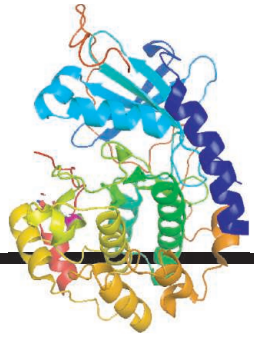
814 SPECIMEN COLLECTION: AN ESSENTIAL TOOL
By L. A. Rocha et al.

815 SPECIMEN COLLECTION: PLAN FOR THE FUTURE
By F.-T. Krell and Q. D. Wheeler

816 RESPONSE
By B. A. Minteer et al.

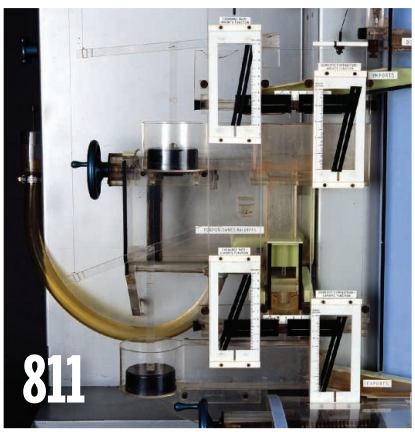
816 ERRATA

Published by AAAS



804, 913, & 917

A kinase mutation linked to adrenal tumors



RESEARCH

IN BRIEF

868 From *Science* and other journals

RESEARCH ARTICLES

871 INFECTIOUS DISEASE
Antibodies to PfSEA-1 block parasite egress from RBCs and protect against malaria infection *By D. K. Raj* et al.

877 DEEP EARTH
Disproportionation of (Mg,Fe)SiO_3 perovskite in Earth's deep lower mantle
By L. Zhang et al.
▸ PERSPECTIVE P. 800

REPORTS

882 COSMOLOGY
Holographic description of a quantum black hole on a computer
By M. Hanada et al.
▸ PERSPECTIVE P. 806

885 CHEMISTRY
Real-space imaging of molecular structure and chemical bonding by single-molecule inelastic tunneling probe
By C. Chiang et al.

889 SUBSURFACE MICROBES
Global rates of marine sulfate reduction and implications for sub–sea-floor metabolic activities *By M. W. Bowles* et al.

892 DEEP EARTH
Melting of subducted basalt at the core-mantle boundary
By D. Andrault et al.
▸ PERSPECTIVE P. 800

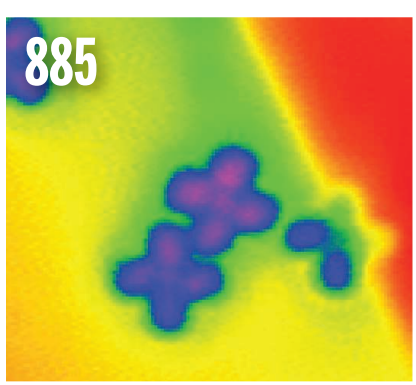
895 CORALS AND CLIMATE
Mechanisms of reef coral resistance to future climate change
By S. R. Palumbi et al.
▸ PERSPECTIVE P. 798

898 EVOLUTION
Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution *By K. J. Mitchell* et al.

901 DECISION-MAKING
FoxP influences the speed and accuracy of a perceptual decision in *Drosophila*
By S. DasGupta et al.

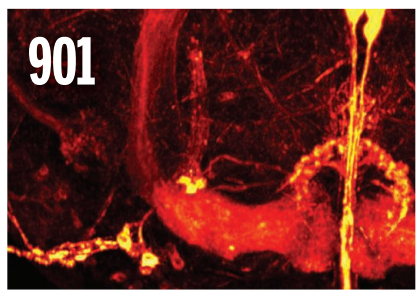
904 NEUROSCIENCE
Rapid Hebbian axonal remodeling mediated by visual stimulation
By M. Munz et al.

909 MICROBIAL GENOMICS
Stop codon reassignments in the wild
By N. N. Ivanova et al.



913 CANCER GENOMICS
Activating hotspot L205R mutation in *PRKACA* and adrenal Cushing's syndrome *By Y. Cao et al.*

917 CANCER GENOMICS
Recurrent somatic mutations underlie corticotropin-independent Cushing's syndrome *By Y. Sato* et al.
▸ PERSPECTIVE P. 804



921 CANCER IMMUNOLOGY
The cellular and molecular origin of tumor-associated macrophages
By R. A. Franklin et al.
▸ PERSPECTIVE P. 801

DEPARTMENTS

781 EDITOR'S NOTE
A new look
By Marcia McNutt

783 EDITORIAL
Inevitable inequality?
By Angus Deaton
▸ THE SCIENCE OF INEQUALITY SECTION P. 818

934 WORKING LIFE
A science career story
By Jim Austin

Science Staff 780
New Products 926
Science Careers 927

Science (ISSN 0000-0000) is published weekly on Fridays, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20009. Periodical mail postage (publication number 0000000) paid at Washington, D.C., and additional mailing offices. Copyright 2013 by the American Association of the Advancement of Science. The title SCIENCE is aregistered trademark of the AAAS. Domestic individual membership and subscription (51 issues): $153 ($74 allocated to subscription.) Domestic institutional subscription (51 issues): $1282; foreign postage extra: Mexico, Caribbean (surface mail): $55; other countries (air mail delivery): $85. First class air mail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST#1254 88122. Publications Mail Agreement Number 1069624. **Printed in U.S.A.**
Change of address: Allow 4 weeks, giving old and new address, and 8 digit account number. **Postmaster:** Send cahnge of address to AAS, PO Box 96718, Washington, D.C. 20090-6178. **Single copy sales:** $10 current issue, $15 back issue pre-paid includes surface postage; bulk rate on request. **Authorization to photocopy** material for internal or personal use under circumstances not falling withiin the fair use provisions of the Copyright Act is granted by AAAS to libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that $30.00 per article is paid directly to CCC, 222 Ros ewood Drive, Danvers, MA 01923. The identifiaction code for *Science* is 0036-8075. *Science* is indexed in the *Readers Guide to Periodical Literature* and in several specailized indexes.

Published by AAAS

EDITOR'S NOTE

A new look

Marcia McNutt is Editor-in-Chief of Science.

With this issue, you will note a new look of *Science*, in both print and online, to make the journal clearer, easier to navigate, and more inviting to the reader. Change can elicit strong reactions, and for that reason I would like to walk you through some of the changes and their rationale.

The tradition for the *Science* cover has been to avoid typography except for its name, its publisher (AAAS), and the date. The motivation had been esthetic: to allow stunning scientific visuals to be unmarred by words. But the lack of scale and context could leave the observer wondering whether a particular cover shows a hyperspectral image of a tectonic terrane or a stained cell. I used to tell my students that a scientific paper should never be written like a mystery novel; it should state the important discovery immediately. In the same vein, we are adding a descriptor that connects the image to the discovery that made it worthy of display on the cover of *Science*. The cover will also highlight other exceptional content, giving readers more reasons to explore the journal.

"The cover will also highlight other exceptional content, giving readers more reasons to explore the journal."

The contents page is now easier to read, with fewer subheadings: just "News," "Insights" (formerly "Commentary"), and "Research." For a quick overview, each research paper's title will have an overline that generally describes the topic of the paper. The choice of these words is unconstrained and can range from interdisciplinary descriptions to specific geographic sites or events (such as "Japan tsunami").

The Research section now opens with "In *Science* Journals" and "In Other Journals," brief descriptions of research that were previously published as "This Week in *Science*" and "Editors' Choice." The goal is to provide a roundup of notable research in the current issue of *Science* and in the *Science* family of journals that week, as well as in other science publications. We hope that this survey draws readers into the research portion of the journal. The names for these sections are also more intuitive: "Editors' Choice" was somewhat "inside baseball" and did not make its intent immediately clear to a reader. For online readers, be sure to check out the "Article Usage Statistics" for each article. This tool is a quick way to track downloads, media, and social media attention.

An additional page called "Working Life" is the home for personal essays, commentaries, and other writing that highlights everything from practical issues in the everyday work life of a science-related career to how to go about preparing for and finding a job.

A more consistent design now better accommodates the images and graphics that enhance the main points in News, Insights, and other portions of *Science*. Visualization has become increasingly important in telling the stories of science. Our goal is to evolve print and online platforms to bring the best visual experience to our readers.

Science's home page and News section online reflect only a "facelift" for now. We will be launching a new Web site for *Science* in the near future, with functionality that will not only augment your awareness and exploration of what *Science* has to offer daily and weekly, but will also, hopefully, elevate your exposure to the sciences overall. Stay tuned, and in the meantime let me know what you think of the new design by commenting* and/or taking our survey.†

– Marcia McNutt

10.1126/science.1255661

*http://comments.sciencemag.org/content/10.1126/science.1255999. †http://bit.ly/1vbmnSE.

PHOTO: (TOP RIGHT) STACEY PENTLAND PHOTOGRAPHY

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

EDITORIAL

Inevitable inequality?

The world is unequal in many dimensions; even life itself is unequally distributed. In the United States and other wealthy nations, only 2 to 6 children out of every 1000 die before age 1, yet there are 25 countries where more than 60 out of 1000 do so. There are 10 countries, all in Africa, where per-capita gross domestic product (GDP) is less than 10% of U.S per-capita GDP. These gaps are a legacy of the Great Divergence that began 250 years ago, in which sustained progress in health and wealth in Europe spread gradually to the rest of the world. Will such gaps continue to be an inevitable consequence of progress?

Between countries, inequalities in per-capita income show little sign of diminishing (although the rapid growth of China and India has pulled more than 2 billion people from destitution to somewhere below the middle of the world income distribution). Life expectancy at birth is rising more rapidly in the least healthy countries, because mortality decline among children raises life expectancy more than does mortality decline among adults (which predominates in rich countries). Countries with the poorest health also tend to have the lowest standards of living, compounding their disadvantage.

Within most countries, income inequality is rising. In the United States, the dispersion of income relative to its mean changed little from the mid-1950s to the mid-1970s, but since then has increased, a trend that is being echoed elsewhere. A general widening of the distribution in the United States had long been noted, based on household survey data, but the subject was given new momentum by the investigative work of Thomas Piketty and Emmanuel Saez. They used tax records to document top incomes, which cannot be captured from sample surveys. Piketty and Saez showed that the share of national income accruing to the top percentiles of the distribution in the United States was U-shaped from 1913 to the present, from 18% in 1913, falling to 7.75% in 1973, then rising to 19% in 2012 (22% if capital gains are included). Some have argued for looking at consumption—what people get, not what people earn—but there are no data on consumption among the very rich. Others have argued that publicly provided goods [Medicare (the U.S. health care program for the elderly), for example] should be accounted for, although such benefits cannot be used for rent or food.

In the United States, the top incomes are mostly salaries, not interest, dividends, or capital gains, although that will change as sustained top incomes generate large fortunes. Top earners are mostly financiers or chief executive officers, with some doctors, lawyers, and sports and entertainment celebrities in the mix. Whether these salaries reflect the contribution that their earners make to society sharply divides economists. If so, inequality can be seen as a reflection of the benevolent incentives that lead people to do the best for themselves and for society. If not, as I believe to be the case, talented young people are diverted from more worthwhile pursuits, which undermines national prosperity. Extreme income inequalities may also be incompatible with a well-functioning democracy; the rich may write the rules in their favor, and they may work against the public provision of health care or education, for which they pay a large share but have little personal need.

The distribution of wealth is more unequal than the distribution of income, and very high incomes will eventually pupate into very large fortunes, ultimately leading to a hereditary dystopia of idle rich. Piketty calls this "patrimonial capitalism," a state that is certain to occur if the rate of return to wealth remains higher than the rate of growth of the economy. This has been true for much of history and raises the question of whether extreme inequality is inevitable. It shouldn't be.

"...very high incomes will eventually pupate into very large fortunes, ultimately leading to a hereditary dystopia of idle rich"

Angus Deaton is the Dwight D. Eisenhower Professor of Economics and International Affairs at Princeton University, Princeton, NJ. E-mail: deaton@princeton.edu

– Angus Deaton

10.1126/science.1255661

PHOTO: (CENTER) GAMMA-RAPHO/GETTY IMAGES; (TOP RIGHT) STACEY PENTLAND PHOTOGRAPHY

Downloaded from www.sciencemag.org on May 26, 2014

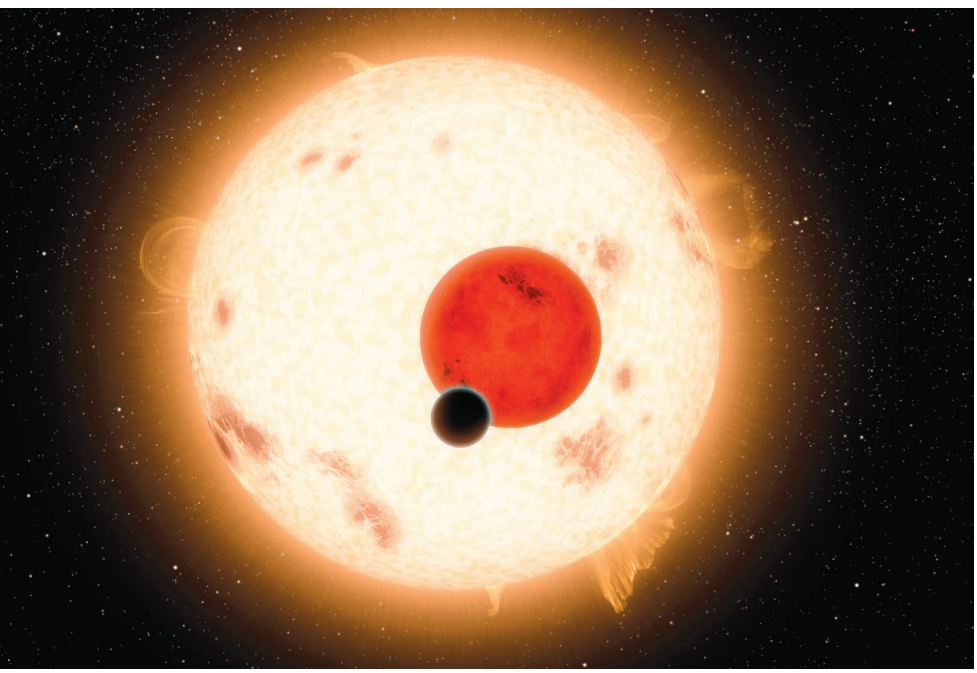
Published by AAAS

NEWS

> “If general relativity can make a 5-year-old cry, science becomes meaningful.”
>
> Physicist **Brian Greene**, speaking at this week’s The Future Is Here Festival about his son’s emotional reaction to the ending of his book *Icarus at the Edge of Time*.

IN BRIEF

Planet hunter gets a second life



A world with two suns: artist’s concept of Kepler-16b, one of the spacecraft’s discoveries.

A year after it lost its bearings in space, NASA’s exoplanet-hunting Kepler spacecraft has received a second lease on life. NASA last week approved $20 million in new funding to keep the Kepler mission going for two more years—with a little help from the sun. Launched in 2009, Kepler has helped astronomers identify scores of planets orbiting nearby stars, including the recent discovery of the first Earth-like planet in a star’s habitable zone. Last year, two of the spacecraft’s four gyroscopic positioning wheels stopped working; aiming the telescope precisely takes three. Now, however, mission managers have devised a plan to aim the telescope with only two working wheels and an assist from the subtle pressure of solar radiation. Kepler will now limit observations to targets along the ecliptic plane—the plane on which the orbital paths of planets in our solar system lie. Kepler’s second run—dubbed the K2 mission—is expected to begin with observations starting 30 May.

AROUND THE WORLD

China embraces open access

BEIJING | The Chinese Academy of Sciences (CAS) has adopted a policy of open access for scientific papers resulting from publicly funded research. “Open access will facilitate the dissemination and utilization of knowledge, [and] turn the knowledge produced by public investment effectively into innovation,” CAS noted in a 15 May statement. CAS researchers will now be required to deposit copies of articles published in academic journals into new repositories managed by their respective institutes. These databases will make articles publicly available online within 12 months. The move puts CAS in line with policies in the United Kingdom and at the U.S. National Institutes of Health that require free access to taxpayer-funded research papers within 6 to 12 months.

U.S., Cuba ocean science summit

WASHINGTON, D.C. | Marine scientists from Cuba and the United States are planning a push to improve collaboration between the two estranged nations. At an unusual 13 May meeting hosted by Senator Sheldon Whitehouse (D–RI), researchers and policymakers from both nations

One potential area of U.S.-Cuba cooperation: mangrove ecosystems, home to largemouth black bass.

discussed existing projects and future possibilities. On the agenda: drafting a new bilateral agreement on marine research and revising trade rules that make it hard for people and scientific equipment to move between the two nations. The idea

PHOTOS: (TOP TO BOTTOM) NASA/JPL-CALTECH; © REINHARD DIRSCHERL/VISUALS UNLIMITED/CORBIS

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

RANDOM SAMPLE

Next stop: the twilight zone

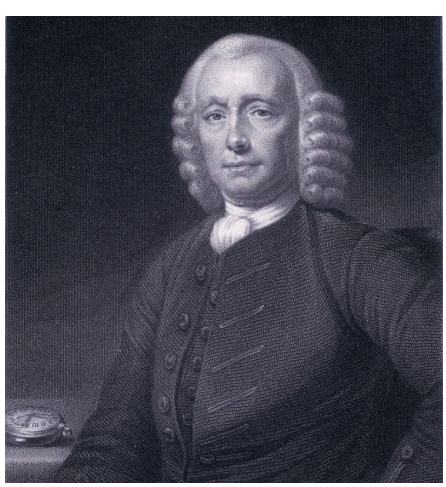
Three years ago, Adriaan "Dutch" Schrier turned his hobby into science. A Curaçao businessman who founded an aquarium and was a scuba diver for almost 50 years, he gave up breathing mixed gases when he turned 60. But, still eager to visit the underwater world, he built the "Curasub," a submersible that seats five. In 2011, Schrier invited scientists at the Smithsonian Institution's National Museum of Natural History to take a peek at the ocean's barely lit twilight zone between 200 and 1000 meters, depths relatively unexplored because they are too deep for divers and not challenging enough for deeper diving submersibles. Since then, researchers have extensively surveyed the reef off Curaçao; discovered dozens of new species such as the yellow-spotted golden bass, described last week in *PLOS ONE* (http://scim.ag/oddlarva); and installed temperature loggers and collection platforms for long-term monitoring. Even after 1200 runs, "every trip I go down I find something completely new," Schrier says. And he likes the solitude: "When I'm cruising down at 600 or 700 feet, there's no stoplights, no traffic; you're the only one down there."

Schrier and Smithsonian zoologist Carole Baldwin in the Curasub.

is to "use marine science as a form of diplomacy," says marine biologist David Guggenheim, president of the nonprofit Ocean Doctor and a meeting organizer. Whitehouse, a vocal supporter of ocean science, has said he'll take the lead on drafting the agreement, which would need White House approval.
http://scim.ag/Cubascience

A new Longitude Prize

LONDON | In 1714, clockmaker John Harrison cracked a seemingly intractable problem: how to find the longitude of a ship at sea. With his marine chronometer, Harrison won the £20,000 Longitude Prize, offered by the British Parliament for what it considered the most challenging conundrum of the 18th century. Now, 300 years later, the United Kingdom is looking for the next John Harrison—and has raised the stakes to £10 million. But first, there needs to be a modern-day challenge to solve. People can vote online or via text for one of six themes—flight, food, antibiotics, paralysis, water, and dementia—beginning on 22 May; the chosen challenge will be announced on 25 June, and experts will then convene to fine-tune exactly what criteria must be met to win the prize, sponsored by the charity Nesta and the U.K. government–funded Technology Strategy Board. Anyone around the world is eligible to enter their invention—but it could be years before a winner is chosen, noted Geoff Mulgan, Nesta's chief executive.


John Harrison won the original Longitude Prize in 1714.

Transparency U-turn?

LONDON | Researchers are worried that the European Medicines Agency (EMA) is backpedaling on a pledge to open clinical trials data to public scrutiny. Draft documents dated 5 May say that registered users would only be allowed to view trial information on screen and would not be permitted to "download, save, edit, photograph, print, distribute or transfer the information." "I am now concerned about what appears to be a significant change in EMA's policy, which could undermine the fundamental right of public access to documents established by EU law," said European Ombudsman Emily O'Reilly in a statement on 16 May. EMA spokesman Martin Harvey-Allchurch denies that the agency is reneging on its promises. He stresses that the plans are still open for discussion before EMA presents the revised draft policy to its managing board on 12 June.
http://scim.ag/EMAopen

PHOTOS: (TOP TO BOTTOM) © BARRY B. BROWN/WILDHORIZONS.COM; © IMAGE ASSET MANAGEMENT LTD./ALAMY

Published by AAAS

BY THE NUMBERS

80,000

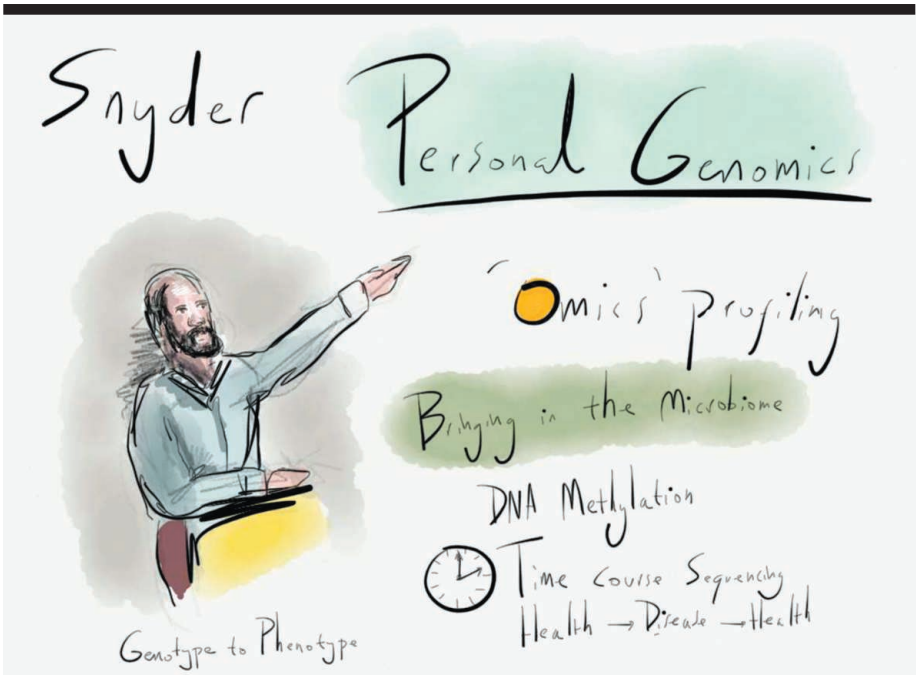
Estimated weight, in kilograms, of a new species of titanosaur unveiled 17 May. The fossils, found in Argentine Patagonia, suggest the dinosaurs may have been the largest to walk the Earth.

6

Number of years by which global average lifespan increased from 1990 to 2012, according to the World Health Organization.

60

Percent by which e-cigarettes improve a smoker's chances of quitting compared with nicotine replacement therapies or willpower alone, according to a 21 May study in *Addiction* based on self-reported data.



Stanford University geneticist Michael Snyder discusses personalized genomics.

Meeting tweeting, in cartoons

For Alex Cagan, 140 characters were just not enough. At the Biology of Genomes meeting earlier this month in Cold Spring Harbor, New York, this graduate student from the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, contributed 54 cartoons to the Twittersphere. During each talk, he drew a caricature of the speaker and added a few choice phrases. At meetings, he likes to draw the speaker whenever he takes notes, regardless of whether he's tweeting. "When I try to remember the talk, if I have notes with the actual person, it triggers my memory," he explains. "When I just have text, it's harder to remember." His tweets were well-received, with one fellow tweeter commenting that the sketches "are wonderful and, unlike most conference tweets, actually informative!"

Gliders sample glacial plumes

PRINCE WILLIAM SOUND, ALASKA | Two sensor-packed remote-controlled wave gliders—bearing a strong resemblance to nerdy yellow surfboards—are part of a new study in the Gulf of Alaska that is helping scientists understand the impact of melting glaciers on ocean acidification. Plumes of fresh glacial water, which peak in summer and fall, have lower alkalinity than seawater, reducing the ocean's ability to buffer against increasing acidification. That can make the waters more corrosive to organisms that build shells. The solar cell–powered gliders will spend 5 months collecting data for the study, funded by the National Oceanic and Atmospheric Administration. A third underwater glider will slide beneath the surface to hunt for freshwater plumes. The robotic fleet, which set sail earlier this month, is scheduled to work through early September.

A wave glider analyzes glacial plumes in Prince William Sound.

Congress pushes for faster cures

WASHINGTON, D.C. | Members of the U.S. House of Representatives have launched a multiyear effort to modernize the country's system for developing and approving drug treatments. The so-called 21st Century Cures Initiative, spearheaded by Energy and Commerce Committee Chair Fred Upton (R–MI) and Representative Diana DeGette (D–CO), held its first hearing this week to get input from experts involved in a September 2012 President's Council of Advisors on Science and Technology report on drug innovation. In coming months, the initiative will solicit feedback from the Food and Drug Administration, the National Institutes of Health, as well as scientists, industry experts, and patients, before considering possible legislation.
http://scim.ag/fastercures

PHOTOS: (TOP TO BOTTOM) ALEX CAGAN; WILEY EVANS/NOAA

Published by AAAS

IN DEPTH

Scientists hope a future mission will ferry Mars rocks back to Earth.

PHOTO: NASA/JPL-CALTECH

PLANETARY SCIENCE

NASA planners gear up for martian sample return

Engineers and scientists designing the Mars 2020 rover keep an eye on geologic diversity—and the clock

By **Eric Hand**

ARLINGTON, VIRGINIA—In the sea of planetary scientists here last week for the first Mars 2020 rover landing site workshop, the engineer stood out. He was decked out in his customary rockabilly regalia: earrings, pompadour, and blue jeans cuffed meticulously above ankle-length black boots. But his nametag was spare. It just read "Adam."

"First name only?" chuckled Al Chen, another engineer at the Jet Propulsion Laboratory (JPL) in Pasadena, California.

Around these parts, Adam—JPL engineer Adam Steltzner—needs no introduction. He led the development of the jet-powered "sky crane" that unspooled NASA's Curiosity rover to the surface of Mars with unprecedented precision in 2012. Now, NASA is planning its next rover, a $1.5 billion machine to be launched in 2020. Steltzner's sky crane will be enlisted again. But NASA has called on Steltzner to stage a different engineering tour de force: gathering a trove of rock and soil samples that will ultimately be returned to Earth, to be inspected for clues to martian geology and signs of ancient life. "It's no accident that Adam is on this," Chen says. "We're bringing the A game."

Scientists want the rover to drill at least 31 rock samples weighing about 15 grams apiece during its 2-year mission and pack them in a honeycomblike repository that a later mission will retrieve. It will have to work faster than the current rover, Curiosity, which has drilled just three samples in its 20 months on Mars. To make the next rover's job easier, the nearly 100 planetary scientists at the workshop focused on potential landing sites that pack maximum geologic diversity into relatively small areas (http://scim.ag/Mars2020). Mission planners are also planning to limit the rover's other scientific tasks, keeping it nimble so it can reach as much of that diversity as possible.

Steltzner's issues begin with the drilling. On Earth, drills are lubricated and kept cool with water, which also flushes out the detritus. On Mars, Steltzner is limited to rotary-percussive drilling, which works without water but carries risks. Laboratory scientists need cohesive, uncontaminated rock samples—but how do you hammer and destroy the material around a sample the size of a stick of chalk without fracturing the thing you want? Temperatures also have to be controlled: If they rise too high, they could alter the hoped-for organic molecules inside the rocks.

Designing the sealed tubes in which samples will await return poses another challenge. Make them out of stainless steel—a strong, ductile material—and you risk contamination from reactions with the metal. Make them out of an inert material like sapphire—perfect for avoiding contamination—and they could be too brittle for a bumpy ride back to Earth.

Above all, Steltzner needs the rover to work quickly. To make more time for sample selection, Steltzner says, engineers might create an onboard "parking lot" for samples so that scientists could evaluate them while the rover is on the go—before committing them to one of the precious caching spots. They might also design the cache—the chamber that holds the sample tubes—so that samples can be ejected, or even provide multiple caches so that scientists can change their minds about which rocks to send home.

Mission planners are also trying to limit the rover's capabilities. A science definition team report, released last July, recommended a $100 million instrument payload that would make brisk measurements. Time-consuming experiments like Curiosity's CheMin—the first x-ray diffraction in-

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

strument flown to another planet, used to analyze the crystal structure of samples—would find no place on board.

That hasn't stopped scientists from proposing even more complicated experiments. In January, NASA received 58 instrument proposals. One would seek to germinate plants as a first step toward increasing the amount of oxygen on the planet. Another team would build a solar-powered helicopter that could fly reconnaissance sorties of hundreds of meters a day.

Payload selection is expected in mid-July. Jack Mustard, a planetary scientist at Brown University who led the science definition team, hopes NASA will opt for a conservative payload—something more along the lines of what is riding aboard the small Opportunity rover, now in its 10th year of operation. The same considerations will drive the choice of a landing site. Sites that require more than 10 kilometers of driving will be problematic, Mustard says. Sites where sample collection can begin immediately after landing are "going to trump."

Chen, taking over for Steltzner in leading the entry, descent, and landing phase of the mission, hopes he can help by delivering the rover to just the right spot. He is considering two modifications to the Curiosity system that could improve landing precision. One is a device that would deploy the entry capsule's parachute based on distance to the target, rather than an estimate of velocity. This range trigger would shrink the landing ellipse—where engineers are confident the rover will end up—from 25 kilometers to 13 kilometers long. According to project manager John McNamee, the trigger would require changing just a few lines of code.

Another change, called terrain relative navigation (TRN), would allow the rover to land at sites that would otherwise be too hazardous. It would outfit the descent module to compare real-time images of the terrain it was approaching with stored images of the landing site, enabling the sky crane to avoid regions with too many rocks. McNamee says the price tag for the TRN is higher—on the order of $10 million—and could ripple through the landing system and affect it in other ways. He wants to be convinced that the change is absolutely necessary before allowing it.

Scientists at the workshop spoke with near unanimity, however: The TRN would open up the most interesting landing sites, which previous missions had discarded for safety reasons. Furthermore, the technology would help guide the subsequent missions needed to fetch the cache. Steltzner concurs. "We all kinda think that the TRN is the steering wheel, the headlights," he says. ■

A 2008 study (*right*) showing that cleanliness influences moral judgments was not replicated in a new study (*left*).

PSYCHOLOGY

Replication effort provokes praise—and 'bullying' charges

Global network fails to confirm 10 of 27 psychology findings, but some call project an inquisition

By **John Bohannon**

After a string of scandals involving accusations of misconduct and retracted papers, social psychology is engaged in intense self-examination—and the process is turning out to be painful. This week, a global network of nearly 100 researchers unveiled the results of an effort to replicate 27 well-known studies in the field. In more than a third of the cases, the result was a complete failure.

As the replicators see it, the failed do-overs are a healthy corrective. "Replication helps us make sure what we think is true really is true," says Brent Donnellan, a psychologist at Michigan State University in East Lansing who has undertaken three recent replications of studies from other groups—all of which came out negative. "We are moving forward as a science," he says.

But rather than a renaissance, some researchers on the receiving end of this organized replication effort see an inquisition. "I feel like a criminal suspect who has no right to a defense and there is no way to win," says psychologist Simone Schnall of the University of Cambridge in the United Kingdom, who studies embodied cognition, the idea that the mind is unconsciously shaped by bodily movement and the surrounding environment. Schnall's 2008 study finding that hand-washing reduced the severity of moral judgment was one of those Donnellan could not replicate.

About half of the replications are the work of Many Labs, a network of about 50 psychologists around the world. The results of their first 13 replications, released online in November, were greeted with a collective sigh of relief: Only two failed. Meanwhile, Many Labs participant Brian Nosek, a psychologist at the University of Virginia in Charlottesville, put out a call for proposals for more replication studies. After 40 rolled in, he and Daniël Lakens, a psychologist at Eindhoven University of Technology in the Netherlands, chose another 14 to repeat.

The output of the new batch of replications, published alongside the previous 13 this week in an issue of *Social Psychology*

PHOTO: (SCALE) FRESHPAINT/ISTOCKPHOTO; (PHOTO ILLUSTRATION) C. SMITH/*SCIENCE*

Published by AAAS

guest-edited by Nosek and Lakens, is less reassuring. All told, the researchers failed to confirm the results of 10 well-known studies, such as the social psychological effects of washing one's hands, holding cups of warm or cold liquid, or writing down flattering things about oneself. In another five cases, the replications found a smaller effect than the original study did or encountered statistical complications it did not report. For embodied cognition and also for behavior priming—the study of how exposure to one stimulus, such as the word "dog," changes one's reaction to another, such as a photo of a cat—the results are particularly grim. Seven of the replications focused on experiments in these areas, and all but one failed.

No one is suggesting misconduct in any of the original studies, but the results are further blows to a field shaken several years ago when a towering figure in priming research, Diederik Stapel, confessed to faking data (*Science*, 7 December 2012, p. 1270). And earlier this month, Jens Förster of the University of Amsterdam, a pioneer of embodied cognition research, was accused by a Dutch government-appointed ethics panel of data manipulation—charges he denies (*Science*, 9 May, p. 566).

Nor should the results be taken as a general indictment of psychological research, because the targeted studies were not a random sample, Nosek says. "They are entirely cherry-picked," he says, based on the importance of the original study and the feasibility of replicating it.

Some of the authors of the targeted studies, however, feel not just singled out but persecuted. Schnall, for example, contends that the replications were not held to the same peer-review standard as her original studies. "I stand by my methods and my findings and have nothing to hide," she says.

The replications did employ an alternative model of peer review, called preregistration, promoted by the Center for Open Science, a nonprofit organization cofounded by Nosek (*Science*, 30 March 2012, p. 1558). Before any data were collected, the replicators submitted their experimental design and data analysis plan to external peer reviewers, including the principal investigator of the original study. The subsequent data analysis and conclusions were reviewed only by Nosek or Lakens.

Schnall contends that Donnellan's effort was flawed by a "ceiling effect" that, essentially, discounted subjects' most severe moral sentiments. "We tried a number of strategies to deal with her ceiling effect concern," Donnellan counters, "but it did not change the conclusions." Donnellan and his supporters say that Schnall simply tested too few people to avoid a false positive result. (A colleague of Schnall's, Oliver Genschow, a psychologist at Ghent University in Belgium, told *Science* in an e-mail that he has successfully replicated Schnall's study and plans to publish it.)

Some replicators leaked news of their findings online, long before publication and in dismissive terms. On his personal blog, Donnellan described his effort to repeat Schnall's research as an "epic fail" in a December post titled "Go Big or Go Home," which was then widely circulated on Twitter. Donnellan defends the early announcement. "I feel badly, but the results are the results," he says.

Schnall, however, says that her work was "defamed." She believes she was denied a large grant in part because of suspicions about her work and says that a reviewer of one of her recently submitted papers "raised the issue about a 'failed' replication." She adds that her graduate students "are worried about publishing their work out of fear that data detectives might come after them and try to find something wrong."

Other researchers whose work was targeted and failed to replicate told *Science* that they have had experiences similar to Schnall's. They all requested anonymity, for fear of what some in the field are calling "replication bullying."

Yet some whose findings did not hold up are putting a positive spin on the experience. "This was certainly disappointing at a personal level," says Eugene Caruso, a psychologist at the University of Chicago Booth School of Business in Illinois, who in 2013 reported a priming effect—exposing people to the sight of money made them more accepting of societal norms—that failed to replicate. "But when I take a broader perspective, it's apparent that we can always learn something from a carefully designed and executed study." Caruso now has a larger and more nuanced version of his study under way.

The replications in psychology reflect a growing trend in science (see table). The field's bruising experience shows that such efforts should be handled carefully, stresses Daniel Kahneman, a psychologist at Princeton University, whose work was successfully replicated by the Many Labs team. "The relationship between authors and skeptics who doubt their findings is bound to be fraught," he says. "It can be managed professionally if the rules that apply to both sides are clearly laid out."

To reduce professional damage, Kahneman calls for a "replication etiquette," which he describes in a commentary published with the replications in *Social Psychology*. For example, he says, "the original authors of papers should be actively involved in replication efforts" and "a demonstrable good-faith effort to achieve the collaboration of the original authors should be a requirement for publishing replications." In the case of this week's replications, "the consultations did not reach the level of author involvement that I recommend." However, he notes that "authors of low-powered studies with spectacular effects should not wait for hostile replications: They should get in front of the problem by replicating their own work."

For his part, Nosek hopes that the tensions will be short-lived growing pains as psychology adjusts to a demand, from within and outside the field, for greater accountability. "Our primary aim is to make replication entirely ordinary," he says, "and move it from a threat to a compliment." ■

Repeat after me

Select efforts in replication of research

EFFORT	REPLICATION TARGET
Reproducibility Project: Cancer Biology	50 high-impact cancer studies published from 2010 to 2012
Reproducibility Project: Psychology	Articles published in 2008 from three psychology journals
Reproducibility Initiative	Hub for authors to request independent replications of their experiments
Many Labs project	Global network for orchestrating large replications
Reproducibility in Computer Science	Checks software code in 613 applied computer science papers
Crowdstorming project	More than 50 analysts address same research question using shared data set

Published by AAAS

COSMOLOGY

Blockbuster claim could collapse in a cloud of dust

Smoking-gun evidence for cosmic inflation may actually be radiation from within our galaxy

By **Adrian Cho**

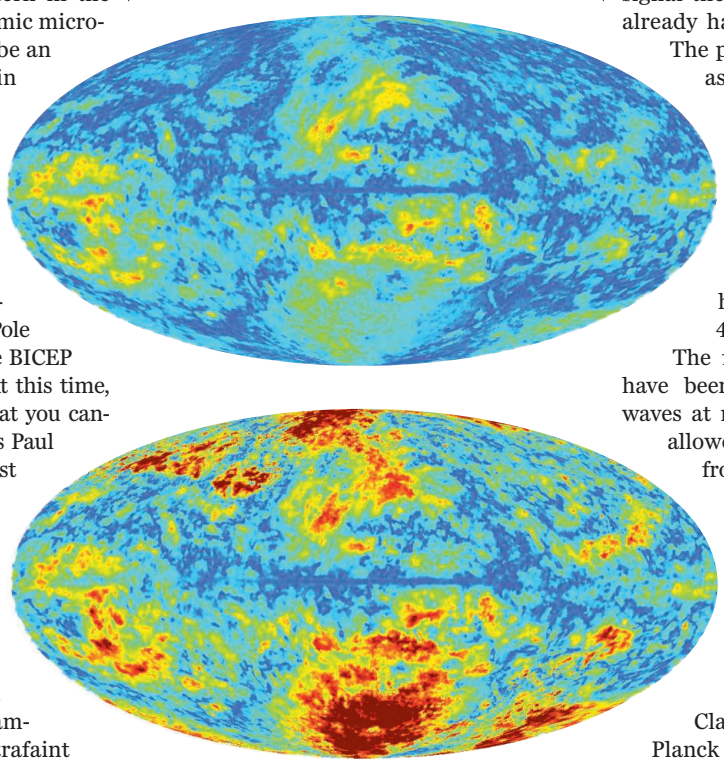
Perhaps it was too good to be true. Two months ago, a team of cosmologists reported that it had spotted the first direct evidence that the newborn universe underwent a mind-boggling exponential growth spurt known as inflation (*Science*, 21 March, p. 1296). But last week a new analysis suggested the signal, a subtle pattern in the afterglow of the big bang, or cosmic microwave background (CMB), could be an artifact produced by dust within our own galaxy.

"We're certainly not retracting our result," says John Kovac, a cosmologist at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and co-leader of the team, which used a specialized telescope at the South Pole known as BICEP2. Others say the BICEP team has already lost its case. "At this time, I think the fair thing to say is that you cannot claim detection—period," says Paul Steinhardt, a theoretical physicist at Princeton University.

From 2010 through 2012, BICEP2 peered at a small patch of the CMB to measure the polarization of the microwaves as it varies from point to point. On 17 March, BICEP researchers announced at a press conference in Cambridge that they had spotted ultrafaint pinwheel-like swirls in the sky. Those swirls, or B modes, are most likely traces of gravitational waves rippling through space and time during the 10^{-32} seconds that inflation lasted, the BICEP team says, and they fulfill a key prediction of the theory of inflation. Many cosmologists hailed the detection as a "smoking gun" for that theory.

But dust within our galaxy can also emit microwaves that mimic the signal. Much or all of the BICEP signal could come from that dust, says Raphael Flauger, a theoretical physicist at the Institute for Advanced Study in Princeton, New Jersey, who performed the new analysis. He presented it at Princeton University on 15 May.

BICEP researchers estimated that "galactic foreground" was negligible. They modeled it several ways, as they report in the paper announcing their claim, which has been submitted to a journal that Kovac declined to name. The most sophisticated model relied on a map of the foreground generated by the European Space Agency's spacecraft Planck, which mapped the CMB across the entire sky from 2009 until last year. Because Planck has not yet released that data, researchers scanned the map from a slide presented at a talk.



A reconstruction of the contaminated foreground map BICEP used (*top*) and the corrected map.

The BICEP team apparently assumed the map shows radiation only from dust inside our own galaxy. In reality, it may also contain an unpolarized haze from other galaxies, which would make the microwaves from within the galaxy look less polarized than they are. So using the map could have led the researchers to underestimate the galactic foreground and overestimate the CMB signal.

To test that idea, Flauger used other Planck data—also scraped from a talk—to correct the map BICEP used (see figure). The foreground appears stronger in the corrected map and could account for the entire BICEP signal, he reported.

BICEP's Kovac says his team always made it clear that they couldn't be sure how much of their signal really comes from the CMB. And he won't put a number on it. "The six models of polarized dust that we use are all quite uncertain," he says, "so the statements that we make about the interpretation are necessarily more qualitative."

Flauger stresses that he hasn't proved that BICEP's signal is spurious. "I'm still hoping that after all I've done there is a signal there," he says. However, the claim already has a couple of strikes against it. The polarization signal is twice as big as an upper limit Planck researchers set indirectly by measuring temperature variations in the CMB. Making the two results jibe would be difficult, researchers say. The size of the signal also causes headaches for theorists trying to explain how inflation happened (*Science*, 4 April, p. 19).

The flap over the BICEP signal may have been predictable. Sampling microwaves at multiple frequencies would have allowed BICEP2 to separate foreground from CMB by itself. But the telescope was designed to maximize overall sensitivity and tracked only one frequency. "All the other experiments that I know of use multiple frequencies," says Charles Bennett of Johns Hopkins University in Baltimore, Maryland.

Clarity may come in October, when Planck researchers plan to release their polarization data. If Planck shows that the foreground is small and the BICEP signal is real, then the BICEP team should still get credit for the discovery, says Marc Kamionkowski, a cosmologist at Johns Hopkins. But David Spergel, a cosmologist at Princeton, says that in that case, the Planck team alone should get the credit.

If Planck shoots down the result, the credibility of science may suffer, Bennett says: "You talk about something like climate change and the public says, 'Yeah, but you guys say you found something and then you take it back all the time.' " ■

PHOTOS: RAPHAEL FLAUGER

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

Wanka trained as a mathematician in East Germany.

PHOTO: KAY NIETFELD/DPA/PICTURE-ALLIANCE/NEWSCOM

RESEARCH FUNDING

Doing the math in Berlin

To keep Germany's science from stalling, research minister Johanna Wanka must break a federal-state impasse

By **Kai Kupferschmidt** *and* **Gretchen Vogel**

BERLIN—When Johanna Wanka was a mathematician in communist East Germany, any foray into politics carried huge risks. In the mid-1980s, she and her husband Gert, also a mathematician, turned the weeklong political indoctrination held every fall at their university in the small town of Merseburg into an open discussion session. It earned them a disciplinary action, effectively freezing their careers. When a close friend was arrested in 1982, Wanka realized that if she and her husband were ever put in prison, their two young children could be sent to an institution. "That was our biggest fear," Wanka told *Science* recently in a wide-ranging interview.

Today, Wanka is at the heart of science politics as federal research minister of the reunited Germany. Later this year, she will literally move close to the pinnacle of German power, when she and her staff relocate from the current ministry building—which served as West Germany's embassy in East Berlin—to a gleaming new one within view of the office of Chancellor Angela Merkel, a physicist who also grew up in the former East.

Politics isn't nearly as terrifying as it was 30 years ago, but after 15 months in office, Wanka, now 63, faces major challenges. Discussions on how to spend €9 billion in promised funding for research and education over the next 4 years are dragging on because of deep disagreements between the federal and *Länder* (state) governments over who decides how the money is used. In the balance hangs the fate of several landmark programs started by Wanka's predecessor, Annette Schavan.

German science leaders want action. In a rare joint statement on 19 May, the heads of the German Research Foundation (DFG), the German Rectors' Conference, and the German Council of Science and Humanities called on politicians to "surmount the paralyzing blockade and to finally act." They asked for a "substantial part" of the €9 billion to go to the country's budget-squeezed universities and to research institutions, and pleaded for a change in the constitution that would allow the federal government to help with long-term financing of universities. Failure to act would set back the research system for years to come, they warned.

Wanka previously served as minister of science and culture in the states of Brandenburg and Lower Saxony; she took her current post in February 2013, when Schavan, a close ally of Merkel, resigned after losing her Ph.D. in a plagiarism scandal (*Science*, 15 February 2013, p. 747). Coming in just months before federal elections, she had little room to maneuver. She kept her post after the center-right Christian Democrats—the party to which she and Merkel belong—formed a new coalition government with the Social Democrats in November.

Wanka found Germany's science landscape in better shape than any time since the second world war. Research spending has increased continuously since Merkel took office, and in 2012, public and private research spending combined, at €79.5 billion, reached 3% of the gross domestic product for the first time. Meanwhile, several high-profile programs have helped make German research more competitive. The widely praised €4.6 billion Excellence Initiative, for instance, gave universities the chance to compete for extra funding and the title of "elite university" (*Science*, 20 October 2006, p. 400). Meanwhile, the Joint Initiative for Research and Innovation has poured money into nonuniversity organizations like the Max Planck Society and the Helmholtz Association, with 5% annual increases from 2011 through 2015.

Science managers are anxious for the good times to continue. The new coalition has set aside an extra €9 billion for research and education, and it has pledged part of that money to keep raising the budgets of nonuniversity organizations. Wanka says she wants to lock in a 3% annual increase for the next 4 or 5 years. That is welcome news, says Karl Ulrich Mayer, president of the Leibniz Association, which is home to 89 research institutes and museums. But between a 5% salary increase negotiated by unions last year and already-planned projects, "there will be a squeeze on real increases," he says.

> ***"Changing the constitution is one of my major goals for this legislative period. But it will be difficult."***
>
> **Johanna Wanka,** Federal Minister of Education and Research

The bigger worry, however, is the state of the universities. Under the German constitution, basic financing of education, including universities, is the responsibility of the *Länder*. The coalition has set aside two-thirds of the €9 billion to help the states. State leaders want a large portion of that money without any strings at-

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

tached, but the federal government wants at least part of it earmarked for universities. Student numbers are rising every year, buildings are old and in need of repair, and many *Länder* are pressed for money. As a result, the gap between research at universities and at institutions like Max Planck is widening, says Horst Hippler, head of the German Rectors' Conference. "The key question is how to finance the universities with federal funds," says Ernst-Ludwig Winnacker, secretary-general of the Human Frontier Science Program Organization and a former DFG president.

The Excellence Initiative was one solution, but its funding runs out in 2017. Figuring out how to build on the program is a top priority, Wanka says. A formal evaluation will start in June, and she is discussing with the *Länder* her ideas for what to do next. Schavan also used another option: allowing universities and nonuniversity research institutes to merge into cooperative units eligible for federal funding. (The Karlsruhe Institute of Technology is one example.)

But many science leaders say these initiatives are only Band-Aids and that a constitutional amendment is necessary. Any federal encroachment on the powers of the *Länder* is contentious, however; Wanka will need to wrangle state politicians into cooperation. "Changing the constitution is one of my major goals for this legislative period," she says. "But it will be difficult." Asked why it is taking so long to pour the coalition agreement into concrete policies, Wanka says setting aside the money was an important first step. "Now we're talking about how these investments should happen. That is complex and needs to be well thought out," she says. "But we are in the home stretch."

Wanka's experience in East Germany may help her navigate the challenge. She understands the universities' predicament, having spent much of her professional life at the University of Applied Sciences in Merseburg, including 6 years as its rector. And as a mathematician, she may have a different style than most politicians. In the run-up to German reunification, many East Germans who worked in science or engineering—Wanka and Merkel are just the most prominent examples—came into politics, says Jörg Hacker, head of the German National Academy of Sciences Leopoldina; they tend to be more observant and analytical, he says. "I think that has really enriched our political culture."

Wanka says politicians with a background in science perhaps approach politics more pragmatically. "We may think more ... like in a chess game." German scientists are waiting expectantly for her next move. ■

AUSTRALIA

Southern sky dims

Jobs and instruments are at risk as government cuts radio astronomy and other science programs

By **Leigh Dayton**

The view of the southern sky got dimmer last week as Australia's national research body, the Commonwealth Scientific and Industrial Research Organisation (CSIRO), announced deep spending cuts—likely 10% to 15%—to astronomy and space science. At worst, the cuts, part of an austerity budget announced by the country's conservative government, could ultimately close one of Australia's five radio observatories. At best, they will mean less observation time and fewer instruments, sapping an area of scientific strength and renewing questions about the government's commitment to science.

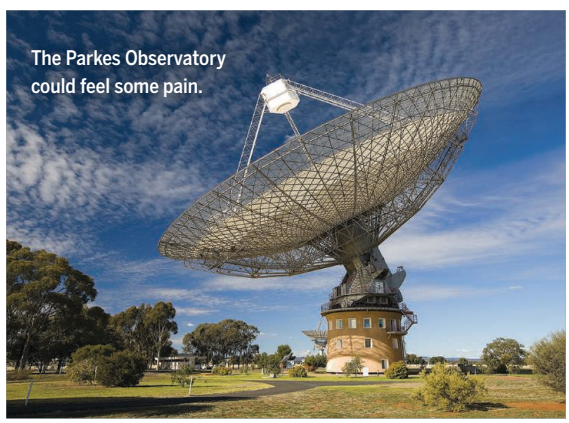
CSIRO chief Megan Clark broke the news in a 14 May memorandum, which came a day after the government announced it was reducing the agency's spending by 5.4%. Radio astronomy, which dominates CSIRO's Astronomy and Space Science (CASS) program, takes one of the heaviest blows. "Obviously, I'm disappointed," says CASS head Lewis Ball. "But I understand that radio astronomy is not immune from the overall funding pressures."

Those pressures mean CSIRO could lose 420 jobs by June 2015, according to Clark's memo. That's on top of 300 jobs already lost. Overall, CSIRO could end up with 2500 fewer staff members than it had in the 1990s. Ball does not yet know how many jobs CASS will lose, or which facilities will see cuts. "My preference is to offer a reduced range of world-class services," he says, rather than close facilities.

The radio astronomy cuts will not affect space tracking tasks for NASA's interplanetary missions, which are conducted at the Canberra Deep Space Communication Complex. And Ball says CSIRO remains "genuinely committed to radio astronomy," particularly to two major radio telescope projects: the $2.5 billion international Square Kilometre Array (SKA) and the Australian Square Kilometre Array Pathfinder.

Construction of SKA, the world's largest radio telescope, is set to begin in 2018 in South Africa and Western Australia. But the new budget commits no funding to the SKA beyond 2016 to 2017. That has raised suspicions about the government's intentions. "Where is ... Australia's share of the startup cost?" asks Kim Carr of Australia's opposition Labor Party. As science minister in the previous government, Carr backed Australia's SKA bid.

Astronomy isn't the only loser. CSIRO cuts will also hit neuroscience, clinical medicine, and energy. Many who lose jobs will head overseas, further eroding Australia's intellectual infrastructure, predicts air quality researcher Michael Borgas, president of CSIRO's Staff Association. "This represents a fundamental lack of governmental understanding of what scientists want to do for the nation," he says.



The Parkes Observatory could feel some pain.

PHOTO: DAVID MCCLENAGHAN/CSIRO

CSIRO's Clark warns that the rollback could put Australia at a competitive disadvantage at a time when other nations are investing in science. Cut too deep, she warns, and "it will take us decades to recover." Ian Chubb, Australia's chief scientist, fears that "it is a distinct possibility that we'll lose our position in international science." The budget squeeze, he says, is a result of "the fact there's no strategic support for science."

Ian Macfarlane, the government's minister for industry, disagrees. The new budget, he notes, includes new funds for biomedical science, a new marine research vessel, and the nation's nuclear research reactor. The government, he says, "is investing strategically in [science] projects that will have long-term, ongoing benefits to the nation." ■

Published by AAAS

FEATURE

A lonely crusade

For years, bioethicist Carl Elliott has insisted that his university failed a young man who died by suicide during a clinical trial

By **Jennifer Couzin-Frankel**

It was an evening in May 2008 when Carl Elliott's life bifurcated into before and after. In an apartment courtyard in a suburb of Cape Town, South Africa, where he and his family were vacationing, Elliott walked in circles, holding his computer high above his head. He was searching for a wireless signal that would transmit a story from the *St. Paul Pioneer Press* back home in Minnesota. He finally managed to access it, and it began without preamble: "Subject 13 was dead."

Sitting outside alone in the dark, Elliott read on. Dan Markingson was 27 years old when he died by suicide, in the bathroom of a halfway house south of St. Paul. Elliott, a tenured bioethicist at the University of Minnesota (UMN), Twin Cities, had never heard of Markingson. But he learned that his office at the university's Center for Bioethics was a 20-minute walk from where Markingson had been hospitalized, across the Mississippi River at the university's medical center, in a unit reserved for patients with psychosis. There, Markingson had signed on to a clinical trial of antipsychotic drugs.

Unwelcome in his department, Carl Elliott works by the window of a local coffee shop.

The *Pioneer Press* alleged a tangled web of conflicts of interest, lack of oversight, a mother begging to withdraw her son from the trial, and questions about whether he was capable of consenting to take part in the first place. Markingson's suicide in May 2004 hadn't made the news. A few years later, following a tip, two reporters began digging and later published their exposé.

From the institution's perspective, there's no evidence that Markingson died because

PHOTO: BEN GARVIN

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

of the clinical trial. An investigation by the U.S. Food and Drug Administration (FDA) absolved the school of any wrongdoing, and a medical board cleared its psychiatrists. The death was tragic, university officials say now, but it's time to let go.

Letting go is not one of Elliott's strengths. Six years after first reading about Dan Markingson, Elliott's beard is fading from light brown to gray, and he has bags under his eyes. His crusade that began that spring continues unabated. Convinced that his university is withholding secrets, Elliott lobbies tirelessly for a thorough investigation that he says hasn't yet happened—to explore whether Markingson was inappropriately recruited into and kept in the drug study, and to determine if there are other stories like his in the university's Department of Psychiatry.

Administrators accuse him of distorting facts to feed an unquenchable agenda. Bioethics colleagues who once rallied around him have turned away, and he rarely steps foot in his own department. His family is resigned to a quest for justice whose end seems always just out of reach. "Wouldn't you like to work at a place that doesn't hate you?" his teenage son asked him recently.

But Elliott will not budge. Seeking justice for Markingson, he also wants to turn a spotlight on psychiatric drug trials at his institution. "I don't think either of us think it's just about the Dan Markingson case from 10 years ago, that's for sure," says bioethicist Leigh Turner, Elliott's only ally in his department. With bulldog tenacity, Elliott is recruiting supporters, a roster that now includes local medical school students and a former governor of Minnesota. More than 170 scholars of law, ethics, and medicine signed a letter last October pressing for the university to investigate Markingson's last months. Some days, resolution feels tantalizingly close. But, Elliott says, "I have no idea" if or when it will come.

PHOTO: BEN GARVIN

RAISED IN A SMALL TOWN in South Carolina, population about 3500, Elliott was the eldest of three boys. "There were a few choices you had—you could become a teacher, a doctor, a minister," says his middle brother Hal. "I think all of us had this idea, our father was a physician, that's what we would do."

Medical school proved disastrous for Elliott. "Wow, this is like being in the military," he remembers thinking. Senior physicians screamed at their students and threw instruments at them in the operating room. The rigid chain of command came as a shock, after a childhood with a father whose solo practice meant no boss to answer to. Elliott knew "about an hour in," he says wryly, that the life of a doctor wasn't for him. He stuck it out, hoping better days were just around the corner.

They weren't. After graduation, he promptly relocated to Glasgow, U.K. There, he began a Ph.D. program in philosophy. "Everybody thought it was insane, including my philosophy supervisor," he says. A medical school professor suggested he see a therapist.

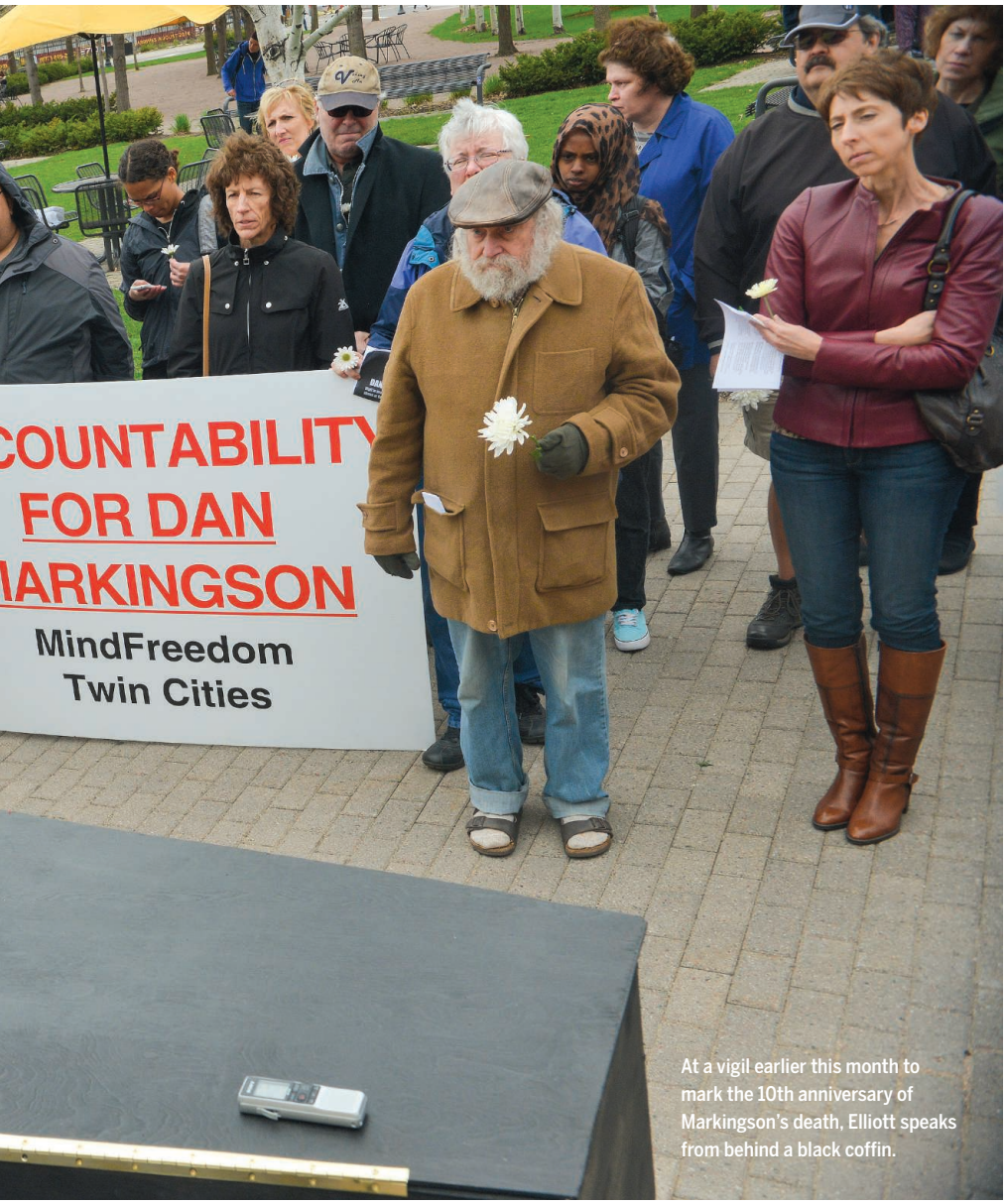
Philosophy suited Elliott. "Hypocrisy is something that is incredibly annoying" to us, Hal says of himself, Carl, and their youngest brother Britt, who works for the Canadian government in Ottawa. (Hal is a psychiatrist at East Tennessee State University, Johnson City.) Even more than his brothers, Carl is "just dogged," Hal says. "If he knows he's right he will just stay there."

In 1990, 3 years after finishing medical school, Elliott completed his dissertation, *Moral Responsibility and Mental Disorders*. It examined under what circumstances mentally ill individuals could be held responsible for crimes they had committed. Elliott's scholarship evolved, but certain themes endured and flourished: flaws in the practice of medicine, vulnerable research subjects, and the influence of the pharmaceutical industry on clinical trials—all of which would be braided together in the Markingson case.

Elliott is drawn to questions of trust, deception, and responsibility in medicine. "Often that takes you into odd, interesting corners," he says.

When the Markingson story broke in 2008, Elliott was on sabbatical. The family's temporary home was decorated with

Published by AAAS



At a vigil earlier this month to mark the 10th anniversary of Markingson's death, Elliott speaks from behind a black coffin.

pictures of his current obsession, a prominent New Zealand psychiatrist convicted of murdering his wife. Elliott was fascinated by how easily the psychiatrist had duped colleagues and continued treating patients; his account of the story, titled *Mind Game*, would later appear in *The New Yorker*. But Markingson soon supplanted the New Zealand psychiatrist in Elliott's mind. The more he learned, the more uneasy he became.

In 2002, the drug giant AstraZeneca launched a national trial, called the CAFE study, to compare three antipsychotic drugs that were already on the market. One of them was the company's product, Seroquel. The goal of the blinded trial was to determine whether Seroquel was as effective as its rivals. The company asked physicians to recruit 400 patients in the midst of their first psychotic episode, each of whom would be randomly assigned to take one of the three drugs for a year. One recruitment site was UMN.

Dan Markingson was involuntarily committed to the UMN Medical Center, Fairview, in November 2003 and subsequently diagnosed with schizophrenia. Florid delusions had left physicians fearful he might be a danger to himself or others. A local court agreed.

Three days later, the court stayed Markingson's commitment order, after those caring for him suggested he was improving. One condition imposed by the court was that Markingson follow the recommendations of his treatment team. The psychiatrist overseeing his care, Stephen Olson, was also leading the CAFE study's UMN outpost and receiving AstraZeneca money to do so. The next day, Markingson agreed to enroll in the trial. The university received about $15,000 for each subject.

A couple of weeks later, Markingson was discharged to a halfway house. His mother, Mary Weiss, worried that he was in no state to give informed consent for the study and feared the mystery drug was not helping him.

In a wrongful death lawsuit filed in 2007, Weiss's attorneys alleged that "Ms. Weiss sent five letters" to Olson and Charles Schulz, a UMN psychiatrist who was leading the study with Olson. Weiss was increasingly desperate that "her son was not improving and had begun to deteriorate," according to the lawsuit. She received a reply only to her fifth letter, sent via certified mail. In that response, less than 2 weeks before Markingson's death, Schulz wrote, "antipsychotic medications do not always lead to a complete remission of symptoms. … Further improvement may be seen over time, as has been shown in most studies."

A voicemail Weiss left at about the same time with the study coordinator, a social worker named Jean Kenney, was chillingly prescient: "Do we have to wait until he kills himself or someone else before anyone does anything?" When Kenney was deposed during the lawsuit, she acknowledged that after contacting the halfway house and being assured Markingson was "fine," she did not pursue the matter.

Olson and Schulz declined to comment for this story. They remain on the faculty of the university's Department of Psychiatry. Kenney left the university in 2005.

The parents of young adult psychiatric patients can present a dilemma to their doctors, says Paul Appelbaum, a psychiatrist at Columbia University. (Appelbaum wrote a declaration supporting the university's Institutional Review Board [IRB] after the lawsuit was filed, but was not otherwise involved.) By law, parents have no right to participate in decisions about treatment or research studies. But "although we can't talk to family members if patients don't want us to, we can always listen," Appelbaum says. Parents may provide valuable insights, about their child's talk of suicide, for example, or about illicit drugs they may be taking.

After almost 6 months in the CAFE study, Markingson stabbed himself to death. Another 280 people did not complete the trial. Volunteers abandoned therapy due to side effects, inadequate efficacy, or just because they wanted to. The results on the remaining 119 were published in *The American Journal of Psychiatry*. Because the proportion who dropped out from each group was roughly the same, study leaders concluded that Seroquel was, as AstraZeneca had hoped, as good as the other two treatments. All three drugs remain popular options today.

Published by AAAS

Markingson's suicide was reported to the university's IRB and to FDA. In January 2005, an FDA inspector spent 8 days at the university and produced a 21-page report. "No evidence of misconduct or significant violation of the protocol or regulations was found," wrote inspector Sharon Matson. "There was nothing different about this subject than others enrolled to indicate he couldn't provide voluntary, informed consent."

In some respects, the CAFE trial is troubling for how ordinary it was, say psychiatrists with no connection to it. It is not unusual for a physician to play dual roles as treating psychiatrist and trial leader and for the study coordinators who enroll patients to also assess their ability to consent. Drug companies commonly pay universities for the volunteers they recruit and retain. These conflicts, routine as they are, "create risk, both for the individual but also for the research," says Steven Hyman, former director of the National Institute of Mental Health and now head of the Stanley Center for Psychiatric Research at the Broad Institute in Cambridge, Massachusetts. He spoke generally because he's not familiar with Markingson's story.

It's critical for psychotic patients to participate in research—and studies have shown that about half of those who are inpatient can reasonably consent. But patients who have been committed involuntarily are akin to other extremely vulnerable research subjects, such as prisoners, Hyman believes. These people "need an absolutely trusting relationship with their treater," Hyman says. "Anything that would undercut their trust, that would lead them to the sense that they are being instrumentalized, is problematic."

In a deposition for Weiss's lawsuit, Olson was asked: "For Dan by court order, you control his freedom, isn't that right?" Olson replied simply, "Yes." But he also stood by the care he had provided. "I don't believe," he said, "that I abused my clinical relationship with Dan."

ELLIOTT ADMITS to an overdeveloped sense of shame, perhaps due to his Southern roots. "Most of us," he believes, feel shame "for things you didn't do yourself, things that were done by your family, your people, your country." Or your university. So Elliott reached out to the person suffering most deeply in the aftermath, Markingson's mother, Mary Weiss. Their friendship has fueled his devotion to the case. Now in her early 70s, Weiss suffered a series of strokes over the past 2 years and is largely confined to her bed. She was not up to speaking for this story.

> ***"Do we have to wait until he kills himself or someone else before anyone does anything?"***
>
> **Mary Weiss**

Mary Weiss's grief over the loss of her son, Dan Markingson, fuels Elliott's determination. "It seems kind of pathetic to complain" about what he's up against, he says. "I haven't lost a child."

A diminutive woman with white hair, Weiss suggested that Elliott review the evidence. She gave her lawyer permission to pass along hundreds of pages of court depositions and her son's medical records. The documents, Elliott felt, suggested that Markingson was profoundly out of touch with reality and had little sense he was gravely mentally ill. How, Elliott wondered, could someone who lacked insight into his illness consent to a study designed to treat it?

Elliott came to believe that every investigation—not only by FDA but also by the Minnesota Board of Medical Practice, the university's IRB, and its general counsel's office—had been flawed or incomplete. FDA did not seek Weiss's perspective, the views of Markingson's caseworker, or interview staff at the halfway house who had interacted with Markingson, for instance. (FDA would not comment on the Markingson case for this story.) Nor did the agency examine conflicts of interest. Weiss's lawsuit was dismissed not on its merits, but because the university's IRB and Board of Regents were deemed immune from liability thanks their role as state employees. (The judge did argue that informed consent was obtained appropriately, because Markingson had signed the consent form and had not been declared mentally incompetent by a court.)

As Elliott combed the documents, he saw a patchwork system, with no one agency tasked with examining it all. The university could make that happen. But universities "have powerful disincentives even to having an investigation," Hyman says. There's "a psychological sense of being beleaguered and circling the wagons," whether or not there's anything to hide.

ELLIOTT'S FIRST SUPPORTER was Turner, a bioethicist who joined the department from McGill University in Montreal, Canada, 2 months after the *Pioneer Press* series appeared. Turner considered Elliott honest to a fault—the type who wouldn't sugarcoat his assessment of your paper, no matter what he thought. He was also swayed by a lengthy feature article Elliott published in 2010 in *Mother Jones* about the Markingson case.

After the two began speaking out, writing editorials in local papers and later contacting politicians and university administrators and posting on a blog, Elliott and Turner heard from other individuals who insisted that they had been harmed in UMN psychiatric drug trials or had witnessed others' mistreatment. One man said he had worked in the psychiatric units of the hospital where Markingson was treated. Another identified herself as a counselor for teenagers. Elliott heard from parents, who said their son or daughter had enrolled in a study under pressure.

These tipsters would not allow their identities to be publicized. Still, Elliott and Turner see no reason to think they're lying and believe their accounts suggest systemic problems with such trials. "A pattern is a much different story, a much more unnerving story, than one young man who happened to die while he was enrolled in a clinical trial," Turner says.

PHOTO: COURTESY OF MARY WEISS

Published by AAAS

At first, Elliott's department rallied around him. In November 2010, eight faculty members, including Elliott and Turner, wrote a letter to the university's Board of Regents, requesting an independent, university-commissioned investigation into the Markingson case. But as Elliott continued lobbying for action, his colleagues withdrew, first, say Elliott and Turner, into silence, and then increasingly into rancor. One yelled at both to resign. Another filed a complaint against Elliott in late 2012, intimating gender discrimination and bullying. The complaint was subsequently dismissed. "It was sort of a shock, betrayal," Elliott says now. "The mere fact of going to work—you are so despised by the people around you."

All five tenured members of the department declined or did not respond to requests to comment for this story, as did two others who moved elsewhere in recent years.

UMN administrators maintain that the Markingson case has been thoroughly explored. Their frustration with Elliott's tenacity is evident. "Calls for an 'independent investigation' of Mr. Markingson's death intentionally ignore the multiple investigations that have already occurred over the past decade," wrote Brian Lucas, senior communications director, in a statement to *Science*. "Mr. Elliott has disregarded all of the findings. … We have little reason to believe Mr. Elliott would not find similar flaws in any additional investigation. … In the ten years that have followed Mr. Markingson's death, Mr. Elliott has consistently ignored evidence and distorted or omitted key facts in pursuing his own agenda." In an earlier e-mail, Lucas noted: "If it wasn't for Carl there wouldn't be an issue."

But that storyline is becoming difficult to sustain as more supporters join him. "I thought, 'We should not just sit here on the sidelines,' " says Trudo Lemmens, a professor of health law and bioethics at the University of Toronto in Canada who has known Elliott for years. He admits his friend has a "sharp pen, a characteristic that may have gotten under the skin of some people."

Lemmens, who came to believe no one had thoroughly probed a case he found deeply disturbing, soon pulled together more than 170 signatories on a letter to the university's president and Faculty Senate, urging an independent investigation. Many of the signers were big names: Jerome Kassirer, a former editor-in-chief of *The New England Journal of Medicine*; Susan Reverby, who uncovered evidence of mistreatment of Guatemalans by U.S. researchers in the 1940s; and Peter Gøtzsche, who directs the Nordic Cochrane Centre in Denmark that focuses on integrity in clinical trials. Whether Markingson killed himself because of the trial is not necessarily the core issue, says Lois Shepherd, a law professor and bioethicist at the University of Virginia in Charlottesville, and one of the signatories. "The issue is, why was he involved in this research and how did the system and the people in the system fail him? And I think it was on multiple levels."

Elliott leaves the building that houses the bioethics department he rarely visits. Carl is "just dogged," says his younger brother.

PHOTO: BEN GARVIN

Another unlikely ally is former Minnesota Governor Arne Carlson. "We're in the midst of a massive cover-up," Carlson declared in a phone interview. The university hired Elliott because it "found him to be one of America's most outstanding bioethicists. The moment he comes up with something that is sensitive to them, he becomes the village idiot." Carlson is lobbying the current governor of Minnesota to step in.

The pressure has had some effect. This past December, in a crowded windowless auditorium, UMN's Faculty Senate voted 67 to 23 to request that an independent panel inquire into "current policies, practices, and oversight of clinical research on human subjects at the University, in particular clinical research involving adult participants with diminished functional abilities." Administrators have solicited bids from contractors for a review. "We want this to be a transparent, open, fair assessment," says Brian Herman, the university's vice president for research. Herman declined to speak to *Science* about past trials or anything pertaining to Elliott and the Markingson case.

The resolution does not call for reexamining the Markingson story—something the Faculty Senate was considered unlikely to support—and Elliott feels little sense of triumph. He has long since stopped trusting his employer. He spends hours in local coffee shops filing open records requests for documents. He has posted more than 200 online. Ever suspicious that Markingson's story isn't unique, Elliott asked the university how many suicides or serious adverse events had occurred in 54 psychiatric clinical trials in the last decade. Gathering those data would cost $9600, the university told him. After Elliott protested that state law allows for visual inspection of documents like these free of charge, the school pledged to begin releasing them. So far, it has shared information from three. A statement from the university's Academic Health Center argued that "such a project would require considerable resources, take months of time, and ultimately be unproductive."

This month marks the 10th anniversary of Markingson's death. On 9 May, medical students, Elliott, Turner, and others—about 50 in all—held a vigil on the concrete plaza outside the university's McNamara Alumni Center, where the Board of Regents was gathered for a routine meeting. Four students brought a plywood coffin that Elliott had hammered together in his backyard up to the sixth floor. "We laid it out directly outside of the Board of Regents room, with the lid open," says Eden Almasude, a 21-year-old first-year medical student.

Almasude had met Elliott a few months earlier, and his passion resonated with her. "You don't find professors, faculty, who are so devoted to a case and so devoted to ideas, these moral, ethical convictions for justice and accountability," she says. "They're willing to become persona non grata."

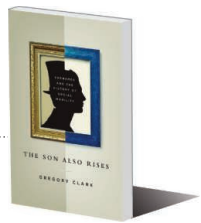
Outside, Almasude, Elliott, and the rest carried placards and called again on the university for an independent investigation into Markingson's death and to share the number of subjects who have died or been harmed in the university's psychiatric drug trials. "What are they hiding?" Elliott wants to know. For now, the coffin is back at his home, perched atop a foosball table in the living room. He's considering what to do with it next. ■

Published by AAAS

INSIGHTS

The story of a statistic, making sense of GDP *p. 811*

Surnames suggest success stems from parents *p. 812*



PERSPECTIVES

Transplanted tabletop corals (Acropora hyacinthus) on top of a Porites coral colony in Ofu lagoon, American Samoa.

PHOTO: STEPHEN R. PALUMBI/DEPARTMENT OF BIOLOGY, STANFORD UNIVERSITY

OCEANS

Lamarck was partially right—and that is good for corals

Tabletop corals can adapt to changing temperature conditions on shorter time scales than previously thought

By **C. Mark Eakin**

Ocean warming is one of the most urgent threats to coral reefs (*1*–*3*). Some taxa may migrate in response to changing environmental conditions (*4*), but corals and other sessile organisms only migrate through larval movement (*5*). This is viable for coral species with planktonic larvae, but not for the many coral species with crawl-away larvae that cannot migrate far. Adult corals must therefore adapt evolutionarily or acclimate physiologically to survive warming. On page 895 of this issue, Palumbi *et al.* (*6*) show that tabletop corals (see the first photo) can both acclimate and adapt to elevated temperatures in American Samoan back-reef pools (see the second photo), where high-temperature extremes are common. If the result holds for other species and locations, it provides hope for coral reefs under global warming.

Evolutionary processes, such as adaptation through acquisition of beneficial mutations and their spread through populations, are much slower than the climate warming of the past century, let alone warming expected in this century. This is especially true for long-lived organisms such as corals. Selection against poorly adapted phenotypes is rapid but results in loss of biodiversity and reduced population sizes. Acclimation is a much faster response to stress that involves physiological adjustment to environmental change. Such physiological plasticity is, however, often limited in its range of potential response. All these mechanisms can combine in unusual ways in corals, because corals are a symbiotic association between cnidarians and endosymbiotic algae in which adaptive and acclimative processes can act on either or both partners. Some recent studies have shown that corals can respond to increased temperatures over relatively short spatial (tens of meters) (*7*)

Published by AAAS

and temporal (decades) (*8*) scales. However, responses at the decadal scale will not be fast enough if climate change proceeds as projected.

So how can corals adapt quickly enough to keep up with climate change? A clue comes from Jean-Baptiste Lamarck's theory of the inheritance of acquired characteristics, published in 1809, the year Darwin was born. Lamarck argued that organisms develop new variations to adapt to their environments and then pass these acquired characteristics to their offspring. Although potentially much more rapid than Darwinian evolution, Lamarck's theories largely passed into disuse. However, recent discoveries have shown that organisms can acquire epigenetic traits, which are "stably heritable phenotypes resulting from changes in a chromosome without alterations in the DNA sequence" (*9*). Epigenetic changes can turn genes on or off or up- or down-regulate protein syntheses, often through changes in the shape of DNA molecules (*9*). Barshis *et al.* have recently shown that some corals can increase the expression of particular genes that help them to maintain physiological resilience during experimental warming; their findings are consistent with heritable traits but heritability has not yet been proven (*10*). Although this activation of previously dormant genes is not quite Lamarck's acquisition of novel characteristics, the thought that Lamarck was not entirely wrong provides hope for corals facing rapidly warming oceans along with other anthropogenic threats.

Palumbi *et al.* now provide evidence that tabletop corals can genetically adapt by changing their gene expression much more quickly than previously thought possible. Many past studies identified transient changes in coral or algal physiological processes or shuffling among the taxa of their endosymbiotic algae (*11*), but these were temporary changes that often reverted within a few months or years after the environmental stress had passed (*12*). The corals in the new study exhibited not only reversible physiological acclimation to high temperatures, but also fixed changes in gene expression. These fixed traits, many consistent with epigenetic up-regulation of acquired heat resistance, were maintained in corals translocated to pools with less extreme temperature variability. The authors estimated the rate required for these epigenetic shifts to be 15 to 24 months—much faster than the multigenerational rate required for evolutionary adaptation. The presence of these phenotypic changes suggests that these adaptations may be heritable, preadapting future generations to thermal stress.

PHOTO: DAN BARSHIS/OLD DOMINION UNIVERSITY

Hope for coral reefs. Palumbi *et al.* show that tabletop corals from back-reef pools in American Samoa (see photo) can adapt to changing thermal conditions on shorter time scales than previously thought. The results provide some hope for coral reef survival under climate change, although it remains to be shown whether corals from other locations also behave in this way.

Like most groundbreaking studies, this one leaves many questions unanswered. The authors did not determine whether the new phenotypic expressions are transgenerational. If the shifts prove to be heritable, some coral species may be better able to adapt than previously expected, and the adaptation may persist in the population. Also, the authors only tested one species of fast-growing coral, and it therefore remains unclear how many coral species have this capacity to adapt. Adaptability is likely to vary across coral taxa, resulting in winners and losers. Moreover, it is unknown whether such phenotypic shifts will only occur in corals found in highly variable conditions such as those found in the backreef pools studied by Palumbi *et al.*

These uncertainties aside, the study by Palumbi *et al.* provides hope to uniquely threatened ecosystems. Will enough coral species adapt sufficiently to maintain coral reefs that are fully functional and provide the key services upon which many species and human communities depend? To what extent can corals adapt through these mechanisms? If too few species are able to adapt sufficiently, coral reef ecosystems may look very different by the time humans deal with the causes of climate change. However, at a time when coral bleaching is increasing in frequency and intensity and there is the potential for another severe El Niño, any sign of greater adaptability of corals to warming oceans is welcome news. ■

Coral Reef Watch, National Oceanic and Atmospheric Administration (NOAA), Center for Satellite Applications and Research, College Park, MD 20740, USA. E-mail: mark.eakin@noaa.gov

REFERENCES AND NOTES

1. O. Hoegh-Guldberg *et al.*, *Science* **318**, 1737 (2007).
2. R. E. Dodge *et al.*, *Science* **322**, 189 (2008).
3. IPCC, *Summary for Policymakers*, in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. Stocker *et al.*, Eds. (Cambridge Univ. Press, Cambridge/New York, 2014).
4. R. M. Pateman, J. K. Hill, D. B. Roy, R. Fox, C. D. Thomas, *Science* **336**, 1028 (2012).
5. H. Yamano, K. Sugihara, K. Nomura, *Geophys. Res. Lett.* **38**, L04601 (2011).
6. S. R. Palumbi, D. J. Barshis, N. Traylor-Knowles, R. A. Bay, *Science* **344**, 895 (2014).
7. D. J. Barshis *et al.*, *Mol. Ecol.* **19**, 1705 (2010).
8. C. A. Logan, J. P. Dunne, C. M. Eakin, S. D. Donner, *Glob. Change Biol.* **20**, 125 (2014).
9. S. L. Berger, T. Kouzarides, R. Shiekhattar, A. Shilatifard, *Genes Dev.* **23**, 781 (2009).
10. D. J. Barshis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1387 (2013).
11. S. F. Gilbert, *in Transformations of Lamarckism: From Subtle Fluids to Molecular Biology*, S. Gissis, E. Jablonka, Eds. (MIT Press, Cambridge, MA, 2011), pp. 283–294.
12. T. C. LaJeunesse *et al.*, *Proc. Biol. Sci.* **277**, 2925 (2010).

ACKNOWLEDGMENTS

I thank my colleagues at NOAA's Coral Reef Watch for their comments. The opinions expressed in this manuscript are solely those of the authors and do not constitute a statement of policy, decision, or position on behalf of NOAA or the U.S. government.

10.1126/science.1254136

Published by AAAS

GEOPHYSICS

Deep mantle matters

Experiments reveal how some deep seismic anomalies near the core-mantle boundary might be generated

By **Quentin Williams**

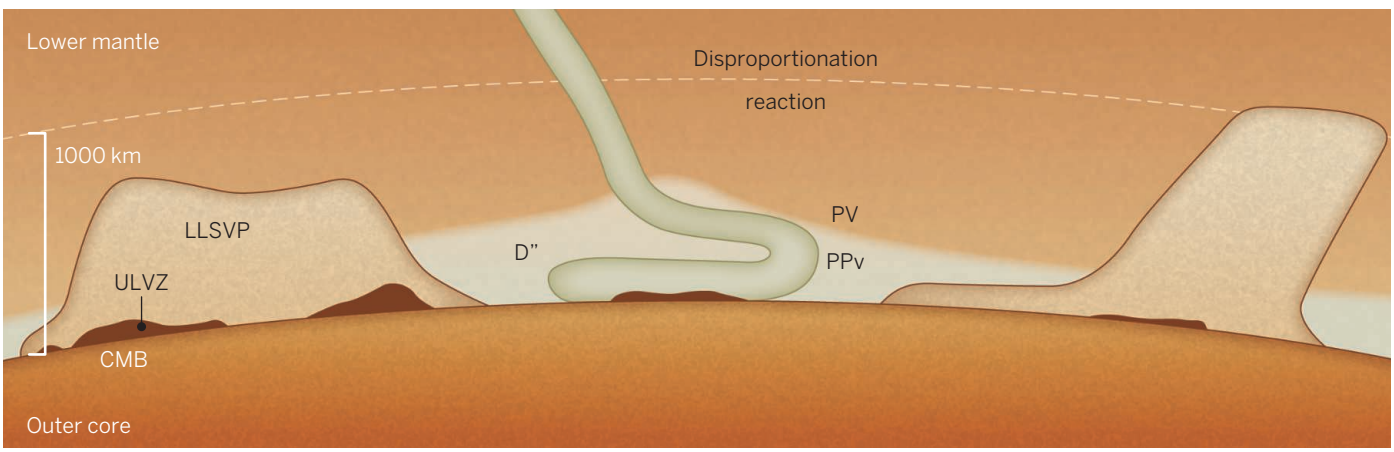
The lower mantle, lying between ~670- and ~2890-km depth, comprises most of the rocky portion of Earth. Convection processes within this region transfer heat from the iron core upward, advecting the heat flow that drives the near-surface tectonic engine. As many of the largest volcanic events appear to be correlated with seismic features in the deep mantle (*1*), the deep lower mantle may also represent an occasional scourge of our surface environment. In this issue, two sets of challenging experiments yield new pictures for how different deep seismic anomalies might be generated. On page 892, Andrault *et al.* (*2*) examine the melting temperature of oceanic crust (basalt) to core-mantle boundary (CMB) pressures and temperature, and use that to explain the genesis of areas with ultralow seismic velocities near the CMB. On page 877, Zhang *et al.* (*3*) report the startling discovery of a new, iron-rich silicate phase that may be a major component of the lowermost ~700 km of Earth's mantle.

Seismic probing has led to a richness of phenomena being recognized in the inaccessible lowermost ~1000 km of Earth's mantle (see the figure). Large, blocky structures beneath Africa and the Pacific with a ~3% decrease in shear velocity extend up to 1000 km above the CMB (large low-shear-velocity provinces, LLSVPs) (*4*). Regions with almost horizontal discontinuities in seismic velocity ~100 to 300 km are present above the CMB with velocity jumps of 2 to 3% (*5*) (these discontinuities define the anomalous D″ zone at the base of the mantle). There are also regions with large reductions in compressional and shear wave velocity in sporadic places in the lowermost ~5 to 40 km of the mantle, with shear velocity decreases as large as 45% (*6*) (ultralow-velocity zones, ULVZs). But all of these features are volumetrically small relative to the bulk of the mantle (LLSVPs are at most a few percent of the lower mantle volume, while ULVZs represent a small fraction of a percent). These features reside within a lower mantle that has long been recognized to be dominated by a simple mix of $(Mg,Fe)SiO_3$-perovskite and (Mg,Fe)O [and with possibly some exotic behavior associated with spin transitions in iron (*7*)].

Modeling and experiments at the extreme pressures and temperatures of the deep mantle (on the order of 100 GPa and 2000 K) reveal links between variations in phase, composition, and temperature and the observed seismic anomalies. One explanation for LLSVPs is that they are associated with slight iron and $(Mg,Fe)SiO_3$ enrichment (*8*), the seismic discontinuities with a pressure-induced transition from $(Mg,Fe)SiO_3$-perovskite to a post-perovskite structure (*9*), and the ULVZ with partial melting and a possible increase in iron content (*10*). But discerning how these anomalies are generated is difficult because our direct information is confined to a seismic snapshot of the present-day structure of this region.

The new studies provide process-oriented constraints on the genesis of ULVZs and chemical variations within the lowermost mantle (*2*) and shift the view on the perceived mineralogic simplicity of much of the lower mantle (*3*). Andrault *et al.* study the melting behavior of basalt (ocean crust) to pressures of the CMB. Basalt is the lowest–melting temperature rock likely to be present near the CMB, and is expected to melt when exposed to the temperatures of Earth's outer core (*11*). The picture that emerges is that subducted plates may carry oceanic crust to depths near the base of the mantle, where it can partially melt, forming ULVZs. But Andrault *et al.* propose that the melt can only be retained within the unmelted residue of the original basalt. Once it migrates into normal mantle, its chemistry is such that it will react and form solid magnesian perovskite, chemically remixing oceanic crust back into the mantle. As a result, ULVZs should be generated from, and associated with, subducted material at depth. Thus, the most abundant magma at Earth's surface may also produce most of the melt near the bottom of Earth's rocky mantle—and plate tectonics may play a fundamental role in the geochemistry of material juxtaposed with Earth's metallic core.

Zhang *et al.* observe the disproportionation of $(Mg_x,Fe_{1-x})SiO_3$-perovskite (where x within Earth's mantle is ~0.9) to nearly pure $MgSiO_3$ and a new, iron-enriched approximately $(Mg_{0.6}Fe_{0.4})SiO_3$ phase. This "H-phase" has a hexagonal structure and occurs at pressures corresponding to depths greater than ~2200 km. The synthesis of this phase occurred at high temperatures (2200 K and above), and this temperature and a low-



The mantle region. Schematic of seismically characterized features occurring in the lowermost ~1000 km of Earth's mantle, and experimentally proposed oceanic crustal-associated ULVZ (*2*) and perovskite disproportionation boundary (*3*). (Note that the ULVZ vertical scale is exaggerated for visibility.) PV, perovskite; PPv, post-perovskite.

ILLUSTRATION: ADAPTED FROM E. GARNERO

Department of Earth and Planetary Sciences, University of California Santa Cruz, Santa Cruz, CA 95064, USA. E-mail: qwilliam@ucsc.edu

Published by AAAS

stress environment appear to be critical for synthesizing this phase.

So why has this new transition not been observed seismically? One possibility is that the velocity change across the transition may be too small and/or the boundary may undulate dramatically in its depth. Alternatively, the temperature of the deep mantle may lie below the temperature of the disproportionation reaction [which would require that the mantle be a few hundred kelvin cooler than currently inferred (*12*)—but this would also imply that disproportionation could have been important in the hotter past]. Another option is that the oxidation state of iron in the mantle may differ from those within the experiments. The provocative aspects of this discovery include not just changing the possible mineralogy of the deeper lower mantle, but also that two phases of markedly different densities are produced. Whether these phases could undergo partial segregation, thus enriching or depleting regions in the H-phase (particularly in an earlier, hotter, less viscous, and possibly partially molten mantle), is unknown. If such segregation did occur, a natural explanation for the genesis of LLSVPs might exist. Depending on its elasticity, an enrichment of H-phase within these regions might provide an avenue to explain their anomalous seismic signature.

Each of these experiments is the direct result of developments in high-pressure, high-temperature techniques and the availability of high-intensity synchrotron sources. Probing the sensitivity of the pressure and temperature of melting and the phase transition to variable oxygen fugacities, shifts in major and minor elemental abundances, and volatile contents holds the prospect of mapping out the likely chemical behavior of the lower mantle. In doing so, the current void of information on the differentiation processes that govern the chemical variations, structural features, and evolution of Earth's deepest rocky reaches will be filled. ■

REFERENCES

1. J. Austermann *et al.*, *Geophys. J. Int.* **197**, 1 (2014).
2. D. Andrault *et al.*, *Science* **344**, 892 (2014).
3. L. Zhang *et al.*, *Science* **344**, 877 (2014).
4. S. Ni, D. V. Helmberger, J. Tromp, *Geophys. J. Int.* **161**, 283 (2005).
5. M. Wysession *et al.*, in *The Core-Mantle Boundary Region*, M. Gurnis, M. Wysession, E. Knittle, B. Buffett, Eds., Geodyn. Ser. (American Geophysical Union, Washington, DC, 1998), vol. **28**, 319–334.
6. M. S. Thorne, E. J. Garnero, G. Jahnke, H. Igel, A. K. McNamara, *Earth Planet. Sci. Lett.* **364**, 59 (2013).
7. J. F. Lin *et al.*, *Rev. Geophys.* **51**, 244 (2013).
8. F. Deschamps, L. Cobden, P. J. Tackley, *Earth Planet. Sci. Lett.* **349-350**, 198 (2012).
9. M. Murakami *et al.*, *Science* **304**, 855 (2004).
10. S. Rost *et al.*, *Nature* **435**, 666 (2005).
11. S. Anzellini *et al.*, *Science* **340**, 464 (2013).
12. T. Katsura, A. Yoneda, D. Yamazaki, T. Yoshino, E. Ito, *Phys. Earth Planet. Inter.* **183**, 212 (2010).

10.1126/science.1254399

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

CANCER IMMUNOLOGY

Identifying the infiltrators

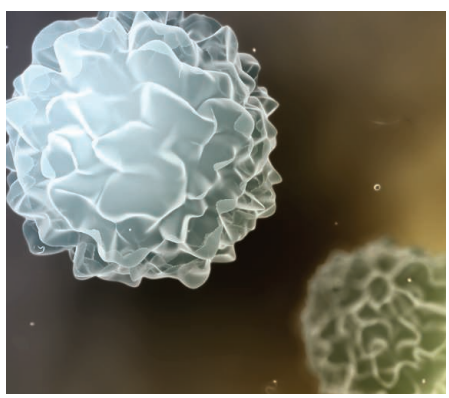
Molecular characterization of macrophages reveals distinct types during tumorigenesis

By **Elisa Gomez Perdiguero** *and* **Frederic Geissmann**

The mammalian immune system both suppresses and tolerates tumors, so understanding this complexity should benefit the development of cancer therapies. Macrophages are proposed to play an important role in suppressing the immune response to cancer cells, but it is not clear where these immune cells come from or whether there are distinct populations of macrophages with specific roles in this setting. On page 921 of this issue, Franklin *et al.* (*1*) forge a more coherent view of macrophages that are associated with tumor growth by assessing their origin, phenotype, and functions in an animal model of breast cancer.

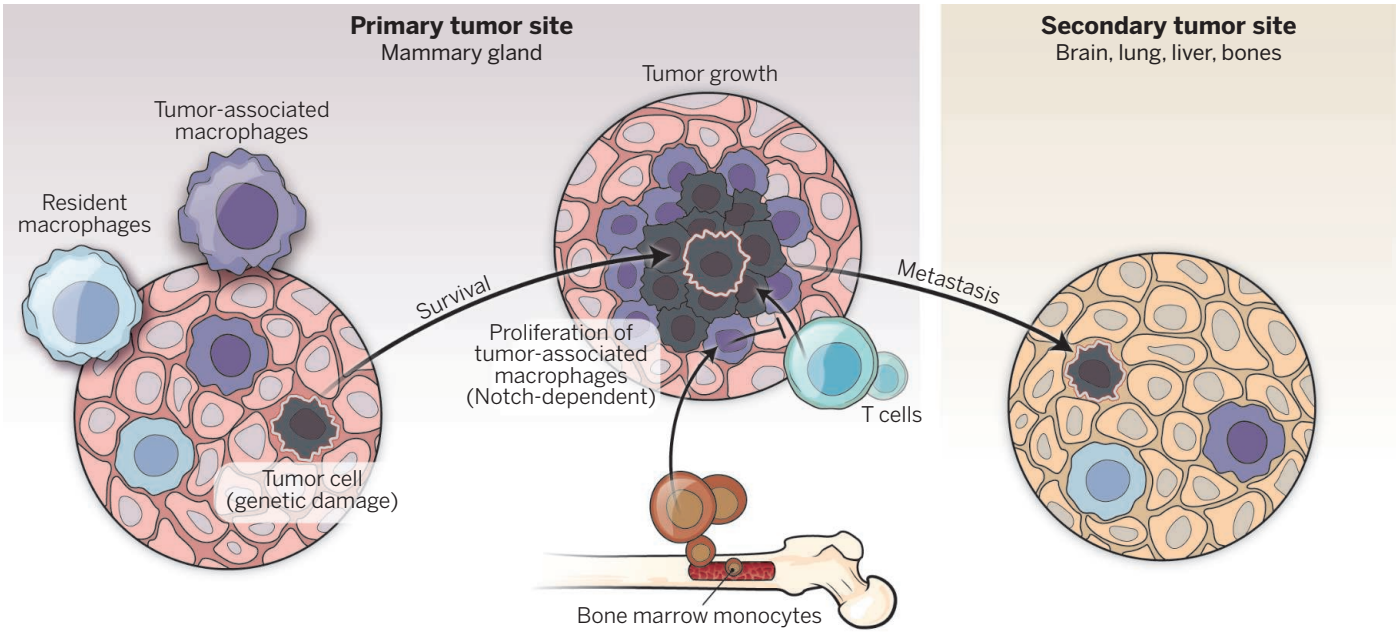
Tumor progression can be divided into three phases—initiation, growth, and metastasis (see the figure). The first phase is characterized by the cell-autonomous accumulation of genetic defects that leads to cell transformation. This is followed by clonal growth of transformed cells within the tissue—the primary tumor site (*2*). Metastasis results from the successful "engraftment" of circulating tumor cells into secondary locations where they proliferate after a dormancy phase in which metastatic cells remain quiescent (*3*). In both primary and secondary tumor sites, the stroma, which includes mesenchymal cells, macrophages, and extracellular matrix (*3*), is thought to play a role in the initial survival and proliferation of transformed cells. However, as a solid tumor grows and tumor cells acquire the potential to escape the primary site, the stroma becomes a more complex environment, with newly formed blood and lymphatic vessels and the recruitment and/or proliferation of lymphoid and myeloid immune cells (*4*). Immune cells are proposed to prevent tumor progression via the elimination of immunogenic tumor cells by T lymphocytes (CD8 subtype), a phenomenon known as immunosurveillance (also called immunoediting). During this process, tumors that display either reduced immunogenicity or enhanced immunosuppressive activity will escape elimination (*5*). Macrophages present in the tumor site can activate the immune response, but are mainly thought to contribute to immunosuppression and tumor progression (*6*, *7*), particularly in the mammary gland (*8*). A high density of macrophages in tumors is also associated with worse overall survival in patients with gastric, urogenital, and head and neck cancers, although it seems to be associated with better overall survival in patients with colorectal cancer (*7*).

Franklin *et al.* carefully explore the contribution of macrophages to tumor growth in mice that develop a mammary cancer



that is genetically driven by the expression of an oncogene. In investigating the development and differentiation of macrophages in the normal mammary gland and during the progression of a mammary tumor, the authors identify a population of macrophages that accumulates during tumor growth called tumor-associated macrophages (TAMs). These cells develop from bone marrow–derived cells with the characteristics of inflammatory monocytes, which are recruited to the tumor where they differentiate into macrophages and subsequently proliferate. Franklin *et al.* observed that when signaling by the protein Notch is prevented in these TAMs, their differentiation is blocked. Interestingly, TAMs are distinct from macrophages present in normal mammary tissue, which develop independently of Notch signaling. Depletion of TAMs led to a reduced tumor burden in the animal and increased the cytotoxic potential of T lymphocytes present in the primary tumor site. Thus, monocyte-derived Notch-dependent TAMs are critical for tumor growth in this mammary gland tumor model, at least in

Published by AAAS



Tumor-associated macrophages. Three stages of tumor progression are shown. Origin and role of TAMs in mammary gland tumors are depicted in the central area. The roles of tissue-resident macrophages and tumor-associated macrophages during initiation stage and the metastatic process are not yet known.

part through the inhibition of a lymphocyte-mediated immune response.

A current model proposes that the activation state of TAMs and their contribution to tumor growth depends on the stage of tumor development (*6*, *9*, *10*). Macrophages are activated by a range of factors including a variety of cytokines, producing different phenotypes. The classical phenotype is known as the "M1" macrophage, which is driven by interferon-γ (IFN-γ) and T helper cells (T_H1 subtype) and is associated with inflammation and tissue destruction. The alternative phenotype, called the "M2" macrophage, is driven by interleukin 4 (IL-4) and T_H2 cells and is associated with tissue repair and remodeling. To which category do TAMs belong? Early-stage tumor macrophage infiltrates have been reported with an M1 phenotype, which could result in tumor elimination (*9*). However, early-stage M1 macrophages also have been reported to favor tumor promotion (*9*). And in late-stage established tumors, an M2 phenotype has been observed, associated with T cell inhibition and tumor growth (*8*–*10*). Franklin *et al.* indicate that monocyte-derived Notch-dependent TAMs in a mammary tumor have neither an M1 nor an M2 gene expression profile and are independent of IL-4 signaling, although they promote tumor growth and inhibit a T cell immune response. This suggests that M1-M2 polarization of macrophages is not essential for TAM functions, at least in this model of mammary cancer.

Center for Molecular and Cellular Biology of Inflammation – CMCBI, Division of Immunology Infection & Inflammatory Diseases, King's College London, UK. E-mail: frederic.geissmann@kcl.ac.uk

The possible roles of Notch-dependent monocyte-derived TAMs in tumor initiation, angiogenesis, and the metastatic process were not investigated by Franklin *et al.* A role for monocytes in the metastasis of mammary tumors was proposed (*11*), and the blockade of delta-like 4 (a Notch ligand) was shown to inhibit tumor growth by promoting nonproductive angiogenesis (*12*). Therefore, it would be interesting to further investigate the roles of Notch-dependent monocyte-derived TAMs in these processes.

Franklin *et al.* report that mammary tissue macrophages (MTMs) present in the tissue in a steady state, and distinct from TAMs, also originate from monocytes but continuously renew in a Notch-independent manner. MTM numbers decrease during tumor progression. Their depletion did not impact tumor burden in the animal. It will be interesting to investigate whether this is also the case in nonmammary tumors or at the site of metastasis, in particular, in tissues such as the liver, epidermis, and brain where resident macrophages do not derive from monocytes. Many tissue-resident macrophages such as Kupffer cells in the liver, Langerhans cells in the epidermis, and microglia in the brain develop during embryogenesis and persist into adulthood independently from hematopoietic stem cells and monocytes (*13*). Activation of Kupffer cells by a cell surface protein called triggering receptor expressed on myeloid cells 1 (TREM-1) (ligand is unknown) was proposed to promote the development of hepatocellular carcinoma (*14*). Langerhans cells can metabolize chemicals into metabolites that induce DNA damage in epithelial cells and promote squamous cell carcinoma (*15*). By contrast, microglia were proposed to suppress brain tumor-initiating cells in the genesis or recurrence of gliomas (*16*). Resident macrophages thus could be involved in tumor initiation or modulate the ability of metastatic tumor cells to engraft or escape dormancy. Furthermore, it would be interesting to investigate whether resident macrophages contribute to tumor growth alongside monocyte-derived TAMs outside the mammary gland.

Tumor macrophages were long suspected to play important, but seemingly complex, roles in tumor progression. The characterization of monocyte-derived TAMs in the growth of mammary tumors is an important step toward a molecular characterization of the role of macrophages in cancer biology. This may open up opportunities to target TAM-specific properties to treat cancer. ■

REFERENCES

1. R. A. Franklin *et al.*, *Science* **344**, 921 (2014).
2. D. Hanahan, R. A. Weinberg, *Cell* **100**, 57 (2000).
3. T. Oskarsson *et al.*, *Cell Stem Cell* **14**, 306 (2014).
4. B. S. Wiseman, Z. Werb, *Science* **296**, 1046 (2002).
5. H. Matsushita *et al.*, *Nature* **482**, 400 (2012).
6. T. A. Wynn *et al.*, *Nature* **496**, 445 (2013).
7. Q. W. Zhang *et al.*, *PLOS ONE* **7**, e50946 (2012).
8. L. M. Coussens, J. W. Pollard, *Cold Spring Harb. Perspect. Biol.* **3**, a003285 (2011).
9. E. Van Overmeire *et al.*, *Front. Immunol.* **5**, 127 (2014).
10. A. Mantovani, A. Sica, *Curr. Opin. Immunol.* **22**, 231 (2010).
11. B. Z. Qian *et al.*, *Nature* **475**, 222 (2011).
12. I. Noguera-Troise *et al.*, *Nature* **444**, 1032 (2006).
13. C. Schulz *et al.*, *Science* **336**, 86 (2012).
14. J. Wu *et al.*, *Cancer Res.* **72**, 3977 (2012).
15. B. G. Modi *et al.*, *Science* **335**, 104 (2012).
16. S. Sarkar *et al.*, *Nat. Neurosci.* **17**, 46 (2014).

10.1126/science.1255117

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

Published by AAAS

ATMOSPHERIC SCIENCE

Record-breaking winters and global climate change

Rising greenhouse gas emissions may have played a role in the severe 2013-2014 winter in the U.S. Midwest

By **Tim Palmer**

Just when it looked like spring was arriving this year, the U.S. Midwest slipped back into winter, and Detroit recorded its snowiest season ever (see the photo). Has global warming gone into reverse, or could human emissions of greenhouse gases actually be responsible for this particular record being broken? Although the chances of cold winters can in general be expected to decrease with global warming, climate change linked to the particular circulation patterns that have prevailed in the past decade or so could have played an important role in this record-breaking winter.

Climate anomalies such as a colder-than-average winter or spring are commonly viewed as part of the natural variability of climate but will become less common as the concentration of greenhouse gases increases in the atmosphere. A simple analogy is to imagine a well-shuffled deck of cards. There is a 0.5 chance of drawing a black card. Now let us throw away five black cards, reshuffle the deck, and draw another card at random. The chance of drawing a black card is now 0.45. Now throw away another five black cards, and the chance is reduced to 0.38. For the purposes of this analogy, a black card represents a cold winter, the shuffling and random drawing is the chaotic variability of climate, and the process of throwing away the black cards mimics the effect of greenhouse gas emissions into the atmosphere. Just like the deck of cards, the chances of cold seasons will in general decrease. Indeed, observational studies show exactly this (*1*).

Given this overall decreasing tendency in cold winters, it seems impossible to argue that the record-breaking snowy winter in the Midwest could be connected to climate change. However, there is a plausible link. To understand this link we must consider the atmospheric circulation patterns that were associated with this winter and ask whether there is evidence that climate change might have increased the likelihood of these patterns. Large-scale atmospheric circulation patterns are controlled by the position of the jet stream. The Northern Hemisphere jet stream flows from west to east at mid-latitudes; it deviates from a line of latitude through a series of ripples called Rossby waves. Regions above which the jet stream is flowing from the north are likely to experience cold weather. Conversely, in regions above which it flows from the south, the weather is likely to be relatively warm. The larger the amplitude of the Rossby waves, the more anomalous the weather is likely to be at the surface.

PHOTO: JARRAD HENDERSON/MCT/NEWSCOM

Oxford Martin Programme on Modelling and Predicting Climate, University of Oxford, Oxford OX1 3BD, UK. E-mail: t.n.palmer@atm.ox.ac.uk

Why does the jet stream across the United States produce cold winters in the Midwest and on the east coast? A key factor is the strength of thunderstorm activity in the tropical West Pacific, associated with above-normal sea surface temperatures (SSTs) in this region (*2*). Latent heat release in this region, which occurs as warm moist air condenses into water droplets in these thunderstorms, acts as a source of energy to excite particularly large-amplitude Rossby waves in the Northern Hemisphere jet stream, with just the right phase to produce cold weather in the Midwest and East Coast. In the 2013–2014 winter, SSTs have been particularly warm in the tropical West Pacific (*3*).

Since the late 1990s, global mean temperatures have been rising quite slowly. This pause or "hiatus" in global warming is linked to an increase in the westward trade winds across the tropical Pacific (*4*). The ocean currents, forced by these intensified trade winds, have been drawing down much of the excess heat associated with human climate change to the deep ocean (*5*, *6*). However, one region that has not been cooling is the tropical West Pacific. Over the period of the hiatus, warm surface waters have been piled up in the tropical West Pacific by the intensified trade winds.

The surface waters in the tropical West Pacific will have been warmed further during this hiatus period through the local effects of man-made enhanced greenhouse gas forcing. Even though this enhanced warming may be small (it is not currently possible to estimate reliably its magnitude), its effect can be important in a region where SSTs are among the highest in the world. Any further warming in this region will lead to a relatively large increase in atmospheric water loading, leading to unusually strong latent heat release. Consistent with this, there was a very active typhoon season over the tropical West Pacific in 2013, including typhoon Haiyan, which devastated parts of the Philippines. These intense tropical weather systems continued into the 2013–14 winter season (*7*). Anomalous latent heat release in the tropical West Pacific can produce a particularly strong Rossby wave response in the Northern Hemisphere (*8*, *9*). The phase of this Rossby wave response is consistent with the cold and snowy season seen in the U.S. Midwest

Record snow. Although it seems counterintuitive, greenhouse gas emissions may have played a role in creating the conditions that led to the extremely icy and snowy 2013–2014 winter in Detroit (see the photo) and many parts of the eastern half of the United States. Seasonal forecasts suggest that the conditions that caused the record-breaking Detroit winter are less likely to occur in the coming winter.

Published by AAAS

and East Coast. If this line of argument is correct, the extremely cold and snowy season in parts of the United States may indeed have been caused at least in part by increased greenhouse gas concentrations.

The linkages described here are fundamentally different to another proposed mechanism, in which reduced Artic sea ice is hypothesized to influence jet stream variability (*10*). The latter hypothesis has been subject to considerable criticism (*11*). By contrast, the linkages proposed here are consistent with dynamical theory (*8*), and global climate models have been shown to be especially sensitive to small perturbations in the tropical West Pacific (*9*).

As this analysis indicates, it is simplistic to say that climate change makes the planet uniformly warmer. Earth's climate is a complex system, and its response to some external forcing will not be linear. Because of this complexity, sophisticated climate models are needed to test the correctness and robustness of climate mechanisms. Running these models is computationally expensive but crucial for advancing understanding of current and future climate.

Current seasonal forecast models suggest that a new (warm) phase of the El Niño/Southern Oscillation phenomenon may begin later this year, when the trade winds will finally weaken. If an El Niño event is on the way, the hiatus period may be coming to a close. If so, the upside is that the residents of the U.S. Midwest will be much less likely to have to suffer very cold winters. The downside is that global temperatures are likely to start to rise again, with many undesirable consequences for humans across the planet (*1*). The only way to reduce the risks associated with man-made climate change, in Detroit or elsewhere, is to cut greenhouse gas emissions. ■

REFERENCES

1. D. L. Hartmann, A. M. G. Klein Tank, M. Rusticucci, lead authors, in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, T. F. Stocker *et al.*, Eds. (Cambridge Univ. Press, Cambridge/New York, 2013), chap. 2.
2. T. N. Palmer, J. A. Owen, *Mon. Weather Rev.* **114**, 648 (1986).
3. Met Office, *The recent storms and floods in the UK*, The Met Office (2014); www.metoffice.gov.uk/media/pdf/1/2/Recent_Storms_Briefing_Final_SLR_20140211.pdf.
4. Y. Kosaka, S. P. Xie, *Nature* **501**, 403 (2013).
5. M. A. Balmaseda, K. E. Trenberth, E. Källén, *Geophys. Res. Lett.* **40**, 1754 (2013).
6. A. Clement, P. DiNezio, *Science* **343**, 976 (2014).
7. T. Hume, CNN, 22 January 2014; http://edition.cnn.com/2014/01/22/world/asia/philippines-agaton-flooding-landslides.
8. A. J. Simmons *et al.*, *J. Atmos. Sci.* **40**, 1363 (1983).
9. T. N. Palmer, D. A. Mansfield, *Nature* **310**, 483 (1984).
10. Q. Tang *et al.*, *Nat. Clim. Change* **4**, 45 (2014).
11. E. Kintisch, *Science* **344**, 250 (2014).

10.1126/science.1255147

MEDICINE

A unified cause for adrenal Cushing's syndrome

Mutations in a signaling protein underlie a class of functional adrenal tumors

By **Lawrence S. Kirschner**

In the cartoon adventure *Falling Hare*, aviator Bugs Bunny is vexed by a series of aerial mishaps, causing his airplane to crash. After spending the episode trying to understand the origin of his problem, Bugs hits on the obvious solution: a gremlin. Four papers—two on pages 913 and 917 in this issue of *Science* (*1*, *2*), a third in the *New England Journal of Medicine* (*3*), and a fourth in *Nature Genetics* (*4*), all from different groups—describe the use of whole exome sequencing (WES) to reach similar conclusions regarding the molecular origin of cortisol-producing adrenocortical adenomas. Just like Bugs' sudden recognition of the gremlin's underlying role in his predicament, the surprisingly consistent answer to the question of the underlying cause of these tumors is no longer obscure; rather, it is a mutation in the catalytic subunit of the signaling subunit of the cyclic-AMP (cAMP)–dependent protein kinase, protein kinase A (PKA).

Disease symptoms. A patient with typical features of Cushing's syndrome, including moon facies, abdominal obesity, striae, and peripheral muscle wasting. The CT image highlights (arrow) a typical adrenocortical adenoma. If this tumor secretes cortisol, there is an approximately 50% chance that it harbors a *PRKACA* L206R mutation.

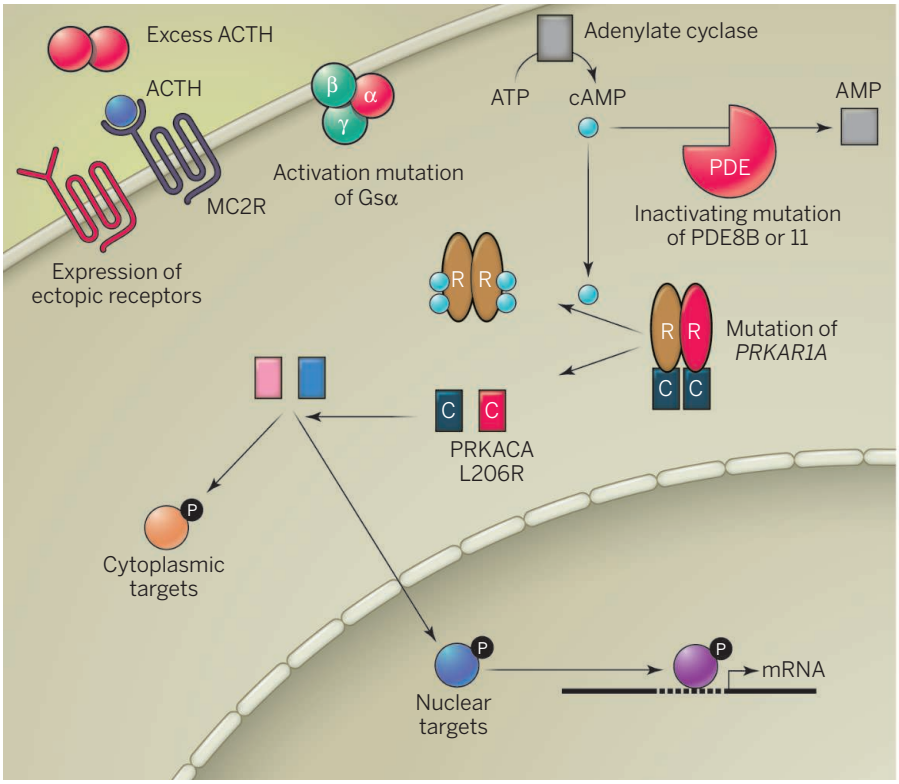
PHOTO: LAWRENCE S. KIRSCHNER/ILLUSTRATION: P. HUEY/*SCIENCE*

The four groups each studied benign cortisol-producing adrenocortical tumors. Like individuals receiving large doses of glucocorticoids for treatment of pulmonary or rheumatologic disorders, patients with endogenous hypercortisolism develop Cushing's syndrome, characterized in part by abdominal adiposity, purplish stretch marks (striae), and an increased risk for diabetes, hypertension, and osteoporosis (see the first figure).

The PKA signaling cascade is a central player in cellular signaling and is conserved in all eukaryotes. The PKA holoenzyme is a heterotetramer consisting of two catalytic (C) subunits and two regulatory (R) subunits. When intracellular cAMP levels rise, two molecules bind cooperatively to each regulatory subunit and release PKA-C, which is then able to phosphorylate targets in the cytoplasm and nucleus. In humans, there are three PKA catalytic subunit genes (*PRKACA*, *-CB*, and *-CG*), as well as a homologous gene on the X chromosome (*PRKX*). There are four regulatory subunits (PRKAR1A, -1B, -2A, and -2B) divided into class I and II based on their physicochemical properties. Of these components, *PRKACA* and *PRKAR1A* are expressed ubiquitously and exhibit the highest protein levels in most tissues; thus, they are thought to be the predominant players in mediating cAMP action. The cAMP/PKA axis is normally tightly regulated upstream at the level of cAMP generation and downstream by cAMP degradation.

The role of PKA signaling in mammalian physiology is complex, and the phenotypic outputs from pathway activation are highly context specific. In the normal adrenal cortex, activation of PKA has long been known to promote both cellular growth and hormone secretion. Pituitary ACTH (adrenocortico-

Published by AAAS



Sending a cellular signal. The cAMP/PKA signaling pathway governs adrenocortical cell physiology. Abnormalities of the cAMP/PKA pathway that have been associated with adrenal hyperproliferation and hyperfunction in human patients are shown in red.

trophic hormone) stimulates cortisol release from the adrenals through the melanocortin 2 receptor (MC2R), a classical Gs-coupled G protein–coupled receptor (GPCR). Patients with chronically elevated ACTH, such as results from ACTH-secreting tumors, develop Cushing's syndrome (hypercortisolism) and hyperplastic adrenals. Aberrant cAMP/PKA signaling in the adrenal is also associated with benign hyperfunctional adrenal neoplasias. Previously identified causes of enhanced PKA signaling in the adrenal cortex include the ectopic expression of GPCRs (*5*), activating mutations of Gs (McCune-Albright syndrome) (*6*), loss of PRKAR1A (Carney complex) (*7*), or phosphodiesterase mutation (*8*) (see the second figure).

Although the new work from the four groups included slightly different tumor spectra, the studies all focused on the analysis of cortisol-producing adrenal adenomas. In these tumors, a recurrent L206R mutation in PRKACA was found with an overall incidence of 48% (range 35 to 69%). Based on PKA's crystal structure, the L206R mutation is proposed to disrupt the interface between C and R subunits (*9*), and three of the four groups confirmed this biochemical effect (*1*, *3*, *4*). Each of the groups also detected PKA hyperactivation, confirming that loss of PKA-R binding enables unregulated PKA-C activity.

These four studies provide a new mechanism by which aberrant activation of PKA in the adrenal cortex causes excessive cortisol secretion and neoplasia, which is a biologically satisfying conclusion. These data contrast with the recent finding of inactivating *ARMC5* mutations in ACTH-independent macronodular adrenal hyperplasia (*10*), a condition with massive adrenal enlargement but inefficient cortisol overproduction. Because ARMC5's mechanism of action is unknown, its connection to PKA is uncertain. Loss of *ARMC5* favors growth over hormone secretion, whereas *PRKACA* mutations favor hormone secretion over growth, although both mutations eventually promote both effects. Clinically, the identification of *PRKACA* mutations is unlikely to alter patient management, because the decision for surgical removal of an adrenal tumor can be made solely based on the presence of hypercortisolism. Furthermore, because PKA signaling is universal in the body, agents that target PKA signaling are likely to have considerable toxicity. Although gene-based therapy might be effective, the cost and complexity of such treatment compared with surgery would likely render the procedure moot in most cases.

Now that *PRKACA* mutations have been detected in adrenocortical tumors, it may be reasonable to search for this mutation in other tumors, including benign thyroid adenomas and growth-hormone–producing pituitary tumors. These two tumor types are driven by Gs signaling, and mutations in the PKA pathway are already known in both (*6*). In terms of cancers that might harbor *PRKACA* mutations, follicular thyroid cancer and osteosarcomas are both promoted in response to PKA activation in mice, and a role for PKA has been confirmed in the corresponding human tumors (*11*, *12*). Thus, these two cancer types should be among the first studied. Although adrenal cancers secrete cortisol in over 50% of cases (*13*), malignant tumors often demonstrate downregulation of the hormone biosynthetic pathway (*14*), making *PRKACA* mutations seem less likely to be found in these cancers. Work ongoing as part of The Cancer Genome Atlas (TCGA) should shed light on this question in the near future.

Finally, another important point to be considered from these studies bears on the utility of WES in tumors. Next-generation sequencing of cancer DNA typically identifies a large number of somatic changes, such that identifying the initiating mutations may be difficult in the setting of a complex and heterogeneous mutational landscape (*15*). When benign tumors are studied, the number of mutations is much lower, and recurrent sequence alterations are much more likely to be related to tumor initiation.

All in all, WES of tumors continues to provide a wealth of information to clinicians and researchers regarding tumorigenesis, and endocrine tumor genetics has been a particularly fruitful area. Sometimes this information provides molecular insights into new pathways for investigation; other times, we are reminded that there are some processes that we really do understand well. ■

Division of Endocrinology, Diabetes, and Metabolism, Department of Internal Medicine and Department of Molecular Virology, Immunology, and Medical Genetics, The Ohio State University, Columbus, OH, USA. E-mail: lawrence.kirschner@osumc.edu

ILLUSTRATION: P. HUEY/*SCIENCE*

REFERENCES

1. Y. Sato *et al.*, *Science* **344**, 913 (2014).
2. Y. Cao *et al.*, *Science* **344**, 917 (2014).
3. F. Beuschlein *et al.*, *N. Engl. J. Med.* **370**, 1019 (2014).
4. G. Goh *et al.*, *Nat. Genet.*; 10.1038/ng.2956 (2014).
5. A. Lacroix *et al.*, *Endocr. Rev.* **22**, 75 (2001).
6. L. S. Weinstein *et al.*, *Endocr. Rev.* **22**, 675 (2001).
7. L. S. Kirschner *et al.*, *Nat. Genet.* **26**, 89 (2000).
8. M. F. Azevedo *et al.*, *Endocr. Rev.* **35**, 195 (2014).
9. C. Kim *et al.*, *Science* **307**, 690 (2005).
10. G. Assié *et al.*, *N. Engl. J. Med.* **369**, 2105 (2013).
11. D. R. Pringle *et al.*, *J. Clin. Endocrinol. Metab.* **99**, E804 (2014).
12. S. D. Molyneux *et al.*, *J. Clin. Invest.* **120**, 3310 (2010).
13. B. Allolio, M. Fassnacht, *J. Clin. Endocrinol. Metab.* **91**, 2027 (2006).
14. F. de Fraipont *et al.*, *J. Clin. Endocrinol. Metab.* **90**, 1819 (2005).
15. B. Vogelstein *et al.*, *Science* **339**, 1546 (2013).

10.1126/science.1254901

Published by AAAS

COSMOLOGY

Testing gauge/gravity duality on a quantum black hole

A numerical test shows that string theory can provide a self-consistent quantization of gravity

By **Juan Maldacena**

In general relativity, gravity is formulated as a classical physics theory, and formulating its fully quantum version is a great challenge. Gauge/gravity duality conjectures that certain special quantum theories (typically gauge theories) are equivalent to theories of gravity in an emergent spacetime that has more dimensions than the ones appearing in the original quantum theory, and thereby provides an important bridge between quantum theory and gravity. A simple example involves a matrix quantum mechanics theory (*1*, *2*) of interacting fundamental particles that gives rise to gravity in a 10-dimensional curved spacetime (*3*). Black holes in this spacetime are described by the quantum system at finite temperature. Testing the conjectured duality is difficult because the quantum system is very strongly coupled—it has no easy mathematical solution—when the standard Einstein-like gravity description is applicable. On page 882 of this issue, Hanada *et al.* (*4*) test this relation by numerically calculating the entropy of a black hole in the gauge theory. They also calculate this entropy with a gravity theory that includes the first quantum gravity correction. The agreement between the two computations is evidence for the validity of the gauge/gravity duality, as well as for the internal consistency of string theory as a quantum theory of gravity.

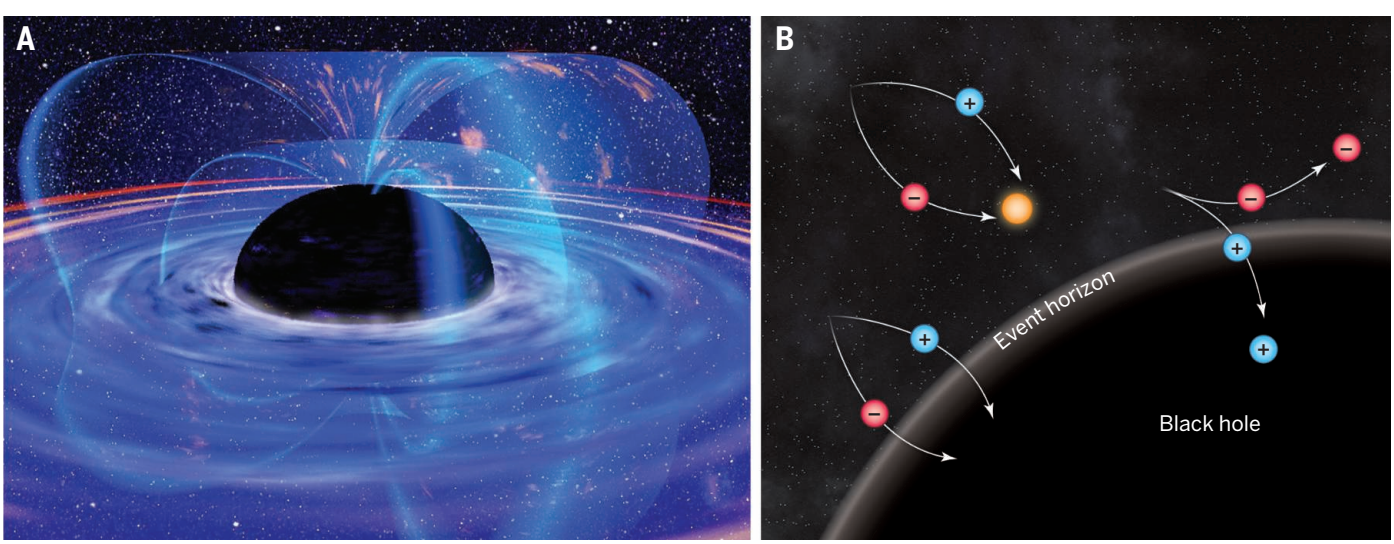
School of Natural Sciences, Institute for Advanced Study, Princeton, NJ 08540, USA. E-mail: malda@ias.edu

The quantum mechanical system used by Hanada *et al.* is conceptually simple. It can be viewed as a set of harmonic oscillators with additional anharmonic terms that are cubic or quartic in the coordinates of the harmonic oscillators. These oscillators are arranged into $N \times N$ matrices and have interactions that are invariant under the special unitary group of order N [SU(N)] symmetry. When N is very large, such theories are expected to have a convergent expansion in N. For example, their thermal free energies have a leading expansion term proportional to N^2 and subleading terms that are suppressed by powers of $1/N^2$.

The alternative description in terms of gravity theory (*3*) can be viewed as a kind of gravitational potential well with spherical symmetry. When N is large, the gravitational theory is weakly coupled and quantum gravity corrections are small. The G_{Newton} expansion in the gravity theory (which is the standard semiclassical expansion in gravity) corresponds to the $1/N^2$ expansion of the collection of interacting oscillators. The bulk theory—a 10-dimensional string theory—can in principle be quantized by using the rules of string theory to calculate the quantum gravity corrections. When the gravity theory is well approximated by an Einstein-type gravity theory (as opposed to a more complicated string theory), then the quantum mechanical system of oscillators is strongly coupled. Under these conditions, the anharmonic terms are very important and cannot be treated as small perturbations. The system of oscillators is providing a so-called "holographic" description of the physics inside the 10-dimensional spacetime.

The quantum mechanical system at finite temperature has a corresponding gravity description that involves a black hole. The entropy of the black hole can be computed with the area formula of Hawking and Bekenstein. In the bulk theory, the computation is relatively simple—it involves finding a black hole solution and computing the area of its horizon. In contrast, the corresponding computation for the strongly interacting oscillators is more difficult. However, it has been done numerically (*5*), and agreement was found for this leading contribution, which is the term in the entropy of order N^2. Hanada *et al.* now go further and compute the order unity (N^0) term that corresponds to a quantum gravity correction in the bulk theory.

In general, such quantum gravity corrections to the entropy also involve the entropy of Hawking radiation (see the figure). However, in this case there is a numerically



Not completely black holes. (**A**) The gravity pull of a black hole is so strong that no particles—not even photons—escape within the event horizon. (**B**) However, Hawking showed that the production of virtual particles in the vicinity of the event horizon, as demanded by quantum field theory, leads to radiation if one particle escapes and the other enters the black hole. Hanada *et al.* calculated quantum corrections to the entropy of certain black holes at finite temperature.

PHOTO: (PANEL A) NASA/ESA/XMM-NEWTON/SCIENCE SOURCE

Published by AAAS

more important correction that is related to the potential divergences in quantum gravity. In the 10-dimensional gravity theory, the first quantum correction would be divergent. In string theory, this divergence is removed by the presence of strings, but a finite local and calculable counterterm remains. This term gives a correction to the black hole entropy, and it is this particular correction that was matched by the numerical computation.

These quantum corrections had been previously computed in the context of extremal and near-extremal black holes [see, e.g., (*6*)]. These are charged black holes with the minimal or near-minimal mass, respectively, for a given charge. They are stable and do not emit Hawking radiation. In some cases, the previous computations even matched further subleading corrections. Hanada *et al.* can now match the corrections in a more generic finite-temperature situation. This calculation numerically tests quantum gravity in a context where string theory is crucial for obtaining the answer. In the usual quantization of gravity as an effective field theory, the correction that was considered would be proportional to an uncalculable counterterm. However, string theory provides a precise value for this counterterm (*7*) and is the one that reproduces a numerical computation with the interacting quantum oscillators.

In the near future, similar methods could also match the term that comes from the entropy of the Hawking radiation itself and eventually could match results for other observables in the thermal ensemble that probe more detailed aspects of the 10-dimensional geometry. This numerical test is further evidence of the internal consistency of string theory (i.e., that it does indeed provide a self-consistent quantization of gravity) and provides further evidence that the gauge/gravity duality is correct. Of course, the 10-dimensional space under consideration here is not the same as the four-dimensional region of the multiverse where we live. However, one could expect that such holographic descriptions might also be possible for a region like ours. ■

REFERENCES

1. J. Polchinski, *Phys. Rev. Lett.* **75**, 4724 (1995).
2. T. Banks, W. Fischler, S. H. Shenker, L. Susskind, *Phys. Rev. D* **55**, 5112 (1997).
3. N. Itzhaki, J. M. Maldacena, J. Sonnenschein, S. Yankielowicz, *Phys. Rev. D* **58**, 046004 (1998).
4. M. Hanada, Y. Hyakutake, G. Ishiki, J. Nishimura, *Science* **344**, 882 (2014).
5. M. Hanada, Y. Hyakutake, J. Nishimura, S. Takeuchi, *Phys. Rev. Lett.* **102**, 191602 (2009).
6. A. Sen, http://arxiv.org/abs/1402.0109 (2014).
7. M. B. Green, M. Gutperle, P. Vanhove, *Phys. Lett. B* **409**, 177 (1997).

10.1126/science.1254597

ILLUSTRATION: C. BICKEL/*SCIENCE*

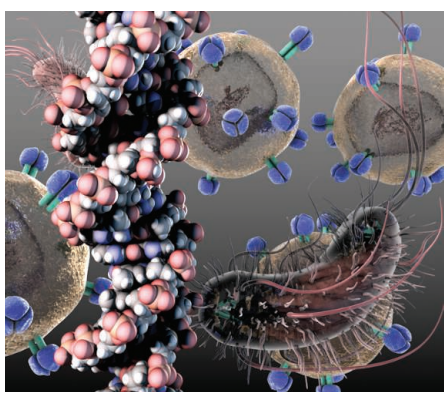
TRANSLATIONAL GENOMICS

Targeting the host immune response to fight infection

Strategies to modify immune responses to infection can be found in our genome

By **J. Kenneth Baillie**

Every year, infectious diseases kill millions of people worldwide. A close look at the modes of death from infection reveals something surprising: Often death is not attributable to a direct effect of the pathogen or of any toxin it produces; rather, it is the consequence of a systemic inflammatory response in the host (*1*). Our own immune system destroys us. This concept was observed long ago, but efforts to find effective therapies that alter the host response to infection to promote survival have largely failed.



Targeting the infectious agent is, at present, the only successful strategy, but the relentless emergence of antimicrobial resistance is an inevitable problem. Some pathogens, such as influenza virus, can evolve de novo resistance with terrifying speed (*2*). With hindsight, it is unsurprising that pharmacological interventions to alter the host response to infection have not been effective. Our immune system has evolved to fight a moving target. Whereas the job of the heart has changed little, and hemoglobin binds the same oxygen, and even the circuitry required to generate consciousness need not be different from that of our early ancestors, immunity must change rapidly, again and again, every time a new pathogen appears or an old pathogen mutates. By its very nature, the immune system is expected to be a mire of complexity, interdependence, and redundancy.

The core problem is finding the right element of an immune response to target with a drug. Even if we were to possess a full knowledge of every connection in an unimaginably vast network of interactions among cells, proteins, and nucleic acids that are required to generate an immune response, we would still be uncertain as to where to intervene to help patients survive infection. The good news is that evolution may have encoded within our genomes a shortcut to these targets.

"The core problem is finding the right element of an immune response to target with a drug ... When there is a way to confer a resistant state on a susceptible individual, the results have been impressive."

Susceptibility to infectious disease is one of the most strongly inherited of all common disease traits (*3*). Specific genetic variants confer susceptibility or resistance to infection. Knowledge of these variants presents us with a challenge: Here is the genetic code required for resistance; all we need to do is find a way to confer it upon susceptible individuals. Genetic susceptibility or resistance tells us the one thing we most want to know about any component of a complex system: What will be the effect of intervening here?

When there is a way to confer a resistant state on a susceptible individual, the results have been impressive. For example, a patient with HIV became resistant to the virus after a bone marrow transplant with cells from a donor whose cells were resistant to the virus (*4*). In this case, the patient

Roslin Institute, University of Edinburgh, Midlothian EH25 9RG, UK, and Intensive Care Unit, Royal Infirmary of Edinburgh, Edinburgh EH16 4SA, UK. E-mail: j.k.baillie@ed.ac.uk

Published by AAAS

Translational genomics. Surveying the genomic variations in patients who are susceptible or resistant to infections should narrow down to key components of a complex immune response, in effect giving us the answer to the critical question: What will be the effect of intervening here?

had also developed leukemia, which was treated by killing his bone marrow before reconstituting his immune system from hematopoietic stem cells bearing the donor's phenotype—HIV resistance. Such extreme measures are not always necessary. A mutation that predisposed patients to chronic infection with Epstein-Barr virus was recently discovered in a gene that encodes a magnesium transporter. Remarkably, treatment with magnesium seems to control the infection (*5*).

Recruiting the large numbers of patients required for genetic studies can be challenging, particularly in acute infections, which are by definition transient. But there are various methods to make the most of what we have. For example, studies of patients with extreme susceptibility to a disease, such as those who become critically ill, have greater statistical power to detect differences. Another consideration is to look in the right places in the genome. Disease variants are greatly enriched in protein-coding regions, and even more so in the regulatory sequences located outside of coding regions, where the activation state of a gene is controlled (*6*). Recent findings of the FANTOM (*7*) and ENCODE (*8*) projects have revealed where regulatory components lie in the human genome and in which conditions and cell types they are active. In some cases, inferences can be drawn from the pattern of gene expression across different cell types to better comprehend the roles of newly discovered genes. These patterns also provide new insights about the roles of the regulatory elements—more abundant than protein-coding genes by a factor of 10—that govern gene expression (*7*). These analyses suggest compelling candidate susceptibility genes.

Rather than studying the response of the entire host, a conceptually narrower approach is to study the host factors that restrict pathogen replication in cultured cells. The full complexity of the immune system is absent from these models, leaving two categories of host genes: innate intracellular defenses, such as the antiviral gene encoding interferon-induced transmembrane protein 3 (*IFITM3*) (*9*) or components of host cells that are required by the pathogen to survive and replicate. For example, viruses co-opt host proteins to perform essential functions in the viral life cycle. For specific pathogens, high-throughput studies that systematically silence large numbers of genes in vitro, one by one, can identify genes in both of these categories (*10*). This has been done with hepatitis C, for example, demonstrating that the host protein cyclophilin A is required for viral replication (*11*). This protein was already the target of a new therapy for hepatitis C, which is now in clinical trials.

These studies capitalize on a uniquely useful feature of intracellular infections—the ability of the pathogen to replicate in cultured cells. Replication has direct relevance to the clinical scenario at the bedside: If a patient's cells can be made less hospitable to the infectious agent, then it is at least plausible, if not likely, that an improvement in their clinical condition will follow.

If identifying a drug target is the goal, it is important to remember that it is easier to break a complex system than to improve it. The genetic causes of major immunodeficiency syndromes tend to be deficiencies in specific critical components of immunity (*12*). Replacing the function of these components, such as in patients lacking a key class of immune cells, is technically challenging and may not lead to improved resistance in individuals who already carry a functional copy of the gene. Patients born lacking the capacity to make adequate antibodies respond very well to antibody replacement (*13*), but (with a few exceptions) treatment with nonspecific antibodies is not effective in immune-competent patients with severe infections (*14*).

Mutations in less critical components of a system may flag better targets. One possibility is the IFITM3 gene, in which a variant associated with less IFITM3 protein production is more common in life-threatening influenza. In cell lines derived from people who possess the susceptible genetic variant, IFITM3 production is restored by treatment with interferon (*15*). Because it is generally easier to create drugs that inhibit proteins, gain-of-function mutations that cause disease are given priority. To efficiently find targets that meet these criteria will require integrating data from the sources such as those described above.

When a patient dies of an infection, we rarely pause to think that a weakness for that specific pathogen may have been present in that individual, right from the start, encoded in their genome. By interrogating these weaknesses, we are finding clues that may, in time, help us to save more people from the same fate. ■

REFERENCES

1. D. C. Angus, T. van der Poll, *N. Engl. J. Med.* **369**, 840 (2013).
2. N. Renzette *et al.*, *J. Virol.* **88**, 272 (2014).
3. T. I. Sørensen, G. G. Nielsen, P. K. Andersen, T. W. Teasdale, *N. Engl. J. Med.* **318**, 727 (1988).
4. G. Hütter *et al.*, *N. Engl. J. Med.* **360**, 692 (2009).
5. B. Chaigne-Delalande *et al.*, *Science* **341**, 186 (2013).
6. R. Andersson *et al.*, *Nature* **507**, 455 (2014).
7. FANTOM Consortium, *Nature* **507**, 462 (2014).
8. ENCODE Consortium, *Nature* **489**, 57 (2012).
9. A. L. Brass *et al.*, *Cell* **139**, 1243 (2009).
10. S. D. Shapira *et al.*, *Cell* **139**, 1255 (2009).
11. Q. Li *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16410 (2009).
12. A. Bousfiha *et al.*, *Clin. Immunol.* **135**, 204 (2010).
13. P. J. Busse, S. Razvi, C. Cunningham-Rundles, *J. Allergy Clin. Immunol.* **109**, 1001 (2002).
14. M. M. Alejandria, M. A. Lansang, L. F. Dans, J. B. Mantaring 3rd, *Cochrane Database Syst. Rev.* **9**, CD001090 (2013).
15. A. R. Everitt *et al.*, *Nature* **484**, 519 (2012).

10.1126/science.1255074

PHOTO: GETTY IMAGES/PHOTOS.COM

Published by AAAS

SOCIOLOGY OF SCIENCE

"Undemocracy": inequalities in science

Inequality, an intrinsic feature of science, has trended upward in recent years

By **Yu Xie**[1,2]

In recent years, academic scholarship and public discourse have become increasingly preoccupied with social and economic inequality, which has risen in many countries. It is surprising that more attention has not been paid to the large, changing inequalities in the world of scientific research. I suggest that although the basic structure of inequalities in science has remained unchanged, their intensities and mechanisms may have been altered by recent forces of globalization and internet technology.

POLICY

By "inequalities," I mean differences in three major domains: resources, research outcomes, and monetary or nonmonetary rewards. At times, my discussion is conceptual and somewhat speculative, as precise and meaningful measurement of these outcomes is difficult. Even though I refer primarily to basic, natural science, most of my general conclusions are also applicable to the social and applied sciences.

AN INTRINSIC PROPERTY OF SCIENCE. Derek Price observed in 1963 that inequality in science is inherently high and called it "undemocracy" (*1*), meaning that a great scientist's value for science far exceeds that of ordinary scientists. Scientific outputs and rewards are much more unequally distributed than other well-being outcomes, such as education, earnings, or health (*1*, *2*).

One source of inequality in science is what Robert Merton called the "Matthew effect" (*3*), referring to Matthew 25:29 in the Christian Bible: "For to all those who have, more will be given, and they will have an abundance; but from those who have nothing, even what they have will be taken away." This "rich get richer" effect means that eminent scientists receive disproportionately greater recognition and rewards than lesser-known scientists for comparable contributions. As a result, a talented few can parlay early successes into resources for future successes, accumulating advantages over time.

Although science rewards all participants through a skewed tier system with the most substantial rewards going to the top performers in a tournament-like reward system, science has attributes that resemble a "winner-takes-all" market: high visibility of top winners, a large contestant base, accumulation of advantages, absence of physical or cultural boundaries, and intense competition (*3*, *4*). Thus, many scientists feel that merely being good at their jobs is not enough. Competition is all about priority, a scientist's claim to be the first to make a big discovery (*5*, *6*).

Whatever their cause, high inequality in scientific rewards is often defended on two grounds. First, given the positive externalities of science (*7*), the more skewed rewards are, the greater the incentive for outstanding scientific work that will ultimately benefit all of humanity (*8*, *9*). Second, as a profession, science is supposed to practice what Merton called universalism (*10*), a norm that dictates that evaluation in science be based solely on merit rather than on functionally irrelevant factors such as gender, race, nationality, age, religion, and class (*2*, *10*). This merit-based system makes inequality seem fair and acceptable.

Before the 19th century, science was mainly a small-scale, personal pursuit enjoyed by a few leisure-class amateurs. Over the next two centuries, it expanded enormously into an institution characterized by certain distinctive features: a huge, well-paid, professional workforce; large-scale government and industrial support; reliance on the university as institution; graduate student labor; and a peer-review system of evaluation (*7*). Advances in Internet technology have facilitated the rapid, broad dissemination of research results (*11*).

Although these features have made scientific production faster and more voluminous, they have also rendered the evaluation of scientists less substance-specific and more "numbers-based." Scientists are increasingly likely to be judged by whatever numbers they can generate in terms of publications, citations, research grants, prestigious awards, research team size, and memberships in elite academies than by their actual scientific contributions (*7*). This tendency may have been amplified by increasing specialization, such that scientists in one specialty area find it difficult to understand content in another. University administrators, faced with uncertainties and competing demands for scarce resources, have strong incentives to use externally generated and validated indicators (*12*).

HAS INEQUALITY IN SCIENCE INCREASED? More empirical research is needed to answer this question, but I believe that two trends have caused inequality in science to rise over time. First, growth in high rewards in science has been limited, occurring much more slowly than the expansion of science itself. It is well known that the number of Nobel Prizes is fixed, although many Nobel Prizes in natural science have been shared in recent decades [see supplementary materials (SM)]. Although the number of academically appointed scientists with Ph.D.'s increased by 150% between 1973 and 2010 in the United States, the number of newly elected mem-

[1]*Institute for Social Research, University of Michigan, Ann Arbor, MI 48104, USA.* [2]*Center for Social Research, Peking University, Beijing 100871, China. E-mail: yuxie@umich.edu*

ILLUSTRATION: PETER AND MARIA HOEY/WWW.PETERHOEY.COM

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

bers each year to the National Academy of Sciences covering all scientific fields stayed at 60 for a long time until it finally went to 72 in 2001 (see SM).

Given science's concentration of high rewards to a select few, the large increase in the population of scientists (*7*) means a higher concentration over time. Furthermore, the tendency toward numbers-based evaluations has made it possible for an outsider—either a scientist in a different field or a nonscientist, who nonetheless may be in position to distribute resources—to pass judgment on any scientist, thus providing an easy pathway for the Matthew effect.

As with income inequality in the United States in general, inequality in academics' salaries has trended upward, both between private and public universities and among universities of each type (*13*). For individual academic scientists, salary inequality has increased substantially since the 1970s, across all ranks and diverse fields (*14*).

Here are two examples of intensified inequality in science today. First, many new science Ph.D. recipients from American universities in recent years have been unable to obtain regular academic positions and instead are taking postdoctoral fellowships or other forms of non–tenure-track employment (*7*). One source of this problem may be the large supply of well-trained foreign students and immigrant scientists (*7*). Of course, the system of postdocs and temporary employees provides benefits to both senior and junior (dependent) scientists. However, the former benefit more, as they are given more credit due to the Matthew effect.

Second, Internet technology, a globalized economy, inexpensive air transportation, and relatively peaceful world politics have created an unprecedentedly interconnected world (*15*). In this new global environment, a successful scientist filled with ideas at a prestigious university in America or Europe can design studies and have them carried out by dependent collaborators in less-developed countries, such as China, where labor-intensive scientific work can be conducted at lower costs. Such collaborations are complementary and can lead to mutual benefits (*16*); at the same time, they amplify inequality across individual scientists. Although global collaboration likely benefits all scientists, benefits vary with a scientist's position in a collaboration network. More-successful scientists are much more likely than less-successful ones to be centrally located in global collaborative networks.

CONTEXTUAL SOURCES OF INEQUALITY. The importance of institutional environment to scientists is well documented (*2*, *17*, *18*). Scientists affiliated with prestigious institutions are more productive and better rewarded than those who are not. Hence, greater institution-level inequality serves to intensify the individual-level inequality of scientists.

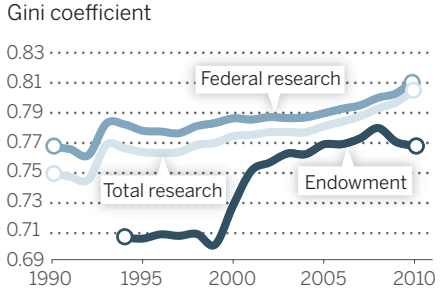
"Scientific outputs and rewards are much more unequally distributed than other well-being outcomes, such as education, earnings, or health."

There is some evidence that institution-level inequality in resources has increased over time. Institutional inequalities are shown in the chart by using the Gini coefficient, which ranges from 0 for absolute equality, to 1 for absolute inequality (see the chart). Despite already high levels of inequality in three resource measures, the Gini coefficients trended upward during the period. Part of the reason for the increases was an expansion in institutions that participate in research. If we restrict our analysis to a limited set of universities that have always been active in research, there is no clear trend (see SM).

Large country-level differences in scientific activities have long been noted (*19*). Historically, the world center of science has shifted several times; in the past nine decades, America has dominated (*7*). However, just as between-country income inequality has narrowed in the world, mostly because of a rise of income in China (*20*), between-country differences in science have also narrowed, thanks to the globalization of science (*7*, *16*).

Globalization of science and the increasing use of Internet technology have conflicting effects on inequality in science. At the levels of individual scientist and institution, they have tended to intensify inequality. At the country level, they have narrowed it. However, the most significant consequences of these two forces for science have been positive overall: the large expansion of science as a collective enterprise and the resulting rapid progress of science on a global scale (*16*). In the long run, resources and rewards must be allocated so that inequality, although incentivizing scientists to make important scientific discoveries, is properly managed and controlled. Especially important is the need to invest sufficient resources in young scientists before they gain recognition. ■

Inequalities in U.S. universities' research funding



RESOURCE INEQUALITY. Gini for resource inequality across U.S. universities, 1900–2010. Data reflect total research expenditure, federal research expenditure, and endowment for U.S. universities. Data from Center for Measuring University Performance, which draws from National Science Foundation (research expenditures) and National Association of College and University Business Officers (endowments). (See SM.)

REFERENCES AND NOTES

1. D. J. Price, *Little Science, Big Science* (Columbia Univ. Press, New York, 1963).
2. J. R. Cole, S. Cole, *Social Stratification in Science* (Univ. Chicago Press, Chicago, 1973).
3. R. K. Merton, *Science* **199**, 55 (1968).
4. R. H. Frank, P. J. Cook, *The Winner-Take-All Society* (Penguin, New York, 1995).
5. H. Zuckerman, *Scientific Elite: Nobel Laureates in the United States* (Free Press, New York, 1977).
6. R. K. Merton, *Am. Sociol. Rev.* **22**, 635 (1957).
7. Y. Xie, A. A. Killewald, *Is American Science in Decline?* (Harvard Univ. Press, Cambridge, MA, 2012).
8. E. P. Lazear, S. Rosen, *J. Polit. Econ.* **89**, 841 (1981).
9. R. G. Ehrenberg, M. L. Bognanno, *J. Polit. Econ.* **98**, 1307 (1990).
10. R. K. Merton, in *The Sociology of Science: Theoretical and Empirical Investigations* (Univ. Chicago Press, Chicago, 1973), pp. 267–278.
11. J. A. Evans, *Science* **321**, 395 (2008).
12. O. Rodríguez-Ruiz, *J. Educ. Adm.* **47**, 250 (2009).
13. R. G. Ehrenberg, *J. Labor Econ.* **21**, 267 (2003).
14. P. Stephan, *How Economics Shapes Science* (Harvard Univ. Press, Cambridge, MA, 2012), p. 41.
15. T. Friedman, *The World Is Flat: A Brief History of the Twenty-First Century* (Farrar, Straus, and Giroux, New York, 2005).
16. Y. Xie, *Issues Sci. Technol.* (Spring) (2014); http://issues.org/30-3/yu_xie/.
17. P. D. Allison, J. S. Long, *Am. Sociol. Rev.* **55**, 469 (1990).
18. J. R. Cole, *The Great American University: Its Rise to Preeminence, Its Indispensable National Role, Why It Must Be Protected* (Public Affairs, New York, 2009).
19. D. A. King, *Nature* **430**, 311 (2004).
20. G. Firebaugh, *The New Geography of Global Income Inequality* (Harvard Univ. Press, Cambridge, MA, 2003).

ACKNOWLEDGMENTS

Financial support provided by the Natural Science Foundation of China (grant no. 71373012), Peking University, and the Population Studies Center at the University of Michigan, which receives core support from the National Institute of Child Health and Human Development (R24HD041028). The author is grateful to C. Glovinsky, J. Hu, Q. Lai, and C. Zhang for research assistance.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/809/suppl/DC1

10.1126/science.1252743

Published by AAAS

BOOKS

ECONOMICS

Gauging national performance

GDP
A Brief but Affectionate History
Diane Coyle
Princeton University Press, 2014. 167 pp.

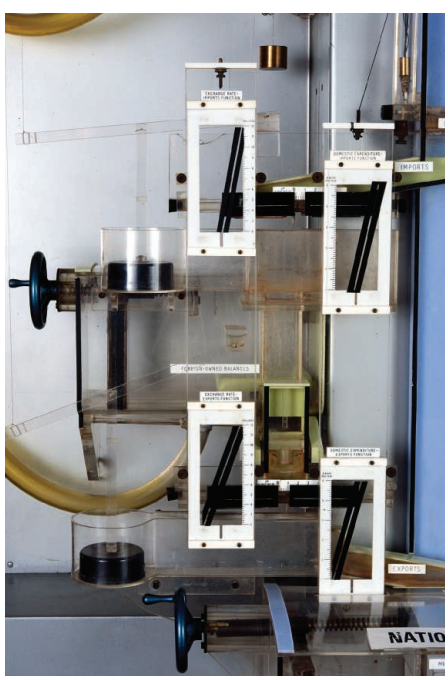
By **Nicholas Oulton**

"GDP rose by 0.3% in the fourth quarter." "China's GDP will be larger than that of the U.S. by 2020." Such statements are frequently encountered, and most educated people understand broadly what they mean. But not many people have an in-depth understanding of the concept of gross domestic product nor of how it is measured in practice. So a book that sets out to explain the basics is welcome.

Diane Coyle opens her book with Andreas Georgiou, someone probably very few people have heard of and quite hard to track down, even on Google. The former head of Greece's national statistical agency, Georgiou is due to be tried for corruption and making false statements, offences that carry a prison sentence of 5 to 10 years. His crime? Telling the truth about the Greek government's budget deficit. The budget deficit as a proportion of GDP was a number critical to Greece's negotiating position when it applied for one of its recent bailouts. Georgiou's refusal to falsify the figures is what landed him in trouble. His case illustrates the importance of GDP in economic policy-making and also the passions that it can arouse.

In six short chapters, Coyle (an economist at Oxford University) discusses how the concept of GDP was developed, starting with the 17th-century English savant William Petty and continuing through the pioneering contributions of Simon Kuznets, Colin Clark, and Richard Stone (Kuznets and Stone each won a Nobel Prize in Economics). She describes how their research led to an international effort under the aegis of the United Nations to define what came to be called the System of National Accounts (SNA), a remarkable example of international cooperation. The SNA has gone through a series of revisions since 1953, the latest of which (2008) is currently being rolled out around the world (with a few obvious exceptions, like North Korea). For a time a rival, the Material Product System, was promoted and enforced by the Soviet Union, but that has been abandoned. So the SNA rules alone, at least for now.

Coyle structures the book around epochs in economic history, seen from a Western perspective. She uses each epoch to illustrate how GDP was employed for policy purposes. Social and economic problems have led to criticisms of the GDP concept and sometimes to revisions to the SNA. The first chapter, "From the eighteenth century to the 1930s: war and depression," explains how these dire events led the U.S. and U.K. governments to start producing official statistics of GDP, building on earlier nonoffficial estimates by Kuznets and Clark. In the chapter, she also lays out the main accounting identities that form the framework of the SNA. The most important is that GDP in current prices equals the sum of final expenditures (consumption plus investment plus government expenditure on goods and services plus exports minus imports), the sum of factor incomes (wages plus profits), or the sum of value added in each industry (sales minus the cost of materials and bought-in services). In principle, expenditure = income = output, though in practice they differ because of errors and omissions in the data.



"The economy as a machine." Bill Phillips built his economic computer (1949) to show the flow of income in the UK economy.

PHOTO: © SCIENCE MUSEUM/SCIENCE & SOCIETY PICTURE LIBRARY

The earliest estimates of GDP were in current prices, which made it difficult to compare different years because there was no correction for inflation. Nor was it obvious how to compare GDPs of different nations measured in their respective currencies. At a high level of generality, the solution to these two puzzles is the same, a price index. Measuring domestic prices is complicated and something that Coyle could have dwelt on more. Within each category of expenditure, say, soft drinks, you select a basket of products. Then for each product, you have to track the prices of examples (say, a six-pack of regular Coca-Cola) in the same outlet over time (usually every month). The average of the prices of the selected products then forms the price index for soft drinks. Exactly the same method is used in cross-country comparisons. But now you have to collect the price of a six-pack of Coca-Cola (and the many other products) in different countries at the same point in time. The result is called a purchasing power parity. It might seem simpler (and cheaper) just to use exchange rates to make international comparisons. Coyle provides a good discussion of why this would not in general be sensible.

Coyle is a friend to GDP. Indeed, she ends the book noting, "GDP, for all its flaws, is still a bright light shining through the mist." But she is certainly not an uncritical friend. Throughout the book, and especially in the last chapter, she offers several criticisms of the GDP concept. First (though more clarification than criticism), GDP is a measure of output, not welfare (I agree). Second, considering GDP as a measure of output, some of the conventions adopted by the SNA's founders are disputable. For example, government activities are treated as part of GDP, although some, like Kuznets, wanted to exclude them on the grounds that they are part of the cost of producing anything at all. Arguably, defense and law and order are necessary costs, not valuable in themselves like food or entertainment. However, these days, at least in developed countries, the largest categories of government expenditure are for health and education. These expenditures are treated as consumption,

The reviewer is at the Centre for Economic Performance, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK. E-mail: n.oulton@lse.ac.uk

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

whereas they can be seen as, at least in part, a form of investment. Coyle also points out that unpaid housework is beyond what the SNA calls the "production boundary" and thus excluded from GDP. In principle, it need not be. We could measure the hours devoted to housework and value each at the wages of child-minders, cleaners, or, perhaps, those doing the housework if they also work outside the home.

A third set of criticisms involves tricky issues of measurement. Coyle argues that GDP tends to be understated, and increasingly so, as societies get richer: The variety of products available is continually increasing. An ever-growing volume of free online services (e.g., Google searches) are not included because their price is zero yet are clearly valued by consumers. And we have only poor measures of the output of service industries, which constitute the bulk of the economy and whose importance is still rising. I find the first two points well taken, but the third seems less valid. Our lack of good price indices makes it difficult to measure the real output of service industries. However, this does not matter for the total of real GDP because the latter is mainly measured from the expenditure side. On that side, most hard-to-measure service industries (e.g., banking, advertising, and accountancy) are not relevant because they sell mainly to other industries. Still, it would be good to more accurately know the output of various industries in order to better see where growth is coming from.

From a welfare point of view, something like net domestic product (i.e., GDP less depreciation) offers a better measure than GDP. But the current definition of depreciation in the SNA only covers assets created by humans (buildings, machinery, etc.)—leaving out depletion of resource stocks like oil and deterioration of environmental assets such as water and air.

Coyle argues that the importance of GDP will decline over the course of the 21st century as the world economy evolves. I am not so sure. I don't see China and India losing interest in GDP any time soon. And if the effects of climate change are as severe as many people forecast, then they are likely to focus more attention on measuring economic performance. Of course, one hopes that the SNA will continue to evolve to take better account of environmental impacts.

Anyone who wants to know how GDP and the SNA have come to play such important roles in economic policy-making will gain from reading Coyle's book. As will anyone who wants to gain more understanding of the concept's strengths and weaknesses. ■

10.1126/science.1251689

ECONOMIC HISTORY

Social mobility: Fixed forever?

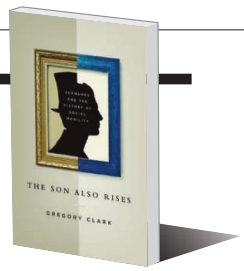
By **Miles Corak**

The Son Also Rises forcefully advances the idea that social position is determined by innate inherited abilities, an idea that is potentially pregnant with policy implications. "Once you have selected your mate," Gregory Clark counsels, "your work is largely done. You can safely neglect your offspring, confident that the innate talents you secured for them will shine through regardless."

With this book, Clark (an economic historian at the University of California, Davis) dons the mantle of Francis Galton, who more than 100 years ago examined the transmission of status across the generations of 19th century England and who is equally known for the statistical methods he developed to study the issue. And in the same way that careful readers have both admired and questioned Galton's work, so too can they legitimately admire *The Son Also Rises* as a work of scholarship while wondering about it as an exercise in scholastic overreach.

Social scientists have examined many different indicators of the intergenerational transmission of status. Clark draws conclusions from his and his collaborators' tracking of the status associated with surnames. Galton used height, but earnings and income are particularly salient in recent research, and occupation, education, wealth, and other indices have also figured prominently. The intergenerational transmission of earnings, for example, varies across countries, with the level of economic development, with the degree of inequality, and in some cases over time, as policies and institutions change. Other indicators have been found to show less variation across time and space. Whereas much of this research looks back only one or two generations, Clark traces the transmission of status over centuries.

The book holds that previous studies are based on observable manifestations of the intergenerational transmission of a latent characteristic, which earnings, education, occupation, and all the other indices measure imperfectly. Clark sometimes refers to this underlying characteristic as "social competence" and at other times outright as "social genotype," but ultimately what he means is "genes." When measured correctly, social competence has a high degree of intergenerational transmission, with about three-quarters of any relative advantage or disadvantage being transmitted from parent to child. This equals or exceeds Galton's estimate for height and is as much as three times some estimates of the transmission of earnings between fathers and sons.



The Son Also Rises
Surnames and the History of Social Mobility
Gregory Clark with others
Princeton University Press, 2014. 380 pp.

But more important, according to Clark, social mobility is an unchanging constant. It did not change with the arrival of free public education, the fall of nepotism in both the private and public sectors, economic growth, the expansion of the franchise in the 19th century, or even the rise of the welfare state and redistributive taxation in the 20th. Social mobility is not related to any of these policies, and it is not tied to the level of inequality. This, he claims, is a law for all times and all places.

I shall call it "Clarke's law," my deliberate misspelling of the author's name paying homage to his use of surnames and changes in surnames and to the use of *e* to represent influences other than family background—what Clark refers to as "luck"—in the mathematical representation of a son's status as a linear function of his father's status. Clarke's law states that 50% to 80% of variation in status is predictable at birth. There has always been social mobility, albeit very slow, but it has little to do with society, institutions, or, by implication, public policy.

The book's first part, an exercise in description, lays out the evidence for this law in countries for which surname information is available for the longest period. The approach involves tracking surnames through time, with due care that spellings may change, and linking them to some measure of social position.

For Sweden, this involves associating uncommon surnames among the Swedish nobility and educated elite of the 17th and 18th centuries to the social status of those carrying these same surnames today, as indicated by the names appearing in lists of physicians, attorneys, university students, and members of the Royal Academy and among the high

The reviewer is at the Graduate School of Public and International Affairs, Social Science Building, 6029, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada. E-mail: mcorak@uottawa.ca

Published by AAAS

Will the descendants of this boy in a grim district of Glasgow be found in similar circumstances?

PHOTO: JEFF J. MITCHELL/GETTY IMAGES

earners in publicly available tax records. The names of titled nobles, for example, appear on the Swedish Bar Association's register of member attorneys at nearly six times their proportion in the general public, and Clark estimates the persistence of high-status occupations, in general, at about 0.75.

Similar methods are applied for the United States and the United Kingdom, using elite names culled, for example, from the American Medical Association's directory of physicians, lists of licensed attorneys, or the list of federal taxpayers published in *The New York Times* in 1923–24.

To put these three modern cases in relief, the author offers surname evidence from medieval England dating back to 1300. Elite names are identified as people associated with Oxford and Cambridge universities, with wills proved in the highest court, with those probates associated with "Sir" or "Gentleman," and with Members of Parliament. The Norman conquerors of England recorded as property owners in the Domesday Book of 1086 are 16 times more likely than other names to be represented in Oxford or Cambridge in 1170 and 25% more likely now: a steady, but very slow, regression to the mean, suggesting a persistence rate of 0.90 from one generation to the next. The bottom line is that "medieval England had mobility rates similar to … those of modern United States and Sweden. In terms of social mobility, then, what did the Scientific Revolution, the Enlightenment, and the Industrial Revolution achieve? Very little."

These chapters rest on excellent research, but they also convey a whiff of overreach. That becomes manifest in the book's second part, which sets out to test the law in diverse countries (India, China, Taiwan, Japan, Korea, and Chile) and also among particular groups (Protestants, Jews, Gypsies, Muslims, and Copts).

Claiming that this dynamic applies to society as a whole rather than just to elites, Clark is sensitive to the need for supporting evidence. He repeatedly seeks confirmation that the data are representative of entire populations. Yet at times the data are very local—for example, the use of two specific regions near Shanghai to study of the evolution of status among Chinese surnames.

Even if it is debatable whether the various data are representative, it may still be appropriate to claim the law is widely applicable—so long as the proportion of advantage passed between generations does not vary across the social hierarchy (i.e., the mobility process is linear). But if Clarke's law varies with social status, then the book is only about the propagation of elite status across the generations.

The research community examining contemporary patterns in social mobility struggles with both representativeness and generality while using much more extensive and more accurate data than can be imagined for past times. That these issues could be resolved more cleanly by even the best economic history is a suggestion that many may find surprising.

Clarke's law can be described as a simple regression to the mean model in the spirit of Galton, one in which two components (reflecting the inheritance of genes and the impact of what some might call luck) shape a child's adult status. But other models can lead to equally high persistence of status. Powerful parents may influence child outcomes directly, rather than through inheritance, as a result of social institutions and things like primogeniture, nepotism, access to select colleges, or lax wealth and estate taxation. The children of the elite will succeed independent of their innate talents because of this type of social tracking.

Or it may be that the host of other influences on child outcomes should not be thought of as a random draw, as simple luck with no multigenerational echo. A longstanding family culture may foster capacities to, for example, teach children to speak with a socially approved accent, buffer them from the downsides of health risks and other threats to their human capital, or simply instill an identity of entitlement. A surname is not just an index of genes but also of social pressures and entitlements that keep some down and keep others from falling down. In this way, what Clark terms luck may also persist across generations and thus also generate persistence in "social competence."

Whether or not Clarke's law is valid or can be generalized, there is, nonetheless, an important lesson for public policy that should not be lost. Economic development, progressive income redistribution, the rise of public education, improvements in health, and the collective insurance of the welfare state are not inconsequential. The suggestion in this book is that they matter less for relative differences among individuals than we might think or hope, but surely they still matter for the absolute level of well-being and the development of human capacity—a message that may be lost by a focus on relative outcomes.

The author suggests his law implies that earnings and other economic outcomes should not be as unequally distributed as they are in some countries, because they reflect the transmission of genes and therefore competencies beyond individual control. But this is no more than a passing suggestion, and readers will suspect that Clarke's law will be interpreted just as forcefully in some public policy communities as an argument for the just deserts of those on top of the social heap.

Ultimately, as interesting as *The Son Also Rises* is as an exercise in historical scholarship, it does not advance contemporary public policy. Galton even began an 1892 book with the same focus as Clark's beginning and ending of *The Son Also Rises*: "I propose to show … that a man's natural abilities are derived by inheritance, under exactly the same limitations as are the form and physical features of the whole organic world. Consequently, … it would be quite practicable to produce a highly-gifted race of men by judicious marriages during several consecutive generations" (*1*). We have heard this advice many times over the decades, and it serves neither good public policy nor good parenting. ■

REFERENCES

1. F. Galton, *Hereditary Genius: An Inquiry into Its Laws and Consequences* (Macmillan, London, 1892).

10.1126/science.1251568

Published by AAAS

LETTERS

Edited by ***Jennifer Sills***

Specimen collection: An essential tool

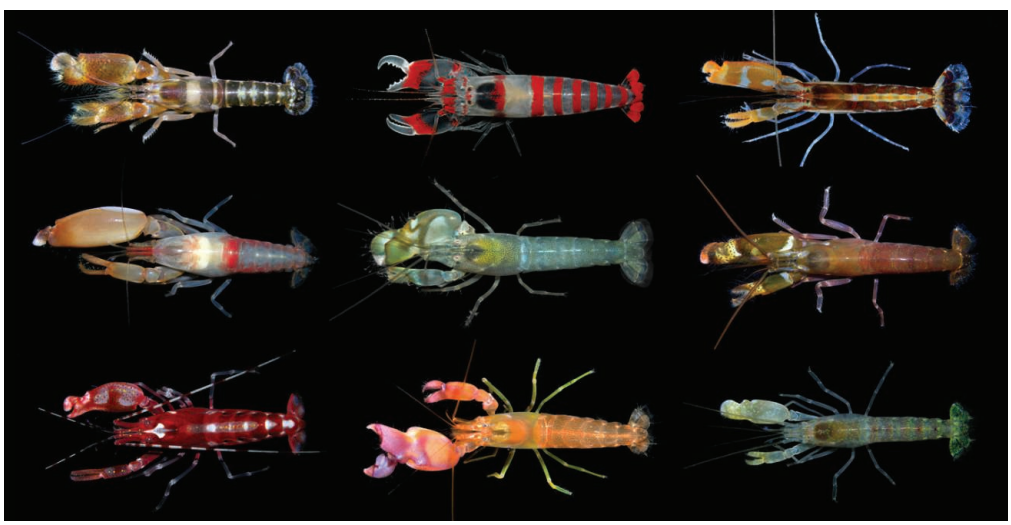
COLLECTING BIOLOGICAL specimens for scientific studies came under scrutiny when B. A. Minteer *et al.* ["Avoiding (re)extinction," Perspectives, 18 April, p. 260] suggested that this practice plays a significant role in species extinctions. Based on a small number of examples (rare birds, frogs, and a few plants), the authors concluded that collection of voucher specimens is potentially harmful to many species, and that alternatives—photographs, audio recordings and nonlethal tissue sampling for DNA analysis—are sufficient to document biological diversity.

The isolated examples that Minteer *et al.* cited to demonstrate the negative impact of scientific collecting have been carefully analyzed, and none of these extinction events can be attributed to that cause (*1*–*3*). For example, only about 102 Great Auk specimens (*Pinguinus impennis*) exist today in scientific collections, many of which are skeletons obtained after extinction, whereas millions were harvested for food, oil, and feathers over millennia (*1*). Similarly, only nine Mexican elf owls (*Micrathene whitneyi graysoni*), endemic to Socorro Island, Mexico, are present in natural history collections. Field notes indicate that this species was common when specimens were collected between 1896 and 1932, and the most likely reasons for extinction around 1970 were habitat degradation and predation by invasive species (*2*).

Scientists have come a long way from the days of indiscriminate collecting. Modern collecting adheres to strict permitting regulations and ethics guidelines, including the general practice of collecting a number of specimens substantially below levels that would affect population demography (*3*, *4*). The suggested alternatives to collecting specimens (photographing, recording calls, or sampling tissue nonlethally) are individually problematic, and even together cannot be used to reliably identify or describe the vast majority of Earth's biodiversity [for example, a large proportion of the world's marine biodiversity is hidden deep in its habitat (see image)]. Moreover, identification is often not the most important reason to collect voucher specimens. Studies of morphological diversity and its evolution are impossible without whole specimens. Preserved specimens also provide verifiable data points for monitoring species health, distribution, and phenotypes through time. Both historical and new collections played a key role in understanding the spread of the chytrid fungus infection, one of the greatest current threats to amphibians (*5*). The decision to ban dichlorodiphenyltrichloroethane (DDT) and other environmental pollutants was the result of the discovery of thinning of bird eggshells collected over an extended period (*6*). One of the negative effects of climate change, declining body size in animals, was only discovered with morphological data from museum specimens (*7*). Furthermore, IUCN Red List criteria require specific and detailed information about life history and biology (such as longevity and growth rate), especially for widely distributed species; therefore, without specimens, the extinction risk of many species cannot be properly assessed (*8*). Most specimens were not collected with these objectives in mind, and this is a hallmark of biological collections: They are often used in ways that the original collector never imagined. With new technologies continuing to emerge (such as stable isotope analyses, massive parallel sequencing, and CT-scan tomography), scientific collections are becoming even more important for studies of ecology, evolution, and conservation (*9*).

The arguments of Minteer *et al.* erroneously portray the critical importance of scientific collecting in a negative light and distract from the primary causes of modern extinction: habitat degradation and loss, unsustainable harvesting, and invasive species (*10*). It is important to distinguish protecting the lives of individuals from conserving populations and species. Individuals are lost every day to predation, natural death, and anthropogenic factors, hence it is the populations we try to save. Halting collection of voucher specimens by scientists would be detrimental not only to our understanding of Earth's diverse biota and its biological processes, but also for conservation and management efforts. Species descriptions, biodiversity inventories, and the identification of areas of endemism are just some of the basic information that can be obtained from specimens and collections-based research. Such knowledge, with its rich temporal and spatial dimensions, has proven fundamental in designing conservation areas and in making environmental impact assessments (*11*). These issues are particularly relevant in many developing nations, which ideally must seek a balance between the conservation of their natural (biological) resources and development. One example comes from the Bird's Head Peninsula of New Guinea, Indonesia, where the discovery and description of small endemic species—undetectable without specimen collection—directly resulted in the creation of several new protected areas and increased support for marine parks (*12*).

With our ever-increasing footprint, humans now affect even the most remote corners of Earth. Because an estimated 86% of species on the planet remain unknown (*13*), our goal should be to document biodiversity as rigorously as possible through carefully planned collections so that it can be effectively preserved



Undercover. Many Alpheidae shrimps live deep in the reef and are impossible to collect nonlethally.

PHOTO: ARTHUR ANKER

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

and understood. Specimens from such collections and their associated data are essential for making informed decisions about management and conservation now and in the future. As a community, we advocate the utmost responsibility and care while making scientific collections (*4*). Furthermore, given increasing rates of habitat loss and global change, we believe that responsibly collecting voucher specimens and associated data and openly sharing this knowledge (for example, through GBIF, iDigBio, and VertNet) are more necessary today than ever before.

L. A. Rocha,[1]* A. Aleixo,[2] G. Allen,[3] F. Almeda,[1] C. C. Baldwin,[4] M. V. L. Barclay,[5] J. M. Bates,[6] A. M. Bauer,[7] F. Benzoni,[8] C. M. Berns,[9] M. L. Berumen,[10] D. C. Blackburn,[1] S. Blum,[1] F. Bolaños,[11] R. C. K. Bowie,[12] R. Britz,[5] R. M. Brown,[13] C. D. Cadena,[14] K. Carpenter,[15] L. M. Ceríaco,[16] P. Chakrabarty,[17] G. Chaves,[11] J. H. Choat,[18] K. D. Clements,[19] B. B. Collette,[20] A. Collins,[20] J. Coyne,[21] J. Cracraft,[22] T. Daniel,[1] M. R. de Carvalho,[23] K. de Queiroz,[4] F. Di Dario,[24] R. Drewes,[1] J. P. Dumbacher,[1] A. Engilis Jr.,[25] M. V. Erdmann,[26] W. Eschmeyer,[1] C. R. Feldman,[27] B. L. Fisher,[1] J. Fjeldså,[28] P. W. Fritsch,[1] J. Fuchs,[29] A. Getahun,[30] A. Gill,[31] M. Gomon,[32] T. Gosliner,[1] G. R. Graves,[4] C. E. Griswold,[1] R. Guralnick,[33] K. Hartel,[34] K. M. Helgen,[4] H. Ho,[35] D. T. Iskandar,[36] T. Iwamoto,[1] Z. Jaafar,[4,37] H. F. James,[4] D. Johnson,[4] D. Kavanaugh,[1] N. Knowlton,[4] E. Lacey,[12] H. K. Larson,[38] P. Last,[39] J. M. Leis,[40] H. Lessios,[41] J. Liebherr,[42] M. Lowman,[1] D. L. Mahler,[25] V. Mamonekene,[43] K. Matsuura,[44] G. C. Mayer,[45] H. Mays Jr.,[46] J. McCosker,[1] R. W. McDiarmid,[4] J. McGuire,[12] M. J. Miller,[41] R. Mooi,[1] R. D. Mooi,[47] C. Moritz,[48] P. Myers,[49] M. W. Nachman,[12] R. A. Nussbaum,[49] D. Ó Foighil,[49] L. R. Parenti,[4] J. F. Parham,[50] E. Paul,[51] G. Paulay,[52] J. Pérez-Emán,[53] A. Pérez-Matus,[54] S. Poe,[55] J. Pogonoski,[39] D. L. Rabosky,[49] J. E. Randall,[56] J. D. Reimer,[57] D. R. Robertson,[41] M.-O. Rödel,[58] M. T. Rodrigues,[23] P. Roopnarine,[1] L. Rüber,[59] M. J. Ryan,[55] F. Sheldon,[17] G. Shinohara,[44] A. Short,[13] W. B. Simison,[1] W. F. Smith-Vaniz,[52] V. G. Springer,[4] M. Stiassny,[22] J. G. Tello,[22,60] C. W. Thompson,[49] T. Trnski,[61] P. Tucker,[49] T. Valqui,[62] M. Vecchione,[20] E. Verheyen,[63] P. C. Wainwright,[25] T. A. Wheeler,[64] W. T. White,[39] K. Will,[12] J. T. Williams,[4] G. Williams,[1] E. O. Wilson,[34] K. Winker,[65] R. Winterbottom,[66] C. C. Witt[55]

[1]California Academy of Sciences, San Francisco, CA 94118, USA. [2]Museu Paraense Emílio Goeldi, Belém, PA, 66040-170, Brazil. [3]Western Australian Museum, Perth, WA, 6986, Australia. [4]Smithsonian Institution, Washington, DC 20560, USA. [5]Natural History Museum, London, SW7 5BD, UK. [6]Field Museum of Natural History, Chicago, IL 60605, USA. [7]Villanova University, Villanova, PA 19085, USA. [8]University of Milano-Bicocca, Milan, 20126, Italy. [9]Utica College, Utica, NY 13502, USA. [10]King Abdullah University of Science and Technology, Thuwal, 23955, Saudi Arabia. [11]Universidad de Costa Rica, San José, 11501-2060, Costa Rica. [12]University of California, Berkeley, CA 94720–3161, USA. [13]University of Kansas, Lawrence, KS 66045, USA. [14]Universidad de los Andes, Bogotá, 4976, Colombia. [15]Old Dominion University, Norfolk, VA 23529, USA. [16]Museu Nacional de História Natural e da Ciência, Lisbon, 7005-638, Portugal. [17]Louisiana State University, Baton Rouge, LA 70803, USA. [18]James Cook University, Townsville, 4811, Australia. [19]University of Auckland, Auckland, 1142, New Zealand. [20]NOAA Systematics Laboratory, Washington, DC 20013, USA. [21]University of Chicago, Chicago, IL 60637, USA. [22]American Museum of Natural History, New York, NY 10024, USA. [23]Universidade de São Paulo, São Paulo, SP, 05508-090, Brazil. [24]Universidade Federal do Rio de Janeiro, Macaé, RJ, 27965-045, Brazil. [25]University of California, Davis, CA 95616, USA. [26]Conservation International, Denpasar, Bali, 80235, Indonesia. [27]University of Nevada, Reno, NV 89557–0314, USA. [28]Natural History Museum of Denmark, Copenhagen, DK-2100, Denmark. [29]Muséum National d'Histoire Naturelle, Paris, 75005, France. [30]Addis Ababa University, Addis Ababa, 1176, Ethiopia. [31]University of Sydney, Sydney, NSW, 2006, Australia. [32]Museum Victoria, Melbourne, 3001, VIC, Australia. [33]University of Colorado, Boulder, CO 80309–0334, USA. [34]Harvard University, Cambridge, MA 02138, USA. [35]National Museum of Marine Biology & Aquarium, Pingtung, 944, Taiwan. [36]Institut Teknologi Bandung, Bandung, 40132, Indonesia. [37]National University of Singapore, 117543, Singapore. [38]Museum and Art Gallery of the Northern Territory, Darwin, 0820, NT, Australia. [39]CSIRO Marine & Atmospheric Research, Hobart, TAS, 7000, Australia. [40]Australian Museum, Sydney, NSW, 2010, Australia. [41]Smithsonian Tropical Research Institute, Balboa, 0843-03092, Panamá. [42]Cornell University, Ithaca, NY 14853, USA. [43]Université Marien Ngouabi, Brazzaville, B.P. 69, Republic of Congo. [44]National Museum of Nature and Science, Tsukuba, 305-0005, Japan. [45]University of Wisconsin-Parkside, Kenosha, WI 53141–2000, USA. [46]Cincinnati Museum Center, Cincinnati, OH 45203, USA. [47]The Manitoba Museum, Winnipeg, MB, R3B 0N2, Canada. [48]Australian National University, Canberra, ACT, 0200, Australia. [49]University of Michigan, Ann Arbor, MI 48109–1079, USA. [50]California State University, Fullerton, CA 92831, USA. [51]The Ornithological Council, Chevy Chase, MD 20815, USA. [52]University of Florida, Gainesville, FL 32611, USA. [53]Universidad Central de Venezuela, Caracas, 1041, Venezuela. [54]Pontificia Universidad Católica de Chile, Santiago 6513677, Chile. [55]University of New Mexico, Albuquerque, NM 87131–0001, USA. [56]Bernice P. Bishop Museum, Honolulu, HI 96817, USA. [57]University of the Ryukyus, Nishihara, 903-0213, Japan. [58]Museum für Naturkunde, Berlin, 10115, Germany. [59]Naturhistorisches Museum der Burgergemeinde Bern, Bern, CH-3005, Switzerland. [60]Long Island University, Brooklyn, NY 11201–8423, USA. [61]Auckland Museum, Auckland, 1142, New Zealand. [62]Centro de Ornitologia y Biodiversidad, Lima, 33, Peru. [63]Royal Belgian Institute of Natural Sciences, Brussels, 1000, Belgium. [64]McGill University, Montreal, QC, H9X 3V9, Canada. [65]University of Alaska Museum, Fairbanks, AK 99775, USA. [66]Royal Ontario Museum, Toronto, ON, M5S 2C6, Canada.

*Corresponding author. E-mail: LRocha@calacademy.org

REFERENCES

1. E. Fuller, *The Great Auk* (H. N. Abrams, New York, 1999).
2. J. P. Hume, M. Walters, *Extinction in Birds* (Bloomsbury, London, 2012).
3. N. J. Collar, *Bird Cons. Int.* **10**, 1 (2000).
4. K. Winker *et al.*, *Auk* **127**, 690 (2010).
5. T. L. Cheng, S. M. Rovito, D. B. Wake, V. T. Vredenburg, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9502 (2011).
6. R. D. Porter, S. N. Wiemeyer, *Science* **165**, 199 (1969).
7. J. L. Gardner, A. Peters, M. R. Kearney, L. Joseph, R. Heinsohn, *Trends Ecol. Evol.* **26**, 285 (2011).
8. Y. Sadovy de Mitcheson *et al.*, *Fish Fish.* **14**, 119 (2013).
9. K. Bi *et al.*, *Mol. Ecol.* **22**, 6018 (2013).
10. M. Clavero, E. Garcia-Berthou, *Trends Ecol. Evol.* **20**, 110 (2005).
11. B. Collen, N. Pettorelli, J. E. M. Baillie, S. M. Durant, Eds., *Biodiversity Monitoring and Conservation: Bridging the Gap between Global Commitment and Local Action* (Wiley, Cambridge, UK, 2013).
12. M.V. Erdmann, in *Still counting…: Biodiversity Exploration for Conservation – the first 20 years of the Rapid Assessment Program*, L. E. Alonso, J. L. Deichmann, S. A. McKenna, P. Naskrecki, S. J. Richards, Eds. (Conservation International, Arlington, VA, 2010).
13. C. Mora, D. P. Tittensor, S. Adl, A. G. B. Simpson, B. Worm, *PLOS Biol.* **9**, e1001127 (2011).

Specimen collection: Plan for the future

WE WISH THAT B. A. Minteer *et al.*'s claim that field biologists routinely collect voucher specimens were true ["Avoiding (re)extinction," Perspectives, 18 April, p. 260]. Any museum curator will tell you that it is a constant struggle to convince them to do so, despite countless publications rendered unreliable because it is impossible to verify species' identities. The necessity of voucher specimens varies by taxon and region, but in general, it is good practice to deposit them and as much data as possible, including DNA and photos in life.

We certainly do not wish to see any species driven to extinction by overcollecting, but submit that this is rare and more associated with commercial or ardent, recreational overcollecting than sensible scientific vouchering (*1*, *2*). If the kill of a single individual increases the extinction risk of a species, then it is well below viable population size and already among the "walking dead."

Dawkins' description of evolution as improbability on a colossal scale is nowhere more evident than in morphology. Whether or not a species survives, museum specimens represent a window on many of its most remarkable novelties. Molecular data, although helpful in identifications, is neither a panacea nor surrogate for museum specimens, especially when it comes to newly discovered species. Describing a new species without depositing a holotype when a specimen can be preserved borders on taxonomic malpractice. Even given good photographs and a tissue sample, there are reasons to collect one or more complete specimens. We do not know what morphological characters will prove important in future studies of species status, phylogenetic

Published by AAAS

relationships, or genetic or epigenetic variation. As taxonomists and ecologists, we do not want to know only that a species exists but to understand what makes it unique compared to related species. Given the importance of the phenotype-environment interface in natural selection, we potentially sacrifice the most important things to know about a species when we forego more than superficial evidence of anatomical details.

With millions of species threatened by extinction, it would be tragic were we left with no more than a few photographs and sequences as evidence they were once here. Given well-preserved specimens, we can continue to marvel at adaptations, discover models for biomimicry, refine theories of character transformations, and verify the state of internal or external structures discovered in related species. As the last generation with the opportunity to explore, discover, and document millions of species evolved over billions of years, we should not be so arrogant as to assume what science of the future may want or need.

Frank-T. Krell[1] and Quentin D. Wheeler[2]

[1]Department of Zoology, Denver Museum of Nature & Science, Denver, CO 80205, USA. [2]College of Environmental Science and Forestry, State University of New York, Syracuse, NY 13210, USA.

*Corresponding author. E-mail: frank.krell@dmns.org

REFERENCES

1. D. A. Norton *et al.*, *Taxon* **43**, 181 (1994).
2. K. Winker *et al.*, *Auk* **127**, 690 (2010).

Response

THE PURPOSE OF OUR Perspective was to raise awareness about an issue that will increase in prevalence as the global biodiversity crisis unfolds: Absent a reliable estimate of population size, is it prudent and ethical to collect a newly observed individual of a species so rare it was thought extinct [e.g., (*1*)]? We support the work of natural history museums, and nowhere in our discussion did we argue that responsible collecting should be halted. Specimen collections provide invaluable contributions to many disciplines beyond taxonomy [e.g., (*2*, *3*)]; moreover, we continue to collect ourselves (J.P.C. and R.P.). We repeatedly emphasized that we were targeting the specific context of small and vulnerable populations only.

We would like to believe that we live in Rocha *et al.*'s world in which the responsible collector follows every regulation and ethical code (where these exist). Our own experience and research, however, paint a more complicated picture. A culture of responsible scientific practice is harder to establish than just following regulatory prescriptions and ethical injunctions (*4*). Rocha *et al.* also introduce a red herring by raising the distinction between individual- and population- or species-level concern in conservation, which we understand and have discussed elsewhere (*5*). It is obvious that our Perspective concerns survival of populations and species; the individual specimen becomes important in our argument because of the small size of populations, especially when (as in the case of rediscovered amphibian populations) such individuals are found coexisting with the lethal pathogen that likely greatly reduced their numbers (*6*).

Nowhere do we claim that scientific collection is a leading driver of extinction. We are aware of the major threats posed by habitat loss and fragmentation, commercial use, exotic species, toxins, infectious diseases, and climate change (*7*). Collectors may have taken the last Auks, but the species was pushed to the brink of extinction by centuries of human overexploitation. Still, the point remains that without a reliable estimate of population size, collecting individuals from a small, isolated population can pose an extinction risk. We believe that it is important to highlight this risk, and to suggest how to mitigate the threat.

We are troubled by Krell and Wheeler's argument, which seems to suggest that collecting in vulnerable populations is justified as a way to preserve the present for a future in which many species will be extinct. Even small populations seem eligible for collecting based on their claim that such species are already among the "walking dead." If collecting a specimen increases extinction risk, however, then it is a threat to biodiversity and should be avoided. Krell and Wheeler object to the "arrogance" of assuming "what science of the future may want or need," but we find more hubris in their suggestion that taxonomists and ecologists should be unconcerned about driving the final nail in a species' coffin.

Cultural change in science can be difficult. Long-established techniques are questioned as alternatives arise. Specimen collection is no exception, especially in light of growing concerns about our entering a sixth mass extinction event (*8*), and we encourage more research into new ways to document Earth's biodiversity. A precautionary approach to scientific collection will help ensure that we do not put additional pressure on already vulnerable populations as we seek to identify organisms new to science, or to confirm a species' welcome return from the dead.

Ben A. Minteer,[1]* James P. Collins,[1]
Robert Puschendorf[2]

[1]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA. [2]School of Biological Sciences, Plymouth University, Drake Circus, Plymouth, Devon PL4 8AA, UK.

*Corresponding author. E-mail: ben.minteer@asu.edu

REFERENCES

1. IUCN Red List of Threatened Species, Craugastor fleischmanni (www.iucnredlist.org/details/56603/0).
2. C. Moritz *et al.*, *Science* **322**, 261 (2008).
3. R. Puschendorf, F. Bolaños, G. Chaves, *Biol. Conserv.* **132**, 136 (2006).
4. B. A. Minteer, J. P. Collins, *Sci. Eng. Ethics* **14**, 483 (2008).
5. B. A. Minteer, J. P. Collins, *ILAR J.* **54**, 41 (2013).
6. M. J. Ryan, F. Bolaños, G. Chaves, *Science* (2010); published online: www.sciencemag.org/content/329/5997/1272/reply.
7. J. P. Collins, M. Crump, *Extinction in Our Times: Global Amphibian Decline* (Oxford Univ. Press, Oxford, 2009).
8. E. Kolbert, *The Sixth Extinction: An Unnatural History* (Henry Holt, New York, 2014).

ERRATA

Editor's note: We are simplifying our procedure for making corrections to articles published in *Science*, while maintaining transparency for our readers. The full text and PDF files will be corrected online as soon as possible, with an explanation at the end of the full text and, for corrections involving data or metadata, in an accompanying online Erratum. A notification that an Erratum has been published online will appear in a subsequent print issue in this space.

Erratum for the Research Article: "Total Synthesis of a Functional Designer Eukaryotic Chromosome" by N. Annaluru *et al.*, *Science* **344**, 1254596 (2014). Published online 18 April; 10.1126/science.1254596

Erratum for the Report: "Wild Pollinators Enhance Fruit Set of Crops Regardless of Honey Bee Abundance" by L. A. Garibaldi *et al.*, *Science* **344**, 1255213 (2014). Published online 2 May; 10.1126/science.1255213

Erratum for the News & Analysis: "Designer Microbes Expand Life's Genetic Alphabet" by R. F. Service, *Science* **344**, 1255780 (2014). Published online 16 May; 10.1126/science.1255780

Erratum for the Report: "I-Love-Q: Unexpected Universal Relations for Neutron Stars and Quark Stars" by K. Yagi and N. Yunes, *Science* **344**, 1250349 (2014). Published online 23 May; 10.1126/science.1250349

Erratum for the Report: "Mapping the Cellular Response to Small Molecules Using Chemogenomic Fitness Signatures" by A. Y. Lee *et al.*, *Science* **344**, 1255771 (2014). Published online 23 May; 10.1126/science.1255771

Published by AAAS

Downloaded from www.sciencemag.org on May 26, 2014

the science of Inequality

What the numbers tell us

By **Gilbert Chin** *and* **Elizabeth Culotta**

In 2011, the wrath of the 99% kindled Occupy movements around the world. The protests petered out, but in their wake an international conversation about inequality has arisen, with tens of thousands of speeches, articles, and blogs engaging everyone from President Barack Obama on down. Ideology and emotion drive much of the debate. But increasingly, the discussion is sustained by a tide of new data on the gulf between rich and poor.

This special issue uses these fresh waves of data to explore the origins, impact, and future of inequality around the world. Archaeological and ethnographic data are revealing how inequality got its start in our ancestors (see pp. 822 and 824). New surveys of emerging economies offer more reliable estimates of people's incomes and how they change as countries develop (see p. 832). And in the past decade in developed capitalist nations, intensive effort and interdisciplinary collaborations have produced large data sets, including the compilation of a century of income data and two centuries of wealth data into the World Top Incomes Database (WTID) (see p. 826 and Piketty and Saez, p. 838).

It is only a slight exaggeration to liken the potential usefulness of this and other big data sets to the enormous benefits of the Human Genome Project. Researchers now have larger sample sizes and more parameters to work with, and they are also better able to detect patterns in the flood of data. Collecting data, organizing it,

INSIDE

NEWS

The ancient roots of the 1% *p. 822*

Our egalitarian Eden *p. 824*

Tax man's gloomy message: the rich will get richer *p. 826*

Physicists say it's simple *p. 828*

Can disparities be deadly? *p. 829*

While emerging economies boom, equality goes bust *p. 832*

Tracking who climbs up—and who falls down—the ladder *p. 836*

RESEARCH REVIEWS

Inequality in the long run *p. 838*

Skills, education, and the rise of earnings inequality among the "other 99 percent" *p. 843*

Income inequality in the developing world *p. 851*

The intergenerational transmission of inequality: Maternal disadvantage and health at birth *p. 856*

On the psychology of poverty *p. 862*

SEE ALSO

▸ EDITORIAL p. 783

▸ PERSPECTIVE p. 809

▸ BOOK REVIEWS pp. 811 & 812

▸ ONLINE PODCAST, VIDEO, AND BLOGS

▸ HTTP://SCIENCEMAG.ORG/SPECIAL/INEQUALITY

PHOTO: (LEFT) STEVE RAYMER/CORBIS

Published by AAAS

developing methods of analysis, extracting causal inferences, formulating hypotheses—all of this is the stuff of science and is more possible with economic data than ever before. Even physicists have jumped into the game, arguing that physical laws may help explain why inequality seems so intractable (see p. 828).

In the United States, the new information suggests a wide rift between top and bottom. Tax data from the WTID suggest that today the top 1% control nearly 20% of U.S. income, up from about 8% in the 1970s. But inequality is increasing within the 99%, too, as a consequence of a growing premium on college and postgraduate education: The fates of the tech-savvy worker at Google and the blue-collar employee at General Motors have been decoupled (see Autor, p. 843). According to surveys by the Census Bureau, in 2012 the richest 20% of Americans enjoyed more than 50% of the nation's total income, up from 43% in 1967. The middle 20%—the actual middle class—received only about 14% of all income, and the poorest got a mere 3% (see graphic).

Flip to a world map, and America's inequality, despite reaching levels last seen in the Gilded Age, turns out to be far from extreme. Many nations, especially emerging economies, have even larger chasms between the super-rich and the poor. One widely used metric, the Gini coefficient, estimates inequality as an index between 0—at which point everyone has exactly equal incomes—to 1, in which a single person takes all the income and the rest get nothing. The U.S. Gini, at 0.40 in 2010, seems relatively high compared with, for example, Japan at 0.32. But South Africa is a sky-high 0.7.

Many assume that governments in emerging economies have chosen to favor growth even at the cost of inequality on the

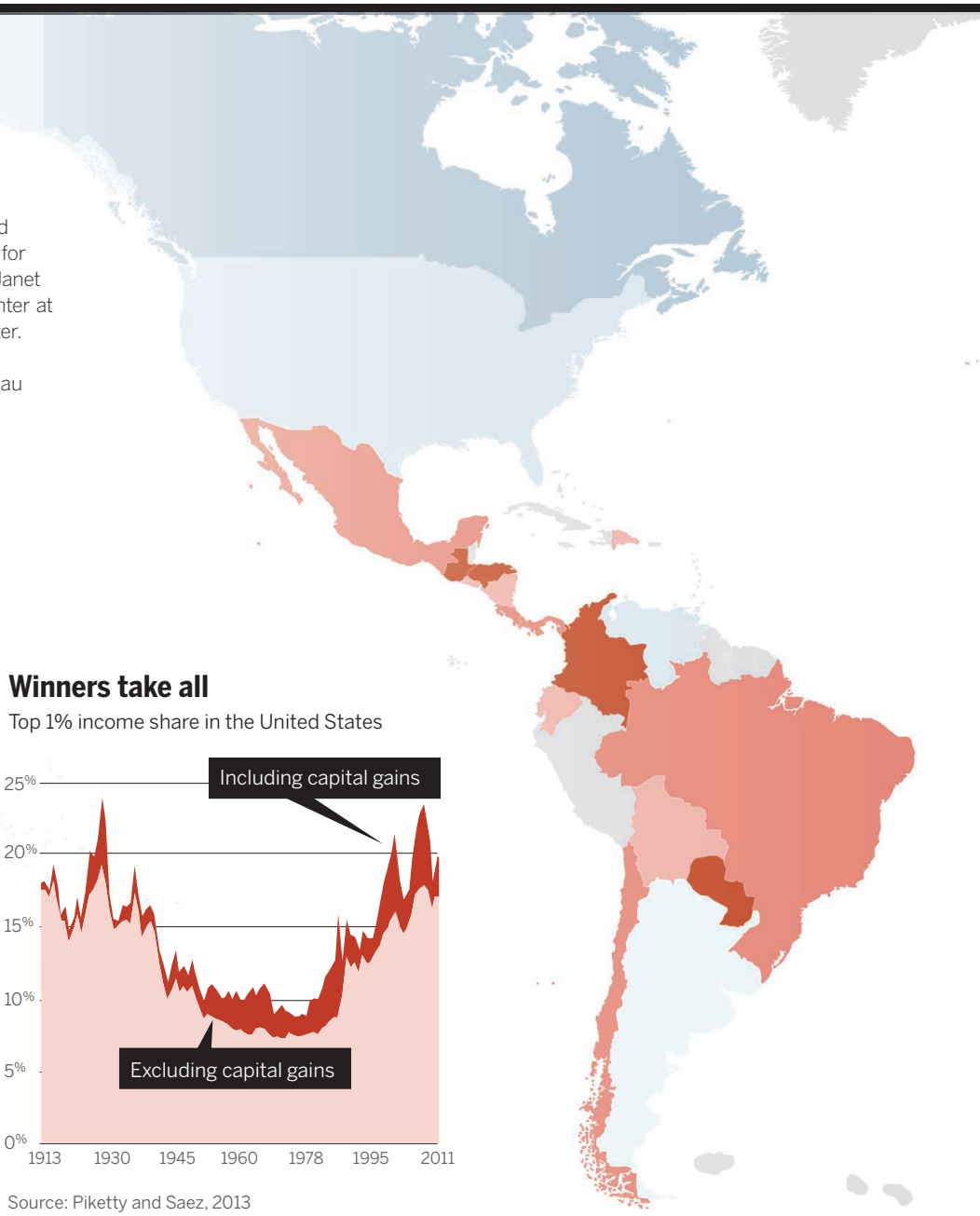
A world of difference

Countries vary widely in inequality

Compiled by **Emily Underwood**

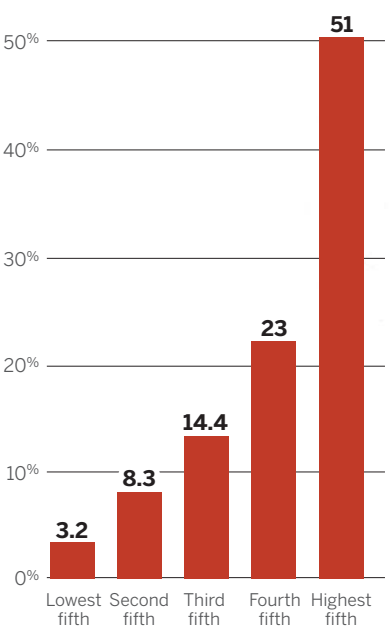
The world Gini data, collected between 2008 and 2012, cover 117 countries and were prepared for *Science* by researchers Branko Milanovic and Janet Gornick of the Luxembourg Income Study Center at the City University of New York's Graduate Center.

U.S. data are based on 2012 U.S. Census Bureau surveys of 122,459 households.



A sharp divide

Shares of U.S. income by quintile, 2012



Source: U.S. Census Bureau

Winners take all

Top 1% income share in the United States

Source: Piketty and Saez, 2013

Published by AAAS

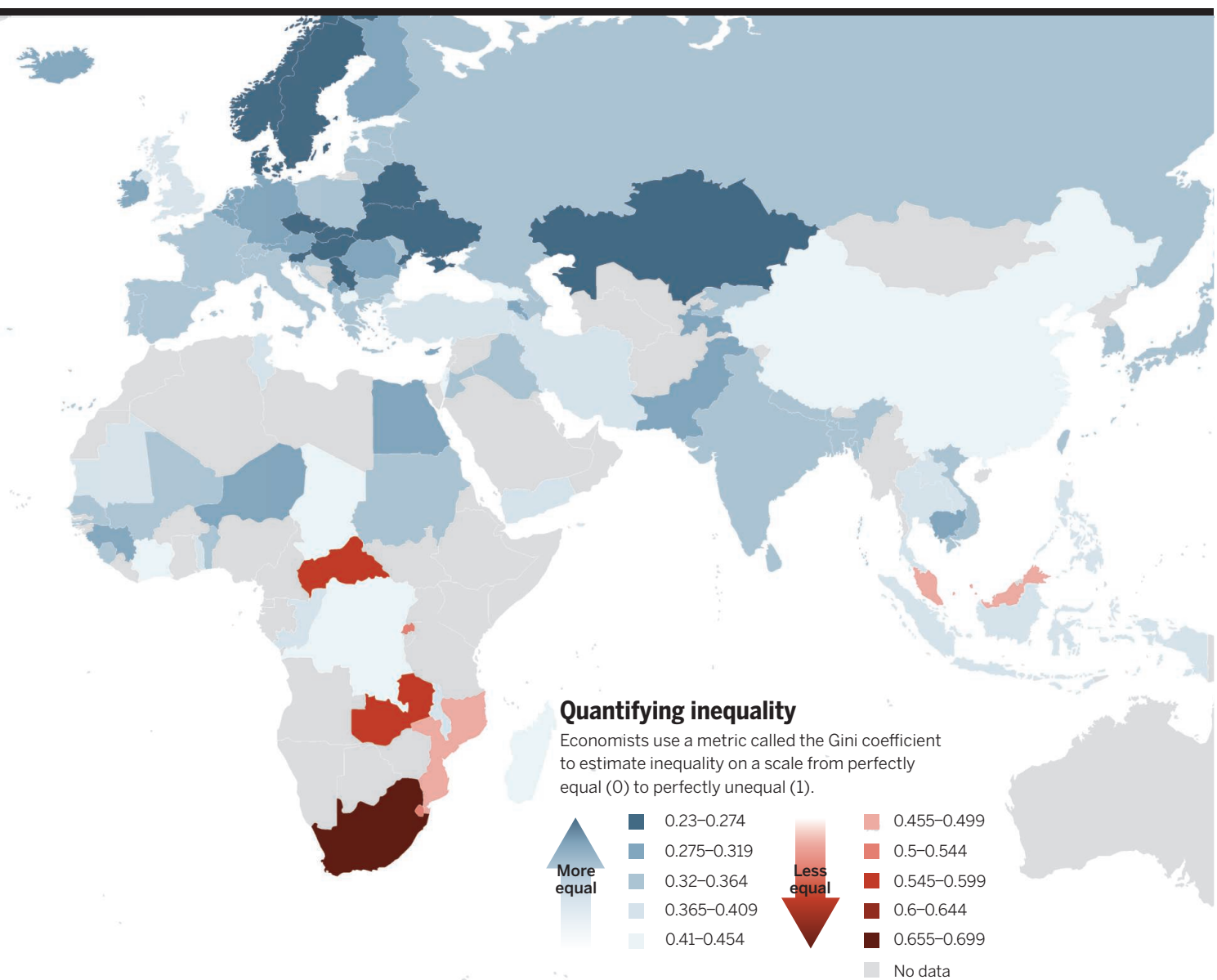
grounds that "a rising tide lifts all boats." But evidence that this trade-off is necessary is sparse, and recent data show that policies to reduce inequality need not stymie growth (see Ravallion, p. 851).

What of those at the bottom? Research has established a base of knowledge about the harmful effects of disadvantageous circumstances on education and health. These influences can begin early in life, even prenatally (see Aizer and Currie, p. 856). But researchers are still exploring whether the stress of being low-ranked itself adds to the poor's burden, causing illness and even early death (see p. 829). In addition, psychological mechanisms may spur a negative feedback loop in which poor individuals behave in ways that help keep them poor (see Haushofer and Fehr, p. 862).

Harsh as life can be for those at the bottom, the opportunity to move up the ladder can compensate. Newly available data from taxes and other records promise to yield insights into intergenerational mobility, in which children advance from their parents' socioeconomic status. But so far, researchers have a relatively limited view of how and why people move into different social, as well as economic, classes (see p. 836 and http://scim.ag/sci_inequality; also see Corak, p. 812).

Few would deny that excessive inequality can be unhealthy for societies and economies, but the new data don't pinpoint a desirable level. They do show that the forces that foster inequality—from the patchy distribution of resources among ancient hunter-gatherers to the sheer earning power of capital today—are many and potent. It is up to society to decide whether, and how, to restrain them (see p. 783).

Gilbert Chin is a senior editor for Science *and Elizabeth Culotta is a deputy news editor for* Science.



Published by AAAS

A Pompeiian fresco shows the Roman 1% living the good life.

Downloaded from www.sciencemag.org on May 26, 2014

The ancient roots of the 1%

Don't blame farming. Inequality got its start among resource-rich hunter-gatherers

By **Heather Pringle**

In 79 C.E., the year Mount Vesuvius destroyed it, Pompeii was not one city but two. Its wealthiest families owned slaves and lived in multistoried, seaside mansions, one of which was more than half the size of the White House. They dined in rooms with costly frescoes, strolled in private gardens, and soaked in private baths. Meanwhile, at least one-third of all Pompeiian households scraped to make ends meet, with families dwelling in single rooms behind workshops, in dark service quarters, or in small houses. Such economic disparities were common in the Roman Empire, where 1.5% of the empire's households controlled 20% of the income by the late 2nd century C.E., according to one recent study.

Inequality has deep archaeological roots. Yet if existing traditional societies are any guide, our hunter-gatherer ancestors were mostly egalitarian (see sidebar, p. 824). How and when did a few members of society begin to amass wealth?

Farming has long been blamed for the rise of inequality. Relying on evidence from the Near East, researchers suggested that the earliest elites emerged after 10,500 years ago, when people successfully domesticated plants and animals and settled in large permanent villages. In this view, agriculture led to the production of surpluses and the emergence of managers, craftspeople, and other specialists, who eventually gained control over extra resources.

Now, analyses of archaeological sites as well as ethnographies of traditional societies are etching a more complex picture, suggesting that some ancient hunter-gatherers may have accumulated wealth and political clout by taking control of concentrated patches of wild foods. In this view, it is the ownership of small, resource-rich areas—and the ease of bestowing them on descendants—that fosters inequality, rather than agriculture itself.

The transition from egalitarianism to societies rife with economic competition and inequality was "the single most critical watershed in the last 2.5 million years of human history," says archaeologist Brian Hayden of Simon Fraser University (SFU), Burnaby, in Canada. Over time, it paved the way for the development of "chiefdoms, states, and ultimately industrial empires."

BEFORE FARMING. Archaeologists have spotted the earliest glimmers of inequality among the Natufians of the Eastern Mediterranean, one of the first peoples to embark on the long transition to farming. Beginning some 14,500 years ago, the Natufians began settling at least part-time in small villages amid rich food resources, regularly supplementing their diet of wild game, fruits, and nuts with wild cereals—a lifestyle that ultimately led to agriculture.

Natufians left some traces of inequality behind. Archaeologists T. Douglas Price of the University of Wisconsin, Madison, and

PHOTO: MICHELE FALZONE/GETTY IMAGES

Published by AAAS

Ofer Bar-Yosef of Harvard University examined published reports for 25 Natufian and later sites dating between about 15,000 and 8000 years ago. The two looked for standard archaeological markers of inequality: disparities in grave goods, house sizes, and the ornamentation of the dead. Not surprisingly, inequality markers became more common between 10,500 and 8200 years ago, as early farmers began sowing domesticated einkorn wheat and other plants and tending domesticated sheep and goats.

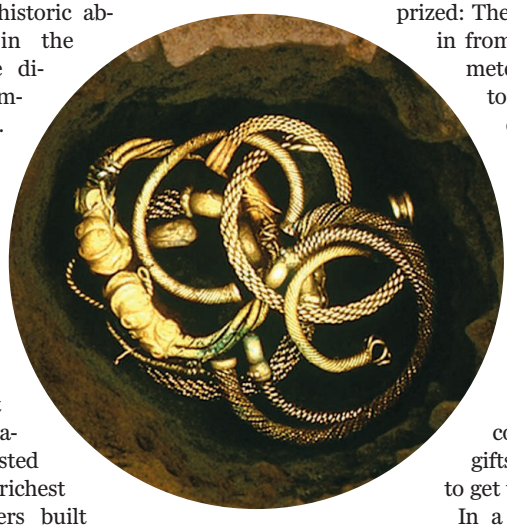
But signals of incipient inequality appeared well before that, between 14,500 and 12,800 years ago, while the Early Natufians were still hunting and gathering, Price and Bar-Yosef reported in a 2010 volume. Some Early Natufian skeletons were richly ornamented, but the vast majority were not. The wealthiest 8%, for example, were decorated with pendants or marine shells such as *Dentalium*, imported or traded from as far as 400 kilometers away. At one site, three male skeletons were buried with *Dentalium* headdresses, one fringed with shells four deep—an impressive display of riches. Natufians also placed carved artworks in a few graves, built houses of varying sizes, and produced large goblet-shaped stone mortars well-suited for preparing or serving food at feasts. Making these mortars "by pecking for many hours is hardly the business of a fully egalitarian society," noted Price and Bar-Yosef.

The findings suggest how people took a first tentative step on the long road to inequality. The Natufians lived in an environment of abundance—wild cereal grains flourished in dense patches in the forest, and game was plentiful. As Price points out, the Natufians apparently "were harvesting wild plants in large quantities and storing cereal grains as well." He thinks that these stored surpluses of wild cereals may have given some Natufian hunter-gatherers an edge over others. "Those surpluses could allow people to begin manipulating things, giving away food and so establishing some dominance behaviors."

HOLDING ON TO WEALTH. Other abundant, storable wild foods can lead to surpluses, too, and private ownership of these natural resources could have boosted inequality, creating a new kind of "transegalitarian" hunter-gatherer society.

Take an ancient village on Keatley Creek in Canada's Northwest Plateau, which was occupied for part of the year by hunter-gatherers between 2500 and 1100 years ago. The village contains more than 115 house pits, the remains of semisubterranean structures with log and earthen roofs, and appears to have had a peak population of as many as 1500 people. The excavation team, led by SFU's Hayden, found that the houses varied dramatically in size, from the square footage of a microapartment to that of a medium-sized house today.

To understand these disparities, Hayden and his colleagues examined ethnographic records of historic aboriginal societies in the region, which were divided into nobles, commoners, and slaves. The highest status families owned certain resources and passed them down to their children: fences for driving deer into hunting traps and, especially, fishing rocks that jutted out into the Fraser River, which hosted some of the world's richest salmon runs. Owners built fishing platforms out from these rocks, and so could fish in deep waters where the biggest salmon swam. Lower status families had to fish from public areas along the riverbanks with dip nets, and could reach only smaller fish. Families then wind-dried their catch and stored it.



Since before the Romans, elites have hoarded wealth in gold.

PHOTO: THE BRITISH MUSEUM

To see if this private ownership of resources extended back in time, Hayden's team analyzed fish vertebrae excavated from house pits of various sizes. As much as 75% of the fish bone in the large house pits came from big, 4- to 5-year-old chinook and sockeye salmon, laden with calorie-rich fat. In contrast, 100% of the bone in the two smallest houses came from smaller, 2- to 3-year-old salmon likely caught along the riverbanks.

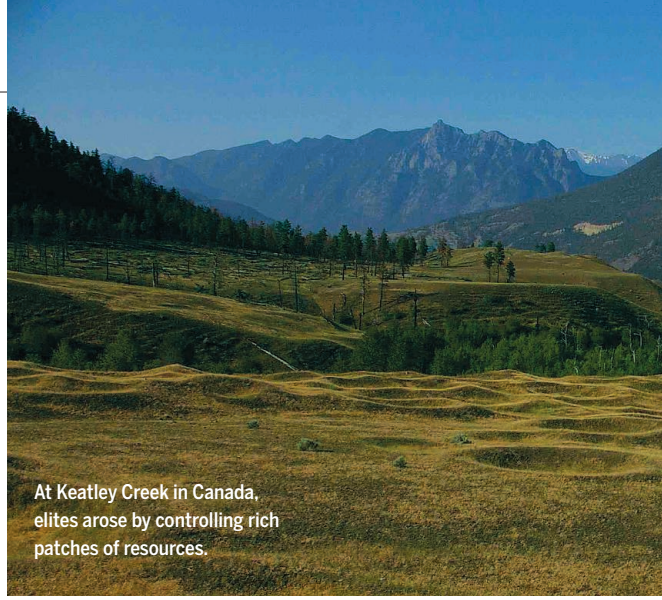
The findings suggest that inequality began at Keatley Creek some 2500 years ago when a few ambitious, aggressive people capitalized on the salmon's bounty, Hayden says. Aggrandizers who wanted more food than their neighbors likely built fishing platforms out over key fishing rocks and claimed private ownership. These aggrandizers controlled bigger food surpluses than others, but no one stopped them—as can happen to those who refuse to share in other hunting and gathering societies—because there was plenty of food for all, Hayden says. "It is no coincidence that the greatest inequalities on the Northwest Plateau emerged at the most productive fishing locations, where huge surpluses were produced ethnographically."

Hayden and his team also found evidence at Keatley Creek of large roasting pits, one of which was large enough to cook food for 500 people. This and other evidence suggested that aggrandizers at Keatley Creek organized feasts resembling historic potlatches, in which a chief cajoled his clan into producing a cornucopia of food as well as obtaining prestige goods such as *Dentalium*, which these people also prized: They brought the shells in from as far as 300 kilometers away, to wear and to give away to a rival clan. Such feasts publicly displayed the host's power and wealth, and forced rival chiefs to compete. Potlatch guests were expected to reciprocate with goods of greater value, and families who couldn't come up with gifts had to go into debt to get them.

In a study published in a 2011 volume called *Guess Who's Coming To Dinner: Feasting Rituals in the Prehistoric Societies of Europe and the Near East*, Hayden argues that elites in Natufian villages may have pursued similar tactics. Like Price, Hayden argues that long before farming, Natufian elites could have amassed large surpluses of food by "owning" natural concentrations of resources, such as groves of pistachio trees, or by constructing drive lanes for hunting gazelles. Massive roasting pits and hearths at some Natufian sites suggest a feasting tradition, too.

Some researchers think Hayden overemphasizes the role of aggressive, competitive individuals in the origins of inequality, and underemphasizes the role of population pressures or resource stresses. Archaeologist Anna Marie Prentiss of the University of Montana, Missoula, contends that in another ancient Northwest Plateau village in Canada, it was a shortage of food, rather than an abundance, that sparked inequality. Data from her team's excavations at the Bridge River site suggest that the first elites emerged after salmon runs declined about 1200 years ago and the village population plummeted, she and colleagues reported online in December in the *Journal of Anthropological Archaeology*. They found that some families responded to scarcity by closing off public access to hunting and fishing resources and holding feasts to attract workers to their depleted households, tactics that allowed them to amass more food than their neighbors. Inequality at Bridge River, Prentiss says, "came about as a byproduct of feeding

Published by AAAS



At Keatley Creek in Canada, elites arose by controlling rich patches of resources.

their families" during lean times.

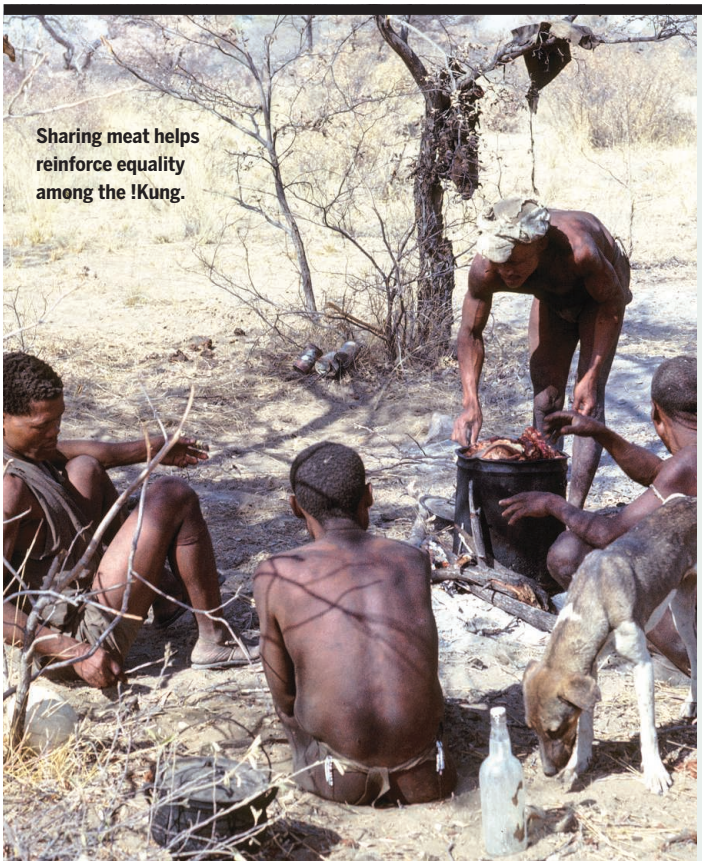
Prentiss's findings are raising questions about when Keatley Creek's elites first emerged. So Suzanne Villeneuve, project director of the Keatley Creek Archaeological Research Project at SFU, and her team are now excavating and analyzing new housepit data to re-evaluate the site's dating. But Hayden insists that in historical hunter-gatherer cultures both in Canada and abroad, aggrandizers build surpluses, amass wealth items, and hold feasts only when food is abundant. "When food is in short supply, no one tolerates other people hoarding," Hayden says. "The majority simply take what they need because their lives depend on it. Scarcity breeds revolts and demands for more equality." In contrast, when times are good—for example in a booming modern economy like China (see p. 832)—people seem more tolerant of inequality.

HOW THE RICH GET RICHER. While archaeologists on Canada's Northwest Plateau probe the origins of wealth, other researchers are examining how it is passed on from generation to generation, perpetuating inequality. Economist Samuel Bowles of the Santa Fe Institute and anthropologist Monique Borgerhoff Mulder of the University of California, Davis, led an international team that studied inheritance in four types of societies: hunter-gatherers, pastoralists, horticulturalists who planted hand-tended gardens, and agriculturalists who used more advanced technology such as plows or organic fertilizers to boost crop yields.

Using historical and ethnographic data on 21 populations around the world, the team examined three kinds of wealth: material riches such as real estate, embodied wealth such as physical strength, and relational riches such as the number of people in a person's social network. They conducted statistical analyses to determine how much of each type of wealth was transmitted. "We counted things like the number of cattle people had and their sons had, and we did the same thing for forms of wealth used by hunters, such as grip strength, which measures how strong your forearms are," Bowles says.

They found that only material forms of wealth, such as land and livestock valued by farmers, were readily handed down to children. "I can pass on my cows to my sons, but if I have some phenotypic trait that accounts for my income, like being physically strong, it is less likely that I will pass that on to my offspring," says Bowles, whose team reported their findings in a paper in *Science* in 2009 and a series of papers in *Current Anthropology* in 2010.

The team found an unexpected difference between horticulturalists and agriculturalists. Horticulturalists, who tended widely scattered fields of domesticated plants, scored little higher than hunter-gatherers in the overall transmission of all forms of wealth. That suggests that just having domesticated crops wasn't enough to fuel enduring inequality. Farmers who practiced intensive agriculture and boosted yields in regions where arable land was scarce readily passed



Sharing meat helps reinforce equality among the !Kung.

Our egalitarian Eden

By **Elizabeth Pennisi**

Seek the richest family in a traditional camp of the Ju/'hoansi/!Kung people of the Kalahari Desert in Africa, and you will almost surely fail. There is no such thing. These hunter-gatherers traditionally moved periodically and had few possessions. What they had, they shared—food, weapons, property, even territory. The poorest looking hut in a camp likely belonged to the leader, explains anthropologist Richard Lee, a professor emeritus at the University of Toronto in Canada, because leaders try to avoid looking superior.

Many anthropologists think this egalitarian lifestyle was an essential feature of hunting and gathering societies. In contrast with both today's titans of Wall Street and the alpha males of the great apes, people in these societies "had an ethic of sharing that was central to their way of life," Lee says. "No one takes precedence over anyone else."

Our species has lived as hunter-gatherers for more than 90% of our history, Lee notes. Today's economic inequality goes back thousands of years (see main story, p. 822) but in evolutionary time it is relatively recent. Although some of our great ape cousins and arguably our ape ancestors lived in sometimes brutal hierarchies, humans adopted an egalitarian way of life for all but the last 10,000 years.

Achieving and sustaining such egalitarianism is not easy, anthropologists say. "The only way you can avoid hierarchies

PHOTOS: (TOP) SUZANNE VILLENEUVE; (BOTTOM) RICHARD B. LEE

Published by AAAS

down their wealth. These farmers could control access to their fields, protect them, and leave them to their heirs, Bowles says.

He thinks resource concentration is a key factor in explaining inequality among both farmers and the ancient salmon fishers. "Those societies had in a natural state exactly the same kind of concentration of resources that farming made possible everywhere," Bowles says. "Farming vastly increased the productivity of small patches of land and a small number of animals." People who owned particularly fertile patches of farmland had a good shot at becoming wealthy and passing on that wealth, in part because the land was defendable against others.

As agricultural societies developed, so did more elaborate hierarchies, evolving into hereditary chiefdoms and eventually kingdoms. In these complex societies, chiefs and kings came up with new strategies for amassing surpluses and concentrating wealth and power. Many chiefs created economic bottlenecks in trade routes, noted economic anthropologist Timothy Earle of Northwestern University in Evanston, Illinois, in a 2011 paper in *Social Evolution & History*. These leaders then collected payments from merchants for safe passage and used the surplus to finance specialized warriors to defend and extend their rule. Material culture also became ever more sophisticated, multiplying into innumerable kinds of highly concentrated and easily transmitted forms of wealth, from copper ingots to gold jewelry. All of these trends led to ever greater levels of inequality.

By the time of the Romans, a yawning gap separated rich from poor. Historian Walter Scheidel of Stanford University in California and biblical studies scholar Steven Friesen of the University of Texas, Austin, used historical records to calculate the Gini coefficient—a standard measure of inequality in modern societies—for the Roman Empire. The coefficient ranges from 0, in which everyone shares equally, to 1, in which one wealthy person has everything and the rest have nothing. The Roman Empire's Gini for income was about 0.43, the pair reported in 2009 in *The Journal of Roman Studies*—close to the 0.49 for pretax income in the United States in 2010. In fact, Rome's super-rich had wealth on the scale of today's billionaires. The income of the wealthy Roman triumvir Marcus Crassus equaled about $1 billion per year today, reported economist Branko Milanovic, of the Luxembourg Income Study Center of the City University of New York in New York City, and his colleagues in a working paper in 2007; that's not quite up to Bill Gates's more than $2 billion per year.

In today's complex world, there's no going back to the egalitarianism of some hunter-gatherers. And yet studies of prehistory may offer some hope for lessening the grip of the 1%, Bowles says. As societies move toward knowledge-based economies, wealth increasingly reflects know-how, social skills, and networking—factors that cannot be transmitted across generations as easily as plots of land or stock portfolios, he says. "So I think the long-term possibilities for a more egalitarian future are certainly there." ■

is that you work very hard to head [them] off," says Christopher Boehm, a cultural anthropologist at the University of Southern California in Los Angeles. Like all animals, humans are born unequal—some run faster, plan better, or make friends more easily than others. And we instinctively want the best for ourselves, even at the expense of others. That sets the stage for some to dominate.

To find out why humans shunned hierarchies for most of our history, anthropologists have studied living hunter-gatherers around the world, including Native Americans and the Ju/'hoansi/!Kung. Iconic studies of these societies show that boasting and other self-aggrandizing behaviors are not allowed. Offenders are teased, ignored, banned from camp, or, in extreme cases, killed. Humility, humor, and strict protocols about distributing meat helped keep people on an even footing, says Boehm, who has surveyed the ethnographic literature. For example, !Kung people traditionally downplay their accomplishments: A hunter will say he's caught only a small skinny animal, even if it's big and meaty, and his comrades will agree. "You have to demean yourself," Boehm says.

Lee and his colleagues, who observed the Ju/'hoansi/!Kung for years, found that to counter differences in hunting prowess, men exchange arrows before they hunt. The owner of the arrow, not the bowman himself, gets the credit and decides how to distribute the meat while everyone looks on.

Other traditional societies have similar customs, Boehm found in an unpublished analysis. Of today's 330 foraging societies, he examined 56 that live in conditions resembling those of Paleolithic hunter-gatherers. Among these groups, having someone other than the successful hunter distribute the meat "is universal," he says.

Several factors prevented the concentration of wealth and reinforced cooperation in these groups. Scattered, unpredictable food resources encouraged nomadism, a lifestyle that ensured no one accumulated many material goods. Cooperative hunting, in turn, yields more than enough meat to go around, and groups that shared equitably were at an evolutionary advantage because all of their members were strong enough to be good hunters or fighters should the need arise, says economist Samuel Bowles of the Santa Fe Institute. And cooperation is self-reinforcing: Sharing the spoils promotes further cooperation and self-sacrifice (*Science*, 4 September 2009, p. 1196). "Inequality may be the enemy of cooperation," Bowles says.

Boehm also stresses that the behavior of the lower ranks can help quell dominant individuals. In a seminal paper back in 1993, he proposed that several low-ranking men could band together to oust a too-greedy or violent alpha male. He confirmed this idea based on a survey of behaviors of 48 seemingly egalitarian groups. This isn't a lack of dominance but a reversal of it, he says: "An apparent absence of hierarchy was the result of followers' dominating their leaders rather than vice versa," he wrote.

Over the millennia, as societies began accumulating surplus food and goods, the inequality held in check for thousands of years re-emerged. Yet humans "have experienced an extraordinary amount of equality," Bowles says. "The question is whether conditions now will allow us to experience the egalitarianism of the past." ■

Published by AAAS

Thomas Piketty foresees a continued rise in inequality.

Downloaded from www.sciencemag.org on May 26, 2014

Tax man's gloomy message: the rich will get richer

With a massive database of income tax records, a French superstar challenges conventional wisdom on inequality

By **Eliot Marshall**

On a rain-soaked 15 April—U.S. income tax day—Thomas Piketty arrived early at a think tank in Washington, D.C., with a stack of boxes containing his new book, *Capital in the Twenty-First Century*. The 43-year-old French economist had come to talk about his radical ideas. The room was packed with young policy wonks and grizzled journalists, lawyers, political aides, and even a former U.S. representative, David Obey, who once chaired a key spending committee in the House of Representatives. Piketty, who sometimes apologizes to audiences for his strong accent, presented his data and predictions for capitalist economies and gently brushed aside criticisms. Dressed in a rumpled jacket and open collar, he looked a bit like a busy graduate student rather than a world-famous economics guru. By the time he was done, his supply of books had sold out. Young and old were lining up for an autograph.

Scenes like this played out many times this spring as Piketty toured North America. Amazon.com now ranks his book its number-one best-seller, and the publisher, Harvard University Press, reports that the first-year sales of *Capital* are more than for any title in its 101-year history.

Despite Piketty's popularity, his message is harsh. He labels as "a fairy tale" the long-accepted idea that wealth and income will be more evenly distributed within nations as they develop, and suggests that even the best run capitalist economies concentrate riches at the top. The reason: In the long run, he says, the return paid to owners of capital is higher than the rate of economic growth.

These provocative conclusions are based primarily on a huge database of tax records that Piketty and a team of 30 researchers around the globe have assembled from more than 20 countries, including the United States. From atop this mountain of data, Piketty is able to offer a 2-century retrospective view of capitalism and make predictions about its future. The database is, Piketty writes, "the largest historical database concerning the evolution of income inequality."

It is also the cornerstone of his credibility. Even economists who disagree with Piketty's message acknowledge the importance of his data. The chair of Harvard University's economics department, N. Gregory Mankiw, called it "[t]he best data we have on the upper tail of the income distribution" in a 2013 essay called *Defending the One Percent* that runs counter to Piketty's views.

Piketty began this project in the late 1990s, when he dug into old tax records for a book on the distribution of wealth in France. He had received his Ph.D. in 1993 at age 22 from the School for Advanced Studies in Social Sciences (EHESS), then moved to the Massachusetts Institute of Technology (MIT) as an assistant professor of economics. But he quit 3 years later—"I did not find the work of U.S. economists entirely convincing," he writes in *Capital*. Back home in Paris, he is now an economics professor at EHESS and the Paris School of Economics.

Piketty unearthed French data on wealth going back to the 1789 French Revolution, as well as a century's worth of income tax data that hadn't been analyzed systematically. Many such records have been ignored, he writes: "The historical and statistical study of tax records falls into a sort of academic no-man's-land, too historical for economists and too economic for historians."

PHOTO: AUDREY CERDAN

Published by AAAS

Extending the tax data analysis to Britain and the United States, Piketty teamed up with economists Anthony Atkinson of the University of Oxford and Emmanuel Saez of the University of California, Berkeley. Saez and colleagues gained rare direct access to U.S. tax records, enabling them to include new data from the world's top economy in the expanding database (see p. 836).

The team sifted through 100 years of tax records and stored standardized data in a resource they call the World Top Incomes Database (topincomes.g-mond.parisschoolofeconomics.eu). In addition to the more than 20 nations now represented, the website claims that about 45 more are "under study" and will be added when the work is complete.

In his book, Piketty combines income, inheritance, and national wealth data to reach a striking conclusion. Capitalism concentrates riches at the top of society, Piketty argues, because the rate of return to capital (labeled r) is higher than the overall rate of economic growth (labeled g) over the long run. This simple formula ($r > g$) means that families who own capital tend to acquire more and more wealth.

This pattern broke down, Piketty concedes, in the mid-20th century. Two world wars and the Great Depression destroyed a huge swath of Western wealth. Following World War II, as nations began to rebuild, they distributed the new wealth more equitably, Piketty finds. Aid programs and progressive tax schemes also decreased inequality, while a concurrent rise in industrial productivity and a population boom boosted economies and benefited the middle class. This golden era lasted roughly from 1950 to 1980, triggering a pause in the concentration of wealth.

Piketty insists the pause was an aberration. Capital resumed its dominant place in the 1980s, and wealth is again being concentrated at the top of society, he demonstrates. Today, Piketty says, inequality in the developed economies and particularly in the United States has reached an "extreme" point that could lead to "terrifying" disparities in the future and threaten democracy.

The concentration of wealth will continue, Piketty says, because economic growth is likely to be no more than 2% a year, limited in part by a widely predicted decline in birth rates. He predicts that the rate of return for capital will remain about what it has been historically, 4% to 5%. "It would be an incredible coincidence if the number of children we have and the number of innovations we make push the growth rate up" enough to counterbalance the return on capital, he says. "There's no logical or historical reason why this should happen," he adds in an e-mail.

Piketty's analysis offers "a new way of looking at the functioning of a capitalist economy."

Branko Milanovic,
Luxembourg Income Study Center

These ideas are important, said economist Robert Solow, a 1987 Nobel Prize awardee and professor emeritus at MIT, who spoke after the first of Piketty's several tax day talks and, in a book review, called his work a "new and powerful contribution" to economics. Economist Branko Milanovic, a former World Bank researcher on inequality now at the Luxembourg Income Study Center of the City University of New York in New York City, says Piketty has pulled several strands of research together into a single framework that offers "a new way of looking at the functioning of a capitalist economy."

In France, Piketty (*far left*) endorsed left-wing political candidates like Ségolène Royal because other choices were worse, he says.

PHOTO: GAMMA-RAPHO/GETTY IMAGES

Critics are sharpening their knives, however. Economist Kevin Hassett of the conservative American Enterprise Institute in Washington, D.C., for example, questions Piketty's reliance on income tax data. Hassett argued at one of the tax day talks that Piketty's approach fails to count the billions of dollars in government "transfers" that flow to citizens for benefits such as medical care and income support. Hassett prefers to rely on U.S. economic consumption data from household surveys run by the U.S. Bureau of Labor Statistics, which he says reveal spending in lower income households that may not even file tax returns. Those surveys don't show an alarming disparity between wealthy and nonwealthy Americans, Hassett says.

Piketty maintains that tax records give a truer picture of wealth and income at the top. On surveys, people tend to understate their income, while taxpayers are likely to be more truthful because they can be punished if they lie, he says. As for transfer payments, Piketty suggested that these funds increasingly pay for health care in the United States. With a smile, he said it would be interesting to know whether such transfers do more to benefit individuals or health care providers.

But Piketty's remedy for inequality—a universal tax on wealth that takes little from the bottom of the scale and a lot from the top—has drawn fire from left and right, in part because it would be so difficult to impose. Legislators don't seem ready to try it, doubters point out, and if they did, the super-rich could threaten to move to more tax-friendly climes. At Piketty's first stop, several economists suggested what they consider better and more achievable ways to counter inequality, such as investing in public education, job creation, and infrastructure. At another seminar, economist Dean Baker, co-director of the Center for Economic and Policy Research in Washington, D.C., suggested other remedies. He would tax financial transactions, limit patents to reduce drug company profits, break up monopolies, and more.

Piketty agreed with those suggestions but said they are not enough to change the fundamental dynamic of $r > g$. But as the first seminar on tax day came to a close, Solow and others mused on the implausibility of Piketty's scheme—requiring all the world's nations to agree in concert to boost taxes on their wealthiest citizens. Sounding a "pessimistic note," Solow reminded Piketty and the audience that the United States "is a country that can't even sustain an inheritance tax."

Piketty anticipated the pessimism. "In 1900," he said, "most people would have said a progressive income tax would never happen." But it happened. ■

Published by AAAS

Physicists say it's simple

If the poor will always be with us, an analogy to the second law of thermodynamics may explain why

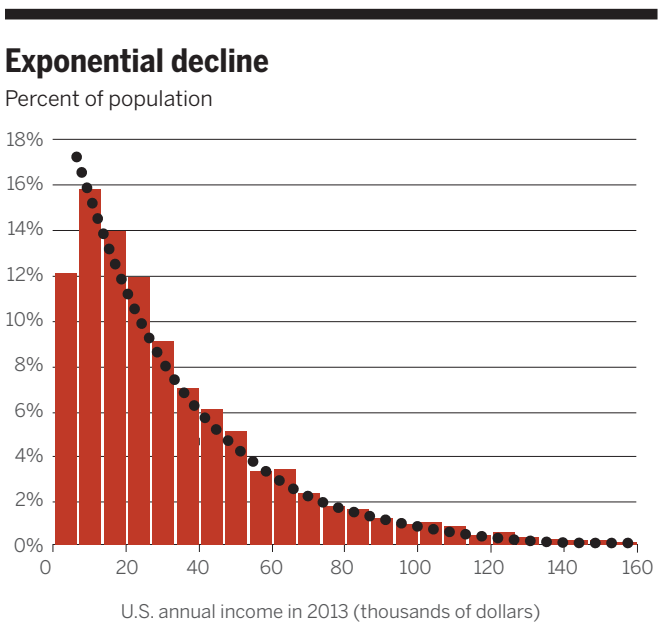
By **Adrian Cho**

The basic inequality that plagues economies the world over may have a simple explanation—at least, according to physicists who've turned to economics. Pick a country, they claim, and you'll find multitudes of people who earn next to nothing, a few who rake in plenty, and a distribution between the extremes that falls exponentially as income increases (see figure). That distribution applies to all but the very rich, they say, and it arises from an analogy to the concept of entropy, a measure of disorder in a physical system such as a gas. Just as a gas evolves to a state of maximum entropy, they argue, random churning in the economy ensures that the income distribution naturally tends to this inequitable form.

The argument suggests that although social and economic policy can help individuals edge ahead or perhaps boost everybody's fortunes, nothing short of radical intervention can overcome the forces of randomness and transform the lopsided distribution. An equal sharing of income, in this view, is as likely as the air in your office collapsing into your empty coffee cup. The reasoning is "not very close to the thinking of economists, but it's pretty persuasive," says Thomas Lux, an economist at the University of Kiel in Germany. But Frank Cowell of the London School of Economics and Political Science says "I'm extremely skeptical" that the argument provides any insight into the economy.

The argument builds on the century-old kinetic theory of gases, in which physicists asked: What is the most probable distribution of the energies of the molecules in a gas? That might seem impossible to determine without tracking exactly how the molecules ping off one another. But the puzzle can be solved by simply counting the ways the gas's overall energy can be divvied among the molecules—a number that defines the gas's entropy. The most likely energy distribution is the one that can be achieved by the most combinations of individual molecular energies. That turns out to be essentially an exponential distribution, with lots of molecules of low energy and a few with high energy.

Victor Yakovenko, a theoretical physicist at the University of Maryland, College Park, applies the same reasoning to income. Suppose you randomly divide $500 million in income among 10,000 people. There's only one way to give everyone an equal, $50,000 share. So if you're doling out earnings randomly, equality is extremely unlikely. But there are countless ways to give a few people a lot of cash and many people a little or nothing. In fact, given all the ways you could divvy out income, most of them produce an exponential distribution of income. So that's what you end up with, even if you start with a different pattern and let random economic activity take over, Yakovenko and a colleague argued in 2000 in *The European Physical Journal B*. "The exponential distribution is what you would call natural inequality—what you would get from entropy," Yakovenko says.



NATURAL INEQUALITY. Econophysicists say the income distribution is, inevitably, a decreasing exponential with few winners and lots of losers.

In 2001, Yakovenko and a colleague argued using tax data that income in the United States and both income and wealth in the United Kingdom follow exponential curves, as they reported in *Physica A*. Such curves also fit income data from Japan, Sweden, and the European Union, others have shown. "I don't know any other place in economics where the theory is so close to reality," says Mauro Gallegati, an economist at the Polytechnic University of Marche in Ancona, Italy.

Some economists are unimpressed. Economists gave up explaining the exact shape of the income distribution decades ago, Cowell says. Now, he says, economists estimate inequality using only raw data, and don't depend on knowing the distribution's precise mathematical form.

An exponential distribution also predicts fewer superrich people than are found in most economies, and the entropy argument does not explain the balance of wealth between the handful of superrich and the masses, says James Foster, an economist at George Washington University in Washington, D.C. The idea also fails to account for well-established correlations between income and education, race, and other factors, Foster says, and so is unhelpful in making policy.

Such criticism misses the point, Lux says. An individual's income can be highly correlated with other factors, he says, even as the distribution of income in a population as a whole appears random—just as the trajectory of a molecule in a gas is in principle predictable even as the motion of all the molecules as a group appears random. "That this is not a contradiction is something that we have to communicate to [traditional] economists," Lux says.

Even the theory's supporters say it must be fleshed out. It assumes that, at least over short times, the total amount of income is fixed, notes Duncan Foley, an economist at the New School for Social Research in New York City. If that's true, economists need to explain how that constraint arises. Still, Foley says he finds the physicists' perspective compelling, if bleak. In trying to achieve a more equal income distribution, "you're kind of fighting against the second law of thermodynamics," he says, "which as we know is generally a losing battle." ■

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

Uninsured people wait for basic health care at a free clinic in Los Angeles.

PHOTO: JOSEPH SOHM/VISIONS OF AMERICA/CORBIS

Can disparities be deadly?

Controversial research explores whether living in an unequal society can make people sick

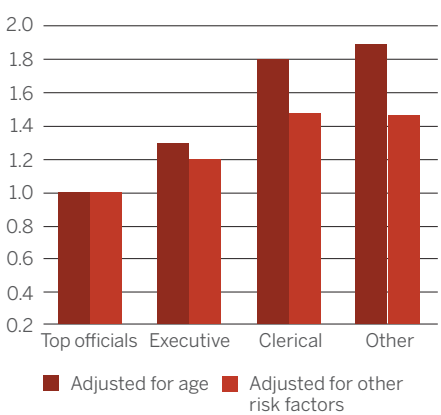
By **Emily Underwood**

Whitehall street, just south of Trafalgar Square in central London, is the heartbeat of the British government. Generations of workers in the highly stratified British Civil Service have marched to work each day in the government offices lining the road, with top bureaucrats working and living in palatial brick mansions built for aristocrats. Over the years, the denizens of Whitehall have fallen prey to the ills of the modern world: Their arteries have filled with fatty plaque; their blood sugar has spiked from diabetes; their lungs have been damaged by emphysema. And with surprising and troubling frequency, lower ranked workers have died earlier from these ailments than have their superiors.

To find out why, thousands of these civil servants, from typists to top officials, have gone to nearby medical clinics to have blood drawn, fill out questionnaires about how much they exercise and smoke, and don scratchy paper gowns for physical exams. Last year marked the 11th wave of data from this ambitious study, which has run for roughly 40 years and sparked an entire research program on the contentious question of whether being low-ranked can make you sick.

Deaths by rank at Whitehall

Relative rate of death over 25 years



Source: Marmot, 2000

HEALTHY AT THE TOP. In the long-running Whitehall studies, civil servants at every occupational grade live longer than their inferiors.

All agree that compared with the wealthy, poor people are less healthy. A child born in Norway can expect to live roughly 30 years longer than one born in Afghanistan. In the United States, on average, people in the highest income group can expect to outlive those in the lowest income group by more than 6 years. Preventable illnesses caused by poor nutrition and lack of education and care account for much of the disparity. Investing in health care and making it widely available can boost the health of those at the bottom. Redistributing wealth to the lower end of the curve helps, too. One simulation by researchers at the University of Otago, Wellington, for example, showed that shifting New Zealanders' incomes toward the mean income by 10% would save about 1100 lives per year.

But epidemiologist Michael Marmot of University College London (UCL), who leads the Whitehall study, argues that there's more to health than money alone. On the basis of his own and other studies, Marmot argues that hierarchy itself is a threat to health, with low-ranking individuals getting sicker and dying younger than higher-ups in part because of the sheer stress of being low on the social ladder.

Some public health experts say their own studies bear out Marmot's claim, but others think that confounding factors could easily account for the Whitehall findings. To these skeptics, focusing on hierarchy distracts from the real challenge of providing better health care to the poor. One

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

question is how being low on the social ladder matters to your health. Another is whether a society's health is worse when the rungs are far apart. The issue bubbles just below the surface in policy debates and has erupted recently in impassioned editorials. Some argue, paraphrasing Roman philosopher Seneca the Younger, that "to be poor in a wealthy society is the worst kind of poverty." But will it send you to an early grave?

WHO DIES FIRST? The Whitehall studies began as a simple search for heart disease risk factors. In the late 1960s, heart disease was thought to prey disproportionately upon upper-class, white-collar workers, because of their high-stakes jobs and type A personalities. After following more than 17,000 40- to 64-year-old male Whitehall employees for a decade, however, researchers at UCL found the opposite. During that period, 1652 men died, and men of the lowest rank were nearly four times more likely to die prematurely of heart disease than those in the highest tier, even though all had free health care.

In 1985, Marmot and his colleagues set out to determine why this might be so. They recruited a second cohort of more than 10,000 white-collar civil servants, including women, and found the same patterns of illness and mortality by rank, with some variations between men and women. Marmot started asking participants to fill out ever more extensive questionnaires, including not only their past medical history and health behaviors, but also their job demands, levels of stress, and social networks and support. As the data rolled in, he found that the psychological effects associated with status and job rank consistently predicted employees' health better than did their salaries, or even health-related behaviors like diet and exercise.

PODCAST

To hear a podcast with author Emily Underwood, see http://scim.ag/pod_6186.

Based on these findings, Marmot developed a theory: When a population moves beyond abject poverty, rank in the social hierarchy, not income, ultimately determines how healthy people are. Some animal studies suggest how status stress might "get under the skin," as epidemiologists put it: Low-ranking baboons and macaques can develop higher levels of stress hormones, atherosclerosis, and hypertension when subject to a dominant male's whims.

If Marmot and others are correct, simply shifting money to the poor won't be enough to boost their health. The health gradient among people who are not poor shows that it's "not only about poverty—we've got to improve society," he says.

From the dangerous streets of Chicago's South Side to the neatly tended homes of a Helsinki suburb, the link between low status and poor health has now been found in many different countries and contexts, says Ichiro Kawachi, a social epidemiologist at Harvard University. "The higher up the gradient you are, the longer you tend to live and the healthier you tend to be," he says.

SCALING UP. More controversial is whether overall population health is worse in more unequal societies. In 2009, Kawachi published a meta-analysis of epidemiological studies linking inequality and health in about 60 million people around the world. He and his colleague found an excess mortality risk of 8% for every 0.05 unit increase in a country's Gini coefficient, the most commonly used statistical measure of the gap between rich and the poor (see p. 818). Although such an effect may seem modest, when extrapolated to the global population it suggests that leveling income inequality could help avert more than 1.5 million deaths per year worldwide—assuming the effect is causal, he says.

In the United States, Kawachi and public health researcher S. V. Subramanian, also at Harvard, have found that income inequality is also strongly correlated with rates of infant mortality, heart disease, and several health conditions across many states and cities, even after controlling for variables such as absolute income in each location, race, age, and education. Measures of social cohesion such as trust also appear to track with inequality, he says. In one of America's most unequal states, Louisiana, for example, people are far more likely to agree with the statement that "most people would try to take advantage of you if they got the chance."

Based on such studies, Kawachi and others argue that inequality breaks down social values, such as trust and support, that protect against both physical and mental illness. In a recent op-ed in *The New York Times*, epidemiologists Richard Wilkinson and Kate Pickett of the University of York in the United Kingdom took the argument even further. They claim that the reason more unequal countries like the United States see higher rates of schizophrenia and other mental illnesses is because inequality causes "social corrosion" that damages the individual psyche.

Others aren't convinced. John Lynch, an epidemiologist at the University of Adelaide in Australia, says that although he started out as a "true believer" in the income inequality hypothesis, a string of negative and equivocal studies turned him into a skeptic. Back in a 2004 paper, for example, Lynch and colleagues reviewed 98 cross-national studies and found "little evidence" of a consistent link between income inequality and health, although the United States displayed a more robust association than others. Working on well-established public health goals such as reducing smoking and improving the living conditions of the poor will likely have more direct health impacts than targeting relative income gaps, he says.

Even if the correlations Kawachi and others have found hold up, there's no strong evidence that income inequality, per se, is directly damaging people's health, says Angus Deaton, an economist at Princeton University. In American cities and states where there are large proportions of African-Americans, for example, racism, poor health care, and political disenfranchisement could just as easily explain poor health outcomes as income inequality, he says. Deaton argues that extreme in-

The gap between what people desire—like these luxury cars in South Africa—and what they can afford may be a source of unhealthy stress.

PHOTO: SCHALK VAN ZUYDAM/AP/CORBIS

Published by AAAS

equality is a risk to health chiefly because it skews politics to favor the rich and powerful in society. "I get angry" over Wilkinson's claim that psychological stress is the primary culprit, because it completely deflects from the real issues," he says.

CAUSE OR CORRELATE? In 2011, Princeton University economists Christina Paxson and Anne Case found another potential explanation for the correlations between rank and health. They reexamined data from the Whitehall II study and found that adults who were healthier as children started at higher grades in the Civil Service, were promoted to higher positions, and maintained better health throughout their lives. Occupational rank was a marker, but not a cause, of poor health in adulthood, Paxson and Case concluded.

Many economists agree that people's health influences their status, rather than the other way around, says Dalton Conley, a sociologist at New York University in New York City. "Economists tend to think that your health predicts where you are on the social scale," he says. If you're sick a lot and miss school, for example, you won't do as well in the labor market. He notes that the initial Whitehall studies also didn't take into account "very controversial" questions about the extent to which genes determine later health and wealth.

Marmot says he's now persuaded that genetics and early-life experiences do play some role in adult health and socioeconomic rank. Still, neither can fully account for the huge difference in mortality and morbidity among Whitehall's occupational grades, he says. Pointing to more than 100 studies based on Whitehall data, Marmot maintains that stressors such as lack of control and harassment at work fall hardest on low-ranking workers and take a fatal toll.

Causality lies at the heart of the issue, so scientists are now looking for mechanisms that could link inequality and health. Biocultural anthropologist Elizabeth Sweet of the University of Massachusetts, Boston, notes that any causal link assumes that people know their place in the hierarchy. "We don't always walk around with our salaries tacked on our foreheads, so how do we get the information to make that social comparison?"

Kawachi suggests that for Americans, their own aspirations may provide the point of comparison. Even though an American born in the bottom fifth of the income distribution has only about an 8% chance of rising to the top fifth—half the likelihood of a child born in Denmark—more than 90% of Americans still believe in the American dream, he says, and the collision of their ideal with reality may take a toll on health. "When you work hard on the assumption that we're building a meritocracy, then fail," the resulting depression and frustration may contribute to the country's high rates of drug abuse, suicide, and violence, he says.

Similarly, Sweet hypothesizes that the gap between the standard of consumption one identifies with success and one's ability to meet that ideal produces measurable stress and health impacts. Through extensive interviews, she and others collect information about the cultural norms of material success in a given community. In rural Brazil, being successful might mean owning a TV, whereas in U.S. suburbs it might mean having the "right" brand of jeans or cellphone. The researchers measure the degree to which an individual is able to "keep up with the Joneses," and compare that with health indicators such as the amount of cortisol in saliva, a marker of stress.

In a study of African-American teenagers in Chicago, Sweet demonstrated that teens who could easily conform to their communities' "ideal" level of consumption had lower blood pressure than teens who couldn't meet those norms. But if they managed to get expensive sneakers and brand-name clothes even though they couldn't really afford them, the students had abnormally high blood pressure. In Sweet's view, this suggests that the tension caused by the gap between what people need and what they can afford can affect health. But she and Kawachi admit that the causal link is tenuous.

Back at Whitehall, civil servants are still striding into work every morning. Some of the original participants have retired and moved to the suburbs. Many others have died, leaving behind reams of data about what they ate, if they exercised, and how often they felt lonely. Marmot and others have produced more than 500 papers based on these workers' experiences and continue to churn out dozens each year. To fully explain the links between inequality, rank, and health, however, may take hundreds more. ■

Published by AAAS

While emerging economies boom, equality goes bust

Inequality spikes in developing nations around the world

By **Mara Hvistendahl**

SHANGHAI, CHINA—Starting in 1949, the Communist government led by Mao Zedong waged war on inequality of all kinds. The administration seized property from privileged classes, imprisoned intellectuals, and appointed teams of workers to run universities. The revolution upended the class structure, and the party campaigned against inherited wealth and gender discrimination. By the time the Cultural Revolution ended and Mao died in 1976, the government had mandated a bland unisex style of dress and effectively abolished property ownership. Society had ostensibly been "leveled off," even if in practice the new system concentrated resources in the hands of party cadres.

Then, beginning in the 1980s, the country pulled an abrupt about-face. China reintroduced land rights, allowed foreign investment, and spurred private enterprise in a few designated areas. Inequality was no longer the enemy; in fact, the government signaled that it was to become the new norm. The reformist leader Deng Xiaoping disparaged Mao's egalitarianism as "everyone eating from the same big pot." Overturning that failed ideal would bring growth to everyone eventually, he suggested: "It is good for some people to get rich first."

Some people did. China now has more than a million millionaires and more than 200 billionaires. Although no country can quite match this meteoric rise, similar stories have played out across the developing world. For example, the Latin American middle class mushroomed from roughly 100 million in 2000 to about 150 million a decade later, according to the World Bank.

But a rash of new studies—based on longitudinal surveys, better cross-sectional data, and renewed attention from scholars—has also laid bare extraordinarily high levels of inequality in these growing economies. In China, the richest 10% now makes 13 times as much as the poorest 10%, compared with five times as much in the United States, according to data from the China Family Panel Studies, run by Peking University's Institute of Social Science Survey in Beijing. With economic development, "the rising tide has indeed raised all boats," notes University of Maryland, College Park, sociologist Reeve Vanneman. "But the big yachts have done better, so overall income inequality is increasing."

That's not what many 20th century economists would have predicted. In a 1954 speech at an American Economic Association meeting, economist Simon Kuznets proposed that the urbanization that accompanies development inevitably triggers a growing income gap, but that societies become more equal as they democratize and adopt social welfare programs. When inequality was plotted against income levels, Kuznets maintained, the relationship looked like an inverted U curve—first rising, then falling. He won a Nobel Prize for his work.

But his analyses were based on data from the United States, the United Kingdom, and Germany in the 20th century. Kuznets himself cautioned that the hypothesis needed further testing. "The Kuznets curve is a perfect example of taking trends observed in wealthy countries and projecting [them]

PHOTO: FRITZ HOFFMANN/IN PICTURES/CORBIS

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

as universal to the world," says Timothy Moran, a sociologist at Stony Brook University in New York.

In fact, the curve's predictions have not held up in many countries. Initial growth between the 1960s and 1990s in the East Asian "tigers"—Hong Kong, Singapore, South Korea, and Taiwan—did not yield a larger income gap. In other industrialized countries such as the United States, meanwhile, inequality is now rising, not falling.

A wave of longitudinal studies tracking income and other metrics has helped flesh out the picture in developing countries like Indonesia, South Africa, India, and China. Those studies reveal growing inequality, which itself may stymie further growth, because poor people without access to good education cannot contribute to economies to their full potential.

But a tour of emerging economies also shows that cultural factors influence how governments react, and whether citizens accept what seems to be an inevitable march toward greater inequality, or protest it.

INDIA: HOW UNEQUAL? India illustrates the daunting task of measuring income and wealth in emerging economies. Half of all households get some income from agriculture, and most receive income from more than one source. A farmer might collect wages or receive payments from a cousin in the city, while a wage earner might also keep farm animals. To capture all earnings, surveyors for the national India Human Development Survey—which examines 41,554 households across the country—personally ask participants about 50 separate indicators of income.

This herculean labor pays off, says Vanneman, a principal investigator on the survey, which is jointly administered by the University of Maryland and the National Council of Applied Economic Research in New Delhi. For example, one previously elusive indicator for India was the Gini coefficient, a common index of income inequality ranging from 0, in which everyone makes the same income, to 1, in which a single rich person would get a country's entire income. Government surveys based on expenditures and excluding income data had found figures in the 0.30s—below the level in the United States of 0.40. Such figures sparked "disbelief" among scholars, Vanneman notes: "Anybody who walks the streets of India cannot believe that inequality in India is as low as the common statistics suggest."

In 2010, the Indian survey found a Gini coefficient of 0.52—close to China's, which scholars most recently estimated at 0.55. At a time when attention is focused on inequality in the developed world, that's a sharp reminder that the worst inequalities are often in emerging economies (see map, pp. 820–821). Inequality in high-income countries "still falls well below levels found in low- and middle-income countries," Vanneman notes.

Laborers work on new construction in the booming city of Chongqing, China.

CHINA: SURFING A RISING TIDE. In China, the market reforms of the past few decades have yielded some spectacular successes, giving rise to the lucky billionaires and also lifting the standard of living for the middle class. Between 2004 and 2009, the percentage of Chinese owning color TVs shot up from 80% to 96% and the percentage owning refrigerators swelled from 37% to 54%, according to surveys by sociologist Martin Whyte of Harvard University and colleagues at Peking University's Research Center for Contemporary China.

Even so, the middle classes in China or India are "still rather poor within global comparisons," Moran cautions. And the dramatic boost in inequality in China now presents a powerful challenge to the Kuznets curve. In a paper published online last month in the *Proceedings of the National Academy of Sciences*, sociologists Yu Xie and Xiang Zhou, both of the University of Michigan, Ann Arbor, plot China's rising income inequality, represented by the average Gini coefficients found by seven independent household surveys, against a Kuznets curve.

Published by AAAS

In 1980, after the storms of the Cultural Revolution, China was well below the level of inequality predicted by the curve, with a Gini coefficient of merely 0.28. But in 2002, the country's Gini intersected the curve and then shot beyond it, Xie and Zhou found (see graph). The relationship between inequality and development in China looks more like a straight diagonal line than an inverted U, with no sign of flattening. Fearing a backlash, Chinese officials have suppressed publication of the Gini coefficient (*Science*, 31 May 2013, p. 1037) and challenged estimates of it that they consider high.

But in fact Chinese appear remarkably tolerant of income gaps. The 2004 round of the survey by Whyte and colleagues polled 3267 Chinese on their attitudes as well as their income. Although respondents valued equality and believed the national income gap was excessive, only 30% supported redistributing wealth from rich to poor. Asked why people are poor, 61% said a lack of ability was an important cause, far higher than in any other country.

In the next round of the survey, done in 2009, the researchers found that despite the rising Gini, even fewer respondents viewed existing inequality as excessive. The findings challenge the notion that "rising income gaps are a major, or even the primary, threat to social order and political stability in China," Whyte says. In a separate study, Xie, who also directs the Center for Social Research at Peking University, found that Chinese largely believe Deng's assertion that development and inequality are necessarily linked—even though economists have mostly disproven that statement.

In 2006, Xie and colleagues polled residents in six provinces, asking them to separately rate levels of development and inequality in five countries: Brazil, China, Japan, Pakistan, and the United States. For level of development, respondents came up with rankings that closely mirrored U.N. estimates. But their guesses for inequality were way off. Instead of corresponding to Gini coefficients for the various countries, respondents believed that the most developed countries have the greatest inequality. Thus, many Chinese view inequality as the price of economic growth and accept it "as a fact of life," Xie says. "That's why there's not as much resentment."

SOUTH AFRICA: ECHOES OF APARTHEID. Halfway around the world from China, South Africa faces similar economic challenges, but has a very different response, perhaps because the countries' starting points were so different. For decades under apartheid, black South Africans faced discriminatory barriers to mobility. As those barriers fell after 1994, expectations for a more level playing field soared. "This is the new South Africa," says Murray Leibbrandt, an economist at the University of Cape Town and a principal investigator on the South African National Income Dynamics Study. "There was almost this irrationality that things were going to be much better moving forward."

By some measures, things did get better: As in China, absolute mobility rose, and most people are better off economically than they were 20 years ago. The share of people living below the poverty line—defined as $60 a month—fell from 57% in 2006 to 46% in 2011. Some expected reshuffling occurred, as skilled black Africans moved up the ladder and low-skilled whites moved down.

But despite significant investment in education and a government vocally committed to fighting inequality, whites continued to earn more than blacks, and income became more concentrated in the top 10th. Between 1993 and 2008, overall income inequality actually increased, with the country's Gini coefficient rising from an already high 0.66 to a staggering 0.70, one of the highest in the world. Demand for highly skilled workers at the top increased, while black citizens at the bottom, burdened with poor education and health, remained relatively worse off. "The disadvantages of apartheid just linger," Leibbrandt says.

South Africans may be less accepting of inequality than the Chinese. In the past 5 years, millions of South Africans have taken to the streets to protest everything from high crime rates to a lack of affordable housing. About 91% say income differences are too large, and two-thirds say the country is going in the wrong direction, according to the South African Social Attitudes Survey.

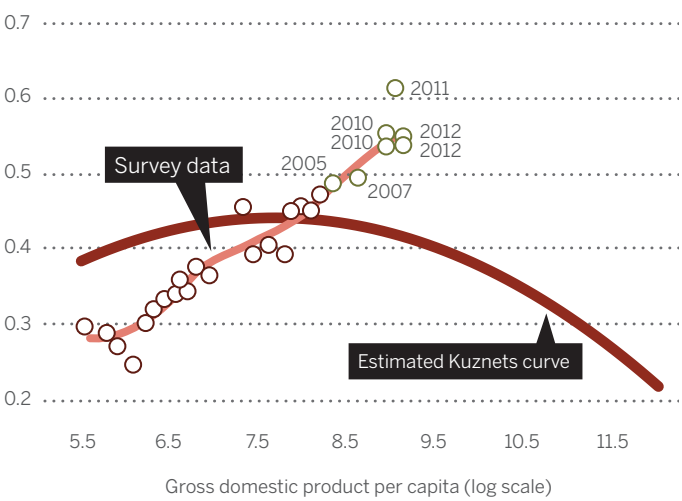
And yet, the hopefulness that dominated post-apartheid has not yet died. The survey also found that 42% of respondents believe that life will improve over the next 5 years. Given trends in absolute mobility, they are probably right.

LATIN AMERICA: SEEKING THE SWEET SPOT. In Latin America, as in South Africa, a colonial past primed nations for inequality. Institutions established by colonial governments allowed elites to consolidate power and excluded indigenous and black populations from land ownership, education, and politics. Thus, the region has historically had very high Ginis: 0.59 for Brazil in 1998 and 0.55 for Mexico in 1996, according to a recent working paper from the World Bank.

Added to this historically large gap between rich and poor is the fact that people born poor tend to stay poor. In Mexico, children of managers are a whopping 15.6 times more likely to hold on to their class status than to change it, according to data from the Mexican Social Mobility Surveys. Those are "near caste-like conditions," wrote sociologist David Grusky of Stanford University in California and colleagues in a working paper last fall. In the United States, by contrast, children of managers are only 2.3 times more likely to end up in the same class. In every category except farming, Mexicans are less mobile than Americans. New York University sociologist Florencia Torche has found a similar lack of social mobility in Chile.

Nevertheless, the gulf between income classes in Latin America has gradually nar-

Inequality and growth in China



Source: Xie and Zhou, 2014

STRAYING FROM THE CURVE. Surveys taken during the last 10 years show that as China continues its rapid economic growth, its inequality continues to shoot upward, in contrast to what a Kuznets curve would predict.

Published by AAAS

An elderly man holds out his begging cup in bustling Hong Kong.

PHOTO: AFP/GETTY IMAGES

rowed, resulting in an impressive decrease in inequality across the region. In Mexico, the Gini coefficient fell 0.07 units between 1996 and 2010, to 0.48. In Brazil, the Gini coefficient dropped 0.05 units from 1998 to 2009, to 0.54. Throughout the 2000s, Ginis fell in 13 of 17 Latin America countries for which the World Bank has reliable data.

So although the region is still battling inequality, such countries are now at something of a sweet spot, says Timothy Smeeding, an economist at the University of Wisconsin, Madison: As with the Asian "tigers" before them, the economy is growing, while inequality is falling. Even if it remains hard for people to move up relative to each other, many people are better off than before because absolute mobility is rising.

GROWTH FOR ALL. Policy measures helped achieve such "pro-equity growth," and scholars from countries like South Africa are studying how it was done. For example, the Brazilian government used grants to boost education. Average years of schooling shot up even among the poor. So when strong economic growth hit in the 1990s, marginalized citizens could get better jobs. From 2002 to 2009, the income of the bottom 10% grew at almost 7% a year, while that of the wealthiest 10% inched up by only 1.1% a year. The lesson for other developing nations, Leibbrandt notes, is the importance of job creation: Improvements in education and health may be good on their own, but they "don't narrow the income distribution until you get some feedback into the labor market."

Bucking theories put forward by proponents of the Kuznets curve, research now suggests that inequality may be a trap for developing countries. Far from boosting development, a large income gap can slow growth and stymie poverty reduction (see p. 851). In an entirely equal society, an increase in gross domestic product benefits everyone to the same degree, explains J. Humberto Lopez, an economist in the World Bank's Latin America and Caribbean region. In an unequal one, those at the top accumulate more income, leaving fewer dollars to boost households at the bottom. So to achieve the same reduction in poverty, highly unequal Brazil now needs to grow at least twice as much as a more equal country like Poland.

Growth suffers as well; in unequal societies, talented people born into poverty have fewer opportunities to contribute. "It's a perfect storm," Lopez says. "High inequality is bad for poverty, high inequality is bad for poverty reduction, and high inequality is not good for growth."

From Latin America's success at easing this trap and other cases, one thing now seems certain: Where inequality does decline, government involvement is key. Without substantial improvements in education and the social welfare system, "it's not natural" that inequality falls on its own, says Gan Li, an economist at Texas A&M University, College Station, and the Southwestern University of Finance and Economics in Chengdu, China.

In China, now that Deng Xiaoping's prediction about some getting rich first has come true, economists hope it, too, will adopt a more "pro-poor" strategy.

Over the past decade, China has boosted investment in social welfare programs, but it hasn't yet reached the spending necessary to begin leveling the playing field, Gan says. "China is at a crossroads," he says. The government could follow the status quo, or it could "follow many other successful countries' paths—and change the system." ■

Published by AAAS

Tracking who climbs up—and who falls down—the ladder

Researchers seek new ways to understand social mobility and opportunity in America

By **Jeffrey Mervis**

Back in 1982, songwriter Billy Joel started off a hit music video with images of happy GIs returning to Allentown, Pennsylvania, at the end of World War II and taking advantage of a booming economy. As the years pass, however, the mood darkens.

The song's message becomes clear when an apologetic foreman hands a pink slip to a young factory worker who's just ended his shift. There was a time, Joel croons, when "every child had a pretty good shot/ To get at least as far as their old man got/ But something happened on the way to that place."

The video pays homage to a once-thriving steel town battered by a spike in energy prices and the brief but painful recession in the early 1980s. It also questions the notion that talent and hard work can create a brighter future regardless of one's social standing at birth, suggesting that America is no longer the proverbial land of opportunity.

Three decades later, in the wake of an even worse recession, President Barack Obama appears to have come to a similar, although more nuanced, conclusion. America faces a "dangerous and growing inequality and lack of upward mobility that has jeopardized middle-class America's basic bargain—that if you work hard, you have a chance to get ahead," Obama warned in a December 2013 speech on the topic. It is, he said, "the defining challenge of our time."

The economists, sociologists, and demographers who study social mobility are thrilled that the president is shining a spotlight on their fields. At the same time, some grumble that Obama has muddied the water by conflating two different ideas—economic inequality and social, or intergenerational, mobility. The former is the gap in wealth between those in the penthouse and the poorhouse, which data show has reached high levels in the United States and is growing around the world (see pp. 826 and 838). The latter is whether children, as adults, "get at least as far as their old man got" on a ladder defined, not only by income but also by social factors including education, occupation, and where you live.

Researchers say the data aren't good enough to support what seems to be accepted wisdom—that social mobility in the United States is declining. In fact, many think our understanding of this complex phenomenon is embarrassingly thin. "President Obama has made a big commitment to equalizing opportunity," says sociologist David Grusky of Stanford University in California. "But he is making proclamations without any good evidence. Do we want to continue to do policy in the total, utter, complete blindness that the president now finds himself in?"

> *"Do we want to continue to do policy in the total, utter, complete blindness that the president now finds himself in?"*
>
> **David Grusky,**
> Stanford University

To truly assess social mobility, researchers need a deep understanding of why some people are afforded an easier route to the good life than others. That means assessing complex variables such as education, occupation, health, neighborhood, social networks, and more. Researchers must also find a way to track large swaths of Americans over successive generations of parents and children, a costly proposition when done by the traditional mode of surveying a sample population.

One new data set, based on an innovative use of U.S. tax data on a scale many times larger than has previously been done, illustrates the possibilities, showing for the first time how economic mobility varies by region of the country. The data and methods used wowed the field, and provide a starting point from which researchers might be able to tap into the records of other government agencies. But such studies will require the cooperation of those agencies and may not be easy to replicate.

ILLUSTRATION: DAVID PLUNKERT

WHAT WE'VE LEARNED. Interest in social mobility has waxed and waned with economic cycles like the one that inspired the Allentown video. But tracking it has always been a challenge. Most of the recent data on U.S. social mobility comes from detailed surveys of relatively small samples of Americans. These studies have yielded some useful information. For example, they've shown how discrimination hampers economic mobility among African-Americans, how a college education can propel a rise through the economic ranks, and that a disturbingly large fraction of Americans—43% in one recent analysis—who grew up on the lower rungs remain there as adults. These studies have also allowed U.S. researchers to become world leaders in survey design and analysis, and other countries now emulate their methods.

But many caveats still apply. "A lot of the data we have now [on social mobility] are essentially high-quality snapshots of individuals," notes economist Ron Jarmin, who manages research and methodology for the U.S. Census Bureau, the government's de facto chief statistical agency. So although researchers have managed to pull out some results on social mobility, the survey data "are not intended or designed for longitudinal analysis," Jarmin says. Without such analysis, it's impossible for researchers to say whether mobility is changing over time, and why.

VIDEO

For a related video on social mobility, see http://scim.ag/vid_6186.

Downloaded from www.sciencemag.org on May 26, 2014

Published by AAAS

That void leaves policymakers at a loss in knowing which factors to address—parental education, occupational status, social standing, housing, health, and so on—if they want to create a more level playing field.

WHAT'S ON FILE. Given limited resources, researchers say the best way to overcome these shortcomings is to supplement survey results with so-called administrative records. Government agencies are awash in data collected for other purposes that could inform studies of mobility. A student loan application, for example, includes information on income, siblings, parental level of education, high school record, neighborhood, and other bits of sociodemographic information. Several Western European countries link such records together and allow researchers to access them. That's one reason we know a lot more about social mobility in, say, Denmark and Sweden than in the United States.

No such massive, linked data sets exist in the United States. But last summer, in a study that has impressed their colleagues, a team of economists led by Raj Chetty of Harvard University and Emmanuel Saez of the University of California, Berkeley, took an important step in creating one.

The study, put up on a website in January, showed for the first time a link between children's hometowns and whether they climb or fall down the economic ladder as compared with their parents. One metric they used calculated the chances that a child born in the early 1980s (as it happens, when Joel wrote his hit song) into a family in the lowest fifth of the income distribution would jump to the top fifth by age 30. They found striking geographical differences. For example, in San Jose, California, nearly 13% made the leap upward, in contrast with only 4.4% of children from Charlotte, North Carolina. Allentown fell in the middle range, with a score of 8.3%.

The findings drew global media attention. But what created a buzz among mobility researchers was how the research team was able to gain direct access to previously unavailable U.S. tax records. Most researchers aren't allowed to handle tax data themselves. Instead, they submit their program for examining such data to the IRS, where agency employees run the program and ship researchers the output. Chetty and Saez, in contrast, effectively became part of the IRS workforce, submitting to training, background checks, and fingerprinting before being allowed to run the data and tweak their models themselves, according to IRS officials (http://scim.ag/sci_inequality).

In this way, Chetty and Saez created a database of children born in the early 1980s. They then followed the children into young adulthood and obtained their incomes at age 30. This data is now part of a larger database covering more than 20 countries, called the World Top Incomes Database, which French economist Thomas Piketty and colleagues have created to study inequality (see p. 826).

Researchers say that Chetty and Saez's data, if linked to administrative records at other U.S. agencies, could yield "a gold mine of information on intergenerational mobility" that researchers never thought would be available. "If you would have asked me a couple of years ago, I would have said that was a pipe dream," says Gary Solon, an economics professor at Michigan State University in East Lansing.

For all its richness, however, the study also comes with notable limitations. One is that a person's economic profile is still being formed at age 30. Collecting data on them at age 40 would yield "a more accurate assessment," Solon says. That could happen only if the IRS allows researchers the same access 10 years from now that it gave the Chetty-Saez team.

The study also captures a cohort of Americans at a particular time. Immigrants who arrived after the first snapshot of people was taken in the 1980s aren't included; people who don't earn taxable income or lack valid Social Security numbers are also omitted. In addition, the IRS data lacks information on race and ethnicity, which may be correlated with mobility rates.

But researchers see promise in Chetty and Saez's approach. "They linked tax records over time and showed some really cool stuff," Jarmin notes. "But what's really exciting is that they have also created a core infrastructure that will allow researchers to link to many other data sets that already exist."

When linked to administrative records, the usefulness of surveys by the Census Bureau and others "grows by orders of magnitude," Jarmin says. Linking Chetty's data on income to the Census Bureau's demographic and housing data, for example, could create a picture of quality of life—a rich, multidimensional view that goes well beyond simple income.

Of course, many hurdles remain before such linked data sets are common. Few agencies are set up to collaborate with other agencies, much less with outside researchers. And agencies would have to remain vigilant against any breaches in confidentiality that could lead to the identification of individuals or companies.

All the same, there are hints that the government is eager to maximize the value of the data it collects. In February, the White House Office of Management and Budget urged agencies to make better use of their troves of economic and demographic data—so long as their actions don't jeopardize privacy.

If the data were connected, mobility researchers would be on their way to a much better understanding of who rises and falls in American society—and why. It's definitely a step in the right direction, says the Census Bureau's Jarmin. "If I ran the world," he says of linking government and survey data, "that's what I'd do." ■

Published by AAAS

REVIEW

Inequality in the long run

Thomas Piketty[1]* and Emmanuel Saez[2]

This Review presents basic facts regarding the long-run evolution of income and wealth inequality in Europe and the United States. Income and wealth inequality was very high a century ago, particularly in Europe, but dropped dramatically in the first half of the 20th century. Income inequality has surged back in the United States since the 1970s so that the United States is much more unequal than Europe today. We discuss possible interpretations and lessons for the future.

The distribution of income and wealth is a widely discussed and controversial topic. Do the dynamics of private capital accumulation inevitably lead to the concentration of income and wealth in ever fewer hands, as Karl Marx believed in the 19th century? Or do the balancing forces of growth, competition, and technological progress lead in later stages of development to reduced inequality and greater harmony among the classes, as Simon Kuznets thought in the 20th century? What do we know about how income and wealth have evolved since the 18th century, and what lessons can we derive from that knowledge for the century now under way? For a long time, social science research on the distribution of income and wealth was based on a relatively limited set of firmly established facts together with a wide variety of purely theoretical speculations. In this Review, we take stock of recent progress that has been made in this area. We present a number of basic facts regarding the long-run evolution of income and wealth inequality in advanced countries. We then discuss possible interpretations and lessons for the future.

Data and Methods

Modern data collection on the distribution of income begins in the 1950s with the work of Kuznets (*1*). Shortly after having established the first national income time series for the United States, Kuznets set himself to construct time series of income distribution. He used tabulated income data coming from income tax returns—available since the creation of the U.S. federal income tax in 1913—and statistical interpolation techniques based upon Pareto laws (power laws) to estimate incomes for the top decile and percentile of the U.S. population. By dividing by national income, Kuznets obtained series of U.S. top income shares for 1913 to 1948.

In the 1960s and 1970s, similar methods using inheritance tax records were developed to construct top wealth shares (*2*, *3*). Inheritance declarations and probate records dating back to the 18th and 19th centuries were also exploited by a growing number of scholars in France, the United States, and the United Kingdom (*4*–*7*).

Such data collection efforts on income and wealth dynamics have started to become more systematic and broader in scope and time only since the 2000s. This is due first to the advent of information technologies, which allow much larger volumes of data to be collected and processed than were accessible to previous generations of scholars. The second reason for this time gap in using tax data is that most modern research on inequality has focused on micro-survey data that became available in the 1960s and 1970s in many countries. Survey data, however, cannot measure top percentile incomes accurately because of the small sample size and top coding. The top percentile plays a very large role in the evolution of inequality that we will discuss. Survey data also have a much shorter time span—typically a few decades—than tax data that often cover a century or more.

Kuznets-type methods to construct top income shares were first extended and updated to the cases of France (*8*, *9*), the United Kingdom (*10*), and the United States (*11*). By combining the efforts of an international team of over 30 scholars, similar series covering most of the 20th century were constructed for more than 25 countries (*12*–*15*). The resulting "World Top Incomes Database" (WTID) is the most extensive data set available on the historical evolution of income inequality. The series is constantly being extended and updated and is available online (http://topincomes.parisschoolofeconomics.eu/) as a research resource for further analysis.

Historical top wealth shares series have also been constructed with similar methods, albeit for a smaller number of countries so far, but with a longer time frame (*16*–*21*). Drawing on previous attempts to collect historical national balance sheets (*22*), long-run series on the evolution of aggregate wealth-income ratios in the eighth largest developed economies were established, some of them going back to the 18th century (*23*).

This Review draws extensively on this body of historical research on income and wealth, as well as on a recently published interpretive synthesis (*24*). We start by presenting three basic facts that emerge from this research program (Figs. 1 to 3), and then turn to interpretations.

Income inequality in Europe and the United States, 1900–2010

Share of top income decile in total pretax income

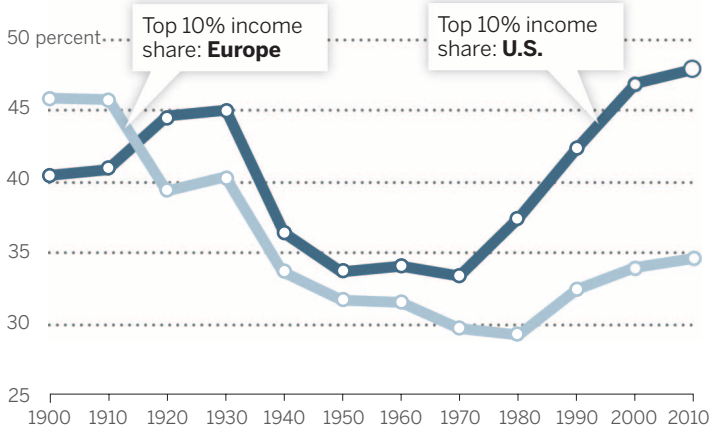


Fig. 1. Income inequality in Europe and the United States, 1900 to 2010. The share of total income accruing to top decile income holders was higher in Europe than in the United States from 1900 to 1910; it was substantially higher in the United States than in Europe from 2000 to 2010. The series report decennial averages (1900 = 1900 to 1909, etc.) constructed using income tax returns and national accounts. See (*24*), chapter 9, Fig. 9.8. Series available online at piketty.pse.ens.fr/capital21c.

Three Facts About Inequality in the Long Run

We find large changes in the levels of inequality, both over time and across countries. This reflects the fact that economic trends are not acts of God, and that country-specific institutions and historical circumstances can lead to very different inequality outcomes.

Income Inequality

First, we find that whereas income inequality was larger in Europe than in the United States a

[1]Department of Economics, Paris School of Economics, Paris, France. [2]Department of Economics, University of California at Berkeley, Berkeley, CA, USA.
*Corresponding author. E-mail: thomas.piketty@psemail.eu

century ago, it is currently much larger in the United States. This is true for every inequality metric. The simplest and most powerful measure, on which we focus in this article, is the share of total income going to the top decile (Fig. 1).

On the eve of World War I (WWI), in the early 1910s, the top decile income share was between 45 and 50% of total income in most European countries. This applies in particular to the United Kingdom, France, Germany, and Sweden, which are the four countries that we use to compute the European average series reported in this article. At the same time, the top decile income share was slightly above 40% in the United States.

One century later, in the early 2010s, the inequality ordering between Europe and the United States is reversed. In Europe, the top decile income share fell sharply, from 45 to 50% to about 30%, between 1914 and the 1950s–1960s. It has been rising somewhat since the 1970s–1980s, and it is now close to 35% (somewhat less in continental Europe and somewhat more in the United Kingdom, which has experienced an evolution closer to that of the United States). That is, the top decile share in Europe is currently almost one-third smaller than what it used to be one century ago. The secular decline in inequality would be even larger if we took into account the rise of taxes and transfers, and measure instead income after taxes and transfers. Total tax revenues and public spending were less than 10% of national income in every country before WWI, and they are now on the order of 30 to 50% of national income in every developed country. Properly attributing taxes, transfers, and public spending to each income decile raises important measurement issues, however, particularly regarding in-kind transfers (such as health, education, or public good spending). In this Review, we therefore focus on the long-run evolution of the inequality of primary income (pretax, pretransfer).

In the United States, the top decile income share in 1910 was lower than in Europe, then rose in the 1920s, fell in the 1930s–1940s, and stabilized around 30 to 35% in the 1950s–1960s, slightly above European levels of the time. It then rose at an unprecedented pace since the 1970s–1980s, and is now close to 50%. According to this measure, primary income concentration is currently higher than it has ever been in U.S. history. It is also slightly higher than in pre-WWI Europe.

Wealth Inequality

Second, we observe the same "great inequality reversal" between Europe and the United States when we look at wealth inequality rather than income inequality. That is, the share of total net private wealth owned by the top 10% of wealth holders was notably larger in Europe than in the United States one century ago, while the opposite is true today (Fig. 2).

There are important differences between income and wealth inequality dynamics, however. First, we stress that wealth concentration is always much higher than income concentration. The top decile wealth share typically falls in the 60 to 90% range, whereas the top decile income share is in the 30 to 50% range. Even more striking, the bottom 50% wealth share is always less than 5%, whereas the bottom 50% income share generally falls in the 20 to 30% range. The bottom half of the population hardly owns any wealth, but it does earn appreciable income: On average, members of the bottom half of the population (wealth-wise) own less than one-tenth of the average wealth, while members of the bottom half of the population (income-wise) earn about half the average income.

In sum, the concentration of capital ownership is always extreme, so that the very notion of capital is fairly abstract for large segments—if not the majority—of the population. The inequality of labor income can be high, but it is usually much less extreme. It is also less controversial, partly because it is viewed as more merit-based. Whether this is justified is a highly complex and debated issue to which we later return.

Next, in contrast to income inequality, U.S. wealth inequality levels have still not regained the record levels observed in Europe before World War I. The U.S. top decile wealth share was about 70 to 80% in from 1870 to 1910, fell to 60 to 70% from 1950 to 1980, and has been rising above 70% in recent decades. Naturally, this means that wealth concentration has been high throughout U.S. history. But this also implies that there has always been a large fraction of U.S. aggregate wealth—about 20 to 30% —that did not belong to the top 10%. As the bottom 50% wealth share has always been negligible, this remaining 20 to 30% fraction corresponds to the share owned by the "middle 40%" (i.e., the intermediate group between the bottom 50% and the top 10%), a social group that one might want to call the "wealth middle class." The important point is that, to a large extent, there has always been a wealth middle class in the United States.

In contrast, wealth concentration was so extreme in pre-WWI Europe that there was basically no wealth middle class. That is, the top decile wealth share was close to 90% (or even somewhat higher than 90%, as in the UK), so that the middle 40% wealth holders were almost as poor as the bottom 50% wealth holders (the wealth share of both groups was close to or less than 5%). Between 1914 and the 1950s–1960s, the top decile wealth share fell dramatically in Europe, from about 90% to less than 60%. It has been rising since the 1970s–80s, and is now close to 65% (somewhat more in the United Kingdom, and somewhat less in Continental Europe). In other words, the wealth middle class now commands a larger share of total wealth in Europe than in the United States—although this share has been shrinking lately on both sides of the Atlantic.

Wealth inequality in Europe and the United States, 1870–2010

Share of top wealth decile in total net wealth

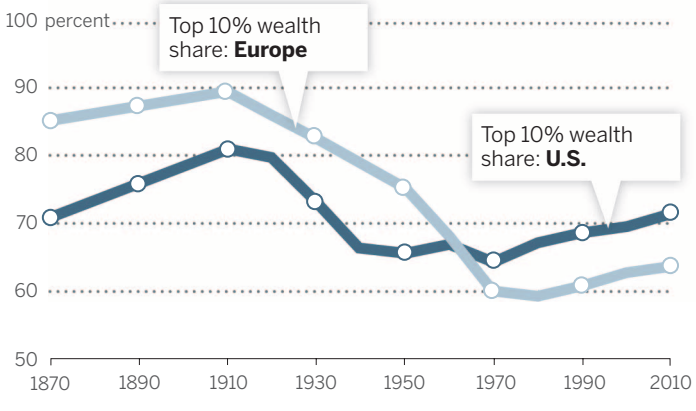


Fig. 2. Wealth inequality in Europe and the United States, 1870 to 2010. The share of total net wealth belonging to top decile wealth holders became higher in the United States than in Europe over the course of the 20th century. But it is still smaller than what it was in Europe before World War I. The series report decennial averages constructed using inheritance tax returns and national accounts. See (*24*), chapter 10, Fig. 10.6. Series available online at piketty.pse.ens.fr/capital21c.

Given that wealth inequality is lower in the United States today than in 1913 Europe, why is U.S. income inequality now as large as (or even slightly larger than) that in 1913 Europe? The reason is that modern U.S. inequality is based more on a very large rise of top labor incomes than upon the extreme levels of wealth concentration that characterized the "patrimonial" (wealth-based) societies of the past. In 1913 Europe, top incomes were predominantly top capital incomes (rent, interest, and dividends) coming from the very large concentration of capital ownership. Top U.S. incomes today are composed about equally of labor income and capital income. This generates approximately the same level of total income inequality, but it is not the same form of inequality.

Wealth-to-Income Ratios

Before further discussing the different possible interpretations for these important transformations, we introduce a third basic fact: If we look at the evolution of the aggregate value of wealth relative to income, we also find large historical

variations, again with striking differences between Europe and the United States (Fig. 3). This ratio is of critical importance for the analysis of inequality, as it measures the overall importance of wealth in a given society, as well as the capital intensity of production.

In every European country for which we have data, and in particular France, the United Kingdom, and Germany, the aggregate wealth-income ratio has followed a pronounced U-shaped pattern over the past century. On the eve of WWI, net private wealth was about equal to 6 to 7 years of national income in Europe. It then fell to about 2 to 3 years of national income in the 1950s. It has risen regularly since then, and it is now back to about 5 to 6 years of national income. Interestingly, we also find a similar pattern for Japan (*23*).

In contrast, the U.S. pattern is flatter: Net private wealth has generally equalled about 4 to 5 years of national income in the United States, with much less variation than in Europe or Japan. The U.S. pattern is also slightly U-shaped—with aggregate wealth-income ratios standing at a relatively lower level in the mid-20th century than at both ends of the century. But it is clearly much less marked than in Europe.

The comparison between Figs. 1 and 3 is particularly striking. Both figures have two U-shaped curves, but these are clearly different. The United States displays a U-shaped pattern for income inequality (mostly driven by the large rise of top labor incomes in recent decades). Europe (and Japan) shows a U-shaped pattern for aggregate wealth-income ratios. The United States is the land of booming top labor incomes; Europe is the land of booming wealth (albeit with a lower wealth concentration than in the United States). These are two distinct phenomena, involving different economic mechanisms and different parts of the developed world.

Interpreting the Long-Run Evidence

We now turn to the discussion of possible interpretations and lessons for the future. We stress at the outset that what we have to offer is little more than an informed discussion. Although we have at our disposal much more extensive historical and comparative data than were available to previous researchers, existing evidence is still far too incomplete and imperfect for a rigorous quantitative assessment of the various causes at play. Several different mechanisms have clearly played an important role in the evolution of income and wealth depicted in Figs. 1 to 3, but it is extremely difficult to disentangle the individual processes. We are not in the domain of controlled experiments: We cannot replay the 20th-century income and wealth dynamics as if the world wars, the rise of progressive taxation, or the Bolshevik revolution did not happen. Still, we can try to make some progress.

Wealth-to-Income Ratios

The relatively easier part of the story is the long-run evolution of aggregate wealth-to-income ratios (Fig. 3). The fall of European wealth-income ratios following the 1914–1945 capital shocks can be well accounted for by three main factors: direct war-related physical destruction of domestic capital assets (real estate, factories, machinery, equipment); lack of investment (a large fraction of 1914–1945 private-saving flows was absorbed by the enormous public deficits induced by war financing; there was also massive dissaving in some cases, e.g., foreign assets were sold to purchase government bonds; the resulting public debt was eventually wiped away by inflation); and a fall in relative asset prices (real estate and stock market prices were both historically very low in the immediate postwar period, partly due to rent control, nationalization, capital controls, and various forms of financial repression policies). In France and Germany, each of these three factors seems to account for about one-third of the total decrease in wealth-income ratios. In the United Kingdom, where domestic capital destruction was of limited importance (less than 10% of the total), the other two factors each account for about half of the decline in the aggregate wealth-income ratio (*23*, *24*).

Why did the postwar recovery of European wealth-income ratios take so much time? The simplest way to understand why capital accumulation is a slow process is to consider the following elementary arithmetic: With a saving rate of 10% per year, it takes 50 years to accumulate the equivalent of 5 years of income.

How is the long-run equilibrium wealth-income ratio determined, and why does it seem to vary across countries and over time? A simple yet powerful way to think about this issue is the so-called Harrod-Domar-Solow formula (*23*). In the long-run, assuming no systematic divergence between the relative price of capital assets and consumption goods, one can show that the wealth-to-income (or capital-to-income) ratio $\beta_t = K_t/Y_t$ converges toward $\beta = s/g$, where s is the long-run annual saving rate and g is the long-run annual total growth rate. The growth rate g is the sum of the population growth rate (including immigration) and the productivity growth rate (real income growth rate per person). This formula holds whether savings are invested in domestic or foreign assets (it also holds at the global level).

That is, with a saving rate $s = 10\%$ and a growth rate $g = 3\%$, then $\beta \approx 300\%$. But if the growth rate drops to $g = 1.5\%$, then $\beta \approx 600\%$. In short: Capital is back because low growth is back.

Intuitively, in a low-growth society, the total stock of capital accumulated in the past can become very important. In the extreme case of a society with zero population and productivity growth, income Y is fixed. As long as there is a positive net saving rate $s > 0$, the quantity of accumulated capital K will go to infinity. Therefore, the wealth-income ratio $\beta = K/Y$ would rise indefinitely (at some point, people in such a society would probably stop saving, as additional capital units become almost useless). With positive but small growth, the process is not as extreme: The rise of β stops at some finite level. But this finite level can be very high.

One can show that this simple logic can account relatively well for why the United States accumulates structurally less capital relative to its annual income than Europe and Japan. U.S. population growth rates exceed 1% per year, thanks to large immigration flows, so total U.S. growth rates—including productivity growth of around 1 to 1.5%—are at least 2 to 2.5% per year, if not 2.5 to 3%. By contrast, population growth in Europe and Japan is now close to zero, so that total growth is close to productivity growth, i.e., about 1 to 1.5% per year. This is further reinforced by the fact that U.S. saving rates tend to be lower than in Europe and Japan. To the extent that population growth will eventually decline almost everywhere, and that saving rates will stabilize, this also implies that the return of high

Wealth-to-income ratios in Europe and the United States, 1900–2010

Market value of net private wealth (% national income)

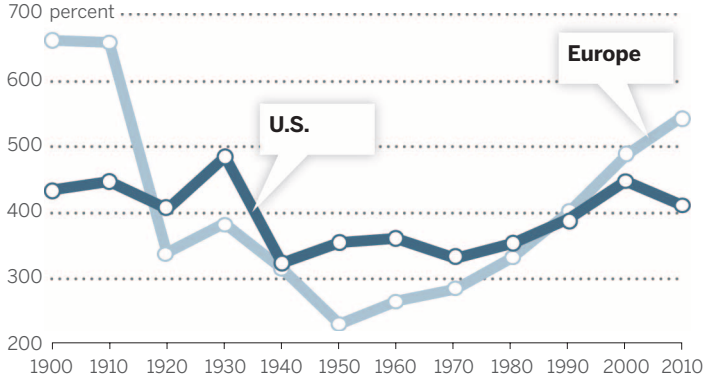


Fig. 3. Wealth-to-income ratios in Europe and the United States, 1900 to 2010. Total net private wealth was worth about 6 to 7 years of national income in Europe before World War I, fell to 2 to 3 years in 1950–1960, and increased back to 5 to 6 years in 2000–2010. In the United States, the U-shape pattern was much less marked. The series report decennial averages (1900 = 1900 to 1909, etc.) constructed using national accounts. See (*24*), chapter 5, Fig. 5.1. Series available online at piketty.pse.ens.fr/capital21c.

capital-to-income ratios will apply at the global level in the very long run (*23*, *24*).

The share of capital income in national income is defined as $\alpha = rK/Y = r\beta$, where r is the average annual real rate of return on wealth. For instance, if $r = 5\%$ and $\beta = K/Y = 600\%$, then $\alpha = 30\%$. Whether the rise in the capital income ratio β will also lead to a rise in α is a complicated issue.

In the standard economic model with perfectly competitive markets, r is equal to the marginal product of capital (that is, the additional output produced by one additional capital unit, all other things being equal). As the volume of capital β rises, the marginal product r tends to decline. The important question is whether r falls more or less rapidly than the rise in β. This depends on what economists define as the elasticity of substitution σ between capital and labor in the production function $Y = F(K, L)$.

A standard hypothesis in economics has been to assume a unitary elasticity, in which case the fall in r exactly offsets the rise in β, so that the capital share $\alpha = r\beta$ is a technological constant. However, historical variations in capital shares are far from negligible: α typically varies in the 20 to 40% range (and the labor share $1 - \alpha$ in the 60 to 80% range). In recent decades, rich countries have experienced both a rise in β and a rise in α, which suggests that σ is somewhat larger than 1. Intuitively, it makes sense to assume that σ tends to rise over the development process, as there are more diverse uses and forms for capital and more possibilities to substitute capital for labor (e.g., replacing delivery workers by drones or self-driving trucks).

Whether the capital share α will keep rising in future decades is an open question. It depends both on technological forces and on the bargaining power of capital and labor and the collective institutions regulating the capital-labor relationship (the simple economic model with perfectly competitive markets is likely excessively naïve). But from a logical standpoint, this is a plausible possibility, especially if the population and productivity growth slowdown pushes the global capital income ratio β toward higher levels.

Wealth Inequality: r > g

We now move to an even more complicated—and arguably more important—issue: the long-run dynamics of wealth inequality (Fig. 2). High capital intensity, as measured by high β and α, is not bad in itself. After all, it would be good to have an infinite quantity of robots producing most of the output, so that we can devote more time to leisure activities. The problem is twofold: Can we all find jobs as a robot designer (or in leisure-related activities), and who owns the robots? In practice, the concentration of capital ownership always seems to be very high—much more than the concentration of labor income (Figs. 1 and 2). The "patrimonial" (wealth-based) societies of Europe one century ago were characterized not only by very high β and α, but also by extreme capital concentration, with a top decile wealth share of around 90%.

How can we account for the very high level of wealth concentration that we observe in historical series, and what does this tell us about the future? The most powerful model to analyze structural changes in wealth inequality is a dynamic model with multiplicative random shocks. That is, assume that the individual-level wealth process has the following general form: $z_{it+1} = \omega_{it} z_{it} + \varepsilon_{it}$, where z_{it} is the position of individual i in the wealth distribution prevailing at time t (i.e., $z_{it} = k_{it}/k_t$ where k_{it} is net wealth owned by individual i at time t, and k_t = average net wealth of the entire population at time t), ω_{it} is a multiplicative random shock, and ε_{it} is an additive random shock.

The shocks ω_{it} and ε_{it} can be interpreted as reflecting different types of events that often occur in individual wealth histories, including shocks to rates of return (some individuals may get returns that are far above average returns; investment strategies may fail and lead to family bankruptcy); shocks to demographic parameters (some families have many children; some individuals die young); shocks to preferences parameters (some individuals like to save, some prefer to consume their wealth); shocks to productivity parameters (capital income is sometimes supplemented by high labor income); and so on.

Importantly, for a given structure of shocks, the variance of the multiplicative term ω_{it} is an increasing function of $r - g$, where r is the (net-of-tax) rate of return and g is the economy's growth rate. Intuitively, a higher $r - g$ tends to amplify initial wealth inequalities: It implies that past wealth is capitalized at a faster pace, and that it is less likely to be overtaken by the general growth of the economy. Under fairly general conditions, one can show that the top tail of the distribution of wealth converges toward a Pareto distribution, and that the inverted Pareto coefficient (measuring the thickness of the upper tail and hence the inequality of the distribution) increases with $r - g$ (*3*, *14*, *24*–*26*).

The dynamic wealth accumulation model with multiplicative shocks can explain the extreme levels of wealth concentration that we observe in the data much better than alternative models. In particular, if wealth accumulation were predominantly driven by lifecycle or precautionary motives, then wealth inequality would not be as large as what we observe (it would be comparable in magnitude to income inequality, or even lower).

The dynamic multiplicative model can also help to explain some of the important historical variations that we observe in wealth concentration series.

Rate of return vs. growth rate at the world level, Antiquity–2100

Annual rate of return or rate of growth

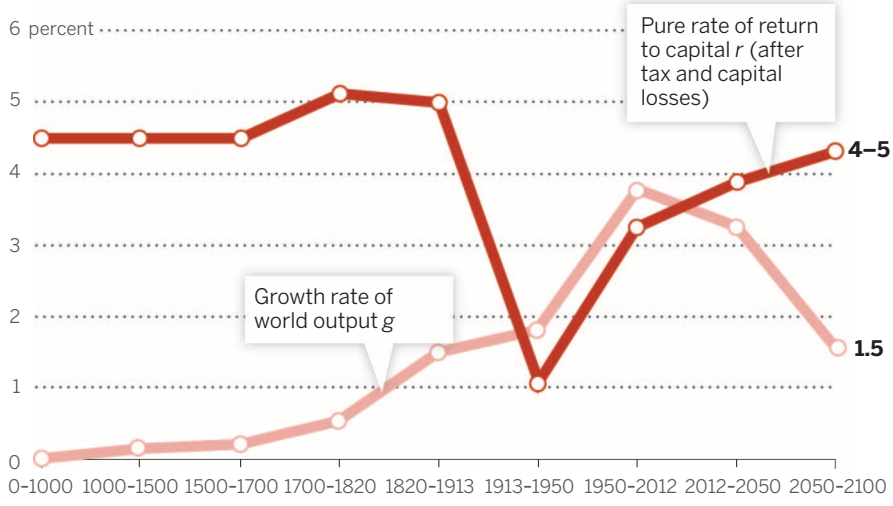


Fig. 4. Rate of return versus growth rate at the global level, from Antiquity until 2100. The average rate of return to capital (after tax and capital losses) fell below the growth rate in the 20th century. It may again surpass it in the 21st century, as it did throughout human history except in the 20th century. The series was constructed using national accounts for 1700 and after and historical sources on growth and rent to land values for the period before 1700. See (*24*), chapter 10, Fig. 10.10. Series available online at piketty.pse.ens.fr/capital21c. The future values for g are based upon UN demographic projections (median scenario) for population growth and on the assumption that between-country convergence in productivity growth rates will continue at its current pace. The future values for r are simply based upon the continuation of current pretax values and the assumption that tax competition will continue. See (*24*), chapter 10, Fig. 10.10. Series available online at piketty.pse.ens.fr/capital21c.

In particular, it is critical to realize that $r - g$ was very large during most of human history (Fig. 4). Growth was very low until the industrial revolution (much less than 1% per year), whereas average rates of return were typically on the order of 4 to 5% per year (historically, in preindustrial agrarian societies, annual rent on land, the main capital asset, was about 4 to 5% of the land value) and taxes were minimal. Growth rates rose substantially during the 18th and 19th centuries, but they remained relatively small (1 to 1.5%) compared to rates of return. This large gap between r and g explains why wealth concentration was so large until World War I and why wealth concentration was smaller in the United States, where population growth was faster.

During the 20th century, growth rates were exceptionally high (in particular due to very high population growth, which even today represents about half of global gross domestic product growth), and rates of return were severely reduced by capital shocks (destructions) and the rise of taxation. Simple simulations show that this effect is quantitatively sufficiently important to explain why wealth concentration did not return to pre-WWI levels in the postwar period.

Other factors might also have played a role. For instance, the rise of the wealth middle class might partly come from the fact that the growth of incomes and living standards eventually induced the rise of middle class saving. However, this process does not seem to have taken place in pre-WWI Europe, because of the powerful unequalizing impact of the $r - g$ factor (*17*, *21*, *24*, *27*).

To the extent that population growth (and possibly productivity growth) will slow down in the 21st century, and that after-tax rates of return to capital will rise (due to rising international tax competition to attract capital, and maybe also to changing technology), it is likely that $r - g$ will increase again in the 21st century, which could lead to a structural rise in wealth concentration.

This model seems to capture relatively well some of the evolutions that we are currently observing at the global level. For instance, if we use the global billionaires rankings published by *Forbes* magazine since 1987, we find that the very top fractiles of the global wealth distribution have been rising on average at about 6 to 7% per year in real terms over the 1987–2013 period, i.e., more than three times as fast as average global wealth (about 2% per year over the same period) (*24*).

We stress, however, that our ability to properly measure and monitor the dynamics of the global distribution of wealth is far from being satisfactory. National statistical institutes as well as international organizations are facing major difficulties in tracking down cross-border wealth, and magazines are ill-equipped to produce rigorous statistics. Despite some recent progress in this area (*28*), our ability to measure global wealth is also severely limited by the rise of tax havens (*29*).

The Dynamics of Income Inequality

We finally return to the most difficult and uncertain part: the long-run dynamics of income inequality (Fig. 1). This is the most difficult part because income inequality combines forces arising from the inequality of capital ownership and capital income (which, as we have just seen, are relatively complex) and forces related to the inequality of labor income (which involve a different set of economic and social processes).

Kuznets posited that income inequality first rises with economic development when new, higher-productivity sectors emerge (e.g., manufacturing industry during the industrial revolution) but then decreases as more and more workers join the high-paying sectors of the economy. Our data show that this is not the reason that income inequality declined in developed countries during the first half of the 20th century. The compression of incomes occurred primarily because of the fall of top capital incomes induced by the world wars, the Great Depression, and the regulatory and fiscal policies developed in response to these shocks. In particular, there was no structural decline in the inequality of labor income (*8–13*, *24*). Kuznets' overly optimistic theory of a natural decline in income inequality in market economies largely owed its popularity to the Cold War context of the 1950s as a weapon in the ideological fight between the market economy and socialism (*24*).

What are the main forces that determine the level of labor income inequality in the long-run? The most widely used economic model is based on the idea of a race between education and technology (*30*). That is, the expansion of education leads to a rise in the supply of skills, while technological change leads to a rise in the demand for skills. Depending on which process occurs faster, the inequality of labor income will either fall or rise.

One proposed explanation for the increase of inequality in recent decades has been the rise in the global competition for skills, itself driven by globalization, skill-biased technical change and the rise of information technologies. Such skill-biased technological progress is not sufficient to explain important variations between countries: The rise of labor income inequality was relatively limited in Europe (and Japan) compared to the United States, despite similar technological changes. In the very long run, European labor income inequality appears to be relatively stable (there is no major downward or upward trend in the wage shares received by the various deciles and percentiles of the wage distribution). This suggests that the supply and demand for skills have increased approximately at the same pace in Europe

Could the particularly large increase in U.S. labor income inequality in recent decades be explained by insufficient educational investment for large segments of the U.S. labor force? In that case, massive investment in higher education would be the right policy to curb rising income inequality (*30*). Although this view is very appealing, it cannot account for all of the facts. In particular, the race between education and technology fails to explain the unprecedented rise of very top labor incomes that has occurred in the United States over the past few decades. A very large part of the rise in the top 10% income share comes from the top 1% (or even the top 0.1%). This is largely due to the rise of top executive compensation in large U.S. corporations (both financial and nonfinancial). We discuss in the supplementary online material how changes in tax policy, as well as social norms regarding pay equality, likely play a key role in shaping labor income inequality.

To summarize: Inequality does not follow a deterministic process. In a sense, both Marx and Kuznets were wrong. There are powerful forces pushing alternately in the direction

Box 1. Income and wealth: definitions

Income is a flow. It corresponds to the quantity of goods and services produced and distributed each year. Income can be decomposed as the sum of labor income (wages, salaries, bonuses, earnings from nonwage labor, and other remuneration for labor services) and capital income (rent, dividends, interest, business profits, capital gains, royalties, and other income derived from owning capital assets). In this Review, we focus on the long-run evolution of the inequality of primary income, defined as income before taxes and government transfers. In contrast, disposable income is defined as income after taxes and government transfers. Although we do not analyze disposable income in this article, comparing inequality of primary income and inequality of disposable income is useful to assess the role of the government in reducing income inequality.

Wealth (or capital) is a stock. It corresponds to the total wealth owned at a given point in time. This stock comes from the wealth appropriated or accumulated in the past. In the context of this article, wealth is defined as nonhuman net worth, i.e., the sum of nonfinancial and financial assets, net of financial liabilities (debt). National wealth is the sum of private wealth (net worth owned by private individuals) and public wealth (net worth owned by the government and other public agencies). In this article, we focus on the level and distribution of private wealth. More details on these definitions, concepts, and corresponding series are provided in (*23*, *24*).

of rising or shrinking inequality. Which one dominates depends on the institutions and policies that societies choose to adopt.

REFERENCES

1. S. Kuznets, *Shares of Upper Income Groups in Income and Savings* (National Bureau of Economic Research, Cambridge, MA, 1953).
2. R. J. Lampman, *The Share of Top Wealth holders in National Wealth, 1922-1956* (Princeton Univ. Press, Princeton, NJ, 1962).
3. A. B. Atkinson, A. J. Harrison, *Distribution of Personal Wealth in Britain, 1923-1972* (Cambridge Univ. Press, Cambridge, 1978).
4. A. Daumard, *Les fortunes françaises au 19e siècle. Enquête sur la répartition et la composition des capitaux privés à Paris, Lyon, Lille, Bordeaux et Toulouse d'après l'enregistrement des déclarations de successions* (Mouton, Paris, 1973).
5. A. H. Jones, *American Colonial Wealth: Documents and Methods* (Arno Press, New York, 1977).
6. P. Lindert, *J. Polit. Econ.* **94**, 1127–1162 (1986).
7. L. Soltow, *Distribution of Wealth and Income in the United States in 1798* (Univ. of Pittsburgh Press, Pittsburgh, PA, 1989).
8. T. Piketty, *Les hauts revenus en France au 20e siècle—Inégalités et redistributions, 1901–1998* (Grasset, Paris, 2001).
9. T. Piketty, *J. Polit. Econ.* **111**, 1004–1042 (2003).
10. A. B. Atkinson, *J. R. Stat. Soc. Ser. A Stat. Soc.* **168**, 325–343 (2005).
11. T. Piketty, E. Saez, *Q. J. Econ.* **118**, 1–41(2003).
12. A. B. Atkinson, T. Piketty, Eds., *Top Incomes over the 20th Century—A Contrast Between Continental European and English Speaking Countries* (Oxford Univ. Press, New York, 2007).
13. A. B. Atkinson, T. Piketty, Eds., *Top Incomes—A Global Perspective* (Oxford Univ. Press, New York, 2010).
14. A. B. Atkinson, T. Piketty, E. Saez, *J. Econ. Lit.* **49**, 3–71 (2011).
15. F. Alvaredo, A. B. Atkinson, T. Piketty, E. Saez, *J. Econ. Perspect.* **27**, 3–21 (2013).
16. W. Kopczuk, E. Saez, *Natl. Tax J.* **57**, 445–487 (2004).
17. T. Piketty, G. Postel-Vinay, J. L. Rosenthal, *Am. Econ. Rev.* **96**, 236–256 (2006).
18. J. Roine, D. Waldenstrom, *Scand. J. Econ.* **111**, 151–187 (2009).
19. H. Ohlson, J. Roine, D. Waldenstrom, in J. B. Davies, Ed., *Personal Wealth from a Global Perspective* (Oxford Univ. Press, Oxford, 2008), pp. 42–63.
20. D. Waldenstrom, *Lifting all Boats? The Evolution of Income and Wealth Inequality Over the Path of Development* (Lund University, Sweden, 2009)
21. T. Piketty, *Q. J. Econ.* **126**, 1071–1131 (2011).
22. R. Goldsmith, *Comparative National Balance Sheets: A Study of Twenty Countries, 1688-1978* (Univ. of Chicago Press, Chicago, IL, 1985)
23. T. Piketty, G. Zucman, *Q. J. Econ.* **129**, in press (2014); http://piketty.pse.ens.fr/files/PikettyZucman2013WP.pdf.
24. T. Piketty, *Capital in the Twenty-first Century* (Harvard Univ. Press, Cambridge, MA, 2014).
25. J. Stiglitz, *Econometrica* **37**, 382–397 (1969).
26. M. Nirei, "Pareto Distributions in Economics Growth Models," Institute of Innovation Research Working Paper No. 09-05, Hitotsubashi University, Tokyo (2009)
27. T. Piketty, G. Postel-Vinay, J. L. Rosenthal, *Explor. Econ. Hist.* **51**, 21–40 (2014).
28. J. Davies, S. Sandstrom, T. Shorrocks, E. Wolff, *Econ. J.* **121**, 223–254 (2011).
29. G. Zucman, *Q. J. Econ.* **128**, 1321–1364 (2013).
30. C. Goldin, L. Katz, *The Race Between Education and Technology* (Harvard Univ. Press, Cambridge, MA, 2008).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/838/suppl/DC1
Supplementary Text
Figs. S1 and S2
References (*31*, *32*)

10.1126/science.1251936

REVIEW

Skills, education, and the rise of earnings inequality among the "other 99 percent"

David H. Autor

The singular focus of public debate on the "top 1 percent" of households overlooks the component of earnings inequality that is arguably most consequential for the "other 99 percent" of citizens: the dramatic growth in the wage premium associated with higher education and cognitive ability. This Review documents the central role of both the supply and demand for skills in shaping inequality, discusses why skill demands have persistently risen in industrialized countries, and considers the economic value of inequality alongside its potential social costs. I conclude by highlighting the constructive role for public policy in fostering skills formation and preserving economic mobility.

Public debate has recently focused on a subject that economists have been analyzing for at least two decades: the steep, persistent rise of earnings inequality in the U.S. labor market and in developed countries more broadly. Much popular discussion of inequality concerns the "top 1 percent," referring to the increasing share of national income accruing to the top percentile of households. Although this phenomenon is undeniably important, an exclusive focus on the concentration of top incomes ignores the component of rising inequality that is arguably even more consequential for the "other 99 percent" of citizens: the dramatic growth in the wage premium associated with higher education and, more broadly, cognitive ability. This paper considers the role of the rising skill premium in the evolution of earnings inequality.

There are three reasons to focus a discussion of rising inequality on the economic payoff to skills and education. First, the earnings premium for education has risen across a large number of advanced countries in recent decades, and this rise contributes substantially to the net growth of earnings inequality. In the United States, for example, about two-thirds of the overall rise of earnings dispersion between 1980 and 2005 is proximately accounted for by the increased premium associated with schooling in general and postsecondary education in particular (*1*, *2*). Second, despite a lack of consensus among economists regarding the primary causes of the rise of very top incomes (*3*–*6*), an influential literature finds that the interplay between the supply and demand for skills provides substantial insight into why the skill premium has risen and fallen over time—and, specifically, why the earnings gap between college and high school graduates has more than doubled in the United States over the past three decades. A third reason for focusing on the skill premium is that it offers broad insight into the evolution of inequality within a market economy, highlighting the social value of inequality alongside its potential social costs and illuminating the constructive role for public policy in maximizing the benefits and minimizing the costs of inequality.

The rising skill premium is not, of course, the sole cause of growing inequality. The decades-long decline in the real value of the U.S. minimum wage (*7*), the sharp drops in non-college employment opportunities in production, clerical, and administrative support positions stemming from automation, the steep rise in international competition from the developing world, the secularly declining membership and bargaining power of U.S. labor unions, and the successive enactment of multiple reductions in top federal marginal tax rates, have all served to magnify inequality and erode real wages among less educated workers. As I discuss below, the foremost concern raised by these multiple forces is not their impact on inequality per se, but rather their adverse effect on the real earnings and employment of less educated workers.

I begin by documenting the centrality of the rising skill premium to the overall growth of earnings inequality. I next consider why skills are heavily rewarded in advanced economies and why the demand for them has risen over time. I then demonstrate the substantial explanatory power of a simple framework that embeds both the demand and supply for skills in interpreting the evolution of the inequality over five decades. The final section considers the productive role that inequality plays in a market economy and the potential risks attending very high and rising inequality; evidence on whether those risks have been realized; and the role of policy and governance in encouraging skills formation, fostering opportunity,

Department of Economics and National Bureau of Economic Research, Massachusetts Institute of Technology, 40 Ames Street, E17-216, Cambridge, MA 02142, USA. **E-mail: dautor@mit.edu**

and countering the possibility that extremes of inequality erode economic mobility and reduce economic dynamism.

The Critical Role of Skills in the Labor Market

There is no denying the extraordinary rise in the incomes of the top 1% of American households over the past three decades. Between 1979 and 2012, the share of all household income accruing to the top percentile of U.S. households rose from 10.0% to 22.5% (*8*, *9*). To get a sense of how much money that is, consider the conceptual experiment of redistributing the gains of the top 1% between 1979 and 2012 to the bottom 99% of households (*10*). How much would this redistribution raise household incomes of the bottom 99%? The answer is $7107 per household—a substantial gain, equal to 14% of the income of the median U.S. household in 2012. (I focus on the median because it reflects the earnings of the typical worker and thus excludes the earnings of the top 1%.)

Now consider a different dimension of inequality: the earnings gap between U.S. workers with a 4-year college degree and those with only a high school diploma (*11*). Economists frequently use this college/high school earnings gap as a summary measure of the "return to skill"—that is, the gain in earnings a worker can expect to receive from investing in a college education. As illustrated in Fig. 1, the earnings gap between the median college-educated and median high school–educated among U.S. males working full-time in year-round jobs was $17,411 in 1979, measured in constant 2012 dollars. Thirty-three years later, in 2012, this gap had risen to $34,969, almost exactly double its 1979 level. Also seen is a comparable trend among U.S. female workers, with the full-time, full-year college/high school median earnings gap nearly doubling from $12,887 to $23,280 between 1979 and 2012. As Fig. 1 underscores, the economic payoff to college education rose steadily throughout the 1980s and 1990s and was barely affected by the Great Recession starting in 2007.

Because the earnings calculations in Fig. 1 reflect individual incomes while the top 1% calculations reflect household incomes, the two calculations are not directly comparable. To put the numbers on the same footing, consider the earnings gap between a college-educated two-earner husband-wife family and a high school–educated two-earner husband-wife family, which rose by $27,951 between 1979 and 2012 (from $30,298 to $58,249). This increase in the earnings gap between the typical college-educated and high school–educated household earnings levels is four times as large as the redistribution that has notionally occurred from the bottom 99% to the top 1% of households. What this simple calculation suggests is that the growth of skill differentials among the "other 99 percent" is arguably even more consequential than the rise of the 1% for the welfare of most citizens.

The median earnings comparisons in Fig. 1 also convey a key feature of rising inequality that cannot be inferred from trends in top incomes: Wage inequality has risen throughout the earnings distribution, not merely at the top percentiles. Figure S1 documents this pattern by plotting, for 12 Organization for Economic Cooperation and Development (OECD) member countries over three decades (1980 to 2011), the change in the ratio of full-time earnings of males at the 90th percentile relative to males at the 10th percentile of the wage distribution. Although the 90/10 earnings ratio differed greatly across countries at the earliest date of the sample—from a low of 2.0 in Sweden to a high of 3.6 in the United States—this earnings ratio increased substantially in all but one of them (France) over the next 30 years, growing by at least 25 percentage points in 10 countries, by at least 50 percentage points in 8 countries, and by more than 100 percentage points in three countries (New Zealand, the United Kingdom, and the United States).

How much does the rising education premium contribute to the increase of earnings inequality? Although data limitations make it difficult to answer this question for most countries, we do know the answer for the United States. Goldin and Katz (*1*) found that the increase in the education wage premium explains about 60 to 70% of the rise in the dispersion of U.S. wages between 1980 and 2005 and, similarly, Lemieux (*12*) calculated that higher returns to postsecondary education can account for 55% of the rise in male hourly wage variance from 1973–1975 to 2003–2005. Firpo *et al.* (*13*) found that rising returns to education can explain just over 95% of the rise of the U.S. male 90/10 earnings ratio between 1984 and 2004. That is, holding the expanding education premium constant over this period, there would have been essentially no increase in the relative wages of the 90th-percentile worker versus the 10th-percentile worker.

I have so far used the terms education and skill interchangeably. What evidence do we have that it is skills that are rewarded per se, rather than simply educational credentials? The Program for the International Assessment of Adult Competencies (PIAAC) provides a compelling data source for gauging the importance of skills in wage determination. The PIAAC is an internationally harmonized test of adult cognitive and workplace skills (literacy, numeracy, and problem-solving) that was administered by the OECD to large, representative samples of adults in 22 countries between 2011 and 2013 (*14*). Figure 2, sourced from (*15*), plots the relationship between adults' earnings and their PIAAC numeracy scores across these 22 countries. The length of each bar reflects the average percentage earnings differential between full-time workers ages 35 to 54 who differ by one standard deviation in the PIAAC score. The whiskers on each bar provide the 95% confidence intervals for the estimates.

College/high school median annual earnings gap, 1979–2012

In constant 2012 dollars

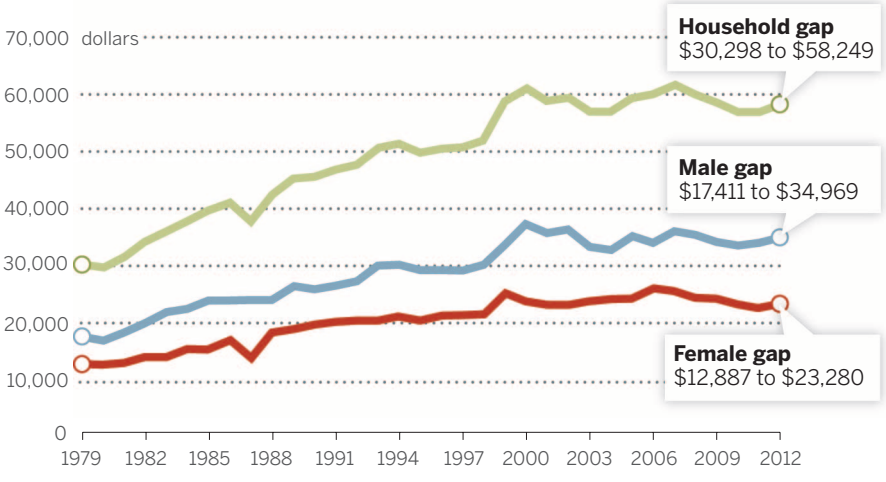


Fig. 1. College/high school median annual earnings gap, 1979–2012. Figure is constructed using Census Bureau P-60 (1979–1991) and P-25 (1992–2012) tabulations of median earnings of full-time, full-year workers by educational level and converted to constant 2012 dollars (to account for inflation) using the CPI-U-RS price series. Prior to 1992, college-educated workers are defined as those with 16 or more years of completed schooling, and high school–educated workers are those with exactly 12 years of completed schooling. After 1991, college-educated workers are those who report completing at least 4 years of college, and high school–educated workers are those who report having completed a high school diploma or GED credential.

This figure conveys three points. First, cognitive skills are substantially rewarded in the labor market across all 22 economies. The average wage premium corresponding to one "unit" (i.e., one standard deviation) increase in measured cognitive skills is 18%. In addition, cognitive earnings premiums differ substantially across countries. The premium is below 13% in Sweden, the Czech Republic, and Norway. It is above 20% in six countries. The United States stands out as having the highest measured return to skill, with a premium of 28% per unit increment to cognitive ability. Concretely, comparing two U.S. workers who are one standard deviation above and one standard deviation below the population average of cognitive ability, we would expect their full-time weekly earnings to differ by 50 to 60%. Notably, the high return to cognitive ability in the United States does not follow automatically from high levels of U.S. earnings inequality. If U.S. wages were determined mainly by luck, beauty, or family connections, we would expect little connection between workers' cognitive ability and their labor market rewards (*16*). Figure 2 demonstrates that this is not the case.

Of course, these data do not explain why the skill premium has risen over time, nor why the United States has a higher skill premium than so many other advanced nations. The next section considers the supply and demand for skill in the labor market—specifically, why they fluctuate over time and how their interaction helps to determine the skill premium. I focus on the United States in this section to allow a deeper exploration of the data.

Education and Inequality

Workers' earnings in a market economy depend fundamentally (some economists would say entirely) on their productivity—that is, the value they produce through their labor. And in turn, workers' productivity depends on two factors. One is their capabilities, concretely, the tasks they can accomplish (i.e., their skills). A second is their scarcity: The fewer workers that are available to accomplish a task, and the more employers need that task accomplished, the higher is workers' economic value in that task. In conventional terms, the skill premium depends upon what skills employers require (skill demand) and what skills workers have acquired (skill supply). To interpret the evolution of this premium, we need to account for both forces.

Skill Demands: The Long View

A technologically advanced economy requires a literate, numerate, and technically and scientifically trained workforce to develop ideas, manage complex organizations, deliver healthcare services, provide financing and insurance, administer government services, and operate critical infrastructure. This was not always the case. In 1900, 4 in 10 U.S. jobs were in agriculture, 11% of the population was illiterate, a substantial fraction of economic activity required hard physical labor, and workers' strength and physical stamina were key job skills (*17*, *18*). Few citizens would have predicted at the time that a century later, health care, finance, information technology, consumer electronics, hospitality, leisure, and entertainment would employ far more workers than agriculture—which employed only 2% of U.S. workers in 2010. As physical labor has given way to cognitive labor, the labor market's demand for formal analytical skills, written communications, and specific technical knowledge—what economists often loosely term cognitive skills—has risen spectacularly.

The central determinant of the supply of skills available to an advanced economy is its education system. In 1900, the typical young, native-born American had only a common school education, about the equivalent of six to eight grades (*19*). By the late 19th century, however, many Americans recognized that farm employment was declining, industry was rising, and their children would need additional education to earn a living. Over the first four decades of the 20th century, the United States became the first nation in the world to deliver universal high school education to its citizens. Tellingly, the high school movement was led by the farm states.

As the high school movement reached its conclusion, postsecondary education became increasingly indispensable to the growing occupations of medicine, law, engineering, science, and management. In 1940, only 6% of Americans had completed a 4-year college degree. From the end of the Second World War to the early 1980s, however, the ranks of college-educated workers rose robustly and steadily, with each cohort of workers entering the labor market boasting a proportionately higher rate of college education than the cohort that preceded it. This intercohort pattern, which was abetted by the Second World War and Korean War GI Bills (*20*) and by huge state and federal investments in public college and university systems, is depicted in Fig. 3A. From 1963 through 1982, the fraction of all U.S. hours worked that were supplied by college graduates rose by almost 1 percentage point per year, a remarkably rapid gain.

After 1982, however, the rate of intercohort increase fell by almost half—from 0.87 percentage points to 0.47 percentage points per year—and did not begin to rebound until 2004, nearly two decades later. As shown in fig. S2, this deceleration in the supply of college graduates is particularly stark when one focuses on young adults with fewer than 10 years of experience—that is, the cohorts of recent labor market entrants at each point in time. Although the supply of young college-educated males relative to young high school–educated males increased rapidly in the 1960s and early 1970s (and indeed throughout the postwar period), this rising tide reached an apex in 1974 from which

Fig. 2. Cross-national differences in wage returns to skills, 2011–2013. Reproduced with permission from Hanushek *et al.* [(*15*), table 2]. Estimates are obtained by regressing the natural logarithm of workers' weekly full-time earnings on test scores while controlling for sex and labor market experience (both a linear and a quadratic term). Regression estimates are performed separately for each country and test scores are normalized with mean zero and unit standard deviation within each country. Estimates that normalize test scores on a common basis across countries, or that use literacy or problem-solving scores rather than numeracy scores, yield qualitatively similar patterns.

it barely budged for the better part of the next 30 years. Among young females, the deceleration in supply was also unmistakable, although not as abrupt or as complete as for males.

The counterpart to this deceleration in the growth of supply of college-educated workers is the steep rise in the college premium commencing in the early 1980s and continuing for 25 years. Concretely, when the influx of new college graduates slowed, the premium that a college education commanded in the labor market increased. The critical role played by the fluctuating supply of college education in the rise of U.S. inequality is documented in Fig. 3B, which plots the college wage premium from 1963 through 2012 (blue line). This premium fluctuated in a comparatively narrow band during the 1960s and 1970s, as rising demand for educated workers was met with rapidly rising year-over-year increases in supply. In 1981, the average college graduate earned 48% more per week than the average high school graduate—a significant earnings gap but not an earnings gulf. When the supply deceleration began in 1982, however, the college premium hit an inflection point. This premium notched remarkably rapid year-over-year gains from 1982 forward, reaching 72% in 1990, 90% in 2000, and 97% in 2005 (*21*, *22*). Thus, the average earnings of college graduates were 1.5 times those of high school graduates in 1982 but were double those of high school graduates by 2005.

Why is this deceleration in supply relevant to the college premium? After all, although the growth of supply slowed in 1982, it was still rising. A likely answer is that the demand for college workers rose in the interim. Throughout much of the 20th century, successive waves of innovation—electrification, mass production, motorized transportation, telecommunications—have reduced the demand for physical labor and raised the centrality of cognitive labor in practically every walk of life. The past three decades of computerization, in particular, have extended the reach of this process by displacing workers from performing routine, codifiable cognitive tasks (e.g., bookkeeping, clerical work, and repetitive production tasks) that are now readily scripted with computer software and performed by inexpensive digital machines. This ongoing process of machine substitution for routine human labor complements educated workers who excel in abstract tasks that harness problem-solving ability, intuition, creativity, and persuasion—tasks that are at present difficult to automate but essential to perform. Simultaneously, it devalues the skills of workers, typically those without postsecondary education, who compete most directly with machinery in performing routine-intensive activities. The net effect of these forces is to further raise the demand for formal education, technical expertise, and cognitive ability (*23*–*27*).

Bringing the Supply-Demand Framework to the Data

The persistently rising demand for educated labor in advanced economies was first noted by the Nobel Prize–winning economist Jan Tinbergen (*28*) and is often referred to as the "education race" model (*19*). Its primary implication is that if the supply of educated labor does not keep pace with persistent outward shifts in demand for skills, the skill premium will rise. In the words of the Red Queen in Lewis Carroll's Alice in Wonderland, "…it takes all the running you can do, to keep in the same place." Thus, when the rising supply of educated labor began to slacken in the early 1980s, a logical economic consequence was an increase in the college skill premium.

To more formally account for the impact of the fluctuating growth rate of supply of college-educated workers on the college wage differential, Fig. 3B depicts the fit of a simple regression model that predicts the college wage premium in each year as a function of two factors: (i) the contemporaneous supply of college graduates, and (ii) a time trend, which serves as a proxy for the secularly rising demand for college-educated

The supply of college graduates and the U.S. college/high school premium, 1963–2012

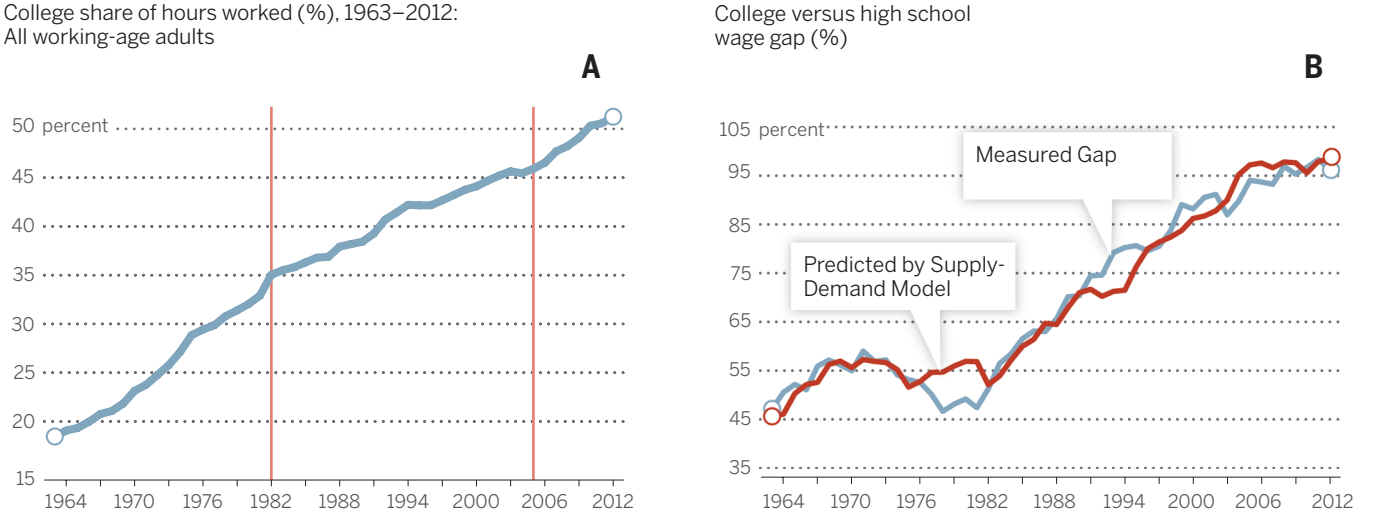


Fig. 3. The supply of college graduates and the U.S. college/high school premium, 1963–2012. (**A**) College share of hours worked in the United States, 1963–2012: All working-age adults. Figure uses March CPS data for earnings years 1963 to 2012. The sample consists of all persons aged 16 to 64 who reported having worked at least 1 week in the earnings years, excluding those in the military. Following an extensive literature, college-educated workers are defined as all of those with four or more completed years of college plus half of those with at least 1 year of completed college. Non-college workers are defined as all workers with high school or less education, plus half of those with some completed college education. For each individual, hours worked are the product of usual hours worked per week and the number of weeks worked last year. Individual hours worked are aggregated using CPS sampling weights. (**B**) College versus high school wage gap. Figure uses March CPS data for earnings years 1963 to 2012. The series labeled "Measured Gap" is constructed by calculating the mean of the natural logarithm of weekly wages for college graduates and non–college graduates, and plotting the (exponentiated) ratio of these means for each year. This calculation holds constant the labor market experience and gender composition within each education group. The series labeled "Predicted by Supply-Demand Model" plots the (exponentiated) predicted values from a regression of the log college/noncollege wage gap on a quadratic polynomial in calendar years and the natural log of college/noncollege relative supply. See text and supplementary material for further details.

workers (*29*). Comparing the fitted values (red series) from this simple supply-demand model alongside the actual data (blue series) reveals an extremely tight correspondence over the course of five decades and three distinct eras: a declining skill premium in the 1970s; an explosive rise in the premium during the 1980s, 1990s, and early 2000s; and, most recently, a plateau commencing after 2005. A key implication of this figure is that a central causal factor behind rising inequality in the United States has been the slowdown in the accumulation of skills by young adults almost 30 years ago. Had the supply of college graduates risen as rapidly in the decades after 1980 as it did in the decades immediately before, it is quite plausible that there would have been no sustained rise in the skill premium in the U.S. labor market.

Of course, this set of facts raises another puzzle: If slackening college supply sparked rising inequality, what caused rising U.S. postsecondary achievement to grind to a sudden halt in 1982? Work by Card and Lemieux (*30*) highlights that one critically important factor was the United States' involvement in the Vietnam War. Because draft-eligible males in the Vietnam era were often able to defer their military service by enrolling in postsecondary schooling, the war artificially boosted college attendance. This created something of a glut of college enrollments in the late 1960s and early 1970s, which in turn depressed the college earnings premium in the 1970s (see Fig. 3) and likely reduced the attractiveness of college-going absent the military draft. Thus, when the war ended in the early 1970s, college enrollment rates dropped sharply, particularly among males. The fall in enrollment produced a corresponding decline in college completions half a decade later, and a surge of inequality followed.

This supply-demand explanation for the rise of U.S. inequality may appear almost too simple to be credible. After all, we are comparing just two economic variables: the college wage premium and the supply of college graduates in the U.S. workforce. But a host of rigorous studies commencing with Katz and Murphy (*31*) confirm the remarkable explanatory power of this simple supply-demand framework for explaining trends in the college versus high school earnings gap over the course of nine decades of U.S. history, as well as across other industrialized economies (most notably, the United Kingdom and Canada) and among age and education groups within countries (*19*, *31*–*36*). The United States was far from the only Western country to experience this surge.

Present discounted value of college relative to high school degree net of tuition, 1965–2008

College/high school difference, 2009 dollars

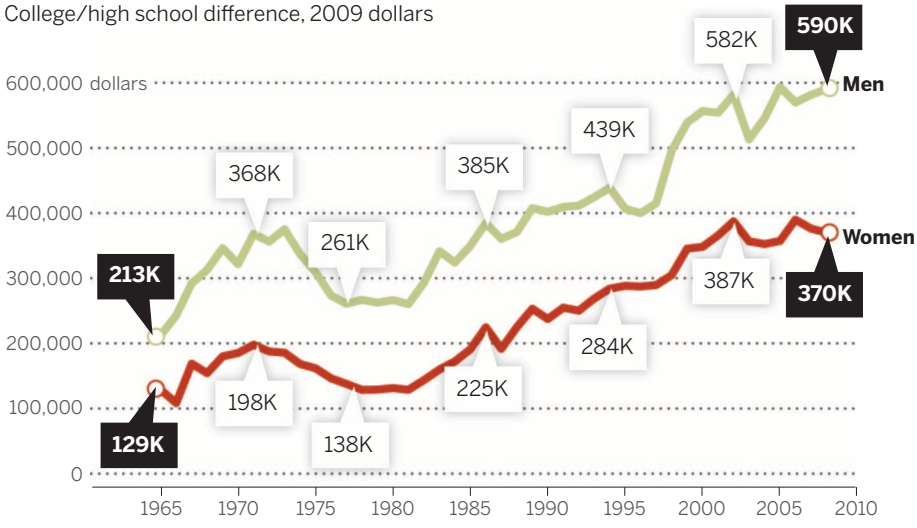


Fig. 4. Present discounted value of college relative to high school degree net of tuition, 1965–2008. Reproduced from Avery and Turner with permission of the American Economic Association (*39*). Expected earnings are calculated from the March Current Population Survey files for full-time, full-year workers using sample weights. The estimates equal what a man or woman would expect to earn working full-time, full-year over a career of 42 years, with a discount rate of 3%, assuming that college graduates delay the start of earnings for 4 years while in school. Earnings expectations are formed in each year by assuming that future high school and college graduates will have future earnings at each age equal to the average earnings of high school and college graduates (respectively) currently observed at each age; for example, expected earnings in 1980 are based on data across ages for 1980. Results for college-educated workers are net of 4 years of tuition and fees associated with appropriate year-specific values for public universities. Plotted points show the difference between expected earnings for college graduates and for high school graduates.

One should not, of course, take this model as irrefutable. A puzzling pattern evident in the data is that the rising demand for skilled workers appears to have slowed in the early 1990s, a phenomenon that is not anticipated by the "education race" model (*37*). This discrepancy underscores that the supply-demand model is necessarily incomplete—in part for the sake of expositional clarity and, in larger part, because our understanding of macroeconomic phenomena is typically imperfect. Nevertheless, the data speak sufficiently clearly to warrant two economic inferences. The first is that although popular accounts frequently assert that the United States is in the midst of a "college bubble"—too many students going to college at too high a cost—abundant economic evidence strongly suggests otherwise. Yes, college tuitions have risen far faster than inflation, and indeed, student debt has risen rapidly, with more than $100 billion in federal student aid dollars loaned in 2012–2013 alone (*38*). But the doubling of the college weekly wage differential over the past 30 years also implies that there have been sizable increases in the lifetime earnings of college graduates relative to high school graduates. How large are these gains? Figure 4, reproduced from (*39*), reports the estimated lifetime college earnings differential net of tuition for cohorts of students entering the labor market between 1965 and 2008. For both males and females, the expected net present value of a college degree relative to a high school diploma roughly tripled in this period, with the fastest gains accruing during the 1980s and 1990s. Note that this growing college/high school gap reflects the rising payoff to the 4-year college degree, the even steeper rise in the premium associated with graduate and professional degrees (see below), and the growing fraction of college graduates who obtain higher degrees; thus, an additional payoff to the college degree is that it opens the door to further specialization. This lifetime earnings differential would, of course, have risen further still if college tuitions had held steady rather than rising. But the inevitable sticker shock that households feel when confronting the cost of college should not obscure the fact that the real lifetime earnings premium to college education has likely never been higher (*40*).

The second positive economic news implied by Fig. 3 above is that the ongoing rise of skill differentials is not inevitable. Prior cohorts of U.S. students, particularly males, were slow to react to the rising return to education during the 1980s and 1990s, but the message appears to have finally gotten through. During the first decade of the 21st century, the U.S. high school graduation rate rose sharply after having been essentially stagnant since the late 1960s (*41*). This unanticipated rise was followed just a few years later by a surge in college completions. Between 2004 and 2012, the supply of new college graduates to the U.S. labor market rose at a rate not seen in several decades (Fig. 3A). As this influx of supply took hold, the college wage premium halted its enduring rise (Fig. 3B). What these observations and our simple

supply-demand model suggest is that the flattening of the college premium after 2005 is in large part a consequence of the quickening pace of educational attainment.

Inequality: Causes for Concern?

A market economy needs some inequality to create incentives. If, for example, students were not ultimately rewarded for spending their early adulthoods pursuing undergraduate, graduate, and professional degrees, or if the hardest-working and most productive workers were paid the same as the median worker, then citizens would have little incentive to develop expertise, to exert effort, or to excel in their work (*42*). Having acknowledged that some inequality is necessary, however, how can we gauge whether there is too much of it? I offer two analytical perspectives on this question.

Earnings Mobility

One metric by which to evaluate the consequences of inequality is via its relationship with economic mobility—that is, the degree to which individual economic fortunes change over time. Of particular interest is the degree of intergenerational mobility, meaning the likelihood that children born to low-income families become high-income adults and vice versa. High levels of economic inequality at a given point in time are not intrinsically inimical to economic mobility; a society with high inequality may be dynamic, with lots of movement up and down the economic ladder, and one with low inequality may be dynastic. But a natural concern is that high inequality at a point in time may serve to reduce mobility over time. If, for example, adults who became wealthy through hard work are able to "buy" success for their children through outsized investments and personal connections, while adults who are unproductive or unlucky in their careers are unable to muster the resources to foster their children's potential, then inequality of incomes could become self-perpetuating even if it originally emanates from high market returns to skill (*43*).

To understand the importance of high and rising U.S. inequality, it is therefore useful to ask how U.S. economic mobility compares to that of other developed countries, and whether U.S. mobility has fallen as inequality has risen. The answers to both questions will surprise many. Contrary to conventional civic mythology, U.S. intergenerational mobility is relatively low. The left panel of Fig. 5, reproduced from (*44*), which plots the relationship between cross-sectional inequality (x axis) and earnings mobility (y axis) among a set of 13 OECD member countries for which consistent data are available, documents that the United States has both the lowest mobility and highest inequality among all wealthy democratic countries. The right panel of Fig. 5, also sourced from (*44*), suggests one proximate explanation for this pattern: Countries with high returns to education tend to have relatively low mobility. Why, if education is "the great equalizer" in the words of Horace Mann, do high educational returns predict low mobility? A key reason is that educational attainment is highly persistent within families. Indeed, two of the strongest predictors of children's ultimate educational attainment are parental education and parental earnings (*45*, *46*). Hence, when the return to education is high, children of better-educated parents are doubly advantaged—by their parents' higher education and higher earnings—in attaining greater education while young and greater earnings in adulthood. Figure 5 therefore lends credence to the concern that rising inequality may erode economic mobility.

Has this erosion occurred? Surprisingly, the best evidence to date suggests that it has not. Evidence from Chetty *et al.* (*46*), documented in the supplementary material, underscores the message from Fig. 5 that there is substantial economic immobility in the United States. Children born three deciles apart in the household income distribution are on average one decile apart in the earnings distribution at age 29 or 30. Similarly, children born three deciles apart in the household income distribution differ by 20 percentage points in their probability of attending college at age 19 (relative to a mean of approximately 55%). Yet these data offer no evidence that mobility has appreciably changed among children born prior to the historic rise of U.S. inequality (1971–1974) and those born afterward (1991–1993). As far as we can measure, rising U.S. income inequality has not reduced intergenerational mobility so far. These findings, which also appear to hold over a longer historical time frame (*47*), suggest that U.S. mobility has not trended downward as many social scientists would have anticipated, and as

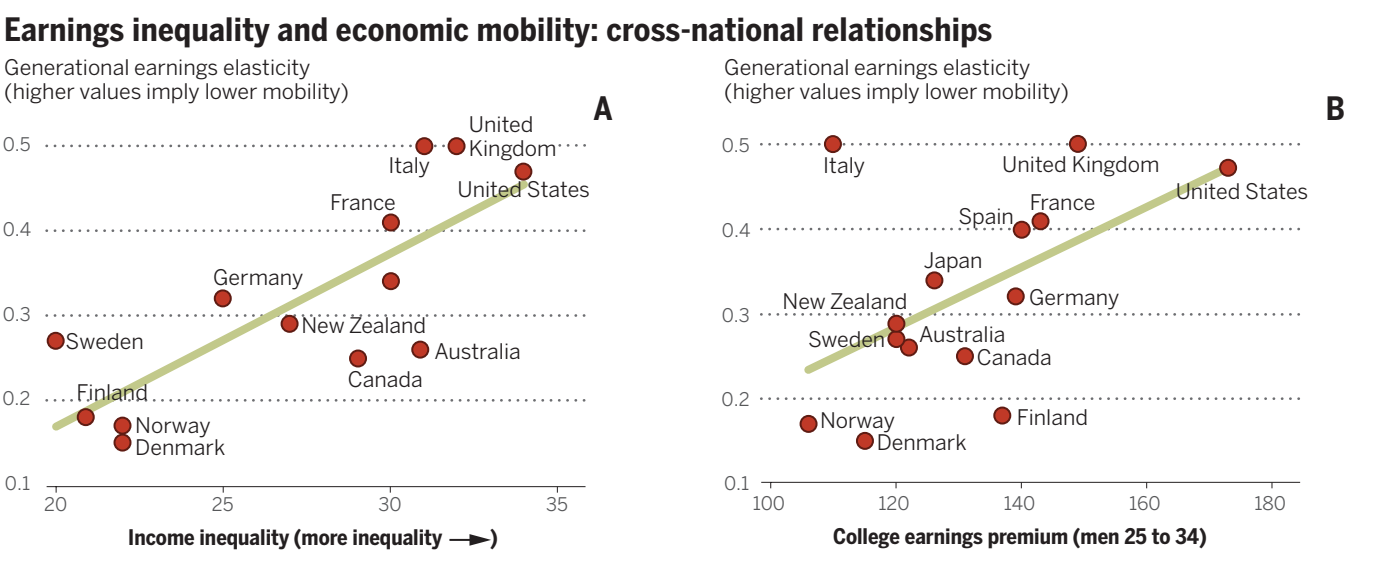


Fig. 5. Earnings inequality and economic mobility: Cross-national relationships. Reproduced from Corak [(*44*), figs. 1 and 4] with permission of the American Economic Association. In both panels, the mobility measure is equal to the intergenerational earnings "elasticity," meaning the average proportional increase in a son's adult earnings predicted by his father's adult earnings measured approximately three decades earlier. A higher intergenerational earnings elasticity therefore implies lower intergenerational mobility. In the left panel, cross-sectional income inequality is measured using a "Gini" index that ranges from 0 to 100, where 0 indicates complete equality of household incomes and 100 indicates maximal inequality (all income to one household). In the right panel, the college earnings premium refers to the ratio of average earnings of men 25 to 34 years of age with a college degree to the average earnings of those with a high school diploma, computed by the OECD using 2009 data. See (*44*) for further details.

many policymakers and popular accounts frequently assume.

It is important to interpret these results in context. The most recent birth cohorts whose adult outcomes can be observed at present were born no later than the early 1990s, which is still relatively early in the rise of U.S. inequality. Another 10 years of data, focusing on children born since 2000, may suggest a different conclusion. Moreover, the fact that mobility has stayed constant while inequality has risen means that the lifetime relative disadvantage of children born to low- versus high-income families has increased substantially; concretely, the rungs of the economic ladder have pulled farther apart but the chance of ascending the ladder has not improved. Finally, it is possible to interpret the fact that mobility has remained unchanged as evidence that U.S. mobility would have declined had it not been for the other compensatory steps taken by the federal government during this period, including, for example, expanding the Earned Income Tax Credit for low-income workers in the 1980s, enlarging the early childhood education Head Start program in the 1990s, and increasing federal student grant and loan programs to support college-going (*48*). Declines in racial and gender discrimination during this period likely also complemented these policies (*49*). A cautious read of the evidence is that although the United States is not a "land of opportunity" by conventional economic mobility metrics, it has not become less so in recent decades.

Real Earnings

A second gauge of economic health is the trajectory of earnings and employment. Here, the data present substantial cause for concern. Although the substantial college wage premium conveys the positive economic news that educational investments offer large returns, this wage premium also masks a discouraging truth: The rising relative earnings of workers with postsecondary education is not simply due to rising real earnings among college-educated workers but is also due to falling real earnings among non–college-educated workers. Between 1980 and 2012, real hourly earnings of full-time college-educated U.S. males rose anywhere from 20% to 56%, with the greatest gains among those with a postbaccalaureate degree (Fig. 6A). During the same period, real earnings of males with high school or lower educational levels declined substantially, falling by 22% among high school dropouts and 11% among high school graduates. Although the picture is generally brighter for females (Fig. 6B), real earnings growth among females without at least some college education over this three-decade interval was extremely modest.

Accompanying the fall in real wages among less educated workers has been a pronounced drop in their labor force participation rates, particularly among less educated males. Between 1979 and 2007, prior to the onset of the Great Recession, the fraction of working-age males in paid employment fell by 12 percentage points among high school dropouts and 10 percentage points among those with exactly a high school diploma. Conversely, employment rates were generally stable for males with postsecondary education and rose for females of all education levels except for high school dropouts.

The causes for the sharp falls in real earnings among non–college-educated workers are multiple. One likely force, as noted above, is the ongoing substitution of computer-intensive machinery for workers performing routine task-intensive jobs. This has depressed demand for workers in both blue-collar production and white-collar office, clerical, and administrative support positions, and has reduced the set of middle-skill career jobs available to non–college-educated workers more generally (*25*). A second factor is the globalization of labor markets, seen particularly in the greatly increased U.S. trade integration with developing countries. Globalization has become particularly important for U.S. labor markets since the early 1990s, when China began its extremely rapid integration into the world trading system. The influx of Chinese goods lowered consumer prices but also fomented a substantial decline in U.S. manufacturing employment, contributing directly to the decline in production worker employment (*50*). A third factor impinging on the earnings of non–college-educated males is the decline in the penetration and bargaining power of labor unions in the United States, which have historically obtained relatively generous wage and benefit packages for blue-collar workers. Over the past three decades, however, U.S. private-sector union density—that is, the fraction of private-sector workers who belong to labor unions—has fallen by approximately 70%, from 24% in 1973 to 7% in 2011 (*51*, *52*).

Notably, these three forces—technological change, deunionization, and globalization—work in tandem. Advances in information and communications technologies have directly changed job demands in U.S. workplaces while simultaneously facilitating the globalization of production by making it increasingly feasible and cost-effective for firms to source, monitor, and coordinate complex production processes at disparate locations worldwide. In turn, the globalization of production has increased competitive conditions for U.S. manufacturers and U.S. workers, eroding employment at unionized establishments and decreasing the capability

Changes in real wage levels of full-time U.S. workers by sex and education, 1963–2012

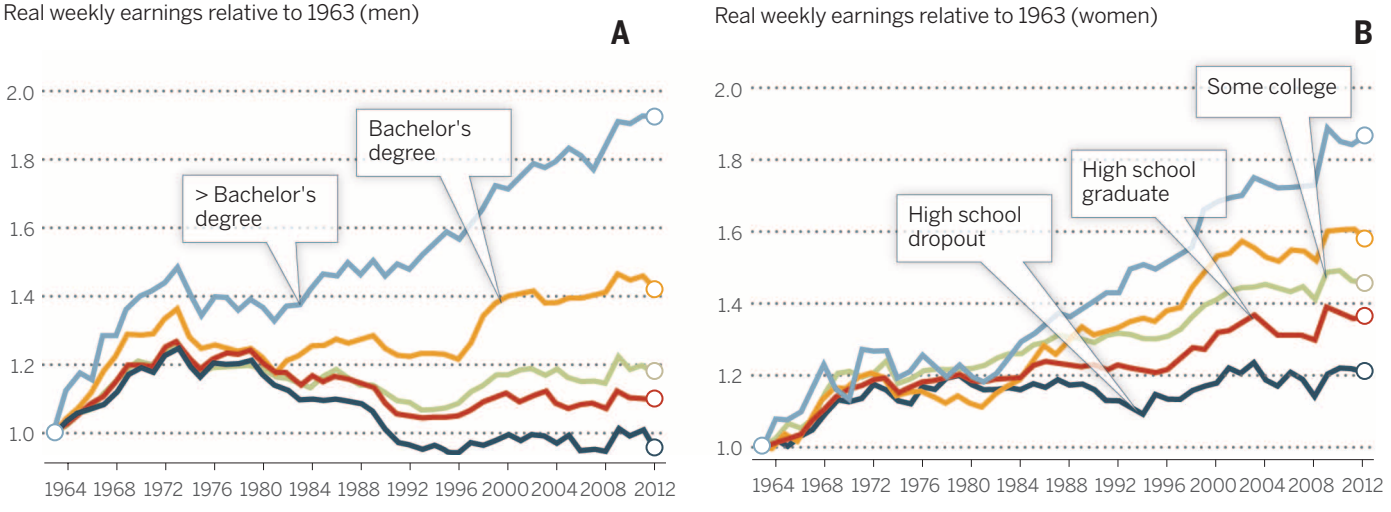


Fig. 6. Change in real wage levels of full-time workers by education, 1963–2012. (**A**) Male workers, (**B**) female workers. Data and sample construction are as in Fig. 3.

of unions to negotiate favorable contracts, attract new members, and penetrate new establishments.

In all cases, the foremost concern raised by these multiple forces impinging on the earnings of workers at different skill levels is not their impact on inequality per se, but rather their adverse effect on the real earnings and employment of less educated workers. These declines in both earnings and employment bode ill for the welfare of non–college-educated U.S. adults and are likely to have broader detrimental social consequences that frequently accompany non-employment: greater criminality, increased social dependency, and (more mundanely) reduced tax receipts.

Do Supply and Demand Make Policy Irrelevant?

One potential interpretation of the evidence above is that, because rising inequality is substantially a consequence of the impersonal forces of supply and demand, public policy has no role to play in shaping the trajectory of inequality or its social impact. This conclusion is incorrect for two reasons. First, there are multiple channels by which policy has contributed to the rise of U.S. inequality, many of which are not fully evident in the education earnings premium. These include the fall over several decades in the real value of the U.S. minimum wage (*7*); the declining prevalence and bargaining power of U.S. labor unions; mounting international competition that places particular pressure on the wages and employment of less educated workers; and sharp reductions in top federal marginal tax rates that have raised after-tax inequality and increased the incentive of highly paid workers to seek still higher compensation. As discussed in the companion paper by Piketty and Saez, there is also disagreement among economists about whether the rising share of household incomes accruing to the top few percentiles of households in numerous developed countries over the past several decades is also primarily a market phenomenon, or instead reflects changing social norms, growing corporate misgovernance, slackening regulatory oversight, or increasing political capture of the policymaking process by elites (*3*–*6*). It would therefore be a vast overstatement to conclude that the rise of U.S. inequality is exclusively due to conventional market forces, or that public policy has not played a role.

But let us assume for the sake of argument that the rise of income inequality is entirely a market phenomenon. Would this imply that there is no role for public policy? A moment's reflection suggests otherwise. As the economist Arthur Goldberger once famously observed, the fact that nearsightedness is substantially a genetic disorder has no bearing on whether doctors should prescribe eyeglasses (*53*). What is relevant is whether the benefits of addressing myopia exceed the costs. In the case of myopia, the availability of eyeglasses make this an easy call.

Although there is no "remedy" for inequality that is as swift or cheap as eyeglasses, prosperous democratic countries have numerous effective policy levers for shaping inequality's trajectory and socioeconomic consequences. Policies that appear most effective over the long haul in raising prosperity and reducing inequality are those that cultivate the skills of successive generations: excellent preschool through high school education; broad access to postsecondary education; and good nutrition, good public health, and high-quality home environments. Such policies address inequality from two directions: (i) enabling a larger fraction of adults to attain high productivity, rewarding jobs, and a reasonable standard of living; and (ii) raising the total supply of skills available to the economy, which in turn moderates the skill premium and reduces inequality (*54*).

Of course, building skills is a multigenerational process and thus has little impact on inequality in the short term. There are, however, numerous nearer-term levers that moderate inequality directly without imposing substantial economic costs: applying progressive tax and transfer policies that fund public investments and foster opportunities for children of all socioeconomic backgrounds; applying well-crafted labor regulations that ensure safe and non-exploitive working conditions; providing wage subsidies such as the Earned Income Tax Credit that increase the payoff to employment for those with limited skills; setting modest but nonzero minimum wage rules; and offering numerous social insurance policies (health and disability insurance, flood insurance, disaster assistance, food assistance) that buffer misfortune for the unfortunate. Although it is outside the scope of this article to evaluate these policies, it is critical to underscore that policy and governance has played and should continue to play a central role in shaping inequality—even when a central cause of rising inequality is the changing supply and demand for skills.

REFERENCES AND NOTES

1. C. D. Goldin, L. F. Katz, *Brookings Pap. Econ. Act.* (fall), 135 (2007).
2. Goldin and Katz (*1*) found that the increase in the education wage premium, particularly the college premium, explains about 60 to 70% of the rise in wage inequality (variance) between 1980 and 2005.
3. F. Alvaredo *et al.*, *J. Econ. Perspect.* **27**, 3–20 (2013).
4. J. Bivens, L. Mishel, *J. Econ. Perspect.* **27**, 57–78 (2013).
5. A. Bonica *et al.*, *J. Econ. Perspect.* **27**, 103–124 (2013).
6. S. N. Kaplan, J. Rauh, *J. Econ. Perspect.* **27**, 35–56 (2013).
7. D. Autor *et al.*, *The Contribution of the Minimum Wage to U.S. Wage Inequality over Three Decades: A Reassessment* (NBER Working Paper 16533, Cambridge, MA, 2010).
8. T. Piketty, E. Saez, *Q. J. Econ.* **118**, 1–41 (2003).
9. These calculations use data from (*8*), with data updated to 2012 available at http://elsa.berkeley.edu/saez/~Tab-Fig.2012prel.xls. Average U.S. household incomes, including the top 1%, rose by 20.2%, while the average household income of the bottom 99% of households rose by only 3.5%.
10. Thus, the top 1% maintains its share of household income at a constant 10.0% while average household incomes rise by 20.2%, as actually occurred.
11. This point is due to Lawrence Katz of Harvard University, who offers these calculations in his graduate labor economics lecture notes.
12. T. Lemieux, *Post-Secondary Education and Increasing Wage Inequality* (Working Paper 12077, National Bureau of Economic Research, 2006).
13. S. Firpo *et al.*, Decomposition methods in economics. In *Handbook of Labor Economics*, D. Card, O. Ashenfelter, Eds. (Elsevier-North Holland, Amsterdam, 2011), vol. 4, pp. 1–102.
14. See www.oecd.org/site/piaac/surveyofadultskills.htm for more information. The PIAAC program will encompass 33 countries, but data for only 22 were available at this writing.
15. E. A. Hanushek, G. Schwerdt, S.Wiederhold, L.Woessmann, *Returns to Skills Around the World: Evidence from PIAAC* (NBER Working Paper 19762, Cambridge, MA, 2013).
16. Hanushek *et al.* (*15*) also found that the correlation between numeracy skills and years of schooling is 0.45. When including both numeracy skills and years of schooling in an earnings regression, they found that both are substantial and significant predictors of earnings, although each is attenuated relative to a model where only one factor is included at a time. This pattern of results suggests, logically, that neither test scores nor years of schooling is a complete measure of labor market skills.
17. T. D. Snyder, *120 Years of American Education: A Statistical Portrait* (National Center for Education Statistics, U.S. Department of Education, 1993).
18. L. D. Johnston, "History lessons: Understanding the decline in manufacturing." *MinnPost*, 22 February 2012; www.minnpost.com/macro-micro-minnesota/2012/02/history-lessons-understanding-decline-manufacturing.
19. C. Goldin, L. F. Katz, *The Race Between Education and Technology* (Harvard Univ. Press, Cambridge, MA, 2008).
20. M. Stanley, *Q. J. Econ.* **118**, 671–708 (2003).
21. B. Pierce, in *Labor in the New Economy*, K. G. Abraham, J. R. Spletzer, M. Harper, Eds. (Univ. of Chicago Press, Chicago, 2010), pp. 63–98.
22. These comparisons hold labor market experience and gender constant. This doubling of the college premium very likely understates the magnitude of the increase in inequality between college-educated and non–college-educated workers. Alongside higher hourly earnings, college-educated workers enjoy greater job stability, lower rates of unemployment, more generous fringe benefits, and better working conditions; Pierce (*21*) found that these differentials have generally increased in the same time period.
23. D. H. Autor *et al.*, *Q. J. Econ.* **118**, 1279–1333 (2003).
24. D. Acemoglu, D. H. Autor, Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, D. Card, O. Ashenfelter, Eds. (Elsevier-North Holland, Amsterdam, 2011), vol. 4, pp. 1043–1171.
25. D. H. Autor, D. Dorn, *Am. Econ. Rev.* **103**, 1553–1597 (2013).
26. M. Goos *et al.*, www.aeaweb.org/forthcoming/output/accepted_AER.php
27. Extensive recent literature, commencing with Autor *et al.* (*23*) and summarized in Acemoglu and Autor (*24*), considers the role of technological change in displacing workers performing routine tasks and complementing workers performing nonroutine tasks. An additional implication of this framework is that an increasing share of employment will be found in comparatively low-skill nonroutine manual tasks that require situational adaptability, visual and language recognition, and in-person interactions but limited formal education (e.g., janitors and cleaners, home health aides, construction laborers, and security personnel). See Autor and Dorn (*25*) and Goos *et al.* (*26*) for evidence that employment in the U.S. and among OECD member countries has increasingly polarized into high-paid, abstract-intensive occupations and low-paid, manual-intensive occupations.
28. J. Tinbergen, *Kyklos* **27**, 217–226 (1974).
29. Details of this model are given in the online supplement.
30. D. Card, T. Lemieux, *Am. Econ. Rev.* **91**, 97–102 (2001).
31. L. F. Katz, K. M. Murphy, *Q. J. Econ.* **107**, 35–78 (1992).
32. L. Katz, D. H. Autor, Changes in the wage structure and earnings inequality. In *Handbook of Labor Economics*, D. Card, O. Ashenfelter, Eds. (Elsevier-North Holland, Amsterdam, 1999), vol. 3, pp. 1463–1555.
33. D. Card, T. Lemieux, *Q. J. Econ.* **116**, 705–746 (2001).
34. D. H. Autor *et al.*, *Rev. Econ. Stat.* **90**, 300–323 (2008).
35. E. Crivellaro, "College wage premium over time: Trends in Europe in the last 15 years." University Ca' Foscari of Venice, Department of Economics Research Paper Series no. 03/WP/2014 (2014); http://dx.doi.org/10.2139/ssrn.2383795.
36. Summarizing evidence on the college premium in 12 European countries between 1994 and 1999, Crivellaro (*35*) found a pattern of increasing skill differentials except in countries that have had a large increase in the relative supply of college

graduates, a pattern consistent with the conceptual model laid out below.
37. Although this deceleration is not evident from Fig. 3, it is detected by the regression equation, as discussed in the online supplement.
38. College Board, *Trends in Student Aid: 2013* (College Board, New York, 2013).
39. C. Avery, S. Turner, *J. Econ. Perspect.* **26**, 165–192 (2012).
40. Three sources of uncertainty should be kept in mind when interpreting these estimates. First, they encompass substantial heterogeneity. Although the average college graduate earns substantially more than the average high school graduate, the least successful college graduates may earn substantially less than the median among high school graduates, and the most successful high school graduates may earn substantially more than the median among college graduates. Second, for students who acquire substantial student debt but do not complete the college degree, it is far from certain that college will prove a good investment. Finally, these calculations assume that the lifetime profile of earnings observed in the year of college graduation will persist throughout the career. As Fig. 3 indicates, this premium has changed substantially over time, so this assumption is only a rough approximation. However, the college premium is so high at present that even with a substantial decline, college would remain an attractive financial proposition on average from a lifetime earnings perspective.
41. R. J. Murnane, *J. Econ. Lit.* **51**, 370–422 (2013).
42. D. Acemoglu, J. Robinson, Why Nations Fail (Crown, New York, 2012).
43. As with cross-sectional inequality, there is no economically "ideal" level of intergenerational mobility. Even in a society with perfect equality of opportunity, one would expect children of successful parents to have above-average success as adults, simply because many attributes that contribute to success (appearance, intellect, athleticism) are partly heritable.
44. M. Corak, *J. Econ. Perspect.* **27**, 79–102 (2013).
45. S. F. Reardon, The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. In *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, G. J. Duncan, R. J. Murnane, Eds. (Russell Sage Foundation, New York, 2011), pp. 91–115.
46. R. Chetty *et al.*, *Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility* (NBER Working Paper No. 19844, Cambridge, MA, 2014).
47. C.-I. Lee, G. Solon, *Rev. Econ. Stat.* **91**, 766–772 (2009).
48. Between 2002–2003 and 2012–2013, the sum of federal Pell Grants and loans for higher education increased by 105%, from $83 billion to $170 billion in constant 2012 dollars [(*38*), table 1].
49. C.-T. Hsieh *et al.*, *The Allocation of Talent and U.S. Economic Growth* (NBER Working Paper No. 18693, Cambridge, MA, 2013).
50. D. H. Autor *et al.*, *Am. Econ. Rev.* **103**, 2121–2168 (2013).
51. D. Card *et al.*, *J. Labor Res.* **25**, 519–559 (2004).
52. B. T. Hirsch, *J. Econ. Perspect.* **22**, 153–176 (2008).
53. A. S. Goldberger, *Economica* **46**, 327 (1979).
54. The extensive involvement of state and federal government in education at all levels also underscores the fact that the distribution of education and skills today is in no sense a "free market" outcome; it is a consequence of both individual and public choices.

ACKNOWLEDGMENTS

I thank D. Acemoglu, L. Katz, J. Van Reenen, M. Tatsutani, and two anonymous referees for valuable comments and advice, and C. Patterson and B. Price for expert research assistance. Supported by NSF grant SES-1227334, Russell Sage Foundation grant 85-12-07, and Alfred P. Sloan Foundation grant 2011-10-12. All data and code that are unique to this article (Figs. 1, 3, and 6; fig. S2) are available from the author. All other figures (Figs. 2, 4, and 5; figs. S1 and S3) are reproduced from other publications, as noted, with permission of the authors.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/843/suppl/DC1
Supplementary Text
Figs. S1 to S3
References (*55*–*61*)

10.1126/science.1251868

REVIEW

Income inequality in the developing world

Martin Ravallion

Should income inequality be of concern in developing countries? New data reveal less income inequality in the developing world than 30 years ago. However, this is due to falling inequality between countries. Average inequality within developing countries has been slowly rising, though staying fairly flat since 2000. As a rule, higher rates of growth in average incomes have not put upward pressure on inequality within countries. Growth has generally helped reduce the incidence of absolute poverty, but less so in more unequal countries. High inequality also threatens to stall future progress against poverty by attenuating growth prospects. Perceptions of rising absolute gaps in living standards between the rich and the poor in growing economies are also consistent with the evidence.

Development economics emerged as a subdiscipline of economics in the 1950s, and its initial focus was on economic growth, with inequality as a secondary concern. The prevailing orthodoxy for many decades was that a period of rising inequality was to be expected in growing poor countries. Rising inequality was seen to be more or less inevitable and not something to worry about, particularly if the incidence of poverty was falling. Another commonly held view was that policy efforts to reduce inequality were likely to impede growth and (hence) poverty reduction.

The existence of high inequality within many developing countries, side by side with persistent poverty, started to attract attention in the early 1970s. Nonetheless, through the 1980s and well into the 1990s, the mainstream view in development economics was still that high and/or rising inequality in poor countries was a far less important concern than assuring sufficient growth, which was the key to poverty reduction. The policy message for the developing world was clear: You cannot expect to have both lower poverty and less inequality while you remain poor, and, if you choose to give poverty reduction highest priority, then focus on growth.

Other objections could still be raised to high income inequality. The classical utilitarian formulation—whereby social welfare is judged by the sum of utilities, assuming diminishing marginal utility of income—pointed to social welfare losses from high inequality at a given mean. But that did not persuade those who believed that there was a trade-off between equity and growth. A moral defense could also be mounted for the view that inequality is not an important issue for a growing developing country by appeal to John Rawls's "difference principle" that (subject to assuring liberty and equal opportunity) higher inequality can be justified as long as it benefits the worst-off group in society (*1*).

The period since 2000 has seen a deeper and more widespread questioning of this long-standing view of pro-poor inequality. New concerns have emerged about the instrumental importance of equity to other valued goals, including poverty reduction and human development more broadly. It appears more likely today that high inequality will be seen as a threat to future development than as an inevitable and unimportant consequence of past progress. The long-standing idea of a substantial growth-equity trade-off has come to be seriously questioned.

This paper reports new estimates of the levels and changes in income inequality measures for the developing world. The new estimates take us up to 2010, embracing the period of higher growth rates in the developing world since the turn of the millennium. In the light of these new data, I revisit past and ongoing debates on inequality in developing countries and the trade-offs with growth and poverty reduction.

Income Inequality Measures

To measure inequality in the developing world as a whole, one ignores country borders—pooling all residents and measuring inequality among them. This overall measure will naturally depend on the inequality between countries as well as that within them. Thus, its evolution over time will depend on whether poorer countries are seeing lower growth rates as well as the things happening within countries—economic changes and policies—that affect inequality.

If we are comparing country or regional performance, then we want to isolate the within-country component of inequality as distinct from that between countries. Although there are many inequality measures, not all of them allow a clean separation of the between and within components. For example, such a decomposition is

Department of Economics, Georgetown University, Washington, DC 20057, and National Bureau of Economic Research, Cambridge, MA 02138, USA. **E-mail: mr1185@georgetown.edu**

not generally possible for the Gini index—a popular inequality index based on the average absolute difference between all random pairs of incomes normalized by the mean. (The exception is when the distributions of different countries do not share any common support, which is unlikely.)

The mean-log deviation (MLD) offers an elegant solution. This is given by the (appropriately weighted) mean across households of the log of the ratio of the overall mean income to household income per person. When all incomes are equal, MLD is 0; the higher the inequality, the higher the MLD. Like the Gini index, MLD is a relative inequality measure in that it depends on ratios of incomes to the mean; this is implied by the scale-invariance axiom of inequality measurement, which says that when all incomes are multiplied by a constant the inequality measure does not change. (We will return to this axiom, which can be questioned.) Also, both measures satisfy the standard transfer axiom for inequality measurement, namely that a small transfer to someone with a lower income will reduce inequality. However, unlike the Gini index, MLD is exactly decomposable by population subgroups.

Inequality is measured by using household surveys. The estimates presented here have their basis in ~900 surveys spanning 1980 to 2010 and 130 countries. The calculations have been done on an updated version of the data set used by (*2*) for measuring poverty, and (when relevant) the same methods have been used. For over half the surveys, household consumption expenditure (including imputed values for consumption from own-production) was used rather than income, although for brevity the word "income" is used for both. For a given economy, income inequality measures are expected to be somewhat higher than for consumption given the scope for smoothing consumption in the presence of income shocks.

Although household surveys are the only source of data on inequality across households, there are two sources of data on the growth rates in average household consumption or income. The same surveys used to measure inequality compose one source, and the other is the household consumption component of domestic absorption in the national accounts. There is no clear way of ranking these measures; in some developing countries, the measures based on national accounts are considered quite unreliable, whereas elsewhere there are serious concerns about the survey-based measures.

Household surveys may well underestimate the extent of inequality, notably through either the rich underreporting their incomes or through selective compliance in the randomized assignments of the survey instrument whereby richer households are less likely to participate. [The latter problem does not imply that inequality will be underestimated; Korinek *et al.* (*3*) found that selective compliance leads to an underestimation of inequality in the United States.] Although there are methods that can be used to address these concerns, they have so far tended to be confined to research studies, and mostly in rich countries. No corrections for these data problems have been made in the primary data used here.

The overall MLD for the developing world in 2010 is estimated to be 0.578. To help interpret this number, imagine a distribution with three incomes: 1, 2, and x; for this distribution to have MLD = 0.578, x would need to be 12.73, that is, the richest of three people would need to have over 12 times that of the poorest and 6 times that of the middle income.

Inequality has fallen over this 30-year period. For the earliest year that the data permit an estimate, 1981, MLD was 0.651. (This is equivalent to x = 14.79 in the distribution 1, 2, and x.) Fig. 1 plots MLD for the developing world as a whole and its between-country component. (The supplementary materials give detailed estimates.)

We see that the overall decrease came with ups and downs and an increase over 2005 to 2010. The variance over time is largely accountable to inequality between countries. Over the period as a whole, the between-country component has fallen, while the within-country component has risen. The latter accounted for 31% of total inequality in the developing world in 1981 but 47% in 2010. However, this pattern has reversed since 2000, with inequality rising between countries but falling on average within countries. Something different seems to be happening since the turn of the millennium. The next section will consider the role played by the higher growth rates since then.

Figure 2 plots the within-country component by region. Latin America and the Caribbean (LAC) have persistently had the highest average inequality of any region. (The mean difference between LAC and other regions falls

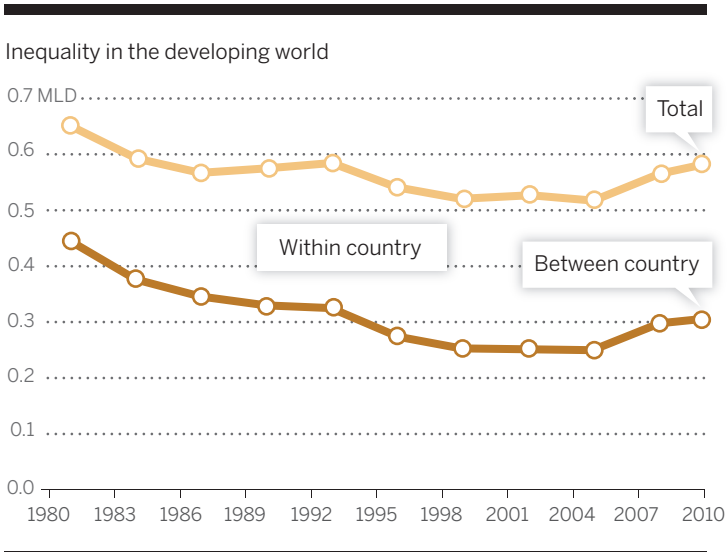


Fig. 1. Inequality in the developing world. Between- and within-country totals are shown.

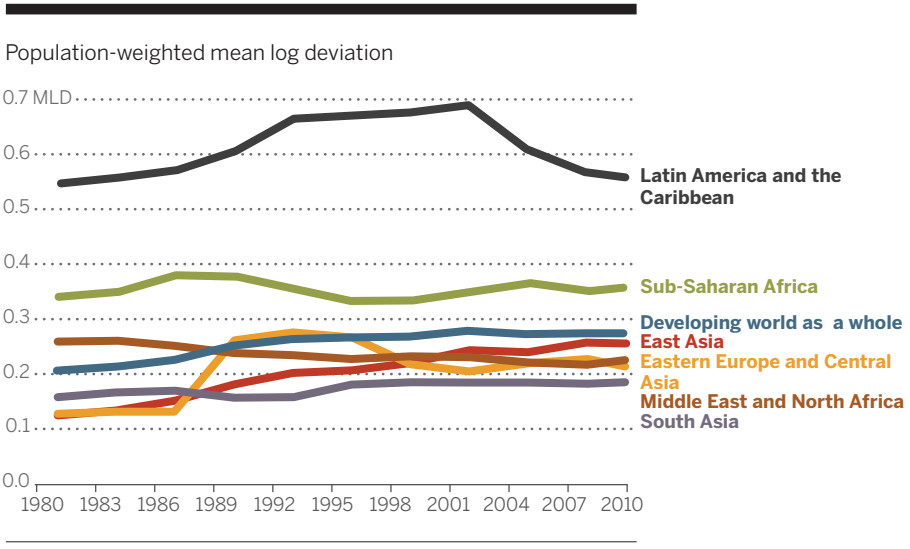


Fig. 2. Population-weighted MLD. Developing-world countries are grouped geographically.

only slightly when one controls for income surveys, which have been more popular in LAC; see supplementary materials.) Inequality was rising in LAC until around 2000 but has fallen noticeably since. Eastern Europe and Central Asia saw a sharp rise in inequality in the 1990s but has seen generally falling inequality since. Sub-Saharan Africa has the second-highest average inequality, although with no clear trend. South Asia has been a region of low inequality, rising somewhat since the early 1990s. East Asia started out as the region with the lowest inequality but has seen a steady rise. The Middle East and North Africa has seen steadily falling inequality.

Inequality and Growth Across Countries

As is well known, the developing world has seen substantially higher growth rates since 2000. Have the processes of stronger economic growth in developing countries, and the acceleration of growth seen since the turn of the century, put upward pressure on inequality?

The recent signs of rising inequality in the developing world as a whole have coincided with the higher growth rates since 2000 (Fig. 3). However, when we unpack the relationship (exploiting the properties of the MLD), we find two opposing forces: As can be seen in Fig. 4, the higher growth rates have come with higher between-country inequality but lower average within-country inequality. Across the developing world as a whole and taking 3-year periods back to 1981, the periods of higher growth have come with higher between-country inequality (r = 0.70), whereas inequality within developing countries has been falling with higher growth (r = −0.63). But the latter relationship is rather flat; going from zero growth rate to a 5% annual rate is associated with a move from an expected rise in MLD of 0.005 (about one-third of the standard deviation in the annualized change in MLD) to no change on average. Growth in average household income appears to be close to inequality-neutral on average.

These are aggregates across countries. I also assembled a data set of the longest spells between two household surveys, and I obtained complete data for about 100 countries. The median year of the first survey is 1989, and it is 2008 for the latest survey. The (unweighted) mean MLD is unchanged at 0.318. (Population-weighted values are slightly lower.) However, the variance has fallen, from a standard deviation of 0.195 to 0.171 for the latest year. Inequality fell over time in 11 of the 13 countries that had MLD greater than 0.576—the aggregate value (including the between country component) in 1990. The annualized change in MLD is negatively correlated with its initial value (r = −0.33, n = 122), and the same is true for the Gini index (r = −0.55; n = 122). This pattern of inequality convergence is consistent with other evidence (*4*, *5*). Although measurement errors exaggerate the signs of convergence, it remains robust to the use of an econometric estimator that addresses this concern (*5*).

These data suggest a small negative correlation between changes in inequality and growth rates among developing countries. The correlation coefficient between the growth rate (annualized log difference) in the survey mean and the annualized change in MLD is −0.24, which is significant at the 5% level. The use of growth rates from the national accounts instead of surveys drops the correlation to −0.04. Use of the Gini index rather than MLD gives similar results. Nor did I find evidence of nonlinearity in the relationship; by regressing the change in inequality on both the growth rate and the annualized change in the squared log mean, I found neither coefficient was significant. (See the supplementary materials for details.) So, again, inequality increases about as often as it falls during spells of growth. This finding is consistent with past evidence that used cross-country comparisons [as surveyed in (*6*)].

It is not then surprising that there is a strong negative correlation between growth rates and changes in absolute poverty. Across the regional averages used above, the correlation coefficient is −0.75 between growth rates in mean household consumption or income and the annualized changes in the log of the poverty rate for a poverty line of $1.25 a day at purchasing power parity. Across countries the corresponding correlation coefficient is −0.85. This is consistent with a large body of cross-country evidence back to around 1990, as reviewed by (*6*).

There are a number of caveats that should be noted about this negative correlation. There are measurement errors to consider. There are some rather extreme values in some of the country data that may well reflect such errors. However, the correlation is still high at −0.70 (n = 88) if one trims extreme values (dropping absolute log differences per year over 0.5). The correlation is also robust to the use of an econometric correction for correlated measurement errors between the poverty measures and the survey means (supplementary materials). The correlation is lower if one uses growth rates from national accounts instead of those from surveys. (The correlation between annualized log differences in the $1.25 a day poverty rate and the growth rates in private consumption per capita from national accounts is −0.56.) Some of the things that are included in the national accounts do not get passed on quickly (or at all) to household living standards. Also, survey dates do not cover entire years as do national accounts. Another caveat is that the correlation tends to be much weaker with use of relative poverty measures, whereby the poverty line rises with average income (consistently with the cross-country relationship between national poverty lines and average income); see (*7*).

An important caveat is that the fact that we do not see rising inequality on average in growing economies does not imply that inequality can safely be ignored. The rest of this paper points to a number of reasons why inequality should be considered a central development issue.

Inequality Matters to Progress Against Poverty

Does higher inequality in poor countries permit faster progress against poverty through economic growth? Most development economists would probably have answered "yes" even 10 years ago, but recent development theories and evidence are more suggestive of a negative answer.

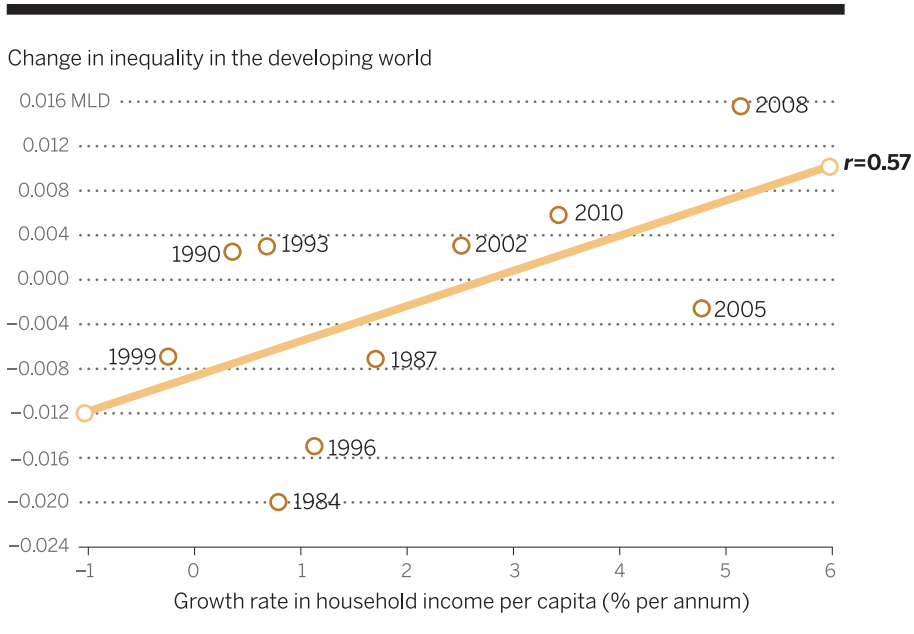


Fig. 3. Change in inequality over time. Change in MLD plotted as a function of the growth rate in household income per capita.

The performance of developing countries against poverty is quite diverse. Inequality comes back into the story in efforts to explain this diverse performance. We have learned that inequality plays three important roles in influencing the pace of progress against poverty. First, changes in inequality during the growth process have implications for how much that growth affects poverty. By using the country-level data set discussed in the last section, I find that, among growing developing countries in terms of mean household income based on the surveys, those experiencing falling inequality see the $1.25 a day poverty rate falling at a median rate of about 1.30% points per year versus a median fall of only 0.42% points per year for countries with rising inequality. Either way, poverty incidence tends to fall but at very different rates.

Second, initial inequality reduces the growth elasticity of poverty reduction—the responsiveness of poverty measures to growth in mean income. This is intuitive: The more unequal the original distribution, the smaller the share of the growth accruing to the poor and the lower the poverty reduction arising from that growth; this was demonstrated empirically by (*8*). The converse holds too: In more unequal societies, the poor tend to be more protected from aggregate economic contractions.

Third, even if inequality does not rise during a period of economic growth, a high initial level of inequality can mean less growth and (hence) less progress against absolute poverty. In the 1990s, we started to see various theoretical arguments that high levels of inequality stifled investment, innovation, and reform. Here, I only provide a sketch of the arguments that have been made; other studies (*9*, *10*) surveyed the arguments and evidence in the literature on how the initial level of inequality influences the subsequent growth rate.

An influential strand of this literature points to the implications of borrowing constraints associated with asymmetric information and the inability to write binding enforceable contracts. Credit market failure that disproportionately affects poor people leaves unexploited opportunities for their investment in physical and human capital, and it is assumed that there are diminishing returns to capital, such that poor people have higher marginal products of capital. (This idea can be extended to also embrace technical innovation, assuming that everyone gets new ideas but that the poor are more constrained in responding.) Then higher current inequality implies lower future mean wealth at a given value of current mean wealth.

Other sources of economic distortions can create costly inequalities. This can happen if a relatively privileged subgroup is able to restrict entry to economic opportunities (including jobs) and thus set the returns to those activities above the market clearing level. Labor-market failures in the form of persistent unemployment can also have lasting adverse consequences for both equity and efficiency. Human capital is developed in part by working; thus, longer spells of unemployment create a de-skilling that makes it harder to get a job. Unemployment can also be associated with psychological distress and depression. This psychological scarring may also make it harder to get a job.

Another class of models is based on the idea that high inequality restricts efficiency-enhancing cooperation such that key public goods are underprovided, including legally secure property rights, or desirable economic and political reforms are blocked. High initial inequality can also induce costly political economy responses.

Some of the literature has pointed to other aspects of the initial distribution of income besides inequality. It has been argued that a larger middle class helps assure a more diversified economy (especially through greater demand for consumer goods) and that the middle class also tends to be a stronger political force for pro-growth economic reforms. Others have argued that higher current wealth poverty (of which access to land is a key factor) impedes growth, such as through access to credit.

A strand of this new empirical literature on economic growth has tested for inequality as an initial condition, and the results have generally supported the view that higher initial inequality impedes growth. The effect is quantitatively large,

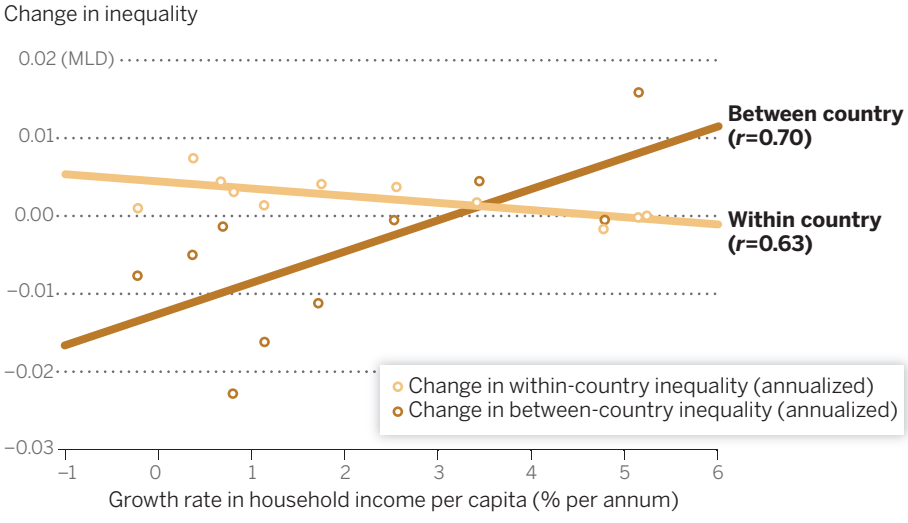


Fig. 4. Changes in inequality between and within countries.

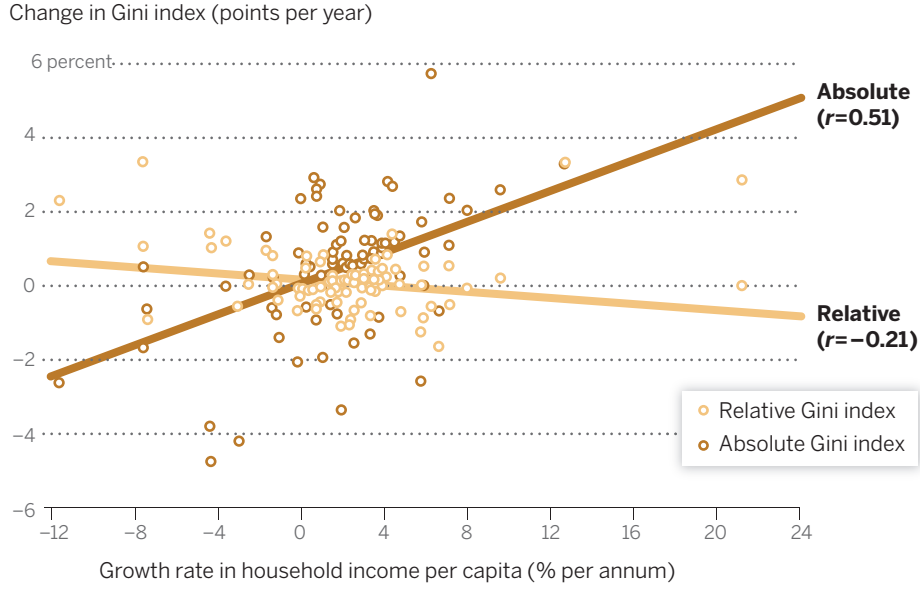


Fig. 5. Changes in Gini index per year. With growth in mean income, the absolute Gini index has tended to rise, whereas the relative index has changed little.

as well as statistically significant; two recent examples are (*11*, *12*).

We have seen that the literature has pointed to various aspects of initial distribution—inequality, the size of the middle class, or poverty—that matter to progress against poverty. These distributional measures tend to be correlated with each other, yet very few studies have tested encompassing models that try to see whether one measure is more important than another. One exception (*13*) finds that high poverty at a given initial mean matters more to rates of growth in mean consumption in developing countries than inequality or measures of the middle class or polarization. This does not imply that inequality is unimportant but rather tells us that inequality matters to growth in poor countries mainly via its bearing on the extent of initial poverty.

So, the arguments and evidence from modern development economics do not suggest that we should expect any substantial aggregate trade-off between progress against absolute poverty and progress in reducing inequality. Indeed, the evidence suggests that falling inequality tends to come with falling poverty incidence (*14*).

Differing Concepts of Inequality

The value judgments made in measuring inequality carry weight for the position one takes on whether economic growth tends to be inequality increasing or not (*15*). So far, this paper has followed the long-standing practice of relying on relative inequality measures, defined in terms of ratios of incomes or consumptions. By contrast, absolute inequality depends on the abolute differences in levels of living. If the distribution of income among two people changes from $1000 and $10,000 to $2000 and $20,000, then relative inequality is unchanged, yet the absolute gap between rich and poor has doubled.

Perceptions on the ground that inequality is rising often appear to be referring to the absolute concept. Amiel and Cowell (*16*) report experiments to identify which concept of inequality is held by people. They found that 40% of the university students surveyed (in the United Kingdom and Israel) thought about inequality in absolute rather than relative terms. [Harrison and Seidl (*17*) report similar findings for a large sample of German university students.] For the purpose of this paper, I fielded a similar (confidential) survey to my class of undergraduates at Georgetown University. From the 130 responses (out of 150 students), the class was roughly evenly split between those who thought about inequality in relative terms and those who thought about it in absolute terms.

It is not that one concept is right and one wrong. They simply reflect different value judgments. The relative inequality concept is implied by the scale invariance axiom, whereas the absolute concept is implied by an alternative axiom called translation invariance. The former axiom has dominated practice in the measurement of inequality by economists and statisticians, but the "axiom" is hardly a self-evident truth.

In this light, let us return to the long-standing debates on growth and equity. Finding that the share of income going to the poor does not change on average with growth does not mean that "growth raises the incomes (of the poor) by about as much as it raises the incomes of everybody else," as claimed by the *Economist* (*18*). Given existing inequality, the absolute income gains to the rich from inequality-neutral growth will of course be greater than the gains to the poor. For example, for the richest decile in India, the income gain from aggregate growth will be about four times higher than the gain to the poorest quintile; it will be 15 to 20 times higher in Brazil or South Africa.

The empirical finding in the literature that growth tends to be inequality-neutral within developing countries will carry little weight for those concerned about absolute inequality, who prefer translation invariance to scale invariance. One expects an absolute measure to rise with growth and fall with contraction. I confirmed this on the aforementioned country-level data set. Changes in the absolute Gini index show a significant positive correlation with growth rates in either survey means (r = 0.51, r = 123) or consumption per capita from the national accounts (r = 0.75, n = 114). Figure 5 plots the relationships. (Absolute Gini indices are scaled by the mean of initial and final years.)

It may well be that past and ongoing debates about the distribution of the gains from growth in the developing world rest in no small measure on this (rarely discussed) conceptual difference in how inequality is defined. Unlike those who see inequality as relative, those who view it in absolute terms will expect to see a trade-off between reducing inequality and reducing poverty. That does not mean that any policy that is good for one is necessarily bad for the other or that it is impossible to have both; the correlation is just that—a correlation. However, it does help us understand why some growth-promoting and poverty-reducing policy reforms may well come in for serious criticism and even be blocked by a non-negligible number of observers concerned about widening gaps in living standards between the rich and the poor. How policy-makers deal with that critique may matter greatly to progress against poverty.

REFERENCES AND NOTES

1. J. Rawls, *A Theory of Justice* (Harvard Univ. Press, Cambridge, MA, 1971).
2. S. Chen, M. Ravallion, *Q. J. Econ.* **125**, 1577–1625 (2010).
3. A. Korinek, J. Mistiaen, M. Ravallion, *J. Econ. Inequal.* **4**, 33–55 (2006).
4. R. Bénabou, in *National Bureau of Economic Research Macroeconomics Annual*, B. Bernanke, J. Rotemberg, Eds. (MIT Press, Cambridge, MA, 1996), pp. 11–92.
5. M. Ravallion, *Econ. Lett.* **80**, 351–356 (2003).
6. F. H. G. Ferreira, M. Ravallion, in *The Oxford Handbook of Economic Inequality*, W. Salverda, B. Nolan, T. Smeeding, Eds. (Oxford Univ. Press, Oxford, 2009), pp. 599–638.
7. S. Chen, M. Ravallion, *Review of Income and Wealth* **59**, 1–28 (2013).
8. M. Ravallion, in *Inequality and Poverty Re-Examined*, J. Micklewright, S. Jenkins, Eds. (Oxford University Press, Oxford, 2007), pp. 37–61.
9. S. Voitchovsky, in *The Oxford Handbook of Economic Inequality*, W. Salverda, B. Nolan, T. Smeeding, Eds. (Oxford Univ. Press, Oxford, 2009), pp. 549–574.
10. M. Ravallion, in *Handbook of Income Distribution Volume 2*, A. B. Atkinson, F. Bourguignon, Eds. (North Holland, Amsterdam, 2014); working paper available at www.nber.org/papers/w19210.
11. A. Berg, J. D. Ostry, J. Zettelmeyer, *J. Dev. Econ.* **98**, 149–166 (2012).
12. D. Herzer, S. Vollmer, *J. Econ. Inequal.* **10**, 489–503 (2012).
13. M. Ravallion, *Am. Econ. Rev.* **102**, 504–523 (2012).
14. M. Ravallion, *J. Econ. Inequal.* **3**, 169–181 (2005).
15. M. Ravallion, in *Brookings Trade Forum 2004*, S. Collins, C. Graham, Eds. (Brookings Institution, Washington, DC, 2004), pp. 1–38.
16. Y. Amiel, F. Cowell, *Thinking About Inequality: Personal Judgment and Income Distributions* (Cambridge Univ. Press, Cambridge, 1999).
17. E. Harrison, C. Seidl, *Public Choice* **79**, 61–81 (1994).
18. *Economist* **27**, 94 (2000).

ACKNOWLEDGMENTS

The author is grateful to the journal's reviewers for helpful comments. The author declares no conflict of interest.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/851/suppl/DC1
Fig. S1
Tables S1 to S3

10.1126/science.1251875

REVIEW

The intergenerational transmission of inequality: Maternal disadvantage and health at birth

Anna Aizer[1,2] and Janet Currie[2,3]*

Health at birth is an important predictor of long-term outcomes, including education, income, and disability. Recent evidence suggests that maternal disadvantage leads to worse health at birth through poor health behaviors; exposure to harmful environmental factors; worse access to medical care, including family planning; and worse underlying maternal health. With increasing inequality, those at the bottom of the distribution now face relatively worse economic conditions, but newborn health among the most disadvantaged has actually improved. The most likely explanation is increasing knowledge about determinants of infant health and how to protect it along with public policies that put this knowledge into practice.

Income inequality in many developed countries has been rising steadily since the late 1970s, with the United States having recently earned the distinction of being the most unequal of all developed countries (*1*). In addition to affecting the current generation, rising inequality may have long-term consequences, affecting the distribution of health, human capital, and income of the next generation. This conclusion is based on evidence about the importance of maternal conditions in determining the health and human capital of their offspring, which in turn affects their future economic status. One important way in which maternal circumstances matter is by affecting health at birth, which is an important predictor of long-term outcomes.

We review existing research on the pathways through which conditions associated with maternal economic disadvantage during the prenatal period affect health at birth and children's outcomes. It is difficult to distinguish between the effects of prenatal conditions and those of genetic inheritance or postnatal "investments" in children. However, mounting evidence suggests that maternal impoverishment during the prenatal period does have a substantial causal impact on infant health, which in turn affects long-term outcomes. Most of this research is based not on randomized controlled trials, which are typically infeasible and/or unethical in this context, but rather on studies of "natural experiments" or on sibling comparisons.

These designs, when well executed, can allow researchers to draw more credible causal inferences than can studies that document simple correlations. For example, although we do not believe genes to be an empirically important cause of inequality, it is theoretically possible that unfavorable genetic inheritance could cause both low maternal income and poor infant health. In such an example, improving income would not improve infant health because both would be driven by an omitted third factor (genes). Natural experiments are events that researchers exploit in an attempt to eliminate confounding due to unobserved or unmeasured variables through a design that mimics the random assignment of controlled experiments. Early studies examining reductions in pollution due to plant closings or recessions provide an important example (*2*, *3*). When an industrial plant closes for economic reasons, there may be a sudden reduction in pollution. As long as the composition of mothers living near the plant does not change too rapidly with the closure, a before and after comparison of women living near the plant and women living further away from the plant can be used to assess the effect of the change in pollution on infant health.

Sibling comparisons control for constant characteristics of the parents (such as genetic inheritance) and hence eliminate this potential source of confounding. For example, to estimate the impact of birth weight (the most common measure of newborn health) on the future outcomes of offspring, one cannot simply compare outcomes of low-birth-weight and normal-birth-weight children because these differences may reflect factors such as genetic inheritance or differential patterns of prenatal investment. If the researcher cannot control for these differences, they will bias the estimated effects of birth weight. Hence, to address the problem of confounding, researchers have used twinning as a natural experiment (*4*–*7*). Twins typically differ in their birth weight because of location in the uterus and/or differences in the placentas (a "natural" source of random variation), but their genetic inheritance and postnatal environments are very similar. As a result, comparing the long-term outcomes of twins born with different birth weights allows one to estimate the causal impact of birth weight on outcomes controlling for these two important confounders. Such studies, which are typically based on large national samples, have shown that children of lower birth weight have substantially worse adult outcomes than those of their own twins or closely spaced siblings in terms of schooling attainment, test scores, employment, use of disability programs, wages, and adult health (*4*–*7*).

This Review describes research, largely based on sibling comparisons or natural experiments, on the ways in which maternal conditions during the prenatal period affect newborn health and later offspring outcomes. Specifically, the focus is on the following four domains of maternal disadvantage: (i) poor health behaviors during the prenatal period (more smoking, overweight, and neglect of prenatal care); (ii) greater exposure to harmful environmental factors, including the direct effect of toxic pollutants, violence, and stress, combined with a lower likelihood of taking action to avoid potential harms; (iii) poorer access to medical care, including contraception, leading to a greater likelihood of unplanned pregnancy; and (iv) worse underlying health, including poorer nutrition. We will show that all of these factors are associated with worse infant health and with poorer future outcomes, although there are policies that have been shown to be effective in breaking the cycle.

Prenatal Conditions and Offspring Outcomes

The Relationship Between Maternal Disadvantage and Health at Birth

Figure 1 illustrates the huge inequality in health at birth as proxied by the incidence of low birth weight, that exists even in a rich country such as the United States (low birth weight, a common measure of health at birth, is defined as birth weight less than 2500 g) (*8*). Because U.S. natality data do not include maternal income, we rely on other measured characteristics that are strongly related to income—namely, maternal race, marital status, and education. African American, single, and less educated mothers are all (independently) more likely to be economically disadvantaged. Moreover, all three groups are more likely to have a low-birth-weight baby, as shown in Fig. 1. In the absence of information on income, we use these three characteristics to proxy for maternal advantage, defining the most economically advantaged as non-Hispanic white, married, college-educated mothers and the most economically disadvantaged as African American, unmarried, and less than a high school education. Using census data from the American Community Survey, we calculated that the average household income for women 18 to 40 years old with at least one child less than 5 years old in 2010 to 2012 was $120,655 for the "advantaged" group, whereas the corresponding figure for the

[1]Department of Economics, Brown University, Providence, RI 02912, USA. [2]National Bureau of Economic Research, Cambridge, MA 02138, USA. [3]Department of Economics, Princeton University, Princeton, NJ 08544, USA.
*Corresponding author. E-mail: jcurrie@princeton.edu

"disadvantaged" group was $16,497, suggesting that our proxies for economic advantage are reasonable. The last pair of bars in Fig. 1 show that the incidence of low birth weight among the most disadvantaged is more than three times that of the most advantaged mothers (*9*).

The difference between the most and least advantaged mothers has declined over the past 20 years (Fig. 2), suggesting that birth weight is indeed malleable and influenced by many factors (*10*). It is remarkable that health at birth continued to improve in the face of growing income inequality over the same period (*11*). How did conditions improve for babies, when they were deteriorating for many of their mothers? Our discussion will shed light on policies that may have helped to counteract the effects of growing income inequality.

Explaining the Maternal Socioeconomic Status and Offspring Outcomes Relationship

The studies surveyed in this section all rely on natural experiments or sibling comparisons to estimate causal effects. They fall into two groups. The first consists of studies of the impact of prenatal conditions on newborn health (usually measured with birth weight). These studies do not show a direct connection between maternal prenatal conditions and later offspring outcomes but instead rely on existing studies that have established a strong relationship between newborn health and long-term outcomes, including intelligence quotient (IQ), education, employment, and income. The second group of studies link prenatal maternal conditions directly with long-term offspring outcomes.

Maternal Behavior

It is perhaps unsurprising that consumption of tobacco, alcohol, and illegal drugs during pregnancy is associated with adverse birth outcomes, or that good nutrition and better access to medical care are associated with positive outcomes. Health behaviors during pregnancy are typically better among more economically advantaged women (*12*). For example, in the United States in 2011 18% of the most disadvantaged women smoked during pregnancy compared with 1% of the most advantaged women (Fig. 3). There are a number of randomized controlled trials of smoking cessation interventions among pregnant women that link reductions in smoking during pregnancy with improved birth outcomes such as prematurity and birth weight. These studies have been reviewed elsewhere (*13*).

Other studies based on natural experiments or sibling comparisons have also found that smoking has a negative impact on newborn health and that this does not simply reflect other adverse circumstances affecting mothers who happen to smoke (*14*–*16*). For example, researchers using data from 1.5 million births in New Jersey between 1989 and 2003 compared the birth weights of sibling pairs in which the mother smoked during one pregnancy but not the other. They estimated that conditional on any smoking, smoking at the mean number of cigarettes smoked per day (10) increases the probability of low birth weight by 1.8 percentage points on a baseline of 8.9% low birth weight in their sample (*15*).

Environmental Factors: Pollution, Violence, and Stress

Another way in which maternal economic status influences newborn health is through exposure to harmful environmental factors such as pollution, violence, and stress. There is growing evidence that even relatively low levels of pollution have negative effects on infant health. Several recent studies consider variation in pollution owing to policy changes as a "natural experiment" that can help to identify the effect of in utero exposure to pollution on birth weight, as well as on later outcomes such as child test scores and adult earnings (*17*–*19*). For example, the introduction of electronic toll-collection devices (E-ZPass) in New Jersey and Pennsylvania greatly reduced auto emissions in the vicinity of toll plazas and reduced the incidence of low birth weight by about one percentage point in the 2 km surrounding the toll plaza (*17*).

Because of lower housing costs, poor women are more likely to live near, and therefore to be exposed to, sources of pollution. A study of all births in five large U.S. states found that African American women and less-educated women (two correlates of economic disadvantage) are more likely to live within 200 m of Superfund hazardous waste sites or factories emitting toxic releases

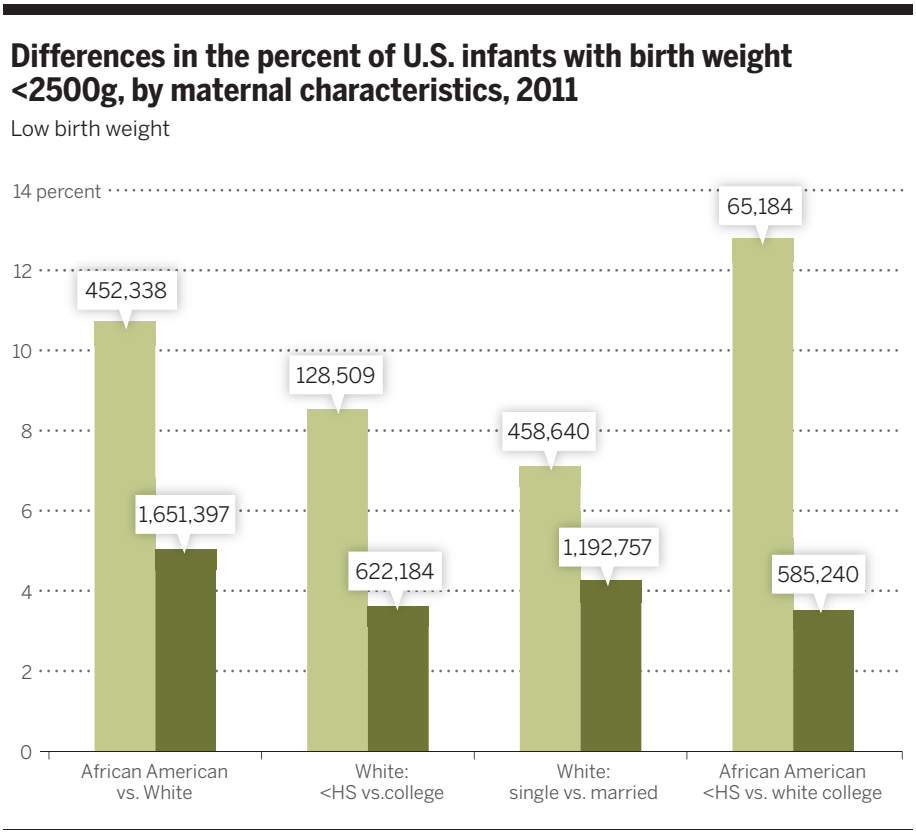


Fig. 1. Differences in the percent of U.S. infants with birth weight <2500 g, by maternal characteristics, 2011. Data was calculated by using singleton births to mothers age 19 to 39 years from the U.S. National Individual-Level Natality Data (birth records). The following states are excluded because education is inconsistently coded over time: Alabama, Alaska, Arizona, Arkansas, Connecticut, Hawaii, Maine, Massachusetts, Minnesota, Mississippi, New Jersey, Rhode Island, Virginia, and West Virginia. We focus on singleton births because multiple births are much more likely to be low birth weight, and many multiple births result from assisted reproductive technology. Sample sizes are printed over each bar. Given the large sample sizes, SEMs are very small (<0.15 percentage points).

and that this pattern holds even within small areas defined by ZIP codes (*20*). Moreover, less-educated and minority women are less likely to move to cleaner areas between births. When Superfund sites are cleaned up, the population living in the area immediately surrounding the site becomes more economically advantaged (again as proxied by race and education). Similarly, a study of all births in New Jersey over a 10-year period found that less-educated women were less likely to move away from water districts with drinking-water quality violations between births (*21*).

Poor women and minority women are also more likely to be the victims of violence, especially intimate partner violence (IPV). More specifically, African American and Hispanic women are twice as likely to be the victims of IPV than are white women, and women with incomes less than $25,000 per year are two and half times as likely to be victimized than are those with incomes greater than $25,000 (*22*).

Changes in local, county-level policing policies have been effective in reducing IPV, and research has used the rollout of these policies to estimate the impact of IPV on birth outcomes (*23*). These policy changes involve reductions in police and prosecutorial discretion as well as increasing the availability of community resources for victims (*24*). Reductions in levels of violence resulting from these policies have been used to show that hospitalization for an assault while pregnant reduces birth weight by 163 g, on average. Thus, although these policies target women, they have substantial positive effects on newborn health and well-being, as well (*23*).

In addition to the direct negative effect of violence on maternal and child health, violence may also indirectly harm pregnant women by increasing their exposure to stress. Although stress is difficult to measure and may have different effects depending on whether it is chronic or episodic, greater exposure to stress among the poor and disadvantaged has been linked with worse physical and mental health. Research in neurobiology has identified abnormal levels and fluctuations in cortisol (often referred to as the "stress hormone") and the dysregulation of the hypothalamic-pituitary-adrenal axis as the primary underlying mechanism (*25*, *26*). More recently, given that cortisol crosses the placenta, researchers have focused on the negative effects of exposure to stress in utero on offspring outcomes. Comparisons of siblings show that those who were exposed to high levels of the stress hormone cortisol in utero have lower IQ levels at age 7 and complete 1 fewer year of schooling (*27*).

An innovative study of all births to mothers with Arabic names in California in the 6 months after the 2001 terrorist attacks in the United States found that the incidence of poor birth outcomes was substantially elevated for this group relative to that for non-Arabic–named mothers. The study argues that Arabic-named mothers suffered particularly strong psychological stress in the aftermath of the attacks, which in turn had a negative impact on their babies (*28*). Children in utero in areas in the path of hurricanes also have worse outcomes than those of unaffected siblings, with the most likely explanation being exposure to maternal stress (*29*, *30*).

Access to Medical Care and Family Planning

Previous work has shown that the Medicaid expansions of the 1980s, which expanded health insurance coverage for pregnant women, reduced the probability of low birth weight (slightly) and infant mortality (more strongly) among low-income women (*31*–*33*). More recent work has focused on Medicaid (and other publicly funded) expansions of coverage for family-planning services. Poor women are more likely to have unplanned pregnancies (*34*) and to space their pregnancies more closely together (*35*). Both factors are thought to reflect inadequate access to family-planning services, and both have been linked to worse birth outcomes (*36*–*38*). Hence, offering poor women access to more comprehensive family-planning technologies may have the potential to improve birth outcomes for poor mothers.

Two recent studies have exploited geographic variation in access to family-planning services in order to estimate their impact on fertility and child outcomes. The first used state-level expansions in Medicaid eligibility for family-planning services to estimate the impact of access to family planning on fertility. The authors found that expanding Medicaid coverage for family planning reduced birth rates among the "near poor" (the poor were already covered) by between 2 and 4% (*39*). The second study used historical data on the timing of access to oral contraception (which differed greatly by state) and on federal family-planning grants to poor areas in order to estimate the impact of these policies on children's outcomes. Both types of policies resulted in declines in fertility (*40*). Although infant health did not improve measurably in areas that received grants to subsidize contraception/family planning, maternal resources available to children did, as did the children's long-term well-being. Children born in areas and years that had access to subsidized family planning saw substantial improvements in educational attainment and earnings in adulthood (*41*, *42*).

Worse Maternal Health, Including Nutrition

Disadvantaged women are in worse underlying health than are their more advantaged counterparts. For example, in 2011 disadvantaged pregnant women were twice as likely to be obese or to have preexisting conditions of diabetes or hypertension as compared with their advantaged counterparts (Fig. 3). Some of this elevated risk may be due to their own histories of disadvantage. This history of disadvantage is what underlies the "weathering hypothesis," first proposed by Geronimus to explain the fact that the African American–white gap in infant mortality increases with maternal age (*43*). Historical U.S. data on mothers born between

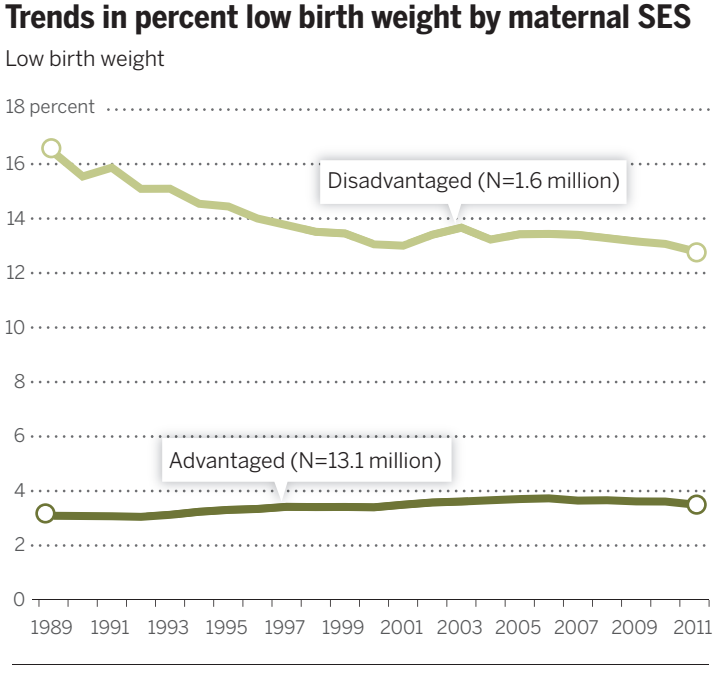


Fig. 2. Trends in percent low birth weight by maternal socioeconomic status (SES). The sample is defined as in Fig. 1. Disadvantaged is defined as African American, less than a high school education, and unmarried. Advantaged is defined as white, college education, and married.

1961 and 1990 shows that mothers who were born in a high disease environment (measured by using high post-neonatal mortality rates) were more likely to have diabetes than were other women at the time they delivered their own infants. Moreover, these women were also more likely to have low-birth-weight babies (*44*, *45*).

In addition to being in worse underlying health, disadvantaged women also have greater exposure to and are more susceptible to contagious diseases such as seasonal influenza. Hence, they may be disproportionately affected by pandemics, which can negatively affect fetal development (*46*). There are a number of factors that can potentially explain disadvantaged women's greater susceptibility, including the fact that disadvantaged women are more likely to live in crowded homes, are more reliant on public transportation, are less likely to be immunized, are less able to stay home from work when ill, and are less likely to believe the influenza vaccine to be effective (*47*). They are also more likely to have asthma—a complicating factor—and less likely to have access to health care.

Exposure to influenza in utero has been linked with both worse newborn health and worse long-term outcomes for offspring. Recent research has documented spikes in prematurity among siblings nearing full term in months with greater prevalence of influenza (*48*). Researchers have also relied on unexpected pandemics to estimate the causal impact of influenza on offspring outcomes (*49*–*51*). For example, the influenza epidemic of 1918 affected different parts of the country at different times. A study using decennial census data from the United States to examine the long-term impacts of in utero exposure to the epidemic shows that affected cohorts had lower earnings and educational attainment and higher probabilities of being disabled decades later (*49*). World War I is a possible confounder because it could have affected the type of women who became pregnant in 1918 (that is, the wives of soldiers may have been less likely to conceive). However, recent research examining the effect of the pandemic in Taiwan (which was not involved in WWI) finds similar effects (*52*). Moreover, the 1957 flu epidemic in Great Britain had negative effects on birth weight, although only for the children of the least healthy mothers: smokers and women of below-average height, which is a marker for poor nutrition in childhood (*51*).

Disadvantaged women are also less likely to gain the recommended weight during pregnancy (30% of disadvantaged versus 16% of more advantaged pregnant women gained less than the recommended amount in 2011) (*53*). A number of randomized controlled trials of nutritional supplementation during pregnancy, mostly in developing countries, have shown that relatively inexpensive nutritional supplements often increase birth weight (*54*). Other studies examine specific micronutrient deficiencies. For example, a study of periodic iodine supplementation programs in Tanzania found that children from cohorts that benefited from supplementation in utero attain a third to a half of a year more schooling than do siblings who did not benefit (*55*).

A number of recent studies have examined the short- and long-term effects of nutritional deprivation during pregnancy that occurs as a result of a fasting during Ramadan among Muslims in both developed and developing countries. Because Ramadan moves through the Gregorian calendar, it is possible to separate the effects of the fast from seasonal effects. Prenatal exposure to Ramadan in early pregnancy is associated with lower birth weight and an increase in the likelihood of disability in adulthood, even in rich countries such as the United States (*56*).

Birth Weight and Parental Investments

Poor health at birth will only have lasting effects if parents are unable or unwilling to offset its impacts through postnatal parental investments. There are two potential mechanisms at work. First, higher-income parents have more money to invest in their children and so may be better able to offset the effects of poor infant health (*57*, *58*). Second, postnatal investments may be more productive for those born healthier; that is, prenatal and postnatal investments may prove to be complementary (*59*).

Existing evidence on the latter pathway is based on studies with long-term follow-up and suggests that newborn health and postnatal investments are indeed complementary (*60*). Perhaps the best example of such a study is the Infant Health and Development Program (IHDP), which randomized low-birth-weight infants to an intensive preschool program and a less intensive control program. The IHDP found significant, positive, and sustained (to age 18 years) effects of the enriched intervention on cognitive test scores, but only in children with birth weights between 2000 and 2500 g (that is, at the higher end of the low-birth-weight distribution). Among children with birth weights below 2000 g, program effects were negligible (*61*). Similarly, a recent study of children from the National Collaborative Perinatal Project found that a preschool intervention in a group of disadvantaged children was effective in raising the IQ of infants with the highest cognitive ability in the group but did not raise the IQ of others, although the intervention may still have had positive effects on noncognitive skills (*62*).

These examples highlight an important caveat to studies of long-term outcomes among low-birth-weight infants, which is that, of necessity, they follow cohorts born many years ago. Medical advances in the treatment of low-birth-weight infants may have increased the efficacy of postnatal interventions among the lightest infants today, relative to what was seen in previous cohorts. At the same time, infants of lower and lower birth weight are surviving, and their long-term prognosis may be worse than that of the lightest infants in earlier cohorts.

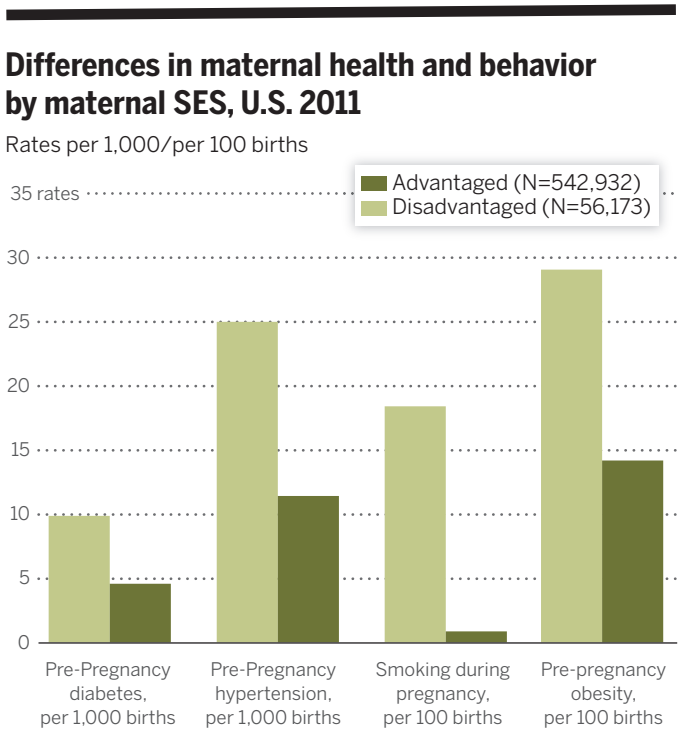


Fig. 3. Differences in maternal health and behavior by maternal SES, United States 2011. The sample is defined as in Fig. 1. Disadvantaged is defined as African American, less than a high school education, and unmarried. Advantaged is defined as white, college education, and married. Given the large sample sizes, SEMs are very small (<0.2 units).

Thus, although considerable uncertainty remains, the available evidence suggests that the initial health disadvantages associated with being born to a poor mother are likely to be exacerbated over time: Children with poorer initial health endowments typically receive fewer postnatal investments, and the investments that they do receive may be less effective. This mechanism can potentially explain the considerable persistence of in utero conditions on later offspring outcomes. It can also explain why the long-term impact of low birth weight is greater when children are born into poverty (*63*).

Given the importance of maternal prenatal conditions in perpetuating inequality across generations, we next consider evidence regarding the effectiveness of public interventions in breaking the cycle of poverty.

Effectiveness of Public Programs in Breaking the Cycle of Poverty

Existing Programs Targeting the Prenatal Period

As discussed above, health insurance, measures to curb domestic violence, and family-planning services have had positive effects on infant health and longer-term outcomes. Food supplementation programs also appear to show consistent benefits, even in rich countries. For example, the Supplemental Feeding Program for Women, Infants, and Children (WIC) was in the 1970s rolled out gradually, which effectively granted access to some mothers before others. For mothers who were high school dropouts and mothers in high-poverty counties, the introduction of WIC in the mother's county of residence reduced the proportion of low-birth-weight births by ~1% (*64*). Conversely, the closing of WIC centers has been associated with reductions in pregnancy weight gain and increases in the incidence of low birth weight among mothers with a high school education or less (*65*).

The Food Stamps program was also rolled out on a county-by-county basis, and its introduction reduced the incidence of low birth weight by 1% among whites and by as much as 1.5% among African Americans (*66*). A follow-up study examining long-term outcomes found that access to food stamps in early childhood led to a significant reduction in the incidence of "metabolic syndrome" (obesity, high blood pressure, and diabetes) as well as an increase in economic self-sufficiency among women (*67*).

Another program that has been shown to be effective in terms of improving short- and long-term outcomes is the Nurse Family Partnership (NFP) program, a modified version of which is being expanded through the Affordable Care Act. NFP programs provide nurse home visits to poor, unmarried, young women who are pregnant for the first time. The visits occur monthly during the pregnancy and during the first 2 years of the child's life. Nurses provide guidance to pregnant women and new mothers regarding healthy behaviors, competent care of children, and personal maternal development. Two randomized controlled trials with long-term follow-up found significant reductions in child abuse and adolescent criminal activity among the children, as well as improvements in the children's academic achievement. Improvements were greatest among children whose mothers had the lowest levels of intelligence and/or poor mental health (*68*, *69*).

Postnatal Interventions

Although our focus has been on the role of maternal prenatal conditions and newborn health, there is evidence that intervening in the postnatal period can also be effective. In particular, early-education programs and income-transfer programs have been shown to improve long-term child outcomes among low-income families (*70*, *71*). Reviews of the evidence suggest that intervening during the preschool years is more effective than post-school intervention, although many evaluations of interventions later in life suffer from short follow-up periods. Specifically, only programs that start before the age of 3 years seem to have long-lasting effects on IQ. (*70*) However, intensive, center-based preschool programs such as the Perry Preschool and Abecederian programs have had large long-term positive effects on children largely because they improved noncognitive or "character" skills rather than through improvements in IQ, which suggests that an exclusive focus on raising IQ may be misguided (*72*).

Means-tested income transfer programs have also been found to improve maternal health and child outcomes. In particular, researchers in the U.S. (*73*–*75*) and Canada (*76*) have exploited changes in cash transfers to low-income working families in order to measure the impact of income transfers on child outcomes. These studies found improvements in maternal mental health as well as children's cognitive test scores, which is an important predictor of long-term economic outcomes. However, women become eligible for these particular programs (among others) only once they have given birth. Given the research on the importance of prenatal conditions, such policies are potentially missing a major window of opportunity.

Conclusions: Implications for Policy and Future Research

Over the past several decades, inequality has been increasing in many developed countries, largely because of economic forces described elsewhere in this issue (*11*, *77*). Other things being equal, one would have expected relative health at birth to deteriorate along with economic conditions for those at the bottom of the distribution. However, other things were not equal: Health at birth bucked this trend and improved steadily for the most disadvantaged in both relative and absolute terms, even during the Great Recession. We now know that the fetus is particularly vulnerable to myriad health insults, but we also know much more about how to protect infants from these shocks.

New knowledge about what matters for fetal health suggests many protective factors, including increasing the prevalence of influenza vaccination for pregnant women, which helps to prevent premature delivery and fetal damage; the expansion of programs such as NFP; supplementing the food and nutrition of pregnant mothers through government programs such as WIC and SNAP (the Supplemental Nutrition Assistance Program, formerly Food Stamps); advances in pollution control that have reduced exposures to harmful toxins; new legal structures that have reduced the incidence of violence against women; the growth of income transfer programs for the near poor and nonpoor (although not the poorest families); and a better understanding of the relationship between contraception, birth spacing, and maternal and fetal health.

Although research on the fetal origins of adult health is exciting and promising, some rather basic questions remain. Looking backward, we can ask, how much of the recent gains in fetal health are due to the various specific factors we have discussed? Which public policies and programs have had the most impact in counteracting the negative effects of growing inequality on infant health, and which are most cost-effective? It would be useful to know, for example, how exposure to diseases such as influenza compares with exposure to common environmental toxicants, but the science is not yet sufficiently advanced to allow these comparisons. Looking forward, we can ask, what are the most promising strategies for continuing to improve fetal health, and how much will this improvement reduce inequality in years to come?

REFERENCES AND NOTES

1. Central Intelligence Agency, *The World Factbook 2013-2014* (Cental Intelligence Agency, Washington, DC, 2013).
2. K. Y. Chay, M. Greenstone, *Q. J. Econ.* **118**, 1121–1167 (2003).
3. J. D. Parker, P. Mendola, T. J. Woodruff, *Epidemiology* **19**, 820–823 (2008).
4. S. E. Black, P. J. Devereux, K. G. Salvanes, *Quarter. J. Econ.* **122**, 409–439 (2007).
5. P. Oreopoulos, M. Stabile, L. Roos, R. Walld, *J. Hum. Resour.* **43**, 88–138 (2008).
6. H. Royer, *Am. Econ. J. Appl. Econ.* **1**, 49–85 (2009).
7. P. Bharadwaj, J. Eberhard, C. Neilson, "Do initial endowments matter only initially? The persistent effect of birth weight on school achievement health at birth, parental investments, and academic outcomes," Working paper, University of California at San Diego, Economics Working Paper Series (2010-09-16).
8. If one looks at all births, the fraction of low birth weight among white college-educated mothers has increased more than suggested with Fig. 2 because these mothers are more likely than others to use assisted reproductive technology.
9. Although much of the literature focuses on low birth weight as a summary of health at birth, disparities are also present if we look at alternative indicators. Prematurity is examined in fig. S1. Infant mortality for African Americans and whites is shown in fig. S2 over the same time period. Information about the mother's education and marital status comes from birth records, so infant mortality by education or marital status groups must be constructed by using linked birth and infant death records, which are missing from some years of the data and are only available up to 2006. Figures S3 and S4 are similar to Figs. 2 and 3, except that they include only mothers 20 to 30 years of age. All of these figures suggest that the trends in low birth weight are not anomalous but are present in other measures of infant health.

10. More detailed overviews of factors that have been shown to influence birth weight and of policies that have been shown to be effective in ameliorating the long-term consequences of low birth weight are available in (*78*).
11. T. Piketty, E. Saez, *Science* **344**, 838–843 (2014).
12. This is not always the case; for example, higher-income women are more likely to drink during pregnancy, and alcohol is known to have potentially negative effects on the fetus, although this has been difficult to establish in population studies. However, a recent Swedish study links increases in in utero alcohol exposure, because of a temporary relaxation of drinking laws, to lower adult educational attainment, cognitive ability, wages, and employment (*79*).
13. C. Chamberlain *et al*., *Cochrane Database Syst. Rev.* **10**, CD001055 (2013).
14. D. S. Lien, W. N. Evans, *J. Hum. Resour.* **40**, 373–392 (2005).
15. J. Currie, M. Neidell, J. F. Schmieder, *J. Health Econ.* **28**, 688–703 (2009).
16. P. Bharadwaj, J. V. Johnsen, K. V. Loken, Smoking Bans, Maternal Smoking and Birth Outcomes, Institute for the Study of Labor (IZA) Discussion Papers 7006 (IZA, Bonn, Germany, 2012).
17. J. Currie, R. Walker, *Am. Econ. J. Appl. Econ.* **3**, 65–90 (2011).
18. N. J. Sanders, *J. Hum. Resour.* **47**, 826–850 (2012).
19. A. Isen, M. Rossin-Slater, W. R. Walker, "Every breath you take—Every dollar you'll make: The long-term consequences of the Clean Air Act of 1970," Working paper 19858, National Bureau of Economic Research (NBER) (2013).
20. J. Currie, *Am. Econ. Rev.* **101**, 1–22 (2011).
21. J. Currie, J. S. Graff-Zivin, K. Meckel, M. J. Neidell, W. Schlenker, "Something in the water: Contaminated drinking water and infant health," Working paper 18876, NBER (2013).
22. J. R. Vest, T. K. Catlin, J. J. Chen, R. C. Brownson, *Am. J. Prev. Med.* **22**, 156–164 (2002).
23. A. Aizer, *J. Hum. Resour.* **46**, 518–538 (2010).
24. Arguably the single most important policy change was Congressional passage of the National Violence Against Women Act of 1994, which provided $1.6 billion in resources for local law enforcement and community groups.
25. B. S. McEwen, P. J. Gianaros, *Ann. N. Y. Acad. Sci.* **1186**, 190–222 (2010).
26. T. Seeman, E. Epel, T. Gruenewald, A. Karlamangla, B. S. McEwen, *Ann. N. Y. Acad. Sci.* **1186**, 223–239 (2010).
27. A. Aizer, L. Stroud, S. Buka, "Maternal stress and child outcomes: Evidence from siblings," Working paper 18422, NBER (2012)
28. D. S. Lauderdale, *Demography* **43**, 185–201 (2006).
29. J. Currie, M. Rossin-Slater, *J. Health Econ.* **32**, 487–503 (2013).
30. Currie and Rossin-Slater (*29*) did not include any direct measures of stress but assumed that mothers in the hurricane's direct path were subject to greater stress than were mothers in similar areas off the path.
31. J. Currie, J. Gruber, *J. Polit. Econ.* **104**, 1263–1296 (1996).
32. W. Lin, *Health Econ.* **18**, 823–841 (2009).
33. They found that increasing Medicaid eligibility by 30 percentage points reduced the probability of a low-birth-weight birth by 1.9% and infant mortality by 8.5%.
34. L. B. Finer, M. R. Zolna, *Contraception* **84**, 478–485 (2011).
35. K. Kost, J. D. Forrest, *Fam. Plann. Perspect.* **27**, 11–17 (1995).
36. A. Conde-Agudelo, A. Rosas-Bermúdez, A. C. Kafury-Goeta, *JAMA* **295**, 1809–1823 (2006).
37. B. P. Zhu, *Int. J. Gynaecol. Obstet.* **89**, S25–S33 (2005).
38. L. J. Smits, G. G. Essed, *Lancet* **358**, 2074–2077 (2001).
39. M. S. Kearney, P. B. Levine, *Rev. Econ. Stat.* **91**, 137–151 (2009).
40. M. J. Bailey, "Fifty years of family planning: New evidence on the long-run effects of increasing access to contraception," Working paper 19493, NBER (2013).
41. These results echo earlier findings that the legalization of abortion in the United States improved the outcomes of surviving cohorts of children (*80*).
42. Evidence that restricting Medicaid funding of abortions increased birth rates is available in (*81*).
43. A. T. Geronimus, *Soc. Sci. Med.* **42**, 589–597 (1996).
44. D. Almond, J. Currie, M. Herrmann, *Labour Econ.* **19**, 475–483 (2011).
45. Similarly, the eradication of malaria in the U.S. South produced gains for the affected cohorts (*82*). Improvements in health in the late 1800s and early 1900s have been associated with improved health at older ages (*83*).
46. S. C. Quinn *et al*., *Am. J. Public Health* **101**, 285–293 (2011).
47. K. G. Wooten, P. M. Wortley, J. A. Singleton, G. L. Euler, *J. Vaccine* **6**, 6927–6934 (2012).
48. J. Currie, H. Schwandt, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 12265–12270 (2013).
49. D. Almond, *J. Polit. Econ.* **114**, 672–712 (2006).
50. R. E. Nelson, *Health Econ.* **19**, 1181–1192 (2010).
51. E. Kelly, *J. Hum. Resour.* **46**, 669–694 (2011).
52. M.-J. Lin, E. Liu, "Does 'in utero' exposure to illness matter? The 1918 influenza epidemic in Taiwan as a natural experiment," Working paper, University of Houston, August 2013.
53. The corresponding figures for infants with gestations of 39 to 41 weeks are 0.243 and 0.138, although excessive weight gain during pregnancy is also associated with poor pregnancy outcomes (*84*). In addition, high weight gain is actually less common in disadvantaged women (*85*).
54. K. Abu-Saad, D. Fraser, *Epidemiol. Rev.* **32**, 5–25 (2010).
55. E. Field, O. Robles, M. Torero, *Am. Econ. J. Appl. Econ.* **1**, 140–169 (2009).
56. D. Almond, B. Mazumder, *Am. Econ. J. Appl. Econ.* **3**, 56–85 (2011).
57. J. Guryan, E. Hurst, M. Kearney, *J. Econ. Perspect.* **22**, 23–46 (2008).
58. W. J. Yeung, M. R. Linver, J. Brooks-Gunn, *Child Dev.* **73**, 1861–1879 (2002).
59. There is a third potential mechanism that has to do with parental preferences. Parents may want to equalize outcomes among their children (in which case, they would invest more in the least healthy); or, they may want to have at least one child who is very successful, in which case they might invest most in the healthiest child. Evidence on this point is somewhat mixed in developed countries (*86*, *87*).
60. F. Cunha, J. Heckman, S. Schennach, *Econometrica* **78**, 883–931 (2010).
61. M. C. McCormick *et al*., *Pediatrics* **117**, 771–780 (2006).
62. A. Aizer, F. Cunha, "The production of human capital: Endowments, investments and fertility," Working paper 18429, NBER (2012).
63. J. Currie, E. Moretti, *J. Labor Econ.* **25**, 231–264 (2007).
64. H. W. Hoynes, M. E. Page, A. H. Stevens, *J. Econ. Perspect.* **20**, 47–68 (2009).
65. M. Rossin-Slater, *J. Public Econ.* **102**, 51–69 (2013).
66. H. Hoynes, D. W. Schanzenbach, D. Almond, *Rev. Econ. Stat.* **93**, 387–403 (2011).
67. H. Hoynes, D. W. Schanzenbach, D. Almond, "Long run impacts of childhood access to the safety net," Working paper 18535, NBER (2012).
68. J. Eckenrode *et al*., *Arch. Pediatr. Adolesc. Med.* **164**, 9–15 (2010).
69. D. L. Olds *et al*., *Pediatrics* **120**, e832–e845 (2007).
70. J. Heckman, R. Pinto, P. Savelyev, *Am. Econ. Rev.* **103**, 2052–2086 (2013).
71. F. Campbell *et al*., *Science* **343**, 1478–1485 (2014).
72. J. Heckman, T. Kautz, "Fostering and Measuring Skills: Interventions that Improve Character and Cognition," Working paper 19656, NBER (2013).
73. G. Dahl, L. Lochner, *Am. Econ. Rev.* **102**, 1927–1956 (2012).
74. W. Evans, C. Garthwaite, "Giving Mom a break: The effect of higher EITC payments on maternal health," Working paper 18206, NBER (2011).
75. H. Hoynes, D. Miller, D. Simon "Income, the Earned Income Tax Credit, and infant health," Working paper 18206 NBER (2012).
76. K. Milligan, M. Stabile, *Am. Econ. J. Econ. Pol.* **3**, 175–205 (2011).
77. D. H. Autor, *Science* **344**, 843–851 (2014).
78. D. Almond, J. Currie, in *Handbook of Labor Economics*, D. Card, O. Ashenfelter, Eds. (Elsevier, Amsterdam, 2011), pp. 1315–1486.
79. J. P. Nilsson, "Alcohol Availability, Prenatal Conditions, and Long-term Economic Outcomes," Working paper, Institute for International Economic Studies, 2014; available at https://sites.google.com/site/nilssonjanpeter.
80. E. Ananat, J. Gruber, P. Levine, D. Staiger, *Rev. Econ. Stat.* **91**, 124–136 (2009).
81. M. P. Bitler, M. Zavodny, *Soc. Sci. Med.* **5**, 918–924 (2010).
82. H. Bleakley, *Am. Econ. J* **2**, 1–45 (2010).
83. J. Ferrie, K. Rolf, *Explor. Econ. Hist.* **48**, 445–460 (2011).
84. D. S. Ludwig, J. Currie, *Lancet* **376**, 984–990 (2010).
85. D. S. Ludwig, H. Rouse, J. Currie, *PLoS Med.* **10**, e1001521 (2013).
86. E. Del Bono, J. Ermisch, M. Francesconi, *J. Labor Econ.* **30**, 657–706 (2012).
87. A. Hsin, *Demography* **49**, 1385–1405 (2012).

ACKNOWLEDGMENTS

The authors thank H. Schwandt for assistance in preparing the figures. J.C. received financial support from the John D. and Catherine T. MacArthur Foundation and for support from U.S. Environmental Protection Agency (EPA) grant EPA G2009-STAR-B1. The authors have no conflicts of interest to report.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/856/suppl/DC1
Figs. S1 to S4

10.1126/science.1251872

REVIEW

On the psychology of poverty

Johannes Haushofer[1,2,3,4]* and Ernst Fehr[3]*

Poverty remains one of the most pressing problems facing the world; the mechanisms through which poverty arises and perpetuates itself, however, are not well understood. Here, we examine the evidence for the hypothesis that poverty may have particular psychological consequences that can lead to economic behaviors that make it difficult to escape poverty. The evidence indicates that poverty causes stress and negative affective states which in turn may lead to short-sighted and risk-averse decision-making, possibly by limiting attention and favoring habitual behaviors at the expense of goal-directed ones. Together, these relationships may constitute a feedback loop that contributes to the perpetuation of poverty. We conclude by pointing toward specific gaps in our knowledge and outlining poverty alleviation programs that this mechanism suggests.

More than 1.5 billion people in the world live on less than $1 a day (purchasing power parity in December 2013 dollars) (*1*). This lack of financial means has far-reaching consequences: In Africa, the average person dies 21 years earlier than in Europe, one-third of the population is illiterate (*1*), and one in three children is stunted in growth (*2*). Economic poverty means living in squalor, dying early, and raising children who face similar prospects.

But does poverty affect people's affective states and their economic choice patterns, i.e., the way they feel and act? Here, we discuss recent findings that suggest that poverty causes negative affect and stress—defined as an organism's reaction to environmental demands exceeding its regulatory capacity—and that this effect may change people's behaviorally revealed preferences. Poverty may, in particular, lower the willingness to take risks and to forgo current income in favor of higher future incomes. This may manifest itself in a low willingness to adopt new technologies and in low investments in long-term outcomes such as education and health, all of which may decrease future incomes. Thus, poverty may favor behaviors that make it more difficult to escape poverty.

Two caveats are in order at the outset. First, poverty is characterized not only by insufficient income but also by dysfunctional institutions, exposure to violence and crime, poor access to health care, and a host of other obstacles and inconveniences. This diversity complicates a single and simple account of the relationship between poverty and psychology. However, a first, useful step can be made by focusing on material poverty as a central feature and powerful predictor of the ancillary features of poverty described above. Second, in asking whether poverty reinforces itself through psychological channels, we are not suggesting that the poor bear blame for their poverty. Rather, an environment of poverty into which one happens to have been born can trigger processes that reinforce poverty. On this view, any one of us might be poor if it were not for certain environmental coincidences.

[1]Abdul Latif Jameel Poverty Action Lab, Massachusetts Institute of Technology, 30 Wadsworth Street, Cambridge, MA 02142, USA. [2]Program in Economics, History, and Politics, Harvard University, Cambridge, MA 02138, USA. [3]Department of Economics, University of Zürich, Blümlisalpstrasse 10, Zürich 8006, Switzerland. [4]Department of Psychology and Woodrow Wilson School of Public and International Affairs, Princeton University, Princeton, NJ 08544, USA.
*Corresponding author. E-mail: joha@mit.edu (J.H.); ernst.fehr@econ.uzh.ch (E.F.)

The Effect of Poverty on Risk-Taking and Time-Discounting

People living in poverty, especially in developing countries, have repeatedly been found to be more risk averse and more likely to discount future payoffs than wealthier individuals. For example, discount rates of poor U.S. households are substantially higher than those of rich households (*3*); likewise, studies of Ethiopian farm households (*4*) and a South Indian sample (*5*) find that lower wealth predicts substantially higher (behaviorally measured) discount rates. Wealthier households or those with higher annual incomes also display lower levels of risk aversion in representative samples (*6*, *7*).

In addition to these correlations between wealth/income and preference measures, there is also evidence suggesting that poverty has a causal effect on risk-taking and time-discounting. In (*7*), the potential reverse causality problem—that low risk aversion may on average lead to higher incomes or wealth—is tackled by using windfall gains as an instrumental variable (IV). The IV estimates show a substantial negative effect of income/wealth on risk aversion. The assumption needed for this approach to work is that windfall gains are positively correlated with household income/wealth—which they are—and that they only affect risk aversion through the income/wealth channel—which is plausible. In another study (*8*), experimentally measured discount rates of Vietnamese respondents were negatively related to income; that is, poorer households were more likely to choose smaller and earlier monetary rewards over larger, delayed ones. Here, the potential reverse causality problem—that high incomes may cause low discount rates—was solved by using rainfall as an instrumental variable for income. Rainfall is significantly correlated with income, and on the assumption that it affects the discounting of future payoffs only through income it is a valid instrument. The IV estimates confirm the negative relationship between the discount rate and income, suggesting that poverty may causally affect time-discounting. In addition, the results show marginally more risk aversion in poorer participants.

Negative income shocks are a pervasive feature of the lives of the poor, and they are particularly vulnerable to these shocks because of limited access to credit markets (*9*, *10*). It is therefore interesting to study the effect of negative income shocks on economic choice. In (*11*), subjects were randomly assigned to income shocks in a laboratory experiment after they had first earned some income in an effort task. The authors compared the discounting of future payoffs of subjects who experienced a negative shock with those of a control group that had not experienced an income shock; importantly, a suitable choice of initial endowments ensured that the two groups had the same absolute income when they performed the discounting task. In addition, the potential reverse causality between income levels and time-discounting could be perfectly controlled in the laboratory setting through exogenous manipulation of income levels. Controlling for absolute income, subjects who received a negative income shock exhibited more present-biased economic behavior than those whom the shock did not affect. No opposite effect was found for positive income shocks. Thus, negative income shocks—a pervasive feature of poverty—appear to increase time-discounting.

In a similar study, subjects were randomly assigned to a smaller ("poor condition") or a larger ("rich condition") budget (*12*) and were then asked to make a series of "purchasing" decisions. Naturally, those with a smaller budget faced more difficult trade-offs because they could afford fewer of the desirable goods. Because decision-making under difficult trade-offs is likely to consume scarce cognitive resources, subjects with a small budget were hypothesized to be impaired in subsequent tasks that require willpower and executive control (*13*). The study indeed found that previous decision-making in the poor condition—but not the rich condition—impaired behavioral control, as measured by the duration of time subjects were able to squeeze a handgrip and their performance in a Stroop task. Thus, poverty appears to affect decision-making by rendering people susceptible to the willpower and self-control depleting effects of decision-making. Because willpower and self-control are hypothesized to be important components of the ability to defer gratification,

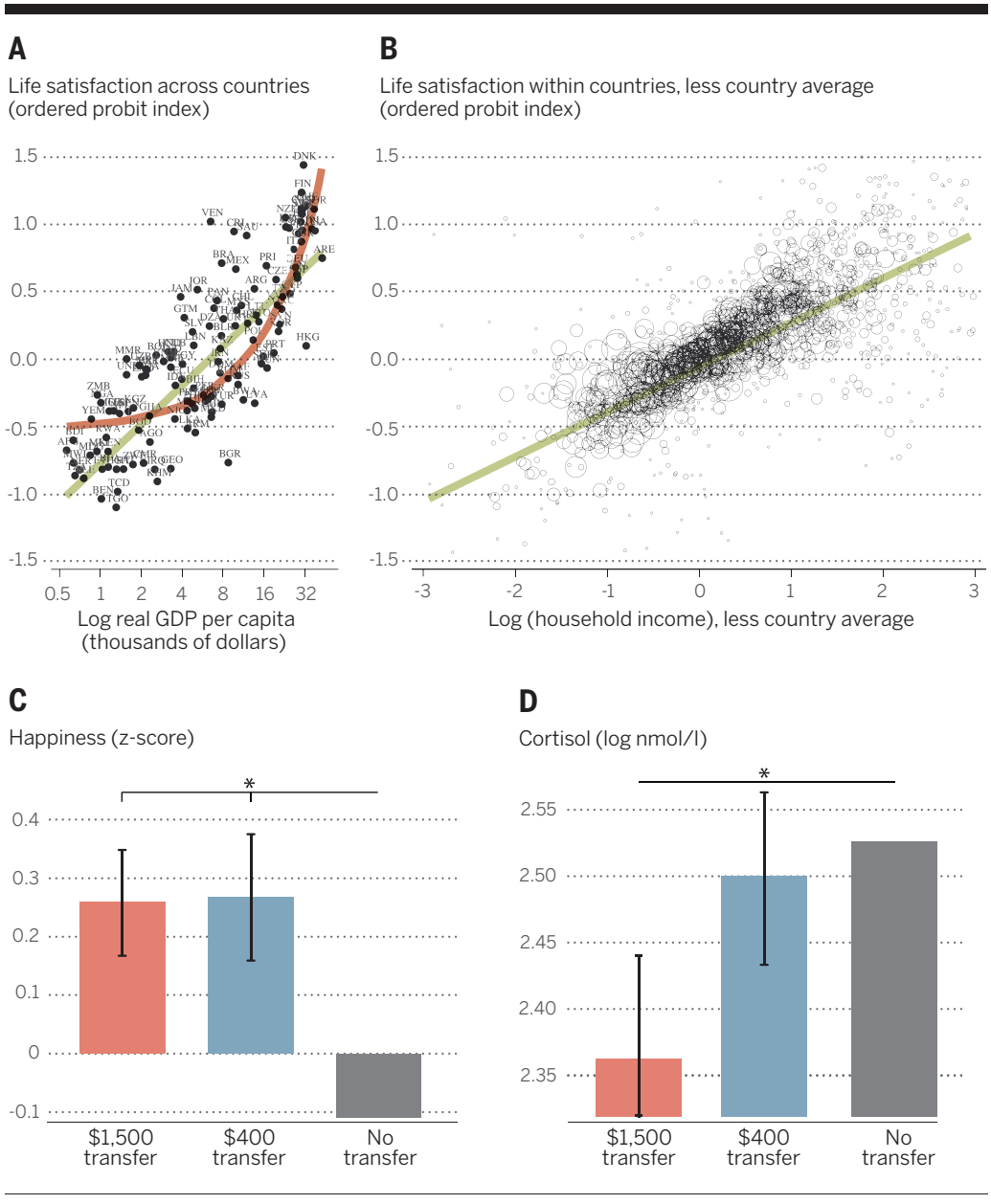


Fig. 1. The relationship between poverty, affect, and stress. The top panels show the relationship between income and life satisfaction, adapted from (*21*), using data from the Gallup World Poll, (**A**) across and (**B**) within countries. We plot standardized responses of 102,583 respondents from 131 countries to the question "Please imagine a ladder with steps numbered from zero at the bottom to ten at the top. Suppose we say that the top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?" In (A), we plot country mean responses against country gross domestic product (GDP) per capita (purchasing power parity in constant 2000 international dollars). The dashed line is fitted from an ordinary least squares (OLS) regression; the dotted line is fitted from a lowess estimation. In (B), each circle represents one income bracket in one country, with its diameter proportional to the population of that income category in that country, and the horizontal axis represents the log of household income after subtracting the country average. (**C**) *Z*-scored happiness responses of $N = 1440$ poor households in Kenya to the happiness question from the World Values Survey ("How happy are you with your life as a whole these days?" on a scale from 1 to 10). Data are from (*32*). Households received unconditional transfers of either $1500 (red) or $400 (blue) or no transfer (gray), and happiness responses were measured about 1 year after the start of the program. (**D**) Levels of the stress hormone cortisol of the same households in Kenya. The error bars in (C) and (D) represent the standard errors of the regression coefficients of the $1500 and the $400 dummy variable in an OLS regression, with happiness or cortisol levels, respectively, as dependent variables. Significant differences ($P < 0.05$) between conditions are marked with an asterisk.

such effects may also affect time-discounting behavior.

Why Does Poverty Affect Risk-Taking and Time-Discounting?

The economic and social conditions under which poor people live may affect discount rates and risk-taking behavior, even though the intrinsic time and risk preferences of the poor may be identical to those of wealthier people. For example, poor people often have no access to formal credit markets (*9*, *10*) and are forced to borrow through informal channels from money lenders, friends or merchants. They often face very high interest rates for credit, and frequently the lenders constrain the amount they lend to them (*9*, *14*), implying that they are much more likely to be liquidity-constrained. Thus, if a poor individual has the choice between a current and a delayed payment in an experiment, he or she may opt for the current payment not because of an intrinsic preference for present payments but because of the credit market imperfections present in informal markets.

In support of this view, a recent study (*17*) measures time preferences of U.S. households shortly before versus shortly after payday. Those surveyed before payday have 22% less cash, and they spend 20% less than those after payday, suggesting that households are liquidity-constrained with regard to money before payday. The study further shows that households surveyed before payday are more present-biased, and this effect is specific to monetary tasks and does not extend to nonmonetary real effort tasks. Because liquidity constraints cannot play a role with regard to effort, this result suggests that liquidity constraints before payday are the source of the apparent present bias for monetary outcomes.

The anticipation of future liquidity constraints may also induce an individual to prefer a safe payment over a risky payment (e.g., in an experiment) (*15*); again, this may occur not because the individual is intrinsically risk averse but because the safe payment helps alleviate liquidity constraints. In addition, poor individuals often face uninsurable, nondiversifiable

"background" risks such as crop failure. They may therefore display less risk-taking behavior with regard to avoidable risks (e.g., in an experiment) even though their risk preferences may not differ from those who are less exposed to background risks (*16*). Indeed, higher background risks have been shown to be associated with higher levels of risk aversion (*7*).

Thus, economic theory and empirical evidence suggest that poor households may display a lower willingness to take risks and to forgo current income for larger future incomes, even though their intrinsic time and risk preferences are not necessarily different from those of richer households. However, we will provide evidence suggesting that this is not the whole story. In a first step, we will show that poverty is associated with negative affect and with stress, and in a second step we will discuss evidence suggesting that negative affect and stress change subjects' risk-taking and time-discounting. In the second part, in particular, we will focus on experiments in which subjects are randomly assigned to treatment conditions and in which the usual economic channels for changes in time and risk-taking behavior—e.g., liquidity constraints or economic background risks—cannot play a role. It is therefore impossible to attribute differences in behavior across treatment to these channels.

The Effect of Poverty on Affect and Stress

Correlations Between Poverty, Affect, and Stress

For several decades, the prevalent view on the relationship between income and psychological well-being was what became known as the Easterlin Paradox (*18*), according to which income, self-reported happiness, and life satisfaction are correlated within but not across countries and are uncorrelated above income levels required to meet basic needs. In addition, higher incomes were thought to be uncorrelated with increased happiness and satisfaction over time. However, larger and newer data sets now suggest that higher incomes are associated with more happiness and life satisfaction both within and across countries, that no saturation point exists (although there are decreasing happiness returns to income), and that as countries grow richer, they also grow happier (*19*–*21*). Fig. 1 shows a correlation between self-reported life satisfaction and income across countries (Fig. 1A) and within countries (Fig. 1B).

In addition to happiness and life satisfaction, poverty is also more broadly related to mental health. According to the 2003 World Health Report, the poorest population quintiles in rich countries exhibit a depression and anxiety disorder prevalence that is 1.5 to 2 times as high as that of the richest quintiles (*22*). A recent comprehensive review of 115 studies (*23*) on the relationship between mental health and poverty in low- and middle-income countries finds a negative association between poverty indicators and good mental health outcomes in 79% of studies. Finally, income and socioeconomic status are also correlated with levels of the stress hormone cortisol. Several studies have shown elevated cortisol levels in persons with lower income and education (*24*, *25*) and lower lifetime economic position as measured by occupational status (*26*, *27*). Similar results have been obtained in infants and children (*27*–*31*).

Together, these findings show that poverty correlates with unhappiness, depression, anxiety, and cortisol levels. But is this relationship causal?

Causal Effect of Poverty on Affect and Stress

The effect of reductions in poverty on affect and stress is usually studied in the context of randomized field experiments or natural experiments such as lottery wins. One such study (*32*) examined the effects of an unconditional cash transfer program in Kenya on psychological well-being. Households were randomly chosen to receive unconditional transfers of either $0, $400, or $1500. Psychological well-being was measured with the happiness and life satisfaction questions from the World Values Survey, and stress and depression were measured using the Center for Epidemiologic Studies Depression Scale, Cohen's Perceived Stress Scale, and levels of the stress hormone cortisol in saliva. The study finds substantial improvements in all of these variables when households receive positive transfers (Fig. 1C), but the stress hormone cortisol was only reduced in those who received large transfers (Fig. 1D). Similarly, several other studies (*33*–*37*) report results from randomized controlled trials that show that cash transfers reduce distress and depression scores (*38*).

Similarly, using natural experiments such as the introduction of guaranteed incomes, lottery payouts, access to a pension scheme, and payouts to Native Americans from a casino opening, several studies find that the resulting increases in income lead to a reduction in hospitalization for mental health problems (*39*), lower consumption of anxiolytics (*40*), and increases in self-reported mental health (*41*–*44*). Less direct alleviations of poverty have also shown effects; several randomized controlled trials report increases in psychological well-being when participants receive health insurance (*45*), improved housing (*46*), and access to water (*47*).

Conversely, the effect of increases in poverty on well-being is usually studied using unexpected shocks such as spells of bad weather for farmers. One such study examined whether random negative income shocks to farmers in Kenya, generated by periods of drought, lead to increases in cortisol levels (*48*). The study finds that farmers have higher levels of cortisol and self-reported stress during drought periods when crops are likely to fail. This relationship does not hold for nonfarmers and is more pronounced among farmers who depend solely on agriculture for their income than among those who also have other sources of earnings. In addition, it is robust to controlling for physical activity, suggesting that changes in labor supply are not the driving factor; the plausibility of this alternative account is further reduced by the fact that the increase in cortisol levels is mirrored by an increase in self-reported stress. Another study (*49*) measured cortisol levels in a sample of 354 Swedish blue-collar workers before and after a subset of these workers lost their jobs. Cortisol levels were significantly higher in those workers who lost their jobs. Importantly, the layoffs were due to a plant closure, arguing against the possibility that job loss might be a consequence rather than a cause of high cortisol levels in individual workers. However, the fact that only one plant was studied and attrition among participants over the course of the study was non-negligible weakens the finding. A further study (*50*) uses declining industries as an exogenous source of variation for job loss and finds an effect of job loss on family mental health using this approach.

These findings thus suggest causal links between poverty, psychological well-being, and stress levels. Altogether, we identified 25 studies that report the effect on psychological well-being of an increase or decrease in poverty, induced either in randomized controlled trials or natural experiments [see the supplementary material (*51*)]. Of these, 18 studies show a positive effect of poverty alleviation on psychological well-being or stress, 5 studies show effects on some psychological variables related to well-being or stress (e.g., certain mental disorders), but not others, and 2 show no results. The mixed or inconsistent findings in these studies may reflect deficiencies or noise of some of the measures used, heterogeneity in the interventions tested, or heterogeneity in the effect of changes in poverty on particular psychological constructs; future studies need to assess these different explanations.

Thus, the large majority of the findings suggests that increases in poverty often lead to negative affect and stress, and decreases in poverty have the opposite effect. We now ask whether negative affect and stress influence risk-taking and time-discounting and could therefore be among the channels through which poverty affects economic behavior.

The Effect of Negative Affect and Stress on Risk-Taking and Time-Discounting

The existence of severe credit constraints and uninsurable background risks implies that the poor are particularly vulnerable to income and health shocks; that is, they are less able to exert control over their life circumstances. As discussed above, this leads to stress and negative affective states such as unhappiness and anxiety, and it raises the question whether such states exert an independent effect on decision-making.

Effects on Risk-Taking

In a recent paper (*52*), subjects were randomly assigned to the threat of receiving unpredictable, randomly administered high or low electrical shocks to their hands during a risk-taking task. The administration of unpredictable shocks is a reliable method for inducing a state of fear and stress (*53*). Subjects in the high-threat condition showed significantly higher risk aversion than those in

the low-threat condition (Fig. 2A). In another study (*54*), subjects' fear was exogenously induced by making them watch a horror video that shows a young man being inhumanly tortured; this fear induction also led to significantly higher risk aversion compared with subjects who saw a control video. Fear induction also led to more risk-averse choices in several other studies (*55*, *56*), and it has also been shown that risk-averse choice can be reduced through cognitive reappraisals that undo the fear effect of a fear-inducing video (*57*). Thus, it is possible not only to increase risk aversion through fear induction but also to reduce risk aversion by reducing fear.

Although the majority of the studies show an unambiguous positive effect of fear and anxiety on risk aversion (*51*), we found one study that does not show such an effect (*58*). However, this study fails to document the specificity of the fear induction and confronts subjects with 100 different choice problems after the fear induction. If induced emotions are not continuously sustained through an appropriate induction procedure—for example, through the threat of aversive shocks—their emotional effect is likely to be short-lived. It may thus be the case that the fear induction was no longer effective for a sizeable part of the 100 choice problems.

Increased risk aversion can also be induced by administering hydrocortisone, which raises cortisol levels in the brain and thus mimics some of the neurobiological effects of stress. In a placebo-controlled experiment (*59*), half of the volunteers received hydrocortisone over a period of 8 days, enabling the study of the acute (on day 1) and the chronic effects (on subsequent days) of the substance. Interestingly, the acute effects of hydrocortisone did not cause changes in risk-taking, whereas the chronic administration led to strong increases in risk aversion: Subjects in the placebo and the acute cortisone condition chose the risky alternative in a risk-taking task in roughly 50% of the cases, but subjects in the chronic hydrocortisone condition chose it only in slightly more than 20% of the cases (Fig. 2B). Other studies (*60*–*63*) have used well-known behavioral stress inductions—the cold pressor task or the Trier Social Stress Test (TSST)—to show that stress typically induces more risk aversion, although this holds only for the domain of gains and not for losses in (*61*) and only for women in (*63*). However, the stress induction did not work for men in the latter study because their cortisol levels in the stress and the control conditions were identical. Thus, taken together, both the evidence from experiments on fear and on stress induction indicates that fear and stress cause higher levels of risk aversion.

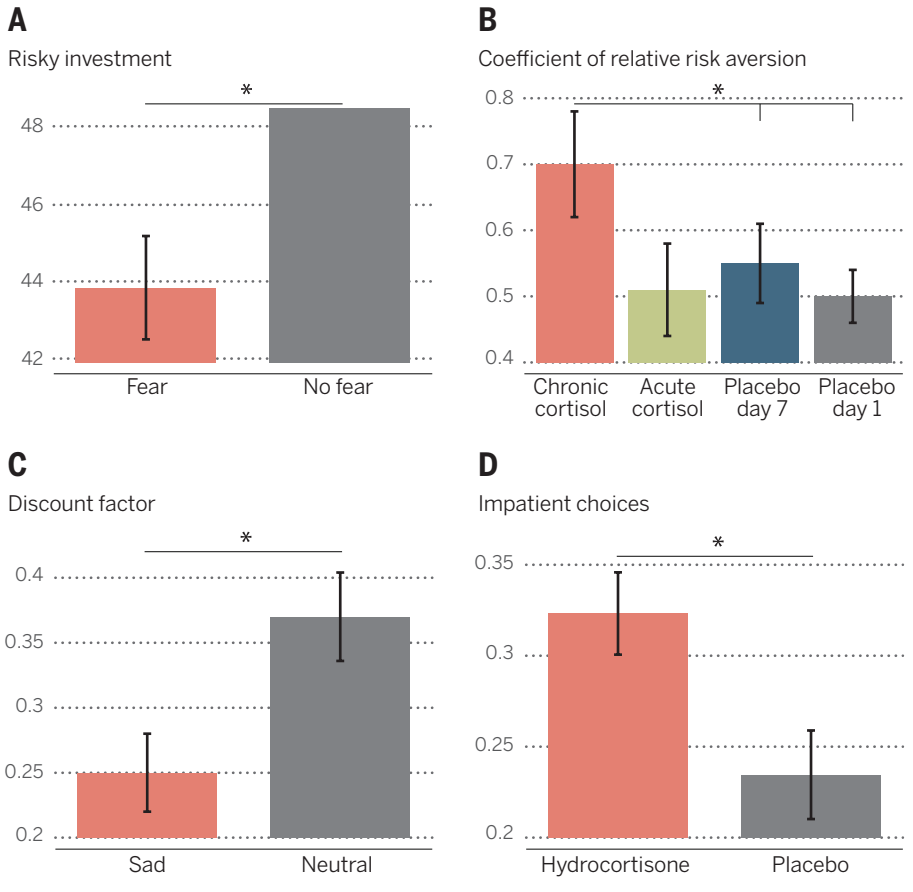


Fig. 2. Effect of negative affect and stress on risk and time preferences. (**A**) Amount invested in the risky asset (out of a total of CHF 24) when subjects (N = 41) faced the threat of receiving a painful electrical shock (fear condition, red bar) and when they received only a mild shock that was not painful (no fear condition, gray bar). Data are taken from (*52*). Subjects who faced the threat of a painful shock were less likely to make risky investments ($P < 0.05$). The error bar indicates the standard error of the regression coefficient for the fear dummy in an OLS regression with risky investment as the dependent variable. (**B**) Coefficient of relative risk aversion (mean ± SEM) of N = 36 subjects that were exposed to either repeated pharmacological elevation of cortisol levels through administration of hydrocortisone over 1 week (red), acute administration (1 day, green), or placebo (day 7, blue; day 1, gray). Data are taken from (*59*). Chronic administration led to an increase in the coefficient of relative risk aversion (CRRA) relative to placebo on both day 1 ($P < 0.05$) and day 7 ($P < 0.05$). (**C**) Discount factors (mean ± SEM) of N = 189 subjects who were exposed to either a sad or a neutral prime. Data are from (*64*). Subjects in the sad condition exhibited lower discount factors ($P < 0.05$), implying greater discounting of the future (because a low discount factor indicates a low valuation of future payoffs relative to present payoffs). (**D**) Share of impatient choices (mean ± SEM) of N = 53 subjects who received either hydrocortisone or placebo. Data are from (*66*). Subjects in the hydrocortisone condition were more impatient ($P < 0.05$) in a discounting task; i.e., they showed greater discounting of future payoffs. Significant differences ($P < 0.05$) between conditions are marked with an asterisk.

Effects on Time-Discounting

A number of recent studies show that negative affect and stress lead to increases in time-discounting (*51*, *64*–*66*). One study (*64*) induced sadness by showing participants film clips that were independently verified to induce the desired emotional state. They subsequently offered subjects choices between smaller amounts of money available immediately or larger amounts available after a delay. This task measures temporal discounting, i.e., the degree to which delayed rewards are devalued. Subjects who had viewed the sadness-inducing film clip were less likely to choose larger, delayed payments than those in the control condition; that is, they discounted future payments more strongly, indicating that sadness reduces patience (Fig. 2C). Conversely, another recent study (*65*) induced positive affect through film clips and found that it increased patience in a similar task.

As in the domain of risk-taking, pharmacological elevation of the stress hormone cortisol through hydrocortisone administration has also been found to increase time-discounting. A recent study administered 10 mg of hydrocortisone or placebo orally to healthy subjects (*66*). After administration, subjects performed a temporal discounting task similar to that described above. Subjects who had been given hydrocortisone

showed an increase in temporal discounting compared with placebo 15 min after administration; that is, they valued the present more highly relative to the future (Fig. 2D). Thus, both negative affect and elevated cortisol levels increase time-discounting, whereas positive affect has the opposite effect (*64*–*67*). Future studies will have to elucidate whether chronic stress in conditions of poverty has similar behavioral effects as acute stress induced under laboratory conditions.

Exactly how might negative affect and stress lead to increased discounting? One possibility lies in the fact that stress has recently been shown to induce a shift from goal-directed to habitual behavior (*68*). If the habitual behavior is to consume immediately, this mechanism would predict that stress should increase temporal discounting by favoring habitual responses. A related possibility is that stress and negative affect may bias attention toward salient cues. If immediate consumption is more salient than delayed consumption, this mechanism would also predict that stress and negative affect should increase time-discounting. In line with this view, Shah *et al.* (*69*) showed that decision-making under scarcity—whether this scarcity is temporal, financial, or of another type—shows signs of the irrationality frequently observed in decision-makers in settings of poverty and that this effect is due to attentional capture by salient cues. More recently, Mani *et al.* (*70*) found that poor individuals (in contrast to the rich) performed worse on tasks measuring intelligence and cognitive control after they had been asked to think about their finances; similarly, farmers performed worse on these tasks before the harvest, when they were relatively poor, than after the harvest. In both cases, material scarcity seems to change people's allocation of attention in ways that are detrimental for their performance. It is possible that similar attentional mechanisms are behind the effect of poverty on risk-taking and time-discounting, in that they induce a focus on immediate and safe payoffs; data on this question are not yet available, however.

Emerging Issues

We have outlined a feedback loop in which poverty reinforces itself through exerting an influence on psychological outcomes, which may then lead to economic behaviors that are potentially disadvantageous. This feedback loop may prolong the climb out of poverty for poor individuals, or even make the escape from poverty impossible if the relationships described above are strong enough.

A number of questions and concerns arise from the previous discussion. First, in our view, the weakest link in the relationship between poverty, psychological outcomes, and economic choice is the effect of stress and negative affect on economic choice. Despite intriguing initial results, it remains incompletely understood exactly which psychological aspects of stress, and which types of negative affect, influence economic behaviors. In addition, the evidence on this link is currently restricted to laboratory studies, and the literature does little to distinguish between the effects of acute and chronic stress on economic choice. Because poverty is usually a chronic condition, future studies should examine the effect of changes in chronic stress on economic choices in the laboratory as well as in field settings. Second, there is still little evidence on the causal effects of different poverty alleviation interventions on life satisfaction and well-being. We do not know whether some interventions work better, per dollar spent, than others. For example, are cash transfers more effective than the provision of health insurance or crop failure insurance? Third, the temporal dimension remains almost entirely unexplored. Little is known about whether poverty alleviation leads to a permanent or only a temporary increase in psychological well-being. To address this problem, repeated surveying after interventions is necessary.

A further open question is whether the relationships outlined above could constitute a poverty trap. For this to be the case, a strong nonlinearity in the relationship between poverty and psychological outcomes, or psychological outcomes and economic choice, would be required (*71*). No evidence is present for the former; existing studies on the relationship between income and psychological outcomes show no strong signs of being nonlinear. In contrast, the famous Yerkes-Dodson law states that stress and performance may exhibit a nonlinear relationship resembling an inverted U (*72*): According to this law, moderate increases in arousal lead to improvements in performance, whereas extreme levels of arousal lead to performance decrements (*73*, *74*). However, little evidence exists on whether this holds for economic behavior; this is a fruitful area for future research.

Finally, what types of welfare programs or interventions would break the relationships discussed above? If the proposed feedback loop holds true, three possibilities seem promising for breaking the cycle and improving welfare: The first is to target poverty directly, the second is to target its psychological consequences, and the third is to target the economic behaviors that result from them. These possibilities are not mutually exclusive, of course, but should be studied in isolation as well as in combination to understand their effect.

With regard to the first possibility—targeting poverty directly—a number of studies have tested the effect of direct poverty alleviation programs on psychological outcomes and economic behavior. Most of these studies examine cash transfer programs, which have produced broadly encouraging results on general welfare in recent years (*32*, *37*, *41*, *75*–*79*). Regarding the third possibility—targeting economic behaviors directly—a number of programs provide small nudges to economic behaviors with large positive welfare consequences—for instance, commitment savings accounts (*80*, *81*), reminders to save (*82*), or the provision of a lockable metal box with a deposit slit at the top (like a piggy bank) (*83*) all led to considerable increases in savings.

In our view, the second possibility, i.e., targeting the psychological consequences of poverty, holds much promise for future work. Although an early randomized controlled trial showed that group interpersonal psychotherapy helped people complete daily economic tasks in Uganda (*84*), research on the economic effects of such interventions is otherwise still in its infancy. Most important, this study targeted depressed individuals, whereas the evidence discussed in this article shows that the debilitating effects of stress and negative affect on economic behavior may occur even in individuals who do not suffer from full-fledged clinical depression. This insight suggests that psychotherapy-like interventions may have economic benefits even in nonclinical populations (*85*).

More broadly, we propose that an increased understanding of the relationship between poverty, its psychological consequences, and their potentially disadvantageous effects on economic choice will lead to poverty alleviation programs that achieve two goals. First, they will take both the psychological costs of poverty and, conversely, the psychological benefits of poverty alleviation into account. Second, they will consider psychological variables as novel intervention targets for poverty alleviation. It is our hope that this will lead to a more refined understanding of poverty and thus contribute to the solution of this lingering global problem.

REFERENCES AND NOTES

1. WHO, *World Health Statistics 2013* (WHO, Geneva, 2013).
2. UNICEF, *The State of the World's Children* (UNICEF, New York, 2012).
3. E. C. Lawrance, *J. Polit. Econ.* **99**, 54–77 (1991).
4. M. Yesuf, R. Bluffstone, Wealth and Time Preference in Rural Ethiopia, Environment for Development Discussion Paper No. EfD DP 08-16 (2008); available at http://www.rff.org/RFF/Documents/EfD-DP-08-16.pdf
5. J. L. Pender, *J. Dev. Econ.* **50**, 257–296 (1996).
6. T. Dohmen *et al.*, *J. Eur. Econ. Assoc.* **9**, 522–550 (2011).
7. L. Guiso, M. Paiella, *J. Eur. Econ. Assoc.* **6**, 1109–1150 (2008).
8. T. Tanaka, C. F. Camerer, Q. Nguyen, *Am. Econ. Rev.* **100**, 557–571 (2010).
9. A. Banerjee, in *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress*, L. Hansen, M. Dewatripont, S. Turnovsky, Eds. (Cambridge Univ. Press, Cambridge, 2003), vol. 3., chap. 1.
10. A. V. Banerjee, E. Duflo, *J. Econ. Perspect.* **22**, 3–28 (2008).
11. J. Haushofer, D. Schunk, E. Fehr, Negative income shocks increase discount rates. University of Zurich Working Paper (2013); available at http://web.mit.edu/joha/www/publications/Haushofer_Schunk_Fehr_2013.pdf
12. D. Spears, *B.E. J. Econ. Anal. Poli.* **11**, 10.2202/1935-1682.2973 (2011).
13. M. Muraven, R. F. Baumeister, *Psychol. Bull.* **126**, 247–259 (2000).
14. A. Banerjee, E. Duflo, Do firms want to borrow more: Testing credit constraints using a targeted lending program. *Rev. Econ. Stud.*; available at http://economics.mit.edu/files/2706
15. C. Gollier, in *Household Portfolios*, L. Guiso, M. Haliassos, T. Jappelli, Eds. (MIT Press, Cambridge, MA, 2001), chap. 1.
16. J. W. Pratt, R. Zeckhauser, *Econometrica* **55**, 143–154 (1987).

17. L. Carvalho, S. Meier, S. W. Wang, Poverty and economic decision making: evidence from changes in financial resources at payday. Center for Economic and Social Research Working Paper (2014); available at http://www.princeton.edu:economics:seminar-schedule-by-prog:behavioral-s14:Poverty-and-Economic-Decision-Making.pdf
18. R. Easterlin, Does economic growth improve the human lot? Some empirical evidence, in *Nations and Households in Economic Growth: Essays in Honour of Moses Abramowitz*, P. A. David, M. W. Reder, Eds. (Academic Press, New York and London, 1974), pp. 89–125.
19. D. W. Sacks, B. Stevenson, J. Wolfers, in *Developmental Challenges in a Postcrisis World*, The World Bank, Ed. (The World Bank, Washington, DC, 2012), pp. 283–315.
20. D. W. Sacks, B. Stevenson, J. Wolfers, *Emotion* **12**, 1181–1187 (2012).
21. B. Stevenson, J. Wolfers, *Brookings Pap. Econ. Act.* **2008**, 1–102 (2008).
22. WHO, *World Health Report 2001: Mental Health: New Understanding, New Hope* (WHO, Geneva, 2001).
23. C. Lund *et al*., *Soc. Sci. Med.* **71**, 517–528 (2010).
24. S. Cohen *et al*., *Psychosom. Med.* **68**, 41–50 (2006).
25. S. Cohen, W. J. Doyle, A. Baum, *Psychosom. Med.* **68**, 414–420 (2006).
26. L. Li, C. Power, S. Kelly, C. Kirschbaum, C. Hertzman, *Psychoneuroendocrinology* **32**, 824–833 (2007).
27. N. S. Saridjan *et al*., *Horm. Behav.* **57**, 247–254 (2010).
28. G. W. Evans, K. English, *Child Dev.* **73**, 1238–1248 (2002).
29. S. J. Lupien, S. King, M. J. Meaney, B. S. McEwen, *Biol. Psychiatry* **48**, 976–980 (2000).
30. S. J. Lupien, S. King, M. J. Meaney, B. S. McEwen, *Dev. Psychopathol.* **13**, 653–676 (2001).
31. E. Chen, S. Cohen, G. E. Miller, *Psychol. Sci.* **21**, 31–37 (2010).
32. J. Haushofer, J. Shapiro, Household response to income changes: Evidence from an unconditional cash transfer program in Kenya. Massachusetts Institute of Technology Working Paper (2013); available at http://web.mit.edu/joha/www/publications/haushofer_shapiro_uct_2013.11.16.pdf
33. S. Baird, J. de Hoop, B. Özler, *J. Hum. Resour.* **48**, 370–403 (2013).
34. F. M. Ssewamala, T. B. Neilands, J. Waldfogel, L. Ismayilova, *Journal of Adolescent Health* **50**, 346–352 (2012).
35. F. M. Ssewamala, C.-K. Han, T. B. Neilands, *Soc. Sci. Med.* **69**, 191–198 (2009).
36. L. C. H. Fernald, R. Hamad, D. Karlan, E. J. Ozer, J. Zinman, *BMC Public Health* **8**, 409 (2008).
37. E. J. Ozer, L. C. Fernald, A. Weber, E. P. Flynn, T. J. VanderWeele, *Int. J. Epidemiol.* **40**, 1565–1576 (2011).
38. One study finds no effect of cash transfers on levels of maternal depression (*77*).
39. E. J. Costello, S. N. Compton, G. Keeler, A. Angold, *JAMA* **290**, 2023–2029 (2003).
40. D. Cesarini, E. Lindqvist, R. Östling, B. Wallace, Estimating the causal impact of wealth on health: Evidence from the Swedish lottery players. *New York University Working Paper* (2013); available at http://webmeets.com/files/papers/res/2014/1050/Health RES.pdf
41. A. Case, Does money protect health status? Evidence from South African pensions, in *Perspectives on the Economics of Aging*, D. A. Wise, Ed. (University of Chicago Press, Chicago, 2004), pp. 287–312.
42. J. Gardner, A. J. Oswald, *J. Health Econ.* **26**, 49–60 (2007).
43. B. H. Apouey, A. Clark, Winning big but feeling no better? The effect of lottery prizes on physical and mental health. *J. Health Econ.*; available at http://ideas.repec.org/p/fem/femwpa/2009.96.html
44. One study finds no effect of lottery winnings on self-reported happiness (*86*).
45. A. Finkelstein *et al*., *Q. J. Econ.* **127**, 1057–1106 (2012).
46. J. Ludwig *et al*., Long-term neighborhood effects on low-income families: Evidence from Moving to Opportunity. NBER Working Paper No. w18772 (2013); available at http://www.nber.org/papers/w18772
47. F. Devoto, E. Duflo, P. Dupas, W. Pariente, V. Pons, Happiness on tap: Piped water adoption in urban Morocco. NBER Working Paper No. w16933 (2011); available at http://www.nber.org/papers/w16933
48. M. Chemin, J. de Laat, J. Haushofer, Poverty and stress: Rainfall shocks increase levels of the stress hormone cortisol. Massachusetts Institute of Technology Working Paper (2013); available at http://web.mit.edu/joha/www/publications/Chemin_deLaat_Haushofer_2013.pdf
49. B. B. Arnetz *et al*., *Psychother. Psychosom.* **55**, 76–80 (1991).
50. S. Mendolia, The impact of job loss on family mental health. School of Economics University of New South Wales Working Paper (2013); available at http://www.iza.org/conference_files/SSch2009/mendolia_s5043.pdf
51. Supplementary materials are available on *Science* Online.
52. A. Cohn, J. Engelmann, E. Fehr, M. A. Maréchal, Evidence for countercyclical risk aversion: An experiment with financial professionals. UBS International Center of Economics in Society Working Paper No. 004 (2013); available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2327557
53. A. Schmitz, C. Grillon, *Nat. Protoc.* **7**, 527–532 (2012).
54. L. Guiso, P. Sapienza, L. Zingales, Time varying risk aversion. NBER Working Paper No. w19284 (2013); available at http://www.nber.org/papers/w19284
55. R. Raghunathan, M. T. Pham, *Organ. Behav. Hum. Decis. Process* **79**, 56–77 (1999).
56. T. Kugler, T. Connolly, L. D. Ordóñez, *J. Behav. Decis. Making* **25**, 123–134 (2012).
57. R. M. Heilman, L. G. Crişan, D. Houser, M. Miclea, A. C. Miu, *Emotion* **10**, 257–265 (2010).
58. A. Conte, M. V. Levati, C. Nardi, The role of emotions on risk preferences: An experimental approach. Jena Economic Research Papers No. 2013-046 (2013); available at http://ideas.repec.org/p/jrp/jrpwrp/2013-046.html
59. N. Kandasamy *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3608–3613 (2014).
60. M. Mather, M. A. Gorlick, N. R. Lighthall, *Psychol. Sci.* **20**, 174–176 (2009).
61. A. J. Porcelli, M. R. Delgado, *Psychol. Sci.* **20**, 278–283 (2009).
62. L. Cingl, J. Cahlikova, Risk preferences under acute stress. IES Working Paper No. 17/2013 (2013); available at http://ideas.repec.org/p/fau/wpaper/wp2013_17.html
63. N. R. Lighthall, M. Mather, M. A. Gorlick, *PLOS ONE* **4**, e6002 (2009).
64. J. S. Lerner, Y. Li, E. U. Weber, *Psychol. Sci.* **24**, 72–79 (2013).
65. J. Ifcher, H. Zarghamee, *Am. Econ. Rev.* **101**, 3109–3129 (2011).
66. S. Cornelisse *et al*., Time-dependent effect of hydrocortisone administration on intertemporal choice. SSRN Working Paper Series (2013); available at http://web.mit.edu/joha/www/publications/Cornelisse_vanAst_Haushofer_Seinstra_Kindt_Joels_2013.pdf
67. In our systematic review, we also identified two studies that found no effect of affect or stress on time preferences. One study (*87*) exposed subjects to an easy or difficult test, thus inducing feelings of relative success or failure, and then measured time-discounting. No effect of test difficulty on time-discounting was found. However, in this study, the time-preference task was administered at the end of a battery of behavioral tests; it is possible that the negative affect induction had already worn off by then. Alternatively, it is possible that the induction of mood through this manipulation is less powerful than movie clips or that subtly different types of affect may differentially affect time preference. Another study used the TSST to induce stress, then measured temporal discounting and found no effect (*88*). A potential explanation for this finding is that the TSST induces acute stress (i.e., concurrent glucocorticoid and noradrenergic activity), whereas hydrocortisone administration lacks some of the components of acute stress (e.g., noradrenergic coactivation). The lack of an effect of the TSST on discounting could thus suggest that acute stress does not affect discounting, whereas chronic stress may. This account is superficially consistent with a recent finding (*89*) showing that the combined administration of hydrocortisone and yohimbine, an α2-adrenoceptor antagonist, has different behavioral consequences than hydrocortisone in isolation.
68. L. Schwabe, O. T. Wolf, *J. Neurosci.* **29**, 7191–7198 (2009).
69. A. K. Shah, S. Mullainathan, E. Shafir, *Science* **338**, 682–685 (2012).
70. A. Mani, S. Mullainathan, E. Shafir, J. Zhao, *Science* **341**, 976–980 (2013).
71. S. Bowles, S. N. Durlauf, K. Hoff, *Poverty Traps* (Princeton University Press, 2006).
72. R. M. Yerkes, J. D. Dodson, *J. Comp. Neurol. Psychol.* **18**, 459–482 (1908).
73. A. F. Arnsten, *Trends Cogn. Sci.* **2**, 436–447 (1998).
74. P. L. Broadhurst, *Acta Psychol.* **15**, 603–604 (1959).
75. C. Blattman, N. Fiala, S. Martinez, Credit constraints, occupational choice, and the process of development: long run evidence from cash transfers in Uganda. *SSRN Scholarly Paper* No. ID 2268552 (2013); available at https://www.poverty-action.org/sites/default/files/ipa_0189_uganda_youth_opportunities.pdf
76. S. de Mel, D. McKenzie, C. Woodruff, *Q. J. Econ.* **123**, 1329–1372 (2008).
77. C. Paxson, N. Schady, *Econ. Dev. Cult. Change* **59**, 187–229 (2010).
78. F. M. Tseng, D. Petrie, Handling the endogeneity of income to health using a field experiment in Taiwan. *Dundee Discussion Papers in Economics* 263 (2012); available at http://ideas.repec.org/p/dun/dpaper/263.html
79. L. C. H. Fernald, P. J. Gertler, L. M. Neufeld, *Lancet* **371**, 828–837 (2008).
80. N. Ashraf, D. Karlan, W. Yin, *Q. J. Econ.* **121**, 635–672 (2006).
81. R. H. Thaler, S. Benartzi, *J. Polit. Econ.* **112**, (S1), S164–S187 (2004).
82. D. Karlan, M. McConnell, S. Mullainathan, J. Zinman, Getting to the top of mind: how reminders increase saving. *NBER Working Paper* No. w16205 (2010); available at http://www.nber.org/papers/w16205
83. P. Dupas, J. Robinson, *Am. Econ. Rev.* **103**, 1138–1171 (2013).
84. P. Bolton *et al*., *JAMA* **289**, 3117–3124 (2003).
85. M. E. Seligman, T. A. Steen, N. Park, C. Peterson, *Am. Psychol.* **60**, 410–421 (2005).
86. P. J. Kuhn, P. Kooreman, A. R. Soetevent, A. Kapteyn, *Am. Econ. Rev.* **101**, 2226–2247 (2011).
87. A. C. Drichoutis, R. M. Nayga Jr., *J. Socio-Economics* **45**, 18–27 (2013).
88. J. Haushofer *et al*., *PLOS ONE* **8**, e78597 (2013).
89. L. Schwabe, M. Tegenthoff, O. Höffken, O. T. Wolf, *J. Neurosci.* **32**, 10146–10155 (2012).

ACKNOWLEDGMENTS

We thank M. Alsan, V. Baranov, L. I. O. Berge, S. Bowles, A. D. Almenberg, C. Fry, H. Gintis, J. Elster, U. Karmarkar, I. Krajbich, J. Lerner, S. Mullainathan, D. Prelec, E. Rothschild, B. Tungodden, and T. Williams for comments. We gratefully acknowledge support from the Cogito Foundation, NIH grant R01AG039297, the Abdul Latif Jameel Poverty Action Lab, and the Harvard Program in Economics, History, and Politics.

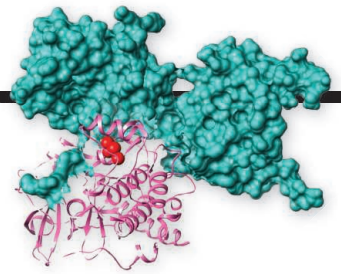
SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/862/suppl/DC1
Supplementary Text
Literature Reviews
Fig. S1
Table S1
References (*90*–*94*)

10.1126/science.1232491

RESEARCH

A potential therapeutic target for adrenal Cushing's syndrome
p. 917



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

An adult kiwi and an elephant bird egg.

EVOLUTION

Ruffling ancient ratite feathers

Biologists have often pointed to the breakup of the supercontinent Gondwana to explain how related species ended up on far-flung continents, but as new research shows, that explanation doesn't fly with ratite birds. Ratite birds are a lineage of large, mostly flightless birds including the African ostrich, the Australian emu, the South American rhea, the diminutive New Zealand kiwi, and the extinct Madagascar elephant bird. Mitchell *et al*. examined the phylogeny of these birds, adding ancient mitochondrial DNA sequences from the extinct elephant bird. It seems that ratites originated from flighted ancestors who evolved large sizes and loss of flight only after flying to their new homes. — LMZ

Science, this issue p. 898.

SUBSURFACE MICROBES

Mapping sub–sea-floor communities

The sea floor is teeming with microbes, whose sheer numbers produce a major effect on the global biogeochemical cycles of carbon, sulfur, and other important nutrients. Bowles *et al*. constructed a map showing how deeply sulfates penetrate marine sediments worldwide and how quickly that sulfate is chemically reduced by microbes in the sub–sea-floor. Globally, almost a third of the organic carbon that reaches the sea floor is consumed during sulfate reduction, and the vast majority of microbial cells in the sub–sea-floor at continental margins get their energy through the biochemical processes of fermentation and methanogenesis. — NW

Science, this issue p. 889.

DEEP EARTH

Delving deeper into the lower mantle

Earth's lower mantle is an enigmatic region, a transition zone between slowly churning solids and a liquid outer core. Large seismic structures and discontinuities in this region are probably due to sharp gradients in temperature, composition, or mineralogy. Teasing apart the precise effects of these factors requires experiments at lower mantle temperatures and pressures (see the Perspective by Williams). Zhang *et al*. found that the major mineral phase of the lower mantle decomposes into two minerals. Andrault *et al*. show how the melting of subducted basalt from the oceanic crust will form pile-like structures on top of the core/mantle boundary. — NW

Science, this issue p. 877, p. 892; see also p. 800.

COSMOLOGY

Confirming cosmic dual conjecture

Quantum mechanics and gravity can seem to contradict each other. Superstring theory may provide a route to reconcile the two, thanks to the gauge/gravity duality conjecture, which allows the system to be described mathematically. However, this conjecture has yet to be formally confirmed. Hanada *et al.* (see the Perspective by Maldacena) performed a simulation of the dual gauge theory in the parameter regime that corresponds to a quantum black hole. Their results agree with a prediction for an evaporating black hole, including quantum gravity corrections, confirming that the dual gauge theory indeed provides a complete description of the quantum nature of the evaporating black hole. — ISO

Science, this issue p. 882; see also p. 806.

CORALS AND CLIMATE

Hot and bothered corals can cope

How well can corals adapt to temperature extremes? Better than anticipated, it turns out. Corals from reef pools with wide temperature fluctuations resist stress better than corals from less extreme pools. Nevertheless, corals transplanted into the hotter and more variable conditions soon acquired thermal tolerance. Palumbi *et al.* (see the Perspective by Eakin) found that the tougher specimens produced more of certain proteins, such as the tumor necrosis factor receptor superfamily, which protected them from the effects of heat. Ramping up heat shock and transport proteins yielded heat tolerance far more rapidly than mutation and adaptation. Hopefully, this ability will allow some mitigation of climate change on coral reefs. — CA

Science, this issue p. 895; see also p. 798.

GUT MICROBIOTA

Bacteria breach intestinal barriers

In an ironic complication of liver cirrhosis, beneficial microbes can escape from the gut and cause serious infections—or

Downloaded from www.sciencemag.org on May 26, 2014

PHOTO: Y. SATO *ET AL.*; KYLE DAVIS AND PAUL SCOFIELD/CANTERBURY MUSEUM

Published by AAAS

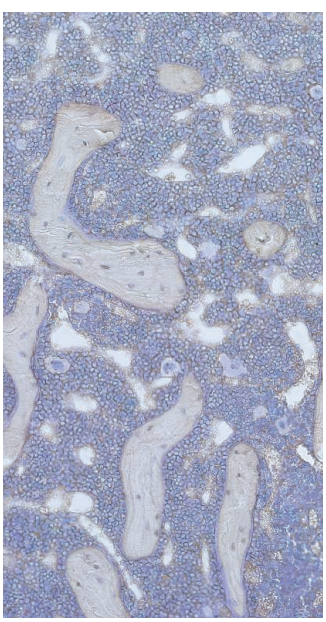
even death. Balmer *et al.* now show that the blood vessels of the healthy liver form a barrier to runaway gut bacteria. However, in animal models of liver disease and gut dysfunction and in patients with nonalcoholic liver disease, the liver is unable to capture these escapees. The bacteria then leak into the blood system, activating immune responses that break down the mutualistic relationship between the gut microbes and the host. This type of breakdown is an important complication of liver disease. — OMS

Sci. Transl. Med. **6**, 237ra66 (2014).

BONE PHYSIOLOGY

Another way of growing strong bones

To stay strong, bones are constantly rebuilding themselves. Thyroid hormones regulate this process by entering cells and binding to nuclear receptors,



Thyroid hormones increase the number of cells (blue) in trabecular bone.

which travel to the nucleus, where they change gene expression. However, these hormones also stimulate rapid cellular changes that do not require gene regulation. Kalyanaraman *et al.* found a different form of nuclear receptor in bone cells. When bound to thyroid hormones, this receptor increased the numbers of bone cells and protected them from death. When the researchers treated mice lacking thyroid hormones with a compound that mimicked the effects of this receptor, their bone cells grew normally. — JFF

Sci. Signal. **7**, ra48 (2014).

CANCER IMMUNOLOGY

Origins of tumor macrophages

To help the immune system fight cancer, it is important to understand the origins and functions of immune cells in tumors and the surrounding tissues. One type of immune cells, macrophages, is present both in tumors and in nearby noncancerous tissue, but the relationship between these two cell populations is unclear. Franklin *et al*. found that tumor-associated macrophages in mouse mammaries differed in form, function, and origin from macrophages found in nearby noncancerous mammary tissue. Moreover, when they removed macrophages from the tumors but not the other mammary tissue, tumors shrank and cytotoxic T cells—another kind of immune cell that kills tumor cells—infiltrated the tumors. Tumor-associated macrophages may thus be an important therapeutic target. — KLM

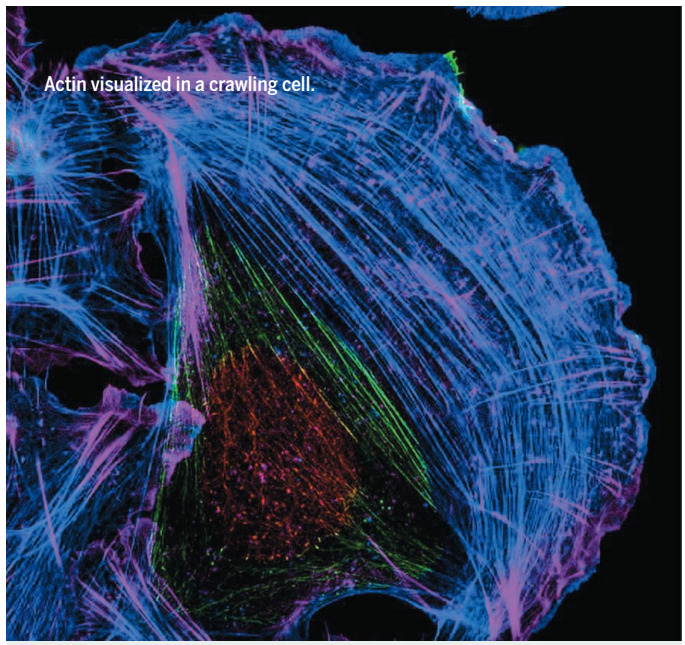
Science, this issue p. 921.

NEUROSCIENCE

Decisions, decisions, decisions...

Flies, like humans, deliberate before making perceptual judgments: They ponder difficult decisions longer than they do easy ones. DasGupta *et al*. measured reaction times in flies choosing between different smells. Mutations in a particular gene, they found, could cause indecision. Mutations in the same gene are implicated in intellectual disability, learning deficits, and language impairment. — PRS

Science, this issue p. 901.



Actin visualized in a crawling cell.

CELL MOTILITY

Cells need to stay in shape, too

To move efficiently, people need to stay in shape—and the same is true for cells. Burnette *et al*. looked at the 3D organization of contractile fibers used by living animal cells as they crawled about on a surface. The cells adopted a wedge-like shape with a wide, flattened front end dragging a slim rear. To keep moving, the cells used a counterbalanced contraction-adhesion system. At the top of the cell, a network of contractile fibers made from actin and myosin (the same proteins used in muscles) coupled to noncontracting stress fibers anchored to the cell's surface. Understanding how cells move is important for understanding normal development, wound healing, and metastasizing tumor cells. — SMH

J. Cell Biol. **205**, 83 (2014).

NEUROSCIENCE

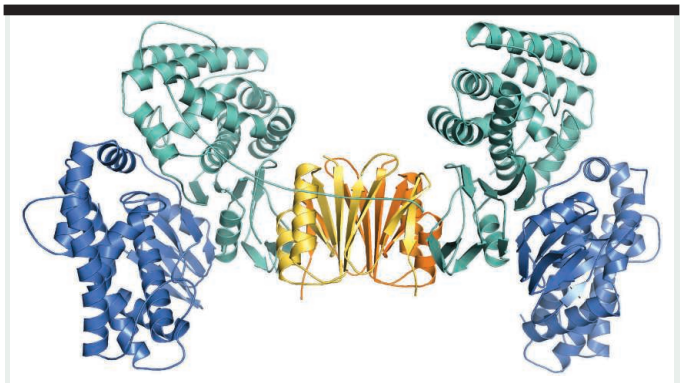
How the brain responds to fairness

Many people consider freedom of choice and fairness fundamental values, but what are their neural bases? To probe the question, Aoki *et al*. had pairs of people put their heads in functional magnetic resonance scanners and then play a game. When both players were offered an equal number of choices, they were more likely to report feeling happy, and their brain scans showed increased activity in the area called the ventromedial prefrontal cortex. In contrast, when the combined absolute number of options available to players increased, so did activity in the ventral striatum. Because these regions have been implicated already in value processing, these results may illuminate how a sense of fairness evolved in the human brain. — PRS

J. Neurosci. **34**, 6413 (2014).

PHOTO: D. BURNETTE *ET AL*., *THE JOURNAL OF CELL BIOLOGY* **205**, 1 (14 APRIL 2014) © THE ROCKEFELLER UNIVERSITY PRESS; KALYANARAMAN *ET AL*.

Published by AAAS



Crystal structure of the Cas1-Cas2 complex of the CRISPR-Cas system.

HOST DEFENSE

Cas proteins help acquire immunity

Bacteria and archaea are under constant attack from foreign genetic elements. The CRISPR-Cas immune system affords protection from such invaders. Upon encountering foreign DNA, the CRISPR-Cas system creates DNA fragments that match the sequences of the invading foreign DNA and then inserts these DNA fragments, or "spacers" into the microbe's genome. That way, the next time the organism encounters something similar, it can quickly recognize it and defend against it through an RNA interference–like mechanism. By solving the crystal structures and performing additional biochemical analysis, Nuñez *et al*. now uncover the specific functions of the enzymes Cas1 and Cas2. Acting as a complex, Cas1 and Cas2 help bacteria and archaea acquire DNA spacers and insert them correctly into the host genome. — VV

Nat. Struct. Mol. Biol. 10.1038/nsmb.2820 (2014).

RNA TRANSLATION

Yeast's translational hopscotch

Ribosomes translate mRNA into proteins sequentially, one codon at a time—except when they don't. Lang *et al.* now report that many of the mitochondrial genes in the yeast *Magnusiomyces capitatus* are infested with short sequence inserts that should kill the translation of the coded protein and, in theory, the yeast as well, but don't. Why not? The *M. capitatus* protein-synthesizing machinery was able to ignore the inserts and make functional proteins. The protein-synthesizing ribosomes recognized a special element in the mRNA that warned of the inserts. The ribosome then "hopped" from the upstream codon to an identical or very similar codon downstream, clean over the insert, leaving the insert out of the protein it made. — GR

Proc. Natl. Acad. Sci. U.S.A. 111, 5926 (2014).

OCEANOGRAPHY

What goes in does not come out

Vast swaths of floating plastic debris in a northern Pacific Ocean region have earned it a grim nickname: the Great Pacific Garbage Patch. But how much plastic really floats in the Pacific? Different research teams use different methodologies, and coverage is often patchy, making data notoriously difficult to obtain and compare. Using data spanning over 40 years, Law *et al.* report maps of plastic debris concentrations in the Eastern Pacific and estimate that at least 21,000 metric tons of microplastic are floating in the region. However, despite increased plastic production over the past decade, they could not detect an increase in plastic debris over time. Data coverage may be insufficient for capturing such trends. — JFU

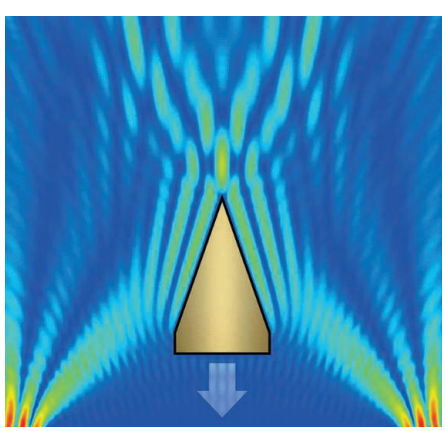
Environ. Sci. Technol. 10.1021 es4053076 (2014).

APPLIED PHYSICS

An attractive-sounding proposition

Demore *et al.* are using sound to build a better tweezer—or at least, to show how it could be done. Today's optical and acoustic tweezers typically use intensity gradients to trap and manipulate the particles. That's because propagating fields such as light and sound are usually associated with positive forces—the radiation pressure they create tends to push objects away. In contrast, Demore *et al.* have figured out how to use sound not just to push, but to pull. They use an array of ultrasonic resonators to show how the wavefront of a propagating beam of sound can be shaped to apply negative radiation pressure, creating a tractor beam. The new technology could give surgeons and astronauts greater dexterity as they perform operations and repairs from a distance. — ISO

Phys. Rev. Lett. **112**, 174302 (2014).



Map of the distribution of pressures within the field produced by the ultrasonic tractor beam.

ASTROPHYSICS

A chilly little neighborhood object

There's a chilly little neighbor lurking near the Sun, just ~7 light years away. Luhman detected the substellar object, a brown dwarf, by tracking the relative motions of objects in infrared images. Unlike the Sun, such low-mass objects cannot sustain the hydrogen fusion necessary to radiate visible light, so they produce energy primarily at thermal wavelengths. The brown dwarf probably has a mass 3 to 10 times that of Jupiter and a temperature near the freezing point on Earth, making it the coldest brown dwarf detected so far. That means astrophysicists will be able to test atmospheric models at new thermal low values. — MMM

Astrophys. J. 10.1088/2041-8205/786/2/L18 (2014).

MARINE BIOLOGY

Why octopuses don't get tied in knots

An octopus's appendages can form seemingly infinite postures and positions, but somehow they avoid becoming hopelessly entangled. The key appears to be chemicals in the skin, By examining amputated arms from the common octopus (*Octopus vulgaris*), Nesher *et al*. found that the animal's suckers would latch on to everything except its own arms. Petri dishes coated with the octopus's skin, whether intact or ground up into a mush, became "immune" to the zombie arms, suggesting that a substance in the skin repels the suckers. As octopuses are known to dine on their comrades, the substance may also prevent them from eating themselves alive. — NA

Curr. Biol. 10.1016/j.cub.2014.04.024 (2014).

PHOTO: J. K. NUÑEZ *ET AL*., *NATURE STRUCTURAL & MOLECULAR BIOLOGY* (4 MAY 2014) © 2014 NATURE PUBLISHING GROUP; MIKE MACDONALD AND CHRISTINE DEMORE

Published by AAAS

IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

SURFACE SCIENCE

Probing bonding profiles with a CO tip

Greater resolution has been achieved in atomic force microscopy by terminating the tip with a sharper probe: an adsorbed CO molecule. Chiang *et al.* now show that the adsorbed CO tip can reveal the bonding within cobalt phthalocyanine molecules absorbed on silver or gold surfaces. Inelastic tunneling spectroscopy reveals variations in the vibration excitation of the CO molecule that can map out the internal bonding of the molecules, as well as hydrogen bonding between molecules. — PDS

Science, this issue p. 885

INFECTIOUS DISEASE

Progress toward an effective malaria vaccine

The history of efforts to develop a malaria vaccine has been long and difficult. Raj *et al.* probed for molecules produced by this blood parasite that are recognized by natural immune responses of people living in malaria-endemic areas of Africa. One, PfSEA-1, blocked parasite exit from red blood cells. Vaccination experiments with mouse malaria showed almost fourfold reduction in parasitemia; moreover, passive transfer of PfSEA-1 antibodies transferred protection from mouse to mouse. Encouragingly, the presence of PfSEA-1 antibodies correlates with significant protection from severe malaria in children and adolescents from Kenya and Tanzania. — CA

Science, this issue p. 871

MICROBIAL GENOMICS

In translation, sometimes stop can mean go

The genetic code appears to be largely conserved across all domains of life. Although limited deviations have been reported, Ivanova *et al*. used metagenomics to survey the prevalence of stop codon reassignment in naturally occurring microbial populations. Certain bacteria and bacteriophages exhibited lineage-specific recoding of their stop codons. In one specific phage, the genome was restructured into two genetic sets. One set of genes was encoded in a way that didn't gel with the host genome and probably helped with infection. A second set of more host-compatible sequences encoded proteins expressed in the later stages of infection. — LMZ

Science, this issue p. 909

NEUROSCIENCE

Firing, wiring, and Hebbian remodeling

Correlated neuronal activity is generally thought to drive circuit remodeling in the central nervous system. This model, first proposed by Hebb, is strongly supported by several lines of evidence, though it has been difficult to directly observe such changes in real time. Munz *et al*. developed an experimental approach to watch structural remodeling of neuronal axons in vivo at high temporal resolution. They measured changes in synaptic efficacy while presenting specific patterned stimuli to test the Hebb model. Although the key predictions of Hebbian developmental plasticity were upheld, the mechanistic details of how this occurred were unexpected. — PRS

Science, this issue p. 904

CANCER GENOMICS

Candidate Cushing's culprit identified

Cushing's syndrome is a rare condition resulting from the excess production of cortisol. About 15% of Cushing's syndrome cases are associated with an adrenocortical tumor. However, the genetic etiology of these adrenocortical tumors is ill defined (see the Perspective by Kirschner). Cao *et al*. and Sato *et al*. both performed whole-exome sequencing of tumors from individuals with adrenal Cushing's syndrome and compared it with the patient's own matched nontumor DNA and identified recurrent mutations in the *protein kinase A catalytic subunit alpha* (*PRKACA*) gene, as well as less frequent mutations in other putative pathological genes. The most common recurrent mutation activated the kinase, which may suggest a potential therapeutic target. — LMZ

Science, this issue p. 913, p. 917; see also p. 804

Published by AAAS

RESEARCH ARTICLES

INFECTIOUS DISEASE

Antibodies to PfSEA-1 block parasite egress from RBCs and protect against malaria infection

Dipak K. Raj,[1] Christian P. Nixon,[1] Christina E. Nixon,[1] Jeffrey D. Dvorin,[2] Christen G. DiPetrillo,[2] Sunthorn Pond-Tor,[1] Hai-Wei Wu,[1,3] Grant Jolly,[4] Lauren Pischel,[1] Ailin Lu,[1] Ian C. Michelow,[1,3] Ling Cheng,[1] Solomon Conteh,[5] Emily A. McDonald,[1] Sabrina Absalon,[2] Sarah E. Holte,[6] Jennifer F. Friedman,[1,3] Michal Fried,[5]* Patrick E. Duffy,[5]* Jonathan D. Kurtis[1,4]*†

Novel vaccines are urgently needed to reduce the burden of severe malaria. Using a differential whole-proteome screening method, we identified *Plasmodium falciparum* schizont egress antigen-1 (PfSEA-1), a 244-kilodalton parasite antigen expressed in schizont-infected red blood cells (RBCs). Antibodies to PfSEA-1 decreased parasite replication by arresting schizont rupture, and conditional disruption of PfSEA-1 resulted in a profound parasite replication defect. Vaccination of mice with recombinant *Plasmodium berghei* PbSEA-1 significantly reduced parasitemia and delayed mortality after lethal challenge with the *Plasmodium berghei* strain ANKA. Tanzanian children with antibodies to recombinant PfSEA-1A (rPfSEA-1A) did not experience severe malaria, and Kenyan adolescents and adults with antibodies to rPfSEA-1A had significantly lower parasite densities than individuals without these antibodies. By blocking schizont egress, PfSEA-1 may synergize with other vaccines targeting hepatocyte and RBC invasion.

Plasmodium falciparum malaria is a leading cause of morbidity and mortality in developing countries, infecting hundreds of millions of individuals and killing up to 1 million children in sub-Saharan Africa each year (*1*, *2*). Children suffer the most from malaria, yet vaccine discovery efforts have not targeted this age group. Of the ~100 vaccine candidates currently under investigation, more than 60% are based on only four parasite antigens (*3*, *4*). New antigen candidates are urgently needed, but strategies to identify novel antigens are limited.

Human residents of endemic areas develop protective immunity that limits parasitemia and disease; thus, naturally acquired human immunity provides an attractive model for vaccine antigen identification. Plasma from some chronically exposed individuals contains antibodies that restrict parasite growth ex vivo (*5*) and after adoptive transfer (*6*). One approach to identifying vaccine antigens is to identify malarial proteins that are only recognized by antibodies in the plasma of chronically exposed individuals who remain resistant to infection but are not recognized by antibodies in the plasma of susceptible individuals.

[1]Center for International Health Research, Rhode Island Hospital, The Warren Alpert Medical School of Brown University, Providence, RI 02903, USA. [2]Division of Infectious Diseases, Boston Children's Hospital and Harvard Medical School, Boston, MA 02115, USA. [3]Department of Pediatrics, Rhode Island Hospital, The Warren Alpert Medical School of Brown University, Providence, RI 02903, USA. [4]Department of Pathology and Laboratory Medicine, Rhode Island Hospital, The Warren Alpert Medical School of Brown University, Providence, RI 02906, USA. [5]Laboratory of Malaria Immunology and Vaccinology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, MD 20892, USA. [6]Fred Hutchinson Cancer Research Center Program in Biostatistics and Biomathematics, Department of Biostatistics and Global Health, University of Washington, Seattle, WA 98109, USA. *These authors contributed equally to this work. †Corresponding author. E-mail: jonathan_kurtis@brown.edu

Identification and in Silico Evaluation of Vaccine Candidates

Using our cDNA library–based differential screening method (*7*) and plasma and epidemiologic data from a Tanzanian birth cohort (*8*), we probed the *P. falciparum* blood-stage proteome with plasma from resistant and susceptible 2-year-old children to identify parasite proteins that are the targets of protective antibody responses. We selected 2-year-olds because, in our cohort, resistance to *P. falciparum* parasitemia is first detected at this age (fig. S1).

We selected 12 resistant and 11 susceptible 2-year-old children with partial matching for gender and village of residence, which may be related to resistance (table S1). Resistance was determined based on the mean parasite density in all blood films collected from children between ages 2 and 3.5 years. We pooled plasma collected at age 2 years (±2 weeks) from the resistant individuals and the susceptible individuals and performed differential screening experiments on a *P. falciparum* 3D7 strain blood-stage cDNA library. We screened 1.25×10^6 clones and identified three clones that were recognized by antibodies in plasma from resistant, but not susceptible, individuals. The sequences of these clones were compared with the published *P. falciparum* genome (www.PlasmoDB.org) and found to encode nucleotides (nt) 2431 to 3249 of *PF3D7_1021800*, a hypothetical gene on chromosome 10; nt 3490 to 5412 of *PF3D7_1134300*, a hypothetical gene on chromosome 11; and nt 201 to 1052 of *PF3D7_1335100*, which encodes merozoite surface protein-7 (MSP-7), a protein involved in red blood cell (RBC) invasion, which is currently under study as a potential vaccine candidate (*9*–*12*).

In silico analysis (www.PlasmoDB.org and www.OrthoMCL.org) predicted that *PF3D7_1021800* contains a 6744–base pair (bp) gene that encodes a 244-kD acidic phosphoprotein (*13*), with three introns near its 3′ end, and syntenic orthologs in all rodent and primate malarias evaluated to date, but not in other genera. Based on in vitro experiments, we designated the protein product of *PF3D7_1021800* as *Plasmodium falciparum* schizont egress antigen-1 (PfSEA-1) and its corresponding gene as *PfSEA1*. PfSEA-1 contains multiple complex repeat regions and extensive regions of low amino acid complexity, with 50% of the protein being composed of four amino acids (Asn, Lys, Glu, and Asp).

PfSEA1 expression increases throughout blood-stage schizogeny, and the gene displays minimal sequence variation in the immunorelevant region recognized in our differential screens (nt 2431 to 3249). A recently reported deep sequencing effort on 227 field samples identified only three nonsynonymous and one synonymous single-nucleotide polymorphisms in the cloned region (*14*).

Conditional Destabilization of PfSEA-1

PfSEA-1 has no significant homology to proteins of known function. Multiple attempts to disrupt *PfSEA1* by homologous recombination were unsuccessful, which suggests that PfSEA-1 is essential for blood-stage replication. Using the conditional destabilization (knockdown) system, we generated a parasite strain with a destabilization domain and hemagglutinin (HA) tag fused to the C terminus of endogenous PfSEA-1 (*15*) and verified the strain by Southern blot and sequencing across the insertion boundary (fig. S2, A and B). After removal of the stabilizing agent, Shield-1, expression of PfSEA-1 decreased by 75% (Fig. 1A), and parasites with destabilized expression of PfSEA-1 had a marked, 80% inhibition of parasite replication as compared with parasites expressing normal levels of PfSEA-1 (Fig. 1B). In addition, parasites grown for a single cycle in the absence of Shield-1 had delayed schizont egress (fig. S2C) accompanied by a 40% decrease in the number of ring-stage parasites 12 hours after schizont rupture (fig. S2D). The addition of Shield-1 did not alter the replication of wild-type parasites (fig. S3), and conditional destabilization of nonessential genes did not produce a replication defect (*15*, *16*).

Expression and Validation of rPfSEA-1

We expressed and purified the polypeptide encoded by the differentially recognized immunorelevant

region (nt 2431 to 3249, amino acids 810 to 1083) in *Escherichia coli* and designated this recombinant protein rPfSEA-1A (fig. S4A). Before evaluation in the entire birth cohort, we performed an initial validation enzyme-linked immunosorbent assay (ELISA) using an independent selection of resistant and susceptible individuals from our Tanzanian birth cohort (table S2). Immunoglobulin G (IgG) antibody recognition of rSEA-1A was 4.4-fold higher in plasma pooled from resistant individuals (n = 11) than in susceptible individuals (n = 14, fig. S4B). This pattern of recognition was not observed for other malaria proteins and controls (fig. S4C), which supported our library screening results.

We have cloned the immunorelevant region into a eukaryotic expression plasmid (VR2001) and immunized mice with either the recombinant protein (rPfSEA-1A) or DNA constructs to generate antisera to PfSEA-1A (anti–PfSEA-1A). To confirm that *PfSEA1* encodes a parasite protein, we probed *P. falciparum* infected and uninfected RBCs with antisera prepared by DNA vaccination. Anti–PfSEA-1A recognized a 244-kD protein in infected but not uninfected RBCs (fig. S5A), and recognized rPfSEA-1A (fig. S5B).

Anti-PfSEA-1 Mediates Growth Inhibition of Parasites in Vitro

We performed growth inhibition assays using anti–rPfSEA-1A prepared by either DNA or recombinant protein immunization (Fig. 2). Parasites were synchronized to the ring stage, cultured to obtain mature trophozoites, and then incubated with anti–rPfSEA-1A or controls for 24 hours, followed by enumeration of newly invaded ring-stage parasites. Anti–rPfSEA-1A generated by DNA plasmid or recombinant protein–based immunization inhibited parasite growth by 58 to 74% across three parasite strains compared with controls (all $P < 0.009$).

We developed monoclonal antibodies (mAbs) in rPfSEA-1A–immunized mice that recognized rPfSEA-1A by ELISA. Some (mAb CF6), but not all (mAb CF3), mAbs mediated up to 75% growth inhibition at a concentration of 250 μg/ml (Fig. 2F), which is substantially lower than the inhibitory concentration of mAb to MSP-1 (*17*). In addition, we purified human polyclonal antibodies to PfSEA-1 from pooled plasma using rPfSEA-1A coupled to Sepharose beads and demonstrated that these human antibodies to PfSEA-1 significantly inhibited parasite growth (fig. S6, A and B).

Stage-Specific Immunolocalization of PfSEA-1

We immunolocalized PfSEA-1 by immunofluorescence confocal microscopy and immunogold

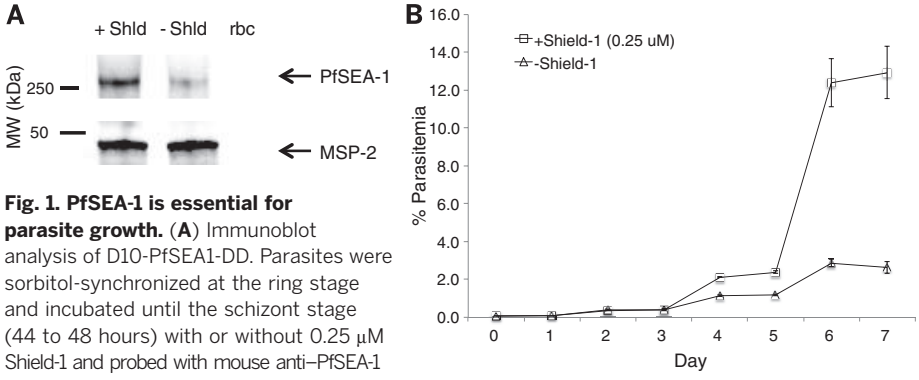


Fig. 1. PfSEA-1 is essential for parasite growth. (**A**) Immunoblot analysis of D10-PfSEA1-DD. Parasites were sorbitol-synchronized at the ring stage and incubated until the schizont stage (44 to 48 hours) with or without 0.25 μM Shield-1 and probed with mouse anti–PfSEA-1 produced by DNA immunization (top panel) or rabbit anti-MSP2 (loading control, bottom panel). Lane 1, Shield-1–treated cultures; lane 2, untreated cultures; lane 3, uninfected RBCs. The top arrow indicates PfSEA-1 fused with a DD tag; the bottom arrow indicates MSP-2. (**B**) Replication curves for D10-PfSEA1-DD parasites. Ring-stage parasites were cultured with or without 0.25 μM Shield-1. Parasitemia was measured by flow cytometry. Each data point represents the mean of three replicates; error bars indicate SEM.

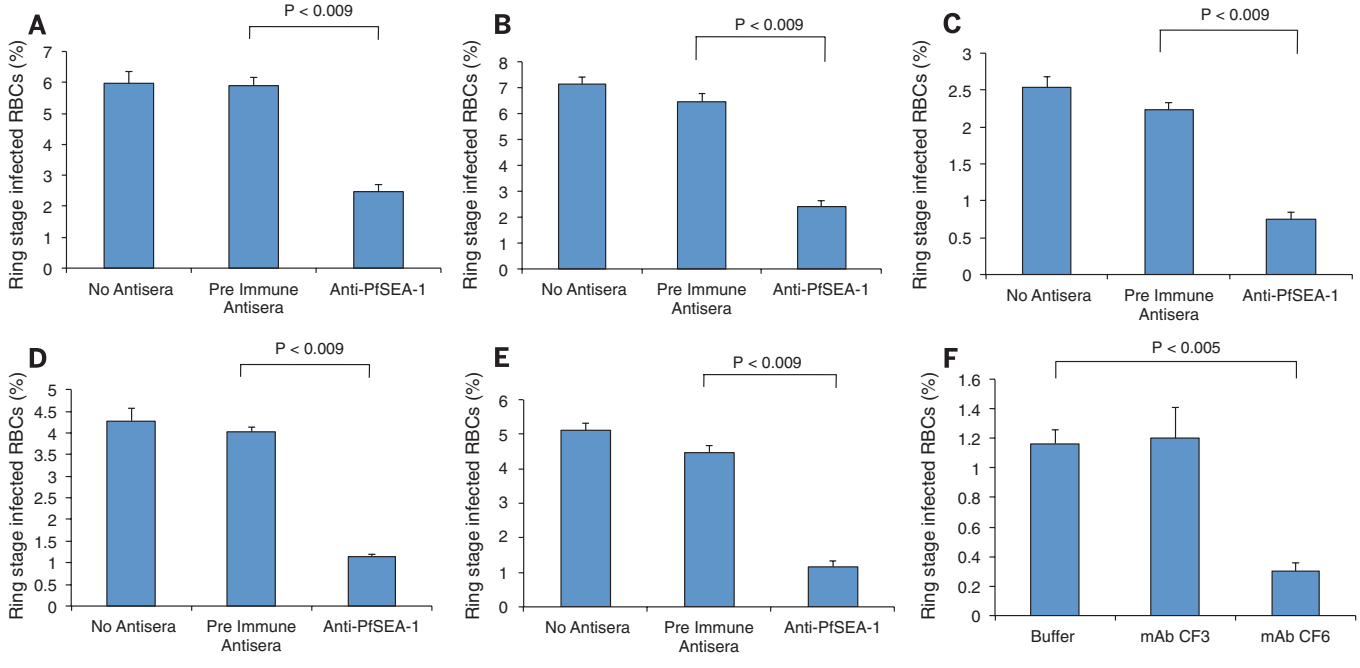


Fig. 2. Antibodies to PfSEA-1 inhibit malaria parasite growth. Polyclonal antibodies to PfSEA-1 generated by DNA (**A** to **C**) and recombinant protein (**D** and **E**) immunization, and mAbs to PfSEA-1 (**F**) inhibit parasite growth and invasion by 58 to 74% for three parasite strains in vitro. Mature trophozoite-stage 3D7 [(A), (D), and (F)], W2 [(B) and (E)], and D10 (C) parasites were were cultured in the presence of mouse antisera to PfSEA-1 [1:10 dilution, (A) to (E)]. Negative controls included no mouse sera and preimmune mouse sera (1:10 dilution). In (F), parasites were incubated in the presence of mAbs to PfSEA-1 CF3 or CF6 (250 μg/ml) or buffer. Parasites were cultured for 24 hours, and ring-stage parasites were enumerated by microscopy. Bars represent the mean of five independent replicates, with each replicate performed in triplicate. Error bars represent SEMs. Comparisons between pre- and postimmune mouse sera by a nonparametric Mann-Whitney U test are indicated. (A) to (C) and (F) are representative of three independent experiments. (D) and (E) are representative of five independent experiments.

transmission electron microscopy (Fig. 3). Anti–PfSEA-1 did not bind to free merozoites, rings, or late trophozoite–stage parasites, but did specifically recognize an antigen expressed by late schizont–infected RBCs (Fig. 3, A and 3). To determine whether antibodies to PfSEA-1 could access PfSEA-1 in vivo, we used a live cell–labeling technique. PfSEA-1 colocalized with glycophorin A in nonpermeabilized, unfixed schizont-infected RBCs (Fig. 3C). We confirmed the stage-specific expression and localization of PfSEA-1 using antibodies to the HA tag in D10-PfSEA1-DD parasites, in which PfSEA-1 is fused with an HA tag (fig. S7).

The accessibility of antibodies to PfSEA-1 in living parasites was further evaluated by immuno–electron microscopy (Fig. 3D and fig. S8). When the same live cell–labeling technique was used, PfSEA-1 localized to the schizont/parasitophorous vacuole membrane, Maurer's clefts, and the inner leaflet of the RBC membrane, whereas glycophorin A was confined to the outer leaflet of the RBC membrane. This pattern of staining was observed in the majority of late schizont–infected RBCs examined. The close juxtaposition of these structures in late schizont–infected RBCs with the RBC outer membrane explains the apparent colocalization of PfSEA-1 with glycophorin A observed by confocal microscopy. The accessibility of antibodies to PfSEA-1 in nonpermeabilized, unfixed schizont–infected RBCs is consistent with the known permeability of parasitized RBCs to macromolecules, including antibodies, at the later stages of schizogony (*18*–*20*).

Anti–PfSEA-1 Mediates Schizont Arrest of Parasites in Vitro

The localization of PfSEA-1 was not consistent with a role in RBC invasion; rather, it suggested a role in parasite egress from RBCs. To determine the mechanism of growth inhibition, we performed schizont arrest assays using anti–rPfSEA-1A prepared by either DNA or recombinant protein immunization, as well as mAb directed against rPfSEA-1A (Fig. 4). Parasites were synchronized to the ring stage at high (3.5%) parasite density, cultured to obtain early schizonts, and then incubated with anti–rPfSEA-1A or controls for 12 hours, followed by enumeration of the remaining schizont-stage parasites. Under these conditions, the majority of schizont-infected RBCs should rupture, releasing merozoites to invade new RBCs and develop into ring-stage parasites. Anti–rPfSEA-1A generated by DNA plasmid– (Fig. 4, A to C) or recombinant protein– (Fig. 4, D and E, and fig. S9) based immunization, as well as mAb anti–rPfSEA-1A (Fig. 4F) inhibited schizont egress in a dose-dependent manner (Fig. 4G), resulting in 4.3- to 6.8-fold higher proportions of schizonts across three parasite strains as compared with controls (all $P < 0.009$). In addition, purified human polyclonal antibodies to PfSEA-1 significantly inhibited schizont egress (fig. S6C).

Vaccination with PbSEA-1 Protects Mice from *P. berghei* ANKA Challenge

To evaluate the protective efficacy of vaccination with SEA-1 in vivo, we selected the *Plasmodium berghei* ANKA strain model, because of its aggressive parasite growth rate, extreme lethality in mice, and the failure of known vaccine candidates [such as PbAMA-1 and PbMSP-1 (*21*)] to afford protection. We cloned the *P. berghei* ANKA strain ortholog of rPfSEA-1A (nt 2173 to 3000, 47% similar and 34% identical amino acid sequence as compared with rPfSEA-1A) into the expression plasmid pET30 and expressed and purified rPbSEA-1A (amino acids 725 to 1000) from *E. coli* (Fig. 5A). We conducted five (four active vaccination with rPbSEA-1A and one passive transfer of antibodies to PbSEA-1A) trials in mice with *P. berghei* ANKA challenge. After vaccination, mice generated robust anti–rPbSEA-1A immunoglobulin G (IgG) responses in each trial (Fig. 5B and fig. S10). In trial 1, BALB/c mice vaccinated intraperitoneally (ip) with TiterMax and challenged ip with 10^6 *P. berghei* ANKA–infected RBCs (iRBCs) had 2.25-fold higher parasitemia on day 7 after challenge, as compared with mice vaccinated with rPbSEA-1A in TiterMax

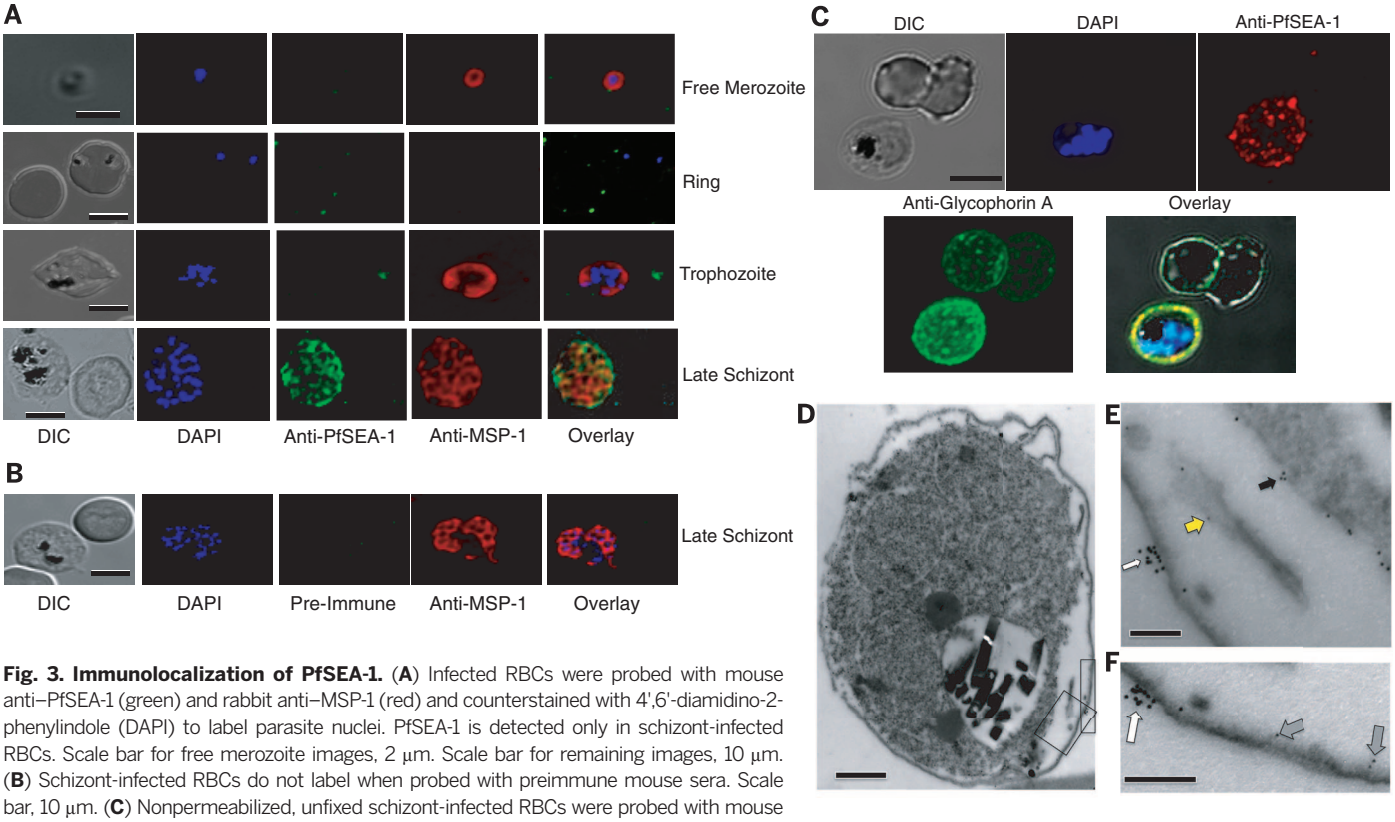


Fig. 3. Immunolocalization of PfSEA-1. (**A**) Infected RBCs were probed with mouse anti–PfSEA-1 (green) and rabbit anti–MSP-1 (red) and counterstained with 4',6'-diamidino-2-phenylindole (DAPI) to label parasite nuclei. PfSEA-1 is detected only in schizont-infected RBCs. Scale bar for free merozoite images, 2 μm. Scale bar for remaining images, 10 μm. (**B**) Schizont-infected RBCs do not label when probed with preimmune mouse sera. Scale bar, 10 μm. (**C**) Nonpermeabilized, unfixed schizont-infected RBCs were probed with mouse anti–PfSEA-1 (red) and rabbit anti–glycophorin A (green) and counterstained with DAPI to label parasite nuclei. PfSEA-1 colocalized with glycophorin A to the surface of schizont-infected RBCs. Scale bar, 5 μm. (**D** to **F**) Nonpermeabilized, unfixed schizont-infected RBCs were probed with mouse anti–PfSEA-1 (5-nm gold particles) and rabbit anti–glycophorin A (10-nm gold particles). PfSEA-1 localized to the schizont/parasitophorous vacuole membrane (black arrow), Maurer's clefts (yellow arrow), and the inner leaflet of the RBC membrane (gray arrow), whereas glycophorin A was confined to the outer leaflet of the RBC membrane (white arrow). Scale bar in (D), 0.5 μm. Scale bars in (E) and (F), 0.1 μm.

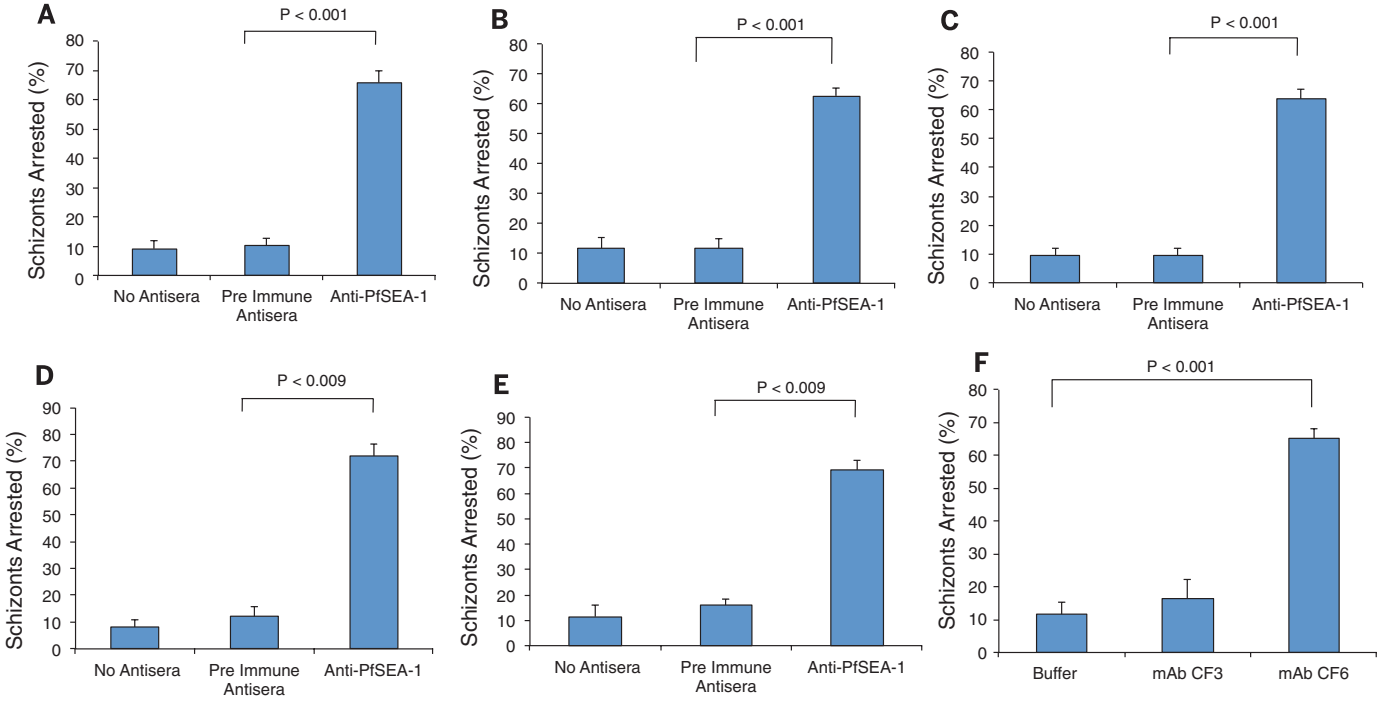
(Mann-Whitney U test, $P = 0.0002$, Fig. 5C). All control mice were killed on day 7 because of high parasitemia and associated illness, in accordance with our animal protocol Center for International Health Research, Rhode Island Hospital (CIHR). In trial 2, C57BL/6 mice vaccinated subcutaneously (sc) with TiterMax and challenged with 200 *P. berghei* ANKA sporozoites intravenously (iv) had 1.52-fold higher parasitemia on day 7 after challenge, as compared with mice vaccinated with rPbSEA-1A in TiterMax (Mann-Whitney U test, $P = 0.002$, Fig. 5D). All mice were killed on day 8 in accordance with our animal protocol (National Institutes of Health). In trial 3, BALB/c mice vaccinated ip with TiterMax and challenged ip with 10^4 *P. berghei* ANKA iRBC had 3.05-fold higher parasitemia on day 7 after challenge, as compared with mice vaccinated with rPbSEA-1A in TiterMax (Mann-Whitney U test, $P = 0.001$, Fig. 5E). All control mice were killed on day 8 because of high parasitemia and associated illness, in accordance with our animal protocol (CIHR). Trials 4 (active vaccination with rPbSEA-1A) and 5 (passive transfer of antibodies to PbSEA-1A) were conducted under a modified animal protocol (CIHR) that allowed the collection of survival outcomes. In trial 4, BALB/c mice vaccinated ip with TiterMax and challenged ip with 10^4 *P. berghei* ANKA iRBC had 3.9-fold as high parasitemia on day 7 after challenge, as compared with mice vaccinated with rPbSEA-1A in TiterMax (Mann-Whitney U test, $P = 0.0009$, Fig. 5F). The median survival of rPbSEA-1A–vaccinated mice was 1.8-fold longer than that of mice vaccinated with TiterMax alone (log-rank test, $P < 0.0001$, Fig. 5G).

To evaluate the role of humoral anti–SEA-1 responses in mediating protection, we conducted passive transfer experiments, followed by *P. berghei* ANKA challenge. In trial 5, BALB/c mice passively transferred with control mouse sera and challenged ip with 10^4 *P. berghei* ANKA iRBC had 2.7-fold higher parasitemia on day 7 after challenge, as compared to mice passively transferred with anti–PbSEA-1A (Mann-Whitney test, $P = 0.009$, Fig. 5H). The median survival of mice treated with anti–rPbSEA-1A was 2.0-fold longer than that of mice treated with control sera (log-rank test, $P < 0.0001$, Fig. 5I).

Our active vaccine studies included two routes of administration (sc and ip), two stages of challenge (sporozoite and iRBC), two doses of iRBC challenge (10^4 and 10^6 iRBC per mouse), two vaccination schedules, two mouse strains (BALB/c and C57BL/6), two methods of enumeration (microscopy and flow cytometry), and two performance sites (CIHR and NIH). In all experiments, rPbSEA-1A generated significant protection. These results represent the first example of significantly reduced parasitemia and delayed mortality after vaccination with a blood-stage antigen in *P. berghei* ANKA (*21*).

Human Antibody Responses to PfSEA-1 *Tanzanian Birth Cohort*

To evaluate the impact of naturally acquired antibodies to PfSEA-1 on clinical malaria, we measured antibody to PfSEA-1 IgG levels using a fluorescent bead–based assay in our birth cohort and related these levels to subsequent malaria outcomes. We measured levels of antibody to PfSEA-1 IgG in available plasma obtained at scheduled, nonsick visits between 1.5 and



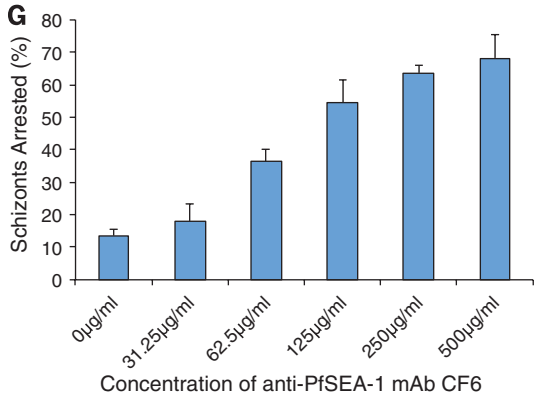


Fig. 4. Antibodies to PfSEA-1 inhibit schizont egress. Antibodies to PfSEA-1 generated by DNA (**A** to **C**) and recombinant protein (**D** and **E**) immunization, as well as mAb to rPfSEA-1A (**F** and **G**), inhibit schizont egress for three parasite strains in vitro. Early schizont stage 3D7 [(A) and (D)], W2 [(B) and (E)], and D10 (C) parasites were incubated in the presence of mouse anti–PfSEA-1 (1:10 dilution). Negative controls included no mouse sera and preimmune mouse sera (1:10 dilution). In (F), 3D7 parasites were incubated in the presence of mAbs to PfSEA-1 CF3 or CF6 (at 250 µg/ml) or buffer. In (G), 3D7 parasites were incubated in the presence of increasing concentrations of mAb to PfSEA-1 CF6 or buffer. Schizonts were counted 12 hours after treatment, and the percent of initial schizonts arrested was calculated. Bars represent the mean of five independent replicates, with each replicate performed in triplicate. Error bars represent SEMs. Comparisons between pre- and postimmune mouse sera and between buffer and mAb-treated cultures by a nonparametric Mann-Whitney U test are indicated. All panels are representative of three independent experiments.

3.5 years of life, encompassing 687 antibody measures on 453 children. The average time interval between each antibody measurement and either a subsequent antibody determination or completion of the study was 39 weeks. For each antibody measurement, the analysis interval for malaria outcomes extended from the time of the antibody measurement until the child had a subsequent antibody determination or completed the study.

Antibodies to PfSEA-1 were detectable in 6.0% of these samples, and children were followed for a total of 25,494 child-weeks of observation. We used Cox proportional hazards models to evaluate the relationship between

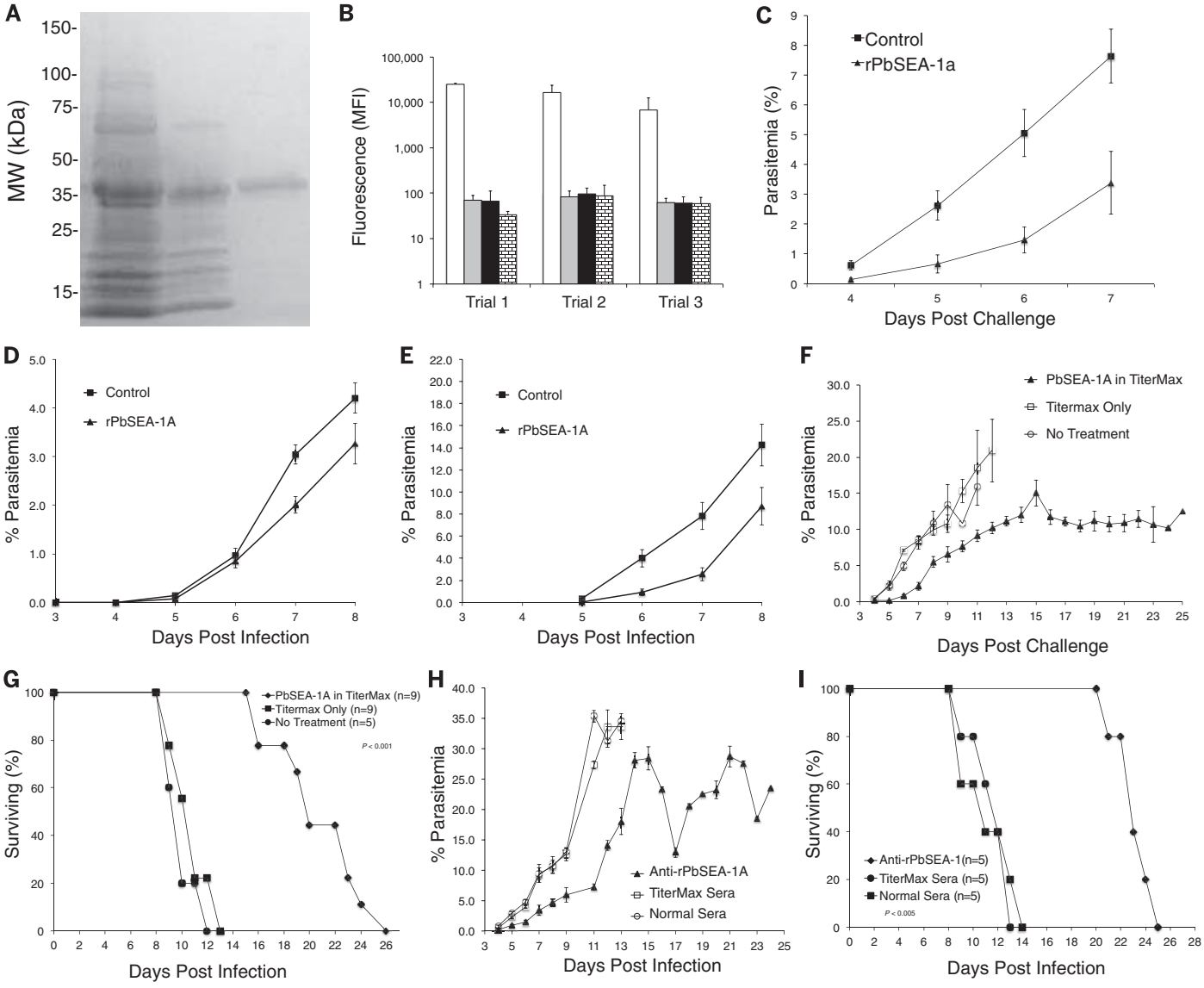


Fig. 5. Vaccination with rPbSEA-1A protects mice from challenge with *P. berghei* ANKA. (**A**) Expression and purification of rPbSEA-1A from *E. coli* soluble lysates. Lane 1, nickel chelate chromatography of soluble *E. coli* lysate; lane 2, hydrophobic interaction chromatography of lane 1; lane 3, anion exchange chromatography of lane 2. (**B**) Antibody response of mice vaccinated with rPbSEA-1A in three vaccine trials. After vaccination, mice generated robust anti–rPbSEA-1A IgG responses in each trial. Bars represent mean fluorescence; error bars represent SEMs. White bars represent anti–rPbSEA-1A IgG responses in mice immunized with rPbSEA-1A, gray bars represent anti-BSA IgG responses in mice immunized with rPbSEA-1A, black bars represent anti–rPbSEA-1A IgG responses in mice immunized with adjuvant alone, and bricked bars represent anti-BSA IgG responses in mice immunized with adjuvant alone. (**C**) Vaccine trial 1: BALB/c mice vaccinated ip with TiterMax (n = 11) and challenged ip with 10^6 *P. berghei* ANKA–infected RBCs had 2.25-fold higher parasitemia on day 7 after challenge (P = 0.0002) as compared to mice vaccinated with rPbSEA-1A (n = 10). (**D**) Vaccine trial 2: C57BL/6 mice vaccinated sc with TiterMax (n = 10) and challenged with 200 *P. berghei* ANKA sporozoites iv had 1.52-fold higher parasitemia on day 7 after challenge (P = 0.002) as compared to mice vaccinated with rPbSEA-1A (n = 10). (**E**) Vaccine trial 3: BALB/c mice vaccinated ip with TiterMax (n = 10) and challenged ip with 10^4 *P. berghei* ANKA–infected RBCs had 3.05-fold higher parasitemia on day 7 after challenge (P = 0.001) as compared to mice vaccinated with rPbSEA-1A (n = 10). (**F** and **G**) Vaccine trial 4: BALB/c mice vaccinated ip with TiterMax (n = 9) and challenged ip with 10^4 *P. berghei* ANKA–infected RBCs had 3.9-fold higher parasitemia on day 7 after challenge (P = 0.0009) as compared to mice vaccinated with rPbSEA-1A in TiterMax. The median survival of rPbSEA-1A vaccinated mice was 1.8-fold longer (P < 0.0001) as compared to that of adjuvant-only control mice (n = 9). (**H** and **I**) Vaccine trial 5: Antisera (rPbSEA-1A or adjuvant alone) or normal mouse sera were passively transferred (0.5 ml, ip) into naïve BALB/c mice (n = 5 mice per group) on day −1 and day +1. On day 0, mice were challenged ip with 10^4 *P. berghei* ANKA–infected RBCs. Control mice had 2.7-fold higher parasitemia on day 7 after challenge (P = 0.009), compared to anti-rPbSEA-1A treated mice. The median survival of rPbSEA-1A vaccinated mice was 2.0-fold longer than that of controls (P < 0.0001).

time-varying levels of antibody to PfSEA-1 (log-transformed) and malaria outcomes. Increasing levels of antibody to PfSEA-1 were associated with significantly decreased risk of severe malaria [HR = 0.669, 95% confidence interval (CI) [0.495, 0.905], P = 0.0092], and these results were essentially identical after adjusting for hemoglobin phenotype (HR = 0.663, 95% CI [0.489, 0.897], P = 0.0078). It was surprising that severe malaria did not occur during periods after children had detectable levels of antibody to PfSEA-1 (0 cases per 1688 child-weeks with detectable antibody to PfSEA-1 versus 45 cases per 23,806 child-weeks with undetectable antibody to PfSEA-1) (Fig. 6A).

To place the PfSEA-1 antibody results in context with other known malarial vaccine candidates, we performed multiplexed bead-based antibody analyses using plasma collected at 2 years of age from n = 225 available samples in our birth cohort. Using this multiplexed platform, we measured IgG antibody levels to MSP-1-19 (3D7 variant), MSP-3 (amino acids 99 to 265), LSA-N (amino acids 28 to 150), LSA-C (amino acids 1630 to 1909), and RAMA-E (amino acids 759 to 840). IgG antibody responses to these well-accepted vaccine candidates were not related to the risk of severe malaria when the same statistical analyses employed for PfSEA-1 were used in these available samples (all P = NS, see supplementary text).

Together, these results demonstrate that antibodies that specifically block schizont egress can predict resistance to severe malaria in humans and provide compelling evidence that PfSEA-1 represents a critical addition to the limited pool of candidate antigens currently being explored for their vaccine potential.

Kenyan Cohort

To further extend the generalizability of our human immunoepidemiologic results, we measured anti–PfSEA-1 IgG responses in an entirely distinct cohort of Kenyan adolescents and young adults. We performed a secondary analysis with data and serum samples that were collected from a cohort of Kenyan males as part of a treatment-reinfection study (*7*, *12*, *22*).

Volunteers were residents of subsistence farming, *P. falciparum*–endemic villages in western Kenya, north of Lake Victoria. The entomological inoculation rate in this area exceeded 300 infectious bites per person per year (*23*), and bed nets had not been introduced into the community. Males (n = 138) aged 12 to 35 years were entered into the study at the beginning of the high transmission season in April 1997. Detectable parasitemia was eradicated by the use of quinine sulfate (10 mg per kilogram of body weight twice daily for 3 days) and doxycycline (100 mg twice daily for 7 days), and individuals were followed with weekly blood smears for 18 weeks. Serum was collected 2 weeks after treatment and stored at −80°C. In this age group, clinical or severe malaria is very uncommon.

In generalized estimating equation (GEE) repeated-measures models, individuals with detectable IgG antibodies to rPfSEA-1A (n = 77 individuals who contributed 1154 weeks of follow-up blood smears) had 50% lower parasite densities over 18 weeks of follow-up than did individuals with no detectable IgG antibodies to rPfSEA-1A (n = 61 individuals who contributed 940 weeks of follow-up blood smears), after adjusting for age, week of follow-up, exposure to *Anopheles* mosquitoes, and hemoglobin phenotype (95% CI [3, 97], P < 0.04, Fig. 6B). We measured IgG antibody levels to MSP-3 (amino acids 99 to 265), MSP-7 (amino acids 117 to 248), LSA-N (amino acids 28 to 150), LSA-C (amino acids 1630 to 1909), and RAMA-E (amino acids 759 to 840) in these same samples. IgG antibody responses to these well-accepted vaccine candidates did not predict resistance to parasitemia in these samples when the same statistical analyses employed for PfSEA-1 were used (fig. S11, all P = NS).

Although residual confounding by unmeasured factors cannot be completely excluded, together with the data from the Tanzanian birth cohort, these data generalize and underscore the relation between anti–PfSEA-1 IgG responses and resistance to parasitemia and severe falciparum malaria in humans.

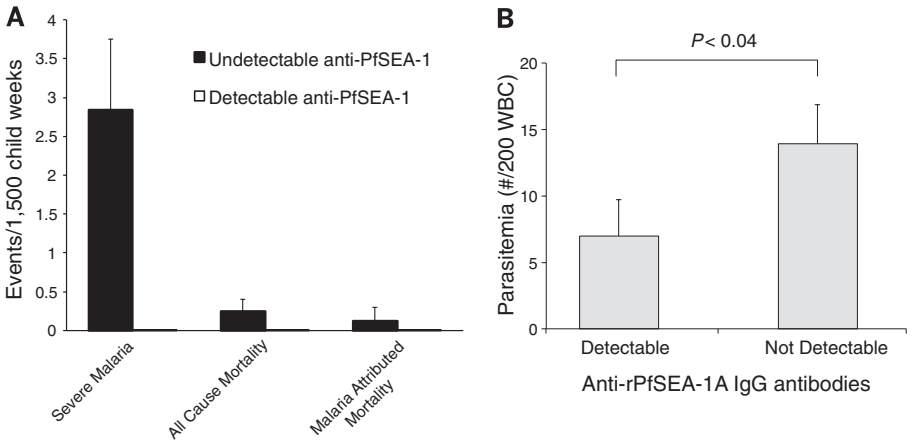


Fig. 6. Antibodies to rPfSEA-1A predict reduced malaria severity and parasitemia. (**A**) Incidence of severe malaria and death in Tanzanian children aged 1.5 to 3.5 years during intervals with detectable and undetectable antibodies to PfSEA-1 (1688 and 23,806 weeks, respectively). No cases of severe malaria or death occurred during intervals with detectable antibodies to rPfSEA-1A. Error bars represent 95% CI, adjusted for repeated measures. (**B**) IgG antibodies to rPfSEA-1A predict decreased *P. falciparum* parasitemia. We enrolled n = 138 Kenyan males aged 12 to 35 years at the start of a high-transmission season, treated them with quinine and doxycycline, and followed them for 18 weeks with weekly blood smears. We collected serum 2 weeks after treatment and measured IgG antibody levels to rPfSEA-1A. In GEE repeated-measures models, individuals with detectable IgG antibodies to rPfSEA-1A had 50% lower parasite densities over 18 weeks of followup as compared to individuals with no detectable IgG antibodies to rPfSEA-1A, after adjusting for potential confounders (P < 0.04). Columns depict least-square mean parasitemia; error bars depict SEM.

Summary and Conclusions

Falciparum malaria remains a leading cause of childhood mortality, and vaccines are urgently needed to attenuate this public health threat. We report the rational identification of novel vaccine candidates by identifying parasite proteins recognized by antibodies expressed by resistant, but not susceptible, children. We identified two previously unknown hypothetical genes as well as MSP-7, a known vaccine candidate (*9*–*12*). We have extensively characterized PfSEA-1, which localizes to the schizont/parasitophorous vacuole membrane, Maurer's clefts, and the inner leaflet of the RBC membrane in schizont-infected RBCs. PfSEA-1 is accessible to antibodies during late schizogeny and displays minimal sequence variation in the immunorelevant region identified by our differential screening experiments (amino acids 810 to 1083). Antibodies to PfSEA-1 significantly attenuate parasite growth by arresting schizont egress from infected RBCs. Concordant with these results, genetically destabilizing PfSEA-1 results in a dramatic 80% inhibition of parasite growth.

In five challenge experiments, rPbSEA-1A or antibodies to PbSEA-1A conferred marked protection against a lethal *P. berghei* ANKA challenge as evidenced by up to a 75% reduction in parasitemia 7 days after challenge. In all five experiments, by day 7 to 8 after challenge, control mice had high parasitemia with associated morbidity, whereas none of the vaccinated mice had high parasitemia or overt morbidity. In experiments with long-term follow-up, both active immunization with rPbSEA-1 and passive transfer of antisera to PbSEA-1 significantly reduced parasitemia and delayed mortality. We report protection from *P. berghei* ANKA by vaccination with a blood-stage antigen, and these data support our ongoing efforts to evaluate PfSEA-1 in the *Aotus* model of *P. falciparum*.

In our Tanzanian birth cohort, antibodies to PfSEA-1 were associated with significant protection from severe malaria, with no cases occurring

while children had detectable antibodies to PfSEA-1. These data suggest that anti–PfSEA-1 responses could augment responses to other targets, such as PfEMP-1, that may protect against severe malaria (*24*). Under conditions of natural exposure, only 6% of 1.5- to 3.5-year-old children in our cohort had detectable antibodies to PfSEA-1. This low natural prevalence suggests that adjuvanted vaccination with PfSEA-1 could have a marked impact on reducing severe malaria in young children.

In a second longitudinal Kenyan cohort, anti–PfSEA-1 antibodies were associated with significant protection against parasitemia in adolescents and young adults. Individuals with detectable IgG anti–rPfSEA-1A antibodies had 50% lower parasite densities over 18 weeks of follow-up compared with individuals with no detectable IgG anti-rPfSEA-1A antibodies. It was noteworthy that the prevalence of detectable anti–rPfSEA-1A antibodies in this age group (56%) was substantially higher than in our birth cohort, which suggests that natural exposure boosts anti–PfSEA-1 antibody levels.

Together, our data validate our field-to-lab-to-field-based strategy for the rational identification of vaccine candidates and support PfSEA-1 as a candidate for pediatric falciparum malaria. By blocking schizont egress, PfSEA-1 may synergize with other vaccines targeting hepatocyte (*25*) and RBC invasion (*26*).

REFERENCES AND NOTES

1. C. J. Murray *et al.*, *Lancet* **379**, 413–431 (2012).
2. WHO, *The World Malaria Report* (WHO, Switzerland, 2012); www.who.int/malaria/publications/world_malaria_report_2012/en/index.html.
3. S. Okie, *N. Engl. J. Med.* **353**, 1877–1881 (2005).
4. WHO, Malaria Vaccine Rainbow Tables (2012); www.who.int/vaccine_research/links/Rainbow/en/index.html.
5. A. N. Hodder, P. E. Crewther, R. F. Anders, *Infect. Immun.* **69**, 3286–3294 (2001).
6. A. Sabchareon *et al.*, *Am. J. Trop. Med. Hyg.* **45**, 297–308 (1991).
7. J. D. Kurtis, D. E. Lanar, M. Opollo, P. E. Duffy, *Infect. Immun.* **67**, 3424–3429 (1999).
8. T. K. Mutabingwa *et al.*, *PLOS Med.* **2**, e407 (2005).
9. R. Spaccapelo *et al.*, *Sci. Rep.* **1**, 39 (2011).
10. N. D. Gómez *et al.*, *PLOS ONE* **6**, e25477 (2011).
11. M. Kadekoppala, R. A. O'Donnell, M. Grainger, B. S. Crabb, A. A. Holder, *Eukaryot. Cell* **7**, 2123–2132 (2008).
12. C. P. Nixon *et al.*, *J. Infect. Dis.* **192**, 861–869 (2005).
13. M. Treeck, J. L. Sanders, J. E. Elias, J. C. Boothroyd, *Cell Host Microbe* **10**, 410–419 (2011).
14. M. Manske *et al.*, *Nature* **487**, 375–379 (2012).
15. J. D. Dvorin *et al.*, *Science* **328**, 910–912 (2010).
16. A. Farrell *et al.*, *Science* **335**, 218–221 (2012).
17. M. J. Blackman, H. G. Heidrich, S. Donachie, J. S. McBride, A. A. Holder, *J. Exp. Med.* **172**, 379–382 (1990).
18. E. S. Bergmann-Leitner, E. H. Duncan, E. Angov, *Malar. J.* **8**, 183 (2009).
19. N. Ahlborg *et al.*, *Exp. Parasitol.* **82**, 155–163 (1996).
20. I. D. Goodyer, B. Pouvelle, T. G. Schneider, D. P. Trelka, T. F. Taraschi, *Mol. Biochem. Parasitol.* **87**, 13–28 (1997).
21. S. Yoshida *et al.*, *PLOS ONE* **5**, e13727 (2010).
22. J. D. Kurtis, R. Mtalib, F. K. Onyango, P. E. Duffy, *Infect. Immun.* **69**, 123–128 (2001).
23. J. C. Beier *et al.*, *Am. J. Trop. Med. Hyg.* **50**, 529–536 (1994).
24. P. C. Bull *et al.*, *J. Infect. Dis.* **192**, 1119–1126 (2005).
25. RTS,S Clinical Trials Partnership, *N. Engl. J. Med.* **365**, 1863–1875 (2011).
26. C. Crosnier *et al.*, *Nature* **480**, 534–537 (2011).

ACKNOWLEDGMENTS

The authors thank the Mother Offspring Malaria Study project staff for their efforts in collecting clinical data, processing samples, and interpreting malaria blood smears, as well as the study participants and their families. Ethical clearance was obtained from the institutitional review boards of the Seattle Biomedical Research Institute and Rhode Island Hospital; the Medical Research Coordinating Committee of the National Institute for Medical Research, Tanzania; and the Kenyan Medical Research Institute. We thank A. Muehlenbachs for assistance in interpreting the electron microscopy images, B. Morrison for assistance with data management, and D. Lanar for the LSA proteins. This work was supported by grants from the U.S. National Institutes of Health (NIH) (grant R01-AI52059) and the Bill & Melinda Gates Foundation (grant 1364) to P.E.D., the Intramural Research Program of the National Institute of Allergy and Infectious Diseases–NIH, and grants from NIH (grant R01-AI076353) and an internal Lifespan Hospital System Research Pilot Award Grant to J.D.K., and grants from NIH to J.D.D. (R01-AI102907 and DP2-AI112219). C.E.N. (grant T32-DA013911) and I.C.M. (grant 1K08AI100997-01A1) were supported by NIH. We also acknowledge research core services provided by the Rhode Island Hospital imaging core (G. Hovanesian) and core services supported by Lifespan/Tufts/Brown Center for Aids Research (P30AI042853) and Center for Biomedical Research Excellence (P20GM103421). The DNA sequence for PfSEA-1 is available in PlasmoDB (www.Plasmodb.org) as ID no. PF3D7_1021800. The work presented in this manuscript has been submitted in partial support of patent application PCT/US12/67404 filed with the U.S. Patent Office. J.D.K., J.F.F., M.F., and P.E.D. designed the study; J.D.K., D.K.R., J.F.F., M.F., and P.E.D. drafted the manuscript; M.F. and P.E.D. conducted the field-based sample and epidemiologic data collection; C.P.N. conducted the differential screening; D.K.R, L.P., A.L., C.E.N., and L.C. performed the in vitro growth assays; D.K.R. and G.J. performed the immunolocalization; D.K.R. performed the gene knockout studies; C.G.D., E.A.M., S.A., and J.D.D. performed the conditional gene knockdown studies; I.C.M. performed the single-nucleotide polymorphism detection; S.P.-T., L.C., and H.-W.W. performed the antibody detection assays; S.P.-T. expressed and purified rPfSEA-1A and rPbSEA-1A; D.K.R., C.E.N., and S.C. performed the vaccination studies; J.F.F., J.D.K., and S.E.H. performed the statistical analyses; and J.D.K., J.F.F., M.F., and P.E.D. supervised the study. The authors declare no competing financial interests.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/334/6186/871/suppl/DC1
Materials and Methods
Supplementary Text
Tables S1 and S2
Figs. S1 to S11
References (*27*–*41*)

30 October 2013; accepted 25 April 2014
10.1126/science.1254417

DEEP EARTH

Disproportionation of $(Mg,Fe)SiO_3$ perovskite in Earth's deep lower mantle

Li Zhang,[1,2]* Yue Meng,[3] Wenge Yang,[1,4] Lin Wang,[1,4] Wendy L. Mao,[5,6] Qiao-Shi Zeng,[5] Jong Seok Jeong,[7] Andrew J. Wagner,[7] K. Andre Mkhoyan,[7] Wenjun Liu,[8] Ruqing Xu,[8] Ho-kwang Mao[1,2]

The mineralogical constitution of the Earth's mantle dictates the geophysical and geochemical properties of this region. Previous models of a perovskite-dominant lower mantle have been built on the assumption that the entire lower mantle down to the top of the D″ layer contains ferromagnesian silicate [$(Mg,Fe)SiO_3$] with nominally 10 mole percent Fe. On the basis of experiments in laser-heated diamond anvil cells, at pressures of 95 to 101 gigapascals and temperatures of 2200 to 2400 kelvin, we found that such perovskite is unstable; it loses its Fe and disproportionates to a nearly Fe-free $MgSiO_3$ perovskite phase and an Fe-rich phase with a hexagonal structure. This observation has implications for enigmatic seismic features beyond ~2000 kilometers depth and suggests that the lower mantle may contain previously unidentified major phases.

The conventional view of materials in Earth's lower mantle, which comprises the largest fraction of our planet (>55% by volume), has evolved greatly in the past 10 years. High-pressure-temperature (*P-T*) experiments of the past century suggested that the predominant phase was ferromagnesian silicate [$(Mg,Fe)SiO_3$] with the orthorhombic perovskite (pv) structure, unchanging over the enormous *P-T* range from the bottom of the transition zone at 670 km depth (24 GPa) to the core-mantle boundary (CMB) at 2900 km (136 GPa). However, laser-heated diamond anvil cell (DAC) technology coupled with x-ray spectroscopy (XRS) and x-ray diffraction (XRD) probes have led to two key discoveries: namely, the phase transition of $(Mg,Fe)SiO_3$ pv to the post-perovskite (ppv) structure (*1*–*3*) near the D″ layer of the lower mantle, and the pressure-induced spin-pairing of the *3d* electrons in Fe (*4*–*6*)

[1]Center for High Pressure Science and Technology Advanced Research (HPSTAR), Shanghai 201203, China. [2]Geophysical Laboratory, Carnegie Institution of Washington (CIW), Washington, DC 20015, USA. [3]High Pressure Collaborative Access Team (HPCAT), Geophysical Laboratory, CIW, Argonne, IL 60439, USA. [4]High Pressure Synergetic Consortium (HPSynC), Geophysical Laboratory, CIW, Argonne, IL 60439, USA. [5]Geological and Environmental Sciences, Stanford University, Stanford, CA 94305, USA. [6]Photon Science, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA. [7]Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455 USA. [8]Advanced Photon Source, Argonne National Laboratory, Argonne, IL 60439, USA.
***Corresponding author. E-mail: zhangli@hpstar.ac.cn**

that dramatically alters the chemical and physical behavior of the Fe-bearing silicates and oxides (*7*). These discoveries have triggered a flurry of new theories for the enigmatic seismic and geodynamic features observed in the deep lower mantle, such as the geothermal and seismic signatures of the spin transition in Fe-bearing minerals, the correlations between the pv-ppv *P-T* boundary and the D″ discontinuity and velocity gradients, and possible mineralogical mechanism for the existence of large low-shear-velocity provinces (LLSVPs) (*1*, *5*, *8*–*11*).

Recent sound-velocity measurements of silicate pv and (Mg,Fe)O ferropericlase (fp) suggested that the (Mg,Fe)SiO_3 pv–dominant petrological model (*12*) would be consistent with the density profile of the lower mantle (*13*). The $MgSiO_3$-$FeSiO_3$ pseudo-binary phase relations, however, have only been well investigated at the *P-T* conditions corresponding to the top half of the lower mantle (*14*, *15*). Beyond 60 GPa, detailed studies of this phase diagram are still scarce and reconnaissance in nature (*16*–*19*).

Disproportionation of (Mg,Fe)SiO_3

To more directly interrogate these phase relations, we brought crystalline $(Mg_{0.85}Fe_{0.15})SiO_3$ (Fs15) orthopyroxene (opx) sample to the pressures (60 to 110 GPa) and temperatures (2000 to 2400 K) of the lower mantle using DAC technology (*20*). We monitored the growth of crystalline phases with in situ XRD under high *P-T* laser-heating conditions as well as after *T*-quench at high *P*.

Compressing Fs15 to 95 GPa at 300 K, the opx powder XRD peaks broaden substantially, but the sample was not amorphized. Instead, the broad peaks resemble the major peaks of pv (fig. S4), indicating a nanocrystalline pv-like structure. The pattern sharpens with moderate heating at 1500 to 2000 K and forms normal crystalline pv, with unit cell parameters consistent with the 15% Fs composition (fig. S4). Increasing temperature to 2300 K above 95 GPa (fig. S4) or heating directly to this *P-T* condition (Fig. 1A), the pv-dominated powder XRD pattern exhibited subtle additional features. At 101 GPa, the pv unit cell of a = 4.362(1) Å, b = 4.611(4) Å, c = 6.343(3) Å, and molecular volume V = 31.89(3) Å^3 (Z = 4) is too small for that expected of $(Mg_{0.85}Fe_{0.15})SiO_3$ but close to that of the pure endmember $MgSiO_3$ pv at the same pressure (*21*). In addition to pv, another set of peaks not corresponding to any previously known phases appeared with particularly conspicuous peaks at 2.55 and 2.40 Å (marked H110 and H101, respectively). These two peaks are close to the diagnostic ppv peaks near 2.5 and 2.4 Å (ppv 022 and 110, respectively), but the high-quality XRD pattern clearly rejects the possibility of ppv. On the basis of three distinct new peaks at 4.413, 2.549, and 2.401 Å and several additional peaks (Tables 1 and 2), we concluded that this phase has a hexagonal unit cell of a = 5.096(2) Å and c = 2.862(3) Å and a molecular volume V = 32.19(4) Å^3 with formula units per unit cell Z = 2 (Tables 1 and 2), referred to as the "H-phase." The slightly larger molar volume of the H-phase as compared with the

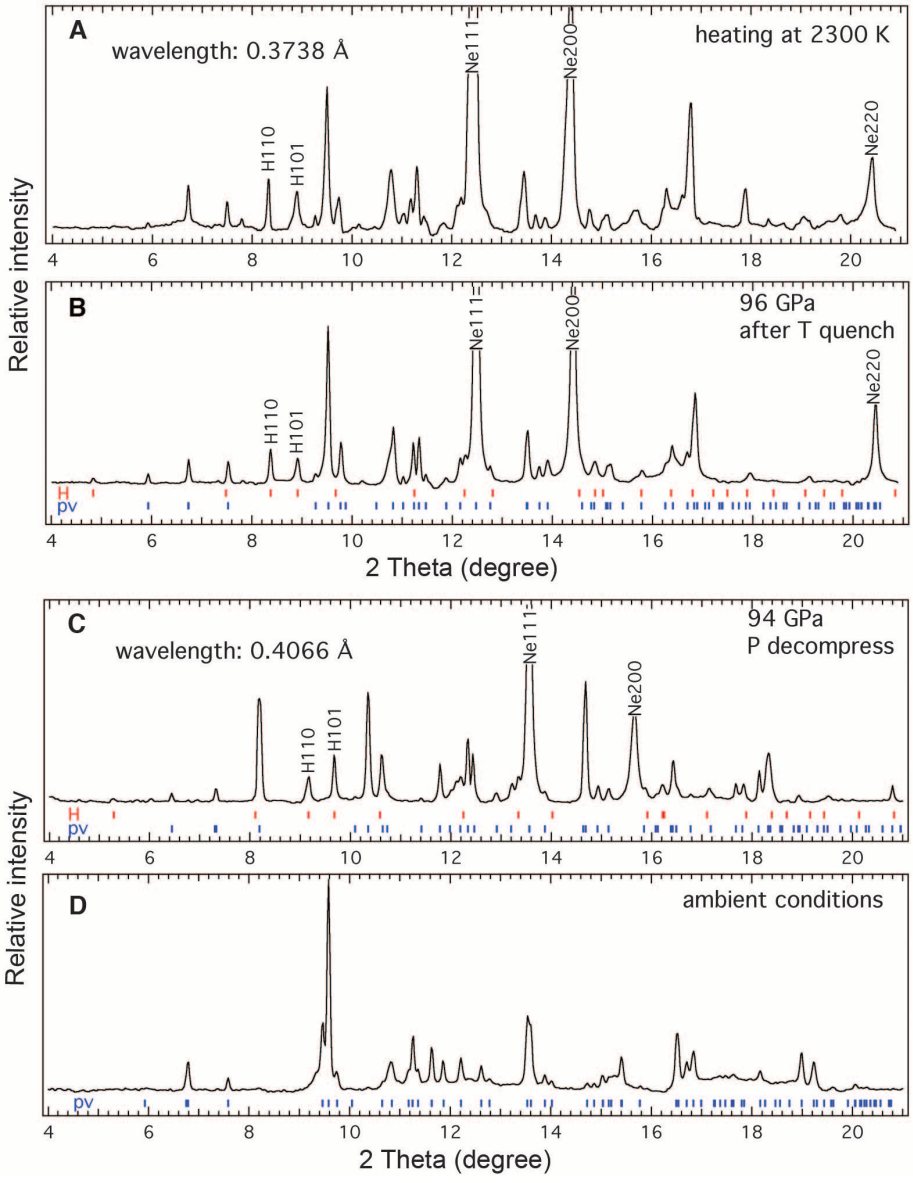


Fig. 1. In situ powder XRD patterns of Fs15 in Ne medium. (**A** and **B**) Coexistence of pv phase and H-phase (A) during laser-heating at 2300 K and (B) after T quench to 300 K and 96 GPa. (**C** and **D**) Another coexisting pv phase and H-phase decompressed from 101 GPa (C) to 94 GPa at 300 K and then (D) to ambient conditions. Because of the multigrain spotty nature of the patterns, the relative peak intensities are affected by the presence or absence of strong spots. The peak at d ~ 2.7 Å, which is 7.8° in (B), belongs to a small amount of SiO_2 ($CaCl_2$ structure) added to the starting material (*20*).

Table 1. Powder XRD data for the H-phase in Fs15 starting composition at 101 GPa. The unit-cell parameters of the H-phase are a = 5.096(2) Å, c = 2.862(3) Å; α = β = 90°, γ = 120° at 101 GPa. The powder XRD pattern was obtained with x-ray wavelength of 0.5164 Å. *h*, *k*, *l* represent Miller indices of reflections; 2θ, Bragg angle; *d*, *d*-spacing; obs, observed; calc, calculated; diff, difference.

h	*k*	*l*	2θ-obs (degrees)	*d*-obs (Å)	*d*-calc (Å)	*d*-diff (Å)
1	0	0	6.708	4.4133	4.4133	0.0000
1	1	0	11.629	2.5487	2.5480	−0.0007
1	0	1	12.346	2.4012	2.4013	0.0001
1	1	1	15.591	1.9036	1.9031	−0.0005
3	0	0	20.218	1.4710	1.4711	0.0001
2	1	1	20.645	1.4409	1.4411	0.0002

Table 2. Single-crystal XRD data for the H-phase in Fs15 starting composition at 101 GPa. The unit-cell parameters of the H-phase are a = 5.096(2) Å, c = 2.862(3) Å; α = β = 90°, γ = 120° at 101 GPa. The single-crystal XRD data of one of the crystallites was obtained for the H-phase with an x-ray wavelength of 0.3738 Å. The Bragg angle 2θ, rotation angle ω, and azimuthal angle η are calculated from the orientation matrix and observed as XRD spots.

h	*k*	*l*	2θ (degrees)	ω (degrees)	η (degrees)	*d* (Å)
−1	0	1	8.93	−1.3	63.4	2.395
−1	1	1	8.93	−16.2	32.8	2.395
0	1	−1	8.93	−11.4	268.8	2.401
1	−1	−1	8.95	0.3	212.8	2.399
1	0	−1	8.95	8.7	243.2	2.402
−1	−1	1	11.29	3.0	87.4	1.897
1	1	−1	11.27	14.3	267.3	1.902
−2	2	1	12.3	12.1	22.1	1.741
0	−2	1	12.27	−12.4	105.9	1.746
0	2	−1	12.25	0.3	285.7	1.751
−1	−2	1	14.92	3.9	101.7	1.436
1	0	−2	15.82	−4.1	240.6	1.358
−2	1	2	17.28	−12.4	47.6	1.242
−1	−1	2	17.29	−15.6	75.8	1.242
1	1	−2	17.22	2.3	255.7	1.248
2	−1	−2	17.25	11.0	227.4	1.245
−2	0	2	17.97	−6.5	63.1	1.194
2	−2	−2	17.95	8.6	211.9	1.198
2	0	−2	17.93	13.7	242.9	1.199
−1	4	−1	19.04	0.5	304.4	1.129
−1	−2	2	19.89	−12.1	88.0	1.082
1	2	−2	19.79	7.8	267.9	1.087
2	1	−2	19.81	16.8	256.4	1.086
−3	2	2	19.88	−3.1	38.2	1.081
0	−4	1	20.94	−8.4	120.3	1.029
0	4	−1	20.82	15.6	300.1	1.033
0	3	−2	20.92	0.4	278.7	1.030

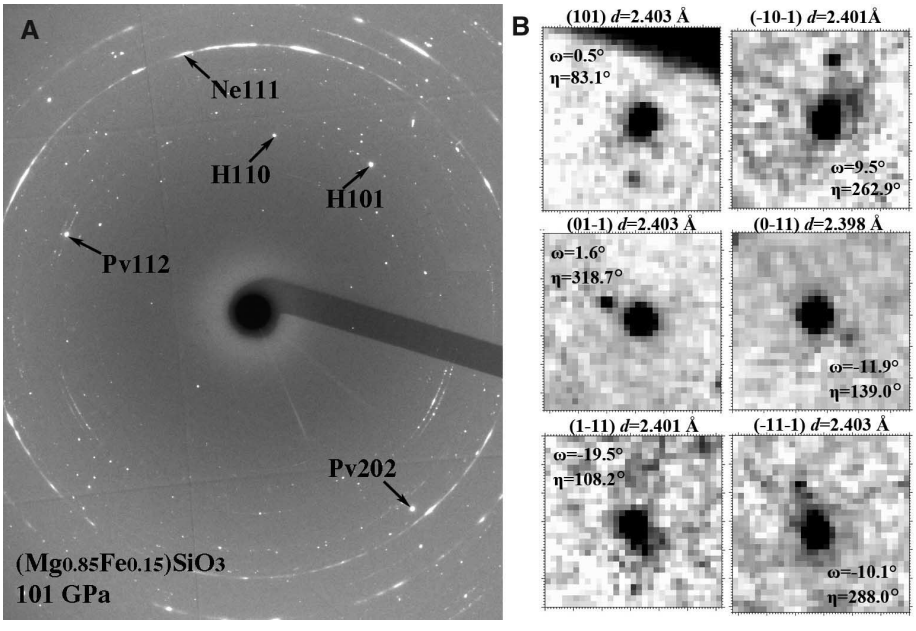


Fig. 2. Selection of $MgSiO_3$ pv and H-phase individual single crystals using the multigrain method. (**A**) A representative spotty multiple-crystal XRD pattern of the H-phase and coexisting pv in Fs15 at 101 GPa and after T quench (x-ray wavelength of 0.3738 Å), and the marked Miller indices show some of the most intense diffraction spots. (**B**) Six {101} diffraction reflections at different rotation angle ω and azimuthal angle η belonging to one selected crystallite of the H-phase at 101 GPa.

coexisting $MgSiO_3$ pv suggests an $(Mg,Fe)SiO_3$ composition with Fe enrichment.

We repeated the Fs15 experiments with Ne pressure medium five times over the same *P-T* region and were able to observe the disproportionation reaction and the formation of the H-phase repeatedly (Fig. 1 and fig. S4). We found that the H-phase grows more readily in the Ne medium, which provides a quasi-hydrostatic environment that is superior to that of other media. To eliminate the possibility of any experimental artifacts introduced by Ne, we have also successfully repeated the observation of the disproportionation and appearance of the H-phase with SiO_2 glass as the pressure media and insulation layers without Ne.

Symmetry and Unit Cell Parameters of H-phase

In the two-phase mixed XRD pattern, the low-symmetry orthorhombic pv has many peaks interfering with the H-phase peaks at the low *d*-spacing region; only three peaks above 2.4 Å are specific to the H-phase. Moreover, powder XRD patterns only contain *d*-spacing information, and successful indexing of multiple peaks could still be inconclusive. The spottiness of the XRD powder pattern may also introduce uncertainty in peak intensity (fig. S8). On the other hand, single-crystal XRD contains three-dimensional orientation and geometrical relationship in addition to *d*-spacings and thus provides a more definitive characterization of the unit cell.

We used a multigrain crystallography method (*22*) to search and determine the orientation matrices of multiple individual crystallites simultaneously within the sample (*20*). We conducted prolonged laser-heating that promoted growth of a coarser-grained multiple-crystal aggregate, including $MgSiO_3$ pv and H-phase crystals at 101 GPa (Fig. 2A). By rotating the ω (vertical) axis of the DAC holder to ±21° at the steps of 0.1°, 154 individual crystallites—each with 11 to 27 reflections consistent with the H-phase—were identified within the x-ray beam spot of 6 by 8 μm (Tables 1 and 2 and table S1). Six {101} diffraction reflections belonging to three Friedel's pairs from another crystallite further confirm these results (Fig. 2B). With the stringent requirement that all reflections from each of the 154 crystallites must satisfy the *d*-spacing, rotation angle ω, and azimuthal angle η predicted by its particular orientation matrix, we confirmed the symmetry and the basic unit cell of the H-phase. For comparison in the same sample at 101 GPa using the same indexing procedure (*20*), we also identified and indexed 130 individual crystallites of the $MgSiO_3$ pv phase to the known *Pbnm* space group, with the maximum of 48 reflections observed in a crystallite.

To verify that all predicted reflections were indeed from the same single crystallite and not fortuitously from the diffraction of another crystallite within the 6 by 8 μm x-ray spot, we were able to cover only about 10 crystallites by using the 0.3 by 0.5 μm focused polychromatic and monochromatic x-ray beam (*23*). The polychromatic Laue XRD spots of different Miller indices appear and disappear

together, indicating that they originate from the same single crystallite. We mapped the relatively large crystallites by using the x-ray beam, monitored the diffraction intensity, and observed that the intensities of the diffraction spots dropped ~50% after moving 0.5 μm in each direction, thus confirming the crystallite size to be typically ~0.5 μm or less (fig. S10).

Fe Depletion in Perovskite

All powder XRD peaks observed at 101 GPa could be assigned to the $MgSiO_3$ pv phase and H-phase. To confirm that the pv is indeed the $MgSiO_3$ endmember, we decompressed and monitored the disproportionation product with in situ XRD. The pv phase remains crystalline, with unit cell consistent with the $MgSiO_3$ pv all the way to ambient pressure (Fig. 1D), and all peaks in the powder XRD pattern can be assigned to the $MgSiO_3$ pv phase with unit cell parameters a = 4.784(2) Å, b = 4.930(2) Å, c = 6.903(4) Å, and V = 162.8(1) $Å^3$, which is in agreement with the established ambient pv values with x = 0.02(2).

We also performed scanning transmission electron microscopy (STEM) analysis on the recovered sample for phase identification and compositional analysis (*20*). First, we prepared a thin specimen with focused ion beam (FIB) and mapped it using energy-dispersive x-ray (EDX) to locate the $MgSiO_3$ region (Fig. 3). The STEM-EDX compositional mapping indicates four regions of distinct compositions marked as positions 1 to 4 on the TEM–bright field (BF) image (Fig. 3A). The selected area electron diffraction (SAED) in position 1 shows the quenched crystalline pv structure (Fig. 3G), and the chemical analysis verified x = 0.02(2) based on multiple analyses (fig. S6); therefore, it can be nominally referred to as $MgSiO_3$. The STEM analysis shows that the composition and structure of the quenched pv match the XRD data of the same pv quenched to ambient conditions, confirming the Fe depletion in the pv as a result of the disproportionation. The overall quantity of SiO_2 in the recovered sample (position 2) is consistent with the excess SiO_2 in the starting composition and the in situ XRD observations (Fig. 3H and fig. S6). In contrast to the preservation of the crystalline $MgSiO_3$ pv, the H-phase was unquenchable and amorphized during decompression; its XRD peaks completely disappeared (Fig. 1D). The amorphous product was highly unstable under ambient conditions and further decomposed during the sample preparation process for STEM (positions 3 and 4). The Fe enrichment in H-phase was confirmed with EDX at ambient conditions (fig. S6), but its unit cell parameters and composition must be characterized in situ at high P. We also tested the reversibility of the disproportionation in the Ne medium by performing heating during decompression and observed that the H-phase clearly diminishes, and only the pv peaks remained at 67 GPa and 2000 K (fig. S7).

Geophysical Implications

The XRD pattern of the disproportionation products only contains the $MgSiO_3$ pv and H-phase, indicating that the H-phase has the opx stoichiometry and Fe enrichment (x > 0.15), as in the following reaction:

$$(Mg_{0.85}Fe_{0.15})SiO_3\ \text{pv} \Rightarrow a MgSiO_3\ \text{pv} + (1-a)(Mg_{1-x}Fe_x)SiO_3\ \text{H-phase} \quad (1)$$

According to the phase rule, at a given P-T condition compositions of both $MgSiO_3$ pv and the H-phase are fixed in the pseudo-binary system. The best approach to determine the H-phase composition x, therefore, would be to synthesize a pure H-phase in situ. We tested another opx starting

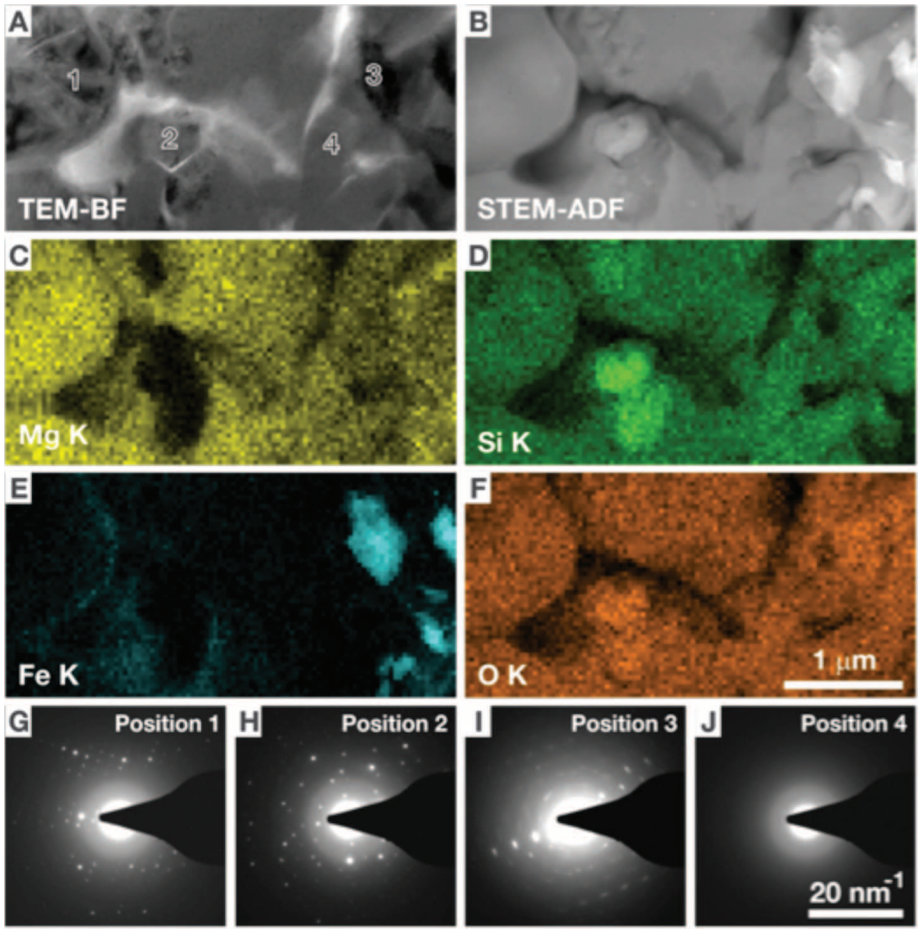


Fig. 3. STEM-EDX analysis of the recovered products. (**A**) TEM-BF image. (**B**) STEM-annular dark-field (ADF) image. (**C** to **F**) Corresponding STEM-EDX maps: (C) Mg K, (D) Si K, (E) Fe K, and (F) O K. STEM-ADF and EDX maps were obtained simultaneously. (**G** to **J**) SAED patterns obtained from the positions indicated in (A).

Table 3. Powder XRD pattern for the H-phase at 85 GPa. The H phase was synthesized in a starting composition $(Mg_{0.60}Fe_{0.40})SiO_3$ (Fs40) and has unit-cell parameters of a = 5.136(6) Å, c = 2.883(3) Å (x-ray wavelength: 0.3978 Å).

h	*k*	*l*	2θ-obs (degrees)	*d*-obs (Å)	*d*-calc (Å)	*d*-diff (Å)
1	0	0	5.124	4.4495	4.4479	−0.0016
1	1	0	8.890	2.5663	2.5680	0.0017
1	0	1	9.434	2.4186	2.4193	0.0007
2	0	0	10.250	2.2266	2.2240	−0.0026
1	1	1	11.934	1.9133	1.9176	0.0043
2	0	1	12.977	1.7601	1.7609	0.0008
2	1	0	13.606	1.6791	1.6812	0.0021
3	0	0	15.413	1.4832	1.4826	−0.0006
2	1	1	15.756	1.4511	1.4523	0.0012
0	0	2	15.863	1.4414	1.4415	0.0001
1	0	2	16.661	1.3728	1.3713	−0.0015
3	0	1	17.340	1.3194	1.3185	−0.0009
2	2	0	17.809	1.2850	1.2840	−0.0010
2	2	1	19.520	1.1733	1.1729	−0.0004

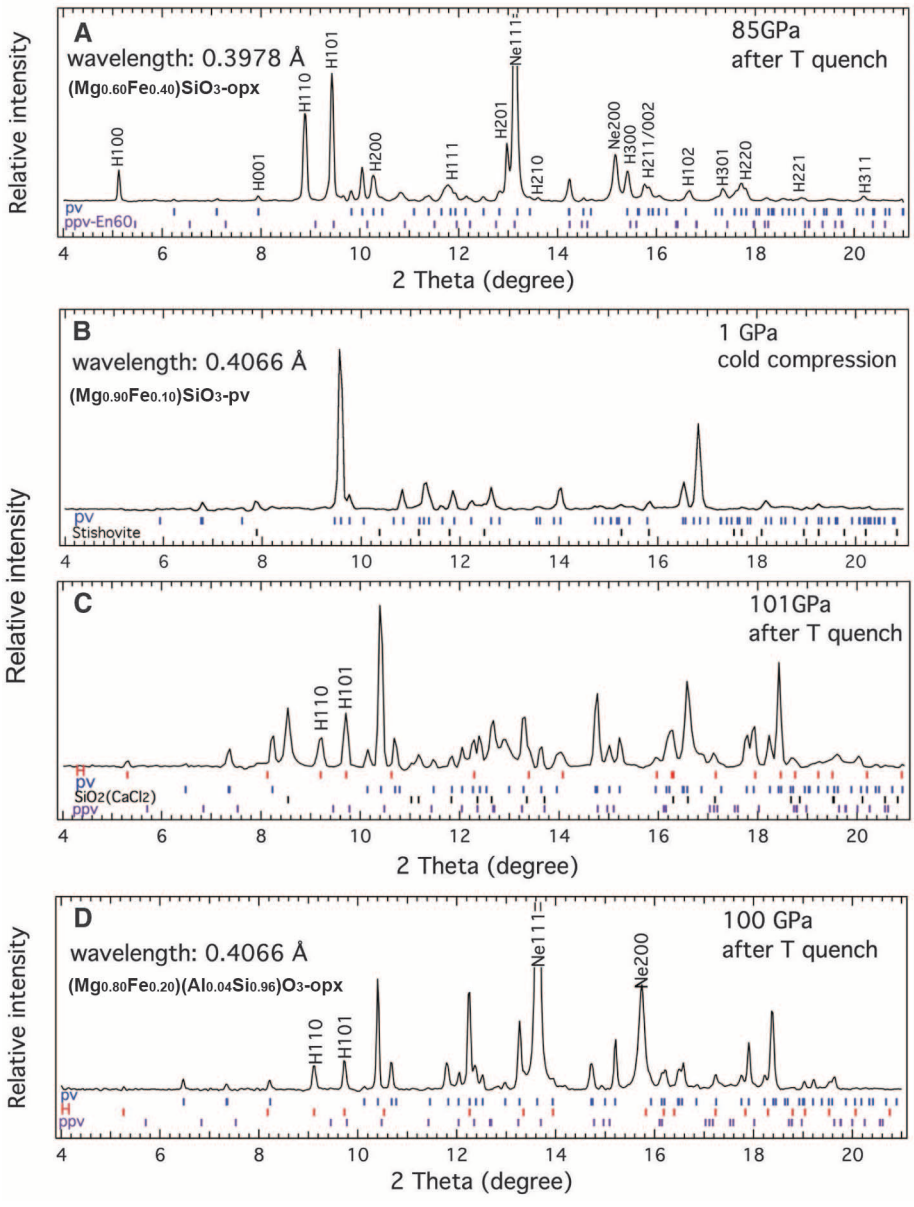


Fig. 4. XRD powder patterns of run products from different starting compositions. (**A**) H-phase–dominant XRD pattern in the Fs40 composition at 85 GPa after *T*-quench to 300 K (Table 3). (**B**) Cold compression of pv10 starting material in SiO_2 glass medium and (**C**) XRD pattern of coexisting pv phase and H-phase after *T*-quench from heating at 2350 to 300 K at 101 GPa. (**D**) XRD pattern of disproportionation products of $(Mg_{0.80}Fe_{0.20})(Al_{0.04}Si_{0.96})O_3$ opx at 100 GPa after *T*-quench to 300 K from heating at 2300 K. The calculated peak positions of the H, pv, and SiO_2 ($CaCl_2$ structure) phases are indicated by small ticks, and the calculated peak positions of the ppv phase at the same *P-T* conditions are also shown for comparison.

material with the composition of $(Mg_{0.60}Fe_{0.40})SiO_3$ (Fs40), compressed it to 60 GPa, and laser heated it to 2100 K. Consistent with the previously known phase relations (*15*), all observed XRD peaks can be assigned to three coexisting phases: SiO_2, (Mg,Fe)Ofp, and (Mg,Fe)SiO_3 pv; no H-phase was observed. When the Fs40 was directly compressed to 85 GPa and heated to 2300 K, however, we observed a powder XRD pattern consisting of predominately H-phase, with only a minor amount of pv peaks (Fig. 4A), indicating that the H-phase composition is close to $(Mg_{0.6}Fe_{0.4})SiO_3$. Refinement of 15 powder XRD peaks of the H-phase yields a hexagonal unit cell of a = 5.136(6) Å, c = 2.883(3) Å, and V = 32.93(5) Å^3 (Z = 2) for the H-phase at 85 GPa (Table 3). The volume reduction of the reaction based on the molar volumes of $MgSiO_3$ pv and the H-phase on the right hand side and the pv on left side (*21*) was obtained as ΔV = −0.27 Å^3 and $\Delta V/V$ = −0.2% for the $(Mg_{0.6}Fe_{0.4})SiO_3$ H-phase. The small, negative ΔV indicates that pressure slightly favors the disproportionation, thermodynamically.

Equation 1 implies that any pv in the composition range of x = 0.02 to 0.40 will disproportionate to $MgSiO_3$ pv and H-phase at the deep lower mantle conditions. To test this conclusion, we used an $(Mg_{0.90}Fe_{0.10})SiO_3$ sample presynthesized in the pv form (pv10) with a multianvil press (Fig. 4B). The pv10 starting material is from the same batch that has been used extensively in previous studies and publications (*6*, *14*, *24*, *25*) and is considered a realistic mineral component for the Earth's interior. Again, we observed the disproportionation to an assemblage of Fe-depleted pv and H-phase (Fig. 4C). The direct transition from the stable phase of pv, rather than from a metastable opx starting material, demonstrates the relative stability. Because we can convert the Fe-bearing pv to coexisting Fe-depleted pv and H-phase, but have not been able to convert H-phase to Fe-bearing pv in this *P-T* range even with extensive heating, we conclude that the disproportionation assemblage is more stable than is Fe-bearing pv.

Most Earth models include excess fp in addition to the major mineral of pv. We conducted an experiment with the starting composition of $MgO:FeO:SiO_2$ = 0.86:0.19:1.0 (pv:fp = 100:5 or opx:olivine = 95:5). The H-phase was again observed at 97 GPa after laser-heating in the fp-saturated environment, confirming the disproportionation of pv in the MgO-FeO-SiO_2 ternary system (fig. S11). We also consider the trivalent minor components of Al and Fe^{3+}, which are coupled according to previous studies (*26*). We performed another set of experiments using an Al-bearing opx starting material $(Mg_{0.80}Fe_{0.20})(Al_{0.04}Si_{0.96})O_3$. Again, we observed the robust disproportionation of pv and formation of H-phase (Fig. 4D). The single-crystal data set for one of the H-phase crystallites in the Al-bearing sample confirms these results (table S2).

Both in situ XRD and ex situ STEM-EDX analysis demonstrate the disproportionation reaction of (Mg,Fe)SiO_3 pv to a nearly Fe-free pv and an Fe-rich H-phase at 95 to 101 GPa and 2200 to 2400 K. Therefore, we suggest that (Mg,Fe)SiO_3 pv may not be the major silicate throughout the lower mantle down to the top of the D″ layer. For the lower mantle below 2000 km, all geodynamic models need to consider the physical properties of the Fe-depleted $MgSiO_3$ pv as a major phase. For instance, the sound-velocity data of $MgSiO_3$ and $MgSiO_3$:Al_2O_3 pv (*12*) can now be applied directly for comparison with seismic data without adjustment for its Fe content. The geochemical importance of high-spin–low-spin transition (*4*–*6*) in (Mg,Fe)SiO_3 pv gives way to other Fe-bearing phases, such as oxides or H-phase, in this region. Likewise, interpretations for key enigmatic features in the deep lower mantle—namely, the sharp boundary on the top of the D″ layer, the anticorrelations between V_S and V_P, the seismic anisotropy, and the LLSVPs in the bottom third of the lower mantle (*8*–*11*, *27*, *28*)—may not be limited to the effects of the pv-ppv transition (*1*–*3*) but should consider the rich three-phase *P-T-x* relationship of pv, ppv, and the H-phase.

We still do not fully understand the detailed crystal chemistry of the H-phase. For instance,

the atomic positions have not been determined within its unit cell, and the small unit cell that we provided could well be the base for larger superlattices (*20*). Its composition and phase relation to pv and ppv needs to be determined as a function of *P-T* for interpreting various features in the deep lower mantle. Last, its elasticity and rheological properties are essential for interpretation of deep Earth seismology.

REFERENCES AND NOTES

1. M. Murakami, K. Hirose, K. Kawamura, N. Sata, Y. Ohishi, *Science* **304**, 855–858 (2004).
2. A. R. Oganov, S. Ono, *Nature* **430**, 445–448 (2004).
3. T. Tsuchiya, J. Tsuchiya, K. Umemoto, R. M. Wentzcovitch, *Earth Planet. Sci. Lett.* **224**, 241–248 (2004).
4. J. Badro *et al.*, *Science* **305**, 383–386 (2004).
5. J. Badro *et al.*, *Science* **300**, 789–791 (2003).
6. J. Li *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 14027 (2004).
7. S. Stackhouse, *Nat. Geosci.* **1**, 648–650 (2008).
8. E. Ohtani, T. Sakai, *Phys. Earth Planet. Inter.* **170**, 240–247 (2008).
9. R. D. van der Hilst, H. Karason, *Science* **283**, 1885–1888 (1999).
10. E. J. Garnero, A. K. McNamara, *Science* **320**, 626–628 (2008).
11. W. J. Su, R. L. Woodward, A. M. Dziewonski, *J. Geophys. Res.* **99**, 6945 (1994).
12. M. Murakami, Y. Ohishi, N. Hirao, K. Hirose, *Nature* **485**, 90–94 (2012).
13. A. Ricolleau *et al.*, *Geophys. Res. Lett.* **36**, L06302 (2009).
14. Y. Fei, Y. Wang, L. W. Finger, *J. Geophys. Res.* **101**, 11525 (1996).
15. Y. Tange, E. Takahashi, Y. Nishihara, K.-i. Funakoshi, N. Sata, *J. Geophys. Res.* **114**, B02214 (2009).
16. W. L. Mao *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9751–9753 (2005).
17. S. Tateno, K. Hirose, N. Sata, Y. Ohishi, *Phys. Earth Planet. Inter.* **160**, 319–325 (2007).
18. J.-F. Lin *et al.*, *Nat. Geosci.* **1**, 688–691 (2008).
19. S. M. Dorfman, Y. Meng, V. B. Prakapenka, T. S. Duffy, *Earth Planet. Sci. Lett.* **361**, 249–257 (2013).
20. Materials and methods are available as supplementary materials on *Science* Online.
21. S. Lundin *et al.*, *Phys. Earth Planet. Inter.* **168**, 97–102 (2008).
22. H. O. Sørensen *et al.*, *Z. Kristallogr.* **227**, 63–78 (2012).
23. L. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6140–6145 (2010).
24. Y. Fei, D. Virgo, B. O. Mysen, Y. Wang, H. K. Mao, *Am. Mineral.* **79**, 826 (1994).
25. J. M. Jackson *et al.*, *Am. Mineral.* **90**, 199–205 (2005).
26. R. Sinmyo, K. Hirose, S. Muto, Y. Ohishi, A. Yasuhara, *J. Geophys. Res.* **116**, B07205 (2011).
27. A. R. Hutko, T. Lay, J. Revenaugh, E. J. Garnero, *Science* **320**, 1070–1074 (2008).
28. M. Panning, B. Romanowicz, *Science* **303**, 351–353 (2004).

ACKNOWLEDGMENTS

We thank Y. Fei for providing the pv10 starting material; J. Shu, M. Somayazulu, E. Rod, G. Shen, S. Sinogeikin, P. Dera, V. Prakapenka, and S. Tkachev for their technical support. We also thank S. Merkel for introducing us to the indexing software. The research is supported by National Science Foundation (NSF) grants EAR-0911492, EAR-1119504, EAR-1141929, and EAR-1345112. This work was performed at HPCAT (Sector 16), Advanced Photon Source (APS), Argonne National Laboratory. HPCAT operations are supported by the U.S. Department of Energy–National Nuclear Security Administration (DOE-NNSA) under award DE-NA0001974 and DOE–Basic Energy Sciences (BES) under award DE-FG02-99ER45775, with partial instrumentation funding by NSF. HPSynC is supported as part of EFree, an Energy Frontier Research Center funded by DOE-BES under grant DE-SC0001057. Portions of this work were performed at GeoSoilEnviroCARS (sector 13), APS, supported by the NSF-Earth Sciences (EAR-1128799) and DOE-GeoSciences (DE-FG02-94ER14466), at 34ID-E beamline, APS, and at 15U1, Shanghai Synchrotron Radiation Facility. Use of the APS facility was supported by DOE-BES under contract DE-AC02-06CH11357. This work was also partially supported by the Materials Research and Engineering Center program of the NSF under award DMR-0819885. Part of this work was carried out in the Characterization Facility of the University of Minnesota. All other data used to support conclusions in this manuscript are provided in the supplementary materials

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/877/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 and S2
References (*29*–*34*)

30 December 2013; accepted 3 April 2014
10.1126/science.1250274

REPORTS

COSMOLOGY

Holographic description of a quantum black hole on a computer

Masanori Hanada,[1,2,3]* Yoshifumi Hyakutake,[4] Goro Ishiki,[1] Jun Nishimura[5,6]

Black holes have been predicted to radiate particles and eventually evaporate, which has led to the information loss paradox and implies that the fundamental laws of quantum mechanics may be violated. Superstring theory, a consistent theory of quantum gravity, provides a possible solution to the paradox if evaporating black holes can actually be described in terms of standard quantum mechanical systems, as conjectured from the theory. Here, we test this conjecture by calculating the mass of a black hole in the corresponding quantum mechanical system numerically. Our results agree well with the prediction from gravity theory, including the leading quantum gravity correction. Our ability to simulate black holes offers the potential to further explore the yet mysterious nature of quantum gravity through well-established quantum mechanics.

In 1974, it was realized that a black hole should radiate particles as a perfect black body due to quantum effects in the surrounding space and that the black hole should eventually evaporate completely (*1*, *2*). This phenomenon, which is known as the Hawking radiation, made more accurate the close analogy between the laws of black hole physics and those of thermodynamics (*3*), but it also caused a long scientific debate (*4*, *5*) concerning the information-loss paradox (*6*, *7*). Suppose one throws a book into a black hole. While the black hole evaporates, all we observe is the blackbody radiation. Therefore, the information contained in the book is lost forever. This statement sharply conflicts with a basic consequence of the law of quantum mechanics that the information of the initial state should never disappear. Then the question is whether the law of quantum mechanics is violated or the above argument should somehow be modified if full quantum effects of gravity are taken into account.

Superstring theory, a consistent theory of quantum gravity, enabled us to understand the statistical-mechanical origin of the black hole entropy for a special class of stable black holes by counting the microscopic states, which look like the same black hole from a distant observer (*8*). The paradox still remains, however, because a complete description of an evaporating black hole is yet to be established. A key to resolve the paradox is provided by the gauge-gravity duality (*9*), which is conjectured in superstring theory. In a particular case, it claims that black holes in gravity theory can be described in terms of gauge theory in one dimension, which is a standard quantum mechanical system without gravity. The conjecture may be viewed as a concrete realization of the holographic principle (*10*, *11*), which states that all of the information inside a black hole should be somehow encoded on its boundary, namely on the so-called event horizon of the black hole. In gauge theory in general, information is conserved during the time evolution as the standard laws of quantum mechanics simply apply. Therefore, it is widely believed that there is no information loss if the conjectured holographic description of a black hole is correct, including full quantum effects of gravity.

[1]Yukawa Institute for Theoretical Physics, Kyoto University, Kitashirakawa Oiwakecho, Sakyo-ku, Kyoto 606-8502, Japan. [2]The Hakubi Center for Advanced Research, Kyoto University, Yoshida Ushinomiyacho, Sakyo-ku, Kyoto 606-8501, Japan. [3]Stanford Institute for Theoretical Physics, Stanford University, Stanford, CA 94305, USA. [4]College of Science, Ibaraki University, Bunkyo 2-1-1, Mito, Ibaraki 310-8512, Japan. [5]Theory Center, High Energy Accelerator Research Organization (KEK), 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan. [6]Graduate University for Advanced Studies (SOKENDAI), 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan.
***Corresponding author. E-mail: hanada@yukawa.kyoto-u.ac.jp**

We performed Monte Carlo simulation of the dual gauge theory in the parameter regime corresponding to a black hole, which is destabilized due to quantum gravity effects. Our results for the mass of the black hole agree precisely with a prediction (*12*) obtained from independent calculations in gravity theory at the leading order of quantum corrections. Thus, we obtain quantitative evidence that the dual gauge theory provides a correct description of the black hole, including quantum gravity effects.

Superstring theory contains closed strings and open strings. The former mediates gravitational force, whereas the latter mediates gauge interactions such as the electromagnetic force. The theory also contains objects called D-branes (*13*), on which open strings can end. The dynamical property of D-branes, including the oscillation of open strings, is described by a gauge theory (*14*), which is a generalization of quantum electrodynamics. As a particular type of D-branes, we consider D-particles, which look like pointlike objects.

Let us consider the low-energy limit, which enables us to view strings as particles. If we furthermore neglect quantum effects, the full superstring theory can be well approximated by a generalized version of Einstein's gravity theory, which describes gravity classically in terms of the curvature associated with the space-time geometry. Within these approximations, N D-particles are expressed as a black hole. When N, the number of D-particles, is large, the size of the black hole is large and the geometry is weakly curved compared with the typical scale of quantum gravity. Hence, quantum gravity effects can be neglected. On the other hand, quantum gravity effects become important as N becomes small.

According to the gauge-gravity duality conjecture (see Fig. 1), superstring theory in the presence of the black hole made of D-particles is equivalent to the gauge theory that describes the system of D-particles (*15*). As the gauge theory is well defined at arbitrary N, it should capture the full quantum nature of superstring theory, including gravity, if the conjecture is true. Although there are many pieces of evidence for the gauge-gravity duality at $N = \infty$, where the classical approximation is fully justified on the gravity side (*16*), very little is known about it at the level of quantum gravity (*17*).

The gauge theory that describes the system of N D-particles is a quantum mechanical system specified with the following action (*18*, *19*)

$$S = \frac{N}{\lambda}\int_0^\beta dt \mathrm{tr}\left\{\frac{1}{2}(D_t X_i)^2 - \frac{1}{4}[X_i, X_j]^2 + \frac{1}{2}\psi_a D_t \psi_a - \frac{1}{2}\psi_a (\gamma_i)_{ab}[X_i, \psi_b]\right\} \quad (1)$$

We have introduced the fields $X_i(t)$ (i = 1, 2,..., 9) and $\psi_a(t)$ (a = 1, 2,..., 16), which are $N \times N$ bosonic and fermionic Hermitian matrices depending on the "imaginary time" t. In the above equation, λ is the coupling constant of the gauge theory, and the repeated indices are implicitly summed over. Intuitively, the diagonal elements of X_i describe the positions of N D-particles in nine spatial directions (*20*), whereas the off-diagonal elements correspond to strings connecting different D-particles. We have also defined the covariant derivative D_t and the 9d gamma matrices γ_i (supplementary text). The range of t is restricted to $0 \le t \le \beta \equiv 1/T$, with T being the temperature of the system, which should be identified with the Hawking temperature of the corresponding black hole on the gravity side.

The partition function Z is defined as the sum of the Boltzmann factors $\exp(-S)$ for all field configurations, where S is the action given by Eq. 1. Then the internal energy, which is the basic quantity we calculate, is defined by $E = -(\partial/\partial\beta)\log Z$. The internal energy actually corresponds to $E = M(T) - M(0)$ on the gravity side, where $M(T)$ denotes the mass of the black hole as a function of T. In what follows, we assume that E and T are made dimensionless by normalizing them with $\lambda^{1/3}$.

From the calculation on the gravity side (*12*), the internal energy is expected to behave as (supplementary text)

$$\frac{1}{N^2}E_{\mathrm{gravity}} = 7.41T^{2.8} - 5.77T^{0.4}\frac{1}{N^2} \quad (2)$$

up to $O(1/N^4)$ terms at sufficiently low temperature [$O(x)$ is a mathematical symbol that implies a quantity of the order of x]. The goal of our study is to see whether the gauge theory can reproduce Eq. 2, including the quantum gravity effects represented by the $O(1/N^2)$ term. On the other hand, the effects of the Hawking radiation, which represent nonlocal quantum effects in the black hole geometry, can be shown to be negligible in the parameter regime considered here (supplementary text). The possibility of probing such effects at much lower temperature shall be discussed as future work toward the end of this article.

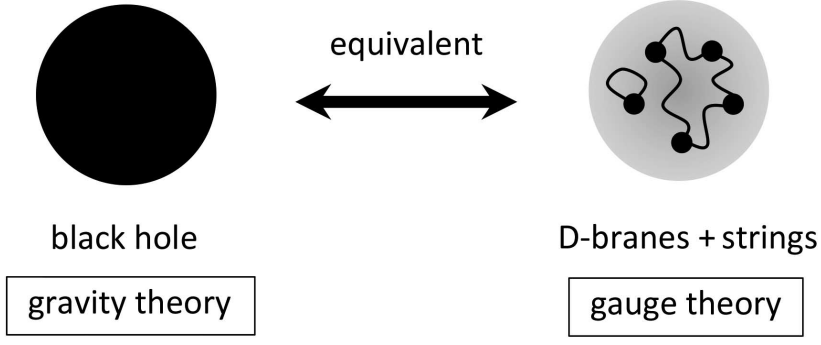


Fig. 1. The gauge-gravity duality conjecture. Black holes in superstring theory are conjectured to be described by the dual gauge theory.

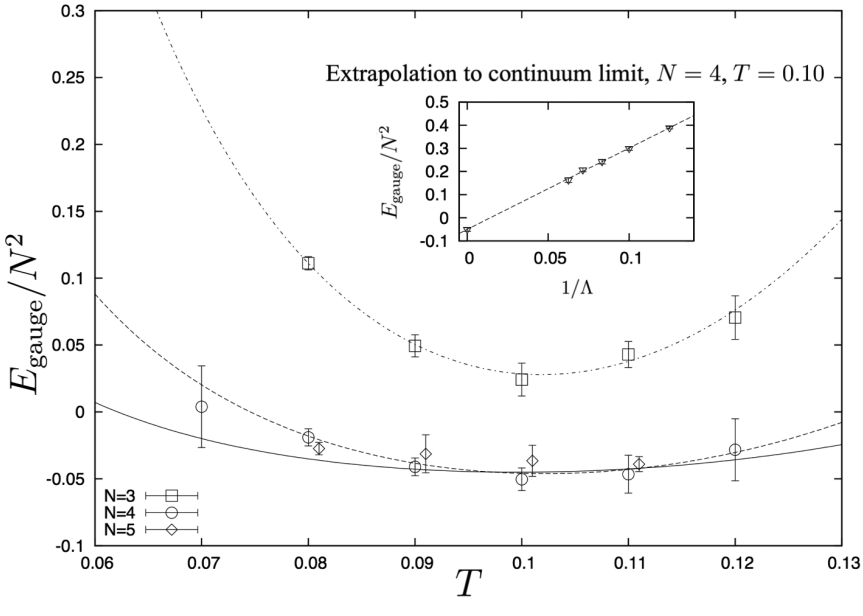


Fig. 2. The internal energy E_{gauge}/N^2 obtained for the metastable bound states in the $\Lambda \to \infty$ limit as a function of T. Results for N = 3 (squares), N = 4 (circles), and N = 5 (diamonds) are shown. The data points and the fitting curve for N = 5 are slightly shifted along the horizontal axis so that the data points and the error bars for N = 4 and 5 do not overlap. Error bars indicate the standard errors. (**Inset**) Extrapolation to $\Lambda \to \infty$ based on the ansatz $E = E_{\mathrm{gauge}} + O(1/\Lambda)$ for N = 4 and T = 0.10.

To put the system in Eq. 1 on a computer, we expand each field in terms of Fourier modes with respect to time t and introduce a cutoff Λ on the frequency after fixing the gauge symmetry appropriately (*21*–*23*). Using the results obtained at various values of Λ, we make an extrapolation to $\Lambda = \infty$. Although the fermionic matrices make the effective Boltzmann weight complex, we simply take the absolute value, which is shown to be a valid approximation in the present case (*24*).

Here, we focus on small values of N (such as $N = 3$, 4, and 5) to probe the quantum gravity effects, which correspond to $1/N^2$ corrections. This causes a new technical difficulty, which was absent in previous work (*22*, *23*) at large N, such as $N = 17$. [See also (*25*, *26*) for related work.] We observe that the eigenvalues of the bosonic matrices X_i start to diverge while we are sampling important field configurations that contribute to the partition function. When N becomes small, the cluster of the eigenvalues becomes metastable, as expected from the instability of the black hole due to the quantum gravity effects discussed above. We therefore had to develop a method (*27*) that enables us to estimate the contribution from the metastable bound states.

In Fig. 2, we plot our results for the internal energy $E_{\rm gauge}$ in the $\Lambda \to \infty$ limit as a function of T for $N = 3, 4, 5$. The curves represent the fits to the behaviors expected from the gravity side, which shall be explained later. At sufficiently low temperature, the internal energy increases as temperature decreases, which implies that the specific heat becomes negative. Such a behavior is possible because we are measuring the internal energy of the metastable bound states.

In Fig. 3, we plot $(E_{\rm gauge} - E_{\rm gravity})/N^2$ against $1/N^4$ for $T = 0.08$ and 0.11. Our data are nicely fitted by straight lines passing through the origin. This implies that our results obtained on the gauge theory side are indeed consistent with the result in Eq. 2 obtained on the gravity side, including quantum gravity corrections. The agreement of similar accuracy is also observed at other values of T.

As a further consistency check, we fit our results for each T by $E_{\rm gauge}/N^2 = 7.41T^{2.8} + c_1/N^2 + c_2/N^4$, leaving c_1 and c_2 as fitting parameters depending on T. In Fig. 4, we plot c_1 obtained by the two-parameter fit against T, which agrees well with $c_1 = -5.77T^{0.4}$. As for the coefficient c_2 of the $O(1/N^4)$ terms, the prediction from the gravity side is given by $c_2 = cT^{-2.6} + \ldots$, where c is an unknown constant. In fact, c_2 can be fitted, for instance, by $c_2 = cT^{-2.6} + \tilde{c}T^{p}$, with $c = 0.0340(11)$, $\tilde{c} = 0.17(19) \times 10^6$, and $p = 4.30(52)$. (The value for $\tilde{c}$ looks huge, but it is actually compensated by the high power of T within the temperature region investigated here.) Therefore, we consider that the T dependence of c_2 is also consistent with the prediction from the gravity side. The curves in Fig. 2 represent $E_{\rm gauge}/N^2 = E_{\rm gravity}/N^2 + (cT^{-2.6} + \tilde{c}T^{p})/N^4$ with the fitting parameters obtained above.

Our work suggests a new approach to the yet mysterious nature of quantum gravity. Now that we have clear evidence for the gauge-gravity duality, including quantum gravity effects, we can turn the table around and attempt to study various issues involving quantum gravity by performing Monte Carlo simulation of the dual gauge theory. For instance, it is speculated that the dual gauge theory studied in this work undergoes a phase transition at the critical temperature $T_c \propto N^{-5/9}$ corresponding to the Gregory-Laflamme transition on the gravity side (*15*). Below this temperature, the dual gauge theory is speculated to describe a Schwarzschild black hole in 11 dimensions, which causes the Hawking radiation. Our method appears capable of reaching this regime, and the obtained results would give us deeper insights into the information loss paradox.

In short, we feel that problems involving quantum gravity have become as tractable as problems involving the strong interaction. The latter can be studied by simulating a gauge theory on a four-dimensional (4D) lattice, and such a method has recently been used to reproduce the mass spectrum of hadrons (*28*) and the nuclear force (*29*). We can now apply essentially the same method to study quantum gravity, which has been thought to be far more difficult.

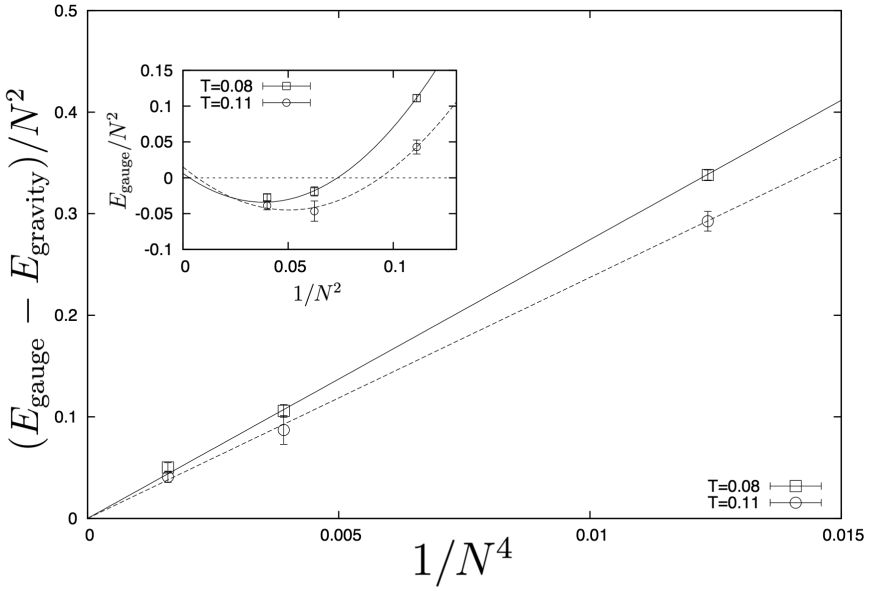


Fig. 3. The difference ($E_{\rm gauge} - E_{\rm gravity}$)/$N^2$ as a function of $1/N^4$. We show the results for $T = 0.08$ (squares) and $T = 0.11$ (circles). Error bars denote the standard errors. (**Inset**) $E_{\rm gauge}/N^2$ plotted against $1/N^2$ for $T = 0.08$ and 0.11. The curves represent the fits to the behavior $E_{\rm gauge}/N^2 = 7.41T^{2.8} - 5.77T^{0.4}/N^2 + C/N^4$ expected from the gravity side, where C is a fitting parameter depending on T.

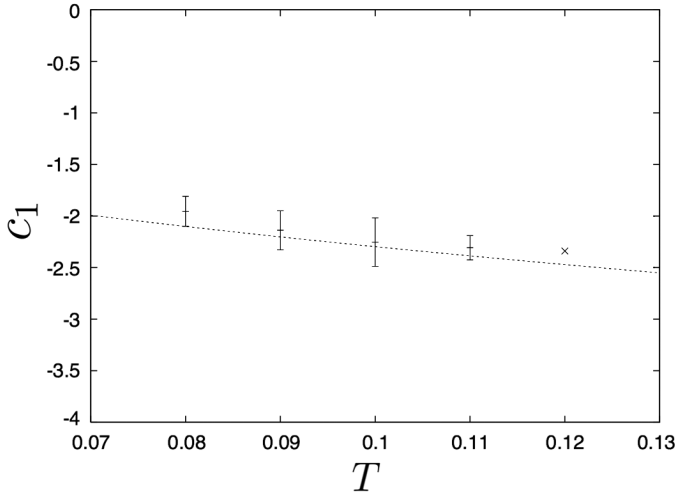


Fig. 4. The coefficient c_1 of the $O(1/N^2)$ term as a function of T. The dotted line represents the prediction $c_1 = -5.77T^{0.4}$ from the gravity side. The data point at $T = 0.12$ does not have an error bar, because only two data points ($N = 3$ and 4) were available for making a two-parameter fit. Error bars indicate the standard errors.

REFERENCES AND NOTES

1. S. W. Hawking, *Commun. Math. Phys.* **43**, 199–220 (1975).
2. S. W. Hawking, *Commun. Math. Phys.* **46**, 206 (1976).
3. J. D. Bekenstein, *Phys. Rev. D* **7**, 2333–2346 (1973).
4. G. T. Horowitz, J. M. Maldacena, *J. High Energy Phys.* **0402**, 008 (2004).
5. A. Almheiri, D. Marolf, J. Polchinski, J. Sully, *J. High Energy Phys.* **1302**, 062 (2013).
6. S. W. Hawking, *Phys. Rev. D* **14**, 2460–2473 (1976).
7. S. W. Hawking, *Commun. Math. Phys.* **87**, 395–415 (1982).
8. A. Strominger, C. Vafa, *Phys. Lett. B* **379**, 99–104 (1996).
9. J. M. Maldacena, *Adv. Theor. Math. Phys.* **2**, 231 (1998).
10. G. 't Hooft, http://arxiv.org/abs/gr-qc/9310026.
11. L. Susskind, *J. Math. Phys.* **36**, 6377 (1995).

12. Y. Hyakutake, *Prog. Theor. Exp. Phys.* **033B04**, 1–27 (2014).
13. J. Polchinski, *Phys. Rev. Lett.* **75**, 4724–4727 (1995).
14. E. Witten, *Nucl. Phys. B* **460**, 335–350 (1996).
15. N. Itzhaki, J. M. Maldacena, J. Sonnenschein, S. Yankielowicz, *Phys. Rev. D* **58**, 046004 (1998).
16. J. A. Minahan *et al.*, *Lett. Math. Phys.* **99**, 33–58 (2012).
17. Although we study dynamical aspects of the duality in the present work, certain kinematical aspects of the duality have been tested recently at the level of quantum gravity (*30*).
18. T. Banks, W. Fischler, S. H. Shenker, L. Susskind, *Phys. Rev. D* **55**, 5112–5128 (1997).
19. B. de Wit, J. Hoppe, H. Nicolai, *Nucl. Phys. B* **305**, 545–581 (1988).
20. Theoretical consistency requires that superstring theory should be defined in 10D space-time. To realize our 4D space-time, the size of the extra six dimensions can be very small without spoiling the consistency.
21. M. Hanada, J. Nishimura, S. Takeuchi, *Phys. Rev. Lett.* **99**, 161602 (2007).
22. K. N. Anagnostopoulos, M. Hanada, J. Nishimura, S. Takeuchi, *Phys. Rev. Lett.* **100**, 021601 (2008).
23. M. Hanada, Y. Hyakutake, J. Nishimura, S. Takeuchi, *Phys. Rev. Lett.* **102**, 191602 (2009).
24. M. Hanada, J. Nishimura, Y. Sekino, T. Yoneya, *J. High Energy Phys.* **1112**, 020 (2011).
25. D. Kabat, G. Lifschytz, D. A. Lowe, *Phys. Rev. Lett.* **86**, 1426–1429 (2001).
26. S. Catterall, T. Wiseman, *Phys. Rev. D* **78**, 041502 (2008).
27. Materials and methods are available as supplementary materials on *Science* Online.
28. S. Dürr *et al.*, *Science* **322**, 1224–1227 (2008).
29. N. Ishii, S. Aoki, T. Hatsuda, *Phys. Rev. Lett.* **99**, 022001 (2007).
30. A. Arabi Ardehali, J. T. Liu, P. Szepietowski, *J. High Energy Phys.* **1306**, 024 (2013).

ACKNOWLEDGMENTS

The authors would like to thank S. Aoki, S. Hartnoll, I. Kanamori, H. Kawai, E. Poppitz, A. Schäfer, S. Shenker, L. Susskind, M. Tezuka, A. Ueda, and M. Ünsal for discussions and comments. M.H. is supported by the Hakubi Center for Advanced Research, Kyoto University and by the National Science Foundation under grant no. PHYS-1066293. M.H. and Y.H. are partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), 25800163, 2013 (M.H.); 19740141, 2007 (Y.H.); and 24740140, 2012 (Y.H.). The work of J.N. was supported in part by Grant-in-Aid for Scientific Research (nos. 20540286 and 23244057) from the Japan Society for the Promotion of Science. Computations were carried out on PC cluster systems in KEK and the Osaka University Cybermedia Center (the latter being provided by the High Performance Computing Infrastructure System Research Project, project ID: hp120162). All the data obtained in the present work are presented in table S1 of the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/882/suppl/DC1
Materials and Methods
Supplementary Text
Fig. S1
Table S1
References (*31*–*35*)

23 December 2013; accepted 31 March 2014
Published online 17 April 2014;
10.1126/science.1250122

CHEMISTRY

Real-space imaging of molecular structure and chemical bonding by single-molecule inelastic tunneling probe

Chi-lun Chiang,[1]* Chen Xu,[1]* Zhumin Han,[1]* W. Ho[1,2]†

The arrangement of atoms and bonds in a molecule influences its physical and chemical properties. The scanning tunneling microscope can provide electronic and vibrational signatures of single molecules. However, these signatures do not relate simply to the molecular structure and bonding. We constructed an inelastic tunneling probe based on the scanning tunneling microscope to sense the local potential energy landscape of an adsorbed molecule with a carbon monoxide (CO)–terminated tip. The skeletal structure and bonding of the molecule are revealed from imaging the spatial variations of a CO vibration as the CO-terminated tip probes the core of the interactions between adjacent atoms. An application of the inelastic tunneling probe reveals the sharing of hydrogen atoms among multiple centers in intramolecular and extramolecular bonding.

The achievement of a mechanistic understanding of chemical and biological functions depends on knowing the geometric structure and the nature of the bonds in the molecules. Consequently, a number of techniques have been extensively developed to attain this knowledge, including x-ray diffraction, electron diffraction, and nuclear magnetic resonance. These techniques, however, do not provide a direct view of the molecules in real space. Nonetheless, they have yielded three-dimensional structures of many complex molecules that enabled the elucidation of their chemical and biological properties. Only recently has the atomic force microscope (AFM) been used to obtain real-space images of the molecular structures of mostly planar molecules (*1*, *2*). The AFM approach allows structural imaging that can discriminate a reactant and its different products (*3*) or reveal hydrogen bonding between molecules (*4*).

The high spatial resolution of the AFM was obtained by functionalizing the tip with a CO molecule (*5*, *6*) and measuring the shift in the resonance frequency of the quartz tuning fork above the adsorbed molecule (*7*). The spatial resolution arises from variations of the force gradient sensed by the CO-tip as it scans over different parts of the molecule. The observed contrast revealing the molecular structure implies that the frequency shift is different over the atoms and the bonds between them, relative to elsewhere. The range of frequency shift is few Hz from the resonance of 20 to 30 kHz.

In comparison, the scanning tunneling microscope (STM) has been shown to reveal the electronic properties of the sample. Good agreements have been obtained between theory and experiment for the molecular orbitals in describing the spatial distributions for the electron density (*8*–*10*) and spin excitation (*11*). These images reflect the electron wave functions that are related to (but do not directly display) the molecular structures. By trapping a hydrogen molecule in the STM junction or transferring a Xe, CO, or CH_4 molecule to the tip, molecular structure could be resolved from the topographic and differential conductance images, and intermolecular bonds were revealed (*12*–*14*).

Here, we demonstrate an approach based on the STM to image the skeletal structure and bonding in an adsorbed molecule by single-molecule inelastic tunneling probe (itProbe). A CO molecule is transferred to the tip, and a vibrational mode of the tip CO senses the bonding between two atoms in an adsorbed molecule. As the CO-terminated tip is scanned over the molecule during imaging, changes in the energy and intensity of the hindered translational vibration of CO are measured by inelastic electron tunneling spectroscopy (IETS) with the STM (*15*). This low-energy CO vibration senses the spatially varying potential energy landscape of the molecule and its surroundings. The range of energy shift is on the order of the vibrational energy of ~3 meV, or equivalently ~0.7 THz.

All of the experiments were performed in ultrahigh vacuum (5×10^{-11} torr); the spectra reported were taken at a sample and STM temperature of 600 mK (*16*). A topographic image taken with a bare Ag tip of cobalt phthalocyanine (CoPc) coadsorbed with CO on Ag(110) is shown in Fig. 1A. Adsorption configurations, labeled CoPc(×) and CoPc(+), are possible on the surface. Each CO molecule is identified by its hindered translational (2.8 meV) and rotational (18.3 and 20.3 meV) modes in the vibrational spectra by STM-IETS (Fig. 1B). The nondegenerate hindered rotation in the two orthogonal directions parallel to the Ag(110) surface is resolved as a peak splitting. The same area imaged after

[1]Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA. [2]Department of Chemistry, University of California, Irvine, CA 92697, USA.
*These authors contributed equally to this work. †Corresponding author. E-mail: wilsonho@uci.edu

transferring a CO molecule (marked by arrow in Fig. 1A) to the tip is shown in Fig. 1C. The presence of CO on the tip is confirmed by STM-IETS taken at any location over the clean Ag surface (Fig. 1D). The energies of the hindered translational (2.1 meV) and rotational (18.2 meV) modes are close to those of CO on the Ag surface as recorded by the bare tip.

Vibrational spectra can be recorded by STM-IETS with high spatial resolution at a chosen position (point spectroscopy) over an adsorbed molecule (*15*, *17*) and with bare, CO-terminated, and ethylene-terminated tips (*18*). In the scanned energy range, vibrational modes are resolved for CO but not for CoPc. As the CO-terminated tip

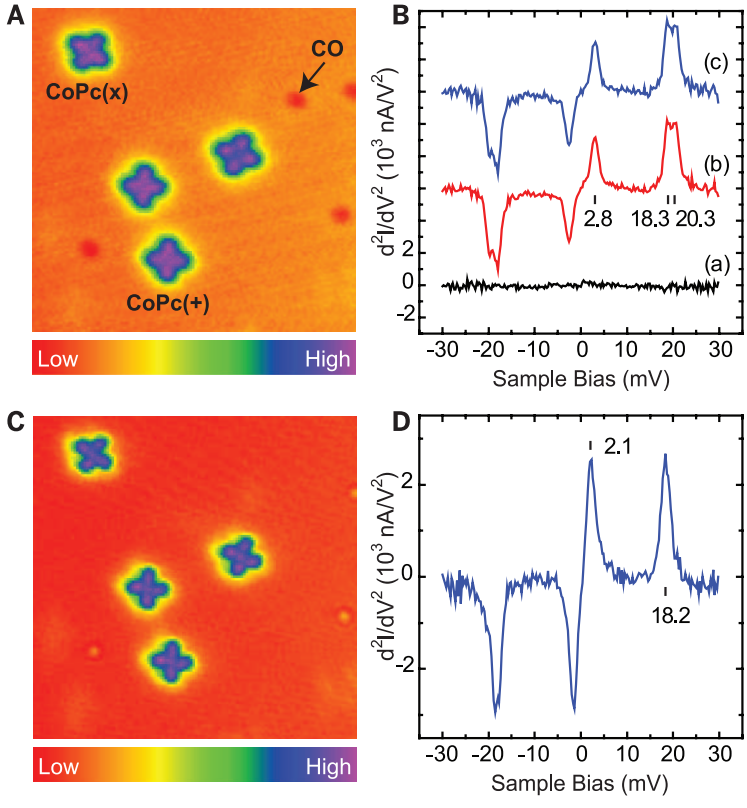


Fig. 1. The creation and characterization of a CO-terminated tip. (**A**) Constant-current topography of CO coadsorbed with CoPc(×) and CoPc(+) on Ag(110) at 600 mK (121.1 Å × 121.1 Å). Tunneling gap set point: V = 0.1 V and I = 0.1 nA. (**B**) Vibrational STM-IETS d^2I/dV^2 spectra taken with a bare tip over the Ag surface (a), CO molecule (b), and the background-subtracted spectrum (c). Bias root mean square (RMS) voltage modulation: 0.6 mV at 471 Hz. Set point: V = 10.0 mV and I = 1.0 nA. (**C**) Constant-current topography of the same area as in (A) after transferring a CO [the one indicated by the arrow in (A)] to the tip. CO is imaged as a protrusion instead of a depression. (**D**) Vibrational STM-IETS d^2I/dV^2 spectrum of CO-terminated tip taken over the Ag surface. Bias RMS voltage modulation: 0.6 mV at 471 Hz. Set point: V = 10.0 mV and I = 1.0 nA.

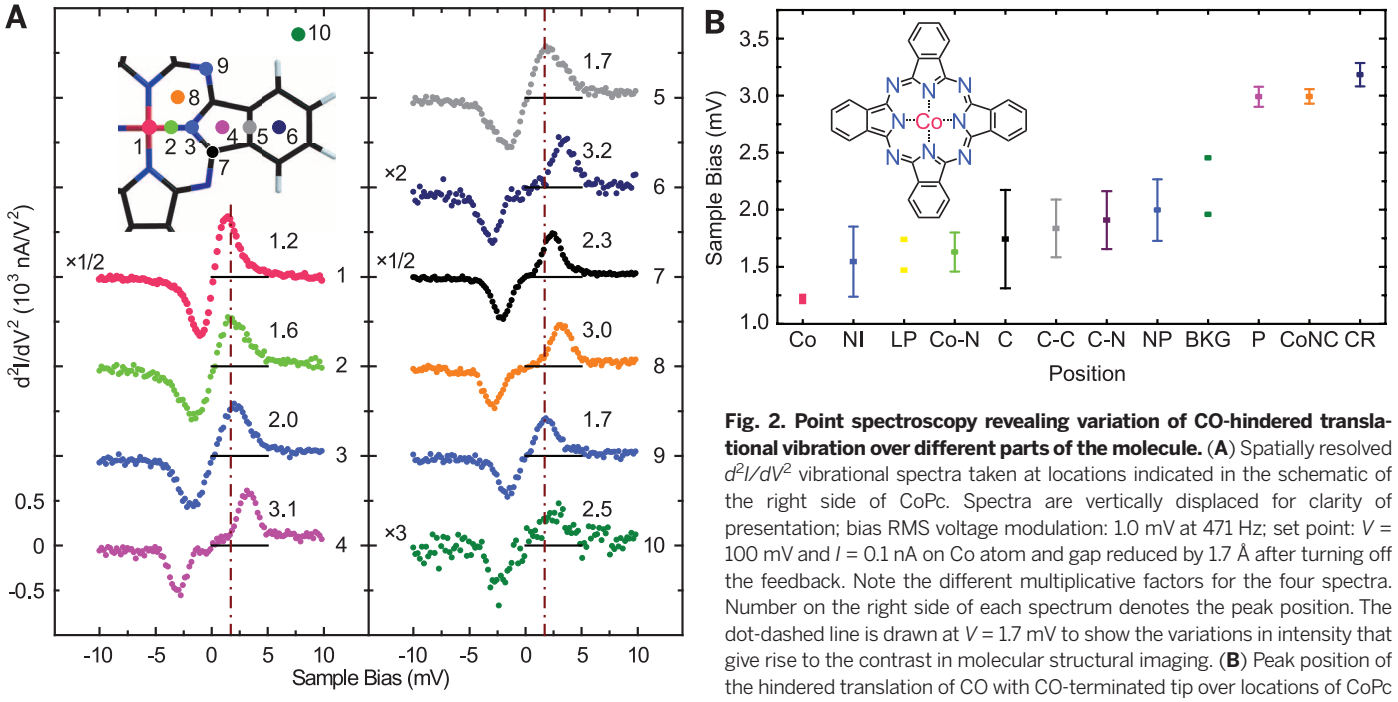


Fig. 2. Point spectroscopy revealing variation of CO-hindered translational vibration over different parts of the molecule. (**A**) Spatially resolved d^2I/dV^2 vibrational spectra taken at locations indicated in the schematic of the right side of CoPc. Spectra are vertically displaced for clarity of presentation; bias RMS voltage modulation: 1.0 mV at 471 Hz; set point: V = 100 mV and I = 0.1 nA on Co atom and gap reduced by 1.7 Å after turning off the feedback. Note the different multiplicative factors for the four spectra. Number on the right side of each spectrum denotes the peak position. The dot-dashed line is drawn at V = 1.7 mV to show the variations in intensity that give rise to the contrast in molecular structural imaging. (**B**) Peak position of the hindered translation of CO with CO-terminated tip over locations of CoPc molecule indicated on the x axis: Co (cobalt atom), NI (imine nitrogen), LP (lone pair on imine nitrogen), Co-N (cobalt-pyrrole nitrogen bond), C (carbon atom), C-C bond, C-N bond, NP (pyrrole nitrogen), BKG (Ag substrate), P (pyrrole ring), CoNC (CoNCNCN ring), CR (six-member carbon ring). Inset shows a schematic diagram of CoPc. There are only two data points for Co, LP, and BKG, hence no error bars.

is scanned over the CoPc, the percent change in the vibrational energy is larger for the hindered translation relative to the hindered rotation. Furthermore, the noise in the current decreases as the bias voltage is lowered. Both factors favor the hindered translational vibration for sensing the CoPc.

The sensitivity of the hindered translational mode of CO to its position over a CoPc molecule is shown by vibrational point spectroscopy in Fig. 2A. A clear vibrational energy upshift by ~2.0 meV is measured as the spectrum is recorded over the central Co atom (point 1; 1.2 meV) instead of over the center of the five-member pyrrole ring (point 4; 3.1 meV), the six-member carbon ring (point 6; 3.2 meV), and the inner six-member CoNC ring (point 8; 3.0 meV). The vibrational energy is <2.5 meV for spectra taken over other points of the molecule. The intensity over the Ag background is weaker than over the molecule. Figure 2A shows the spectra taken over 10 selected points of the molecule and the Ag background (a more comprehensive set of spectra is given in fig. S1; 57 locations over the molecule are labeled in fig. S2E). The bias voltage at the peak of each spectrum in fig. S1 is plotted in Fig. 2B; the error bars show the spread in the peak energy centered over the average value. Possible sources for this spread include (i) incommensurability between molecule and substrate, (ii) atomic unevenness of molecule and substrate, (iii) spatial variation in the tip-CO tilt, and (iv) imprecise positioning of the tip over the molecule.

The CO-terminated tip can be used for energy-resolved spatial imaging of the molecule. When the energy is chosen within the CO-hindered translational vibration and <2.5 meV, the image reveals the skeletal structure of CoPc using either the constant-height or constant-current mode (*16*), as shown in Fig. 3, A and B, for CoPc(+). A structural image for CoPc(×) is shown in Fig. 3C; the arrows point to the four lone pairs of the imine nitrogen, each forming intramolecular hydrogen bonds with the two nearest C-H bonds. The CoPc(+) in Fig. 3B is shown schematically in Fig. 3D; the dashed lines indicate the intramolecular hydrogen bonds. Such bonding has been tentatively discussed in structural images taken with the AFM (*19*). In contrast, these intramolecular hydrogen bonds appear to be absent when the CoPc is adsorbed on the Au(110)2×1 reconstructed surface (Fig. 3E) where every other row of Au atoms is missing. The C-H bonds of CoPc on the Au(110)2×1 project outward, as shown by dashed lines in Fig. 3F. The schematic of CoPc is superimposed and compared to the structural images in fig. S3. The molecular structure and bonds are imaged with different resolutions using constant-height versus constant-current modes, as shown in Fig. 3 and further supported in figs. S4 and S5. Evidently, images of the skeletal structure and spectroscopic sensitivity to different atoms, bonds, and regions in the molecule are achieved by the itProbe. Bonds, rather than densities of states and molecular orbitals, are imaged by itProbe and reflect localized interactions or bonds between adjacent atoms.

Some C-H bonds of CoPc on the Au(110)2×1 surface are imaged as "V" shapes in Fig. 3E and marked in Fig. 3F, exhibiting bifurcation of

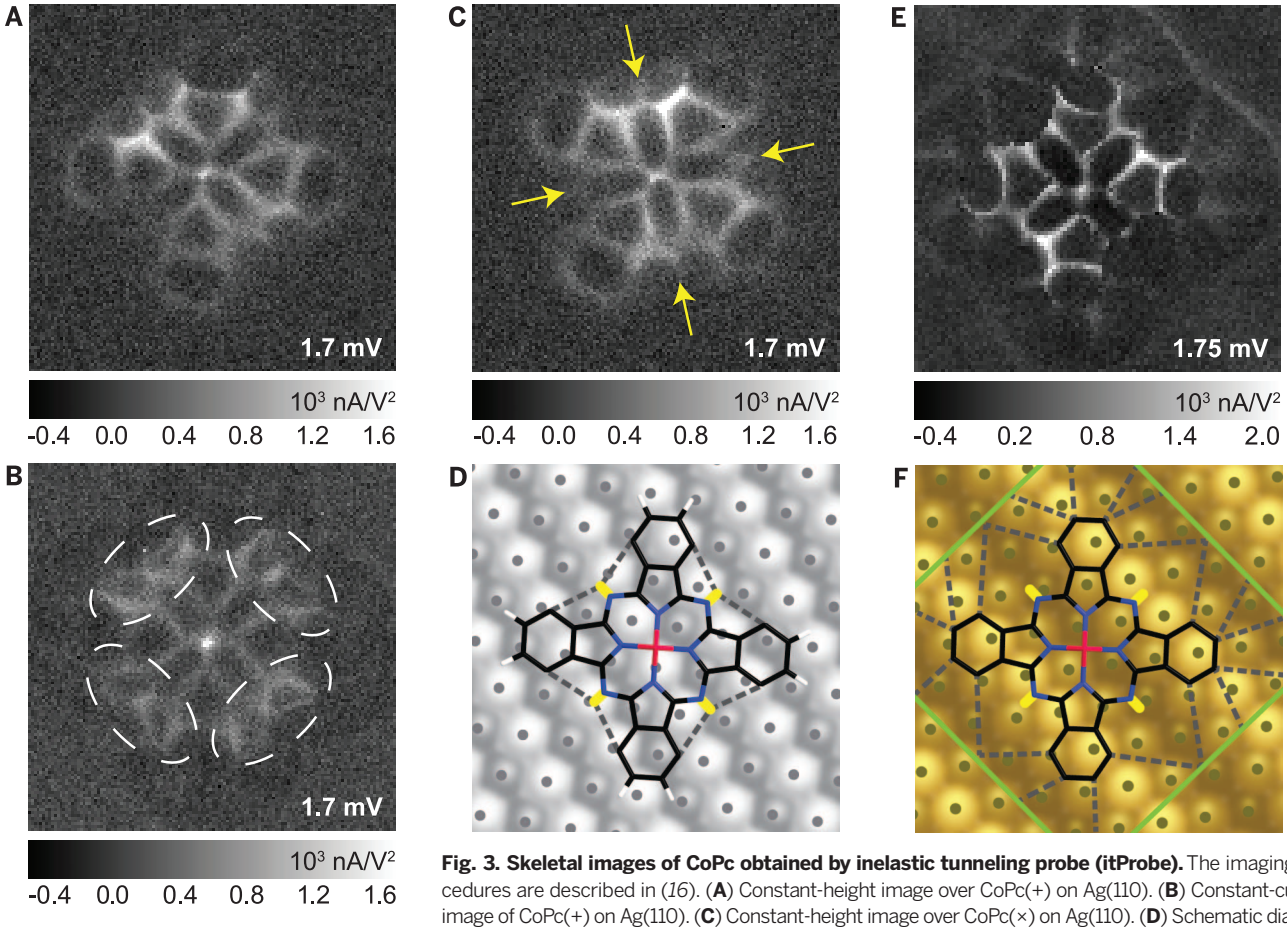


Fig. 3. Skeletal images of CoPc obtained by inelastic tunneling probe (itProbe). The imaging procedures are described in (*16*). (**A**) Constant-height image over CoPc(+) on Ag(110). (**B**) Constant-current image of CoPc(+) on Ag(110). (**C**) Constant-height image over CoPc(×) on Ag(110). (**D**) Schematic diagram showing the skeletal structure of CoPc(+) and the intramolecular hydrogen bonds highlighted in dashed ellipses in (B). The underlying Ag(110) surface and the adsorption geometry of CoPc(+) are determined from atomically resolved topographic images taken with a CO-terminated tip and with each CO bonded on top of an Ag surface atom. Area in (A) to (C) is 128 × 128 pixels and 20.2 Å × 20.2 Å. (**E**) Constant-current image over CoPc(+) on Au(110)2×1 surface (96 × 96 pixels and 21.2 Å × 21.2 Å). A rectangular boundary is resolved and is attributed to the range of electronic influence of adsorbed CoPc. A similar boundary is present on the Ag(110) surface but is less prominent than that on the Au(110) surface. (**F**) Schematic diagram showing skeletal structure of CoPc(+) on Au(110)2×1 surface. The missing surface atom rows are along the dark rows. The Co atom is over the missing row but is displaced to one side. The depicted adsorption geometry is obtained from an atomically resolved topographic image obtained with a CO-terminated tip.

hydrogen bonds (*20*, *21*). This striking feature suggests that the C-H bonds can be aligned in two directions induced by interactions with the underlying Au substrate. Evidently, a hydrogen atom can be involved in a multicenter bonding. In contrast, the structural images for CoPc on Ag(110) reveal internal bonding between each of the four lone pairs of the imine nitrogen with its two nearest-neighbor C-H bonds, forming three-center two-hydrogen bonds. These intramolecular hydrogen bonds are supposedly disfavored because of the nonlinear geometry of the C-H bond and the imine nitrogen lone pair (*21*). The images show curved bonds due presumably to this nonlinear geometry.

The multicenter hydrogen bonding is also evident in intermolecular interactions. Two CoPc molecules can be found to be adjacent to each other, such as in Fig. 4A on Ag(110). The hydrogen bonding driven by the lone pair electrons of the imine nitrogen and nearby C-H bonds is imaged by itProbe in Fig. 4B. A close-up image in Fig. 4C reveals the intramolecular and intermolecular hydrogen bonds at high spatial resolution. In Fig. 4D, the interactions among the imine nitrogen and four nearby carbon atoms are mediated by four hydrogen atoms in a five-center four-hydrogen bonding. The structural images obtained by itProbe reflect the time-averaged density of the hydrogen bonds.

The structural images were obtained by scanning in the horizontal (x) and vertical (y) directions. No noticeable differences were seen (figs. S7 and S8). The CO on the tip, however, is induced to tilt as it is brought close to the surface during imaging (*22*). This tilt gives rise to a constant offset between the structural image and the topography taken with the tip farther away from the surface where the induced tilt is absent. The amount of tilt varies for different CO-terminated tips; this offset is 1.96 Å for the specific CO-terminated tip shown in fig. S9.

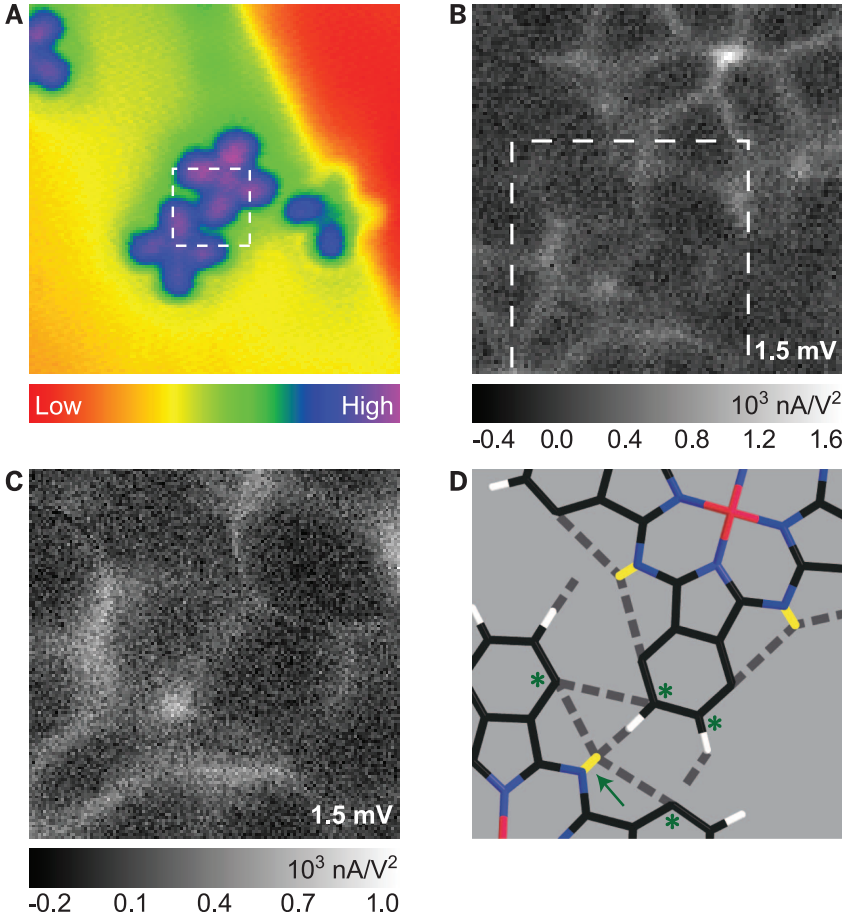


Fig. 4. Imaging the chemical bonding between a pair of CoPc on Ag(110) surface. (**A**) Constant-current topographic image (128 × 128 pixels or 60.6 Å × 60.6 Å) with sample bias V = 0.1 V and tunneling current I = 0.1 nA. (**B**) Constant-current itProbe image (100 × 100 pixels or 12.6 Å × 12.6 Å) at 1.5 mV over the white square area shown in (A). (**C**) Constant-current itProbe image (128 × 128 pixels or 8.1 Å × 8.1 Å) at 1.5 mV over the white square area shown in (B). (**D**) Schematic diagram of the image in (B), showing in dashed lines the intermolecular and intramolecular bonds involving hydrogen, and in yellow the lone pairs of imine nitrogen. Four hydrogen atoms are shared among five centers [four carbon atoms (indicated by asterisk) and one imine nitrogen (indicated by arrow)] in the five-center four-hydrogen bonds.

The contrast revealing the skeletal structure of the molecule (fig. S10) depends on the sample bias used in the imaging. The hindered translational mode of CO is low in energy (soft) and particularly sensitive to its environment, relative to the hindered rotational mode at higher energy. High-energy resolution is required for the STM to measure low-energy modes and is achieved by lowering the temperature to 600 mK (0.28 meV temperature broadening) and the modulation to 1 mV and below (*11*, *23*).

The capability of the STM to image molecular structure and chemical bonds broadens its previous applications to determine the electronic and vibrational properties of single molecules. This added capability provides much-needed knowledge to understand the relation between the structure and function of molecules. In addition to the electronic and vibrational properties, it may be possible to observe changes in geometric structure when the molecule's environment is altered by intermolecular interactions, such as in the formation of self-assembled nanostructures or in reactions on surfaces. Real-space spectroscopy and imaging by the itProbe could lead to a deeper understanding of the nature of different types of chemical bonds and related chemistry.

REFERENCES AND NOTES

1. L. Gross, F. Mohn, N. Moll, P. Liljeroth, G. Meyer, *Science* **325**, 1110–1114 (2009).
2. L. Gross *et al.*, *Science* **337**, 1326–1329 (2012).
3. D. G. de Oteyza *et al.*, *Science* **340**, 1434–1437 (2013).
4. J. Zhang *et al.*, *Science* **342**, 611–614 (2013).
5. L. Bartels *et al.*, *Phys. Rev. Lett.* **80**, 2004–2007 (1998).
6. H. J. Lee, W. Ho, *Science* **286**, 1719–1722 (1999).
7. F. J. Giessibl, *Appl. Phys. Lett.* **76**, 1470–1472 (2000).
8. J. Repp, G. Meyer, S. M. Stojković, A. Gourdon, C. Joachim, *Phys. Rev. Lett.* **94**, 026803 (2005).
9. S. W. Wu, N. Ogawa, W. Ho, *Science* **312**, 1362–1365 (2006).
10. P. Liljeroth, J. Repp, G. Meyer, *Science* **317**, 1203–1206 (2007).
11. U. Ham, W. Ho, *J. Chem. Phys.* **138**, 074703 (2013).
12. C. Weiss *et al.*, *Phys. Rev. Lett.* **105**, 086103 (2010).
13. C. Weiss, C. Wagner, R. Temirov, F. S. Tautz, *J. Am. Chem. Soc.* **132**, 11864–11865 (2010).
14. G. Kichin, C. Weiss, C. Wagner, F. S. Tautz, R. Temirov, *J. Am. Chem. Soc.* **133**, 16847–16851 (2011).
15. B. C. Stipe, M. A. Rezaei, W. Ho, *Science* **280**, 1732–1735 (1998).
16. See supplementary materials on *Science* Online.
17. W. Ho, *J. Chem. Phys.* **117**, 11033–11061 (2002).
18. J. R. Hahn, W. Ho, *Phys. Rev. Lett.* **87**, 196102 (2001).
19. L. Gross *et al.*, *Nat. Chem.* **2**, 821–825 (2010).
20. I. Rozas, I. Alkorta, J. Elguero, *J. Phys. Chem. A* **102**, 9925–9932 (1998).
21. T. Steiner, *Angew. Chem. Int. Ed.* **41**, 48–76 (2002).
22. A. J. Weymouth, T. Hofmann, F. J. Giessibl, *Science* **343**, 1120–1122 (2014).
23. L. J. Lauhon, W. Ho, *Rev. Sci. Instrum.* **72**, 216–223 (2001).

ACKNOWLEDGMENTS

Supported by the Chemical Science, Geo- and Bioscience Division, Office of Science, U.S. Department of Energy, under grant DE-FG02-04ER15595.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/885/suppl/DC1
Materials and Methods
Figs. S1 to S10
Reference (*24*)

14 March 2014; accepted 29 April 2014
10.1126/science.1253405

SUBSURFACE MICROBES

Global rates of marine sulfate reduction and implications for sub–sea-floor metabolic activities

Marshall W. Bowles,[1*†] José M. Mogollón,[1,2,3†] Sabine Kasten,[1,2] Matthias Zabel,[1] Kai-Uwe Hinrichs[1]

Sulfate reduction is a globally important redox process in marine sediments, yet global rates are poorly quantified. We developed an artificial neural network trained with 199 sulfate profiles, constrained with geomorphological and geochemical maps to estimate global sulfate-reduction rate distributions. Globally, 11.3 teramoles of sulfate are reduced yearly (~15% of previous estimates), accounting for the oxidation of 12 to 29% of the organic carbon flux to the sea floor. Combined with global cell distributions in marine sediments, these results indicate a strong contrast in sub–sea-floor prokaryote habitats: In continental margins, global cell numbers in sulfate-depleted sediment exceed those in the overlying sulfate-bearing sediment by one order of magnitude, whereas in the abyss, most life occurs in oxic and/or sulfate-reducing sediments.

Sulfate reduction is a ubiquitous microbial process in oceanic sediments and an important pathway for carbon oxidation and redox cycling (*1*, *2*). Nevertheless, the currently estimated global sulfate-reduction rate (SRR) of 75 teramoles (Tmol) of sulfate per year (*1*) is based on coarse spatial averaging and is not consistent with the up-to-date assessment of the global organic matter flux to marine sediments of 79 to 192 Tmol carbon per year (*3*, *4*). Net sulfate reduction follows a two carbon–to–one sulfur stoichiometric ratio [$C_6H_{12}O_6 + 3SO_4^{2-} \rightarrow 3S^{2-} + 6CO_2 + 6H_2O$; for example, see (*5*)]. Therefore, the current estimates for the rate of subsurface sulfate reduction and for the organic carbon flux to the sediment suggest that either insufficient organic carbon reaches the sediment to account for sulfate reduction or that most (78%) organic matter is channeled toward sulfate reduction. Nevertheless, the organic carbon reaching the sediment must also foment other prominent redox reactions such as carbon respiration (*4*, *6*). Moreover, a sizable portion of sedimentary organic matter successfully survives early diagenesis and is buried (*7*). This discrepancy in global geochemical cycles gives impetus for an amended global view on sulfate reduction, which can be merged with recently revised global prokaryotic abundances to properly assess the activity of sulfate-reducing microorganisms at a global scale (*5*, *8*–*11*).

We used currently available sulfate-concentration profiles from multiple scientific ocean drilling programs (*12*) to estimate global net SRRs (Fig. 1A). These profiles were best described by assuming that sulfate concentrations exponentially decrease with depth. A total of 199 sulfate profiles (Fig. 1B) with a mean error square value <4 mM^2 based on a least-squares regression were selected for the global SRR analysis. We then used depth-decay constants (*b*) extracted from these profiles to train an artificial neural network (ANN) using high-resolution (1 × 1 degree) satellite observations and water-column chemistry maps (e.g., surface water chlorophyll A, particulate organic carbon, and bottom water O_2) (*12*–*14*) (table S1).

The ANN predicts depth-decay constants ranging 16 orders of magnitude from 5.8×10^{-13} to 3.2×10^{2} m^{-1}, while depth-integrated SRRs calculated by a steady-state diffusion, advection, and reaction function (eq. S8) (*12*) ranged from 5.8×10^{-12} to 8.2 mmol cm^{-2} $year^{-1}$. The highest SRRs were predicted in shelf environments, and the lowest SRRs were calculated in the nutrient-poor oceanic gyres (Fig. 1A and Table 1). These general trends are corroborated by a previous prediction of global, depth-integrated SRRs de-

[1]MARUM Center for Marine Environmental Sciences, University of Bremen, Bremen, Germany. [2]Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany. [3]Department of Earth Sciences–Geochemistry, Faculty of Geosciences, Utrecht University, Netherlands.
*Corresponding author. E-mail: bowlesmw@uni-bremen.de
†These authors contributed equally to this work.

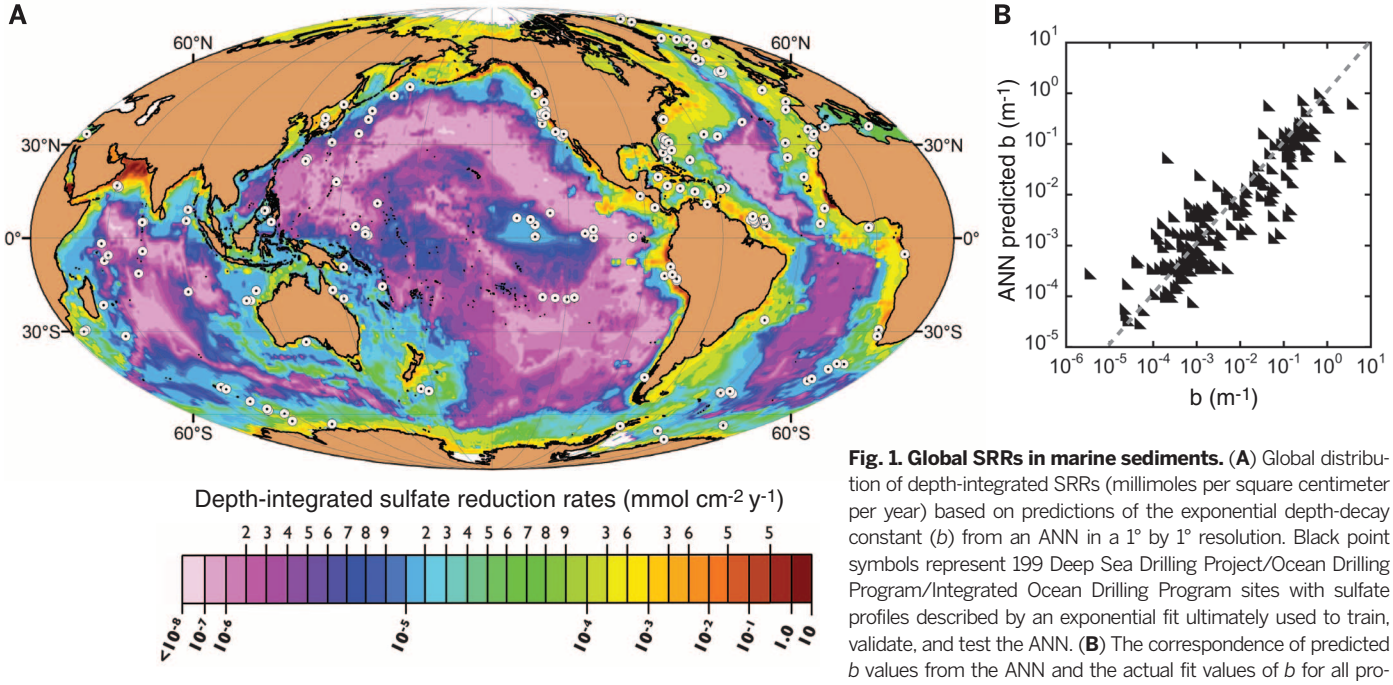


Fig. 1. Global SRRs in marine sediments. (**A**) Global distribution of depth-integrated SRRs (millimoles per square centimeter per year) based on predictions of the exponential depth-decay constant (*b*) from an ANN in a 1° by 1° resolution. Black point symbols represent 199 Deep Sea Drilling Project/Ocean Drilling Program/Integrated Ocean Drilling Program sites with sulfate profiles described by an exponential fit ultimately used to train, validate, and test the ANN. (**B**) The correspondence of predicted *b* values from the ANN and the actual fit values of *b* for all profiles (correlation coefficient $R = 0.88$).

rived from and mainly reflecting the distributions of primary productivity (*2*). Furthermore, these trends are consistent with observations in previous studies of global sulfate profiles from oceanic deep drilling programs (*5*), as well as global compilations of radiotracer gross SRR measurements (*15*). Although the ANN is trained and validated solely with deep-sea drilling data (Fig. 1B) (*12*), it replicates exponential sulfate depth-decay coefficients from several published short cores (<15 m) (fig. S4) taken below the shelf break and captures deep-sea locations characterized by high regional SRRs coupled to methane oxidation [e.g., 0.05 to 0.4 mmol cm^{-2} $year^{-1}$ in the Arabian Sea (*16*)].

From an environment-specific perspective, the ANN predicts a fourth of the previously estimated area-weighted and depth-integrated SRR for shelf sites: 0.097 mmol cm^{-2} $year^{-1}$ (*15*) versus 0.025 mmol cm^{-2} $year^{-1}$ (Table 1). All average depth-integrated rates determined here are considerably lower than previously reported (*15*) but appear to agree better with regionally averaged sulfate penetration depths. For example, assuming the previous upper continental slope average areal rate of 7.4×10^{-2} mmol cm^{-2} $year^{-1}$ (*15*), a tortuosity-corrected diffusion coefficient of 120 cm^2 $year^{-1}$, and a porosity of 0.8, the average sediment depth where sulfate depletes to 0.3 mM in sediments would be only 35 cm or 1.74 m, assuming a linear or an exponential profile, respectively. The calculations here yield an average depth-integrated rate for the upper slope of 1.7×10^{-2} mmol cm^{-2} $year^{-1}$. This rate would produce 0.3 mM sulfate concentrations at a depth of 1.2 or 20 m below sea floor (mbsf) for these same cases, respectively. The main reason for this difference is that the previously compiled global SRRs did not account for large sea-floor surface areas (e.g., ~70% of the continental shelf) consisting of organic-poor relict sands (*17*) and, thus, were averaged with a bias toward high-activity, organic-rich sites (*8*, *18*). This and other geochemical and depositional heterogeneities observed in coastal sediments may also explain some of the larger deviations between fitted and ANN-predicted *b* values for some shallow-water cores (Fig. 1B). For instance, order-of-magnitude variations in SRRs have been estimated within a single muddy basin [e.g., within 3000 km^2 in Arkona Basin (*19*)] and likewise for small gassy basins [e.g., 8 km^2 in Aarhus Bay (*20*)].

Table 1. Weighted average depth-integrated SRR for different water depths. The total area covered here is ~349 by 10^6 km^2 or ~97% of the total ocean.

Region (water depth)	Weighted average depth-integrated SRR ($\times 10^{-3}$ mmol cm^{-2} $year^{-1}$)	Total SRR (Tmol $year^{-1}$)	% Total SRR (Tmol $year^{-1}$)	Area ($\times 10^6$ km^2)
Inner shelf (0–50 m)	39.3	2.9	26	7.6
Outer shelf (50–200 m)	16.2	2.1	19	13
Slope (200–2000 m)	11.2	3.3	29	29
Rise (2000–3500 m)	4.4	2.7	24	63
Abyss (>3500 m)	0.09	0.2	2	237
Global total	3.2	11.3	100	349

The ANN-based global net SRR estimate (11.3 Tmol sulfate $year^{-1}$) (Table 1) is roughly 15% of previous estimates for gross SRR (*1*). Although calculating global net as opposed to gross SRR could explain this divergence, it is highly unlikely that gross SRRs account for more than 78% of the global organic carbon flux to the sea floor. In contrast, the 11.3 Tmol sulfate $year^{-1}$ predicted by the ANN is equivalent to 22.6 Tmol organic carbon $year^{-1}$ [assuming a 2 C–to–1 S stoichiometry (*5*, *12*)] or a more realistic 11 to 29% of the estimated global organic carbon flux to the sea floor.

This substantial diminution in global SRRs inherently affects previous conceptions of global microbial process distributions in sub–sea-floor sediments. Subsurface microorganisms largely depend on harvesting energy from the organic matter reaching the sea floor. This amount is minor in comparison to the carbon supplied to seawater prokaryotes via photosynthesis (4.3 Pmol C $year^{-1}$) (*21*). In spite of this sharp contrast in carbon availability, the marine subsurface total prokaryotic biomass is approximately equal to that of seawater (*8*).

Coupling the ANN-derived global SRR maps to global sub–sea-floor biomass maps (*8*) allows for the calculation of potential cell-specific rates, which can help us to further elucidate the activities of sulfate-reducing microorganisms across various global sedimentary environments. These microorganisms thrive in anoxic surficial sediments where sulfate and labile organic substrates coincide. Within inner-shelf sediments (<50 m water depth), which typically receive the highest inputs of labile organic matter, area-weighted SRR averages (Fig. 2, A and B) exhibited the highest rates (1.5 nmol cm^{-3} day^{-1}), leading to submicromolar sulfate concentrations by 6 mbsf. Furthermore, inner-shelf sediments comprise the highest prokaryotic cell abundances (Fig. 2C) and cell-specific rates (Fig. 2D) around 0.1 fmol $cell^{-1}$ day^{-1}. These data are in strong contrast to those of deep-water environments, which receive considerably less organic carbon. Peak SRRs in abyss sediments (>3500 m water depth) are a fraction

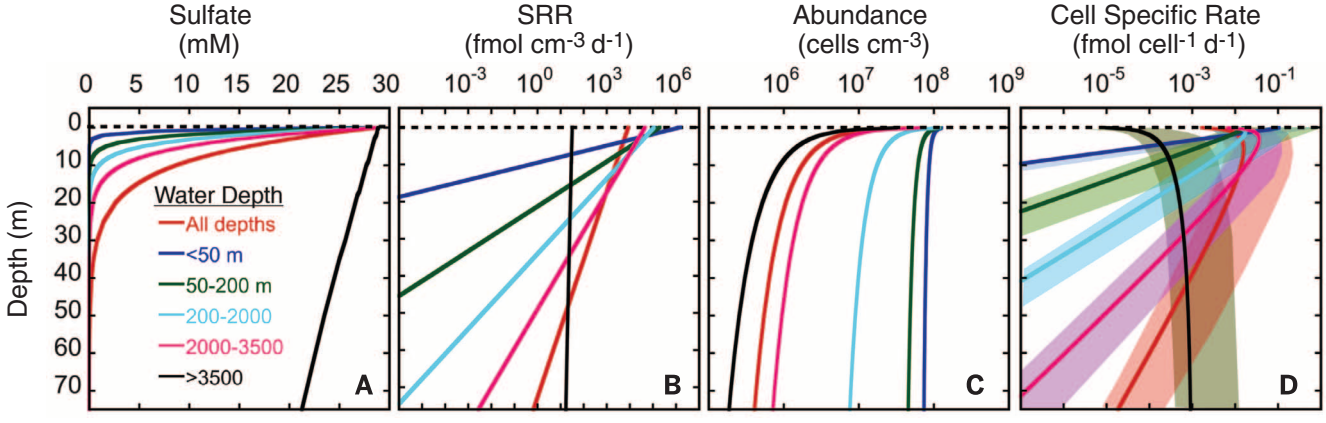


Fig. 2. Subsurface profiles of area-weighted parameters in various oceanic depth zones. (**A**) Sulfate profiles (millimolar), (**B**) SRRs (femtomoles per cubic centimeter per day), (**C**) cellular abundances (cells per cubic centimeter) (*8*), and (**D**) cell-specific rates (femtomoles per cell per day). Thick lines represent a 10% sulfate-reducing microorganism contribution to the total population (baseline scenario discussed in the text), whereas the shaded region for each line in (D) represents the range of 1 to 30% contribution of sulfate reducer to the total population.

Table 2. Global analysis of SRRs with respect to published organic carbon fluxes to the sea floor for various water depth environments. NR, concentration not reached. The global values in bold represent totals, whereas those in italics represent area weighted averages. SR, sulfate reduction.

Region (water depth)	Sediment depth to 0.1 mM sulfate (m)	% Prokaryotes above 0.1 mM sulfate	Organic C flux (*3*) (Tmol year^{-1})	% Organic C remineralized by SR	Organic C flux (*4*) (Tmol year^{-1})	% Organic C remineralized by SR
Inner shelf (0–50 m)	3.8	0.25	63*	16*	92	6.3
Outer shelf (50–200 m)	9.9	0.86			43	9.7
Slope (200–2000 m)	17	5.6			30	22
Rise (2000–3500 m)	26	31	17†	74†	26‡	22‡
Abyss (>3500 m)	NR	100				
Global total	*47*	*45*	**80**	**28**	**191**	**12**

*Values represent the total for the entire global shelf. †Values represent the total for the entire global area below the shelf break. ‡Values represent the total for the entire global area with water depths exceeding 2000 m.

of the shallow-water counterparts, at 0.03 pmol cm^{-3} day^{-1}. Furthermore, cell abundances are generally lower, with the cell-specific rates reaching a maximum of 9×10^{-4} fmol cell^{-1} day^{-1}.

Simultaneous measurements of SRRs and sulfate-reducing microorganism abundances are rare, but the existing data are consistent with our model (*22*–*26*). The majority of these data exist for relatively shallow-water, high-productivity sites (e.g., Aarhus Bay), with surficial cell-specific SRRs around 0.1 fmol cell^{-1} day^{-1} and reaching 1.0×10^{-3} fmol cell^{-1} day^{-1} by ~1 mbsf (*11*). These data are within the range of our areal weighted average for 0.1 mbsf of 0.1 fmol cell^{-1} day^{-1} (<50 m water depth) (Fig. 2). In our modeled shallow-water environments (i.e., inner and outer shelves), high SRRs lead to peak cell-specific rate values near the sediment-water interface. These cell-specific rates taper quickly to zero as sulfate becomes exhausted. Our results show that peak values for the slope also occur near the sediment-water interface (with a slight increase with sediment depth) and gradually decrease as sulfate approaches zero. The abyss, however, does not reach a peak cell-specific rate within the top 80 mbsf, and values remain one order of magnitude lower than the peak values for the other environments. Notably, for the assumed fractions of sulfate reducers within the total microbial community (1 to 30%), the general trends persist for cell-specific SRRs in different environments (Fig. 2D).

Results of the global survey show a distinct trend between environments in the continental margin and the abyss. The abyss (>3500 m water depth) is typically characterized by organic-poor sediments that allow for deep sulfate penetration. This prevalence of sulfate at great sedimentary depths indicates that most cells within the habitable deep sedimentary biosphere (down to 4000 mbsf or the specific basement depth) are found in either oxic or sulfate-reducing settings (Table 2). Nevertheless, within the other environments on the continental margins, sulfate is removed at comparatively shallow sediment depths (<100 m) (Table 2). The sulfate-methane transition zone (SMTZ) is a distinct geochemical horizon that represents an important transition from sulfate-reducing (overlying the SMTZ) to methanogenic sediments (underlying the SMTZ) (*1*, *19*, *20*). Limited data from deep-sea cores at these sites suggest that acetate and hydrogen can be abundant and thus serve as substrates for a vast methanogenic subsurface (*27*–*29*).

Collectively, these observations indicate that although the lack of reduced substrate limits sulfate reduction in deep-sea sediments, the continental margins harbor an expansive biosphere below the SMTZ where traditional, energy-rich electron acceptors are exhausted. Thus, this fraction of the microbial biosphere is largely fermentative and methanogenic (Table 2 and figs. S5 and S6). Roughly estimating the SMTZ at the depth at which sulfate depletion reaches 0.1 mM sulfate, habitable sediments located below the SMTZ would make up a total global subsurface volume of 10^8 km^3 (32% of total), hosting ~50% of the sub–sea-floor biomass (*12*). However, ~90% of cells in the subsurface at the continental margins (<3500 m water depth) would be situated below the SMTZ (*12*).

REFERENCES AND NOTES

1. B. B. Jørgensen, S. Kasten, in *Marine Geochemistry*, M. Zabel, H. Schulz, Eds. (Springer, Berlin, 2006), pp. 271–309.
2. D. E. Canfield, *Am. J. Sci.* **291**, 177–188 (1991).
3. P. Regnier *et al.*, *Nat. Geosci.* **6**, 597–607 (2013).
4. J. P. Dunne, J. L. Sarmiento, A. Gnanadesikan, *Global Biogeochem. Cycles* **21**, GB4006 (2007).
5. S. D'Hondt, S. Rutherford, A. J. Spivack, *Science* **295**, 2067–2070 (2002).
6. H. E. Hartnett, R. G. Keil, J. I. Hedges, A. H. Devol, *Nature* **391**, 572–575 (1998).
7. K. Wallmann *et al.*, *Energies* **5**, 2449–2498 (2012).
8. J. Kallmeyer, R. Pockalny, R. R. Adhikari, D. C. Smith, S. D'Hondt, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16213–16216 (2012).
9. R. J. Parkes *et al.*, *Nature* **436**, 390–394 (2005).
10. B. B. Jørgensen, A. Boetius, *Nat. Rev. Microbiol.* **5**, 770–781 (2007).
11. T. M. Hoehler, B. B. Jørgensen, *Nat. Rev. Microbiol.* **11**, 83–94 (2013).
12. See supplementary materials on *Science* Online.
13. G. C. Feldman, C. R. McClain, *Ocean Color Web*, AquaMODros. Inf. Serv., N. Kuring, S. W. Bailey, Eds. (NASA Goddard Space Flight Center, Greenbelt, MD, 2013); http://oceancolor.gsfc.nasa.gov/cgi/l3
14. S. Levitus, T. P. Boyer, *World Ocean Atlas 1994: Volume 4: Temperature*, NOAA Atlas NESDros. Inf. Serv. 4 (U.S. Government Printing Office, Washington, DC, 1994).
15. D. E. Canfield, E. Kristensen, B. Thamdrup, *Aquatic Geomicrobiology* (Academic Press, Amsterdam, 2005).
16. D. Fischer *et al.*, *Nat. Geosci.* **6**, 647–651 (2013).
17. K. O. Emery, *AAPG Bull.* **52**, 445–464 (1968).
18. J. M. Mogollón, A. W. Dale, J. B. Jensen, M. Schlüter, P. Regnier, *Geo-Mar. Lett.* **33**, 299–310 (2013).
19. J. M. Mogollón, A. W. Dale, H. Fossing, P. Regnier, *Biogeosciences* **9**, 1915–1933 (2012).
20. A. W. Dale *et al.*, *Geology* **37**, 235–238 (2009).
21. G. A. Knauer, in *Interactions of C, N, P and S Biogeochemical Cycles and Global Change*, NATO ASI Series I 4, 211-231, R. Wollast, F.T. Mackenzie, L. Chou, Eds. (Springer, Berlin, 1993).
22. K. Sahm, B. J. MacGregor, B. B. Jørgensen, D. A. Stahl, *Environ. Microbiol.* **1**, 65–74 (1999).
23. K. Ravenschlag, K. Sahm, C. Knoblauch, B. B. Jørgensen, R. Amann, *Appl. Environ. Microbiol.* **66**, 3592–3602 (2000).
24. J. Leloup *et al.*, *Environ. Microbiol.* **9**, 131–142 (2007).
25. J. Leloup *et al.*, *Environ. Microbiol.* **11**, 1278–1291 (2009).
26. L. Holmkvist, T. G. Ferdelman, B. B. Jørgensen, *Geochim. Cosmochim. Acta* **75**, 3581–3599 (2011).
27. V. B. Heuer, J. W. Pohlman, M. E. Torres, M. Elvert, K.-U. Hinrichs, *Geochim. Cosmochim. Acta* **73**, 3323–3336 (2009).
28. S. L. D'Hondt *et al.*, *Proceedings of the Ocean Drilling Program, Initial Reports*, *201* (Ocean Drilling Program, College Station, TX, 2003); 10.2973/odp.proc.ir.201.2003.
29. Y.-S. Lin *et al.*, *Geochim. Cosmochim. Acta* **77**, 186–201 (2012).

ACKNOWLEDGMENTS

All data used in this study are publically available via Janus and Pangea (www.pangea.de). We thank Deep Sea Drilling Project/Ocean Drilling Program/Integrated Ocean Drilling Program, Janus, and Pangea for compiling and providing these data sets. We thank A. Boetius, J. J. Middelburg, and S. Joye for helpful comments during the development of this work. Primary support for this work was provided by the Research Center/Cluster of Excellence "The Ocean in the Earth System" (MARUM) funded by the Deutsche Forschungsgemeinschaft in the framework of a Postdoctoral Fellowship awarded to M.W.B. Additional funding was provided by the European Research Council (ERC) under the European Union's Seventh Framework Programme–"Ideas" Specific Programme, ERC grant agreement no. 247153, the Helmholtz Association (AWI Bremerhaven), and Utrecht University through its strategic theme Sustainability, sub-theme Water, Climate, and Ecosystems.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/889/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
Table S1 to S4
References (*30*–*44*)

2 December 2013; accepted 29 April 2014
Published online 8 May 2014;
10.1126/science.1249213

DEEP EARTH

Melting of subducted basalt at the core-mantle boundary

Denis Andrault,[1]* Giacomo Pesce,[1] Mohamed Ali Bouhifd,[1] Nathalie Bolfan-Casanova,[1] Jean-Marc Hénot,[1] Mohamed Mezouar[2]

The geological materials in Earth's lowermost mantle control the characteristics and interpretation of seismic ultra–low velocity zones at the base of the core-mantle boundary. Partial melting of the bulk lower mantle is often advocated as the cause, but this does not explain the nonubiquitous character of these regional seismic features. We explored the melting properties of mid-oceanic ridge basalt (MORB), which can reach the lowermost mantle after subduction of oceanic crust. At a pressure representative of the core-mantle boundary (135 gigapascals), the onset of melting occurs at ~3800 kelvin, which is ~350 kelvin below the mantle solidus. The SiO_2-rich liquid generated either remains trapped in the MORB material or solidifies after reacting with the surrounding MgO-rich mantle, remixing subducted MORB with the lowermost mantle.

Numerous seismological studies have demonstrated the complexity of the lowest 150- to 300-km-thick mantle layer situated just above the core-mantle boundary (CMB). In many areas, there is an intermittent stratification, with 1.5 to 3% velocity discontinuities, as well as lateral shear-wave anisotropy (*1*). These anomalies could arise from mineralogical heterogeneities (*2*), magma ocean crystallization (*3*), the descent of subducted slabs deep into the lower mantle (*4*), and/or chemical reactions with the outer core. In addition, ultra–low velocity zones (ULVZs), with shear-wave velocity reduction of more than 10%, have been detected in specific mantle regions (*5*). Their size is limited to ~40 km thickness and ~100 km across, and they are ~10% denser than the surrounding mantle. These ULVZs could be due to partial melting occurring in a steep temperature gradient when approaching a very hot outer core. Between 5 and 30% partial melting could attenuate *P* and *S* seismic wave velocities to a similar amplitude as that reported for ULVZs (*6*). However, it was recently shown that partial melting of the bulk lower mantle cannot produce a residue with an equilibrium partial melt that is sufficiently dense to adequately represent the ULVZ mush (*7*).

Partial melting of pyrolitic or chondritic mantle would be possible if the CMB temperature were higher than ~4150 K (*8*, *9*). Although such a high temperature is not precluded, it would require a very hot core. It is more than 1000 K above the most recent determination of the mantle adiabat extrapolated to the CMB (*10*). Moreover, the presence of a very hot core today makes it difficult to explain how a geodynamo could have been maintained for prolonged geological periods: An even hotter core would be required in the past, or an extremely high concentration [400 to 800 parts per million (ppm)] of K in the core (*11*). Alternatively, partial melting in the lowermost mantle could occur below 4150 K if the mantle composition were drastically altered. Chemical differences between peridotite, pyrolite, and chondritic-type mantle would produce only minor variations, because they contain similar mineralogical assemblages. However, a local concentration of fusible (such as alkalis) or volatile (H_2O or CO_2) elements, together with a change in mineralogy, could have a marked effect on the solidus temperature. In particular, a small water concentration can be enough to reach the saturation limit of the bulk mantle (*12*). Consequently, the melting curve of hydrous pyrolite has much lower melting temperatures (*13*).

We investigated the melting behavior of a natural mid-oceanic ridge basalt (MORB) collected at a 2800-m depth during the Searise-1 research cruise (table S1). When a MORB eventually reaches Earth's lower mantle, it is composed of four coexisting phases: (i) Fe-rich silicate perovskite (Mg-Pv); (ii) free silica, in contrast with free (Mg, Fe)O ferropericlase (Fp) in the bulk mantle (fig. S1); (iii) Ca-bearing silicate perovskite (Ca-Pv); and (iv) enough Al, Ca, Na, and K to generate cage structures such as calcium ferrite (CF), hollandite, or an aluminous phase (*14*–*16*). Our starting material contained ~0.3 weight % H_2O and traces of CO_2 (fig. S9), which are common volatile concentrations for natural MORBs. At lower mantle pressures, the basaltic portion of the slab contains minerals that can carry water, such as the Al-bearing stishovite (*17*), δ-AlOOH (*18*), and phase H (*19*).

High pressures and temperatures were provided by a laser-heated diamond anvil cell [fig. S2 (*8*)]. The sample behavior was followed continuously using in situ x-ray diffraction (Fig. 1 and supplementary materials). Upon heating, we first observed the appearance of a continuous diffraction ring on the charge-coupled device detector, at d_{hkl} distances typical of Mg-Pv. This

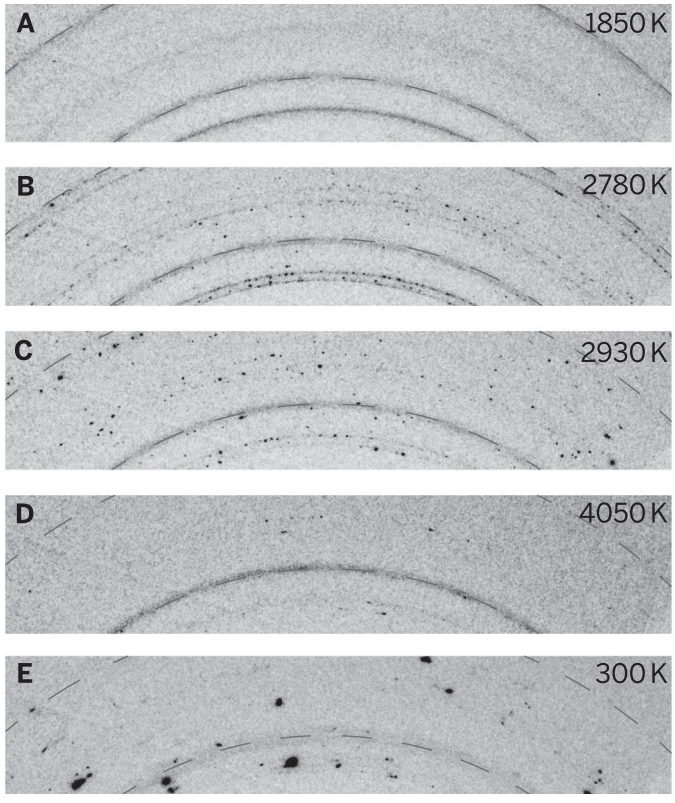


Fig. 1. Sequence of x-ray diffractions recorded with increasing temperatures from 1850 to 4050 K at 56 GPa. Diffraction peaks of the KCl pressure medium are superimposed in blue. Upon heating, (**A**) we first observed the appearance of weak Mg-Pv diffraction peaks, (**B**) followed by crystallization of all other phases, including Ca-Pv and stishovite. Above 2930 K, (**C**) diffraction peaks become suddenly much sharper. Also, all diffraction peaks for a given sample drastically change position on images recorded successively at the same pressure and temperature conditions. Above 4050 K, (**D**) only a few diffraction peaks remain visible. The quench image (**E**) contains several new peaks, showing that the lack of diffraction peaks at the highest temperature was not due to a loss of sample, but rather to complete sample melting.

[1]Laboratoire Magmas et Volcans, Université Blaise Pascal, CNRS, IRD, Clermont-Ferrand, France. [2]European Synchrotron Radiation Facility, Grenoble, France.
*Corresponding author. E-mail: denis.andrault@univ-bpclermont.fr

ring occurred above 1800 and 2500 K, at 40 and 135 GPa, respectively. The relatively large peak profile is typical of a powder with very small grain size (Fig. 1A). At higher temperatures, Ca-Pv and silica appeared simultaneously (fig. S3 and table S2). The latter adopts either the structure of rutile (stishovite), $CaCl_2$, or α-PbO_2 (seifertite), as a function of pressure (fig. S4). Diffraction peaks of the post-perovskite phase of $MgSiO_3$ also appeared above 115 GPa. Diffraction features typical of CF and/or the so-called new aluminous phase remained weak because of major overlaps of their diffraction peaks with those from the other phases present in the sample (*15*).

Less than 100 K above the temperature at which Ca-Pv and silica crystallized from the glass, the initially continuous diffraction rings evolved suddenly into a discontinuous juxtaposition of spots (Fig. 1B), and the peak shape sharpened drastically, showing a net discontinuity in the rate of grain growth. At slightly higher temperatures, diffraction images acquired repeatedly at the same temperature, with a time interval of 20 s (acquisition time), showed large diffraction spots in totally different positions, indicating fast grain rotation. These changes in the sample properties were associated with a flattened temperature profile, although laser power was increased continuously (fig. S5). This effect could result from a change in laser absorption due to the structural transformation (*20*). Our results show that the temperature gap between rapid grain growth and fast grain rotation is less than 200 K (fig. S3). In theory, the extensive grain rotation should take place close to the solidus temperature. The sudden change in sample behavior is most probably due to the appearance of a small amount of liquid at grain boundaries. Because the very first degree of melting can be difficult to detect, we bracketed the solidus using these two major criteria—rapid grain growth and fast grain rotation—where the solidus temperature profile should plot (fig. S3).

The solidus increases continuously from about 2100 (±150) to 3200 (±150) K with an increase in pressure from 20 to 80 GPa (Fig. 2). At the latter pressure, which corresponds to a mantle depth of 1900 km, the MORB solidus is identical to that of a chondritic-type mantle (*8*). With a further increase in pressure to 135 GPa, the MORB solidus temperature increased to 3800 (±150) K, whereas that of the bulk mantle increased to 4150 (±150) K. The flattening of the MORB solidus could be linked to a change in liquid composition with pressure. Extrapolation of our solidus melting curve to ambient pressure yields a melting temperature of ~1300 K, in agreement with previous studies (*21*). However, we could not reproduce the rapid increase of solidus temperature at the low pressures observed in previous studies using a large-volume press (*21*, *22*). Because analyses were performed on quenched samples, for which it is difficult to observe low degrees of partial melting, we believe that the solidus temperature was possibly overestimated in these previous studies. Our solidus curve also plots at much lower temperatures than a previous study performed up to 65 GPa, using a laser-heated diamond anvil cell (*22*). In that study, the melting criterion based on the change of temperature with laser power, for a MORB sample sandwiched between two Re foils, may have hampered detection of the solidus. The maximum discrepancy between the different studies is ~400 K and tends to decrease with increasing pressure above 20 GPa. At 60 GPa, the discrepancy is down to ~250 K between the previous optical (*22*) and our in situ measurements.

To verify that the 350 K melting temperature reduction between MORB and the chondritic-type mantle was not an experimental artifact, we loaded both compositions in the same pressure chamber (fig. S2). Using the same melting criteria as in a previous study of mantle melting (*8*), we observed the solidus of chondritic mantle at temperatures of 2980 (±150), 3450 (±150), and 3680 (±150) K, for pressures of 58, 85, and 107 GPa, respectively, falling within ±60 K of our previous study (*8*). Given the higher amount of fusible elements in MORB, one could have expected a lower solidus temperature as compared to chondritic mantle. However, the role of SiO_2 excess in MORB and Fp excess in the mantle cannot be neglected (fig. S1).

Upon further increase of the laser power, the sample temperature eventually rose above the plateau temperature, allowing us to record diffraction patterns up to ~5000 K. We rarely observed the total disappearance of all diffraction peaks from the sample. Instead, diffraction images often contained a couple of diffraction peaks appearing at random azimuthal positions on the images (Fig. 1D). Convection in the liquid sample could have induced some crystallization at the sample/KCl interface over a very short time scale. Also, the sample temperature was probably slightly below the liquidus temperature, because we observed a clear rim of solid phase just around the melt. We define the liquidus as the temperature at which complete loss of the continuous structure of the diffraction lines was achieved. This temperature corresponds to a clear loss of the three-dimensional solid structure of the sample, with free rotation of a couple of grains in a predominantly liquid fraction.

The temperature gap between the solidus and liquidus is found to increase progressively from 1500 to 2500 K with increasing pressure from 40 to 140 GPa. This gap is much larger than that reported for the chondritic-type mantle. It results in a liquid composition at the solidus, which is very different from the bulk MORB composition (*23*).

Three lines of evidence point to an increase of the SiO_2 content in the liquid with increasing pressure during partial melting of MORB. First, the MORB solidus curve always plots 1000 to 1500 K below the lowest melting curve for pure $CaSiO_3$, $MgSiO_3$, or SiO_2 phases (*24*–*26*) (Fig. 2). Although SiO_2 is the most refractory of these MORB phases at low pressures, it becomes the least refractory above 60 GPa. This suggests a displacement of the pseudo-eutectic composition from Mg-rich liquid (with silica phase on the liquidus) toward Si-rich liquid (with Ca-Pv and/or Mg-Pv on the liquidus) in the (Mg-Pv,

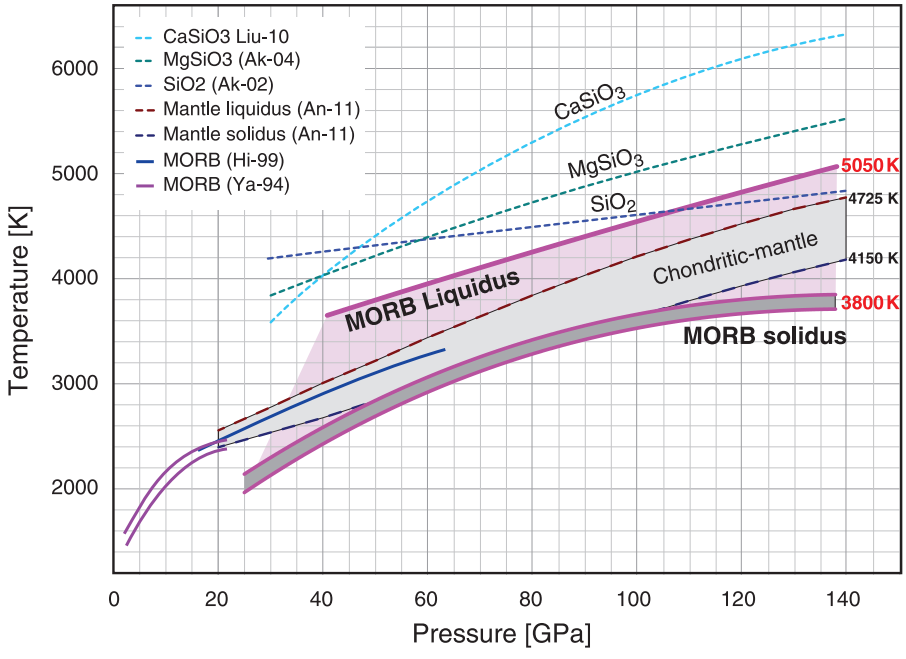


Fig. 2. The MORB solidus and liquidus compared with melting curves of the bulk mantle. At the CMB boundary, the MORB solidus and liquidus reach 3800 (±150) K and 5050 (±300) K, respectively (fig. S3). The curvature of the MORB solidus is much more pronounced than that of chondritic-type mantle (*8*), which induces a lower melting point for the former at 135 GPa, the CMB pressure. The MORB solidus also plots 1000 to 1500 K below the melting curves of pure SiO_2 and $MgSiO_3$, as determined from shockwave experiments (*21*, *22*, *24*–*26*).

Ca-Pv, SiO_2) ternary diagram (fig. S1). A second argument is based on the last phase(s) to disappear from the diffraction patterns at temperatures approaching those of the liquidus. At pressures lower than ~70 GPa, clear peaks of SiO_2 and $CaSiO_3$ remained visible on the diffraction patterns at temperatures between the solidus and liquidus. At higher pressures, SiO_2 peaks disappeared and instead those of Mg-Pv appeared (fig. S6). This indicates a sample undersaturated in SiO_2, with Mg-Pv and Ca-Pv grains coexisting with the liquid at the highest pressures.

A third argument comes from mineralogical and chemical analyses of the recovered samples. After the laser was shut down, we maintained the pressure and made diffraction maps of the quenched molten region of the samples. This corresponds to regions of 30 × 30 or 40 × 40 μm^2, with step resolutions of 2 or 3 μm. The diffraction maps can be treated in series to extract the phase content as a function of the position within the sample (fig. S7). This type of quenched sample generally shows a rim of liquidus phase(s) around the quenched liquid (*13*, *27*). The mineralogical maps of our samples show clearly that the quenched liquid was depleted in the SiO_2 phase and concentrated in Mg-Pv at pressures below ~70 GPa, and the opposite at higher pressures. This indicates an increasing SiO_2 content in the liquid with increasing pressure (fig. S1).

We then made chemical maps of our recovered samples, using a scanning electron microscope (SEM) (supplementary materials, Fig. 3, and fig. S8). At pressures below ~70 GPa, we observed a large amount of MgO in the center of the sample where the liquid had solidified, and the liquid was surrounded by a SiO_2-rich rim. At intermediate pressures, the maps showed similar MgO and SiO_2 content in the liquid as compared to the unheated part of the sample. At the highest pressures, the quenched liquid was enriched in SiO_2 whereas the outer rim was enriched in MgO (fig. S8). MORB partial melting in the lower mantle should always produce liquids with (Mg + Fe + Ca)/Si ratios less than 1, because MORB contains free SiO_2 in addition to Mg-Pv and Ca-Pv refractory phases.

Following the hypothesis that the ULVZ at the D″ layer is associated with partial melting, our experimental results provide explanations for the seismological features. First of all, the nonubiquitous nature of the ULVZ could be due to specific regions where MORB slabs have reached the CMB. The thicknesses of the ULVZ (~40 km) and MORB portion of the subducted slabs [~6 km (*28*)] are of the same order of magnitude. Mantle resistance to the slab penetration, or folding and piling-up at the CMB, could certainly induce thickening of the MORB layer. If temperature in the D″ region is between 3800 and 4150 K, the MORB should be the only material to undergo partial melting (the harzburgitic layer of the subducted slab being more refractory than the pyrolitic or chondritic-type mantle). Because of the large temperature gap found between the MORB solidus and liquidus, low degrees of partial melting are expected in the D″ temperature range. If such liquid eventually percolates and diffuses out of the MORB pile, it will react with the excess (MgFe)O Fp present in the surrounding lower mantle to form $(Mg,Fe)SiO_3$ Mg-Pv. For this reason, the mobility of the liquids extracted from the MORB piles must be extremely limited. Thus, it is not likely that large-scale pockets of liquid will form. It would also explain why seismic shear waves (*Vs*) can propagate through the partially molten ULVZ (*5*).

Also, the decrease of Fp content in the lowermost mantle, resulting from reaction between the bulk mantle and SiO_2-rich liquid, would drive mantle composition toward the perovskitic end-members, in agreement with a recent report (*29*). The loss of Si from the MORB would drive it toward perovskitic end-members (fig. S1). The reaction would eventually stop when the SiO_2 excess in the MORB was exhausted. This would result in the disappearance of the MORB signature, except for minor and trace elements, which could remain concentrated around the ULVZ. This scenario would lead to lower mantle homogenization, in contrast to the chemical segregation generally induced by mantle partial melting. Alternatively, if the MORB proportion is high, Fp could become reactively exhausted, which would imply a lowermost mantle saturated in SiO_2. The answer to these two alternative scenarios depends on how well the underlying harzburgite remains attached to the MORB layer and is entrained into the lowermost mantle. If it is sufficiently abundant, then its reaction with the liquid originating from MORB melting should produce a typical mantle composition such as pyrolite.

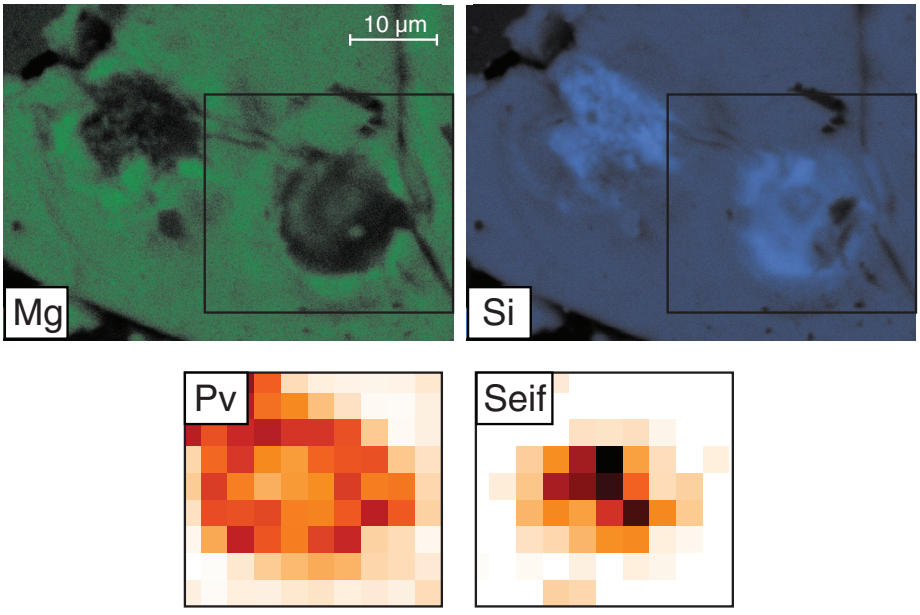


Fig. 3. Phase relations in the sample recovered after partial melting at 120 GPa. The chemical maps (Mg and Si) and mineralogical maps [Mg-Pv and seifertite (Seif)] measured by SEM and x-ray diffraction, respectively, show higher SiO_2 and seifertite contents and lower MgO and Mg-Pv contents in the sample region corresponding to the quenched liquid.

REFERENCES AND NOTES

1. T. Lay, Q. Williams, E. J. Garnero, *Nature* **392**, 461–468 (1998).
2. M. Murakami, K. Hirose, N. Sata, Y. Ohishi, *Geophys. Res. Lett.* **32**, L03304 (2005).
3. S. Labrosse, J. W. Hernlund, N. Coltice, *Nature* **450**, 866–869 (2007).
4. R. D. van der Hilst, H. Karason, *Science* **283**, 1885–1888 (1999).
5. L. Wen, D. V. Helmberger, *Science* **279**, 1701–1703 (1998).
6. S. Rost, E. J. Garnero, Q. Williams, M. Manga, *Nature* **435**, 666–669 (2005).
7. C. W. Thomas, P. D. Asimow, *J. Geophys. Res. Solid Earth* **118**, 5738–5752 (2013).
8. D. Andrault *et al.*, *Earth Planet. Sci. Lett.* **304**, 251–259 (2011).
9. G. Fiquet *et al.*, *Science* **329**, 1516–1518 (2010).
10. T. Katsura, A. Yoneda, D. Yamazaki, T. Yoshino, E. Ito, *Phys. Earth Planet. Inter.* **183**, 212–218 (2010).
11. T. Nakagawa, P. J. Tackley, *Geochem. Geophys. Geosyst.* **11**, Q06001 (2010).
12. N. Bolfan-Casanova, H. Keppler, D. C. Rubie, *Geophys. Res. Lett.* **30**, 1905 (2003).
13. R. Nomura *et al.*, *Science* **343**, 522–525 (2014).
14. T. Irifune, A. E. Ringwood, *Earth Planet. Sci. Lett.* **117**, 101–110 (1993).
15. A. Ricolleau *et al.*, *J. Geophys. Res. Solid Earth* **115**, B08202 (2010).
16. N. Miyajima *et al.*, *Am. Mineral.* **86**, 740–746 (2001).
17. A. R. Pawley, P. F. McMillan, J. R. Holloway, *Science* **261**, 1024–1026 (1993).
18. A. Sano *et al.*, *Geophys. Res. Lett.* **35**, L03303 (2008).
19. M. Nishi *et al.*, *Nat. Geosci.* **7**, 224–227 (2014).
20. Z. M. Geballe, R. Jeanloz, *J. Appl. Phys.* **111**, 123518 (2012).
21. A. Yasuda, T. Fujii, K. Kurita, *J. Geophys. Res. Solid Earth* **99**, 9401 (1994).
22. K. Hirose, Y. W. Fei, Y. Z. Ma, H. K. Mao, *Nature* **397**, 53–56 (1999).
23. P. D. Asimow, M. M. Hirschmann, E. M. Stolper, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **355**, 255–281 (1997).

24. J. A. Akins, S. N. Luo, P. D. Asimow, T. J. Ahrens, *Geophys. Res. Lett.* **31**, L14612 (2004).
25. J. A. Akins, T. J. Ahrens, *Geophys. Res. Lett.* **29**, 31-1–31-4 (2002).
26. Z. J. Liu *et al.*, *Solid State Commun.* **150**, 590–593 (2010).
27. D. Andrault *et al.*, *Nature* **487**, 354–357 (2012).
28. R. S. White, T. A. Minshull, M. J. Bickle, C. J. Robinson, *J. Petrol.* **42**, 1171–1196 (2001).
29. M. Murakami, Y. Ohishi, N. Hirao, K. Hirose, *Nature* **485**, 90–94 (2012).

ACKNOWLEDGMENTS

We thank M. Benbakkar, J. L. Devidal, A. Fabbrizio, G. Garbarino, P. Parisiadis, V. Svitlyk, and F. van Wyk de Vries for their help and three anonymous reviewers for fruitful comments. The MORB starting material was kindly provided by P. Schiano. This work is supported by Institut National des Sciences de l'Univers, l'European Synchrotron Radiation Facility, and the project Oxydeep of the Agence Nationale de la Recherche. This is Laboratory of Excellence ClerVolc contribution no. 99. All experimental data are presented in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/892/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 and S2
References (*30*–*34*)

6 January 2014; accepted 3 April 2014
10.1126/science.1250466

CORALS AND CLIMATE

Mechanisms of reef coral resistance to future climate change

Stephen R. Palumbi,* Daniel J. Barshis,† Nikki Traylor-Knowles, Rachael A. Bay

Reef corals are highly sensitive to heat, yet populations resistant to climate change have recently been identified. To determine the mechanisms of temperature tolerance, we reciprocally transplanted corals between reef sites experiencing distinct temperature regimes and tested subsequent physiological and gene expression profiles. Local acclimatization and fixed effects, such as adaptation, contributed about equally to heat tolerance and are reflected in patterns of gene expression. In less than 2 years, acclimatization achieves the same heat tolerance that we would expect from strong natural selection over many generations for these long-lived organisms. Our results show both short-term acclimatory and longer-term adaptive acquisition of climate resistance. Adding these adaptive abilities to ecosystem models is likely to slow predictions of demise for coral reef ecosystems.

Reef-building corals have experienced global declines resulting from bleaching events sparked by pulses of warm-water exposure (*1*–*4*). However, corals in naturally warm environments can have high resistance to bleaching temperatures and can survive heat exposure that would bleach conspecifics in cooler microclimates (*5*, *6*). Similarly, recent discovery of populations of acidification-resistant corals show that physiological or evolutionary mechanisms of environmental accommodation exist (*7*, *8*). Such populations are ideal test sites for research into the mechanisms of coral response to climate change.

Corals in adjacent backreef pools in the U.S. National Park of American Samoa on Ofu Island experience strong differences in temperature (*9*, *10*). In the highly variable (HV) pool, temperatures often exceed the local critical bleaching temperature of 30°C, reaching 35°C during strong noontime low tides (*6*). By contrast, the moderately bariable (MV) pool rarely experiences temperatures above 32°C. Corals in the HV Pool have higher growth rates (*9*, *10*), higher survivorship, and higher symbiont photosynthetic efficiency during experimental heat stress than conspecifics from the MV pool (*6*). These pools provide a powerful system to test the speed and extent of coral acclimatization and adaptation to warm-water conditions in the context of future climate change.

Department of Biology, Stanford University, Hopkins Marine Station, Pacific Grove, CA 93950, USA.
***Corresponding author. E-mail: spalumbi@stanford.edu** †Present address: Department of Biological Sciences, Old Dominion University, Norfolk, VA 23529, USA.

To test corals in their native habitats for physiological resistance to heat stress, we collected branches of the tabletop coral *Acropora hyacinthus* [cryptic species E (*11*)] and exposed them to experimental bleaching conditions. *A. hyacinthus* is a cosmopolitan species that constitutes a large percentage of hard coral cover on Pacific reefs and shows high levels of bleaching and mortality during large-scale bleaching events (*4*). We chose *A. hyacinthus* for this study because it is a dominant reef-builder and is especially sensitive to environmental stress, making its relative ability to acclimate or adapt extremely important to the future of coral reef ecosystems as climate change proceeds. We subjected branches of corals to a prescribed ramp in water temperature of 29° to 34°C for 3 hours, followed by an incubation for 3 hours at 34°C. These conditions mimic the natural increase in temperature observed in the HV pool during a tidal cycle. Experiments on fragments of tagged and monitored colonies showed that individuals native to the HV pool exhibit higher resistance to thermal stress, measured by retention of chlorophyll derived from photosynthetic symbionts, than corals from the MV pool (Fig. 1). The average retention of chlorophyll a after experimental heat stress was 80% in HV pool corals (Fig. 1C) but only 45% in MV pool corals (Fig. 1A, *t* test, $P < 0.00001$) compared with controls.

To test for acclimatization, we transplanted coral colonies of *A. hyacinthus* reciprocally from their native locations in the HV and MV pools to three transplant sites within each pool. We transplanted 6 colonies from the HV pool and 12 from the MV pool. After 12, 19, and 27 months, we tested transplanted colonies for thermal resistance. For 11 separate colonies, 22 of 23 paired bleaching

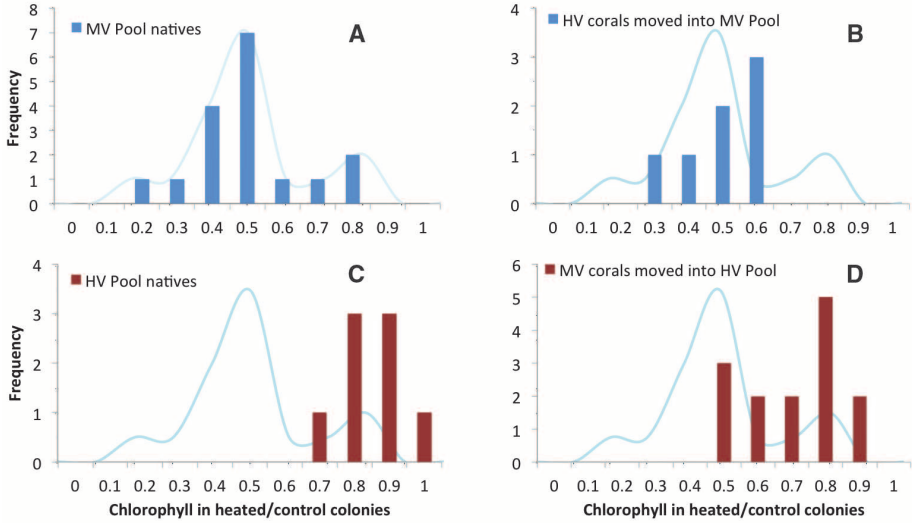


Fig. 1. Chlorophyll retention in coral colonies exposed to experimental heat stress compared with nonstressed controls. Upper panels show results from corals native to the moderately variable (MV) pool (**A**) and from corals moved into the MV pool (**B**). The lower panels are from corals native to the highly variable (HV) pool (**C**) and from corals moved into the HV pool (**D**). The smoothed curve reflects the distribution in (A) and is included in other panels for reference.

experiments show that corals acquired at least part of the heat sensitivity of the pool they were transplanted into. The experiments showed higher chlorophyll a retention during heat stress in colonies transplanted to the HV pool than in the same colony transplanted to the MV pool ($P < 0.0001$, paired t test, Fig. 2). Bleaching resistance did not vary with the time of transplant or season ($P > 0.80$).

Although the MV pool corals acquired heat resistance when moved to the HV pool, they did not achieve the resistance of native HV corals. MV pool corals transplanted to the HV pool retained less chlorophyll in bleaching experiments than corals native to the HV pool (compare Fig. 1, C and D; i.e., 67.5% and 80%, respectively; Student's t test, $P < 0.05$). By contrast, HV pool corals transplanted into the MV pool dropped their chlorophyll a retention to the same level as that of the MV pool natives (compare Fig. 1, A and B; i.e., 47% and 45%, respectively).

To understand the physiological changes associated with acquired heat resistance, we investigated gene expression in reciprocally transplanted corals. Six corals were sampled at noon on 30 August 2011, each from reciprocal transplant sites in the HV and MV pools. Reads from 12 mRNA extractions (one from each coral from each pool) were mapped to the *A. hyacinthus* assembly developed by Barshis *et al.* (*12*). Data pipelines and statistical procedures followed those of De Wit *et al.* (*13*). Transcriptome profiles of transplanted colonies show strong evidence for acclimatization. A two-way analysis of variance (ANOVA) of 16,728 coral genes detected 74 that changed significantly when comparing genetically identical coral fragments between the two pools [average 3.4-fold expression difference, false discovery rate (FDR)–corrected P value < 0.05, table S3). Among the 55 contigs with annotations, we found several transcription factors and cell signaling proteins (14 contigs), heat shock and chaperonin proteins (9 contigs), tumor necrosis factor receptor (TNFR)–associated factor (TRAF)–type proteins, cytochrome P450, and fluorochromes (Fig. 3A) that appeared to be involved in heat acclimation.

The ANOVA test also highlights 71 contigs with differential expression depending on the origin of colonies (two-way ANOVA experiment-wide FDR $P < 0.05$, table S2). These genes showed differences in expression levels depending strictly on the pool of origin, not on the final transplant site. In our previous work, members of the TNFR superfamily, which is involved in eukaryotic immune function and apoptosis (*14*, *15*), were constitutively up-regulated in native HV pool corals (*13*). In the current data set, one *TNFR* gene exhibited a 9.9-fold difference in expression in coral colonies native to the HV pool, whether they were living in the HV or MV pools (Fig. 3B), when compared with corals native to the MV pool. These results indicated that mediators of coral thermal resistance have fixed constitutive expression levels in either pool, perhaps representing signs of genetically based local adaptation.

Our transcriptome data also allowed us to test for changes in the proportions of symbiotic algae during coral acclimatization. *Symbiodinium* clades C and D are common in corals in American Samoa, and corals in warmer microclimates tend to have clade D symbionts (*16*). Though most colonies are dominated by one or the other of these clades, all colonies that we have tested host background populations of the other clade (*17*). To determine if coral host acclimation was accompanied by changes in symbiont proportion, we estimated the proportion of clade C and D by counting the transcriptome reads that mapped exclusively to small artificial test contigs of *ITS1*, *ITS2*, and the chloroplast *23S* gene, which distinguish clade C and D *Symbiodinium* (*17*). In our experiments, symbiont type explained a negligible fraction of the variation in bleaching resistance in common garden conditions (squared correlation coefficient R^2 = 0.15 and 0.06 in the HVP and the MVP transplants, respectively, $P > 0.30$, fig. S2). Our data also showed little shift in clade C versus clade D proportions as a result of transplantation. When clade C–dominated MV pool corals are moved to the HV pool, their proportion of clade D shifts from <1% to about 2%. Likewise, clade D–dominated HV corals moved to the MV pool maintain only about 4% clade C symbionts (Fig. 4). Additionally, there were no gene expression changes in *Symbiodinium* between transplants to the different pools, suggesting that little acclimation by symbionts occurred in the altered environmental conditions (*18*).

These experiments showed that some corals are capable of broad acclimatization to microclimate and developed enhanced resistance to bleaching without changing symbionts. We can place this capacity for acclimatization in an evolutionary context by comparing the natural phenotypic difference that we find between pools with the shift due to acclimatization after transplantation. Phenotypic change in evolutionary biology is often measured by the intensity parameter I, which is defined as the change in mean phenotype before versus after natural selection, divided by the standard deviation (SD) (*19*). In our case, we applied this concept to the spatial differences between pools instead of to temporal differences. The average phenotypic change from corals native to the HV pool versus the MV pool was 0.35 with a SD of 0.14 (compare Fig. 1, A and C). Thus, the phenotypic shift between pools was 2.5 SDs.

Here, I measures the difference in phenotype between coral populations subjected to different environments and is affected by phenotypic change caused by the fixed effects between the microclimates (denoted I_F), as well as by the acclimatization of individuals (denoted I_A), $I = I_F + I_A$. In this case, fixed effects include evolutionary adaptation, shifts in symbiont type, developmental changes, and epigenetics. Our transplant experiment allowed us to measure I_A as the average change in mean phenotype among acclimating individuals (0.214) divided by the SD (0.137), resulting in $I_A = 1.56$ (Fig. 1, A versus B and C versus D). This in turn allowed us to estimate I_F to be 0.94 SD units ($= I - I_A$).

These estimates indicated that these corals did acclimate to higher temperatures. The change in phenotype due to acclimatization (1.56 SD units) was similar to, but higher than, the change that we estimated is caused by fixed effects between pools (0.94 SD units).

We have provisionally extended this analysis to gene expression to illustrate the dual roles of acclimation and adaptation in gene expression shifts. There are 141 contigs for which corals native to the HV pool showed expression levels that differed by more than 1 SD from those of corals native to the MV pool and that were significant in the above two-way ANOVA analysis. Considering gene expression as a phenotype, these loci had an intensity $I > 1.0$. We also measured the contribution of acclimation to this phenotypic change by measuring gene expression in MV pool corals in the HV pool and vice versa. For these 141 contigs, the change caused by acclimation (I_A) averaged 42% of the phenotypic difference (fig. S4 and table S4), with the rest being caused by fixed effects (I_F). Loci with strong fixed effects included *TNFR* (as above) and a series of cellular transport loci with unknown function in corals. As in the

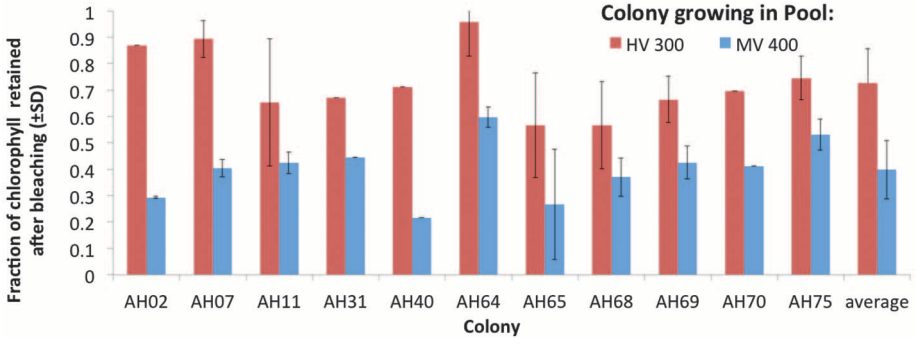


Fig. 2. Response of corals to reciprocal transplantation. The degree of bleaching experienced by a coral colony is measured by the ratio of chlorophyll that remains in experimentally heat-stressed colonies (N = 2 to 4 replicates per colony) compared to non–heat-stressed controls. Data are from 11 transplants into both the MV pool (blue) and HV pool (red). Error bars are SDs for colonies that were stressed-tested on two to three separate dates each.

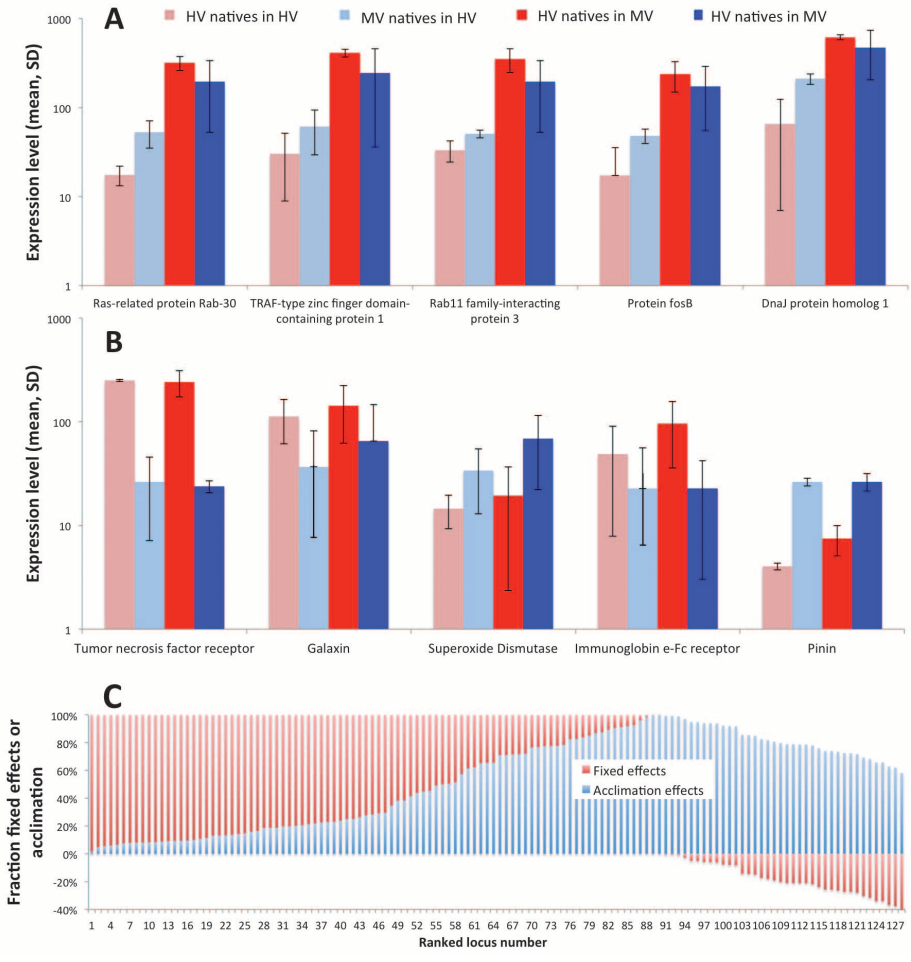


Fig. 3. Changes in gene expression between colonies on the basis of their transplant site (A) or original location (B). Light red represents corals native to the HV pool sampled after living in the HV pool. Light blue: MV natives sampled in MV pool. Dark red: HV natives in MV pool. Dark blue: MV natives in MV pool. Error bars are SDs. (**C**) The fraction of the gene expression difference between pools that is explained by fixed effects (red columns) versus acclimation (blue columns) for 128 loci that have large differences in expression between natives of the MV and HV pools. These 128 contigs correspond to the loci in table S4 and are listed by increasing level of acclimation between pools. Negative fixed effects occur when acclimation of transplanted corals exceeds the differences seen between natives. In our data, this occurs largely (33 of 35 cases) because transplanted HV pool corals in the MV pool have higher gene expression than MV pool natives.

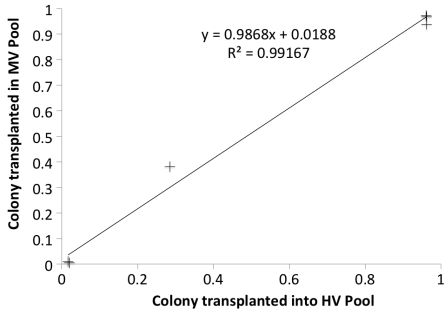


Fig. 4. High similarity of symbiont clade D proportion in six coral colonies transplanted to each of the HV pool and MV pool after 12 months of post-transplant growth.

case of bleaching phenotypes, fixed effects could be due to adaptive evolution, but could also be affected by developmental changes, epigenetics, and symbiont type. It is also possible there are residual effects that might eventually be erased by longer residence time. However, our experiments were conducted on branches that grew in situ after transplantation, and so our data are from tissues that did not experience the native pool environment. Loci with strong components of acclimation include *TRAF*, the signaling transducers for TNFR, as well as proteins encoded by *Ras* and *Rab*, transcription factors, and heat shock proteins.

The possibility that corals can acclimate to local differences has been suggested for decades: Conspecific corals at different latitudes show bleaching temperatures 1° to 2°C above local mean summer maximum sea-surface temperature (*20*, *21*), despite substantial differences in mean summer maximum temperatures. However, the mechanism generating this pattern has not been rigorously tested. Early analyses of the threat to corals from climate change (*1*, *2*) emphasized the potential for coral acclimatization or local adaptation to alter predictions of climate change effects. New models show that coral adaptation over a 40-year time frame could substantially change predictions for coral reef demise (*22*). But because future adaptation over many generations has been considered too slow for long-lived species such as corals, the role of individual acclimatization in coral environmental tolerance has been central to the debate on the future of reefs. The corals in our experiment achieved a larger bleaching phenotype shift (I_A = 1.56) due to acclimatization within 15 to 24 months than the shift due to fixed effects between habitats, a substantial acceleration in the rate of local matching of coral phenotype to thermal environment. Our results show that acclimatization can allow corals to acquire substantial high-temperature resistance more quickly than strong natural selection would produce.

We do not yet know how many coral species can acclimate or evolve. Several dozen coral species live and grow in the overheated backreef pools of Ofu (*23*), but whether all individual colonies have equal acclimatization ability, or if there is an upper thermal limit to acclimatization or adaptation (*24*), remain unknown. It is also probable that multiple stressors—from acidification and heat, for example—can reduce the ability of corals to respond (*25*). Thus, acclimatization alone cannot be expected to completely overcome the threat to corals from widespread bleaching events, especially if the onset of high-temperature stress is abrupt and sustained. In this regard, the tempo and severity of heat anomalies will be critical for effective coral acclimatization.

Persistence of populations through climate change demands biogeographic shifts of species (*26*–*28*), evolutionary adaptation of populations (*29*), or local acclimatization of individuals (*2*). For long-lived, sessile foundation species that create ecosystem habitat such as forest trees or reef-building corals, range shifts and evolution are predicted to be slow (*2*, *30*). Consequently, the rate and scope of acclimatization in these species to future environmental conditions is central to understanding the impact of climate change. For the fast-growing, shallow-water species that we studied, acclimatory and adaptive responses allowed them to inhabit reef areas with water temperatures far above their expected tolerances. How well other corals can similarly respond, and what the limits of these responses are, will determine how well current models accurately predict the future demise of coral reefs.

REFERENCES AND NOTES

1. O. Hoegh-Guldberg, *Mar. Freshw. Res.* **50**, 839 (1999).
2. T. P. Hughes *et al.*, *Science* **301**, 929–933 (2003).
3. T. P. Hughes, N. A. J. Graham, J. B. C. Jackson, P. J. Mumby, R. S. Steneck, *Trends Ecol. Evol.* **25**, 633–642 (2010).
4. K. E. Carpenter *et al.*, *Science* **321**, 560–563 (2008).
5. T. A. Oliver, S. R. Palumbi, *Mar. Ecol. Prog. Ser.* **378**, 93–103 (2009).
6. T. A. Oliver, S. R. Palumbi, *Coral Reefs* **30**, 429–440 (2011).

7. K. E. Shamberger *et al.*, *Geophys. Res. Lett.* **41**, 499–504 (2014).
8. K. E. Fabricius *et al.*, *Nat. Clim. Change* **1**, 165–169 (2011).
9. L. W. Smith, D. Barshis, C. Birkeland, *Coral Reefs* **26**, 559–567 (2007).
10. L. W. Smith, H. H. Wirshing, A. C. Baker, C. Birkeland, *Pac. Sci.* **62**, 57–69 (2008).
11. J. T. Ladner, S. R. Palumbi, *Mol. Ecol.* **21**, 2224–2238 (2012).
12. D. J. Barshis *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1387–1392 (2013).
13. P. De Wit *et al.*, *Mol. Ecol. Resour.* **12**, 1058–1067 (2012).
14. M. Karin, E. Gallagher, *Immunol. Rev.* **228**, 225–240 (2009).
15. H. M. Shen, S. Pervaiz, *FASEB J.* **20**, 1589–1598 (2006).
16. T. A. Oliver, S. R. Palumbi, *Coral Reefs* **30**, 241–250 (2011).
17. J. T. Ladner, D. J. Barshis, S. R. Palumbi, *BMC Evol. Biol.* **12**, 217 (2012).
18. D. J. Barshis, J. T. Ladner, T. A. Oliver, S. R. Palumbi, *Mol. Biol. Evol.* (2014).
19. J. Endler, *Natural Selection in the Wild* (Princeton Univ. Press, Princeton, NJ, 1986).
20. S. L. Coles, P. L. Jokiel, C. R. Lewis, *Pac. Sci.* **30**, 155 (1976).
21. P. Jokiel, S. Coles, *Coral Reefs* **8**, 155–162 (1990).
22. C. A. Logan, J. P. Dunne, C. M. Eakin, S. D. Donner, *Glob. Change Biol.* **20**, 125–139 (2014).
23. P. Craig, C. Birkeland, S. Belliveau, *Coral Reefs* **20**, 185–189 (2001).
24. J. H. Stillman, *Science* **301**, 65 (2003).
25. J. E. Carilli, R. D. Norris, B. A. Black, S. M. Walsh, M. McField, *PLOS ONE* **4**, e6324 (2009).
26. C. Parmesan *et al.*, *Nature* **399**, 579–583 (1999).
27. C. Parmesan, G. Yohe, *Nature* **421**, 37–42 (2003).
28. T. L. Root, D. P. MacMynowski, M. D. Mastrandrea, S. H. Schneider, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 7465–7469 (2005).
29. C. Parmesan, *Annu. Rev. Ecol. Evol. Syst.* **37**, 637–669 (2006).
30. O. Honnay *et al.*, *Ecol. Lett.* **5**, 525–530 (2002).

ACKNOWLEDGMENTS

The transcriptome data are archived at NCBI's Gene Expression Omnibus database (series record GSE56278: www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE56278). Funding was provided by The Gordon and Betty Moore Foundation through grant GBMF2629, the Schmidt Ocean Institute, and a fellowship from the NSF to N.T.K. We thank M. van Oppen, I. Baums, M. De Salvo, L. Bay, M. Morikawa, and O. Hoegh-Guldberg for comments on earlier versions of the manuscript. We also thank the U.S. National Park of American Samoa for permission to work on Ofu reefs and C. Caruso for logistical and research help.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/895/suppl/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S4
References (*31*–*35*)

27 January 2014; accepted 1 April 2014
Published online 24 April 2014;
10.1126/science.1251336

EVOLUTION

Ancient DNA reveals elephant birds and kiwi are sister taxa and clarifies ratite bird evolution

Kieren J. Mitchell,[1] Bastien Llamas,[1] Julien Soubrier,[1] Nicolas J. Rawlence,[1]* Trevor H. Worthy,[2] Jamie Wood,[3] Michael S. Y. Lee,[1,4] Alan Cooper[1]†

The evolution of the ratite birds has been widely attributed to vicariant speciation, driven by the Cretaceous breakup of the supercontinent Gondwana. The early isolation of Africa and Madagascar implies that the ostrich and extinct Madagascan elephant birds (Aepyornithidae) should be the oldest ratite lineages. We sequenced the mitochondrial genomes of two elephant birds and performed phylogenetic analyses, which revealed that these birds are the closest relatives of the New Zealand kiwi and are distant from the basal ratite lineage of ostriches. This unexpected result strongly contradicts continental vicariance and instead supports flighted dispersal in all major ratite lineages. We suggest that convergence toward gigantism and flightlessness was facilitated by early Tertiary expansion into the diurnal herbivory niche after the extinction of the dinosaurs.

Despite extensive studies, the evolutionary history of the giant flightless ratite birds of the Southern Hemisphere landmasses and the related flighted tinamous of South America has remained a major unresolved question. The ratites and tinamous, termed "palaeognaths" due to their shared basal palate structure, form the sister taxon to all other living birds (neognaths). The living ratites are one of the few bird groups composed largely of giant terrestrial herbivores and include: the emu and cassowary in Australia and New Guinea, the kiwi in New Zealand, the ostrich in Africa, and the rhea in South America. In addition, two recently extinct groups included the largest birds known: the moa from New Zealand (height up to 2 to 3 m, 250 kg in weight) (*1*) and elephant birds from Madagascar (2 to 3 m in height, up to 275 kg in weight) (*2*, *3*). Ratites have been believed to have originated through vicariant speciation driven by the continental breakup of the supercontinent Gondwana on the basis of congruence between the sequence of continental rifting and the presumed order of lineage divergence and distribution of ratites (*4*, *5*).

New Zealand is the only landmass to have supported two major ratite lineages: the giant herbivorous moa and the chicken-sized, nocturnal, omnivorous kiwi. Morphological phylogenetic analyses initially suggested that these two groups were each other's closest relatives (*6*, *7*), presumably diverging after the isolation of an ancestral form following the separation of New Zealand and Australia in the late Cretaceous ~80 to 60 million years ago (Ma) (*8*). However, subsequent studies suggest that kiwi are more closely related to the Australasian emu and cassowaries (*9*, *10*), whereas the closest living relatives of the giant moa are the flighted South American tinamous (*11*–*14*). The latter relationship was completely unexpected on morphological grounds and suggests a more complex evolutionary history than predicted by a model of strict vicariant speciation. By rendering ratites paraphyletic, the relationship between the moa and tinamous also strongly suggests that gigantism and flightlessness have evolved multiple times among palaeognaths (*12*, *13*).

Perhaps the most enigmatic of the modern palaeognaths are the recently extinct giant Madagascan elephant birds. Africa and Madagascar were the first continental fragments to rift from the supercontinent Gondwana, separating from the other continents (and each other) completely during the Early Cretaceous (~130 to 100 Ma) (*15*). Consequently, the continental vicariance model predicts that elephant birds and ostriches should be the basal palaeognath lineages (*16*). Most molecular analyses recover the ostrich in a basal position, consistent with a vicariant model. However, the phylogenetic position of the elephant birds remains unresolved, as cladistic studies of ratite morphology are sensitive to character choice and may be confounded by convergence (*17*), whereas DNA studies have been hampered by the generally poor molecular preservation of elephant bird remains (*18*).

We used hybridization enrichment with in-solution RNA arrays of palaeognath mitochondrial genome sequences and high-throughput sequencing to sequence near-complete mitochondrial genomes from both elephant bird genera: *Aepyornis* and *Mullerornis*. Phylogenetic analyses placed the two taxa, *Aepyornis hildebrandti* (15,547 base pairs) and *Mullerornis agilis* (15,731 base pairs), unequivocally as the sister taxa to the kiwi (Fig. 1 and fig. S1). This result was consistently retrieved, regardless of phylogenetic method or taxon sampling, and was strongly supported by topological tests (*19*). To our knowledge, no previous study has suggested this relationship, probably because of the disparate morphology, ecology, and distribution of the two groups. Elephant birds were herbivorous, almost certainly diurnal, and among the largest birds

[1]Australian Centre for Ancient DNA, School of Earth and Environmental Sciences, University of Adelaide, North Terrace Campus, South Australia 5005, Australia. [2]School of Biological Sciences, Flinders University, South Australia 5001, Australia. [3]Landcare Research, Post Office Box 40, Lincoln 7640, New Zealand. [4]South Australian Museum, North Terrace, South Australia 5000, Australia.
*Present address: Allan Wilson Centre for Molecular Ecology and Evolution, Department of Zoology, University of Otago, Dunedin, New Zealand. **†Corresponding author. E-mail: alan.cooper@adelaide.edu.au**

known, whereas kiwi are highly derived omnivores, nocturnal, and about two orders of magnitude smaller. Elephant birds more closely resemble the moa, and analyses of morphology have suggested a close relationship between these taxa (*17*). However, adding morphological characters to our molecular data set increased support for the relationship between elephant birds and kiwi (figs. S2 and S3) and allowed for identification of several distinctive character states that diagnose this clade [see list in (*19*)].

Speciation by continental vicariance provides a poor explanation of the close relationship between elephant birds and kiwi. Madagascar and New Zealand have never been directly connected, and molecular dates calculated from the genetic data suggest that kiwi and elephant birds diverged after the breakup of Gondwana (Fig. 1 and fig. S4). However, mean node age estimates among palaeognath lineages are sensitive to taxon sampling (Fig. 2), so molecular dating provides limited power for testing hypotheses about ratite biogeography. Depending on taxon sampling, estimates for the basal divergence among palaeognaths are equally consistent with the separation of Africa ~100 Ma (*15*) and the Cretaceous-Tertiary (KPg) boundary (~65 Ma) (Fig. 2). Thus, topological comparisons may be a more robust tool to test hypotheses of vicariance and connection.

The phylogenetic placement of the elephant bird as sister to the kiwi creates a marked discordance between the order of continental breakup (Fig. 3, A and B) and the sequence of palaeognath divergences (Fig. 3C). Instead, it appears that the common ancestor of elephant birds and kiwis was probably flighted and capable of long-distance dispersal, which is supported by a small, possibly flighted kiwi relative from the Early Miocene of New Zealand (*20*). Together, the phylogenetic position of the flighted tinamous and apparent flighted ancestor of the kiwi and elephant bird imply that every major ratite lineage independently lost flight (Fig. 1). We suggest that flighted dispersal was the primary driver of the distribution of palaeognath lineages and that the discordance between distribution and phylogeny is more consistent with lineage turnover in a phylogenetically diverse, flighted, and widespread clade. Early Tertiary palaeognaths were capable of long-distance dispersal, with remains found well outside the range of modern ratites, including the flighted lithornithids in North America and Europe and the flightless *Palaeotis* and *Remiornis* in Europe (Fig. 3A) (*21*). Rapid diversification through flighted dispersal also provides an explanation for the short and often poorly supported internodes amongst basal extant ratite lineages (*13*, *14*).

Early ratite evolution appears to have been dominated by flighted dispersal and parallel evolution, with flightlessness evolving a minimum of six times and gigantism a minimum of five (Fig. 1) (*22*), suggesting that adaptations for cursoriality may have confounded phyloge-

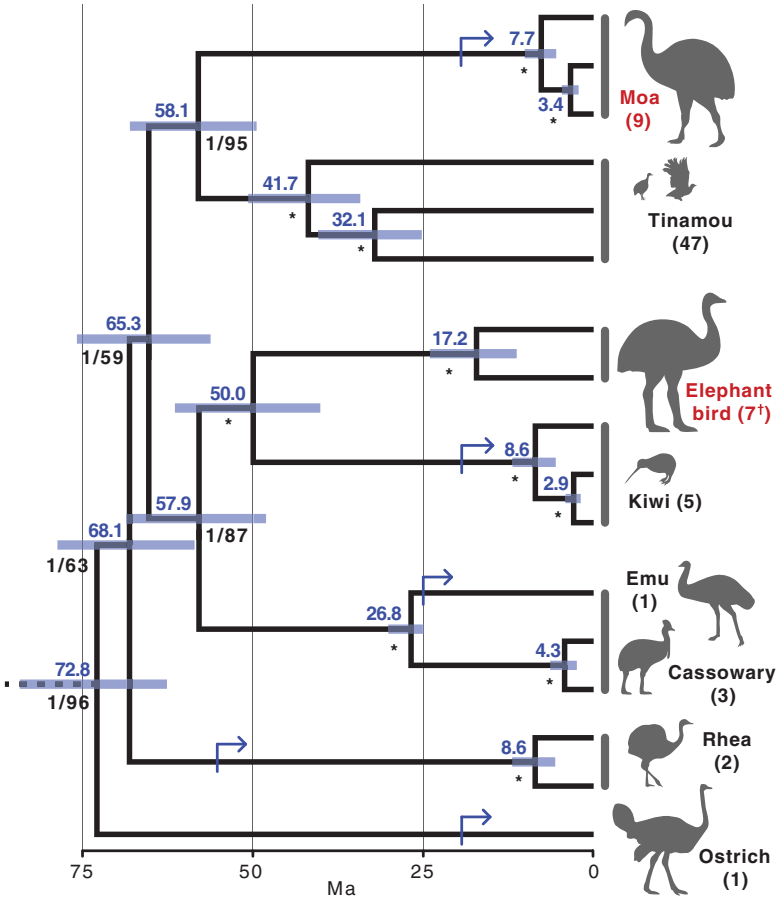


Fig. 1. Phylogenetic position of the elephant birds from mitochondrial sequence data. Bayesian posterior probabilities and maximum likelihood bootstrap are presented in black below each branch; asterisks denote branches that received maximum possible support (bootstrap = 100%, Bayesian posterior probability = 1.0). Divergence dates [blue numbers above branches; blue bars represent 95% highest posterior density (HPD) intervals] were inferred with six well-supported node age constraints (table S5). Blue arrows mark the minimum date for the evolution of flightlessness in lineages for which fossil evidence is available (*21*, *22*). The scale is given in millions of years before the present. Silhouettes indicate the relative size of representative taxa. Species diversity for each major clade is presented in parentheses, with extinct groups shown in red. The dagger symbol (†) indicates that the number of elephant bird species is uncertain.

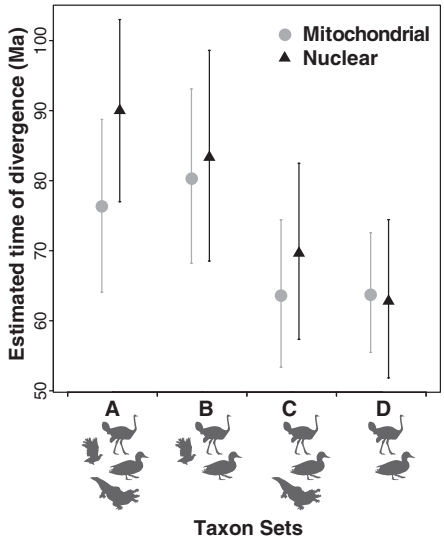


Fig. 2. Sensitivity of palaeognath age estimation to taxon sampling and genetic loci used. Mean and 95% HPD intervals are displayed for the age (basal divergence) of crown palaeognaths, as inferred under several data set permutations. The *y* axis represents time before the present in millions of years, whereas age estimates for individual data sets are arrayed on the *x* axis for both mitochondrial (circles) and nuclear loci (triangles). Results are presented for the following taxon sets: A, a full taxon set including ratites, tinamous, neognaths, and crocodilians; B, ratites, tinamous, and neognaths only; C, ratites, neognaths, and crocodilians only; and D, ratites and neognaths only. Taxon sets A and C are each calibrated with six fossil node constraints, whereas taxon sets B and D are calibrated with a subset of four relevant constraints (table S5). Taxon sets are represented visually with silhouettes of an ostrich (ratites), a flying tinamou (tinamous), a duck (neognaths), and an alligator (crocodilians). Analyses excluding the rate-anomalous tinamous (taxon sets C and D) retrieve a young age near the KPg boundary.

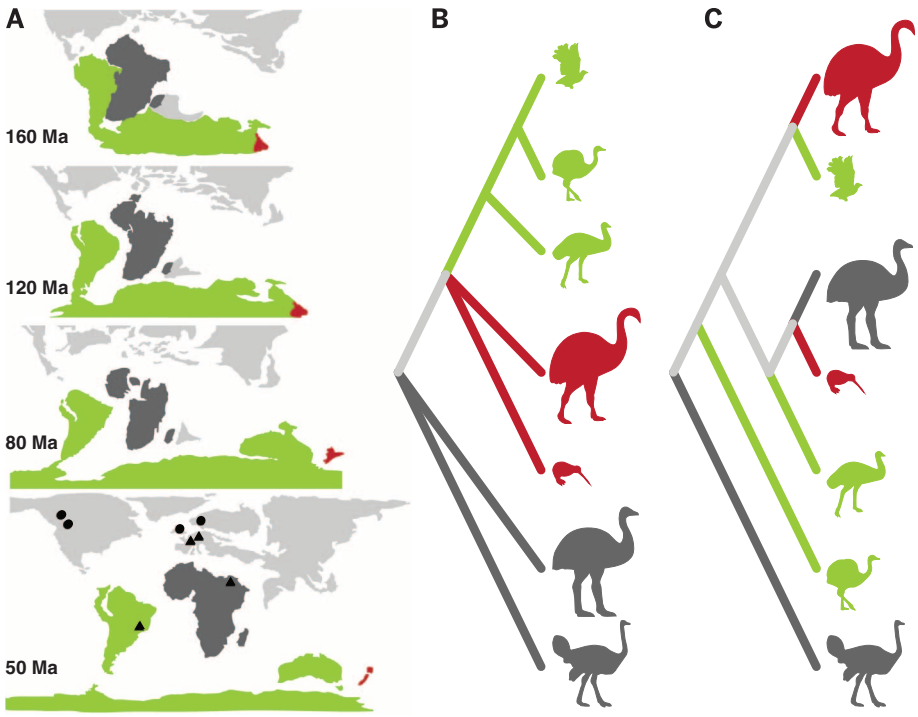


Fig. 3. Conflict between inferred palaeognath phylogeny and the topology predicted by continental vicariance. (**A**) Relative position of continents during the Late Cretaceous and Tertiary. Continental landmasses are colored according to order of severance from the remaining Gondwanan landmass: Africa and Madagascar first (dark gray; 100 to 130 Ma), followed by New Zealand (red; 60 to 80 Ma), then finally Australia, Antarctica, and South America (green; 30 to 50 Ma). Palaeognath-bearing fossil localities from the late Palaeocene and Eocene (*21*, *22*) are represented by circles (flighted taxa) and triangles (flightless taxa). (**B**) Predicted phylogeny of ratites under a model of speciation governed solely by continental vicariance. (**C**) Palaeognath phylogeny as inferred in the present study (see Fig. 1).

netic inference. Elsewhere, avian gigantism and flightlessness are almost exclusively observed in island environments in the absence of mammalian predators and competitors (e.g., the dodo). However, each of the landmasses occupied by ratites (excluding New Zealand) is now home to a diverse mammalian fauna. We suggest that the initial evolution of flightless ratites began in the ecological vacuum after the KPg mass-extinction event and the extinction of the dinosaurs (*12*, *21*). Most mammals appear to have remained relatively small and unspecialized for up to 10 Ma after the KPg extinction (*23*), potentially providing a window of opportunity for the evolution of large flightless herbivores in continental bird lineages. The early Tertiary fossil record supports this interpretation, with geographically widespread flighted palaeognath fossils (Fig. 3A) (*22*) and the appearance of other flightless avian herbivores such as gastornithids in Europe and North America, dromornithids in Australia, and *Brontornis* from South America (*21*). After the early Tertiary, the increasing prevalence of morphologically diverse mammalian competitors is likely to have prevented flightlessness from developing in other continental bird lineages.

The kiwi and tinamous are the only recent palaeognath lineages to not exhibit gigantism, and both taxa co-occur with a second palaeognath lineage (moa and rhea, respectively) that is both much larger and not their closest relative. We suggest that the disparity in size between co-occurring lineages may be a result of the relative timing of arrival of ancestral flighted palaeognaths coupled with competitive exclusion: The first palaeognath to arrive on each landmass monopolized the available niche space for large flightless herbivores and omnivores, forcing subsequent arrivals to adopt an alternative role and remain much smaller. For example, the South American ancestors of the rhea lineage (*Diogenornis*) were already large and flightless at 55 Ma (*21*) when the tinamou lineage originated. The absence of sympatric lineages of small palaeognaths on other landmasses in the recent past may reflect unavailability of alternative niches upon arrival (e.g., due to diversification of herbivorous mammals during the early Tertiary) or subsequent competition with mammals and/or neognathous birds. It is presumably the latter that has necessitated the maintenance of flight in the tinamous.

REFERENCES AND NOTES

1. M. Bunce *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 20646–20651 (2009).
2. T. H. Worthy, R. N. Holdaway, *The Lost World of the Moa* (Indiana Univ. Press, Bloomington, IN, 2002).
3. J. P. Hume, M. Walters, *Extinct Birds* (Bloomsbury, London, 2012).
4. J. Cracraft, *J. Zool.* **169**, 455–543 (1973).
5. K. Lee, J. Feinstein, J. Cracraft, in *Avian Molecular Evolution and Systematics*, D. Mindell, Ed. (Academic Press, New York, 1997), pp. 173–208.
6. T. J. Parker, *Trans. Zool. Soc. London* **13**, 373–431 (1895).
7. J. Cracraft, *Ibis* **116**, 494–521 (1974).
8. W. P. Schellart, G. S. Lister, V. G. Toy, *Earth Sci. Rev.* **76**, 191–233 (2006).
9. A. Cooper *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 8741–8744 (1992).
10. A. H. Bledsoe, *Ann. Carnegie Mus.* **57**, 73 (1988).
11. O. Haddrath, A. J. Baker, *Proc. Biol. Sci.* **279**, 4617–4625 (2012).
12. M. J. Phillips, G. C. Gibb, E. A. Crimp, D. Penny, *Syst. Biol.* **59**, 90–107 (2010).
13. J. Harshman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13462–13467 (2008).
14. J. V. Smith, E. L. Braun, R. T. Kimball, *Syst. Biol.* **62**, 35–49 (2013).
15. J. R. Ali, D. W. Krause, *J. Biogeogr.* **38**, 1855–1872 (2011).
16. P. Johnston, *Zool. J. Linn. Soc.* **163**, 959–982 (2011).
17. T. H. Worthy, R. P. Scofield, *N.Z. J. Zool.* **39**, 87–153 (2012).
18. A. Cooper *et al.*, *Nature* **409**, 704–707 (2001).
19. For details, see materials and methods on *Science* Online.
20. T. H. Worthy *et al.*, in *Paleornithological Research 2013 - Proceedings of the 8th International Meeting of the Society of Avian Paleontology and Evolution*, U. B. Göhlich, A. Kroh, Eds. (Natural History Museum Vienna, Vienna, Austria, 2013), pp. 63–80.
21. G. Mayr, in *Paleogene Fossil Birds* (Springer, Berlin, 2009), pp. 25–34.
22. See supplementary text on *Science* Online.
23. K. Black, M. Archer, S. Hand, H. Godthelp, in *Earth and Life*, J. Talent, Ed. (Springer Netherlands, 2012), pp. 983–1078.

ACKNOWLEDGMENTS

This study was funded by the New Zealand Marsden Fund and the Australian Research Council. Grid computing facilities were provided by eRSA (e-research South Australia) and CIPRES (Cyberinfrastructure for Phylogenetic Research). We thank the Museum of New Zealand Te Papa Tongarewa (P. Millener, S. Bartle, A. Tennyson), Natural History Museum Oslo (N. Heintz), and National Museum of Natural History Paris (C. Lefèvre) for samples. We acknowledge D. Penny, M. Phillips, D. Burney and R. Ward for valuable advice and assistance. All data associated with this study are available on GenBank (accession nos. KJ749824 and KJ749825) and DRYAD (doi:10.5061/dryad.2727k).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/898/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 to S10
References (*24*–*76*)

10 February 2014; accepted 21 April 2014
10.1126/science.1251981

DECISION-MAKING

FoxP influences the speed and accuracy of a perceptual decision in *Drosophila*

Shamik DasGupta, Clara Howcroft Ferreira, Gero Miesenböck*

Decisions take time if information gradually accumulates to a response threshold, but the neural mechanisms of integration and thresholding are unknown. We characterized a decision process in *Drosophila* that bears the behavioral signature of evidence accumulation. As stimulus contrast in trained odor discriminations decreased, reaction times increased and perceptual accuracy declined, in quantitative agreement with a drift-diffusion model. *FoxP* mutants took longer than wild-type flies to form decisions of similar or reduced accuracy, especially in difficult, low-contrast tasks. RNA interference with *FoxP* expression in αβ core Kenyon cells, or the overexpression of a potassium conductance in these neurons, recapitulated the *FoxP* mutant phenotype. A mushroom body subdomain whose development or function require the transcription factor FoxP thus supports the progression of a decision toward commitment.

Integrator models of perceptual decision-making (*1*–*5*) predict that reaction times will vary with the quality of sensory information: Easy decisions, based on clear evidence, will be fast; difficult decisions, based on uncertain evidence, will be slow. We tested this prediction in fruit flies, using a reaction time version of an olfactory discrimination task (*6*–*8*). Flies were analyzed individually in narrow chambers, which were perfused with odor-air mixtures whose convergence defined a 7-mm-wide decision zone (Fig. 1A). With odors present, the flies slowed upon entry into the decision zone, paused near the interface (Fig. 1B and fig. S1, A and B), and exited after committing to a choice (Fig. 1B and fig. S1). We quantified the time between entry and exit as the reaction time (Fig. 1A).

Centre for Neural Circuits and Behaviour, University of Oxford, Tinsley Building, Mansfield Road, Oxford, OX1 3SR, UK.
*Corresponding author. E-mail: gero.miesenboeck@cncb.ox.ac.uk

Flies were trained to avoid a specific concentration of 4-methylcyclohexanol (MCH) and had to distinguish the reinforced concentration from a lower concentration of the same odor. The difficulty of discrimination was titrated by varying the MCH concentration ratio during testing (Fig. 1). At a reinforced MCH intensity of ~12 parts per billion, corresponding to 10^{-4} volumes of saturated vapor per volume of air, flies achieved accuracies of nearly 100% at large concentration differences (concentration ratio of comparison to reinforced odor of 0.1 to 0.2) but performed randomly when the odor concentrations differed by only 10% (concentration ratio 0.9) (Fig. 1C). Reaction times increased as a function of difficulty (Fig. 1, D and F, and fig. S1C), suggesting that the flies compensated for low stimulus contrast in difficult tasks by gathering information for longer. Because overall odor concentrations were highest in the low-contrast tasks that took the longest to complete (Fig. 1, D and F, and fig. S1C) and because only relative stimulus contrast, not absolute stimulus intensity, affected reaction times (Fig. 2, B and G), our data favor integration over probability summation (*9*): If the probability of odor detection were rate-limiting, reaction time would be expected to correlate inversely with stimulus intensity (*5*, *9*).

A drift-diffusion model of evidence integration (*1*, *3*–*5*, *10*) (Fig. 1E) (see Materials and Methods) captured the empirical relationship between difficulty and performance. Drift-diffusion models decompose reaction times into decision and residual times. The residual time encompasses sensory and motor latencies, procrastination, and time required to indicate a choice. The decision

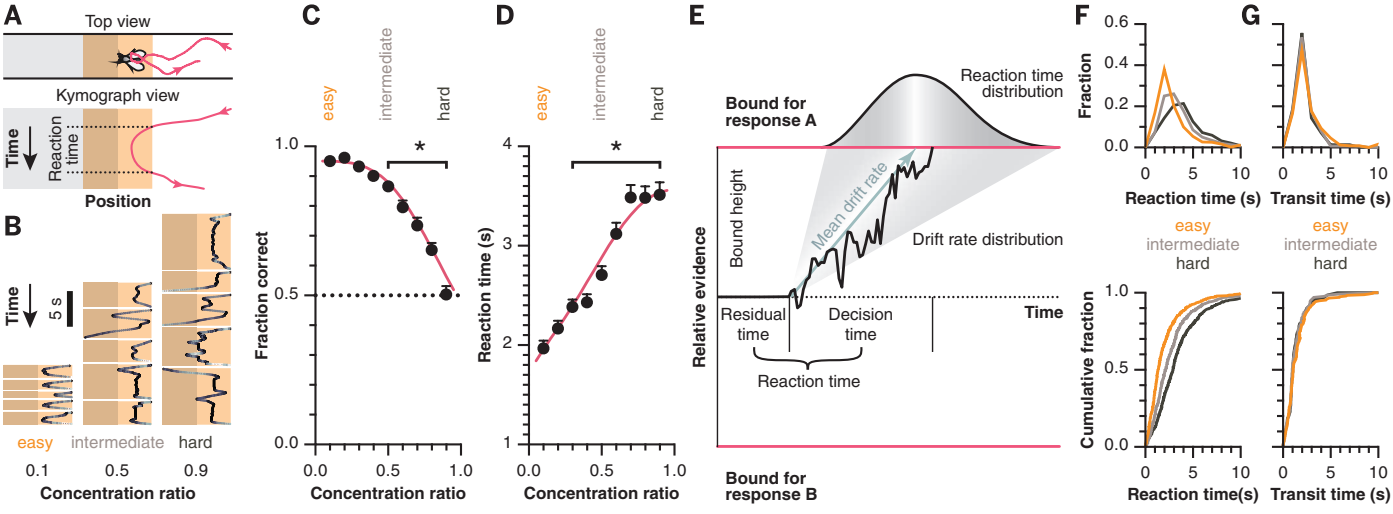


Fig. 1. Analysis of decision accuracies and reaction times. (**A**) Movement trajectory (red) near the odor interface. Orange color indicates the 7-mm-wide decision zone and gray shading on the left the aversively reinforced odor MCH, which must be distinguished from a lower MCH concentration on the right. The kymograph shows the fly's position on the long chamber axis over time. Because the centroid of the fly's silhouette rather than the position of the antennae is monitored, the trajectory stops short of the odor interface. (**B**) Example kymographs of wild-type flies discriminating different MCH intensities. (**C**) Accuracy (mean ± SEM; N = 142 to 173 flies per data point) versus difficulty. Asterisks indicate significant differences from concentration ratio 0.1 ($P < 0.0001$). The red line depicts the psychometric function predicted by a drift-diffusion model (*10*). (**D**) Reaction time (mean ± SEM; N = 592 to 828 decisions per data point) versus difficulty. Asterisks indicate significant differences from concentration ratio 0.1 ($P < 0.0005$). The red line depicts the chronometric function predicted by the drift-diffusion model (*10*). (**E**) Drift-diffusion model of evidence accumulation to response-bound A. (**F** and **G**) Frequency (top) and cumulative frequency distributions (bottom) of reaction times (F) and transit times (G) at odor concentration ratios corresponding to easy (0.1), intermediate (0.5), and hard (0.9) difficulty levels. Reaction time distributions differ between difficulty levels ($P < 0.0001$), but transit time distributions do not ($P > 0.3158$).

time constitutes the integration period per se. Its duration depends on the mean rate at which evidence accrues (the drift rate) and the level of the decision criterion (the bound height) (Fig. 1E). Intuitively, the weaker the sensory evidence, the lower will be the drift rate, the longer the response time, and the poorer the decision accuracy. The model we use (*10*) formalizes this intuition by scaling the drift rate in proportion to stimulus contrast, which we quantify as |log (odor concentration ratio)|. Estimation of the three free model parameters (drift rate, bound height, and residual time) from reaction time measurements generates a prediction of the corresponding decision accuracies. The modeled chronometric and psychometric functions provided satisfying simultaneous fits to our performance data (Fig. 1, C and D).

At all difficulty levels, the reaction time distributions exhibited positive skew (Fig. 1F), a characteristic of information accumulation to threshold (*1*, *3*, *5*, *11*). The drift-diffusion model explains the origin of the asymmetry: Equal differences in drift rate generate unequal differences in reaction time at the intersection with the response bound (Fig. 1E). In contrast, transit times of nondecision zones located off center had nearly symmetrical distributions that did not vary with task difficulty (Fig. 1G).

To examine whether the relationship between task difficulty and reaction time generalized, we designed two tasks other than intensity discrimination. In a masking odor task (fig. S2A), shock-reinforced MCH was presented in a ubiquitous background of 3-octanol (OCT). Difficulty was adjusted by varying the level of masking odor while keeping constant the concentration of cue. Flies took longer to respond to cues hidden in a high level of background than to salient cues and did so with lower accuracy (fig. S2A). In binary mixture discriminations (*6*, *7*, *12*) (fig. S2B), the closer the proportions of MCH and OCT, the lower the accuracy and speed of discrimination (fig. S2B).

A pilot analysis of 41 strains carrying candidate mutations implicated the transcription factor FoxP in the decision process. *FoxP*$^{5\text{-}SZ\text{-}3955}$ mutants learned to distinguish the shock-reinforced concentration of MCH with the same accuracy as wild-type flies (Fig. 2, A and F) but took longer to decide (Fig. 2, B to D and G to I). The defect was subtle in easy discriminations (concentration ratio 0.1 to 0.4) but glaring in difficult tasks (concentration ratio 0.7 to 0.9).

Mutating *FoxP* might alter any one of several processes that affect performance in our assay: the abilities to learn from shock reinforcement, walk to and from the odor interface, detect olfactory cues, and decide. Learning and locomotor deficits could be ruled out by examining the accuracy scores (Fig. 2, A and F) and transit time distributions (Fig. 2, E and J), respectively; both were identical in *FoxP*$^{5\text{-}SZ\text{-}3955}$ mutants and wild-type flies. Mutants detected odors with the same sensitivity as wild-type controls: Diluting all odors 1000-fold had similar effects on either genotype (compare Fig. 2, A to E, with Fig. 2, F to J). Where mutant and wild-type flies clearly differed, however, was in the dependence of reaction time on stimulus contrast: In mutants, narrowing the odor concentration difference caused disproportionate increases in reaction time (compare red and black curves in Fig. 2, B to D, and Fig. 2, G to I). A drift-diffusion model (*10*) (fig. S3, A and B) identified two changes that can account for this

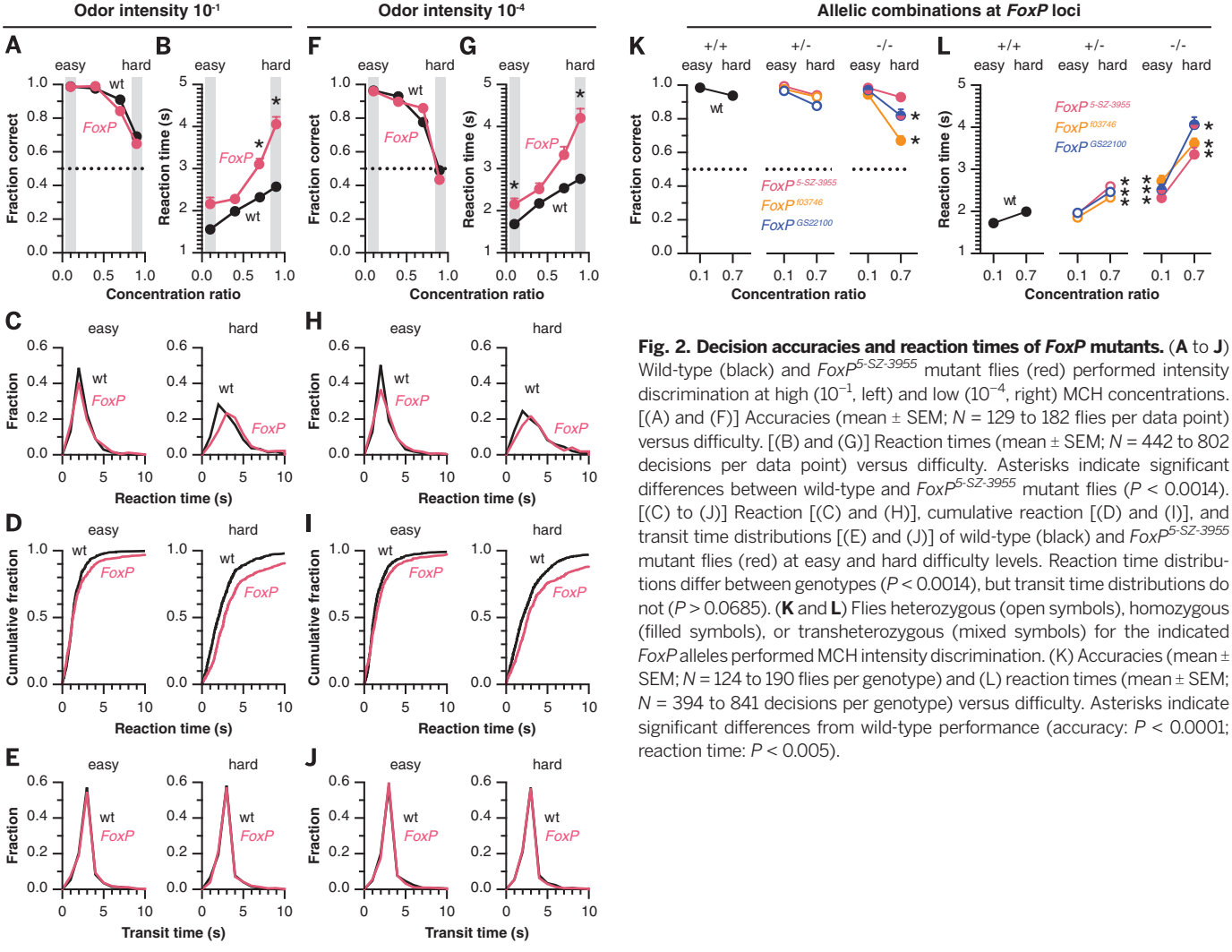


Fig. 2. Decision accuracies and reaction times of *FoxP* mutants. (**A** to **J**) Wild-type (black) and *FoxP*$^{5\text{-}SZ\text{-}3955}$ mutant flies (red) performed intensity discrimination at high (10^{-1}, left) and low (10^{-4}, right) MCH concentrations. [(A) and (F)] Accuracies (mean ± SEM; N = 129 to 182 flies per data point) versus difficulty. [(B) and (G)] Reaction times (mean ± SEM; N = 442 to 802 decisions per data point) versus difficulty. Asterisks indicate significant differences between wild-type and *FoxP*$^{5\text{-}SZ\text{-}3955}$ mutant flies ($P < 0.0014$). [(C) to (J)] Reaction [(C) and (H)], cumulative reaction [(D) and (I)], and transit time distributions [(E) and (J)] of wild-type (black) and *FoxP*$^{5\text{-}SZ\text{-}3955}$ mutant flies (red) at easy and hard difficulty levels. Reaction time distributions differ between genotypes ($P < 0.0014$), but transit time distributions do not ($P > 0.0685$). (**K** and **L**) Flies heterozygous (open symbols), homozygous (filled symbols), or transheterozygous (mixed symbols) for the indicated *FoxP* alleles performed MCH intensity discrimination. (K) Accuracies (mean ± SEM; N = 124 to 190 flies per genotype) and (L) reaction times (mean ± SEM; N = 394 to 841 decisions per genotype) versus difficulty. Asterisks indicate significant differences from wild-type performance (accuracy: $P < 0.0001$; reaction time: $P < 0.005$).

phenotype: a 38% drop in drift rate (fig. S3C) and a—perhaps compensatory—increase in the height of the response bound (fig. S3D). The reduction in drift rate suggests that *FoxP* mutants are impaired in the accumulation and/or retention of sensory information in the buildup to a choice.

We confirmed the *FoxP* mutant phenotype with two independently generated alleles (Fig. 2, K and L, and fig. S4). Heterozygous carriers of any one of these alleles performed like wild-type controls in easy discriminations (concentration ratio 0.1) (Fig. 2, K and L) but displayed prolonged reaction times in difficult tasks (concentration ratio 0.7) (Fig. 2L). Homozygous or transheterozygous carriers of two mutant alleles exhibited pronounced difficulty-dependent speed and, in some allelic combinations, also accuracy deficits (Fig. 2, K and L). The association of similar phenotypes with different mutant alleles, and the lack of complementation between alleles (Fig. 2, K and L), tie the defect in decision formation firmly to the *FoxP* locus.

To identify, label, and manipulate sites of *FoxP* action in the brain, we used a *FoxP* promoter fragment to direct the expression of GAL4. *FoxP-GAL4*–driven transgene expression was confined to two subsets of Kenyon cells (KCs), the principal intrinsic neurons of the mushroom bodies (*13*, *14*): ~80 KCs whose axons extend into the cores of the α and β lobes, and ~100 KCs innervating the γ lobes (Fig. 3, A and B). Given their positions as third-order olfactory neurons, the *FoxP-GAL4*–expressing KCs could transmit sensory data to downstream integrators. Alternatively, the *FoxP-GAL4*–positive KCs themselves could integrate olfactory signals, or the representations of momentary and accumulated sensory evidence might be entwined within the KC population. Because both representations require externally

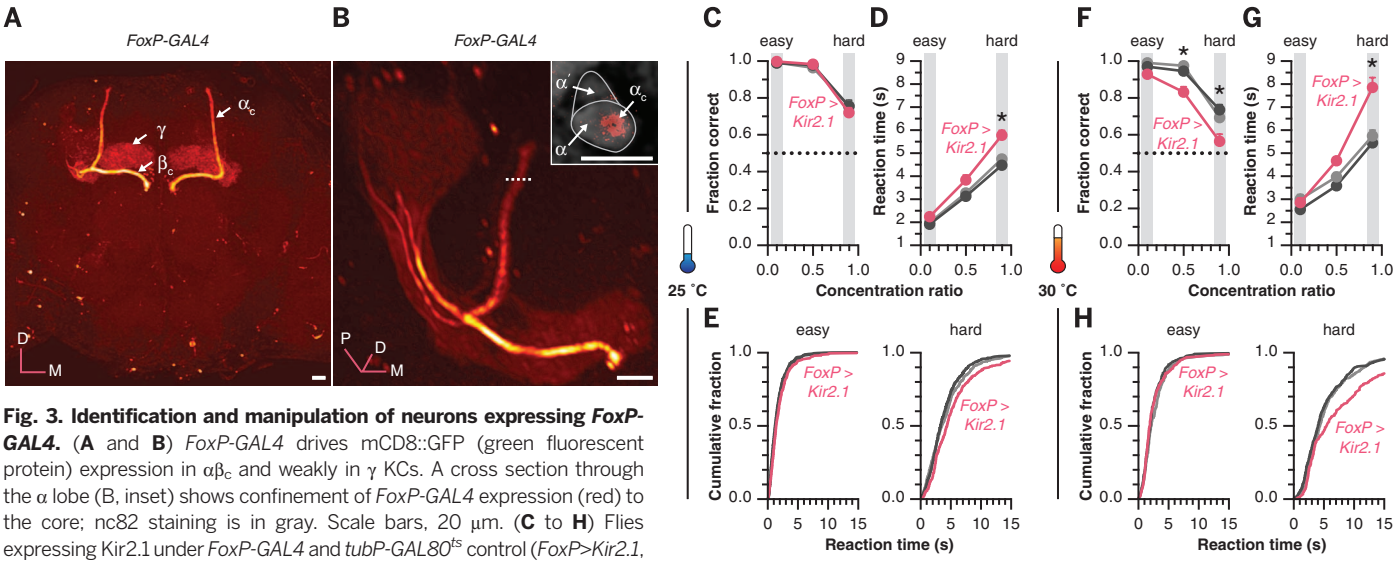


Fig. 3. Identification and manipulation of neurons expressing *FoxP-GAL4*. (**A** and **B**) *FoxP-GAL4* drives mCD8::GFP (green fluorescent protein) expression in $\alpha\beta_c$ and weakly in γ KCs. A cross section through the α lobe (B, inset) shows confinement of *FoxP-GAL4* expression (red) to the core; nc82 staining is in gray. Scale bars, 20 μm. (**C** to **H**) Flies expressing Kir2.1 under *FoxP-GAL4* and *tubP-GAL80^{ts}* control (*FoxP>Kir2.1*, red) and parental controls (*FoxP-GAL4*, dark gray; *UAS-Kir2.1*, light gray) were maintained at the indicated temperatures for 18 hours before performing MCH intensity discrimination. [(C) and (F)] Accuracy (mean ± SEM; N = 81 to 175 flies per genotype) versus difficulty. The asterisk indicates a significant genotype effect ($P < 0.0307$). [(D) and (G)] Reaction times (mean ± SEM; N = 261 to 737 decisions per genotype) versus difficulty. Asterisks indicate significant genotype effects ($P < 0.0002$). [(E) and (H)] Cumulative reaction time distributions of flies maintained at 25°C (E) and 30°C (H). Irrespective of temperature, the distributions differ between experimental flies (red) and parental controls (gray) at the hard ($P < 0.0005$) but not the easy difficulty level ($P > 0.2139$).

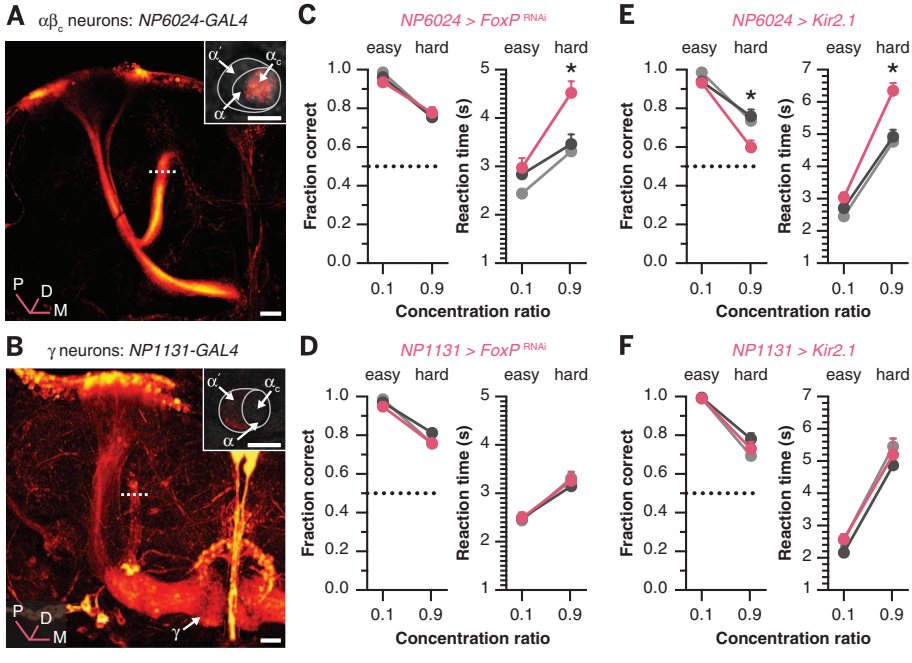


Fig. 4. Manipulation of *FoxP* expression and excitability in KC subsets. (**A** and **B**) *NP6024-GAL4* (A) labels $\alpha\beta_c$ KCs; *NP1131-GAL4* (B) labels γ and some α′β′ KCs. Insets show cross sections through the α lobes: *GAL4*-driven GFP expression in red; mushroom-body staining with *mb247-LexA:LexAop-rCD2::RFP* in gray. Scale bars, 10 μm. (**C** and **D**) Flies expressing *FoxP^{RNAi}* in $\alpha\beta_c$ (C) or γ KCs (D) performed MCH intensity discrimination. Accuracies (mean ± SEM; N = 119 to 178 flies per genotype) and reaction times (mean ± SEM; N = 410 to 803 decisions per genotype) at easy and hard difficulty levels. Red and gray symbols denote experimental flies and parental controls, respectively (*X-GAL4*, dark gray; *UAS-FoxP^{RNAi}*, light gray). The asterisk indicates a significant genotype effect on reaction times ($P < 0.0031$). (**E** and **F**) Flies expressing Kir2.1 in $\alpha\beta_c$ (E) or γ KCs (F) performed MCH intensity discrimination. Accuracies (mean ± SEM; N = 51 to 149 flies per genotype) and reaction times (mean ± SEM; N = 136 to 414 decisions per genotype) at easy and hard difficulty levels. Red and gray symbols denote experimental flies and parental controls, respectively (*X-GAL4*, dark gray; *UAS-Kir2.1*, light gray). Asterisks indicate significant genotype effects on accuracies ($P < 0.032$) and reaction times ($P < 0.0007$).

or recurrently evoked electrical activity (*15*, *16*), reducing the excitability of KCs is predicted to prolong reaction times.

We therefore targeted the inwardly rectifying potassium channel Kir2.1 under *FoxP-GAL4* control to αβ core ($\alpha\beta_c$) and γ KCs while tuning expression levels with temperature-sensitive GAL80^{ts} (*17*). Flies expressing low Kir2.1 levels behaved like homozygous *FoxP*$^{5\text{-}SZ\text{-}3955}$ mutants: Reaction times increased in a difficulty-dependent manner relative to parental controls (Fig. 3, D and E), but accuracy was maintained (Fig. 3C). Boosting the expression of Kir2.1 exacerbated this phenotype: Despite a further increase in reaction times (Fig. 3, G and H), *FoxP>Kir2.1* flies now performed near chance level in difficult discriminations (Fig. 3F), echoing the accuracy defects of severe *FoxP* alleles (Fig. 2K).

To bolster and refine our identification of *FoxP-GAL4*–positive KCs as sites of FoxP action, we compared the consequences of introducing Kir2.1 with those of reducing *FoxP* expression, using a panel of *GAL4* lines whose expression domains included all or parts of the *FoxP-GAL4* pattern: *OK107-GAL4* targets all KCs (*18*), *NP6024-GAL4* and *NP7175-GAL4* label $\alpha\beta_c$ neurons (*14*, *19*), and *NP1131-GAL4* marks γ KCs (*14*, *20*) (Fig. 4, A and B). Knockdown of *FoxP* in $\alpha\beta_c$, but not γ, KCs prolonged reaction times in difficult discriminations (Fig. 4, C and D, and fig. S4B), mirroring the differential impact of Kir2.1 on these neuronal populations (Fig. 4, E and F). Attempts to disrupt *FoxP* expression with the help of *FoxP-GAL4* itself produced marginal effects (fig. S5A), probably due to inadequate *FoxP*^{RNAi} levels. Consistent with this interpretation, significant decision phenotypes were seen with all three of the *GAL4* drivers capable of expressing high levels of *FoxP*^{RNAi} in $\alpha\beta_c$ KCs (Fig. 4C and fig. S5, B and C).

The evolution of a decision toward commitment requires the progression of neural activity from a choice-neutral to a choice-specific state. Mutations in *FoxP* evidently slow this progression, at least in part by interfering with the function of $\alpha\beta_c$ neurons. The same neurons have been implicated in value-based decisions, such as choices between odors associated with punishments of differing severity (*19*). It remains untested whether value judgments also incur a difficulty-dependent cost of decision time. Nonetheless, the available evidence suggests that ~80 *FoxP-GAL4*–positive $\alpha\beta_c$ KCs form part of a versatile decision circuit that processes sensory information in one context and remembered value in another.

As a transcription factor (*21*, *22*), FoxP could act during development to specify synaptic connections and/or throughout life to regulate neuronal function. Vertebrate *FoxP* homologs have been linked to both types of processes (*23*–*26*) and have been attributed critical roles in cognitive development (*24*, *27*, *28*), vocal communication (*21*, *26*, *29*), and motor control (*23*, *25*). A potential commonality between these processes and decision-making is their unfolding over time: Neurons representing trains of thought, strings of syllables, chains of motor commands, or accumulating evidence must all step through ordered activity sequences. It is therefore tempting to speculate that understanding the function of an ancestral *FoxP* gene (*22*) might reveal fundamentals of temporal processing (*30*).

REFERENCES AND NOTES

1. R. Ratcliff, P. L. Smith, *Psychol. Rev.* **111**, 333–367 (2004).
2. M. N. Shadlen, W. T. Newsome, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 628–633 (1996).
3. J. I. Gold, M. N. Shadlen, *Annu. Rev. Neurosci.* **30**, 535–574 (2007).
4. R. Bogacz, E. Brown, J. Moehlis, P. Holmes, J. D. Cohen, *Psychol. Rev.* **113**, 700–765 (2006).
5. R. D. Luce, *Response Times* (Oxford Univ. Press, New York, 1986).
6. N. Uchida, Z. F. Mainen, *Nat. Neurosci.* **6**, 1224–1229 (2003).
7. N. M. Abraham *et al.*, *Neuron* **44**, 865–876 (2004).
8. A. Claridge-Chang *et al.*, *Cell* **139**, 405–415 (2009).
9. A. B. Watson, *Vision Res.* **19**, 515–522 (1979).
10. J. Palmer, A. C. Huk, M. N. Shadlen, *J. Vis.* **5**, 376–404 (2005).
11. J. D. Roitman, M. N. Shadlen, *J. Neurosci.* **22**, 9475–9489 (2002).
12. D. Rinberg, A. Koulakov, A. Gelperin, *Neuron* **51**, 351–358 (2006).
13. K. Ito, W. Awano, K. Suzuki, Y. Hiromi, D. Yamamoto, *Development* **124**, 761–771 (1997).
14. N. K. Tanaka, H. Tanimoto, K. Ito, *J. Comp. Neurol.* **508**, 711–755 (2008).
15. D. A. Robinson, *Annu. Rev. Neurosci.* **12**, 33–45 (1989).
16. X.-J. Wang, *Neuron* **36**, 955–968 (2002).
17. S. E. McGuire, P. T. Le, A. J. Osborn, K. Matsumoto, R. L. Davis, *Science* **302**, 1765–1768 (2003).
18. J. B. Connolly *et al.*, *Science* **274**, 2104–2107 (1996).
19. E. Perisse *et al.*, *Neuron* **79**, 945–956 (2013).
20. H. Qin *et al.*, *Curr. Biol.* **22**, 608–614 (2012).
21. C. S. Lai, S. E. Fisher, J. A. Hurst, F. Vargha-Khadem, A. P. Monaco, *Nature* **413**, 519–523 (2001).
22. M. E. Santos, A. Athanasiadis, A. B. Leitão, L. DuPasquier, E. Sucena, *Mol. Biol. Evol.* **28**, 237–247 (2011).
23. M. Groszer *et al.*, *Curr. Biol.* **18**, 354–362 (2008).
24. G. Konopka *et al.*, *Nature* **462**, 213–217 (2009).
25. C. A. French *et al.*, *Mol. Psychiatry* **17**, 1077–1085 (2012).
26. G. M. Sia, R. L. Clem, R. L. Huganir, *Science* **342**, 987–991 (2013).
27. F. F. Hamdan *et al.*, *Am. J. Hum. Genet.* **87**, 671–678 (2010).
28. G. Konopka *et al.*, *Neuron* **75**, 601–617 (2012).
29. S. Haesler *et al.*, *PLOS Biol.* **5**, e321 (2007).
30. M. D. Mauk, D. V. Buonomano, *Annu. Rev. Neurosci.* **27**, 307–340 (2004).

ACKNOWLEDGMENTS

We thank M. Shadlen for discussions and code; J. Flint, A. Lin, and S. Waddell for comments; and M. Ramaswami for flies. This work was supported by the Wellcome Trust, the Gatsby Charitable Foundation, NIH, and the Oxford Martin School (G.M.); Human Frontier Science Program and Marie Curie Actions Fellowships (S.D.G.); and Fundação Champalimaud and Fundação para a Ciencia e Tecnologia (C.H.F.).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/901/suppl/DC1
Materials and Methods
Figs. S1 to S5
References (*31*–*33*)

11 February 2014; accepted 25 April 2014
10.1126/science.1252114

NEUROSCIENCE

Rapid Hebbian axonal remodeling mediated by visual stimulation

Martin Munz,[1]* Delphine Gobert,[1]* Anne Schohl,[1] Jessie Poquérusse,[1,2] Kaspar Podgorski,[3] Perry Spratt,[1] Edward S. Ruthazer[1]†

We examined how correlated firing controls axon remodeling, using in vivo time-lapse imaging and electrophysiological analysis of individual retinal ganglion cell (RGC) axons that were visually stimulated either synchronously or asynchronously relative to neighboring inputs in the *Xenopus laevis* optic tectum. RGCs stimulated out of synchrony rapidly lost the ability to drive tectal postsynaptic partners while their axons grew and added many new branches. In contrast, synchronously activated RGCs produced fewer new branches, but these were more stable. The effects of synchronous activation were prevented by the inhibition of neurotransmitter release and *N*-methyl-D-aspartate receptor (NMDAR) blockade, which is consistent with a role for synaptic NMDAR activation in the stabilization of axonal branches and suppression of further exploratory branch addition.

Neuronal activity and molecular cues cooperate to form precise neuronal circuits (*1*, *2*). Experimental blockade of action potential firing or synaptic transmission (*3*–*5*), particularly involving *N*-methyl-D-aspartate receptors (NMDARs) (*6*–*9*), degrades axonal projections in the developing nervous system. The precise pattern of neuronal firing is believed to be important for instructing connection refinement because disrupting the temporal correlation of firing between neighboring neurons, even while sparing overall activity levels, results in axons with diffuse terminal arbors (*10*, *11*). Hebbian plasticity, an appealing model for activity-dependent refinement of circuits, posits that synapses may be strengthened or stabilized when the presynaptic cell participates in making its postsynaptic partner fire (*12*). Convergent inputs firing synchronously

[1]Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, 3801 Rue University, Montreal, QC H3A 2B4, Canada. [2]Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA. [3]Brain Research Centre, University of British Columbia (UBC), 2211 Westbrook Mall, Vancouver, BC V6T 2B5, Canada.
*These authors contributed equally to this work. **†Corresponding author. E-mail: edward.ruthazer@mcgill.ca**

would cooperatively excite the postsynaptic neuron to fire. Thus, Hebbian plasticity in principle should aggregate coactive inputs, effectively leading to circuit refinement (*6*, *13*, *14*). The detailed mechanisms by which such remodeling actually occurs remain poorly understood.

The developing retinotectal system of the albino *Xenopus laevis* tadpole is amenable both to live imaging and in vivo electrophysiological characterization. RGC axons in *Xenopus* normally project to the contralateral optic tectum. Occasionally, a single mistargeted ipsilaterally projecting retinal ganglion cell (RGC) axon is observed (Fig. 1, A and B) (*15*). Using post mortem intraocular 1,1′-dioctadecyl-3,3,3′,3′-tetramethylindocarbocyanine perchlorate (DiI) injections to label all RGCs (fig. S1) in stage-46 to -48 tadpoles when the retinotectal projection is established but still refining (*16*), we detected no ipsilateral RGC axon in the majority of cases (61%). However, animals with one (21%), two (9%), or more (9%) ipsilaterally projecting axons occasionally were observed (Fig. 1C). Ipsilaterally projecting axons are thus unlikely to represent a specific class of RGCs but rather reflect random pathfinding errors at the optic chiasm. We exploited the fact that the lone ipsilateral and surrounding contralateral RGCs could be independently visually stimulated in order to test the role of correlated activity on synaptic maintenance and axonal refinement.

We first tested whether ipsilaterally projecting RGCs form functional synapses onto tectal neurons. To identify postsynaptic partners of a single ipsilateral RGC (Fig. 1D), we made transgenic tadpoles expressing photoactivatable green fluorescent protein (PA-GFP) (*17*, *18*) under the neuronal β-tubulin promoter and then bulk-electroporated plasmid encoding tdTomato into one eye so as to label an ipsilaterally projecting axon (*16*). Scanning 780-nm light in a volume immediately surrounding the ipsilateral axon (Fig. 1E) resulted in photoactivated PA-GFP backfilling a few tectal cells from their dendrites (Fig. 1F). In vivo perforated patch recordings of compound synaptic currents (CSCs) from these cells, in response to alternating light flashes presented to each eye through optical fibers (Fig. 1D), revealed that all visually responsive neurons could be driven through the contralateral eye (Fig. 1G). Some cells also responded to stimulation of the ipsilateral eye (Fig. 1H). The ipsilateral CSC (5.31 ± 1.06 pC) was consistently smaller than the contralateral CSC (11.80 ± 2.33 pC).

We next investigated how correlated visual stimulation modifies synaptic strength. Ten min of baseline light-evoked CSCs were measured by alternating a 10-ms light flash to each eye every 30 s, holding the cells at –60 mV (*16*). Cells were then switched to current clamp mode so as to allow spiking, and one of two training paradigms was applied: "synchronous stimulation," in which both eyes were stimulated together every 2 s, or "asynchronous stimulation," in which the eyes were stimulated 1 s apart (Fig. 1I). Both eyes experienced light flashes every 2 s, with only the relative timing differing across paradigms. After 10 min of training stimulation, CSCs were again measured at –60mV in order to test the relative efficacy of the two eyes at driving a response. This cycle of training and testing was repeated three times per cell. After the first 10 min of training, asynchronous stimulation resulted in a loss of

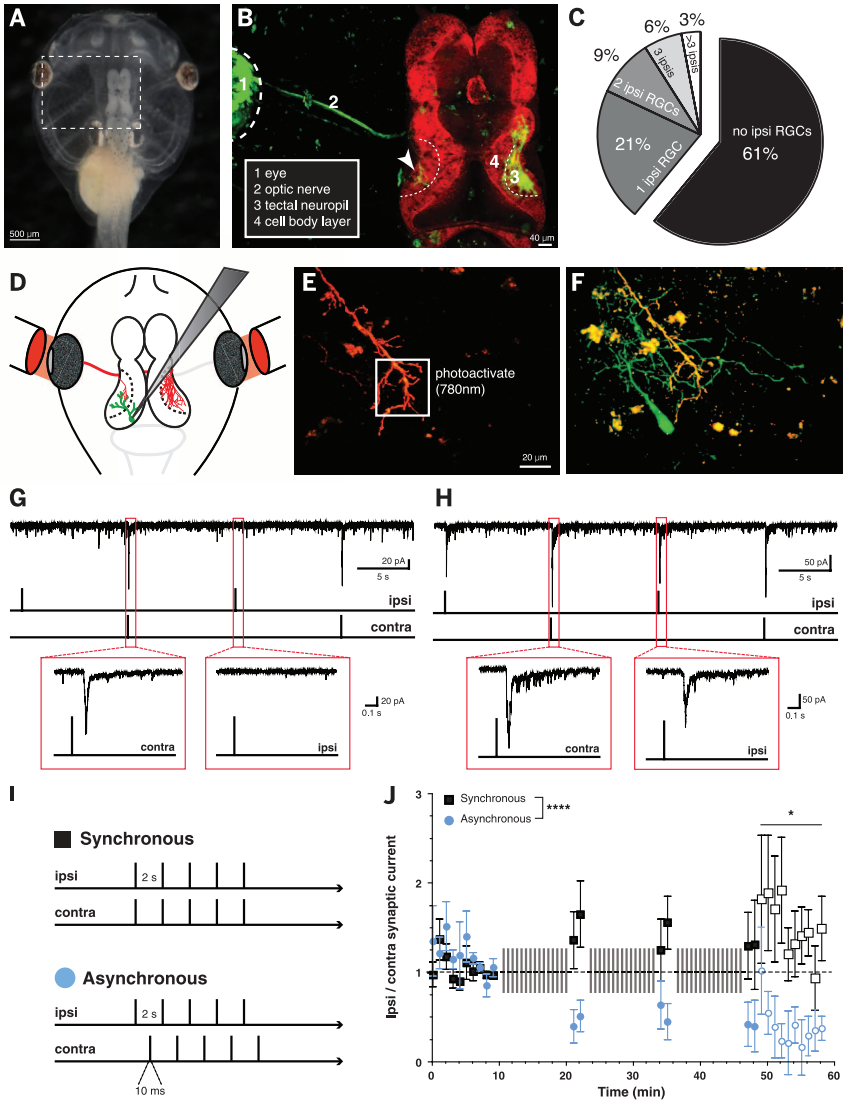


Fig. 1. Sporadic ipsilateral RGC axons can integrate into the retinotectal circuit. (**A**) Albino *Xenopus laevis* tadpole. (**B**) Retinotectal projection visualized by means of retinal electroporation of EGFP plasmid. Brain is stained with boron-dipyrromethene (red). Most RGC axons project contralaterally. Arrowhead indicates ipsilaterally projecting RGC axon. (**C**) Percent of tadpoles with ipsilaterally projecting RGC axons labeled by means of intravitreal DiI injection. (**D**) Schematic of targeted recordings with optical fibers to stimulate each eye. (**E**) Ipsilateral RGC axon (tdTomato) indicating photoactivation site. (**F**) Photoactivation back labels potential postsynaptic partners. (**G** and **H**) Visually evoked synaptic currents in tectal cells that are (G) nonresponsive or (H) responsive to light flashes in the ipsilateral eye. (**I**) Visual stimulation training protocols. A 10-ms flash was presented every 2 s to each eye, either in or out of phase. (**J**) Ratio of ipsilateral to contralateral eye–evoked currents (normalized to last 3 min of baseline) is depressed through 10 min of asynchronous visual stimulation [closed symbols; $^{****}P < 0.0001$, interaction by two-way analysis of variance (ANOVA) mixed design, $n = 12$ cells with synchronous and 10 cells with asynchronous stimulation]. This depression persists after training (open symbols; $^{*}P = 0.027$, main effect by two-way ANOVA mixed design, $n = 7$ cells with synchronous and 5 cells with asynchronous stimulation).

synaptic strength for the ipsilateral eye, whereas synchronous stimulation maintained the relative contribution of the ipsilateral eye at or above baseline levels (Fig. 1J and fig. S2). By the third round of visual training, five of the eight cells with asynchronously stimulated inputs, but none of the synchronously stimulated cases (n = 10 cells) were weakened to <20% of baseline synaptic strength. These changes stably persisted for at least 13 min after training stimulation (Fig. 1J, open symbols).

We next examined whether correlated activity also regulates the growth of developing RGC axons. After retinal electroporation of enhanced GFP (EGFP) plasmid, we selected animals with single EGFP-labeled ipsilaterally or contralaterally projecting axons (*16*). Using a video display directly beneath rearing tanks, free-swimming tadpoles were continuously presented with full-field stroboscopic flashes (0.5 Hz) to synchronize both eyes, or large moving black dots to independently activate the two eyes (Fig. 2A). Dot-rearing leads to asynchronous stimulation of the ipsilateral RGC axon but synchronous activation of contralateral RGC axons relative to other contralateral eye inputs. RGC axons were imaged in vivo with two-photon laser scanning microscopy daily for 5 days (Fig. 2B). All three groups that were reared under synchronizing conditions exhibited comparable axon growth and branch elaboration (Fig. 2, C and D). In comparison, dot-reared ipsilateral RGC axons, which experienced asynchronous stimulation relative to neighboring inputs, grew faster, with larger, more diffuse arbors by 2 days of stimulation.

To investigate axonal growth and branching at higher temporal resolution, we collected two-photon time-lapse images of individual ipsilateral axon arbors in tadpoles every 10 min for 5.5 hours while simultaneously presenting stimuli to the eyes through a pair of optical fibers (Fig. 3, A to C) (*16*). One of two stimulation protocols was applied (Fig. 3B). In the first protocol [dark-asynchronous-synchronous (DAS)], baseline images in darkness were collected for 90 min, after which each eye was stimulated at 0.5 Hz with a 5-ms light flash. Asynchronous stimulation (1 s apart) was presented for the first 2 hours, followed by synchronous stimulation for the last 2 hours. The second protocol [dark-synchronous-asynchronous (DSA)] was identical, except that the 2-hour asynchronous stimulation followed the synchronous stimulation.

Under the DAS protocol, asynchronous visual stimulation rapidly produced a robust increase in axon branch dynamics compared with that in darkness (Fig. 3D and movie S1). Within 20 min of asynchronous stimulation, the rate of new branch additions significantly increased (Fig. 3, E and F). Branches were also eliminated more rapidly during asynchronous stimulation (Fig. 3, G and H), which is consistent with an overall increase in dynamic remodeling. Furthermore, asynchronous stimulation significantly augmented the elongation lengths of branch tips as compared with growth in darkness (Fig. 3, I and J).

Changing from asynchronous to synchronous stimulation, without altering stimulation intensity or frequency, significantly decreased the numbers of branches added (Fig. 3, E and F) and lost (Fig. 3, G and H), in addition to reducing branch elongation (Fig. 3, I and J, and movie S2). The decrease in newly added branch tips was more gradual, only reaching significance by 40 min after the onset of synchronized stimulation. Consistent with these observations, animals that were presented first with synchronous stimulation (DSA protocol) did not show a significant increase in branch additions (Fig. 3, E and F) or elongation (Fig. 3, I and J) until asynchronous stimulation was presented.

Although synchronous stimulation decreased the addition of new branches compared with asynchronous stimulation (Fig. 3F), new branches formed during synchronous stimulation were in fact more stable than those added during asynchronous stimulation. They had longer lifetimes, with a significantly larger fraction surviving longer than 30 min (Fig. 3, K and L). Thus, activation of a RGC rapidly results in highly dynamic exploratory branches when its firing is mismatched to that of surrounding inputs, but coactivation with its neighbors suppresses this probing and stabilizes contacts.

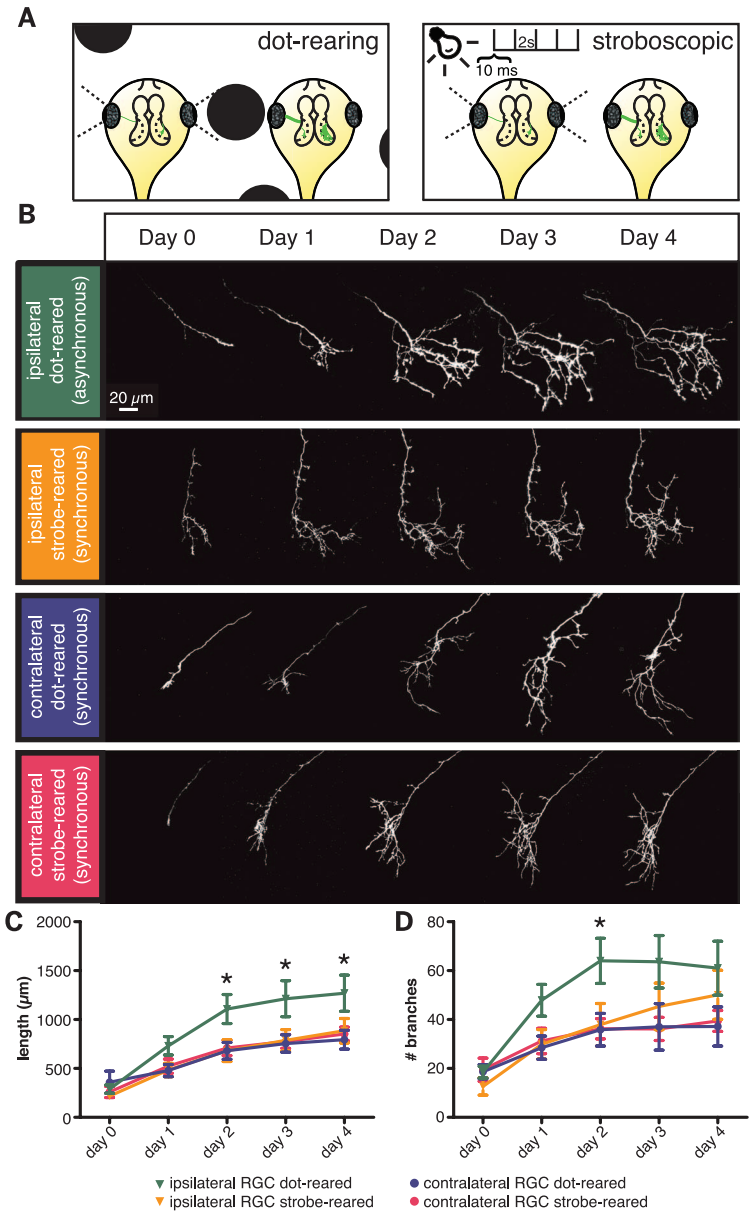


Fig. 2. Rearing tadpoles with asynchronous binocular stimulation enlarges ipsilateral RGC axon arbors. (**A**) Animals were reared for 4 days with stroboscopic or moving-dot stimuli. (**B**) Ipsilaterally and contralaterally projecting RGC axons from animals reared under each experimental condition, imaged once daily. (**C**) Total axonal arbor size and (**D**) branch-tip number were greatest in the asynchronous stimulation condition (ipsilateral axon, dot-reared) (*P < 0.05, two-way ANOVA mixed design, indicating a difference between ipsilateral dot-reared and all other groups by Tukey's post-test, n = 8 axons for ipsilateral strobe, 8 axons for ipsilateral dots, 6 axons for contralateral strobe, and 12 axons for contralateral dots).

Axonal growth may be controlled by retrograde cues from postsynaptic cells (*19*). One option is that RGCs up-regulate dynamic growth in response to increased firing, but synaptic activity correlated with the postsynaptic cell leads to branch-suppressing or -stabilizing retrograde signals that inhibit this growth. Another possibility is that asynchronous input may be detected by the postsynaptic neuron, prompting it to deliver growth-promoting signals to the dissenting axon terminal. These two putative mechanisms predict opposite outcomes of blocking synaptic transmission between an axon and its partner, with increased branching under the first model and reduced branching under the second.

We tested the effects of blocking synaptic transmission by expressing tetanus neurotoxin light chain fused to EGFP (TeNT-Lc:EGFP) (*5*) in the ipsilaterally projecting RGC axon (Fig. 4A and movie S3). Because synaptic NMDARs have been proposed to act as correlation detectors because of their voltage-dependent response to glutamate (*9*) and have previously been implicated in activity-dependent retinotectal map refinement (*7*, *13*), we also tested the effects of NMDAR blockade by treating animals with the blood-brain barrier permeant noncompetitive NMDAR antagonist MK-801 (Fig. 4B and movie S4). Because synchronized stimulation requires ~1 hour to achieve its full effect on branch addition (Fig. 3E) and loss (Fig. 3G) rates, we separately analyzed dynamics during the first and second hours of each stimulation period (Fig. 4, C and D). Control ipsilateral axons significantly increase dynamic branch additions (Fig. 4C) and losses (Fig. 4D) during the first hour of asynchronous stimulation and return to baseline levels by the second hour of synchronous stimulation. In contrast, both TeNT-Lc:EGFP–expressing and MK-801–treated axons showed their largest increases in branch additions and losses during synchronous stimulation. To further ascertain the relative responses of axons to asynchronous versus synchronous stimulation on a cell-by-cell basis, we divided the mean rate of branch addition or loss during the last 90 min of asynchronous stimulation by that during the last 90 min of synchronous stimulation so as to generate addition and loss ratios for each cell (Fig. 4, E and F). Expression of TeNT-Lc:EGFP and blockade of NMDARs both reduced these ratios to ~1, indicating that these axons could not differentiate synchronous and asynchronous stimulation. We also examined the stability of branches in TeNT-Lc:EGFP–expressing (Fig. 4G) and MK-801–treated (Fig. 4H) ipsilateral RGC axons. Branch survival times did not significantly differ between stimulation conditions, which is in clear contrast to control axons, which form longer lasting branches under conditions of synchronous stimulation (Fig. 3, K and L). Taken together, these findings suggest that NMDAR-mediated synaptic transmission leads to increased branch stability and a reduction in branch dynamics during synchronous activation through the action of a branch-suppressing signal.

These experiments demonstrate how correlated neural activity helps orchestrate the morphological remodeling of developing axons into precisely organized maps. Sensory stimulation promotes exploratory branching and outgrowth of RGC axons within their target structure. Axons that may have extended into inappropriate territory, where their firing patterns do not match those of nearby inputs, would fail to maintain

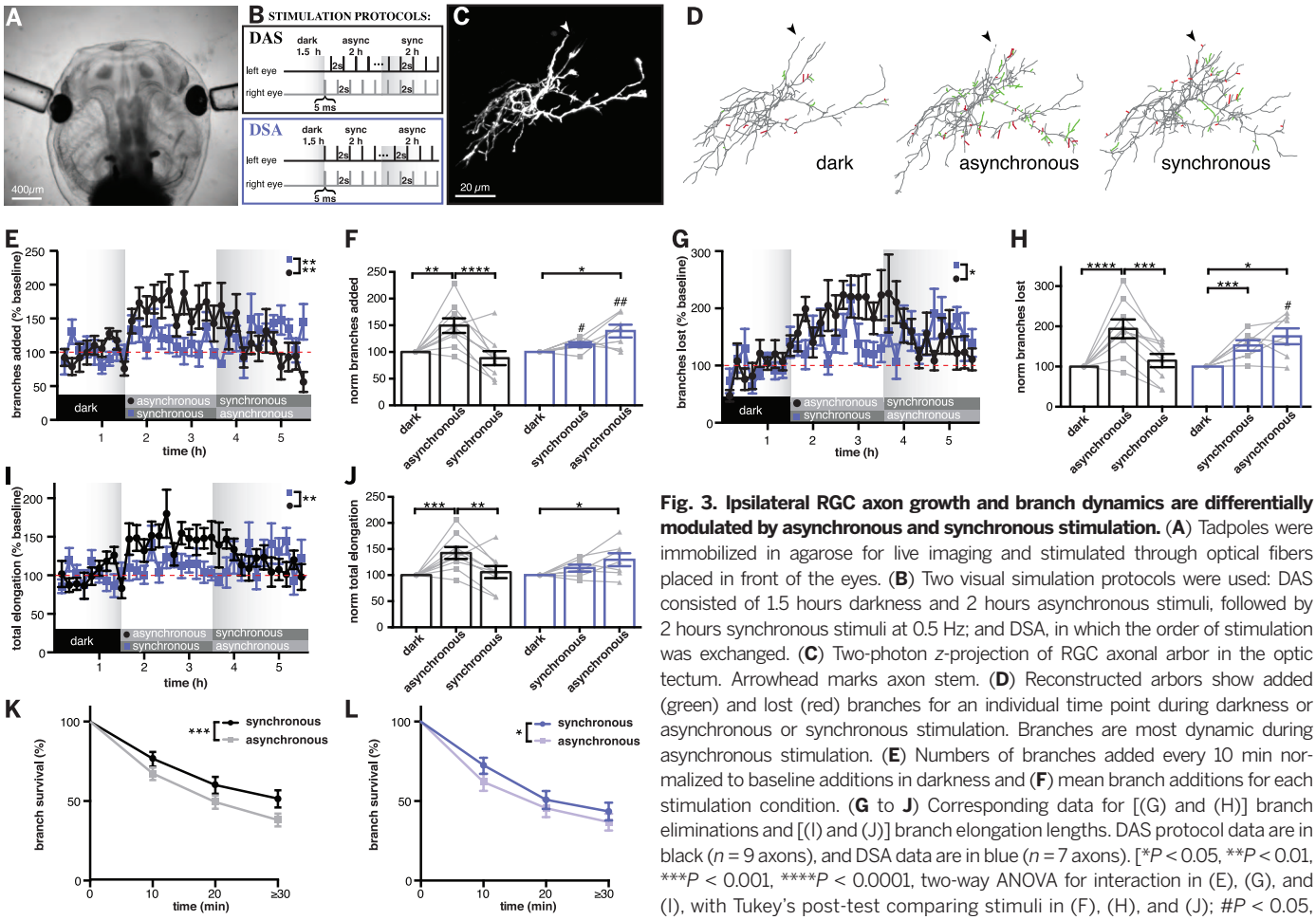


Fig. 3. Ipsilateral RGC axon growth and branch dynamics are differentially modulated by asynchronous and synchronous stimulation. (**A**) Tadpoles were immobilized in agarose for live imaging and stimulated through optical fibers placed in front of the eyes. (**B**) Two visual simulation protocols were used: DAS consisted of 1.5 hours darkness and 2 hours asynchronous stimuli, followed by 2 hours synchronous stimuli at 0.5 Hz; and DSA, in which the order of stimulation was exchanged. (**C**) Two-photon *z*-projection of RGC axonal arbor in the optic tectum. Arrowhead marks axon stem. (**D**) Reconstructed arbors show added (green) and lost (red) branches for an individual time point during darkness or asynchronous or synchronous stimulation. Branches are most dynamic during asynchronous stimulation. (**E**) Numbers of branches added every 10 min normalized to baseline additions in darkness and (**F**) mean branch additions for each stimulation condition. (**G** to **J**) Corresponding data for [(G) and (H)] branch eliminations and [(I) and (J)] branch elongation lengths. DAS protocol data are in black ($n = 9$ axons), and DSA data are in blue ($n = 7$ axons). [$^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$, two-way ANOVA for interaction in (E), (G), and (I), with Tukey's post-test comparing stimuli in (F), (H), and (J); $\#P < 0.05$, $\#\#P < 0.01$, Sidak post-test comparing protocols.] (**K** and **L**) Branch survival fraction shows more stable branches (≥30 min) formed under synchronous than asynchronous stimulation, with either the (K) DAS or (L) DSA stimulation protocols [(K) $n = 551$ asynchronous and 326 synchronous branches from 9 cells; (L) $n = 321$ asynchronous and 229 synchronous branches from 7 cells]. $^{*}P < 0.05$, $^{***}P < 0.001$ by log-rank test. Error bars indicate [(E) to (J)] SEM and [(K) and (L)] 95% confidence interval (CI).

stable functional and structural contacts and continue actively elaborating in search of appropriate partners. In contrast, axons that form synaptic contacts onto partners that receive other highly coactive inputs will engage cooperative mechanisms to stabilize those contacts and the branches on which they reside. Our experiments confirm that glutamate release and activation of NMDARs are critical steps in initiating a branch-stabilizing signal, although its molecular identity remains unknown. Because a single axon firing out of synchrony with numerous synchronized inputs does not appear to benefit from the stabilization signals that these coactive axons presumably receive, the stabilizing signal must be very precisely spatially or temporally restricted, which rules out long-lived, highly diffusible molecules as plausible candidates. The idea that this signal likely originates in the postsynaptic cell is supported by earlier experiments that demonstrated that calcium/calmodulin–dependent protein kinase II activity in tectal neurons can retrogradely modulate RGC axon growth (*20*). However, our data do not exclude a potential contribution by the surrounding glial cells (*21*) or the possibility that putative presynaptic NMDARs (*22*, *23*) on RGCs may be involved.

Our observation that visual stimulation drives a rapid increase in branching and growth is consistent with earlier studies in the retinotectal projections of zebrafish and mouse, in which suppression of RGC firing through expression of inward-rectifying potassium channels inhibited the dense elaboration of branches (*4*, *24*). Similarly, the enlarged arbors reported in zebrafish RGCs expressing TeNT-Lc match our findings that this treatment prevents the down-regulation of branching during correlated activity (*5*). These authors argued for a model based on activity-dependent competition to explain their data; however, our results suggest that correlation detection may offer an important alternative explanation for these findings. Activity-dependent competition is useful to explain pathological conditions such as amblyopia but likely plays a minor role in normal map development. A correlated stimulation approach similar to ours was recently reported that used transgenic mice expressing channelrhodopsin-2 in RGCs for stimulation during early postnatal development. The consequences of synchronous and asynchronous stimulation in that system were highly consistent with our findings, with synchronous stimulation leading to ectopic ipsilateral eye projections presumably stabilized in contralateral eye territory (*11*).

Here, we present live observations of axonal structural plasticity directed by patterned visual stimuli in vivo that support the Hebbian prediction that coactive inputs are stabilized (*25*). Our data also demonstrate the up-regulation of exploratory growth over days in the absence of correlated firing (*26*) and growth suppression when inputs are coactive (*27*). The speed with which physiological visual stimuli can drive such changes—reducing the strength of synaptic currents evoked through the ipsilateral eye after just 10 min of asynchronous visual stimulation and significantly increasing the rate of new branch addition in under 20 min—was unanticipated.

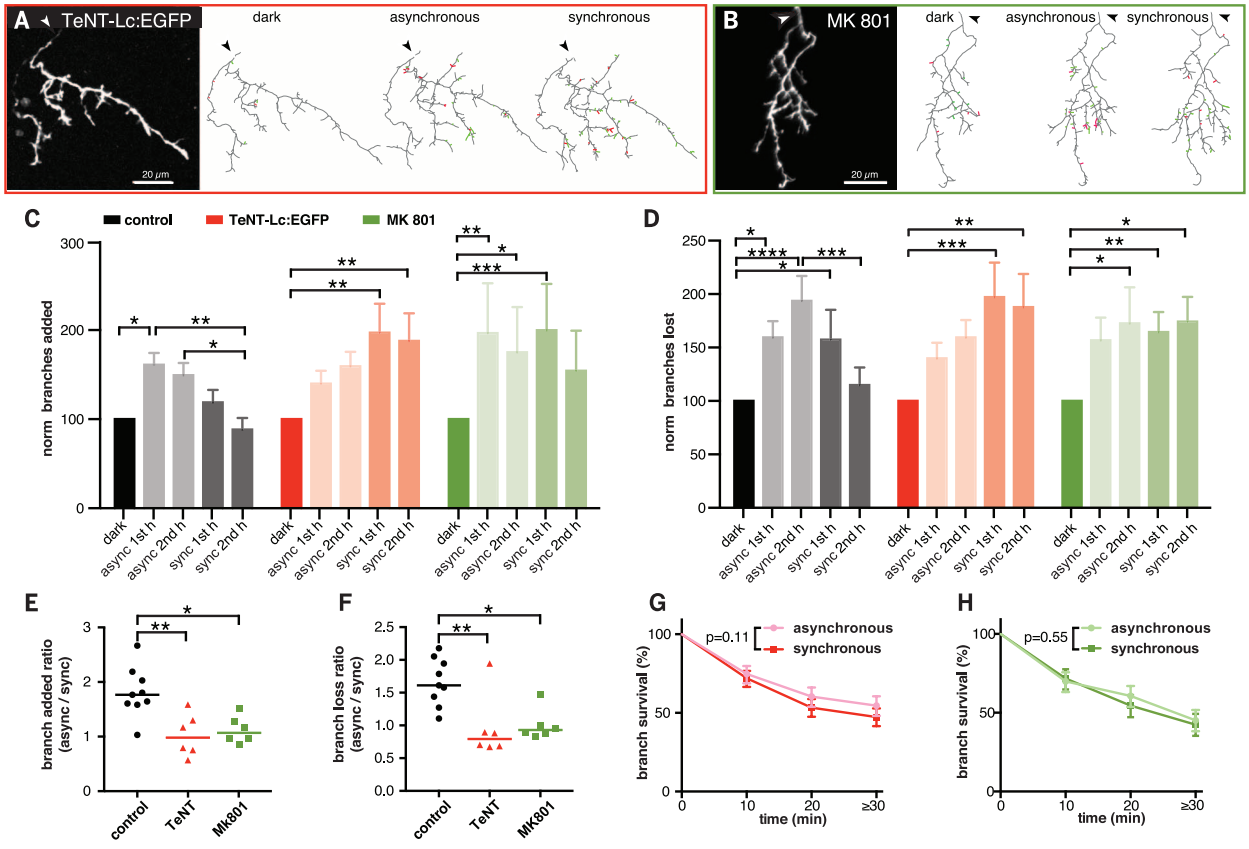


Fig. 4. Synaptic neurotransmission and NMDARs mediate arbor stabilization during synchronous stimulation. (**A** and **B**) Cells (A) expressing TeNT-Lc:EGFP or (B) bathed in MK-801, with example reconstructions under different stimulation conditions, as in Fig. 3. (**C** and **D**) Transmitter release and NMDAR activation are both necessary to reduce branch dynamics during synchronous stimulation. (C) Mean branch additions for each hour of stimulation normalized to darkness, for control (n = 9), TeNT-Lc:EGFP–expressing (n = 6), and MK-801–treated (n = 6) ipsilateral axons. (D) Corresponding data for branches lost (*P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001, two-way ANOVA mixed design with Tukey's post-test). (**E**) Ratios of numbers of branches added for each cell during asynchronous versus synchronous stimulation. (**F**) Ratios for branches lost (*P < 0.05, **P < 0.01, Kruskal-Wallis test with Dunn's post-test). (**G** and **H**) Survival plots for branches formed during asynchronous and synchronous stimulation in (G) TeNT-Lc:EGFP–expressing or (H) MK-801–treated EGFP-expressing cells. Unlike controls (Fig. 3, K and L), synchronous stimulation did not significantly enhance branch stability in these cells (log-rank test; n = 242 asynchronous and 300 synchronous branches from six TeNT-Lc:EGFP–expressing axons, n = 206 asynchronous and 191 synchronous branches from six MK-801–treated axons). Error bars indicate [(C) and (D)] SEM and [(G) and (H)] 95% CI.

Identifying the specific signals that mediate correlation-dependent structural plasticity will be greatly facilitated by exploiting the experimental protocol presented here.

REFERENCES AND NOTES

1. E. S. Ruthazer, H. T. Cline, *J. Neurobiol.* **59**, 134–146 (2004).
2. J. Cang, D. A. Feldheim, *Annu. Rev. Neurosci.* **36**, 51–77 (2013).
3. D. W. Sretavan, C. J. Shatz, M. P. Stryker, *Nature* **336**, 468–471 (1988).
4. J. Y. Hua, M. C. Smear, H. Baier, S. J. Smith, *Nature* **434**, 1022–1026 (2005).
5. N. B. Fredj *et al.*, *J. Neurosci.* **30**, 10939–10951 (2010).
6. H. T. Cline, M. Constantine-Paton, *Neuron* **3**, 413–426 (1989).
7. I. Rajan, S. Witte, H. T. Cline, *J. Neurobiol.* **38**, 357–368 (1999).
8. D. K. Simon, G. T. Prusky, D. D. O'Leary, M. Constantine-Paton, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10593–10597 (1992).
9. M. F. Bear, A. Kleinschmidt, Q. A. Gu, W. Singer, *J. Neurosci.* **10**, 909–925 (1990).
10. O. S. Dhande *et al.*, *J. Neurosci.* **31**, 3384–3399 (2011).
11. J. Zhang, J. B. Ackman, H.-P. Xu, M. C. Crair, *Nat. Neurosci.* **15**, 298–307 (2012).
12. D. Hebb, *The Organization of Behavior* (John Wiley and Sons, New York, 1949).
13. H. T. Cline, E. A. Debski, M. Constantine-Paton, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4342–4345 (1987).
14. L. A. Kirkby, G. S. Sack, A. Firl, M. B. Feller, *Neuron* **80**, 1129–1144 (2013).
15. D. S. Sakaguchi, R. K. Murphey, *J. Neurosci.* **5**, 3228–3245 (1985).
16. Materials and methods are available as supplementary materials on *Science* Online.
17. G. H. Patterson, J. Lippincott-Schwartz, *Science* **297**, 1873–1877 (2002).
18. M. Peter *et al.*, *PLOS ONE* **8**, e62132 (2013).
19. J.-P. Changeux, A. Danchin, *Nature* **264**, 705–712 (1976).
20. D. J. Zou, H. T. Cline, *J. Neurosci.* **19**, 8909–8918 (1999).
21. M. Tremblay *et al.*, *J. Neurosci.* **29**, 14066–14076 (2009).
22. H. Liu *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 8383–8387 (1994).
23. R. S. Larsen *et al.*, *Nat. Neurosci.* **14**, 338–344 (2011).
24. I. Benjumeda *et al.*, *J. Neurosci.* **33**, 18208–18218 (2013).
25. E. S. Ruthazer, C. J. Akerman, H. T. Cline, *Science* **301**, 66–70 (2003).
26. S. B. Udin, *Nature* **301**, 336–338 (1983).
27. E. A. Debski, H. T. Cline, M. Constantine-Paton, *J. Neurobiol.* **21**, 18–32 (1990).

ACKNOWLEDGMENTS

We thank the laboratory of K. Haas (UBC) for assistance with the MATLAB analysis tool Dynamo for dynamic morphometric analysis. We also thank M. Meyer for providing us with the tetanus toxin light chain plasmid (5UAS-TeNT-EGFP) and D. Freiheit for photography of tadpoles. We are grateful to S. Glasgow for advice on statistical analysis and to H. Cline and W. Sossin for useful comments on our manuscript. This work was supported by grants from the Canadian Institutes for Health Research and the Natural Sciences and Engineering Research Council (NSERC) of Canada to E.S.R. and by fellowships from the Deutscher Akademischer Austausch Dienst and the NSERC CREATE Neuroengineering Training Program to M.M., the Fonds de la recherche en santé du Québec to D.G., and NSERC to P.S. The authors declare that they have no conflicts of interest.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/904/suppl/DC1
Materials and Methods
Figs. S1 and S2
References
Movies S1 to S4

31 January 2014; accepted 25 April 2014
10.1126/science.1251593

MICROBIAL GENOMICS

Stop codon reassignments in the wild

Natalia N. Ivanova,[1]* Patrick Schwientek,[1]* H. James Tripp,[1] Christian Rinke,[1] Amrita Pati,[1] Marcel Huntemann,[1] Axel Visel,[1,2,3] Tanja Woyke,[1] Nikos C. Kyrpides,[1] Edward M. Rubin[1,2]†

The canonical genetic code is assumed to be deeply conserved across all domains of life with very few exceptions. By scanning 5.6 trillion base pairs of metagenomic data for stop codon reassignment events, we detected recoding in a substantial fraction of the >1700 environmental samples examined. We observed extensive *opal* and *amber* stop codon reassignments in bacteriophages and of *opal* in bacteria. Our data indicate that bacteriophages can infect hosts with a different genetic code and demonstrate phage-host antagonism based on code differences. The abundance and diversity of genetic codes present in environmental organisms should be considered in the design of engineered organisms with altered genetic codes in order to preclude the exchange of genetic information with naturally occurring species.

Since the discovery of the genetic code and protein translation mechanisms (*1*), a limited number of variations of the standard assignment between unique base triplets (codons) and their encoded amino acids and translational stop signals have been found in bacteria and phages (*2*–*7*). Given the apparent ubiquity of the canonical genetic code, the design of genomically recoded organisms with noncanonical codes has been suggested as a means to prevent horizontal gene transfer between laboratory and environmental organisms (*8*–*10*). It is also predicted that genomically recoded organisms are immune to infection by viruses, under the assumption that phages and their hosts must share a common genetic code (*6*). This paradigm is supported by the observation of increased resistance of genomically recoded bacteria to phages with a canonical code (*9*). Despite these assumptions and accompanying lines of evidence, it remains unclear whether differential and noncanonical codon usage represents an absolute barrier to phage infection and genetic exchange between organisms.

Our knowledge of the diversity of genetic codes and their use by viruses and their hosts is primarily derived from the analysis of cultivated organisms. This is due to our limited access to genome sequences from uncultivated organisms, which are estimated to account for 99% in prokaryotes (*11*). Advances in single-cell sequencing and metagenome assembly technologies have enabled the reconstruction of genomes of uncultivated bacterial and archaeal lineages (*12*–*14*) and the discovery of a previously unknown reassignment of TGA *opal* stop codons to glycine (*4*, *5*, *14*). These initial findings suggest that large-scale systematic studies of uncultivated microorganisms and viruses may reveal the extent and modes of divergence from the canonical genetic code operating in nature.

To explore alternative genetic codes, we carried out a systematic analysis of stop codon reassignments from the canonical TAG *amber*, TGA *opal*, and TAA *ochre* codons in assembled metagenomes and metatranscriptomes from environmental and host-associated samples, single-cell genomes of uncultivated bacteria and archaea, and a collection of viral sequences (Fig. 1A) (*15*). All sequence data were obtained from the Integrated Microbial Genomes (IMG) database (*16*). This global collection of sequences comprised 1776 samples from 145 studies, including 750 samples obtained from 17 human body sites (fig. S1) (*17*). In total, 5.6 terabases of sequence data, including 450 Gb of contiguous sequences (contigs) greater than 1 kb, were analyzed. All samples were classified into human-associated, other host–associated, soil, marine, or freshwater environments according to their metadata (*15*, *18*).

We used a statistic of increased coding potential under alternate genetic codes as calculated by *ab initio* gene finder Prodigal (*19*), which was selected for its low rate of false-positive predictions (*15*). Contigs showing significantly higher coding potential when annotated with modified translation tables were forwarded to filtering and quality control to confirm stop codon reassignment through multiple sequence alignments to known homologs from the National Center for Biotechnology Information protein database (Fig. 1A) (*15*).

By applying this approach to 450 Gb of contigs larger than 1 kb in size, we identified 31,415 contigs with evidence of stop coding reassignment, adding up to a total of 198 Mb of recoded DNA (Fig. 1A). No recoding was observed in the metatranscriptome data. Varying ratios of reassigned to total contigs were observed in samples from terrestrial and aquatic environments and from human mouth, throat, and stool microbiomes

[1]Department of Energy Joint Genome Institute (DOE JGI), Walnut Creek, CA 94598, USA. [2]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]School of Natural Sciences, University of California, Merced, CA 95343, USA.
*These authors contributed equally to this work. **†Corresponding author. E-mail: emrubin@lbl.gov**

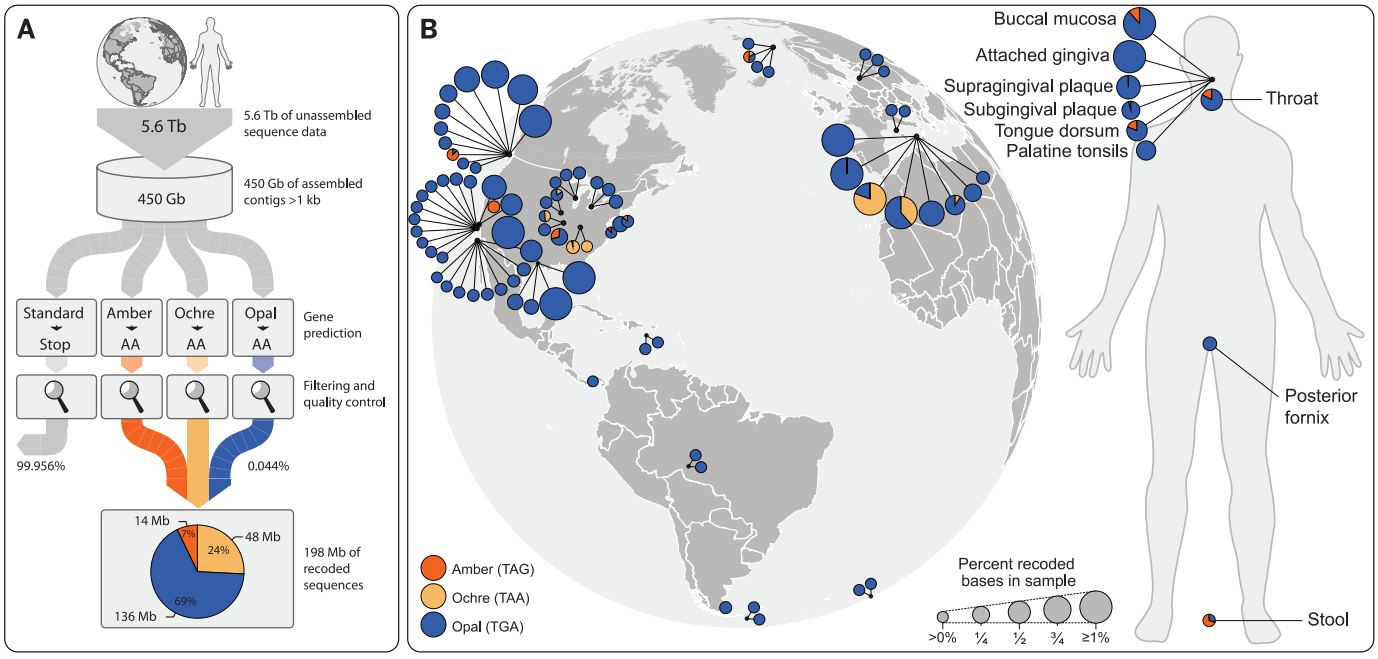


Fig. 1. Recoded DNA sequences identified worldwide. (**A**) Workflow used to identify contigs that contain stop codon reassignment. (**B**) Map showing the locations of 82 environmental samples around the globe together with nine sample sites (derived from 212 samples) of the human body for which recoded sequences have been identified.

(Fig. 1B and fig. S1). The greatest reassignment ratio was in a groundwater sample from a sulfidic aquifer, where 10.4% of all the assembled contigs displayed evidence that one of the three stop codons had been reassigned (table S2). High ratios of contig recoding were also detected in human oral microbiome (table S2).

Reassignment of all three stop codons was found but with different preferences by domain and habitat (Fig. 2). We observed distinct patterns of stop codon reassignment in the three domains of life, with bacteria showing only *opal* reassignments, *ochre* reassignments restricted to eukaryotes, and archaea devoid of codon reassignments (*15*). Among viruses, we found both *amber* and *opal* reassignments. These observations are restricted to DNA viruses only because of the scarcity of sequence information for RNA viruses of bacteria and archaea (Fig. 2) (*15*). Metagenomes of human body sites showed a high rate of reassignments compared with most other sampling sites. Only 10% of all contigs examined in this study originated from human body sites, but they represented 51% of all contigs with codon reassignments. The majority of the remaining stop codon reassignments were found in freshwater environments (44%), representing 13% (56.0 Gb) of all examined metagenomes. In contrast, marine samples contributed only 4% of recoded sequence, although they represent 48% (211.6 Gb) of the total data set (*15*). This suggests that codon reassignments are more abundant in freshwater than in marine samples.

Among bacteria, previous reports of recoding were restricted to the reassignment of *opal* stop codon (*3*–*5*, *20*). Despite our extensive sampling of bacterial sequences, we also observed reassignment

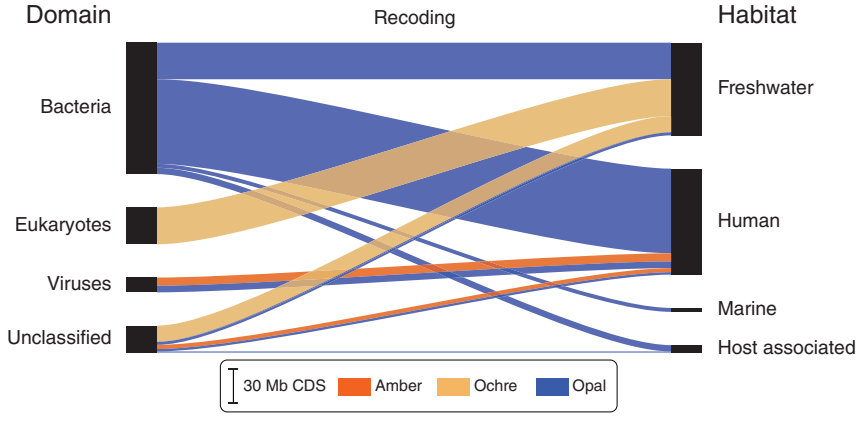


Fig. 2. Stop codon reassignment by taxonomy and habitat. Relative abundance of *amber*, *ochre*, and *opal* stop codon reassignments among bacteria, eukaryotes, and viruses in metagenomes of different habitats. For the sake of clarity, contig sets with less than 1 Mb total length for each combination of domain, stop codon, and habitat were excluded; see fig. S5, A and B (*15*).

exclusively for *opal* codons. *Opal* reassignments to Trp have been previously observed in *Mollicutes* (*20*) and *Candidatus* Hodgkinia cicadicola (*3*), and *opal* reassignments to Gly have been observed in uncultivated representatives of candidate phyla SR1 and Gracilibacteria (*4*, *5*). Our extensive survey suggests that *opal* reassignment in bacteria is likely limited to the same specific lineages (Fig. 3). The multiple SR1 and Gracilibacteria sequences enabled us to explore the evolutionary origin of stop codon reassignment in these closely related, uncultivated bacterial lineages. A maximum likelihood phylogenetic tree revealed that *opal* reassignment occurred in the last common ancestor of these sister lineages after its separation from the Peregrines (PERs) group and before the divergence of SR1 and Gracilibacteria (Fig. 3). The same phylogenetic analysis performed for the *opal* reassigned members of the class *Mollicutes* indicates a single reassignment event within the last common ancestor of the *Mycoplasmatales* and *Entomoplasmatales*.

Although the average GC content of the entire data set was 55%, recoded bacterial sequences had an average GC content of 32%, consistent with previous studies (*21*). In recoded contigs, we observed a shift to low-GC synonymous codons and/or a shift to low-GC nonsynonymous codons for amino acids with similar chemical properties, supporting the hypothesis that changes in the genomic GC content correlate with and may drive

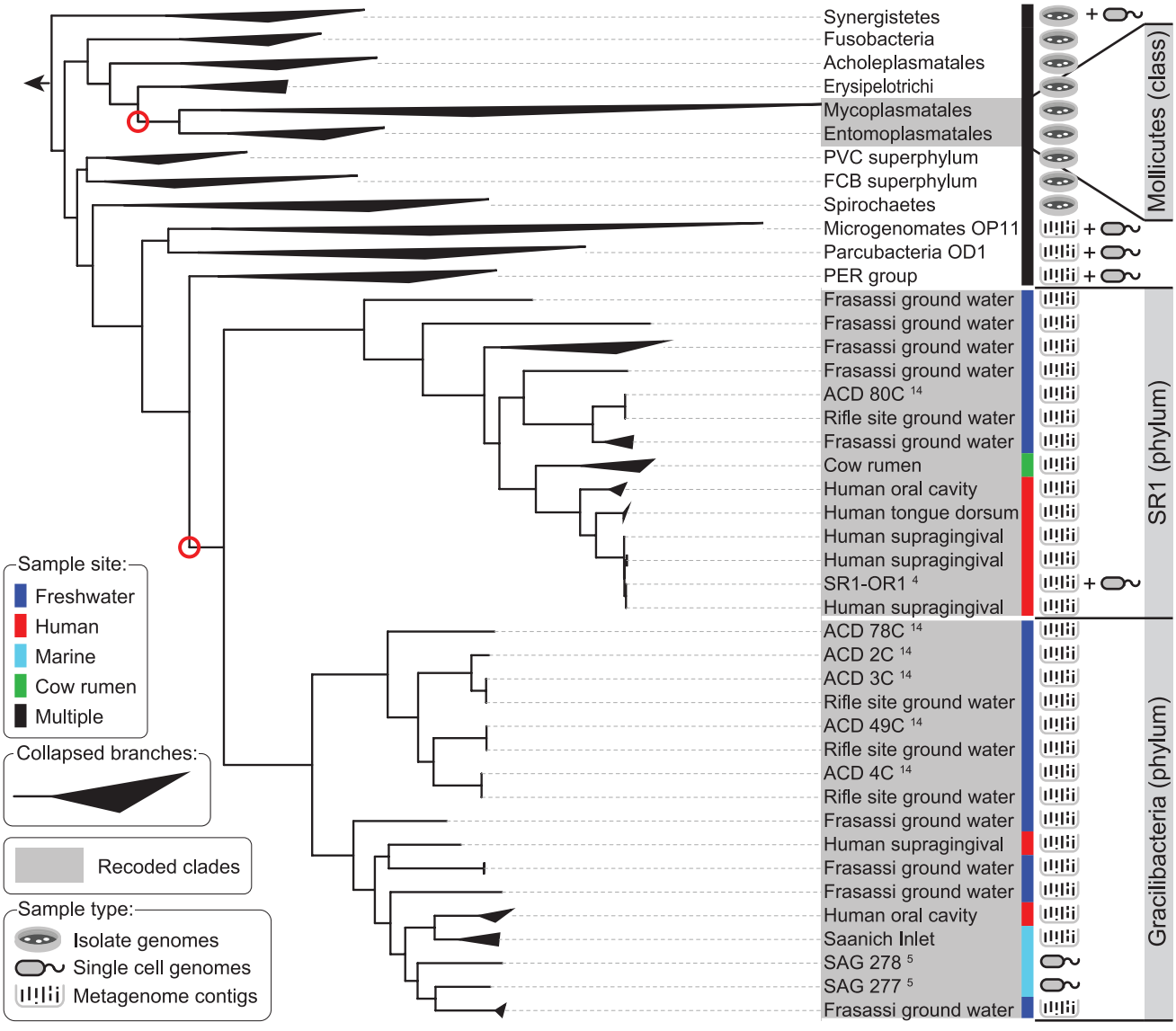


Fig. 3. A maximum likelihood phylogenetic tree of bacterial stop codon reassigned sequences, based on concatenated alignments of protein-coding marker genes. The arrow at the root of the tree points to the outgroup (Terrabacteria). The tree shows the recoded taxonomic groups *Mycoplasmatales* and *Entomoplasmatales* (*opal* to Trp), SR1 (*opal* to Gly), and Gracilibacteria (*opal* to Gly) along with non-recoded reference phyla. The highly reduced alphaproteobacterial *Candidatus* Hodgkinia cicadicola genome was not included. The red circles denote two reassignment events. PVC, *Planctomycetes*, *Verrucomicrobia*, and *Chlamydiae*; FCB, *Fibrobacteres*, *Chlorobi*, and *Bacteroidetes*. Sequences published in (*4*, *5*, *14*).

reassignment of stop codons (figs. S3 and S4). In addition, low-GC organisms used the *ochre* stop codon (TAA) to a higher extent (84% in organisms with GC content <32% versus 41% in organisms with GC content >64%) than *amber* (TAG) and *opal* (TGA). In extreme cases of low GC content, nearly all genes terminate in an *ochre* stop codon.

Although our pipeline for alternative genetic code detection was initially developed for prokaryotic genes, it can identify eukaryotic sequences with reassigned stop codons. In agreement with previous reports (*2*), among eukaryotic sequences we observed recoding for *opal* reassignment in mitochondrial sequences and *ochre* and *amber* reassignment in nuclear sequences. Nuclear sequences with stop codon reassignments appear to belong to the representatives of *Ciliophora* (table S3), and *ochre* reassignments were found exclusively in freshwater samples (Fig. 2).

We identified 19 complete and nearly complete DNA phages with *amber* stop codon reassignments in 177 out of 784 human microbiome samples (*15*). Previous reports of alternative genetic codes in DNA viruses are limited to *opal* reassignment to Trp in *Mycoplasma* phages (*7*, *22*). In our study, we identified two phages with *opal* reassignment to Gly, a phage with *amber* reassignment to Ser, and 14 phages with *amber* reassignment to Gln (table S8)—a code previously observed only in nuclear genes of eukaryotes (*23*). These reassignments are supported by protein alignments (figs. S6 to S8), as well as by the presence of *amber*-recognizing Gln-tRNA$_{CUA}$ and Ser-tRNA$_{CUA}$ in the phage contigs (figs. S9 and S10). We infer from the genome structure of recoded DNA phages that they belong to the order of *Caudovirales*. None of the *amber*-reassigned phage sequences was embedded into recognizable bacterial sequences, and no integrases were detected in the phage genomes, suggesting that these DNA phages are lytic.

Because phages largely depend on the translational machinery of their hosts, it has been suggested that they must use the same genetic code (*6*, *9*). Evidence supporting the matching usage of genetic codes between an *opal*-reassigned phage and its host was obtained by looking for footprints of phage infections in phage-derived spacers of the CRISPR (clustered regularly interspaced short palindromic repeat) adaptive immune system of bacteria (*24*). Out of 26 unique spacers found on CRISPR-harboring contigs with *opal* reassignment,

two spacers had an exact full-length match in the sequence of *opal*-recoded phages (table S9). Alignment of protein-coding genes on both contigs confirmed that they have *opal* to Gly reassignment.

The observation of *amber* reassignments in phages raises questions about the genetic code of their target hosts, given the apparent absence of *amber*-recoded bacterial genomes from environments in which *amber*-recoded phages were present (Fig. 2). This raises the possibility that genetic code differences between phages and hosts do not constitute an obligate barrier to phage infection. By analyzing 29,017 spacers found in CRISPR elements from 553 human oral and stool samples, we identified five spacers (each 33 to 37 bp long) that were identical to sequences from three different *amber*-recoded phage genomes (table S9). The contigs containing the CRISPR spacers also included bacterial genes with the full complement of canonical stop codons. The identified bacterial genes were nearly identical to genes from two *Prevotella* strains that were isolated from human airways and subgingival plaque and shown to have a standard genetic code. These data suggest that *amber*-reassigned phages can infect hosts with different genetic codes, in this case the standard code.

To gain further insight into mechanisms that may enable *amber*-recoded phages to infect hosts with different genetic codes, we examined the assembled genomes of *amber*-recoded phages. In several of these phage genomes, we identified genes for peptide chain release factor 2 (RF-2), which terminates translation at *ochre* and *opal* stop codons. Sequenced isolate phages lack genes for release factors, apparently harnessing the host-encoded release factors. The presence of RF-2 in a phage genome suggests that the phage may infect a host lacking RF-2; a hallmark of *opal*-reassigned bacterial genomes (*3*, *25*). Consistent with this possibility, the human oral cavity environments where *amber*-recoded RF-2–containing phages were detected lack *amber*-recoded bacteria but are enriched for *opal*-recoded bacteria. A further atypical feature noted in the genome of one of these phages (phage 2) is a bimodal pattern of *amber* reassignment across the genome (Fig. 4A). Initial annotation of this phage genome suggests that it is a lytic phage broadly related to T4 (fig. S11), in which *amber* has been reassigned to code for Gln.

This phage also contains a noncanonical Gln-tRNA$_{CUA}$, but closer examination of *amber* distribution across its genome reveals two large domains with distinct gene content and codon usage. The low-*amber* (LA) domain contains genes often found in early-stage phage infection, such as DNA polymerase. The LA domain also contains

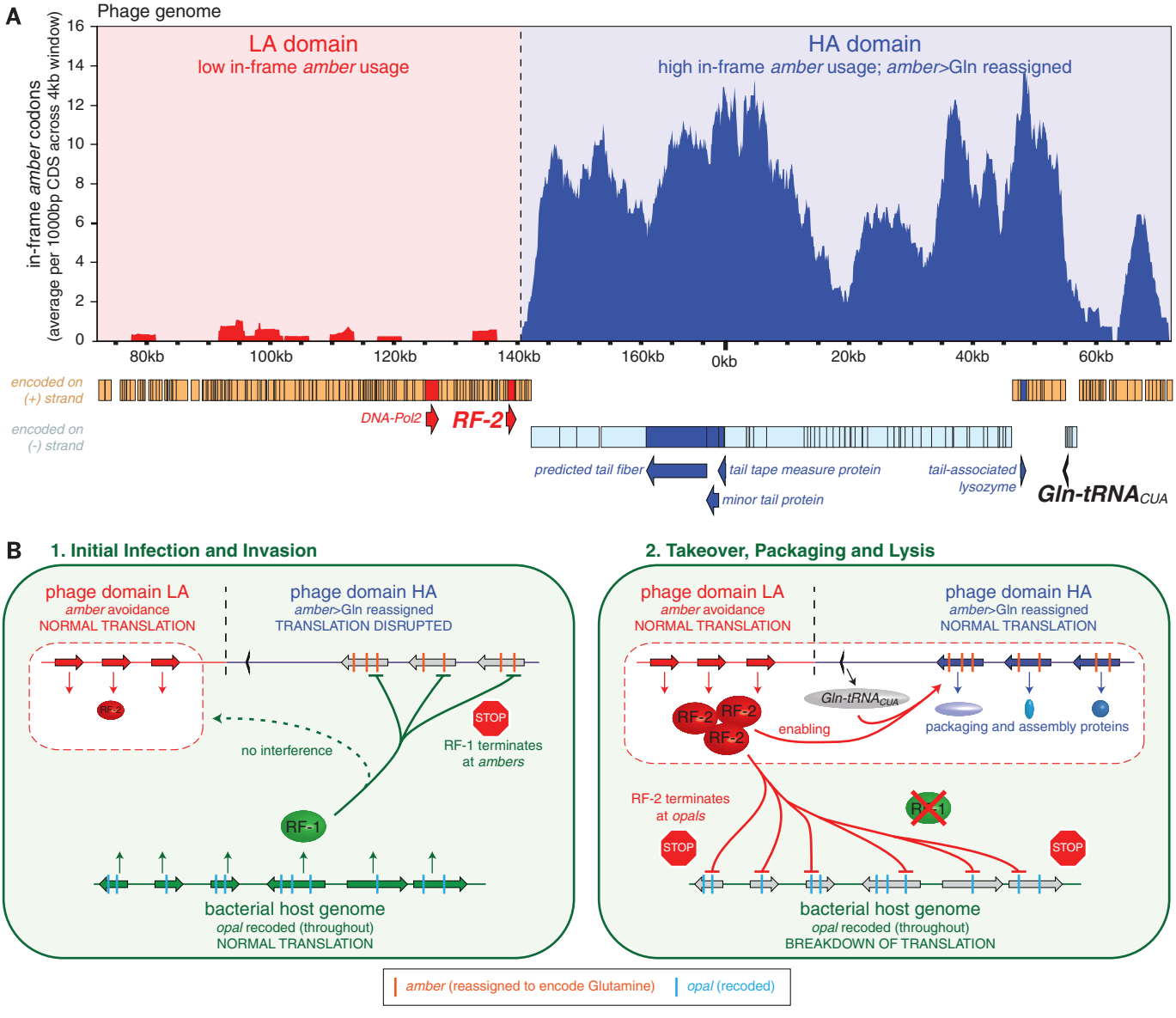


Fig. 4. Phage infections across genetic code boundaries. (**A**) Genome of phage 2. The phage genome is broadly divided into two domains with strong bias in codon utilization as well as strand preference. (**B**) Model of infection of *opal*-recoded hosts by *amber*>Gln-recoded phages.

the RF-2 gene required for normal translation of *amber*-recoded genes. Open reading frames in the LA domain are almost entirely devoid of in-frame *amber* codons and instead rely nearly exclusively on canonical glutamine codons to encode for glutamine (Fig. 4A). In contrast, the high-*amber* (HA) domain with frequent in-frame *amber* codons contains genes often associated with late stages of phage infection, such as packaging and assembly components (e.g., predicted tail fiber protein, minor tail protein, tail tape measure protein, or tail-associated lysozyme) (Fig. 4A). This distinct codon utilization, combined with the presence of RF-2 and a Gln-tRNA$_{CUA}$ in the *amber*-recoded phage, suggests that the *amber*-recoded phage actively interferes with the translation of its presumed *opal*-recoded host through a proposed mode of phage-host antagonism (Fig. 4B). In this model, upon initial phage infection abundant host-derived RF-1 (the releasing factor that terminates peptide chain elongation at *amber* codons) interferes with the translation of *amber*-containing phage HA domain genes, so they are initially not expressed. In contrast, critical *amber*-free phage LA domain genes can be normally translated. Next, phage-derived expression of RF-2 increasingly interferes with translation of *opal*-recoded host genes. Last, the simultaneous depletion in host-derived RF-1 and the increasing availability of phage-derived Gln-tRNA$_{CUA}$ enable the efficient production of assembly and packaging proteins from the phage HA-domain. Although direct in vivo observations of such processes remain to be established, this evidence supports a mechanism of phage-host antagonism in which the host's viability is undermined by the phage through the targeted codon-based disruption of the translation of the host's genetic code.

This survey of environmental sequence data revealed the abundance and diversity of stop codon reassignments in prokaryotes and phages. Several lines of evidence suggest that phages are not obligated to adapt to the codon usage of their hosts and that phages can exploit differences in codon usage to manipulate their hosts. Recently, genomically recoded organisms were created in an attempt to isolate the organism's genetic information from horizontal transfer to natural organisms and viruses (*9*). The diversity and abundance of recoding among uncultured environmental microbes and their phages suggests that even synthetic genomically recoded organisms (*9*) may not be immune to the exchange of genetic information with microbes and phages that populate many ecosystems.

REFERENCES AND NOTES

1. M. Nirenberg *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **53**, 1161–1168 (1965).
2. R. D. Knight, S. J. Freeland, L. F. Landweber, *Nat. Rev. Genet.* **2**, 49–58 (2001).
3. J. P. McCutcheon, B. R. McDonald, N. A. Moran, *PLOS Genet.* **5**, e1000565 (2009).
4. J. H. Campbell *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 5540–5545 (2013).
5. C. Rinke *et al.*, *Nature* **499**, 431–437 (2013).
6. L. A. Shackelton, E. C. Holmes, *J. Theor. Biol.* **254**, 128–134 (2008).
7. L. L. Voelker, K. Dybvig, *Gene* **233**, 101–107 (1999).
8. M. J. Lajoie *et al.*, *Science* **342**, 361–363 (2013).
9. M. J. Lajoie *et al.*, *Science* **342**, 357–360 (2013).
10. J. S. Dymond *et al.*, *Nature* **477**, 471–476 (2011).
11. M. S. Rappé, S. J. Giovannoni, *Annu. Rev. Microbiol.* **57**, 369–394 (2003).
12. M. Albertsen *et al.*, *Nat. Biotechnol.* **31**, 533–538 (2013).
13. T. Woyke *et al.*, *PLOS ONE* **4**, e5299 (2009).
14. K. C. Wrighton *et al.*, *Science* **337**, 1661–1665 (2012).
15. Materials and methods are available as supplementary materials on *Science* Online.
16. V. M. Markowitz *et al.*, *Nucleic Acids Res.* **40**, D123–D129 (2012).
17. Human Microbiome Project Consortium, *Nature* **486**, 215–221 (2012).
18. N. Ivanova *et al.*, *Environ. Microbiol.* **12**, 1803–1805 (2010).
19. D. Hyatt *et al.*, *BMC Bioinformatics* **11**, 119 (2010).
20. J. M. Bové, *Clin. Infect. Dis.* **17** (suppl. 1), S10–S31 (1993).
21. S. Osawa, T. H. Jukes, K. Watanabe, A. Muto, *Microbiol. Rev.* **56**, 229–264 (1992).
22. A. H. Tu, L. L. Voelker, X. Shen, K. Dybvig, *Plasmid* **45**, 122–126 (2001).
23. S. U. Schneider, M. B. Leible, X. P. Yang, *Mol. Gen. Genet.* **218**, 445–452 (1989).
24. L. A. Marraffini, E. J. Sontheimer, *Nat. Rev. Genet.* **11**, 181–190 (2010).
25. Y. Inagaki, Y. Bessho, S. Osawa, *Nucleic Acids Res.* **21**, 1335–1338 (1993).

ACKNOWLEDGMENTS

We thank the DOE JGI production sequencing, IMG, and Genomes OnLine Database teams for their support and J. Kim, A. Tadmor, and A. Nord for reviewing the manuscript. The work conducted by the DOE JGI was supported in part by the Office of Science of DOE under contract DE-AC02-05CH11231. Supporting data can be accessed through www.jgi.doe.gov and can be downloaded from http://portal.nersc.gov/dna/microbial/prokpubs/recoding.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/909/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S12
Tables S1 to S11
References (*26*–*53*)

10 January 2014; accepted 23 April 2014
10.1126/science.1250691

CANCER GENOMICS

Activating hotspot L205R mutation in *PRKACA* and adrenal Cushing's syndrome

Yanan Cao,[1]* Minghui He,[2]* Zhibo Gao,[2]* Ying Peng,[1] Yanli Li,[1] Lin Li,[2] Weiwei Zhou,[1] Xiangchun Li,[2] Xu Zhong,[1] Yiming Lei,[2] Tingwei Su,[1] Hang Wang,[2] Yiran Jiang,[1] Lin Yang,[2] Wei Wei,[1] Xu Yang,[2] Xiuli Jiang,[1] Li Liu,[2] Juan He,[1] Junna Ye,[1] Qing Wei,[3] Yingrui Li,[2] Weiqing Wang,[1]† Jun Wang,[2,4,5,6,7]† Guang Ning[1,8]†

Adrenal Cushing's syndrome is caused by excess production of glucocorticoid from adrenocortical tumors and hyperplasias, which leads to metabolic disorders. We performed whole-exome sequencing of 49 blood-tumor pairs and RNA sequencing of 44 tumors from cortisol-producing adrenocortical adenomas (ACAs), adrenocorticotropic hormone–independent macronodular adrenocortical hyperplasias (AIMAHs), and adrenocortical oncocytomas (ADOs). We identified a hotspot in the *PRKACA* gene with a L205R mutation in 69.2% (27 out of 39) of ACAs and validated in 65.5% of a total of 87 ACAs. Our data revealed that the activating L205R mutation, which locates in the P+1 loop of the protein kinase A (PKA) catalytic subunit, promoted PKA substrate phosphorylation and target gene expression. Moreover, we discovered the recurrently mutated gene *DOT1L* in AIMAHs and *CLASP2* in ADOs. Collectively, these data highlight potentially functional mutated genes in adrenal Cushing's syndrome.

Cushing's syndrome is caused by excessive glucocorticoid production, which may lead to a series of metabolic disorders such as obesity, glucose intolerance, and hypertension. Adrenal Cushing's syndrome results from autonomous production of cortisol from adrenocortical tumors (ACTs), which are most common in adult females. Cortisol-producing ACTs include benign adrenocortical adenoma (ACA), malignant adrenocortical carcinoma (ACC), and rare forms of bilateral adrenal hyperplasia (BAH) and adrenocortical oncocytoma (ADO) (*1*). BAHs consist of macronodular and micronodular hyperplasias, such as adrenocorticotropin-independent macronodular adrenocortical hyperplasia (AIMAH) and primary pigmented nodular adrenocortical disease (PPNAD).

Several genetic alterations have been described in inherited or sporadic BAHs. The inactivating mutations in *PRKAR1A* [cyclic adenosine monophosphate (cAMP)–dependent protein kinase regulatory subunit type Iα] are dominant in

[1]Shanghai Clinical Center for Endocrine and Metabolic Diseases, Shanghai Key Laboratory for Endocrine Tumors, Rui-Jin Hospital, Shanghai Jiao-Tong University School of Medicine, Shanghai, China. [2]BGI-Shanghai, BGI-Shenzhen, Shenzhen, China. [3]Department of Pathology, Rui-Jin Hospital, Shanghai Jiao-Tong University School of Medicine, Shanghai, China. [4]Department of Biology, University of Copenhagen, Copenhagen, Denmark. [5]King Abdulaziz University, Jeddah, Saudi Arabia. [6]Macau University of Science and Technology, Macau, China. [7]Department of Medicine, University of Hong Kong, Hong Kong. [8]Laboratory of Endocrinology and Metabolism, Institute of Health Sciences, Shanghai Institutes for Biological Sciences (SIBS), Chinese Academy of Sciences (CAS), and Shanghai Jiao Tong University School of Medicine (SJTUSM), Shanghai, China.
*These authors contributed equally to this work. **†Corresponding author. E-mail: guangning@medmail.com.cn (G.N.); wangj@genomics.org.cn (J.W.); wqingw@hotmail.com (W.W.)**

patients with Carney complex and PPNAD. Germline or somatic mutations in phosphodiesterase genes (*PDE11A* and *PDE8B*) and in *GNAS* are associated with adrenal hyperplasia (*2*, *3*). However, the genetic architecture of adrenal Cushing's syndrome remains largely uncharacterized, hampering the development of diagnostic and therapeutic approaches for Cushing's syndrome.

To investigate the genetic lesions in adrenal Cushing's syndrome, we performed whole-exome sequencing of DNA from 49 cortisol-producing ACTs and matched blood pairs, including 39 ACAs, 7 AIMAHs, and 3 ADOs (table S1). The histological subtypes of all tumors were confirmed by pathological examination. The average sequencing depth was 169× (82× to 277×), and 95.61% (91.20 to 99.50%) of the target regions were covered by at least 10× (table S2 and fig. S1). Meanwhile, we conducted RNA sequencing (RNA-seq) on 44 ACTs, including 36 whole-exome sequenced tumors, and obtained an average of 12.6 (9.4 to 17.4) Gb of sequenced data per sample (table S3).

We identified a total of 425 candidate somatic mutations among the 49 samples, including 97 synonymous, 298 missense, 15 nonsense, 3 splicing, 4 frameshift, and 8 in-frame, resulting in an average of 8.7 (0 to 28) somatic mutations in exonic regions (tables S4 and S5). We validated 55 of 57 predicted somatic mutations (96.5%) by Sanger sequencing and also verified a subset of highly expressed mutations using RNA-seq data (table S4). The predominant substitution in these mutations was the C:G > T:A transversion, consistent with most cancer types (fig. S2).

We defined 12 potentially functional mutated genes, including 8 recurrently mutated genes, 3 consensus cancer genes with potentially damaging mutations, and the only mutated gene in AIMAH3 (*4*). *PRKACA* (cAMP-dependent protein kinase catalytic subunit α) was the only significantly mutated gene in our study (false discovery rate $q = 3.77 \times 10^{-11}$) (*4*). Strikingly, a hotspot somatic c.T617G/p.L205R mutation in *PRKACA* was discovered in 55.1% of sequenced ACTs (27 of 49) and exclusively in 69.2% of ACAs (27 of 39). Further screening revealed this somatic c.T617G/p.L205R mutation in 65.5% of ACAs (57 of 87) and none in AIMAHs (n = 13), ACCs (n = 16), and ADOs (n = 3) by Sanger sequencing (fig. S3

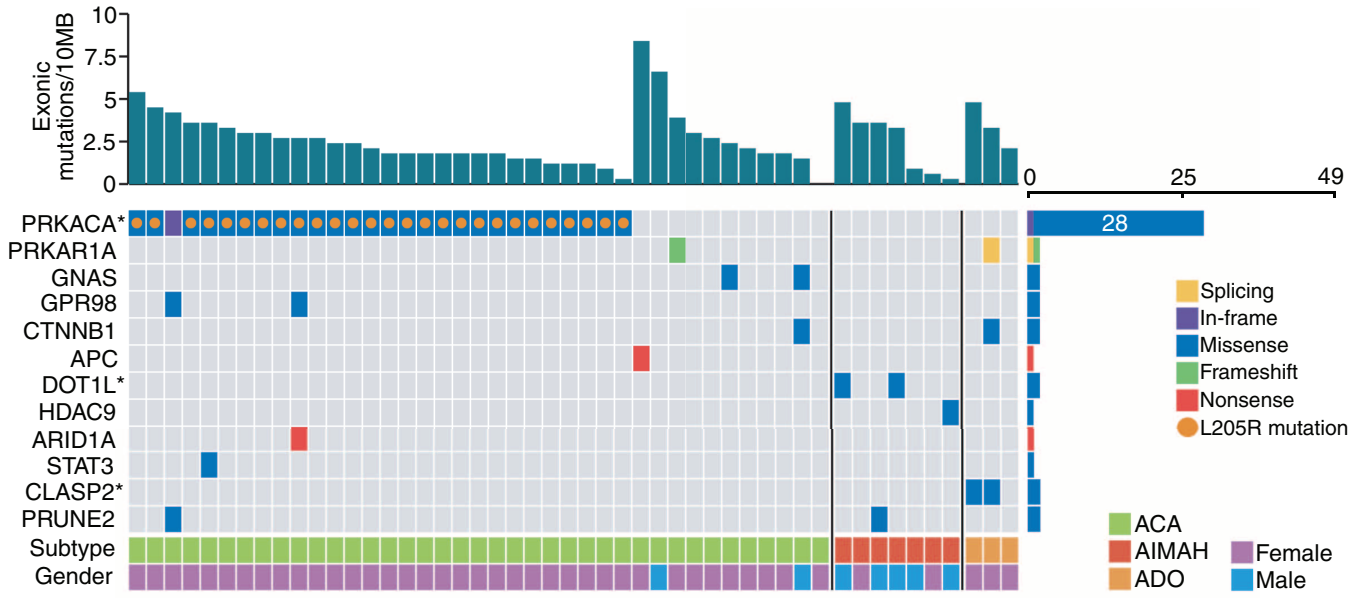
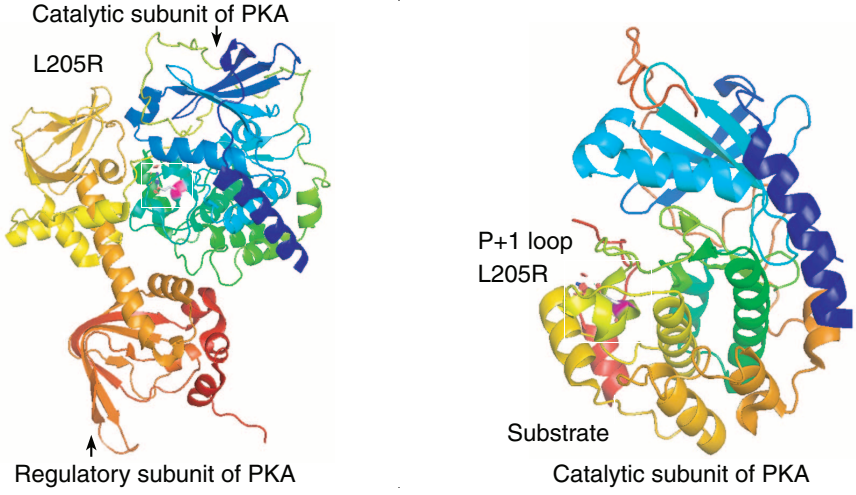


Fig. 1. *PRKACA* L205R mutations in ACAs and landscape of somatic mutations in cortisol-producing ACTs. Candidate cancer genes (*4*) are listed vertically on the left. Columns represent samples, with the average mutation rate in exon regions at the top. Samples are arranged from left to right by subtypes and mutation numbers. Histological subtypes and gender of each patient are reported at the bottom (green, ACA; red, AIMAH; orange, ADO; purple, female; blue, male). Colored rectangles indicate mutation category observed in the cohort. Asterisks indicate significantly mutated genes in tumor subtypes (*PRKACA*, 28/39 in ACA; *DOT1L*, 2/7 in AIMAH; *CLASP2*, 2/3 in ADO).

Fig. 2. L205R mutation occurs in the catalytic subunit of PKA (PKA C). Crystal structures of the interaction between catalytic and regulatory subunits of PKA (left, PDB: 2QCS) and a binding substrate peptide (right, PDB: 1L3R) are shown. The L205R mutation in the P+1 loop is shown in pink; the substrate peptide derived from inhibitor PKI is in red.



and table S6). The L205R ($Lys^{205} \rightarrow Arg$) mutations were mainly detected in adult females (Fisher's exact test, $P = 0.017$), and there were no significant differences in serum cortisol, ACTH, and urinary free cortisol levels between ACA patients with and without the *PRKACA* mutation (tables S7 and S8). In addition, only one somatic *PRKAR1A* mutation was observed in an ACA. Thus, the *PRKACA* L205R mutation was believed to be the dominant genetic alteration in sporadic ACAs.

Activating *CTNNB1* mutations are frequently observed in benign and malignant ACTs (*5*). Germline *APC* mutations in patients with familial adenomatous polyposis could result in the emergence of ACT (*6*). We identified activating *CTNNB1* mutations ($Ser^{33} \rightarrow Phe$, $Ser^{45} \rightarrow Ala$) in ACA and ADO, and an *APC* truncating mutation in ACA (fig. S4 and table S4). In addition, activating mutations in Arg^{201} and Gln^{227} of *GNAS* (encodes stimulatory G protein α subunit), resulting in PKA activation, have been reported in McCune-Albright syndrome and in adrenal and pituitary tumors (*7*, *8*). We found two activating *GNAS* mutations ($Arg^{201} \rightarrow His$, $Arg^{201} \rightarrow Cys$) in two *PRKACA* wild-type ACAs (Fig. 1 and table S4).

We also uncovered novel mutated genes in cortisol-producing ACTs. In the ACAs, we identified mutations in cancer-related genes, including *ARID1A* and *STAT3*. We also detected recurrently mutated *GPR98*, colocalizing with *PRKACA* mutations in ACAs. In the AIMAHs, we identified two mutations in the highly conserved methyltransferase domain of *DOT1L*, a histone H3 Lys^{79} methyltransferase (Fig. 1 and fig. S5). DOT1L regulates gene transcription and cell proliferation and mediates leukemic transformation in MLL (mixed lineage leukemia)–rearranged leukemias (*9*, *10*). Moreover, the histone deacetylase *HDAC9* was mutated in one AIMAH3 (table S4). These data suggest that deregulated chromatin modification might contribute to tumorigenesis

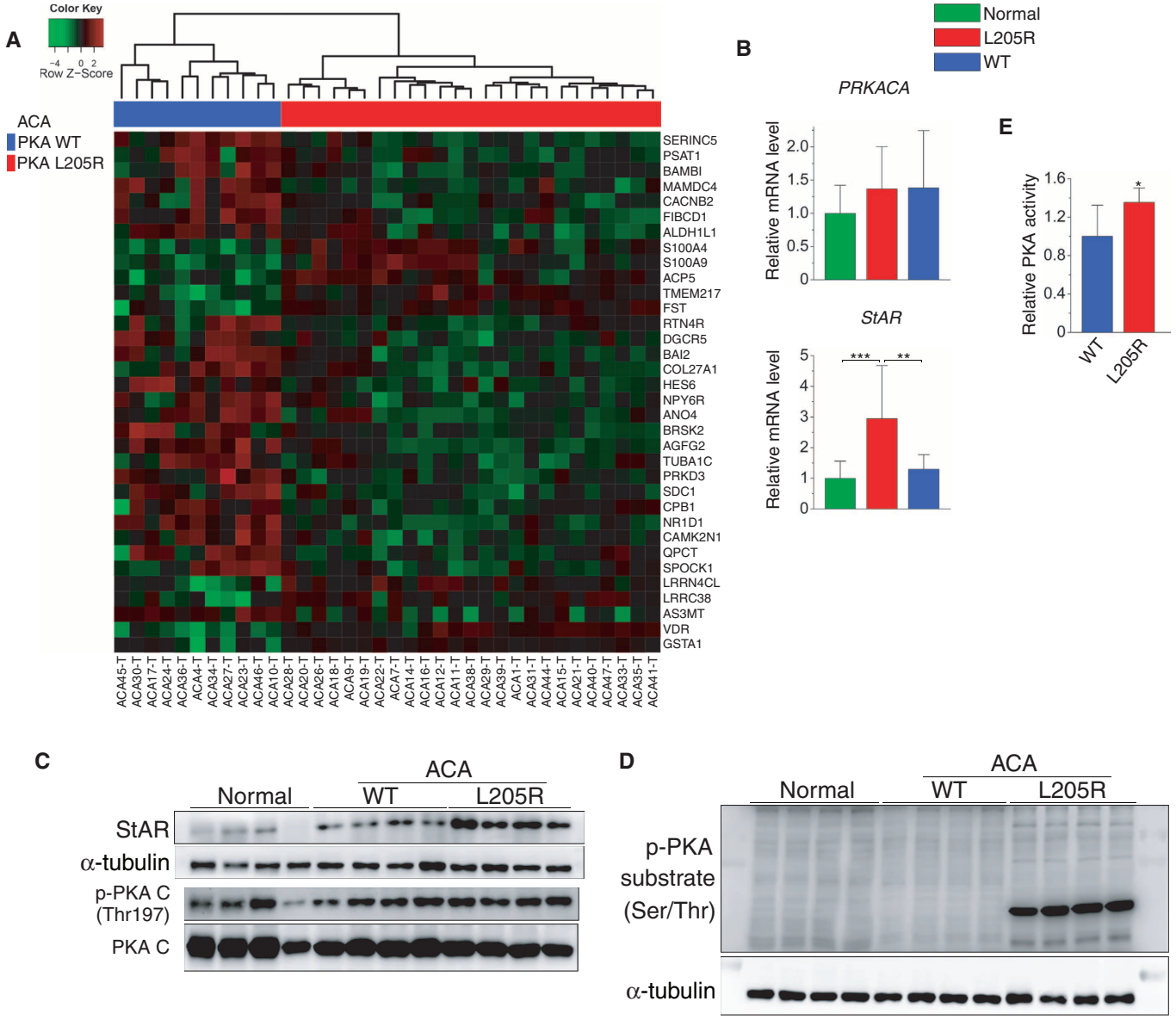


Fig. 3. *PRKACA* L205R mutation activates PKA signaling in ACAs. (**A**) Heat map of significantly differentially expressed genes identified by RNA-seq in ACAs. Columns represent samples; genes are in rows. The mutation status of *PRKACA* for each sample is reported at the top (blue, wild type; red, L205R). Colors of rectangles represent gene expression levels (red, up-regulation; green, down-regulation; black, no change). Samples are arranged by the clustering of gene expression patterns. (**B**) Quantitative reverse transcription PCR analysis of *PRKACA* and *StAR* in normal human adrenocortical tissues (green), L205R mutant ACAs (red), and wild-type ACAs (blue) (normal, $n = 10$; L205R, $n = 20$; wild type, $n = 9$). (**C**) Elevated protein level of StAR in L205R mutant ACAs. (**D**) The phosphorylation of PKA substrates is enhanced in L205R mutant ACAs. (**E**) The PKA activity is significantly higher in L205R mutant ACAs ($n = 5$). Data are means ± SD; *t* test, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

of AIMAH. A recent study revealed recurrent *ARMC5* mutations in AIMAHs (*11*), whereas *DOT1L* and *HDAC9* mutations have not been reported. Functional adrenocortical oncocytoma causing Cushing's syndrome is exceptionally rare. We found *CLASP2* mutated in two of three ADOs, in which we predict the CLIP-associating proteins (CLASPs) to serve a functional role in cell division (*12*). In addition, we also observed *PRKAR1A* and *CTNNB1* mutations in ADOs and *PRUNE2* mutations in both ACA and AIMAH (Fig. 1). Taken together, our findings indicate that multiple genes may be responsible for adrenal Cushing's syndrome.

The L205R mutation in PKA catalytic subunit (PKA C) appears to be a dominant genetic characteristic of ACAs, although PKA C mutation has not been associated with a human disease. Structure analyses of PKA C shows that Lys^{205} is a component of the P+1 loop (Lys^{198}-Lys^{205}), which controls the specific binding between the kinase and its substrates and mediates the communication between catalytic residues and their substrates (*13*–*20*). The highly conserved P+1 loop and activation loop constitute a conserved protein kinase core and contain the activation segment [DFG (Asp-Phe-Gly) to APE (Ala-Pro-Glu) motif] (fig. S6). Previous studies have demonstrated that the $Tyr^{204} \rightarrow$ Ala mutation in the P+1 loop altered the enzyme catalysis and phosphoryl transfer of substrate (*16*–*18*). Lys^{198}, Pro^{202}, and Lys^{205} in the P+1 loop form a hydrophobic binding pocket for accommodating the hydrophobic side chain of the P+1 residue in the substrate consensus (*17*). Crystal structure analyses of the complex between PKA C and regulatory subunit (PKA R) and of PKA C in complex with substrate peptide (Fig. 2) suggested that L205R changed the surface for substrate binding (*15*, *19*, *20*). Together, the structure-based analyses indicate that the L205R mutation might alter the PKA activity by affecting the substrate contact dynamics and catalytic efficiency of PKA C in a manner that contributes to ACA formation.

We analyzed RNA-seq data and obtained 232 genes differentially expressed between *PRKACA* mutant and wild-type ACAs (Fig. 3A and table S9) (*4*). Of note, we did not observe a significantly expressed change of *PRKACA*, which was further supported by quantitative polymerase chain reaction (qPCR) quantification (Fig. 3B). Pathway analysis of these 232 genes using DAVID and WebGestalt (tables S10 and S11) nominated the associated Gene Ontology terms of "biosynthesis and metabolism of steroid and cholesterol" and "response to chemical stimulus." The genes *StAR*, *MC2R*, *GSTA1*, *CXCL2*, and *S100A8/9* involved in these two terms were significantly up-regulated in the *PRKACA* L205R mutant ACAs (fig. S7). These genes may participate in tumor growth, survival, and adrenal steroidogenesis (*21*–*24*). Especially, the mRNA and protein levels of StAR were markedly elevated in the *PRKACA* L205R mutant ACAs relative to normal adrenocortical tissues and *PRKACA* wild-type tumors (Fig. 3, B and C). Furthermore, Western blot analysis showed significantly increased phosphorylation of PKA substrates in ACAs with the L205R mutation (Fig. 3D). Together with the findings that PKA activation can promote *StAR* expression as well as phosphorylate and activate StAR (*25*), our results indicate that the L205R mutation triggers the activation of PKA, which promotes tumorigenesis and steroidogenesis through phosphorylation of substrates.

To further investigate the function and characteristic features of the *PRKACA* L205R mutation, we performed gain-of-function experiments in 293T cells. Overexpression of L205R mutants significantly enhanced the phosphorylation of PKA substrates relative to the wild type (Fig. 4A). The activity of wild-type and mutated PKA C could be suppressed by the selective inhibitor H89 (fig. S8). L205R mutants induced more phosphorylation of CREB (cAMP response element binding protein), which initiates transcription of target genes such as *StAR* and *MC2R* (Fig. 4B) (*25*). In addition, the phosphorylation of PKA C at Thr^{197}, which is critical for catalytic activity of PKA (*26*), was facilitated by the L205R mutation (Fig. 4C). Coimmunoprecipitation assays showed that mutated PKA C was pulled down by PKA R, indicating that the interaction was not interrupted by the L205R mutation (Fig. 4D).

Furthermore, coexpression of L205R mutants and PKA R exhibited more activity than the wild type upon phosphorylation of PKA substrates and CREB in 293T cells, although protein levels of phosphorylated PKA C were comparable (Fig. 4, E and F). PKA activity assays demonstrated that the L205R mutants induced a significant increase in PKA activity, whereas the substrate competitive peptide inhibitor (PKI) could block the catalytic activity of mutated PKA C (Fig. 4, G and H). Taken together, these results indicated that

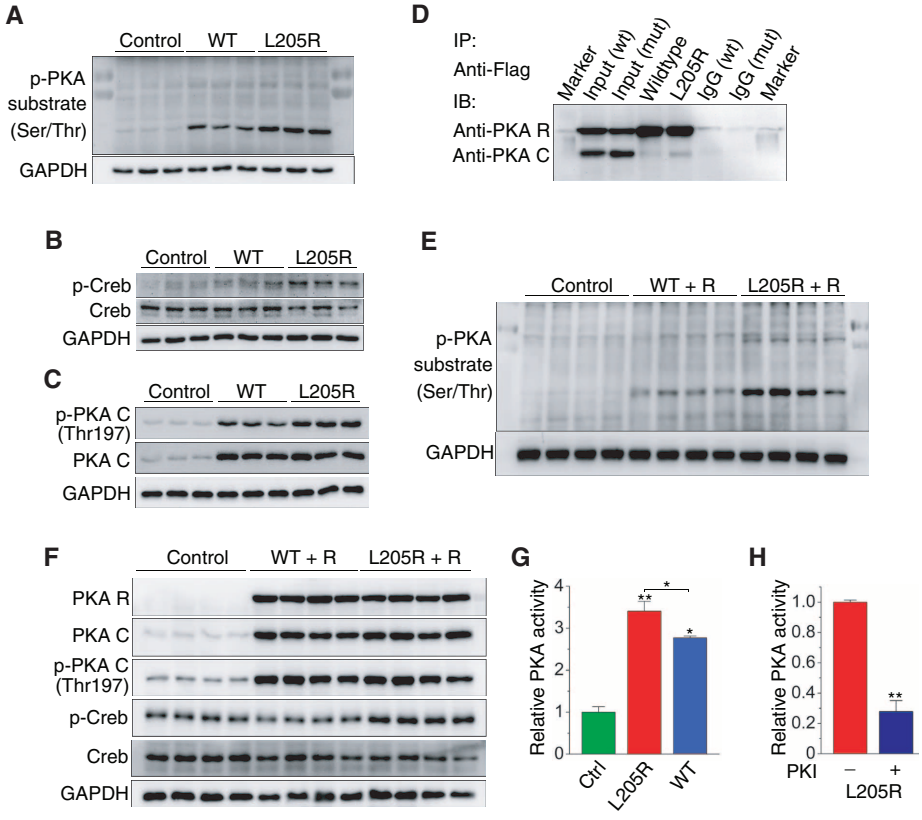


Fig. 4. *PRKACA* L205R mutation promotes PKA activation and substrate phosphorylation. (**A**) Overexpression of L205R-mutated PKA C increases phosphorylation of PKA substrates relative to the wild type in 293T cells. (**B**) Overexpression of L205R-mutated PKA C promotes phosphorylation of CREB in 293T cells. (**C**) The phosphorylation of Thr^{197} is elevated in L205R-mutated PKA C. (**D**) Coimmunoprecipitation of wild type, L205R-mutated PKA C, and Flag-tagged regulatory subunit of PKA (PKA R) in 293T cells. Protein lysates were prepared and immunoprecipitated with antibody to Flag or control immunoglobulin G, respectively. (**E** and **F**) Coexpression of PKA R and PKA C in 293T cells. L205R-mutated PKA C leads to more phosphorylation of PKA substrates (E) and CREB (F). (**G** and **H**) L205R-mutated and wild-type PKA C were overexpressed in 293T cells for PKA activity assay. The activity of L205R-mutated PKA C is significantly higher (G). The peptide inhibitor PKI inhibits the activity of L205R-mutated PKA C (H).

the L205R mutation enhanced the phosphorylation catalysis of PKA.

The identification of activating mutations in oncogenes is particularly valuable in developing targeted therapies for cancer. PKA signaling plays a pivotal role in various physiological and pathological processes, including development, metabolism, and tumorigenesis. Although inactivating mutations in *PRKAR1A* and phosphodiesterase genes are known in adrenal hyperplasia, we revealed the activating *PRKACA* L205R mutation and several potential functional mutated genes in ACTs. Future investigations on mutagenesis and hormones or environmental initiators of the L205R mutation may provide clues for prevention or treatment of Cushing's syndrome.

REFERENCES AND NOTES

1. J. Newell-Price, X. Bertagna, A. B. Grossman, L. K. Nieman, *Lancet* **367**, 1605–1617 (2006).
2. C. A. Stratakis, S. A. Boikos, *Nat. Clin. Pract. Endocrinol. Metab.* **3**, 748–757 (2007).
3. L. S. Kirschner *et al.*, *Nat. Genet.* **26**, 89–92 (2000).
4. See supplementary materials on *Science* Online.
5. A. Berthon, A. Martinez, J. Bertherat, P. Val, *Mol. Cell. Endocrinol.* **351**, 87–95 (2012).
6. S. Gaujoux *et al.*, *Clin. Cancer Res.* **16**, 5133–5141 (2010).
7. L. S. Weinstein *et al.*, *N. Engl. J. Med.* **325**, 1688–1695 (1991).
8. M. C. Fragoso *et al.*, *J. Clin. Endocrinol. Metab.* **88**, 2147–2151 (2003).
9. J. Min, Q. Feng, Z. Li, Y. Zhang, R.-M. Xu, *Cell* **112**, 711 (2003).
10. Y. Okada *et al.*, *Cell* **121**, 167–178 (2005).
11. G. Assié *et al.*, *N. Engl. J. Med.* **369**, 2105–2114 (2013).
12. N. Galjart, *Nat. Rev. Mol. Cell Biol.* **6**, 487–498 (2005).
13. M. D. Uhler *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1300–1304 (1986).
14. B. Nolen, S. Taylor, G. Ghosh, *Mol. Cell* **15**, 661–675 (2004).
15. P. Madhusudan, P. Akamine, N. H. Xuong, S. S. Taylor, *Nat. Struct. Biol.* **9**, 273–277 (2002).
16. M. J. Moore, J. A. Adams, S. S. Taylor, *J. Biol. Chem.* **278**, 10613–10618 (2003).
17. J. Yang, L. F. Ten Eyck, N. H. Xuong, S. S. Taylor, *J. Mol. Biol.* **336**, 473–487 (2004).
18. J. Yang *et al.*, *J. Mol. Biol.* **346**, 191–201 (2005).
19. C. Kim, N. H. Xuong, S. S. Taylor, *Science* **307**, 690–696 (2005).
20. C. Kim, C. Y. Cheng, S. A. Saldanha, S. S. Taylor, *Cell* **130**, 1032–1043 (2007).
21. D. Sarkar *et al.*, *J. Clin. Endocrinol. Metab.* **86**, 1653–1659 (2001).
22. S. Acharyya *et al.*, *Cell* **150**, 165–178 (2012).
23. E. Meimaridou *et al.*, *Endocr. Dev.* **24**, 57–66 (2013).
24. D. Lin *et al.*, *Science* **267**, 1828–1831 (1995).
25. P. R. Manna, M. T. Dyson, D. M. Stocco, *Mol. Hum. Reprod.* **15**, 321–333 (2009).
26. S. S. Taylor, R. Ilouz, P. Zhang, A. P. Kornev, *Nat. Rev. Mol. Cell Biol.* **13**, 646–658 (2012).

ACKNOWLEDGMENTS

Supported by Natural Science Foundation of China grants 30900702, 81130016, and 81270859, and by Guangdong Innovative Research Team Program grant 2009010016. The sequencing data have been deposited in the European Genome-phenome Archive (accession no. EGAS00001000712).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/913/suppl/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S11
References (*27*–*40*)

9 December 2013; accepted 21 March 2014
Published online 3 April 2014;
10.1126/science.1249480

CANCER GENOMICS

Recurrent somatic mutations underlie corticotropin-independent Cushing's syndrome

Yusuke Sato,[1,2]* Shigekatsu Maekawa,[2]* Ryohei Ishii,[3] Masashi Sanada,[1] Teppei Morikawa,[4] Yuichi Shiraishi,[5] Kenichi Yoshida,[1] Yasunobu Nagata,[1] Aiko Sato-Otsubo,[1] Tetsuichi Yoshizato,[1] Hiromichi Suzuki,[1] Yusuke Shiozawa,[1] Keisuke Kataoka,[1] Ayana Kon,[1] Kosuke Aoki,[1] Kenichi Chiba,[5] Hiroko Tanaka,[6] Haruki Kume,[2] Satoru Miyano,[5,6] Masashi Fukayama,[4] Osamu Nureki,[3] Yukio Homma,[2]† Seishi Ogawa[1]†

Cushing's syndrome is caused by excess cortisol production from the adrenocortical gland. In corticotropin-independent Cushing's syndrome, the excess cortisol production is primarily attributed to an adrenocortical adenoma, in which the underlying molecular pathogenesis has been poorly understood. We report a hotspot mutation (L206R) in *PRKACA*, which encodes the catalytic subunit of cyclic adenosine monophosphate (cAMP)–dependent protein kinase (PKA), in more than 50% of cases with adrenocortical adenomas associated with corticotropin-independent Cushing's syndrome. The L206R PRKACA mutant abolished its binding to the regulatory subunit of PKA (PRKAR1A) that inhibits catalytic activity of PRKACA, leading to constitutive, cAMP-independent PKA activation. These results highlight the major role of cAMP-independent activation of cAMP/PKA signaling by somatic mutations in corticotropin-independent Cushing's syndrome, providing insights into the diagnosis and therapeutics of this syndrome.

Cushing's syndrome is a systemic disorder associated with various constitutive symptoms—such as hypertension, impaired glucose tolerance, central obesity, osteoporosis, and depression—that are ascribed to cortisol overproduction (*1*–*3*). Cortisol biosynthesis is primarily regulated by corticotropin secreted from the anterior pituitary gland. Corticotropin acts by binding to the melanocortin-2 receptor, increases cyclic adenosine monophosphate (cAMP) production, and activates cAMP-dependent protein kinase [protein kinase A (PKA)] (*4*). While aberrant corticotropin secretion from pituitary adenoma or other ectopic sites is the leading cause of Cushing's syndrome (corticotropin-dependent Cushing's syndrome), excess cortisol is autonomously produced by adrenocortical tumors in patients with corticotropin-independent Cushing's syndrome (*2*, *5*, *6*). Although mutations in the regulatory subunit type I alpha of PKA (*PRKAR1A*) (*7*, *8*) guanine nucleotide-binding protein subunit alpha (*GNAS*) (*9*), phosphodiesterase-8B (*PDE8B*) (*10*, *11*) and -11A (*PDE11A*) (*12*), and armadillo repeat containing 5 (*ARMC5*) (*13*) are responsible for rare syndromic or hereditary disorders with bilateral adrenocortical hyperplasia, molecular pathogenesis of cortisol-producing adrenocortical adenomas, which account for a large portion of corticotropin-independent Cushing's syndrome (*6*), are less studied.

To investigate genetic lesions in corticotropin-independent Cushing's syndrome, we performed whole-exome sequencing (WES) of eight adrenocortical tumors and matched normal specimens (fig. S1 and table S1). We identified a total of 45 validated nonsynonymous and 59 putative synonymous somatic mutations (tables S2 and S3) (see the supplementary materials). Remarkably, the gene encoding the catalytic subunit (C subunit) of PKA (*PRKACA*) was recurrently mutated in four out of the eight cases, resulting in an identical c.T617G mutation predicted to cause a conversion of leucine to arginine at amino acid position 206 (p.L206R). Together with a *GNAS* mutation (p.R201C), five of the eight cases had somatic mutations in genes involved in the cAMP/PKA signaling pathway. No allelic imbalances were observed at both gene loci in single-nucleotide polymorphism (SNP) array analysis (fig. S2), indicating that these mutations were heterozygous, which was consistent with the observation that, in exome sequencing, variant allele frequencies (VAF) of *PRKACA*/*GNAS* mutations were comparable to those of other somatic mutations (fig. S3). Mutations were not detected in any other known causative genes, including *PRKAR1A*, *PDE11A*, *PDE8B*, and *ARMC5*.

[1]Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [2]Department of Urology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [3]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, Tokyo, Japan. [4]Department of Pathology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [5]Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [6]Laboratory of Sequence Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.
*These authors contributed equally to this work. **†Corresponding author. E-mail: sogawa-tky@umin.ac.jp (S.O.); homma-uro@umin.ac.jp (Y.H.)**

We performed follow-up sequencing of *PRKACA* and *GNAS* as well as previously reported genes (*PRKAR1A*, *PDE11A*, *PDE8B*, and *ARMC5*) in an additional 57 cases (see the supplementary materials). The L206R mutation in *PRKACA* was found in 30 out of the 57 follow-up cases, of which 24 cases were confirmed as being somatic, whereas *GNAS* mutations were found in 10 cases, with somatic origin being confirmed in six cases (table S1, Fig. 1, and fig. S4). No mutations were found in the previously reported genes. The eight samples double-negative for *PRKACA* and *GNAS* mutations were tested for mutations in seven additional genes that were mutated and expressed in three double-negative exome cases, but no more recurrent mutations were identified. Combined with the four *PRKACA* and one *GNAS* mutations in the discovery cases, *PRKACA* and *GNAS* were mutated in 34 (52.3%) and 11 (16.9%) out of the 65 cases with corticotropin-independent Cushing's syndrome, respectively, where both mutations were completely mutually exclusive (Fisher's exact test, $P = 9.46 \times 10^{-5}$) (Fig. 1). The somatic origin was confirmed for 28 out of 28 *PRKACA* and 6 out of 6 *GNAS* mutations thus far tested. In addition, VAFs of *PRKACA* and *GNAS* mutations in deep sequencing were distributed between 0.08 and 0.35, whereas those of most heterozygous SNPs (83%) were between 0.4 and 0.6 (fig. S5), indicating that most of these mutations were somatic in origin.

Patients with mutated *PRKACA* showed significantly higher cortisol levels on the 1-mg dexamethasone suppression test (DST) compared with wild-type *PRKACA* and *GNAS* (*t* test, $P = 2.60 \times 10^{-3}$) (Fig. 2A). *PRKACA*-mutated adenomas had a significantly smaller tumor diameter than those with no known mutations (*t* test, $P = 4.93 \times 10^{-5}$) (Fig. 2B), suggesting that *PRKACA*-mutated adenomas may have higher cortisol production. *GNAS*-mutated adenomas also showed higher cortisol levels on 1-mg DST and have a smaller tumor size. Seventy-six percent of the patients with clinical Cushing's syndrome had a mutation of either gene, whereas only two of nine patients with subclinical Cushing's syndrome had mutated *PRKACA* or *GNAS* genes (Fisher's exact test, $P = 2.87 \times 10^{-3}$) (Fig. 2C and table S4), indicating that these mutations were enriched for clinical Cushing's syndrome.

PRKACA is the catalytic (C) subunit of the tetrameric PKA holoenzyme, which binds tissue-specific dimeric regulatory (R) subunits (*14*–*16*). In the native state, the C subunit is kept inactivated through the binding of the R subunit, which masks the catalytic site of the C subunit (Fig. 3A). However, when intracellular cAMP is up-regulated by external stimuli, each R subunit binds two cAMP molecules, which causes a conformational change in the R subunit to promote dissociation of the C subunit from the PKA complex (*14*–*16*), allowing for the translocation of the dissociated, and thereby activated, free C subunit to the nucleus and phosphorylation of its target substrates therein (*17*, *18*). The highly conserved L206 residue of PRKACA, which resides within the P+1 loop of the C subunit, lies on the surface of the large lobe of PRKACA and is thought to be essential for the catalytic activity of the kinase (Fig. 3A, and fig. S6) (*14*–*16*). In the absence of cAMP, the inhibitory region of the R subunit docks to the active site cleft of the C subunit, including the P+1 loop. The L206 residue is located at the interface between the C subunit and the inhibitory region in the R subunit to form a hydrophobic interaction with the I99 residue of the R subunit (Fig. 3B). Thus, the substitution from the small hydrophobic leucine to a large hydrophobic arginine is predicted to cause steric hindrance and abolish the binding of the C and R subunits (Fig. 3C), resulting in constitutive, cAMP-independent activation of PKA.

In fact, when the PKA complex was reconstituted in vitro using purified proteins (fig. S7), PRKAR1A binds the wild-type C subunit and suppresses its PKA activity in the absence of cAMP, and the suppression is recovered in the presence of cAMP with substantially reduced interaction with the wild-type C subunit (Fig. 4, A and B). In contrast, the R subunit can no longer bind the L206R PRKACA mutant with constitutive PKA activation, regardless of the presence or absence of cAMP (Fig. 4, A and B). The loss of binding to the R subunit and consequent cAMP-independent PKA activation for the mutant PRKACA was also demonstrated in vivo. When expressed in human embryonic kidney 293T (HEK293T) cells (fig. S8), wild-type PRKACA, but not the L206R PRKACA mutant, coimmunoprecipitated with PRKAR1A (Fig. 4C). Both mock- and wild-type *PRKACA*-transduced cells showed increased PKA activity accompanied by an elevated phosphorylation of cAMP response element–binding protein (pCREB), one of the major downstream targets of PKA activation (Fig. 4, D and E, and fig. S9), on cAMP induction by forskolin treatment (see the supplementary materials), whereas mutant *PRKACA*-transduced cells demonstrated a higher basal level of PKA activity and CREB phosphorylation, regardless of forskolin treatment (Fig. 4, D and E, and fig. S9). To confirm the cAMP-independent activation of the mutant PRKACA, we examined the effect of two PRKACA inhibitors (H89 and KT5720) and a competitive inhibitor of cAMP binding for the R subunit (Rp-cAMPS) on the activity of the L206R mutant (see the supplementary materials). PKA activation in the L206R mutant–transduced cells in the absence of forskolin was suppressed by H89 and KT5720 (*t* test, $P = 1.60 \times 10^{-3}$ and $P = 1.86 \times 10^{-3}$, respectively) but not in Rp-cAMPS–treated cells, supporting further that the consequence of the L206R mutation is constitutive, cAMP-independent activation of PKA (Fig. 4F and fig. S10).

Finally, as predicted from low stability of free C subunits, primary *PRKACA*-mutated tumors showed significantly lower PRKACA protein expression compared with unmutated tumors (*t* test, $P = 5.70 \times 10^{-3}$) and normal adrenocortical tissues (*t* test, $P = 2.09 \times 10^{-2}$) (Fig. 4G and fig. S11), although no significant difference was observed for downstream signaling, such as pCREB or steroidogenic acute regulatory protein (StAR). The reduced intracellular expression of the mutant PRKACA protein was also observed for

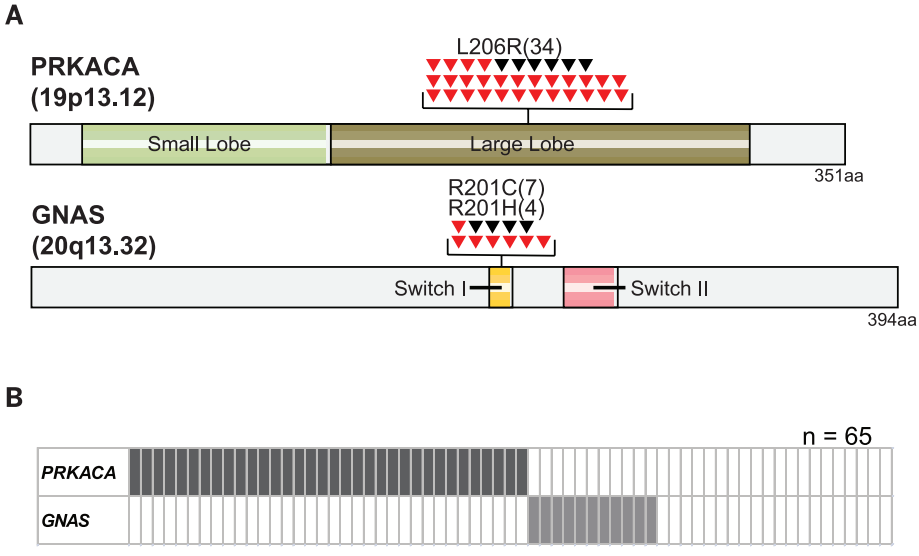


Fig. 1. Recurrent mutations in *PRKACA* and *GNAS*. (**A**) Mutations in *PRKACA* (top) and *GNAS* (bottom) identified in 65 patients with corticotropin-independent Cushing's syndrome (arrowheads). Confirmed somatic mutations are indicated in red. (**B**) Mutually exclusive distribution of *PRKACA* and *GNAS* mutations.

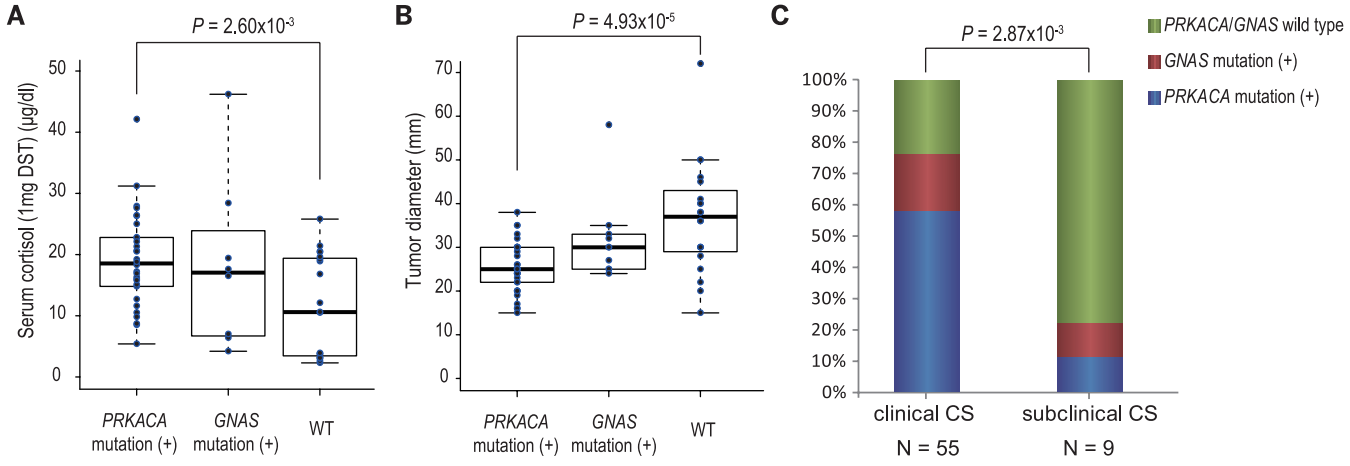
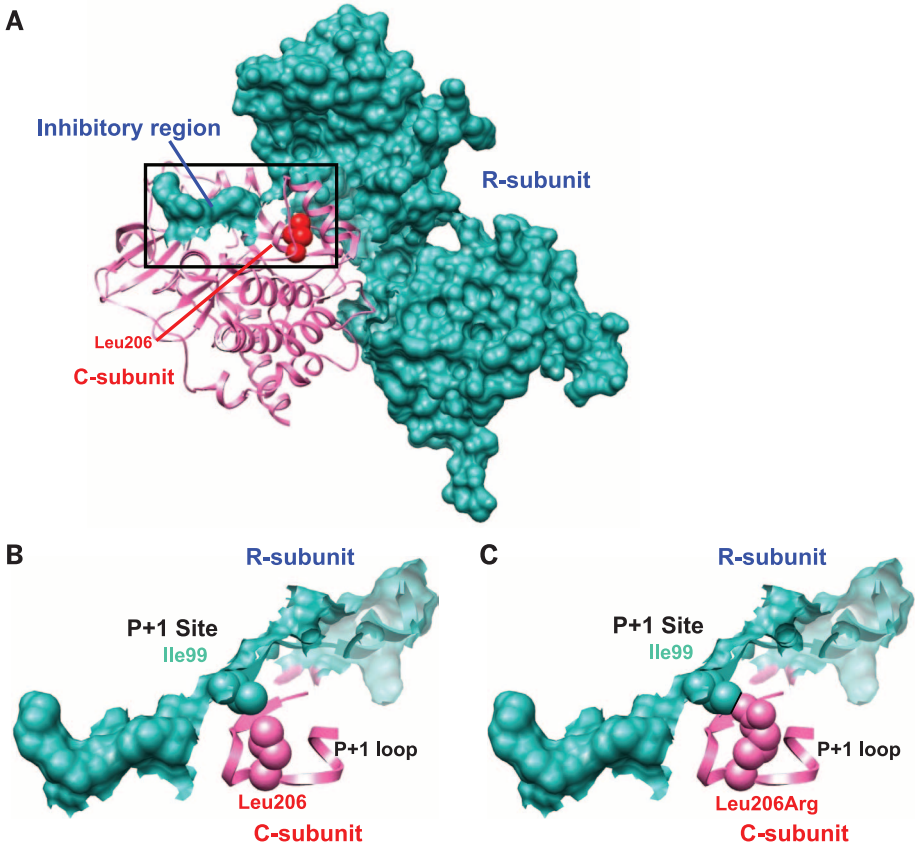


Fig. 2. Relationship between mutation status and clinical features. (**A** and **B**) Serum cortisol level of 1-mg dexamethasone (A) and the diameter of adrenocortical adenoma (B) according to the mutation status of *PRKACA*/*GNAS* genes. (**C**) Clinical and subclinical Cushing's syndrome as attributed to mutated *PRKACA*, *GNAS*, or other, currently undetected, cause.

Fig. 3. Effect of L206R mutation on three-dimensional structures of PKA. (**A**) Three-dimensional structure of the PKA complex, composed of C (PRKACA) (pink) and R (PRKAR1A) (cyan) subunits, is depicted using the University of California–San Francisco Chimera program, based on the Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB ID: 2QCS). L206 is shown in red. (**B** and **C**) A predicted effect of the L206R mutation within the P+1 loop on the interaction with the R subunit (C) in comparison with wild-type PRKACA (B).



exogenously introduced PRKACA in different cell types (Fig. 4H and fig. S12). However, compared with the wild-type protein, mutant PRKACA was more enriched in the nuclear than in the cytoplasmic fraction (Fig. 4H).

In conclusion, frequent somatic mutations in *PRKACA* and *GNAS* genes underlie corticotropin-independent Cushing's syndrome (fig. S13). Strikingly, in accordance with a recent report (*19*) *PRKACA* mutations were found in more than 50% of the current cohort. All the mutations were the identical L206R substitution, which prevents binding to the inhibitory R subunits and results in constitutive, cAMP-independent activation of PKA. *GNAS* mutations were also detected in a substantial fraction of the present cohort (11/65 or 16.9%), in which all were confined to the R201 residue. Steroid hormone biosynthesis in steroidogenic cells is primarily regulated through activation of the cAMP/PKA signaling pathway (*20*). *PRKACA* and *GNAS* mutations in adrenocortical cells affecting this pathway are thus a likely factor responsible for the excessive production of cortisol and together account for as many as 70% of the current cohort of the patients. In contrast, the remaining 30% of the patients with no mutations in either gene tended to show lower cortisol levels on 1-mg DST and had a larger tumor size compared with *PRKACA*-mutated patients. In addition, the identified mutations were confined almost exclusively to patients with clinical Cushing's syndrome, suggesting that double-negative cases have distinct pathogenesis, with driver mutations still to be identified. It should be warranted to identify the genetic basis of double-negative cases in future studies.

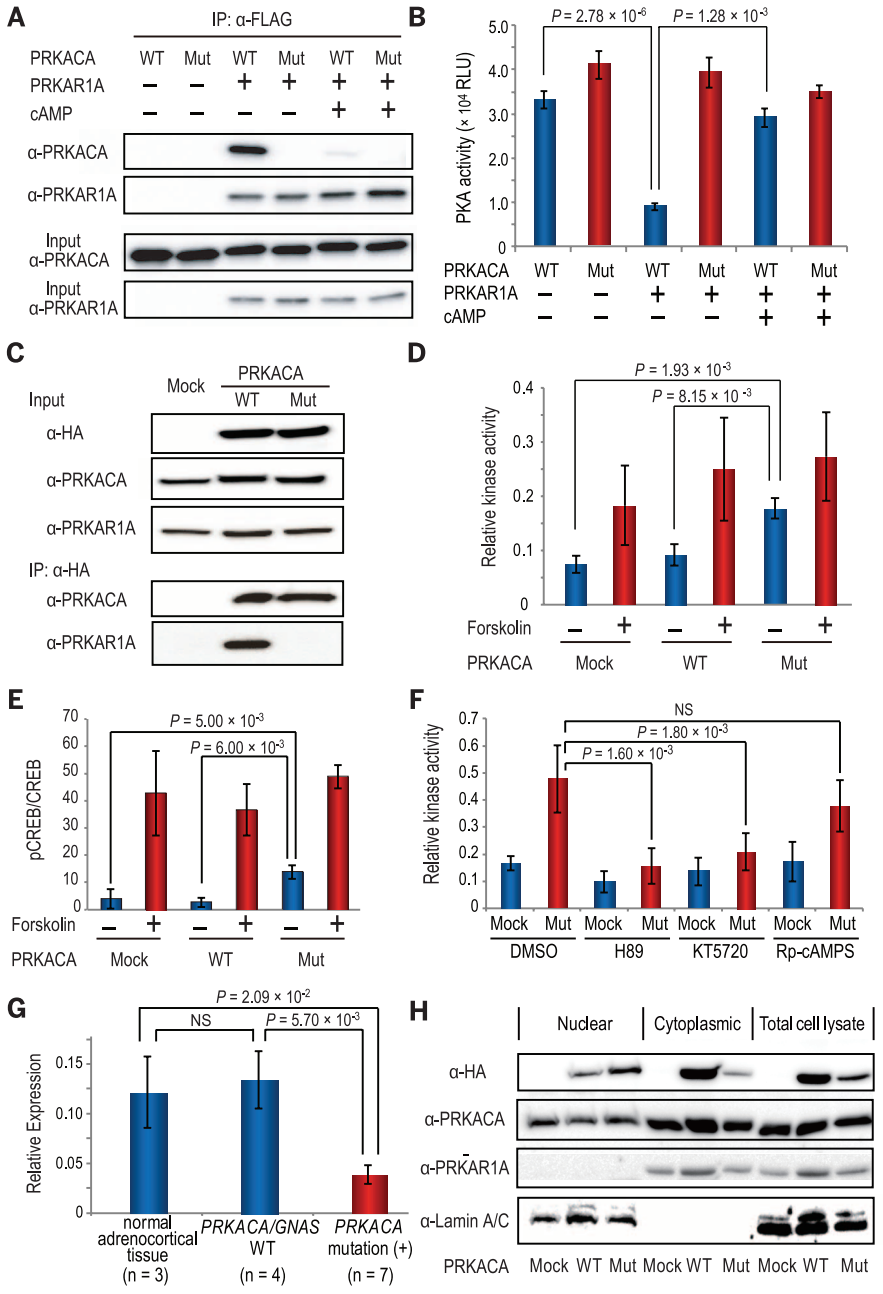


Fig. 4. Functional characterization of PRKACA mutant. (**A**) Immunoblot analysis of antibody to FLAG immunoprecipitates of the PKA reconstructed in vitro with large-scale–purified proteins, in which PRKACA and PRKAR1A subunits are tagged with hemagglutinin (HA) and FLAG, respectively. (**B**) In vitro PKA activities of wild-type and mutant PRKACA in the presence or absence of PRKAR1A and cAMP with purified proteins. (**C**) Immunoblot analysis of the total cell lysates (input) and antibody to HA immunoprecipitates from HEK293T cells stably transduced with either mock, wild-type, or mutant PRKACA tagged with HA with indicated antibodies. (**D**) PKA activities in HEK293T cells stably transduced with either mock, wild-type, or mutant PRKACA in the presence or absence of forskolin stimulation. (**E**) Relative expression of pCREB to total CREB in HEK293T cells stably transduced with either mock, wild-type, or mutant PRKACA as determined by densitometry of the immunoblots. (**F**) The effect of inhibition of PKA (H89 and KT5720) and cAMP (Rp-cAMPS) in HEK293T cells transduced with mock or mutant PRKACA. (**G**) PRKACA expression standardized for mRNA expression in *PRKACA*-mutated and unmutated adenomas and matched normal adrenocortical tissues. (**H**) Western blot analysis of indicated fractions of cell lysates from wild-type and mutant PRKACA-transduced HEK293T cells using indicated antibodies. Standard errors and significant differences are indicated [(C) to (G)].

REFERENCES AND NOTES

1. L. K. Nieman *et al.*, *J. Clin. Endocrinol. Metab.* **93**, 1526–1540 (2008).
2. B. A. Hatipoglu, *J. Surg. Oncol.* **106**, 565–571 (2012).
3. J. Newell-Price, X. Bertagna, A. B. Grossman, L. K. Nieman, *Lancet* **367**, 1605–1617 (2006).
4. K. G. Mountjoy, L. S. Robbins, M. T. Mortrud, R. D. Cone, *Science* **257**, 1248–1251 (1992).
5. D. N. Orth, *N. Engl. J. Med.* **332**, 791–803 (1995).
6. C. A. Stratakis, *Endocr. Dev.* **13**, 117–132 (2008).
7. L. S. Kirschner *et al.*, *Nat. Genet.* **26**, 89–92 (2000).
8. L. S. Kirschner *et al.*, *Hum. Mol. Genet.* **9**, 3037–3046 (2000).
9. L. S. Weinstein *et al.*, *N. Engl. J. Med.* **325**, 1688–1695 (1991).
10. A. Horvath *et al.*, *Nat. Genet.* **38**, 794–800 (2006).
11. A. Horvath *et al.*, *Cancer Res.* **66**, 11571–11575 (2006).
12. A. Horvath, V. Mericq, C. A. Stratakis, *N. Engl. J. Med.* **358**, 750–752 (2008).
13. G. Assié *et al.*, *N. Engl. J. Med.* **369**, 2105–2114 (2013).
14. M. J. Moore, J. A. Adams, S. S. Taylor, *J. Biol. Chem.* **278**, 10613–10618 (2003).
15. C. Kim, N. H. Xuong, S. S. Taylor, *Science* **307**, 690–696 (2005).
16. C. Kim, C. Y. Cheng, S. A. Saldanha, S. S. Taylor, *Cell* **130**, 1032–1043 (2007).
17. M. R. Montminy, L. M. Bilezikjian, *Nature* **328**, 175–178 (1987).
18. A. T. Harootunian *et al.*, *Mol. Biol. Cell* **4**, 993–1002 (1993).
19. F. Beuschlein *et al.*, *N. Engl. J. Med.* **370**, 1019–1028 (2014).
20. A. M. Lefrancois-Martinez *et al.*, *J. Biol. Chem.* **286**, 32976–32985 (2011).

ACKNOWLEDGMENTS

We thank Y. Mori, M. Nakamura, N. Mizota, S. Ichimura, and M. Yamakawa for their technical assistance. The retroviral vector, pGCDNsamIRES-EGFP, was kindly provided by M. Onodera (National Research Institute for Child Health and Development). This work was supported by the Japan Society for the Promotion of Science (JSPS) through Grants-in-Aid for Scientific Research (KAKENHI) grant number 22134006 and the Funding Program for World-Leading Innovative Research and Development on Science and Technology (FIRST Program). Data for exome sequencing, as well as SNP array analysis, are found in the European Genome-phenome Archive (EGA) under accession EGAS00001000661. Author contributions: Y.Sa., S.Ma., K.Y., Y.N., T.Y., H.S., and A.K. performed library preparation and DNA sequencing. Y.Shira., Y.Shio., K.C., H.T., and S.Mi. were committed to bioinformatics analyses of resequencing data. Y.Sa. and A.S.-O. performed SNP array analysis. Y.Sa., S.Ma., M.S., R.I., K.K., and O.N. performed the functional analyses of PRKACA mutant. T.M., H.K., M.F., and Y.H. provided specimens and were also involved in planning the project. Y.Sa., S.Ma., M.S., R.I., and S.O. generated figures and tables and wrote the manuscript. S.O. led the entire project. All authors participated in the discussion and interpretation of data and results. Y.Sa., S.Ma., M.S., Y.H., and S.O. are inventors on a patent applied for by Kyoto University that covers the inspection method for Cushing's syndrome and biomarker and theraputic agent thereof (2014-37189).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/917/suppl/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 to S8
References (*21*–*26*)

17 February 2014; accepted 11 April 2014
10.1126/science.1252328

CANCER IMMUNOLOGY

The cellular and molecular origin of tumor-associated macrophages

Ruth A. Franklin,[1,2] Will Liao,[3] Abira Sarkar,[1] Myoungjoo V. Kim,[1,2] Michael R. Bivona,[1] Kang Liu,[4] Eric G. Pamer,[1] Ming O. Li[1]*

Long recognized as an evolutionarily ancient cell type involved in tissue homeostasis and immune defense against pathogens, macrophages are being rediscovered as regulators of several diseases, including cancer. Here we show that in mice, mammary tumor growth induces the accumulation of tumor-associated macrophages (TAMs) that are phenotypically and functionally distinct from mammary tissue macrophages (MTMs). TAMs express the adhesion molecule Vcam1 and proliferate upon their differentiation from inflammatory monocytes, but do not exhibit an "alternatively activated" phenotype. TAM terminal differentiation depends on the transcriptional regulator of Notch signaling, RBPJ; and TAM, but not MTM, depletion restores tumor-infiltrating cytotoxic T cell responses and suppresses tumor growth. These findings reveal the ontogeny of TAMs and a discrete tumor-elicited inflammatory response, which may provide new opportunities for cancer immunotherapy.

Macrophages are tissue-resident innate immune cells important in homeostasis and host defense against pathogens (*1*). These functionally diverse phagocytes differentiate from yolk sac–derived embryonic precursors and locally self-renew both during steady state (*2*–*4*) and helminth infection (*5*). Additionally, bone marrow–derived monocytes give rise to macrophages in the intestine and the dermis (*6*, *7*), as well as during acute infection and inflammation (*8*). However, the precise ontogeny and function of macrophages in chronic disorders, such as cancer, are incompletely understood (*9*).

To investigate myeloid cells during cancer progression, we utilized the MMTV-PyMT (PyMT) mammary tumor model (*10*). Myeloid cells made up more than 50% of $CD45^+$ tumor-infiltrating leukocytes and consisted of three major populations (I, II, and III), distinguishable by morphology and cell surface expression of major histocompatibility complex class II (MHCII), CD11b, Ly6C, Ly6G, CD11c, CD115, and F4/80 (fig. S1). Populations II and III phenotypically resembled $Ly6C^+$ inflammatory monocytes and neutrophils, respectively, while population I expressed classical dendritic cell (DC) markers MHCII and CD11c and the macrophage marker F4/80. Owing to the ambiguity of characterizing cell populations with surface markers (*11*, *12*), we sought to define these cells on the basis of transcriptional phenotype (*13*). Using principal component analysis of DC and macrophage populations from the ImmGen Project (*14*, *15*), we defined "population I" cells as tumor-associated macrophages (TAMs) because they clustered with macrophage subsets (Fig. 1A). A support vector machine learning algorithm corroborated this classification (fig. S2). Moreover, cells of population I did not express the DC lineage–specific transcription factor *Zbtb46* or DC markers c-Kit, CD26, BTLA, and Flt3, but expressed the macrophage transcription factor *Mafb* and macrophage markers CD64 and MerTK (*14*, *15*) (Fig. 1, B and C). Furthermore, Flt3L-deficient PyMT mice, which lack cells of the classical DC lineage (*16*), showed no defect in population I, confirming a pre-DC–independent origin of TAMs (fig. S3).

Macrophages populate mammary tissues during steady state and are required for mammary gland development (*17*). Upon tumor growth, we observed a decrease in the proportion of $MHCII^{hi}CD11b^{hi}$ cells found in untransformed wild-type (WT) mammary glands and an increase in TAMs (Fig. 1D). We defined $MHCII^{hi}CD11b^{hi}$ cells as "mammary tissue macrophages" or "MTMs" because they also phenotypically resembled macrophages (fig. S4). TAM expansion was associated with the growth of individual tumors (Fig. 1, E and F), demonstrating that $CD11b^{lo}$ TAMs, but not $CD11b^{hi}$ MTMs, are bona fide tumor-associated macrophages that accumulate with increased tumor burden.

Tissue-resident macrophage expansion or differentiation of macrophages from blood-borne precursors could account for TAM accumulation. To distinguish between these mechanisms, we connected congenically marked PyMT mice using parabiosis (fig. S5A). We observed $Ly6C^+$ inflammatory monocytes, MTMs, and TAMs from both parabionts in developing tumors (fig. S5, B and C), demonstrating that TAMs and MTMs required input from the circulation. The chimerism of inflammatory monocytes and T cells (fig. S4, C and D) was in accordance with published studies (*2*, *18*, *19*). This was in contrast to red pulp macrophages, which are maintained independently from monocytes (*2*) and consequently exhibited minimal chimerism (fig. S5C).

Circulating monocytes are critical progenitors for macrophages (*20*). To determine whether $Ly6C^+CCR2^+$ inflammatory monocytes contributed to TAMs and MTMs, we crossed PyMT mice to $Ccr2^{-/-}$ mice, which exhibit reduced numbers of circulating inflammatory monocytes due to impaired bone marrow egress (*21*). At 16 weeks, MTMs were significantly reduced in $Ccr2^{-/-}$ PyMT mice (Fig. 2A and fig. S6), implying that MTMs are constitutively repopulated by inflammatory monocytes. Owing to the loss of both the monocyte and MTM populations (Fig. 2A and fig. S6), a concomitant increase in TAMs would be expected if their maintenance was independent of $CCR2^+$ monocytes. However, the TAM percentage in $Ccr2^{-/-}$ PyMT mice was not significantly different compared to controls (Fig. 2A and fig. S6). Similar trends were observed in 20-week-old mice (Fig. 2A and fig. S6), suggesting that inflammatory monocytes contribute to MTMs, and to a lesser extent, TAMs.

To determine whether inflammatory monocytes were required for TAM maintenance, we generated $CCR2^{DTR}$ PyMT mice expressing diphtheria toxin receptor (DTR) under control of the *Ccr2* locus (*22*). DT treatment resulted in 96% depletion of tumor-associated monocytes (Fig. 2B and fig. S7), compared to 80% depletion in $Ccr2^{-/-}$ mice (Fig. 2A and fig. S6). With this more potent depletion strategy, both MTM and TAM numbers were significantly reduced (Fig. 2B and fig. S7), suggesting that TAMs are derived from $CCR2^+$ monocytic precursors, but require less input from the blood compared to MTMs. We considered that a higher proliferative capacity of TAMs compared to MTMs might account for their differing precursor requirement. Indeed, TAMs expressed higher levels of Ki67 staining and EdU incorporation relative to MTMs (Fig. 2, C and D).

To investigate whether monocytes could differentiate into TAMs in vivo, we transferred $CCR2^+$ bone marrow cells isolated from $CCR2^{GFP}$ reporter mice (*23*) into congenically marked $CCR2^{DTR}$ PyMT mice depleted of endogenous monocytes. At days 5, 7, and 11 after transfer, we observed transferred cells in developing tumors (fig. S8A). Five and 7 days after transfer, we detected up-regulation of F4/80, CD11c, and MHCII and down-regulation of Ly6C and CD11b on the transferred cells (Fig. 2E and fig. S8B). Additionally, transferred cells expanded and were $Ki67^+$ 11 days after transfer (Fig. 2, F and G). Transfer of hematopoietic stem cell–depleted $CCR2^+$ bone marrow monocytes ($CCR2^+Flt3^-c\text{-}Kit^-$) provided similar results (fig. S9). Collectively, these observations demonstrate that tumor growth induces the differentiation of $CCR2^+$ monocytes into TAMs.

We performed gene-expression profiling to further distinguish TAMs from MTMs. As expected, the integrin CD11b (*Itgam*) was expressed at lower levels in TAMs than in MTMs (Fig. 3A). However, several other integrins and the integrin receptor Vcam1 were up-regulated in TAMs (Fig. 3A). "M2" or alternatively activated macrophages

[1]Immunology Program, Memorial Sloan Kettering Cancer Center (MSKCC), New York, NY 10065, USA. [2]Graduate Program in Immunology and Microbial Pathogenesis, Weill Cornell Graduate School of Medical Sciences, Cornell University, New York, NY 10065, USA. [3]New York Genome Center, New York, NY 10022, USA. [4]Department of Microbiology and Immunology, Columbia University, New York, NY 10032, USA.
*Corresponding author. E-mail: lim@mskcc.org

(AAMs) have been proposed to be associated with tumor progression (*24*). Surprisingly, we found that the TAM population did not express AAM markers such as Ym1, Fizz1, and Mrc1; instead, MTMs more closely resembled AAMs (Fig. 3A and fig. S10). In line with the expression data, Vcam1 and Mrc1 (CD206) proteins were detected in TAMs and MTMs, respectively (Fig. 3B). Combining these markers with our monocyte transfer system, we addressed whether TAMs differentiated from MTMs. The lack of Mrc1 expression on transferred monocytes at all examined time points suggested that TAM differentiation from monocytes was a distinct pathway, rather than MTM conversion (Fig. 3C). Additionally, we detected Vcam1 up-regulation on TAMs as a late differentiation event (Fig. 3C). Altogether, we identified sequential phenotypic changes in monocytes during TAM differentiation (fig. S11).

The finding that TAMs did not resemble AAMs was unexpected, because the type 2 cytokine interleukin-4 (IL-4) produced by T cells and/or tumor cells has been implicated in TAM polarization (*25*, *26*). Additionally, in other M2-polarizing environments, IL-4 is crucial for the expansion of tissue-resident macrophage populations (*5*). However, we found that *Il4*$^{-/-}$ PyMT mice had normal proportions of CD11b^{lo}Vcam1^{+} TAMs (fig. S12, A to C). Furthermore, TAM differentiation was intact in the absence of lymphocytes (fig. S12, D to F). These observations suggest that TAMs are not AAMs, and their differentiation is not secondary to tumor-elicited adaptive immune responses.

While investigating the mechanisms of TAM differentiation, we observed that TAMs display a gene expression signature (*27*) associated with the Notch signaling pathway (fig. S13). Notch signaling is a conserved developmental pathway instrumental in hematopoietic cell fate specification (*28*). In DCs, canonical Notch signaling mediated by the key transcriptional regulator RBPJ controls lineage commitment and terminal differentiation (*27*, *29*, *30*). To explore whether Notch signaling played a role in TAM differentiation, we used CD11c^{cre} mice that efficiently deleted floxed DNA sequences to a greater extent in TAMs than MTMs, but not in monocytes or neutrophils (fig. S14). CD11c^{cre}*Rbpj*$^{fl/fl}$ PyMT mice exhibited a selective loss of MHCII^{hi}CD11b^{lo} TAMs (Fig. 4A). However, a MHCII^{hi}CD11b^{hi} population still remained (Fig. 4A and fig. S15, A and B). Transcriptional profiling comparing this population to WT TAMs confirmed a loss of the Notch-dependent program in RBPJ-deficient cells (fig. S15C). MHCII^{hi}CD11b^{hi} macrophages in CD11c^{cre}*Rbpj*$^{fl/fl}$ PyMT mice did not express the TAM marker Vcam1 or the MTM marker Mrc1 in most cells (Fig. 4B), suggesting that the majority are TAM intermediates. Moreover, we observed increased Ly6C^{-}MHCII$^{-/lo}$ TAM precursors (fig. S11) in CD11c^{cre}*Rbpj*$^{fl/fl}$ PyMT mice compared to WT PyMT mice (fig. S15, D and E). These data reveal that in the absence of RBPJ, inflammatory monocytes are unable to terminally differentiate into TAMs.

The Mrc1^{+} cells found within the CD11b^{hi} population in CD11c^{cre}*Rbpj*$^{fl/fl}$ PyMT mice (Fig. 4B) indicated that MTM differentiation was not compromised. To address the specificity of this RBPJ-dependent pathway during tumorigenesis, we analyzed non-PyMT mice. As expected, MHCII^{hi}CD11b^{hi} MTMs from WT and CD11c^{cre}*Rbpj*$^{fl/fl}$ mammary glands expressed Mrc1 (fig. S16). MHCII^{hi}CD11b^{lo} myeloid cells were present in mammary glands from WT mice (Fig. 1D). However, these cells did not express Vcam1 (fig. S16A), and their differentiation was not affected in CD11c^{cre}*Rbpj*$^{fl/fl}$ mice (fig. S16B), suggesting that they are distinct from MHCII^{hi}CD11b^{lo} TAMs.

Associated with the TAM differentiation defect, CD11c^{cre}*Rbpj*$^{fl/fl}$ PyMT mice had reduced tumor burden (Fig. 4C). In *Ccr2*$^{-/-}$ PyMT mice,

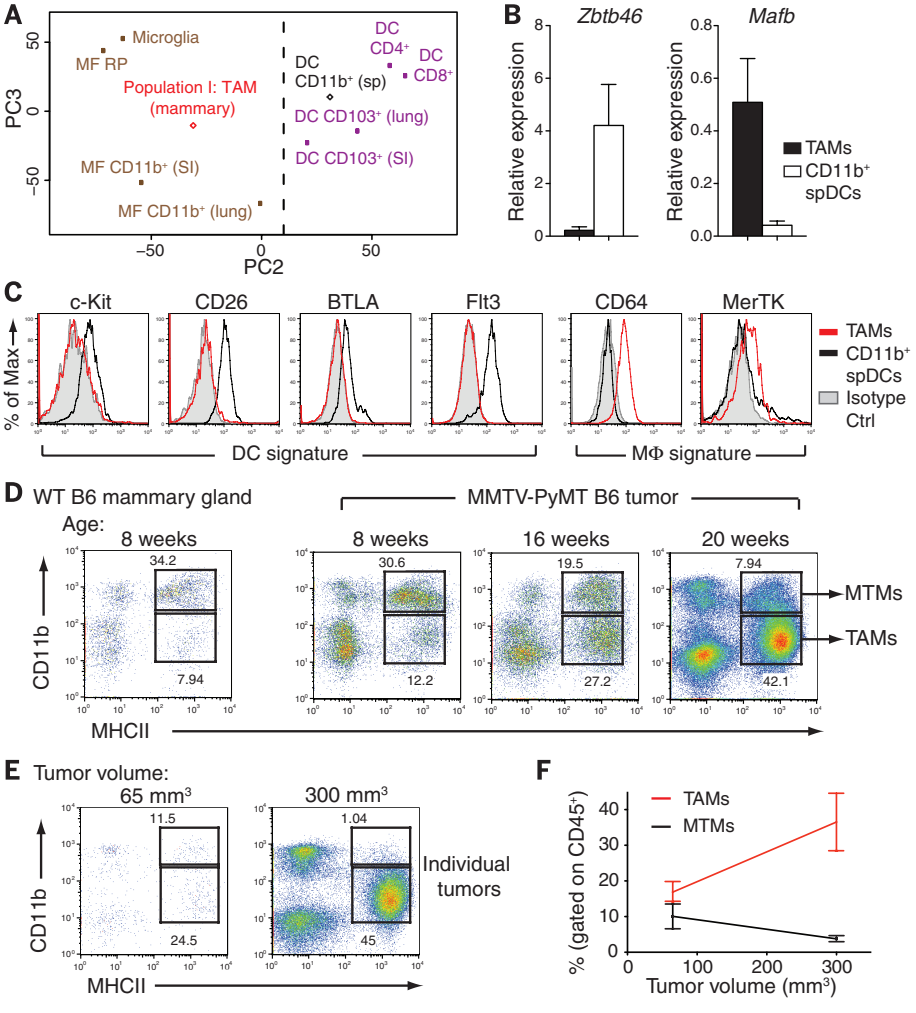


Fig. 1. Macrophages constitute the dominant myeloid cell population in mammary tumors. (**A**) Principal component analysis of population I: TAM (mammary) and CD11b^{+} splenic DC (sp) gene expression compared to populations collected by the Immunological Genome Project (GSE15907). Expression data are pooled from two replicate microarray experiments. (**B**) Expression of *Zbtb46* and *Mafb* mRNA in sorted TAMs and CD11b^{+} splenic DCs (spDCs) relative to *Actb* as determined by quantitative polymerase chain reaction (n = 3). (**C**) Flow cytometric analysis of DC and macrophage (mΦ) signature surface markers expressed on TAMs and CD11b^{+} spDCs. Data are representative of two independent experiments. (**D**) Flow cytometric analysis of myeloid cell populations found in wild-type (WT) mammary glands and pooled tumors from 8-, 16-, and 20-week-old PyMT mice. Data are representative of three independent experiments. (**E**) Flow cytometry of TAM and MTM populations in individual tumors from the same mouse. (**F**) Pooled data of individual tumors as in (E) from multiple mice (n = 3). Data are shown as mean ± SEM. Dot plots are gated on CD45^{+} leukocytes.

which have reduced MTMs (Fig. 2A and fig. S6), tumor development was unaffected (fig. S17A), implying a nonredundant function for RBPJ-dependent TAMs in promoting tumor growth. In the PyMT model, CD11c$^+$ myeloid cells act as antigen-presenting cells, forming stable, yet unproductive, interactions with tumor-infiltrating T cells (*12*). We hypothesized that one tumor-promoting function of TAMs may be their control of the adaptive immune response. Granzyme B (GzmB) is a cytolytic molecule important for tumor immunosurveillance. Conversely, Programmed Death-1 (PD-1) is an inhibitory co-receptor denoting "exhausted" T cells. As PyMT tumors progressed, an increase in PD-1$^+$GzmB$^-$CD8$^+$ T cells was observed (Fig. 4D), paralleling TAM expansion, with PD-1$^+$ cells making up ~50% of late-stage tumor-infiltrating CD8$^+$ T cells (Fig. 4E). In CD11c^{cre}*Rbpj*$^{fl/fl}$ PyMT mice, the PD-1$^+$ population decreased, whereas the GzmB$^+$ population increased (Fig. 4, F and G). In contrast, the T cell phenotype in Ccr2$^{-/-}$ mice was unchanged (fig. S17, B to D). Classical DCs also delete RBPJ in the CD11c^{cre}*Rbpj*$^{fl/fl}$ system, so we examined Flt3L-deficient PyMT mice that lack cells of the DC lineage, but have a comparable TAM population (fig. S3). *Flt3l*$^{-/-}$ PyMT mice showed no difference in tumor growth or CD8$^+$ T cell phenotype (fig. S17, E to H). These data suggest a specific function for RBPJ-dependent TAMs in promoting tumor immune tolerance in part by modulating the CD8$^+$ T cell response (fig. S18).

Other RBPJ-dependent TAM functions, including nonimmune regulatory roles, require further investigation. In addition, TAM differentiation is not completely abolished in the absence of RBPJ, raising the possibility that the remaining TAM intermediates may have tumor-promoting activities. Furthermore, to what extent TAM differentiation from monocytes occurs in other murine and in human tumors remains to be determined. Nonetheless, our findings suggest that a better understanding of this distinct tumor-elicited inflammatory response may create new opportunities for cancer treatment.

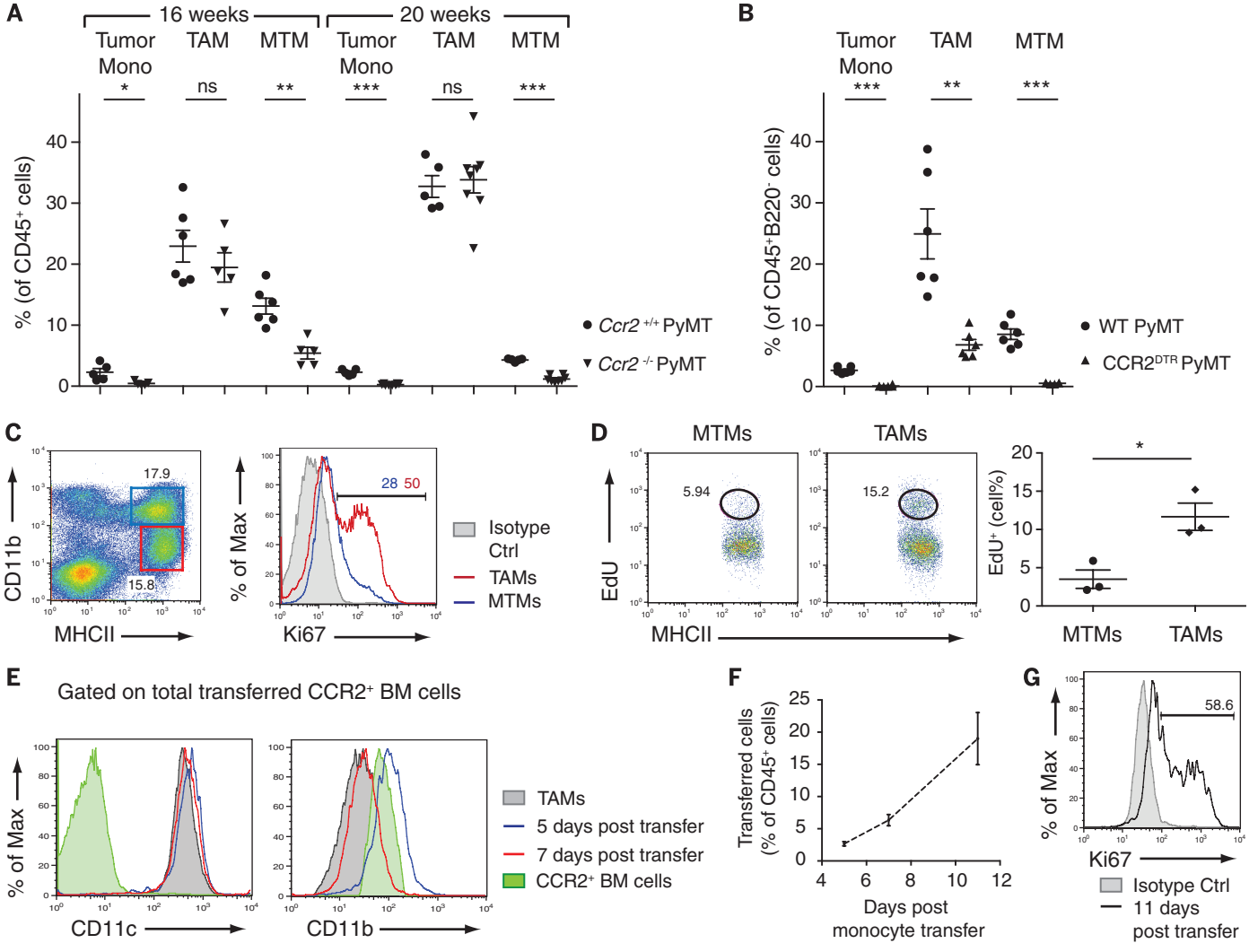


Fig. 2. TAMs differentiate from CCR2$^+$ inflammatory monocytes. (**A**) Flow cytometric analysis of tumor monocytes (mono), TAMs, and MTMs from 16- and 20- week-old *Ccr2*$^{+/+}$ PyMT and *Ccr2*$^{-/-}$ PyMT mice (n = 5 to 8). Data are pooled from five independent experiments. (**B**) Flow cytometric analysis of tumor monocytes, TAMs, and MTMs from WT PyMT and CCR2^{DTR} PyMT mice after DT treatment (mice were treated intraperitoneally every 3 days, seven treatments total) (n = 6). Data are pooled from four independent experiments. (**C**) Ki67 staining in TAMs and MTMs from 16-week-old PyMT mice. Data are representative of four independent experiments. (**D**) EdU incorporation in MTMs and TAMs from 16-week-old PyMT mice 20 hours after intraperitoneal EdU injection. MTM and TAM populations are first gated on CD45$^+$ cells and then gated as shown in (C). Data are representative of two independent experiments. (**E**) Surface expression of CD11c and CD11b on transferred CCR2$^+$ bone marrow cells from CCR2^{GFP} mice 5 and 7 days after transfer, gated on total transferred cells. (**F**) Percentage of total CD45$^+$ leukocytes that are of CCR2^{GFP} donor origin as identified by congenic marker 5, 7, and 11 days after transfer (n = 3 per time point). Data are pooled from three independent experiments. (**G**) Cell proliferation of transferred cells 11 days after transfer. All comparisons were made using student's t test, and data are shown as mean ± SEM. Statistical significance is indicated by *P < 0.05, **P < 0.01, ***P < 0.001; ns, not statistically significant.

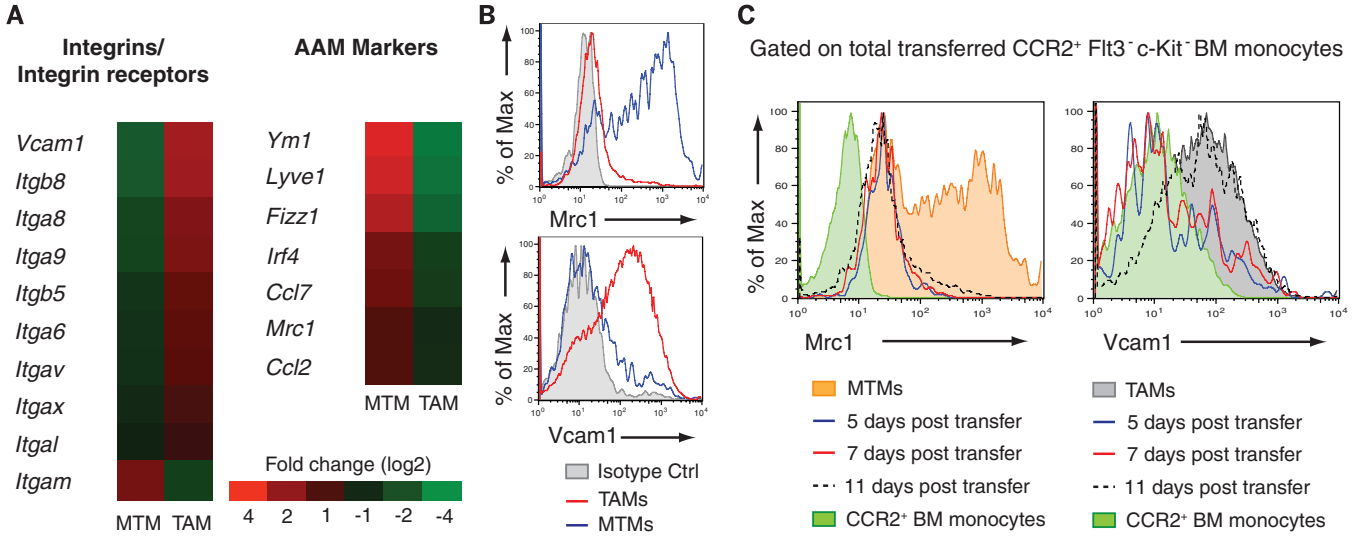


Fig. 3. TAMs are not phenotypically AAMs and can be identified by Vcam1 expression. (**A**) Gene expression of sorted TAMs and MTMs from 16-week-old PyMT mice. Data are pooled from three replicate microarray experiments (two MTM samples and three TAM samples). Differentially expressed genes were determined with a *P*-value threshold of 0.05. Fold change is depicted with a $\log_2$ scale. (**B**) Flow cytometry of Vcam1 and Mrc1 expression on TAMs and MTMs. Data are representative of more than five independent experiments. (**C**) Mrc1 and Vcam1 expression on transferred CCR2$^+$Flt3$^-$c-Kit$^-$ bone marrow monocytes and their progenies at 5, 7, and 11 days after transfer into DT-treated CCR2^{DTR} PyMT recipients (n = 2 mice per time point). Mrc1 and Vcam1 expression on MTMs or TAMs, respectively, are shown for comparison.

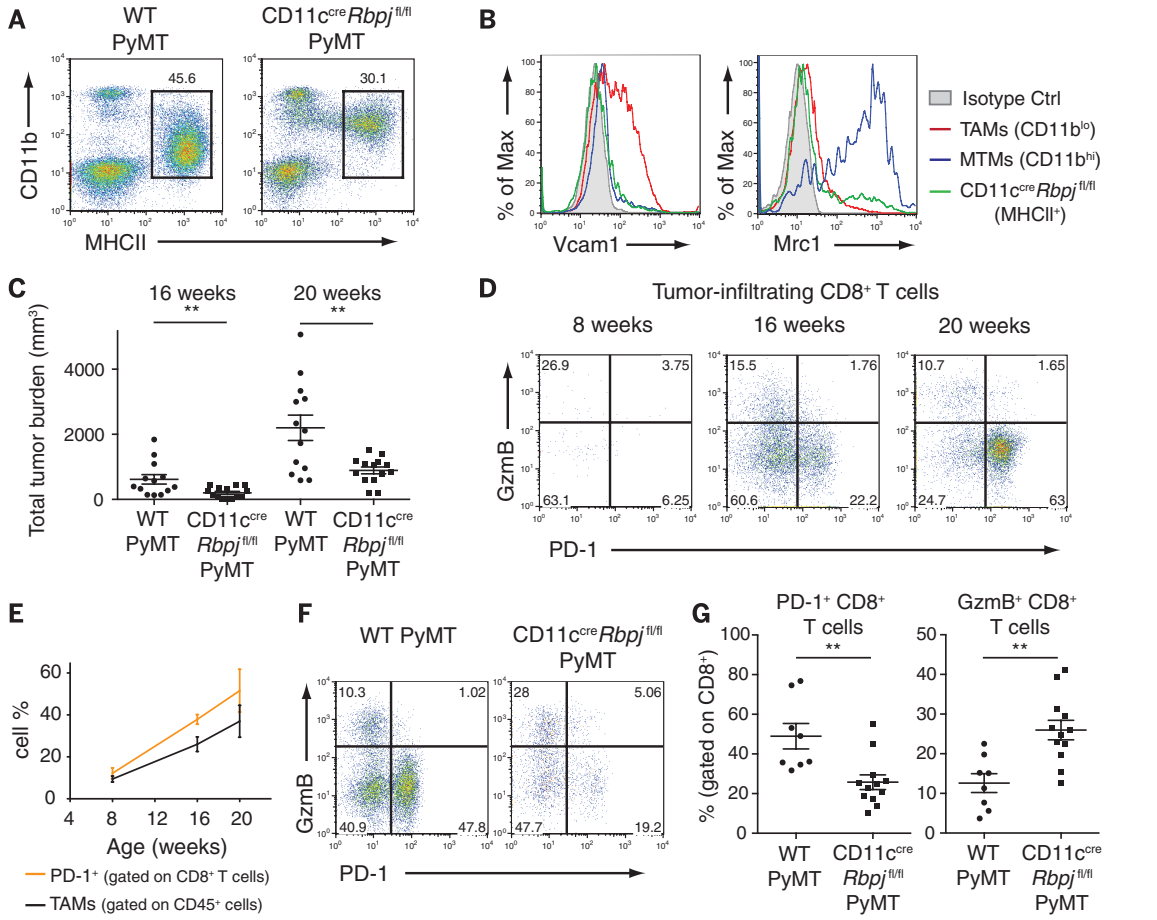


Fig. 4. RBPJ-dependent TAMs modulate the adaptive immune response. (**A**) Flow cytometric data of myeloid populations in 20-week-old CD11c^{cre} *Rbpj*$^{fl/fl}$ PyMT and WT PyMT mice. Plots are gated on CD45$^+$B220$^-$ cells. Data are representative of more than five independent experiments. (**B**) Flow cytometry of Vcam1 and Mrc1 expression on TAMs and MTMs from WT PyMT mice and MHCII$^+$ cells from CD11c^{cre} *Rbpj*$^{fl/fl}$ PyMT mice. Data are representative of three independent experiments. (**C**) Total tumor burden of WT PyMT and CD11c^{cre}*Rbpj*$^{fl/fl}$ PyMT mice at 16 and 20 weeks of age (n = 13 to 14). (**D**) Flow cytometric analysis of GzmB and PD-1 expression in CD8$^+$ T cells infiltrating PyMT tumors at 8, 16, and 20 weeks. Data are representative of three independent experiments. (**E**) Quantification of TAMs (gated on CD45$^+$ cells) and PD-1$^+$CD8$^+$ T cells (gated on CD45$^+$TCRβ$^+$CD8$^+$ cells) from 8-, 16-, and 20-week-old PyMT mice (n = 3 per time point). (**F**) PD-1 and GzmB expression in CD8$^+$ T cells from WT PyMT or CD11c^{cre}*Rbpj*$^{fl/fl}$ PyMT mice. Data are representative of more that five independent experiments. (**G**) Quantification of PD-1$^+$ or GzmB$^+$ CD8$^+$ T cells as in (F) (n = 8 to 12). Results represent pooled data and are shown as mean ± SEM. Student's t test was performed and statistical significance is indicated by $^{**}P < 0.01$.

REFERENCES AND NOTES

1. P. J. Murray, T. A. Wynn, *Nat. Rev. Immunol.* **11**, 723–737 (2011).
2. D. Hashimoto *et al*., *Immunity* **38**, 792–804 (2013).
3. C. Schulz *et al*., *Science* **336**, 86–90 (2012).
4. S. Yona *et al*., *Immunity* **38**, 79–91 (2013).
5. S. J. Jenkins *et al*., *Science* **332**, 1284–1288 (2011).
6. C. Jakubzick *et al*., *Immunity* **39**, 599–610 (2013).
7. E. Zigmond *et al*., *Immunity* **37**, 1076–1090 (2012).
8. C. Shi, E. G. Pamer, *Nat. Rev. Immunol.* **11**, 762–774 (2011).
9. T. A. Wynn, A. Chawla, J. W. Pollard, *Nature* **496**, 445–455 (2013).
10. C. T. Guy, R. D. Cardiff, W. J. Muller, *Mol. Cell. Biol.* **12**, 954–961 (1992).
11. E. Y. Lin, A. V. Nguyen, R. G. Russell, J. W. Pollard, *J. Exp. Med.* **193**, 727–740 (2001).
12. J. J. Engelhardt *et al*., *Cancer Cell* **21**, 402–417 (2012).
13. A. T. Satpathy, X. Wu, J. C. Albring, K. M. Murphy, *Nat. Immunol.* **13**, 1145–1154 (2012).
14. E. L. Gautier *et al*., *Nat. Immunol.* **13**, 1118–1128 (2012).
15. J. C. Miller *et al*., *Nat. Immunol.* **13**, 888–899 (2012).
16. H. J. McKenna *et al*., *Blood* **95**, 3489–3497 (2000).
17. V. Gouon-Evans, M. E. Rothenberg, J. W. Pollard, *Development* **127**, 2269–2282 (2000).
18. C. Jakubzick *et al*., *J. Immunol.* **180**, 3019–3027 (2008).
19. K. Liu *et al*., *Science* **324**, 392–397 (2009).
20. F. Geissmann *et al*., *Science* **327**, 656–661 (2010).
21. N. V. Serbina, E. G. Pamer, *Nat. Immunol.* **7**, 311–317 (2006).
22. T. M. Hohl *et al*., *Cell Host Microbe* **6**, 470–481 (2009).
23. N. V. Serbina, T. M. Hohl, M. Cherny, E. G. Pamer, *J. Immunol.* **183**, 1900–1910 (2009).
24. A. Sica, A. Mantovani, *J. Clin. Invest.* **122**, 787–795 (2012).
25. D. G. DeNardo *et al*., *Cancer Cell* **16**, 91–102 (2009).
26. H. W. Wang, J. A. Joyce, *Cell Cycle* **9**, 4824–4835 (2010).
27. A. T. Satpathy *et al*., *Nat. Immunol.* **14**, 937–948 (2013).
28. F. Radtke, H. R. MacDonald, F. Tacchini-Cottier, *Nat. Rev. Immunol.* **13**, 427–437 (2013).
29. M. L. Caton, M. R. Smith-Raska, B. Reizis, *J. Exp. Med.* **204**, 1653–1664 (2007).
30. K. L. Lewis *et al*., *Immunity* **35**, 780–791 (2011).

ACKNOWLEDGMENTS

We thank T. Honjo for the *Rbpj* $^{fl/fl}$ mouse strain and B. Reizis for the CD11c^{cre} mouse strain. We also thank J. Joyce and the M. Li lab for their insightful discussions. The data presented in this paper are tabulated in the manuscript and in the supplementary materials (Gene Expression Omnibus accession no. GSE56755). MSKCC has filed a provisional patent application with the U.S. Patent and Trademark Office (application no. 61/935,318). The application is directed toward methods and compositions for targeted cancer therapy and to identify potential responders to targeted therapy based on the presence of specific tumor-associated macrophages. R.A.F and M.O.L are listed as inventors on this patent application. This work was supported by the Cancer Research Institute Tumor Immunology Predoctoral Fellowship Training Grant (R.A.F), NIH grant AI101251 (K.L.), Cancer Research Institute Clinic and Laboratory Integration Program Grant (M.O.L.), and the American Cancer Society Research Scholar Award (M.O.L.).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/344/6186/921/suppl/DC1
Materials and Methods
Figs. S1 to S18
References (*31*–*33*)

21 February 2014; accepted 16 April 2014
Published online 8 May 2014;
10.1126/science.1252510

Imaging Cytometer

The next generation SpectraMax MiniMax 300 Imaging Cytometer enables both cellular visualization and first-of-its-kind stain-free cell-based analysis on the field-upgradable SpectraMax i3 Multi-Mode Microplate Reader. Brightfield and fluorescence-based green and red channel cellular image acquisition and analysis is made simple using the SoftMax Pro software workflow. Combining cellular imaging with stain-free cell counting and confluency measurements with microplate-based applications, scientists now have more ways to compress their workflows and increase efficiency. The SpectraMax MiniMax 300 Imaging Cytometer now features the patent-pending StainFree Cell Detection algorithm, which enables cell confluency and cell counting measurements on an imaging plate reader without the need for destructive stains, saving researchers valuable time and money. With two additional fluorescence detection channels, green and red, researchers may now perform and analyze a wide range of cellular viability and cell toxicity assays, including ratiometric assays such as transfection efficiency.

Molecular Devices
For info: 800-635-5577
www.moleculardevices.com

Amino Acid Analysis Reagent

The EZ Nin Reagent is a ready-to-use ninhydrin formulation for use in ion exchange chromatography systems with post-column derivatization for improved amino acid analysis. The EZ Nin Reagent has been developed to save laboratory time and waste; it requires no mixing or nitrogen and does not degrade with time or exposure to oxygen, providing reliable analysis over time without loss of sensitivity. Challenges encountered with standard ninhydrin include the long preparation process, instability of the reagent, degradation and waste, and batch-to-batch variability. Each batch of EZ Nin Reagent is identically prepared, enabling the reagent to be topped up on the instrument, eliminating waste or the risk of running out during a series of analyses. EZ Nin Reagent is methanol and hydrindantin free. Biochrom's formulation combines ease of use with unprecedented chemical stability, providing laboratories with the confidence they require when evaluating concentrations of amino acids for diagnosis and monitoring of inborn errors of metabolism.

Biochrom
For info: +44-(0)-1223-423723
www.biochrom.co.uk

Cell/Particle Counter

The latest Multisizer 4e Coulter Counter features a new 10 µm aperture that enables users to obtain accurate count, size, and mass distribution for particles and cells ranging from 0.2 µm to 1,600 µm. Digital Pulse Processing provides high-resolution analysis in as many as 400 channels for unparalleled performance in pharmaceutical quality control, bioprocessing, and industrial applications. Incorporating the latest technology and software advances, the Multisizer 4e with unique Digital Pulse Processing delivers dynamic size measurements in real time. High-speed signal digitalization allows the use of various pulse parameters for more accurate cell and particle characterization. Response is unaffected by particle color, shape, composition, or refractive index. Biotech applications include monitoring of transfection; counting mitochondria, organelles, and other intracellular components; and characterizing changes in cell size when changes are made to cultures. In these applications the volumetric analysis of the Multisizer 4e provides a more accurate and informative measurement than measuring size alone.

Beckman Coulter Life Sciences
For info: 800-526-3821
www.beckmancoulter.com

Aptamer Kits

New aptamer kits are now available for cell isolation and flow cytometry. In addition to their discriminate molecular recognition properties, aptamers offer advantages over antibodies as they can be engineered completely in a test tube, are readily produced by chemical synthesis, possess desirable storage properties, and elicit little or no immunogenicity in therapeutic applications. The AptoPrep cell isolation kit products combine the advantage of the specificity of aptamer to target cell with the convenience of isolation/elution of a particular cell from complex cell samples. The new kits offer unique characteristics and provide advantages over alternative methods. AptoPrep cell isolation kits deliver positive isolation in as little as 30 minutes. Mild and unique cell releasing technology allows users to isolate bead- and aptamer-free cells in as little as one hour. Supplied as a complete kit (biotinylated aptamer, streptavidin magnetic bead, and buffers), AptoPrep cell isolation products deliver unmatched convenience, increased consistency and reliability.

AMS Biotechnology
For info: +44-(0)-1235-828200
www.amsbio.com/aptamers.aspx

Barcode Cell-Tracking Library

The complex lentiviral-based barcode library allows users to track clonal populations of cells over time or through different treatments. The cell-tracking process begins with the introduction of a packaged lentiviral library encoding a highly heterogeneous population of barcodes into a population of founder cells. After selection, treatment, or differentiation, barcode representation in cells provides data on which clones from founder cells survived, thrived, or died out. The Barcode-Cell Tracking library contains barcodes compatible with the Illumina HiSeq Sequencing platform. For laboratories needing support for this analysis, Cellecta can provide HT Sequencing and Barcode Enumeration of samples. Users provide frozen cell pellets. Cellecta does the rest, including extracting genomic DNA, amplifying the barcodes, performing HT sequencing on the Illumina HiSeq, and enumerating the barcodes from raw sequencing data.

Cellecta Team
For info: 877-938-3910
www.cellecta.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/products/newproducts.dtl for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

Downloaded from www.sciencemag.org on May 26, 2014

By **Jim Austin**

A science career story

I was valedictorian of a large public high school in Florida, attended a top liberal arts college—Swarthmore—and majored in physics. After a short, post-college stint as a small-town journalist, I entered the physics Ph.D. program at the University of North Carolina, Chapel Hill. I connected early with a research project that allowed me to publish often and well, although I did not love the work. I finished my Ph.D. fairly quickly—faster than I needed to really because my wife, a chemist, was still in graduate school. So I stayed in the same lab for a postdoc, doing the same work, funding the position with a grant proposal written by me and submitted in my adviser's name. From graduate school on, I do not remember receiving a single piece of career advice.

A year or so into my postdoc, my adviser retired. I took over his lab, earning a very long title: visiting research assistant professor. I inherited some '80s-vintage electronics and a '60s-era lab with a desk in the corner.

As my wife approached the end of her Ph.D., we began to consider our "two-body problem." We agreed that we would accept the first good offer either of us received.

My job-market timing could not have been worse. The dissolution of the Soviet Union and the Eastern bloc sent many physicists and other scientists streaming west. Big corporate labs were downsizing and moving away from basic research, sending veteran physicists onto the academic job market. The early '90s employment crisis among young Ph.D. physicists made news. (You can read about it in *Science* at http://scim.ag/1qg8PzT.)

I appeared successful. I was running my own funded lab and publishing in good journals. But it wasn't long before I realized that I wasn't competitive for tenure-track positions at the institutions where I wanted to work, including the one where I was already working. One rejection letter among the many I received thanked me for being part of a "remarkable cohort" of more than a thousand applicants for a single faculty post. I believe the number actually exceeded 1300.

My wife applied for just one job and got the offer. I became the trailing spouse, following her to Maine, where she took up a faculty post. I took pride in defying science's gender stereotypes.

It was surprisingly easy to walk away from the career that I had worked so long and hard to attain. The hard part came later as I lost my knowledge of science and saw my mathematical facility dwindle, and as I struggled to fashion an identity that wasn't linked to professional attainment.

I appeared successful. I was running my own funded lab and publishing in good journals.

I turned to writing, and when I wasn't writing, I was repairing and maintaining a passive-solar house in the country: shoveling snow, hauling tons of firewood up a steep hill, planting a vegetable garden that didn't get enough sun, and fulfilling various back-to-nature clichés. For a while, I taught writing part-time to undergraduates; it was ideal training for the work that would come later. We soon had a son, and (except for the breastfeeding part) I was the primary parent.

In 1999 I founded an Internet publication, The Post-Careerist, which was focused on living a rich post-professional life. Through connections I made online, I became science editor at a pioneering Internet publication, BlueEar.

Then, via an acquaintance in Pakistan—this was early online networking, before LinkedIn and Facebook—I heard about a writer/editor position at *Science*'s Next Wave, *Science* Careers' predecessor. I sent an e-mail and within weeks found myself with a full-time job for the first time in years. A few years later, I became the editor of *Science* Careers.

My science career story is hardly unusual. Indeed, what's remarkable is how much it shares with so many other nontraditional career stories: uncertainty, exploration, a difficult transition, self-invention, and (eventually) satisfaction.

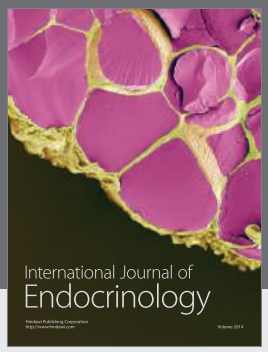
So what's your story? Send stories, perspectives, opinions, and observations on careers in the sciences to me at SciCareerEditor@aaas.org.

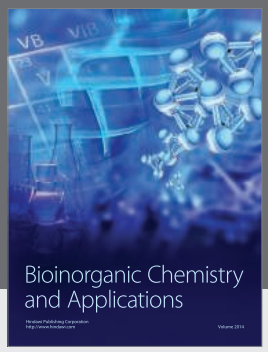
Jim Austin is the editor of Science *Careers. @SciCareerEditor on Twitter.*

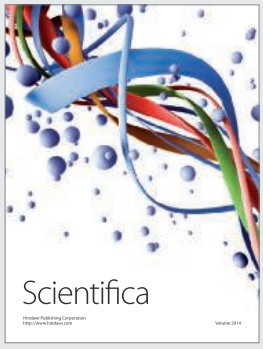
ILLUSTRATION: MARC ROSENTHAL

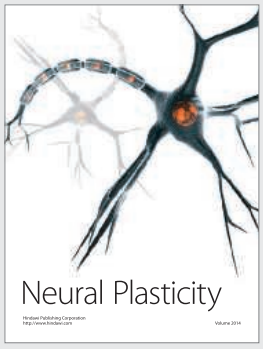
Downloaded from www.sciencemag.org on May 26, 2014

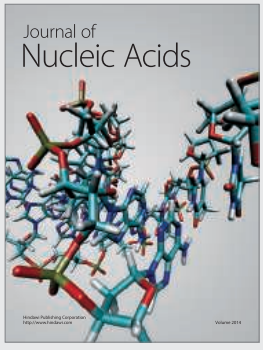
Published by AAAS

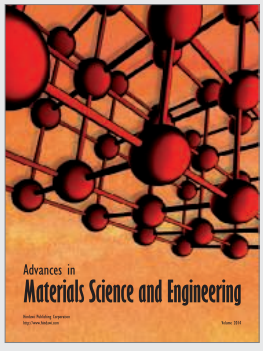


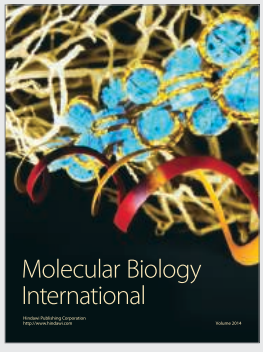


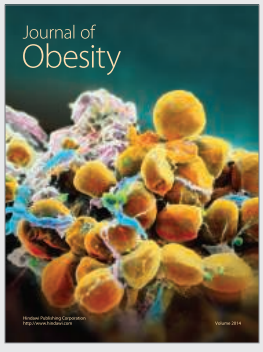


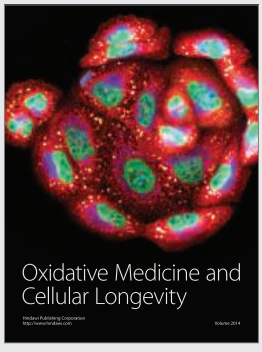


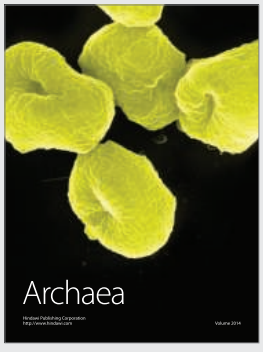






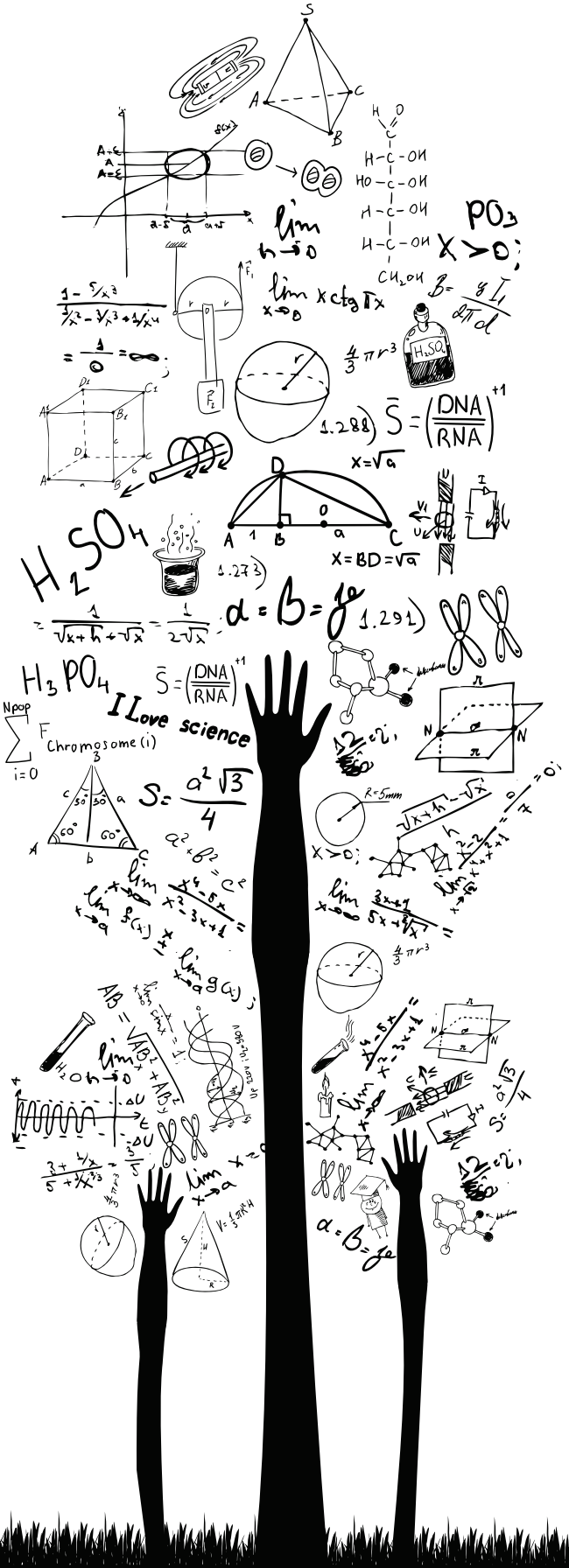






Hindawi

Submit your manuscripts at
http://www.hindawi.com



AAAS members are helping students reach new heights.

With more than 100 volunteers spending a day a week in K-12 classrooms in the D.C. area, and a budding program in Seattle, AAAS is committed to working with scientists and engineers to make scientific literacy possible for all students.

Now, thanks to a generous donation from a AAAS Fellow, we are growing these efforts into a nationwide network. AAAS is pleased to announce the launch of the pilot stage of the AAAS National STEM Volunteer Program. AAAS National STEM Volunteer seed grants have just been awarded to seven members to work with non-profits in their communities to bring scientists and engineers into local classrooms in an ongoing and meaningful way.

The awardees are: Dr. Julie Luft, University of Georgia; Dr. Ronald Evans, Salk Institute for Biological Studies; Dr. Philip Bell, Institute for Science and Math Education, University of Washington; Dr. Barry Shur, University of Colorado, Denver; Dr. Audrey Gasch, Foundation for Madison's Public Schools; Dr. Stefi Baum, Rochester Institute of Technology; and Dr. Edward Bilsky, University of New England.

In addition to funding, AAAS will provide resources for developing, implementing and sustaining the volunteer effort.

Please join us in congratulating all of the awardees and in thanking the nearly 70 AAAS members who applied to be a part of the AAAS National STEM Volunteer Program.

AAAS
NATIONAL STEM
VOLUNTEER PROGRAM

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

To learn more, visit: http://membercentral.aaas.org/volunteer/ANSVP.

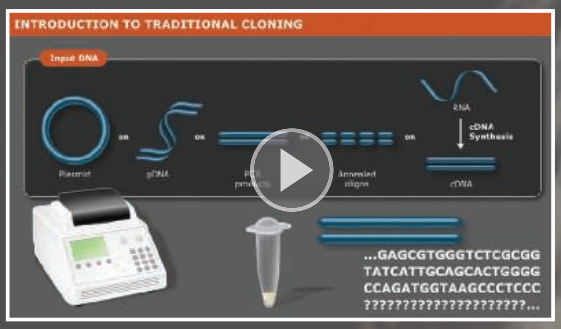
NEW ENGLAND BioLabs® Inc.

Clone with Confidence.

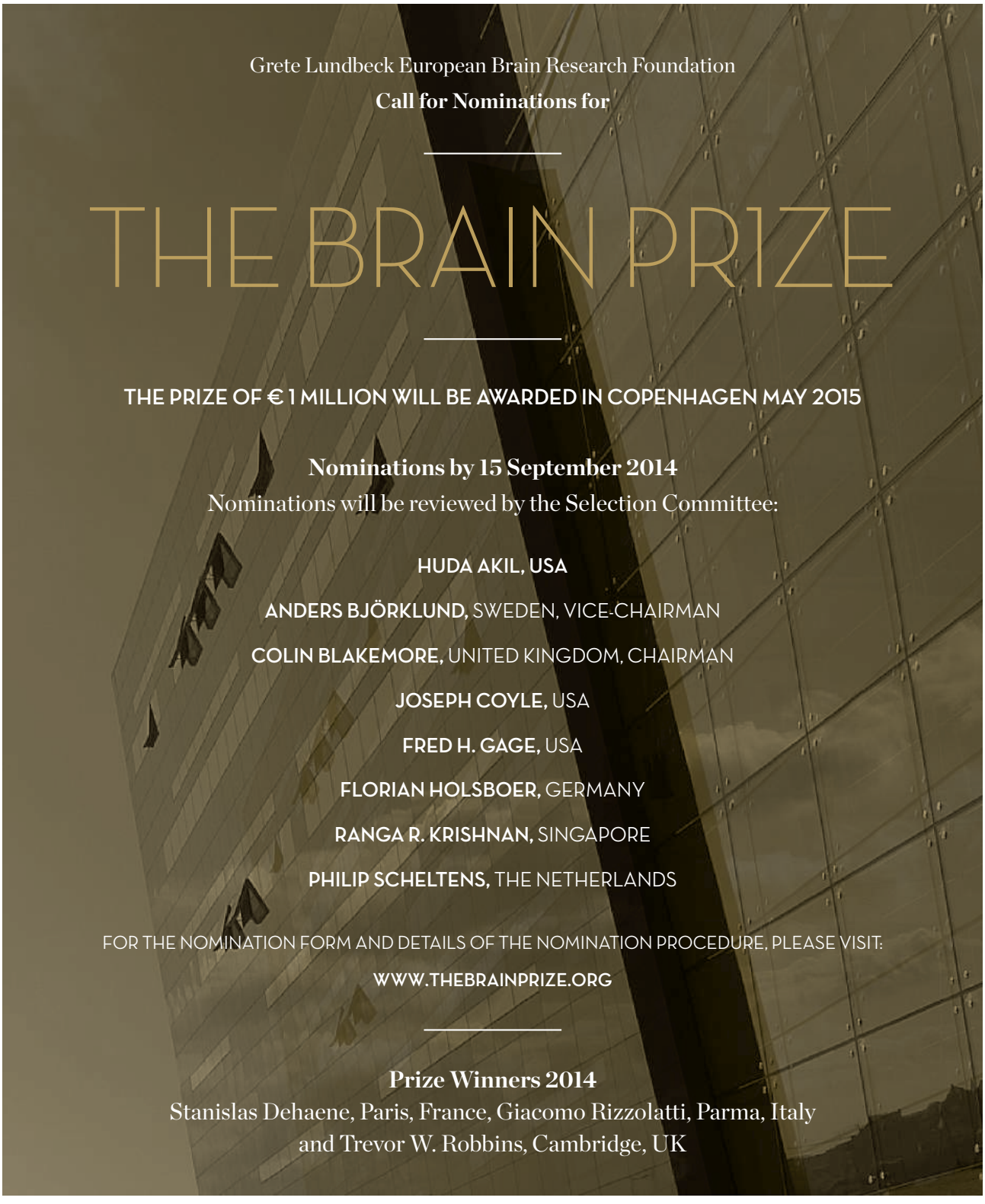
Whether you are performing your first cloning experiment, or constructing multi-fragment gene assemblies, NEB® has the solution for you. Our high quality reagents are available for every workflow, and include specialized enzymes, competent cells, and novel solutions – such as Gibson Assembly®. When you are looking to clone with confidence, think of NEB.

Explore the wise choice at **CloneWithNEB.com.**

Visit CloneWithNEB.com to view online tutorials describing various cloning workflows.



NEW ENGLAND BIOLABS® and NEB® are registered trademarks of New England Biolabs, Inc.
GIBSON ASSEMBLY® is a registered trademark of Synthetic Genomics, Inc.



Grete Lundbeck European Brain Research Foundation

Call for Nominations for

THE BRAIN PRIZE

THE PRIZE OF € 1 MILLION WILL BE AWARDED IN COPENHAGEN MAY 2015

Nominations by 15 September 2014

Nominations will be reviewed by the Selection Committee:

HUDA AKIL, USA

ANDERS BJÖRKLUND, SWEDEN, VICE-CHAIRMAN

COLIN BLAKEMORE, UNITED KINGDOM, CHAIRMAN

JOSEPH COYLE, USA

FRED H. GAGE, USA

FLORIAN HOLSBOER, GERMANY

RANGA R. KRISHNAN, SINGAPORE

PHILIP SCHELTENS, THE NETHERLANDS

FOR THE NOMINATION FORM AND DETAILS OF THE NOMINATION PROCEDURE, PLEASE VISIT:

WWW.THEBRAINPRIZE.ORG

Prize Winners 2014

Stanislas Dehaene, Paris, France, Giacomo Rizzolatti, Parma, Italy
and Trevor W. Robbins, Cambridge, UK

GRETE LUNDBECK
EUROPEAN
BRAIN RESEARCH
FOUNDATION

THE BRAIN PRIZE

The Brain Prize recognizes and rewards outstanding contributions to European neuroscience, from basic to clinical

Webinar

Solid Tumors Reveal Their Secrets

Predictive and Prognostic Evidence from Copy Number Analysis

Wednesday, June 11, 2014

12 noon Eastern, 9 a.m. Pacific, 5 p.m. UK, 6 p.m. Central Europe

Speakers

Paul C. Boutros, Ph.D.
Ontario Institute of Cancer Research
Toronto, Canada

Ajay Pandita, D.V.M., Ph.D.
Core Diagnostics
Palo Alto, CA

REGISTER NOW!
webinar.sciencemag.org

Recent research has revealed the importance of copy number analysis in the study of solid tumors. New evidence argues that tumors can be classified in those driven by either mutations (M class) or copy number aberrations (C class). C class tumors include breast, ovarian, and squamous cell lung cancers as well as, some believe, prostate cancer. Other publications have noted that predictive copy number changes are more frequent than predictive somatic mutation changes in solid tumor samples. These results seem to indicate that there is some risk to limiting tumor profiling to somatic mutations, and emphasize the importance of whole genome copy number analysis in identifying clinically relevant prognostic and predictive markers. This process is technically challenging, especially when using formalin-fixed, paraffin-embedded (FFPE) tissue or heterogeneous samples where only a small fraction of cells may be aberrant. Next generation sequencing technologies have limited ability to detect clinically relevant lower level amplifications, copy neutral loss of heterozygosity, and homozygous deletions, even at significant depth of coverage. This webinar will explore the importance of copy number analysis and highlight new technologies that are making this process more accessible, especially in solid tumor samples.

During the webinar, the speakers will:

- Discuss the importance of copy number analysis in solid tumors and its relevance for identifying predictive and prognostic biomarkers in cancers
- Provide their own data on copy number changes found in solid tumors and how these analyses have informed their research
- Highlight the technologies they have used to identify clinically relevant and informative copy number changes in solid tumor samples
- Answer your questions live during the webinar!

Webinar sponsored by

affymetrix
Biology for a better world

Brought to you by the *Science*/AAAS Custom Publishing Office

Science
AAAS

@SciMagWebinars

immunogenomics 2014

September 29 - October 1, 2014

HudsonAlpha Biotechnology Campus
Huntsville, Alabama, USA

Join us at the intersection of Immunology and Genomics!

Keynote Speakers

Christophe Benoist, Harvard Medical School
Mary Ellen Conley, University of Tennessee, College of Medicine, Memphis
Mark Davis, Howard Hughes Medical Institute; Stanford University School of Medicine

Session Chairs

Lou Bridges, University of Alabama at Birmingham, AL, USA • **Jean-Laurent Casanova**, Howard Hughes Medical Institute, Rockefeller • **Alain Fischer**, Imagine Institute, Necker Hospital, Paris, France • **Dan Littman**, New York University, NY, USA • **Sara Marsal**, University of Barcelona, Spain • **Kristen Mueller**, *Science* • **Harlan Robins**, University of Washington/Fred Hutchinson Cancer Research Center

Session Speakers

Scott Boyd, Stanford University School of Medicine, CA • **Zhijian James Chen**, UT Southwestern, TX • **Yanick Crow**, University of Manchester, UK • **Malek Faham**, Sequenta, CA • **Wendy Garrett**, Harvard University School of Public Health, MA • **Peter Gregersen**, Feinstein Institute for Medical Research, NY • **Jian Han**, HudsonAlpha Institute for Biotechnology, AL • **Lionel Ivashkiv**, Hospital for Special Surgery, NY • **Thirumala-Devi Kanneganti**, St. Jude Children's Research Hosp. • **Julian Knight**, Merton College , University of Oxford, UK • **Mike Lenardo**, National Institutes of Health, Washington D.C. • **Xochitl Morgan**, Harvard University School of Public Health, MA • **Barbara Methe**, J. Craig Venter Institute, MD • **Kees Murre**, University of California, San Diego, CA • **Richard M. Myers**, HudsonAlpha Institute for Biotechnology, AL • **Soumya Raychaudhuri**, Broad Institute, MA • **Lars Steinmetz**, European Molecular Biology Laboratory, Germany • **Barbara Stranger**, University of Chicago, IL USA

Platinum Sponsors

BAY CITY CAPITAL

iRepertoire, Inc.

Gold Sponsors

UAB COMPREHENSIVE ARTHRITIS, MUSCULOSKELETAL, AND AUTOIMMUNITY CENTER

Silver Sponsors

GenapSys

illumina

Bronze Sponsors

SEQUENTA

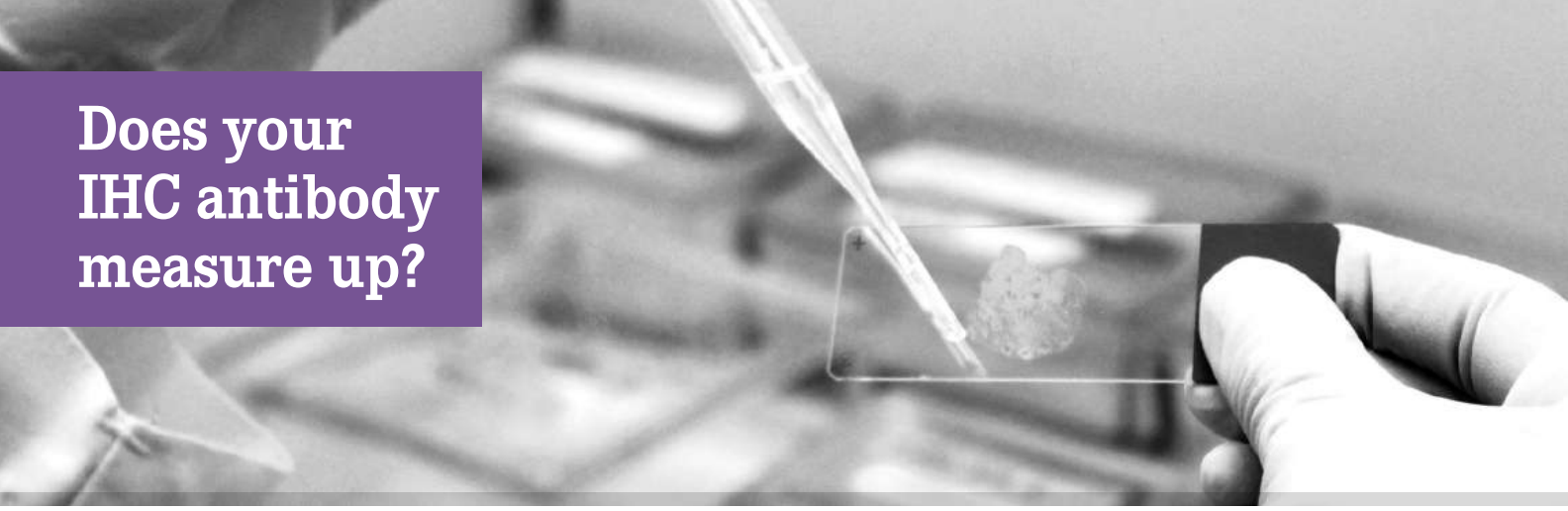
follow @immunogenomics on twitter

Register today at immunogenomics.com

presented by

HUDSONALPHA INSTITUTE FOR BIOTECHNOLOGY

Science AAAS



Does your IHC antibody measure up?

Rigorously Validated IHC Antibodies Against High Impact Research Targets

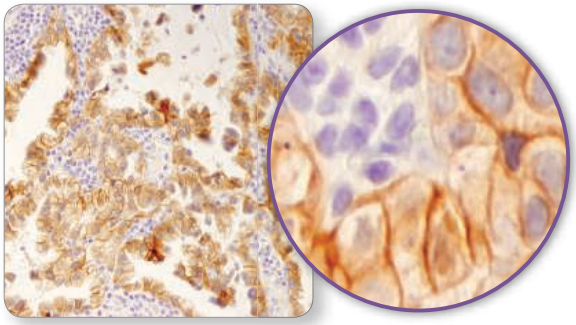
from Cell Signaling Technology

Paraffin-embedded human lung and ovarian carcinoma tissues using CST™ HER3/ErbB3 (D22C5) XP® Rabbit mAb #12708 or Company 2 mouse monoclonal antibody.

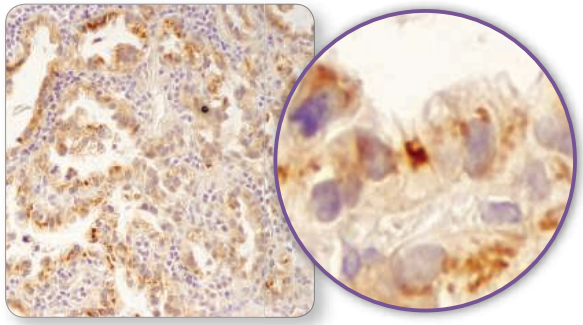
CST HER3/ErbB3 (D22C5) XP® Rabbit mAb #12708

Company 2 Mouse Antibody

Lung Carcinoma

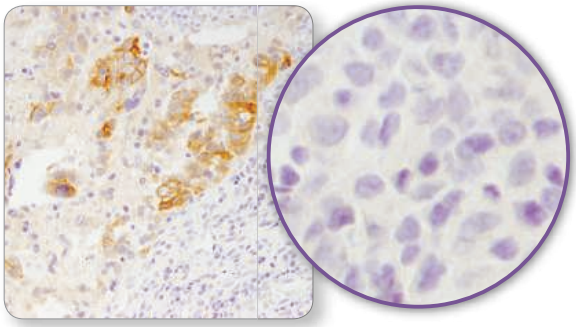


Strong, specific plasma membrane-associated HER3 staining in the epithelium

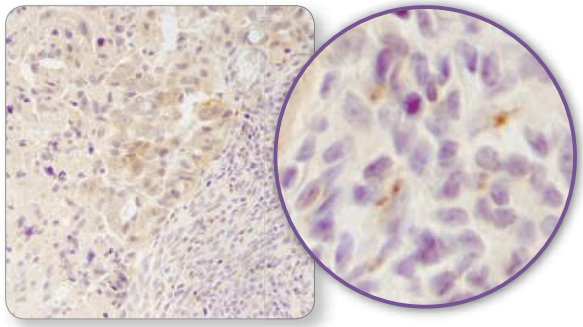


Non-specific, non-membranous staining in the epithelium

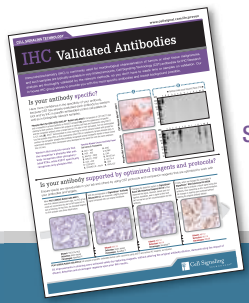
Ovarian Carcinoma



Strong, specific signal in tumor cells, and no staining seen in stromal cells



Weak, signal in tumor cells, and non-specific staining in stromal cells



WB IP IHC IF F ChIP

Skeptical? Go to **www.cellsignal.com/ihcscience** to see more

For Research Use Only. Not For Use In Diagnostic Procedures.

© 2014 Cell Signaling Technology, Inc. Cell Signaling Technology®, CST™, and XP® are trademarks of Cell Signaling Technology, Inc.

14PADIHC_NONE0152ENG_00

Cell Signaling TECHNOLOGY®

Canadian Food Inspection Agency | Agence canadienne d'inspection des aliments

Research Scientist

The Canadian Food Inspection Agency (CFIA) is the largest science-based federal regulatory agency in Canada. It is responsible for delivering all federally mandated programs for food inspection, plant and animal health. The CFIA relies on high-quality, timely and relevant science to make informed regulatory decisions. The CFIA's Science Branch has a network of 13 laboratories across the country and approximately 1000 dedicated staff that deliver on activities related to laboratory diagnostic tests, research, technology and methods development, surveillance and science advice. Four of these laboratories are recognized as reference laboratories for the World Organization of Animal Health (OIE) and are considered to be international centres for designated animal diseases.

The CFIA wishes to hire Animal Health Research Scientists who are recognized as an authority and visionary in areas of bacteriology, virology, immunology, pathology or parasitology and have strategically integrated leading edge scientific and technical objectives into developing Animal Health programs that have long-term impact on organizational priorities and influenced future directions of research.

The successful candidate will have a Ph.D. with a specialization in a relevant biological science (e.g. bacteriology, virology, immunology, pathology, and parasitology) and experience related to conducting and managing research in Animal Health.

For further details including all required selection criteria and how to apply, please visit: www.jobs.gc.ca and search on Reference Number **ICA14J-010156-000020** (Selection Process Number **14-ICA-CLSTF-OE-VAR-HRSD-009**) or follow the direct link: https://emploisfp-psjobs.cfp-psc.gc.ca/psrs-srfp/applicant/page1800?poster=240623&toggleLanguage=en&psrsMode=1&noBackBtn=true

Please note that all candidates' applications will be reviewed regardless of their individual level or tenure of Productivity/Recognition and Experience against the screening requirements.

The Canadian Food Inspection Agency (CFIA) does not give a preference to Canadian citizens. All interested individuals are encouraged to apply.

Please note that this selection process will remain open until a suitable candidate is identified. However, applications received by **June 9, 2014** will be given first considerations.

More information on the Canadian Food Inspection Agency can be found on our website at **www.inspection.gc.ca**

Canada

CNPEM
Brazilian Center for Research in Energy and Materials

SEARCH FOR A NEW CEO
Campinas, São Paulo

The Brazilian Center for Research in Energy and Materials (CNPEM), located in the city of Campinas – SP, Brazil, is a private nonprofit Organization that manages for the Brazilian Ministry of Science, Technology and Innovation (MCTI) four national laboratories: **Brazilian Synchrotron Light Laboratory** (LNLS), **Brazilian Biosciences Laboratory** (LNBio), **Brazilian Bioethanol Science and Technology Laboratory** (CTBE) and **Brazilian Nanotechnology Laboratory** (LNNano).

In order to find a new Chief Executive Officer (CEO), CNPEM is looking for a leader with innovative vision, to lead and help define its future strategies. He will lead a staff of 500 professionals, 70% of them directly linked to the Center scientific and development activities. He (or she) must have intellectual leadership skills, ability to understand and communicate complex scientific ideas, and be a team builder, enabling talent, knowledge and innovation development. He (or she) will also be the CNPEM´s public voice, interacting with the scientific community, with public and private operation companies, and with public policy agents.

Abilities needed for the position:

- Leadership experience of ten or more years in R&D organizations.
- PhD Research Degree.
- Advocacy skills to be used on behalf of CNPEM projects, initiatives and fundraising.
- Fluency in the Portuguese language

The Chief Executive Officer serves a three-year term, renewable. The Center will provide assistance with the expense of moving and establishing residence in Campinas

Please access http://www.cnpem.br/cnpem/selecao-diretor-geral/ to find more information about the recruiting process. For further information about the position or to show interest, please contact selecaodiretorgeral@cnpem.br. Detailed information about the CNPEM´s campus and activities are available at www.cnpem.br. Applications must be sent until June 30 2014.

Brazilian Center for Research in Energy and Materials (CNPEM)
Rua Giuseppe Máximo Scolfaro, 10.000 - Polo II de Alta Tecnologia
Caixa Postal 6192 - 13083-970 - Campinas/SP
www.cnpem.br

Science Careers Advertising

For full advertising details, go to ScienceCareers.org and click For Employers, or call one of our representatives.

Tracy Holmes
Worldwide Associate Director
Science Careers
Phone: +44 (0) 1223 326525

THE AMERICAS

E-mail: advertise@sciencecareers.org
Fax: 202-289-6742

Tina Burks
Phone: 202-326-6577

Nancy Toema
Phone: 202-326-6578

Marci Gallun
Sales Administrator
Phone: 202-326-6582

Online Job Posting Questions
Phone: 202-312-6375

EUROPE / INDIA / AUSTRALIA / NEW ZEALAND / REST OF WORLD

E-mail: ads@science-int.co.uk
Fax: +44 (0) 1223 326532

Axel Gesatzki
Phone: +44 (0)1223 326529

Sarah Lelarge
Phone: +44 (0) 1223 326527

Kelly Grace
Phone: +44 (0) 1223 326528

JAPAN

Yuri Kobayashi
Phone: +81-(0)90-9110-1719
E-mail: ykobayas@aaas.org

CHINA / KOREA / SINGAPORE / TAIWAN / THAILAND

Ruolei Wu
Phone: +86-1367-1015-294
E-mail: rwu@aaas.org

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. *Science* reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. *Science* encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

Science Careers
From the journal *Science* AAAS

ScienceCareers.org

Center for Immunology & Microbial Disease
Albany Medical College

Faculty Position

The Center for Immunology & Microbial Disease at Albany Medical College invites applications for a tenure-track, junior or senior faculty position from individuals who have a doctoral degree, postdoctoral experience, and demonstrated research productivity. Those with an interest in host-pathogen interactions are particularly encouraged to apply. The successful candidate will be expected to establish an independent, extramurally-funded research program and participate in the teaching of medical and graduate students. The basic science departments at Albany Medical College are organized as interdisciplinary research centers and the Center for Immunology & Microbial Disease has a focus on microbial pathogenesis and immune defense, particularly as related to biothreat agents and emerging infections. The new faculty recruit will receive a competitive salary, an attractive start-up package, laboratory space in our newly constructed research building, and access to all departmental core services including the Center's fully-staffed Immunology and ABSL-3/BSL-3 Cores. Albany Medical College is located in a mid-sized city within the upstate New York Capital Region, and has easy access to Boston, New York City, and the Adirondack Mountains.

Applicants should send their curriculum vitae, a statement of research plans, and three letters of reference to:

Faculty Search Committee
Center for Immunology & Microbial Disease
Albany Medical College
47 New Scotland Avenue, MC-151
Albany, NY 12208

For further information about the Center, visit **www.amc.edu/Academic/Research/imd.htm**

An Equal Opportunity/Affirmative Action Employer. Women and minorities are encouraged to apply.

THE UNIVERSITY OF HONG KONG
香 港 大 學

Tenure-Track Assistant Professor in Chemical Biology (Ref.: 201400438)

Applications are invited for tenure-track appointment as Assistant Professor in Chemical Biology in the Department of Chemistry. The position will be tenable from January 1, 2015 or as soon as possible thereafter, on a three-year term basis, with the possibility of renewal and with consideration for tenure during the second three-year contract.

Applicants should have a Ph.D. degree with a strong background and research record in the general area of chemical biology. The appointee is expected to develop original and independent research programs, and excel in both undergraduate and postgraduate teaching. A suitable start-up fund for research will be provided to the appointee. Information about the Department can be obtained at http://www.chemistry.hku.hk.

A globally competitive remuneration package commensurate with qualifications and experience will be offered. At current rates, salaries tax does not exceed 15% of gross income. The appointment will attract a contract-end gratuity and University contribution to a retirement benefits scheme, totalling up to 15% of basic salary, as well as annual leave, and medical benefits. Housing benefits will be provided as applicable.

Applicants who have responded to the previous advertisement (Ref.: 201301014) need not re-apply.

Applicants should send a completed application form, together with an up-to-date C.V., a research proposal, and a statement of teaching philosophy by e-mail to scchem@hku.hk. They should also arrange for submission, to the same e-mail address as stated above, three reference letters from senior academics. One of these senior academics should be asked to comment on the applicant's ability in teaching. Please indicate clearly the reference number in the subject of the e-mail. Application forms (341/1111) can be obtained at http://www.hku.hk/apptunit/form-ext.doc. Further particulars can be obtained at http://jobs.hku.hk/. **Closes July 31, 2014**.

The University thanks applicants for their interest, but advises that only shortlisted applicants will be notified of the application result.

The University is an equal opportunity employer and is committed to a No-Smoking Policy

FUNDING OPPORTUNITIES
U.S. Department of Defense
Defense Medical Research and Development Program
Peer Reviewed Medical Research Program

The Peer Reviewed Medical Research Program (PRMRP) funds exceptional research with the goal to improve the health and well-being of all military service members, Veterans, and beneficiaries. The PRMRP received ***$200 million*** in fiscal year 2014 (FY14), and seeks grant applications in the following ***topic areas***:

Acupuncture
Arthritis (other than post-traumatic osteoarthritis and rheumatoid arthritis)
Chronic Migraine and Post-Traumatic Headaches
Congenital Heart Disease
DNA Vaccine Technology for Post-Exposure Prophylaxis
Dystonia
Epilepsy
Food Allergies
Fragile X Syndrome
Hereditary Angioedema
Illnesses Related to Radiation Exposure (excludes cancer)
Inflammatory Bowel Disease
Interstitial Cystitis
Lupus
Malaria
Metabolic Disease
Neuroprosthetics
Pancreatitis
Polycystic Kidney Disease
Post-Traumatic Osteoarthritis
Psychotropic Medications
Respiratory Health (excludes lung cancer and mesothelioma)
Rheumatoid Arthritis
Segmental Bone Defects
Tinnitus

The FY14 PRMRP Program Announcements and General Application Instructions are anticipated to be posted on Grants.gov by ***May 1, 2014:***

- Clinical Trial Award
- Discovery Award
- Focused Program Award
- Investigator-Initiated Research Award
- Technology/Therapeutic Development Award

All applications must conform to the Program Announcements and General Application Instructions that will be available for electronic downloading from the Grants.gov website (all viewable under CFDA number 12.420). Execution management support will be provided by the Congressionally Directed Medical Research Programs.

http://cdmrp.army.mil
http://cdmrp.army.mil/funding/prmrp.shtml

HEAD
DEPARTMENT OF ANATOMY AND CELL BIOLOGY
University of Illinois at Chicago
College of Medicine

The University of Illinois at Chicago (UIC) College of Medicine, one of the most diverse medical schools in the nation, invites applications and nominations for the position of Head of the Department of Anatomy and Cell Biology. We are looking for an outstanding individual with a clear vision on how to build and lead a world-class department with a tradition of excellence in the neurosciences. Significant resources are available to expand the department and support that vision. UIC is a major public research university and the College of Medicine is currently expanding, particularly in areas of translational research. UIC has a large and varied research portfolio, outstanding research resources and close ties between basic and clinical departments, particularly those in the neurosciences. The Head will also oversee significant departmental responsibilities for medical and graduate education. Information on the Department's interests and composition can be viewed at http://www.anatomy.uic.edu/. Candidates should have a PhD and/or M.D. degree, a well funded and established research program in an area of neuroscience, and strong leadership and administrative skills. The candidate should currently hold or qualify for the rank of Professor (or equivalent) and is expected to expand the departmental research portfolio while maintaining faculty involvement and excellence in the department's educational programs. Informal enquiries may be addressed to the chair of the search committee, **Dr. Jack Kaplan**, at **kaplanj@uic.edu**. For fullest consideration interested individuals should submit a curriculum vitae by **August 1, 2014** to:

Jack H. Kaplan, PhD, FRS
Benjamin Goldberg Professor & Head
Department of Biochemistry & Molecular Genetics
Attn: Ms Jacqueline Taylor
University of Illinois at Chicago
Office of the Dean, Rm. 131
1853 W. Polk Street, M/C 784
Chicago, IL 60612

The University of Illinois at Chicago is an Affirmative Action/Equal Opportunity Employer, Women and Minorities are encouraged to apply.

About NUAA

Nanjing University of Aeronautics and Astronautics(NUAA) is a research-oriented national key university of "211 Project". It also enjoys a well-balanced development of multiple disciplines in engineering, technology, natural sciences, economy, management and social sciences with the characteristics of aeronautics, astronautics and civil aviation. NUAA is qualified to be "Dominant Discipline Innovation Platform of 985 Project" and to independently recruit and receive international students who are granted the Chinese Government Scholarship. Now NUAA consists of 16 colleges with more than 3,000 staff members and approximately 26,000 degree students.

Academia and education at NUAA represent strong capacity among all the universities in China. It has acquired national status through the quality of its excellence research work, especially in the areas of Aerospace Engineering, Mechanics,Electromechanics, Economy and Management, etc.

NUAA gives a warm welcome to excellent experts, scholars and young students from both home and abroad, who are willing to serve the country, dedicate themselves to the development of aerospace science and make contributions to the industrialization, information technology of China. NUAA will provide teachers and researchers with a good academic environment, satisfactory working and living conditions and a stage on which they can put their talents to good use.

Contacts

Ms. Zhao Haiyan, Mr. Cao Yunxing
Personnel Division, NUAA
Address: 29# Yudao St. Nanjing, Jiangsu Province, Postcode: 210016
Tel: +86-25-84892461
Fax: +86-25-84895923
Email: zhaohaiyan@nuaa.edu.cn
Web:http://www.nuaa.edu.cn/nuaanew http://rsc.nuaa.edu.cn

南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

Recruitments of talents abroad by Nanjing University of Science and Technology

http://rczp.njust.edu.cn/urp-portal/portal/group/Recruit

Nanjing University of Science and Technology (NUST) is one of the first national "211 Project" universities affiliated with the Ministry of Industry and Information Technology. It has become a multi-disciplinary and coordinated developing engineering-based university along with science, liberal arts, economics, management, law, education, etc. It is an ideal place for research work for its strong scientific research ability, prominent advantages, perfect construction of infrastructural facilities and it's also a pleasant place to live in since the beautiful scenery of Dr. Sun Yat-sen's Mausoleum which is only less than one mile away.

Please refer to the following application guidelines and we welcome your applications.

Majors for recruitment:

Related disciplines of Ordnance Science and Technology, Mechanical Engineering, Instrumentation Science and Technology, Chemical Engineering and Technology, Chemical, Electronic Science and Technology, Information and Communication Engineering, Optics, Optical Engineering, Computer Science and Technology, Control Science and Technology, Electrical Engineering, Transportation Engineering, Aerospace Science and Technology, Power Engineering and Thermal Physics, New Energy, Mechanics, Mathematics, Physics, Civil Engineering, Materials Science and Engineering, Environmental Science and Engineering, Biomedical Engineering, Law, Intellectual Property, Public Administration, Sociology, Economics, Management Science, (Applied) Linguistics and Literature, Art and Design.

Position and requirements:

Recruiting position: "Zijin scholars" distinguished professor, "Young Talents Professors", Professor, Associate Professor, Assistant Professor.

Basic qualities: Overseas talents with a doctorate degree, passion for education, high academic achievement and strong research capability and, good professionalism, academic character and team spirit as well. Among them, the candidates for "Young Talent Professor" staff position should be under 35 years old.

Related treatment:

The full-time employed teachers will be directly categorized into national institution, enjoying free medical care, pensions and other state welfare, and the family issues such as children's nursery and schooling will be addressed. We provide the high-level personnel and young talents with transitional housing, financial relief, research funding and other supports with generous salaries, comfortable working and living conditions. For high-level talents, we help settle down in terms of team building, work, housing, etc. Specific treatment will be determined by personal discussion.

1. For the full-time "Thousands of Plans" and other leading talents, NUST will provide 600 thousand to 1 million RMB annual salary and no less than 8 million RMB research start-up funds.
2. For the "Thousands Youth Talents" and other related talents, we provide no less than 300 thousand RMB annual salary, 1 million RMB financial relief and 2-4 million RMB research start-up funds.
3. For the "Zijin scholars" distinguished professor, we provide 500 thousand to 2 million RMB financial relief, 1 to 5 million RMB research funds and stipulated wages and other benefits along with certain amount of professor allowances per year.
4. For the "Young Talented Professors", we provide 300 thousand RMB annual salary, with certain financial relief and start-up funds.
5. For the professor, associate professor, assistant professor, we will provide salary and financial relief which is competitive in the same region, as well as starter home, and provide appropriate amount of research funds.

Way for recruitment:

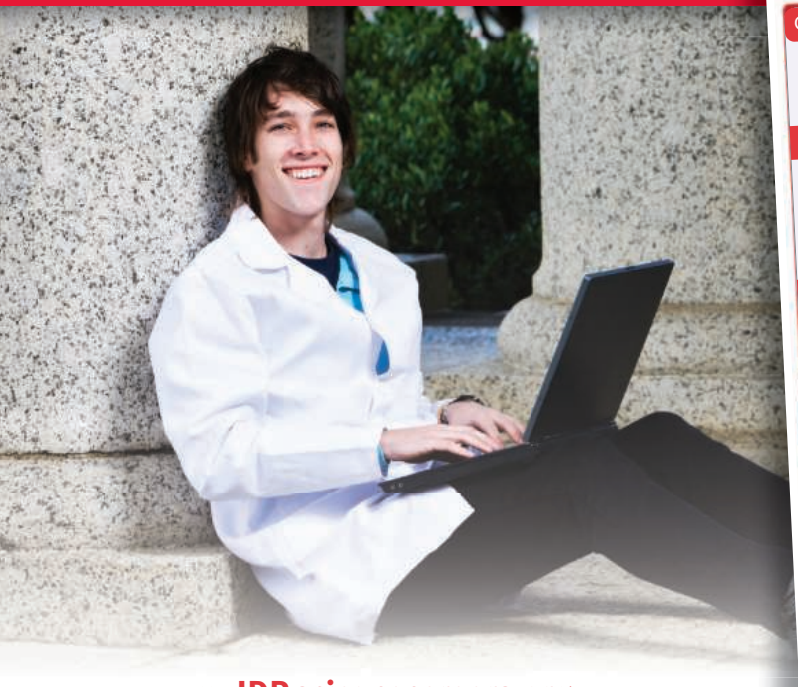
The recruitments of talents abroad by NUST is under way regularly, please log in NUST Recruitment Network http://rczp.njust.edu.cn/urp-portal/portal/group/Recruit to have a registration, or contact us directly.

Contact us:

JI Wenchao Meng Yang
Tel: 86-25-84316943
Mail: rcb@njust.edu.cn
NUST website: www.njust.edu.cn

For your career in science, there's only one **Science**

A career plan customized *for* you, *by* you.



myIDP.sciencecareers.org

Recommended by leading professional societies and endorsed by the National Institutes of Health, an individual development plan will help you prepare for a successful and satisfying scientific career.

my IDP Science Careers

In collaboration with FASEB, UCSF, and the Medical College of Wisconsin and with support from the Burroughs Wellcome Fund, AAAS and *Science* Careers present the first and only online app that helps scientists prepare their very own individual development plan.

Visit the website and start planning today!
myIDP.sciencecareers.org

In partnership with:

advancing the life sciences 1912-2012 FASEB Federation of American Societies for Experimental Biology

UCSF University of California San Francisco

Medical College of Wisconsin

Burroughs Wellcome Fund

UK
UNIVERSITY OF KENTUCKY
College of Medicine

Tenure Track Faculty Position

The University of Kentucky College of Medicine is seeking to recruit a faculty member who will have a joint appointment in the Division of Infectious Diseases in the Dept. of Internal Medicine and the Dept. of Microbiology, Immunology and Molecular Genetics. The successful candidate should have a M.D. or Ph.D. or an equivalent degree, with a strong research program that has active extramural grant support and complements interests in both units. Physician scientists are especially encouraged to apply and will be provided at least 75% protected time for research. The UK College of Medicine, a major regional Medical Center, offers competitive compensation based on rank and experience and an excellent benefits package. Additional information about the position and departments can be found at **http://microbiology.med.uky.edu/MIMG_Position-Open**

The University of Kentucky is located in the heart of the Bluegrass Region. Lexington is one of the top ranked college towns in the country for quality of life (http://livability.com/top-10/top-10-college-towns-2013/lexington/ky). The university is home to an NIH Clinical and Translational Science award and an NCI recognized Markey Cancer Center, and has state of the art core facilities. NIH funding of approximately $50 million to the College of Medicine places it among the top 38 public medical schools nationwide.

Interested persons should send their curriculum vitae, statement of research interests and the names of three references to:

Chair, Search Committee
Dept. of Microbiology, Immunology and Molecular Genetics
The University of Kentucky College of Medicine
Lexington, KY
by email to **kfres1@email.uky.edu**

We will begin to consider applicants on **May 15, 2014**, and will continue the search until an appropriate candidate is identified.

The University of Kentucky is an Equal Opportunity/Affirmative Action Employer. We encourage applications from individuals with diverse experiences and backgrounds.

Institut Pasteur

RESEARCH GROUP IN EPIGENETICS

The Institut Pasteur in Paris announces an international call for outstanding mid-career candidates to establish a research group in epigenetics. Successful applicant will be integrated into the cutting edge interdisciplinary environment offered by an internationally renowned institute combining fundamental and translational research, in an attractive location in central Paris, in close proximity to other major research centers. Candidates specializing in the field of DNA methylation, chromatin modifications, chromatin remodelling, non-coding RNAs, long-range interactions or inheritance of epigenetic traits are encouraged to apply.

Applications will be evaluated on the basis of scientific excellence. A highly attractive package to match the experience of the successful applicant will be provided.

Applicants should send their formal application by e-mail to the Director of Scientific Evaluation, Prof. Alain Israël, at Institut Pasteur (**epigen@pasteur.fr) by July 15, 2014**. Short-listed candidates will be invited for interview in September 2014. Further information on the Institute and on-campus facilities can be found at **http://www.pasteur.fr**

The application should comprise the following in a single PDF file:

1. A brief introductory letter
2. A Curriculum Vitae, a list of 10 selected publications and a full publication list
3. A description of past and present research activities (up to 3 pages)
4. The proposed research project (up to 6 pages, including a summary)

Applications from outstanding junior scientists are also welcome. In this case, please provide name and e-mail address of at least three referees.

AAAS is here –

helping scientists achieve career success.

Every month, over 400,000 students and scientists visit ScienceCareers.org in search of the information, advice, and opportunities they need to take the next step in their careers.

A complete career resource, free to the public, *Science* Careers offers hundreds of career development articles, webinars and downloadable booklets filled with practical advice, a community forum providing answers to career questions, and thousands of job listings in academia, government, and industry. As a AAAS member, your dues help AAAS make this service freely available to the scientific community. If you're not a member, join us. Together we can make a difference.

To learn more, visit aaas.org/plusyou/sciencecareers

AAAS + U = Δ

PENNSTATE HERSHEY
College of Medicine

TENURE TRACK FACULTY POSITION
Pediatric Hematology/Oncology
Penn State Hershey Children's Hospital
Pennsylvania State University College of Medicine

The Division of Pediatric Hematology/Oncology at the Penn State Hershey Children's Hospital invites applications from outstanding early-stage investigators for a Tenure-Track position in the Pediatric Molecular Oncology Program at the rank of Assistant Professor. The mission of the Pediatric Molecular Oncology Program supported by the Four Diamonds Fund is to understand pediatric cancer at the molecular level and to develop novel therapeutic approaches. The Four Diamonds Pediatric Cancer Research Program includes research in transcriptional and epigenetic changes in leukemias, solid tumors, and autophagy. Successful applicants are expected to have a Ph.D., M.D., and/or equivalent degree, a strong publication record, and a future research plan that will integrate into the dynamic and growing research program for pediatric cancer. A competitive start-up package, newly renovated laboratory space, and strong core facilities, together provide an outstanding research environment.

The Division of Pediatric Hematology/Oncology has 12 full-time Pediatric Hematologists/Oncologists and 6 Ph.D. faculty. Over 100 new oncology patients are seen annually, and the division is an active member of COG and POETIC. Research is supported by the Penn State Cancer Institute, the Penn State Hershey Institute for Personalized Medicine, an NIH-funded CTSA, and Pediatric Clinical Trials Office. Interested applicants should submit a single merged pdf file containing a cover letter, curriculum vitae, and statement of research plans to: **Hong-Gang Wang, Ph.D., (huw11@psu.edu), Director of the Pediatric Molecular Oncology Program and Barbara Miller, M.D. (bmiller3@psu.edu), Chief, Division of Pediatric Hematology/Oncology, Penn State Children's Hospital, the Pennsylvania State University College of Medicine, Box 850, MC H085, Hershey, Pennsylvania 17033**. Applications accepted until position filled.

Penn State Milton S. Hershey Medical Center and the College of Medicine are Equal Opportunity/Affirmative Action employers and encourage applications from women and members of minority groups.

POSITIONS OPEN

NYU School of Medicine
NYU LANGONE MEDICAL CENTER

THE COLTON PROFESSOR of Autoimmunity
Endowed Professorship
NYU Division of Rheumatology

The NYU School of Medicine has established an endowed Professorship, the Colton Professor of Autoimmunity, in the Division of Rheumatology and a newly created Colton Center for Autoimmunity. The Colton Professor will be a leader in the field of autoimmunity who will work with a team of scientists across the disciplines of immunology, microbiology, genetics, and bioinformatics to advance translational science in the field of rheumatic diseases. The candidate should be a recognized leader in the field and have a strong track record of NIH funding. To apply or for additional information, contact: **Jill P. Buyon,** M.D., Professor Director, Division of Rheumatology NYU School of Medicine, Department of Medicine **e-mail: jill.buyon@nyumc.org, telephone: 212-598-6110.**

TENURE AND TENURE-TRACK FACULTY POSITIONS

The School of Medicine at the University of Louisville invites applications for faculty position(s) in the Department of Microbiology and Immunology at the **ASSISTANT, ASSOCIATE,** or **FULL PROFESSOR** level. Scientists interested in all areas of microbiology or immunology are encouraged to apply. Preferential consideration will be given to applicants with a research program that emphasizes the study of host-microbe interactions including studies on microbiota and pathogenic microorganisms. For more details on these positions as well as faculty and research programs, please visit **website: http://louisville.edu/medicine/departments/microbiology.** Applicants should submit: (1) a cover letter, (2) a summary of present and future research plans, (3) curriculum vitae, and (4) the names/e-mail addresses of three references at **website: https://highereddecisions.com/uofl/emp_apply_login.asp** (**Job Id: UL157**).

☑ More scientists agree—we are the most useful website.

Science Careers
From the journal *Science* AAAS

www.ScienceCareers.org

Download your free copy today.

ScienceCareers.org/booklets

From technology specialists to patent attorneys to policy advisers, learn more about the types of careers that scientists can pursue and the skills needed in order to succeed in nonresearch careers.

Science Careers
From the journal *Science* AAAS

Women in Science Booklet

Science and the L'Oréal Foundation present

Read inspiring profiles of women making a difference in biology.

Free download at ScienceCareers.org/LOrealWIS