Mexico City favors a scientist for mayor

New books to beat the summer heat

Carbon-based molecules on Mars

Science

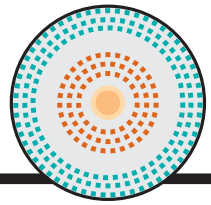
$15
8 JUNE 2018
sciencemag.org

AAAS

GREAT APE GENOMES

Deciphering humanity from our closest relatives

CONTENTS



1075
Rethinking models of embryogenesis

8 JUNE 2018 • VOLUME 360 • ISSUE 6393

1058
Welcome back, beavers.

NEWS

IN BRIEF

1050 News at a glance

IN DEPTH

1052 SHE STUDIED MEXICO CITY. CAN SHE LEAD IT, TOO?
Environmental engineer Claudia Sheinbaum Pardo may become the megacity's next mayor
By L. Wade

1053 CHINA TAKES NEW STEPS TO LURE SCIENCE TALENT FROM ABROAD
Generous funding and the chance to lead projects are helping attract foreign scientists *By D. Normile*

1054 NASA CURIOSITY ROVER HITS ORGANIC PAY DIRT ON MARS
Carbon molecules in rocks from ancient lakebed resemble kerogen, a "goopy" fossil fuel building block on Earth *By P. Voosen*
▸ PERSPECTIVE P. 1068; REPORTS PP. 1093 & 1096

1055 NO BIAS FOUND IN NIH REVIEWS
But race or gender could still influence funding
By J. Kaiser

1056 DIG SEEKS SITE OF FIRST ENGLISH SETTLEMENT IN THE NEW WORLD
Researchers restart their quest for the Lost Colonists' town *By A. Lawler*

1057 MIDDLEWEIGHT BLACK HOLES FOUND AT LAST
Discovery could explain the origin of million-solar-mass monsters at galactic centers *By D. Clery*
▸ PODCAST

FEATURE

1058 BEAVERS, REBOOTED
Artificial beaver dams are a hot restoration strategy, but the projects aren't always welcome *By B. Goldfarb*
▸ VIDEO

INSIGHTS

BOOKS *ET AL.*

1062 SUMMER READING, SCIENCE STYLE
▸ PODCAST

PERSPECTIVES

1068 ORGANIC MOLECULES ON MARS
Data from the Curiosity rover provide evidence for organic molecules in ancient martian rocks and in the atmosphere
By I. L. ten Kate
▸ NEWS STORY P. 1054; REPORTS PP. 1093 & 1096

1069 HONEY BEES ZERO IN ON THE EMPTY SET
Honey bees join a select number of animals shown to understand the concept of nothing *By A. Nieder*
▸ REPORT P. 1124

1070 TARGETING MICROBIAL PATHOGENS
Plants secrete extracellular vesicles to prevent fungal infection
By B. P. H. J. Thomma and D. E. Cook
▸ REPORT P. 1126

1072 WHAT PRECIPITATION IS EXTREME?
How extreme precipitation is defined affects the conclusions drawn about the way it changes with warming
By A. G. Pendergrass

1073 THE EXPANDING MATERIALS MULTIVERSE
Heat capacity and Raman experiments point to fractionalized excitations in a dipole liquid *By B. J. Powell*
▸ REPORT P. 1101

1075 EXPLORING EARLY HUMAN EMBRYO DEVELOPMENT
Stem cell–derived models enable understanding of human embryogenesis
By J. Rossant and P. P. L. Tam

1077 STANLEY FALKOW (1934–2018)
Founder of molecular pathogenesis who discovered how antibiotic resistance spreads *By D. Monack and E. Strauss*

POLICY FORUM

1078 GENEALOGY DATABASES AND THE FUTURE OF CRIMINAL INVESTIGATION
The police can access your online family-tree research—and use it to investigate your relatives *By N. Ram* et al.

1078

IMAGES: (TOP TO BOTTOM) N. DESAI/*SCIENCE*; SARAH KOENIGSBERG, COURTESY OF THE BEAVER BELIEVERS; LUSH/123RF

Published by AAAS

CONTENTS

1069 & 1124
Bee math

8 JUNE 2018 • VOLUME 360 • ISSUE 6393

LETTERS

1080 GREENLAND SLED DOGS AT RISK OF EXTINCTION
By C. Sonne et al.

1080 THE MISUNDERSTOOD SIXTH MASS EXTINCTION
By G. Ceballos and P. R. Ehrlich

1081 INJUSTICES OF FOREIGN INVESTMENT IN COAL
By S. Bashir et al.

IMAGES: (TOP TO BOTTOM) A. KITTERMAN/*SCIENCE*; CARSTEN EGEVANG

RESEARCH

IN BRIEF

1082 From *Science* and other journals

RESEARCH ARTICLES

1085 GREAT APE GENOMICS
High-resolution comparative analysis of great ape genomes *Z. N. Kronenberg* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: dx.doi.org/10.1126/science.aar6343

1086 IMMUNOLOGY
Pulmonary neuroendocrine cells amplify allergic asthma responses *P. Sui* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: dx.doi.org/10.1126/science.aan8546

1087 METABOLISM
A *LIMA1* variant promotes low plasma LDL cholesterol and decreases intestinal cholesterol absorption *Y.-Y. Zhang* et al.

REPORTS

PLANETARY SCIENCE

1093 Background levels of methane in Mars' atmosphere show strong seasonal variations *C. R. Webster* et al.

1096 Organic matter preserved in 3-billion-year-old mudstones at Gale crater, Mars *J. L. Eigenbrode* et al.
▸ NEWS STORY P. 1054; PERSPECTIVE P. 1068

1101 ORGANIC MATERIALS
Evidence for a quantum dipole liquid state in an organic quasi–two-dimensional material *N. Hassan* et al.
▸ PERSPECTIVE P. 1073

1105 APPLIED OPTICS
Imaging-based molecular barcoding with pixelated dielectric metasurfaces *A. Tittl* et al.

1109 PHOTOSYNTHESIS
Structure of the maize photosystem I supercomplex with light-harvesting complexes I and II *X. Pan* et al.

1113 APPLIED OPTICS
A silicon Brillouin laser *N. T. Otterstrom* et al.

1116 SOCIAL SCIENCE
Experimental evidence for tipping points in social convention *D. Centola* et al.

1120 TETRAPOD EVOLUTION
A tetrapod fauna from within the Devonian Antarctic Circle *R. Gess and P. E. Ahlberg*

1124 EVOLUTIONARY COGNITION
Numerical ordering of zero in honey bees *S. R. Howard* et al.
▸ PERSPECTIVE P. 1069

1126 PLANT SCIENCE
Plants send small RNAs in extracellular vesicles to fungal pathogen to silence virulence genes *Q. Cai* et al.
▸ PERSPECTIVE P. 1070

1130 PLANT GENETICS
A selfish genetic element confers non-Mendelian inheritance in rice *X. Yu* et al.

1133 MEDICINE
Noninvasive blood tests for fetal development predict gestational age and preterm delivery *T. T. M. Ngo* et al.
▸ PODCAST

DEPARTMENTS

1049 EDITORIAL
Still not ready for Ebola
By Thomas R. Frieden

1142 WORKING LIFE
The power of persistence
By Asia L. Dowtin and Delphis F. Levia

ON THE COVER

A male Bornean orangutan (*Pongo pygmaeus*). Long-read sequencing and assembly of multiple ape genomes—including those of an orangutan, a gorilla, a chimpanzee, and two humans—provide a more complete view of genetic changes that occurred in each lineage. Human-specific nucleobase deletions and duplications associate with genes differentially expressed during human brain development, in contrast to the expression patterns observed for other apes. See page 1085.
Photo: Michael Leach/Science Source

Science Staff 1048
New Products 1138
Science Careers 1139

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Domestic individual membership, including subscription (12 months): $165 ($74 allocated to subscription). Domestic institutional subscription (51 issues): $1808; Foreign postage extra: Mexico, Caribbean (surface mail) $55; other countries (air assist delivery): $89. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request, GST #125488122. Publications Mail Agreement Number 1069624. **Printed in the U.S.A.**
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. **Postmaster:** Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090–6178. **Single-copy sales:** $15 each plus shipping and handling; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act is granted by AAAS to libraries and others who use Copyright Clearance Center (CCC) Pay-Per-Use services provided that $35.00 per article is paid directly to CCC, 222 Rosewood Drive, Danvers, MA 01923. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

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

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

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

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

Published by AAAS

EDITORIAL

Still not ready for Ebola

Ebola is back. The disease that killed more than 11,000 people in West Africa just a few years ago has returned, striking the Democratic Republic of the Congo (DRC). Last week, intervention with a new vaccine was initiated to help contain the outbreak, adding another tool to a response that is exponentially better than it was 4 years ago. But we are not out of the woods. No matter how long this Ebola outbreak continues, the world faces critical tests in its battle against deadly pathogens.

Unlike 2014, when delayed recognition of Ebola led to undetected disease spread, the DRC promptly admitted it had an outbreak and called for assistance. The World Health Organization (WHO) did not try to minimize the problem. Rapid response units from within the DRC and around the world deployed immediately. And, fortunately, the DRC has well-trained epidemiologists as well as laboratory capacity to test for the virus.

The Ebola vaccine is a very important tool, but it is not a game-changer. Comprehensive detection and response is still necessary to stop this and other outbreaks. And an outbreak in a dense urban setting is far different from a rural outbreak. A single patient can infect dozens of others in taxis, buses, and crowded housing and health-care facilities, making outbreak control more difficult. Nigeria contained Ebola in Lagos, but only with an extraordinary effort that relied on highly experienced polio eradication teams operating within a structured incident-management system.

Globally, we must address three issues to tackle Ebola and other deadly pathogens. One is community engagement. Lack of trust between responders and communities has resulted in patients fleeing isolation, as well as likely missed cases and contacts. Ebola emerged in a remote community; it is essential to understand community perspectives and structure and to gain trust and enlist the community's strengths to stop the disease.

Another issue is WHO's effectiveness. The African Regional Office of WHO now has many staff with the needed technical and operational excellence, and the Geneva-based emergency program is more effective than before. But WHO country offices in DRC and elsewhere are not nearly as effective as they need to be. Tedros Adhanom Ghebreyesus, coming up on his first anniversary as WHO Director General, has unveiled a potentially transformative general program of work. His leadership will be essential for these ambitious goals and inspiring rhetoric to overcome operational and managerial weaknesses at WHO headquarters in Geneva, as well as in some regional and many country offices.

A DRC facility expecting Ebola cases sprays a visitor with chlorine.

"...as long as some countries remain at risk, none of us is safe."

Most important, the entire world needs to support countries, including DRC, that have undergone voluntary external assessments of preparedness, known as Joint External Evaluations (JEEs). The JEE is an objective, transparent, meaningful report card on a country's ability to find, stop, and prevent outbreaks. By the end of 2018, approximately 100 countries will have undergone the rigorous JEE process. Thousands of technical experts—the vast majority coming from within these countries themselves—and billions of dollars are urgently needed to close the thousands of life-threatening gaps identified, and resources that have been committed need to be rapidly and effectively deployed.

The United States, historically a leader on global health security, now risks falling behind in pandemic preparedness. The proposed federal budget would slash U.S. global health security funding by two-thirds and require the Centers for Disease Control and Prevention (CDC) to leave the field open to microbes in dozens of countries. As the latest Ebola outbreak reminds us, if the CDC's funding is not protected, the agency will not be able to help protect us. Because an outbreak can spread from a remote area to any major city in the world in 36 hours or less, we are all at risk. And as long as some countries remain at risk, none of us is safe.

–Thomas R. Frieden

Thomas R. Frieden is president and chief executive officer of Resolve to Save Lives, an initiative housed at Vital Strategies, and is the former director of the U.S. Centers for Disease Control and Prevention. trfrieden@resolvetosavelives.org

PHOTOS: (INSET) KENNY-KATOMBE BUTUNKA/REUTERS/NEWSCOM; (TOP RIGHT) VITAL STRATEGIES

10.1126/science.aau3345

1000

Wild mountain gorillas recorded by a recent census in East Africa. The portion found in the Virunga volcano area—604—more than doubled since the 1980s, despite poaching and other pressures on the critically endangered species.

IN BRIEF

Edited by **Jeffrey Brainard**

CLIMATE SCIENCE

Carbon capture could be cost-effective

A carbon capture plant like the one in this rendering could create feedstock for vehicle fuels.

Carbon dioxide (CO_2) can be sucked out of the atmosphere for less than $100 per ton, a price tag less than one-sixth the cost of previous estimates for the geoengineering technique known as direct air capture, scientists report this week in *Joule*. David Keith, a physicist at Harvard University and co-founder of the startup Carbon Engineering, and his colleagues analyzed the output of a pilot plant in British Columbia in Canada that uses a four-step chemical process to capture CO_2. A liquid takes in the gas initially, then releases it as a concentrated stream of gas that the plant combines with hydrogen to make gasoline and other fuels. Keith's team projected that a large commercial plant using similar technology could capture CO_2 for between $94 and $232 per ton. At that price, Carbon Engineering claims fuels could be made for about $1 per liter. Fuels based on carbon removal could be cost-effective in regions such as California where regulations allow them to command a premium price, says Stephen Pacala, an ecologist at Princeton University who is leading a study about CO_2 removal technologies for the National Academies of Sciences, Engineering, and Medicine.

CRISPR scientist gets his due

MOLECULAR BIOLOGY | Virginijus Šikšnys, whose role in the invention of the revolutionary genome editor CRISPR has often been overlooked, received some vindication on 30 May when he was named a co-winner of the prestigious Kavli Prize in Nanoscience. Šikšnys, a biochemist at Vilnius University's Institute of Biotechnology in Lithuania, will share the $1 million award with two researchers whose roles in creating CRISPR have received far more attention, Jennifer Doudna of the University of California, Berkeley, and her collaborator, Emmanuelle Charpentier of the Max Planck Institute for Infection Biology in Berlin. Šikšnys first showed that the CRISPR system, a bacterial immune mechanism, could be transferred from one bacterium to another. And like Doudna and Charpentier, he independently designed a way to steer the CRISPR complex to specific targets on a genome, which he called "directed DNA surgery."

A big gift for firearm research

PUBLIC HEALTH | A Texas foundation last week gave $20 million to fund research on preventing gun violence, an unprecedented gift in a controversial area. The money from the Houston, Texas–based Laura and John Arnold Foundation will be disbursed over 5 years, through a collaboration overseen by the RAND Corporation, a think tank in Santa Monica, California. U.S. federal agencies have funded relatively few such studies since Congress banned spending that promotes or advocates for gun control. The National Institutes of Health has financed 14 projects worth $6.1 million in 2017 and 2018 that contain the words "firearm" or "gun violence" in their titles or descriptions. The Arnolds were moved to invest in gun research by the 14 February school shooting in Parkland, Florida, which killed 17 people, a foundation spokesperson said. Over the past several decades, other private foundations have contributed millions for research to prevent gun violence—most prominently the Joyce Foundation of Chicago, Illinois, which has spent $28 million since 1993.

IMAGE: CARBON ENGINEERING

Published by AAAS

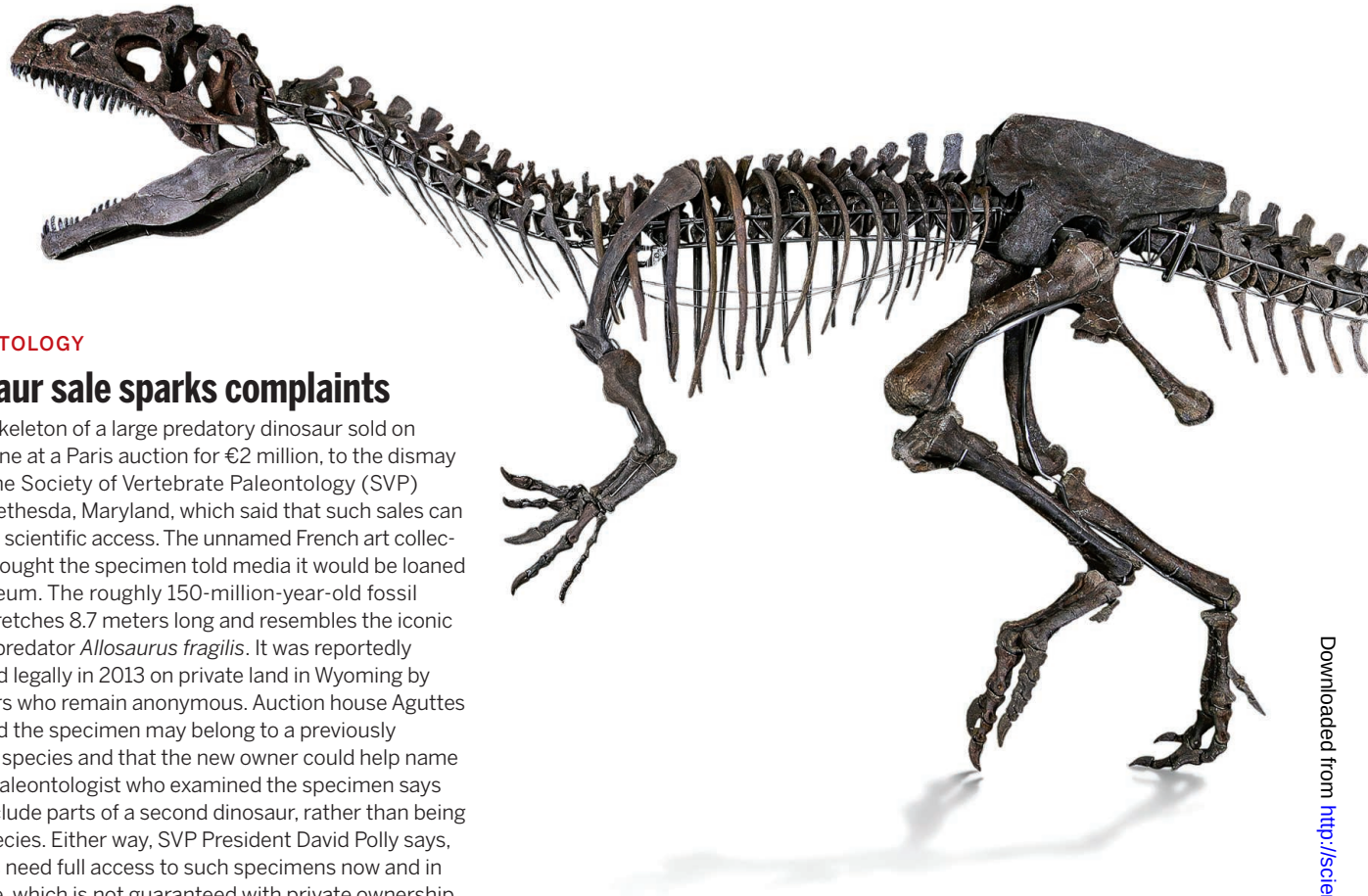


PHOTO: AGUTTES

PALEONTOLOGY

Dinosaur sale sparks complaints

The skeleton of a large predatory dinosaur sold on 4 June at a Paris auction for €2 million, to the dismay of the Society of Vertebrate Paleontology (SVP) in Bethesda, Maryland, which said that such sales can limit scientific access. The unnamed French art collector who bought the specimen told media it would be loaned to a museum. The roughly 150-million-year-old fossil (right) stretches 8.7 meters long and resembles the iconic Jurassic predator *Allosaurus fragilis*. It was reportedly unearthed legally in 2013 on private land in Wyoming by excavators who remain anonymous. Auction house Aguttes suggested the specimen may belong to a previously unknown species and that the new owner could help name it. But a paleontologist who examined the specimen says it may include parts of a second dinosaur, rather than being a new species. Either way, SVP President David Polly says, scientists need full access to such specimens now and in the future, which is not guaranteed with private ownership.

Chile to create science ministry

SCIENCE POLICY | Chile's Congress last week approved the creation of a science ministry. Researchers hope it will give science more prominence and better-coordinated policies. They also hope its budget will match the government's ambitions to "bring Chile towards an information and knowledge society," as Gonzalo Blumel, Chile's minister secretary-general of the presidency, put it in a statement after the vote. So far, the country's science funds have been managed mostly by the National Commission for Scientific and Technological Research, an agency of the Ministry of Education whose budget was about $520 million last year.

Polio resurfaces in Nigeria

PUBLIC HEALTH | Polio is back in Nigeria, after 2 years without a sighting of the virus. Last week, the World Health Organization (WHO) in Geneva, Switzerland, reported that a child had been paralyzed in the northeastern state of Jigawa. This time the culprit is not the wild virus that resurfaced in nearby Borno state in 2016, sparking a huge response, but a weakened type 2 poliovirus used in the live vaccine that mutated and regained its ability to paralyze and spread. Type 2 is the only one of the three polio serotypes that has been eradicated in the wild, so to reduce the risk of these vaccine-derived outbreaks, all countries switched in 2016 to a live vaccine that lacks the type 2 component but works against the other two still circulating. Experts suspect that some use of the older vaccine has continued, perhaps giving rise to the new case. Genetic data show this type 2 poliovirus has been circulating undetected for about a year—evidence that despite Nigeria's ramped-up efforts, vaccination coverage remains low in places and polio surveillance is still weak. The key for Nigeria, says WHO polio spokesperson Oliver Rosenbauer, "is to not take the foot off the accelerator and to keep doing what they are doing but for longer."

NASA seeks cap on big scopes

ASTRONOMY | Since 2016, four teams have been planning designs for a giant NASA orbiting observatory for launch in the 2030s. Last week, the agency told three of the four teams to include at least one design that would cost less than $5 billion—a sign that the overbudget and delayed $8 billion James Webb Space Telescope, due for launch in 2020, has taken a toll on NASA's astrophysics division. The contenders include x-ray and far-infrared observatories, a telescope to characterize exoplanets, and a general-purpose observatory called the Large UV/Optical/IR Surveyor (LUVOIR). The cost cap could prompt some hasty redesigns. But LUVOIR, at least, might get a pass because reducing its size would make it similar to the planned exoplanet telescope.

Google to drop drone contract

MILITARY AFFAIRS | Google told employees on 1 June it will not seek to renew a controversial contract with the U.S. Department of Defense to develop artificial intelligence for analyzing drone footage. The decision, reported by Gizmodo, follows protests and resignations by Google employees who worried that the work would support drone targeting and attacks, although the company had described the work as nonoffensive. Meanwhile, Project Maven, the Pentagon project that issued Google's contract, is expected to continue the work using other contractors.

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

IN DEPTH

SCIENCE AND POLITICS

She studied Mexico City. Can she lead it, too?

Environmental engineer Claudia Sheinbaum Pardo may become the megacity's next mayor

By **Lizzie Wade**, *in Mexico City*

"Clau-dia! Clau-dia! Clau-dia!" The crowd swarms around the smiling woman, chanting her name as she makes her way from her car to the stage at a recent campaign rally in a ramshackle neighborhood of cinder block buildings. Voters jostle to clasp her hands, look in her eyes, and tell her about their troubles.

The enthusiasm still astonishes Claudia Sheinbaum Pardo. "It's like I'm an actress, or someone actually famous!" she says. Until 3 years ago, she worked quietly as an environmental engineer at the National Autonomous University of Mexico (UNAM) here. Now, with a 20-point lead in the polls ahead of the 1 July elections, she seems set to become mayor of this city of nearly 9 million people.

Sheinbaum Pardo considers herself a researcher first and foremost. Her work on energy science and engineering—with a focus on vehicle emissions and climate change mitigation—is respected both in Mexico and abroad, and she's a member of the Mexican Academy of Sciences and a former member of the United Nations Intergovernmental Panel on Climate Change (IPCC). Many say she's uniquely positioned to understand and tackle the myriad problems afflicting this megacity, especially its stuffed-to-the-gills public transportation, epic traffic snarls, and worsening water crisis. "I think she's one of the few people who are aware of the major challenges of the city," says David Bonilla, an economist who studies transportation at UNAM and the University of Oxford in the United Kingdom and has never collaborated with Sheinbaum Pardo. "I can't think of somebody as knowledgeable as her in public policy [in Mexico]."

Critics worry about Sheinbaum Pardo's close ties with Andrés Manuel López Obrador, a leftist populist and the leading candidate in Mexico's presidential election, also taking place on 1 July. She first entered government in 2000, when López Obrador was mayor of Mexico City and appointed her environmental minister. (They met through a family friend, and she shared his progressive politics.) When he founded the National Regeneration Movement in 2014, she followed him to the new party.

Critics liken López Obrador to the late Venezuelan President Hugo Chávez and warn he will weaken Mexico's economy by scaring away investment, but he remains wildly popular here. Sheinbaum Pardo speaks highly of his administrative experience and progressive policies. She acknowledges that voters here "see me as a reflection of him," which partially explains her own popularity.

Yet she has even deeper ties to science. Her mother is a chemist, now emerita at UNAM; her brother, a physicist, helped convince her to study physics as an undergraduate student at UNAM in the 1980s. She completed master's and doctorate degrees in energy engineering, also at UNAM. The combination prepared her well for policymaking, she says. "Training in physics makes you always look for the root causes. Why is something happening? That's fundamental for politics," she says. "And then engineering is much more focused on the 'how.' How can I solve it?"

Sheinbaum Pardo spent 4 years as a Ph.D. student at the Lawrence Berkeley National Laboratory in Berkeley, California, comparing energy consumption in Mexico and other industrialized countries. She "has always approached her academic research work with keen curiosity, intense motivation … and a commitment to use information and analysis to inform evidence-based public policy," says Berkeley Lab energy scientist Lynn Price, with whom she has collaborated. In 1995, Sheinbaum Pardo joined the faculty at UNAM's Institute of Engineering.

After she became Mexico City's environment minister 5 years later, she oversaw two major transportation projects: the introduction of the Metrobus, a rapid transit bus with dedicated lanes; and the construction of the second story of the *Periférico*, Mexico City's ring road. When López Obrador narrowly lost the 2006 presidential election, Sheinbaum Pardo returned to UNAM. She published in top journals, co-authored sections of IPCC's fourth and fifth assessment reports, and doubled down on researching what Adalberto Noyola Robles, a fellow UNAM environmental engineer, calls the "truly unimaginable" number of problems

PHOTO: HAZEL CÁRDENAS

"I just want to make a difference for the city I live in," mayoral candidate Claudia Sheinbaum Pardo says.

in Mexico City, which has a metropolitan population of more than 20 million people. She saw the city's water crisis up close when she re-entered city politics in 2015 as the president of Tlalpan, a southern city district where taps routinely run dry.

Now, Sheinbaum Pardo is making water and mobility centerpieces of her campaign. Mexico City occupies a former lake, drained by the Spanish during the colonial period. Today, urban sprawl has covered almost the entire former lakebed, and most of the city's water is pumped from beneath it. "We've overexploited the aquifer, and as a result, the city is sinking," Sheinbaum Pardo says.

The unstable ground makes earthquakes more dangerous; during the destructive temblor on 19 September 2017, an elementary school collapsed in Tlalpan. Previous administrations have postponed tackling the problem, says Noyola Robles, a water expert. "Claudia understands the issue. I think her proposals will be solid and feasible." She has proposed overhauling the distribution network to fix a plague of leaks, building treatment plants to recycle water, investigating sources of water outside the city, and subsidizing rainwater collection systems.

Mexico City also lags in public transportation. Those who can afford it buy cars; 70% of the city's greenhouse gas emissions come from vehicles, Sheinbaum Pardo says. She proposes investing in bus lines, light-rail trains, and even cable cars, while reducing use of overcrowded informal systems, such as collective vans; she also wants stricter emission standards for cars. Both her transportation and water plans, she says, aim to reduce inequality in access and services.

Sheinbaum Pardo's academic background comes through in her detail-heavy presentations and stump speeches. Competing candidates have called her "arrogant." Besides assailing her close ties to López Obrador, they point to the school collapse and increasingly visible drug crime in southern Mexico City as failings of her administration in Tlalpan. Still, a recent poll found that 40% of those surveyed planned to vote for her; the second and third place candidates didn't crack 20%.

Will Sheinbaum Pardo parlay her likely stint as mayor into a national political career? She won't say, but insists that she would be happy to return to her research at UNAM. She continues to advise a handful of graduate students, squeezing in the work on Sunday afternoons. "I'm not particularly attracted to a political career," she says. "I just want to make a difference for the city I live in." ■

PHOTO: XINHUA/ALAMY STOCK PHOTO

INTERNATIONAL MOBILITY

China takes new steps to lure science talent from abroad

Generous funding and the chance to lead projects are helping attract foreign scientists

By **Dennis Normile**, *in Shanghai, China*

When astronomer Marko Krčo was offered a chance to help commission the world's largest radio telescope, he didn't hesitate. "It's a once in a lifetime opportunity," says Krčo, who has Serbian and U.S. citizenship and earned his Ph.D. from Cornell University. In 2016, Krčo became a postdoc at the Chinese Academy of Sciences's National Astronomical Observatories in Beijing; he spends much of this time in a remote corner of Guizhou province in southwest China, where the Five-hundred-meter Aperture Spherical radio Telescope was completed in 2016 (*Science*, 30 September 2016, p. 1488). "Whether professionally or privately, every day yields a new challenge or a new insight," Krčo says.

The Chinese government, eager to sustain the country's rapid emergence as a scientific superpower, is opening the door wider for people like him. On 22 May, the Ministry of Science and Technology issued guidelines that encourage science ministries and commissions to consult foreign experts and attract non-Chinese to full-time positions within China. In a striking change, foreign scientists are now allowed to lead public research projects.

In the past decade, China has aimed to build up its scientific capacity by luring back some of the tens of thousands of Chinese scientists working abroad. The latest measures emphasize that non-Chinese talent is also welcome. Drafted in December 2017 but not previously made public, they are "a confirmation of things that have been going on for a while," says Denis Simon, an expert on China's science policy at Duke Kunshan University, a Chinese branch campus of the Durham, North Carolina–based Duke University.

Simon says foreign scientists are drawn by China's increased spending on R&D, which is rising twice as fast as its economic growth. Increasingly ambitious big science projects, such as a massive particle accelerator now under study, are a lure as well, says Cao Cong, a science policy specialist at the University of Nottingham Ningbo, a Chinese affiliate of the U.K. university. The opportunity for foreign scientists to serve as principal investigators for publicly funded programs is a significant new incentive, says Liang Zheng, who studies science and technology policy at Tsinghua University in Beijing.

"There is really only one reason why I moved: the money," says 35-year-old U.S. ecologist Luke Gibson, who trans-

New big science projects, like the world's largest radio telescope, make China an attractive destination.

Published by AAAS

ferred from The University of Hong Kong (HKU) to Southern University of Science and Technology (SUSTech) in Shenzhen, a booming city just across the border from Hong Kong, last September. His startup package at SUSTech totals 10 million yuan ($1.6 million), more than 40 times his research support at HKU. "It's rare to find such an enormous level of support," he says. Roughly half comes from the national government's Thousand Talents Plan, aimed at bringing in overseas talent, with matching funding from the Shenzhen government and SUSTech. The support means he can hire four postdocs and extend his ecological studies to the Tibetan Himalayas, the Qinghai-Tibet Plateau, the karst mountain region of south central China, and the Pearl River delta. And whereas Gibson was on a 3-year fixed-term appointment at HKU, at SUSTech he is a tenure track associate professor.

Foreign academics can also join the faculty of one of nine overseas universities that now have Chinese mainland campuses, typically set up with local institutions. Most teach in English, making it easy for non-Chinese academics to feel at home. Duke Kunshan, for example, "recruited faculty from all over the world," says Simon, who is the university's executive vice chancellor. Roughly two-thirds of the more than 40 inaugural faculty members are non-Chinese and, like their counterparts at local universities, they can apply for national and local government research grants.

Relocating to China comes with challenges. Gibson teaches in English, but needs Chinese language help handling administrative matters and grant applications. Restricted access to internet sites such as Google is also a hurdle. "My research and my teaching regularly rely on access to online resources and search platforms [that are] blocked in China, so this is an impediment to my work," Gibson says. But he has found workarounds. China shut down many virtual private networks, which provide access to blocked overseas sites, but a few remain. "There's a saying: 'Everything in China is difficult, but nothing is impossible,' which I think reflects the situation very accurately," Gibson says.

China's push to bring in foreign talent comes at a time when U.S. President Donald Trump is reportedly considering limiting student and academic visas in certain high-tech areas. That would be "one of the most egregious mistakes" the U.S. could make, Simon says. The one-way transfer of knowledge and expertise from the United States to China is a thing of the past, he says: "China increasingly has something to offer us in our own research endeavors." ■

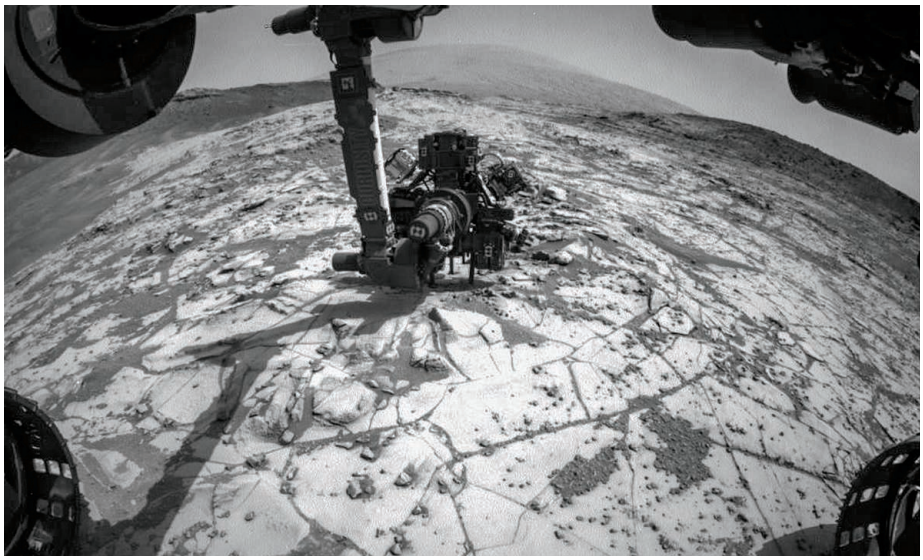


PHOTO: NASA/JPL-CALTECH

PLANETARY SCIENCE

NASA Curiosity rover hits organic pay dirt on Mars

Carbon molecules in rocks from ancient lakebed resemble kerogen, a "goopy" fossil fuel building block on Earth

By **Paul Voosen**

In its quest to find molecules that could point to life on Mars, NASA's Curiosity rover has struck a gusher. Since Curiosity landed in 2012, it has sifted samples of soil and ground-up rock for signs of organic molecules—the complex carbon chains that on Earth form the building blocks of life. Past detections have been so faint that they could be just contamination (*Science*, 27 March 2015, p. 1402). Now, samples taken from two different drill sites on an ancient lakebed have yielded complex organic macromolecules that look strikingly similar to the goopy fossilized building blocks of oil and gas on Earth. At a few dozen parts per million, the detected levels are 100 times higher than previous finds.

Although the team cannot yet say whether these molecules stem from life or a more mundane geological process, they demonstrate that organics can be preserved for billions of years in the harsh martian surface environment, says Jennifer Eigenbrode, a biogeochemist at NASA's Goddard Space Flight Center in Greenbelt, Maryland, who led a study on p. 1096. "We're in a really good position to move forward looking for signs of life."

Ever since it reached its primary target, Aeolis Mons, a 5000-meter-tall mountain rising from the floor of Gale crater, Curiosity has spent much of its time driving on a mudstone formed by sediments that settled to the bottom of a lake some 3 billion years ago, when Mars was a more clement place. Mudstones are ideal for trapping and preserving organic molecules. Because ultraviolet radiation and oxidizing compounds in the martian soil would destroy any compounds exposed at the surface, Curiosity's scientists used a robotic drill to penetrate several centimeters into the mudstone. They delivered the fresh grit to an oven inside the rover's belly.

To unlock organic molecules from the samples, the oven baked them to temperatures of between 600°C and 860°C—the range where a known contaminant disappeared—and fed the resulting fumes to a mass spectrometer, which can identify molecules by weight. The team picked up a welter of closely related organic signals reflecting dozens or hundreds of types of small carbon molecules, probably short rings and strands called aromatics and aliphatics, respectively. Only a few of the organic molecules, sulfur-bearing carbon rings called thiophenes, were abundant enough to be detected directly, Eigenbrode says.

The mass patterns looked like those generated on Earth by kerogen, a goopy fossil fuel building block that is found in rocks

A mudstone drill site called Mojave on Mars was one of two yielding organic molecules resembling kerogen.

such as oil shale—a result the team tested by baking and breaking organic molecules in identical instruments on Earth, at Goddard. Kerogen is sometimes found with sulfur, which helps preserve it across billions of years; the Curiosity scientists think the sulfur compounds in their samples also explain the longevity of the Mars compounds.

Earth's kerogen was formed when geologic forces compressed the ancient remains of algae and similar critters. It's impossible to say whether ancient life explains the martian organics, however. Carbon-rich meteorites contain kerogenlike compounds, and constantly rain down on Mars. Or reactions driven by Mars's ancient volcanoes could have formed the compounds from primordial carbon dioxide. Monica Grady, a planetary scientist at The Open University in Milton Keynes, U.K., believes the compounds somehow formed on Mars because she thinks it's highly unlikely that the rover dug into a site where an ancient meteorite fell. She also notes that the signal was found at the base of an ancient lake, a potential catchment for life's remains. "I suspect it's geological. I hope it's biological," she says.

Curiosity has one last tool to help the team find out: nine small cups containing a solvent that frees organic compounds bonded in rock, eliminating the need to break them apart—and potentially destroy them—at high temperatures. In December 2016, rover scientists were finally prepared to use one of the cups, but just then the mechanism to extend the rover's drill stopped working reliably. The rover began exploring an iron-rich ridge, leaving the mudstone behind. In April, after engineers found a way to fix the drill problem, the team made the rare call to go backward, driving back down the ridge to the mudstone to drill its first sample in a year and half. If the oven and mass spectrometer reveal signs of organics in the sample, the team is likely to use a cup. "It's getting so close I can taste it," says Ashwin Vasavada, Curiosity's project scientist at the Jet Propulsion Laboratory in Pasadena, California.

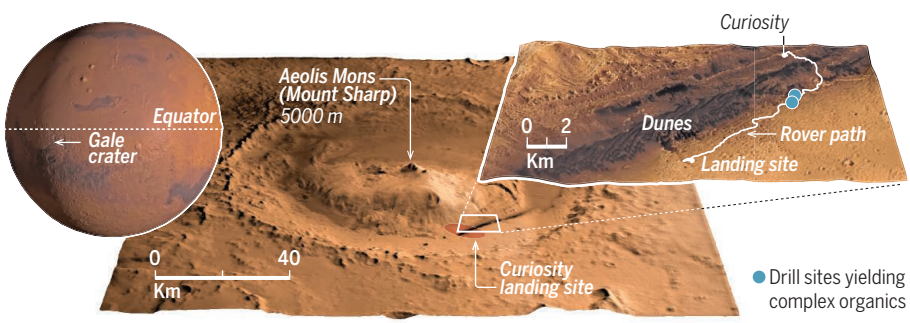
The discovery could give a boost to future Mars exploration plans. Europe's ExoMars rover, due for launch in 2020, will drill deeper than Curiosity, to soil depths better protected from radiation. But detection of past life may ultimately take the precision analysis of labs on Earth, Grady says. "We've got to bring a sample back." In such labs, technicians can dissolve away nonorganic molecules and take a full index of the remaining organic ones, including, say, fatty acids with an even number of carbon atoms—a hallmark of life. Another clue has added to the incentive: In a separate study on p. 1093, Curiosity scientists report that traces of methane in the martian atmosphere rise and fall with the seasons. Nonbiological processes could explain the signal—but so could seasonally varying microbes (*Science*, 5 January, p. 16).

Fortunately, NASA's next rover, Mars 2020, is set to collect some 30 rock cores for return to Earth in subsequent missions. Plans to retrieve those rocks are far from finalized—or financed—but NASA's case has gotten much stronger with the organics discovery, says George Cody, a geochemist at the Carnegie Institution for Science in Washington, D.C. "If somebody asks me to go to Congress and defend a sample return mission, this paper makes that job much easier."

That the rover team found anything at all speaks well to their planning and execution, Cody adds. Imagine drilling in Chile's Atacama Desert, which is often used as a Mars analog. "You'd be damn lucky to detect an ancient kerogen." But the Curiosity team managed it—on Mars. "They lucked out. We lucked out." ■

The path to carbon

Since landing in Gale crater in 2012, the Curiosity rover has reached the base of a 5000-meter-tall mountain. Samples drilled from an ancient lakebed have yielded organic molecules that could stem from biology or geology.



CREDITS: (GRAPHIC) A. CUADRA AND N. DESAI/*SCIENCE*; (DATA) NASA

GRANTMAKING

No bias found in NIH reviews

But race or gender could still influence funding

By **Jocelyn Kaiser**

An investigator's race or gender didn't appear to influence reviewers' scores in an unusual experiment that altered the names on real National Institutes of Health (NIH) grant proposals.

A team led by psychologist Patricia Devine of the University of Wisconsin (UW) in Madison substituted common names of white men, black men, white women, and black women for actual names on 48 proposals for R01 grants—NIH's standard research awards—funded in 2012. When 412 scientists reviewed the grants, they gave similar scores to all four test groups, Devine's team reports in a preprint posted 25 May in PsyArXiv.

The pattern held regardless of the research topic, the proposal quality, and whether the reviewer was a white man. "We just didn't find any evidence that when you randomly assign race and gender names to identical proposals that there was any bias favoring the white male PI [principal investigator]," Devine says. The NIH-funded study is part of an ongoing effort to check for bias after a troubling 2011 study found that blacks are less likely to win NIH grants than whites (*Science,* 19 August 2011, p. 1015).

Devine doesn't think her results rule out bias during the NIH grantmaking process. "I know the problem to be real," she says. "But [this] may not be the place where the bias shows itself." Bias could affect the mentoring that applicants receive in how to write a proposal, she notes, or emerge when reviewers meet to give final scores to each application.

Raynard Kington, a former NIH deputy director and a co-author of the 2011 study finding lower funding rates for blacks, says the UW team's strategy was "reasonable." But changing only the names leaves out other information that could influence a reviewer to discriminate, such as where the investigator was trained, says Kington, now president of Grinnell College in Iowa.

NIH officials declined to comment on the preprint. The agency is also funding a study that stripped all personally identifiable information from 1200 proposals and is recruiting a new set of reviewers to see how their scores compare with the earlier assessment. ■

An English rescue team arrived on Roanoke in 1590, but found only a single word carved in a tree by the abandoned town, as depicted in this 19th century illustration. Archaeologists hope to pinpoint the site of the long-elusive town.

IMAGE: SARIN IMAGES/GRANGER

ARCHAEOLOGY

Dig seeks site of first English settlement in the New World

Researchers restart their quest for the Lost Colonists' town

By **Andrew Lawler**, *on Roanoke Island in North Carolina*

In 1587, more than 100 men, women, and children settled on Roanoke Island in what is now North Carolina. War with Spain prevented speedy resupply of the colony—the first English settlement in the New World, backed by Elizabethan courtier Sir Walter Raleigh. When a rescue mission arrived 3 years later, the town was abandoned and the colonists had vanished.

What is commonly called the Lost Colony has captured the imagination of generations of professional and amateur sleuths, but the colonists' fate is not the only mystery. Despite more than a century of digging, no trace has been found of the colonists' town—only the remains of a small workshop and an earthen fort that may have been built later, according to a study to be published this year. Now, after a long hiatus, archaeologists plan to resume digging this fall. "I firmly believe that our program of re-excavation will provide answers to the vexing questions that past fieldwork has left us," says archaeologist Eric Klingelhofer, vice president for research at the nonprofit First Colony Foundation in Durham, North Carolina.

The first colonists arrived in 1585, when a voyage from England landed more than 100 men here, among them a science team including Joachim Gans, a metallurgist from Prague and the first known practicing Jew in the Americas. According to eyewitness accounts, the colonists built a substantial town on the island's north end. Gans built a small lab where he worked with scientist Thomas Harriot. After the English assassinated a local Native American leader, however, they faced hostility. After less than a year, they abandoned Roanoke and returned to England.

A second wave of colonists, including women and children, arrived in 1587 and rebuilt the decaying settlement. Their governor, artist John White, returned to England for supplies and more settlers, but war with Spain delayed him in England for 3 years. When he returned here in 1590, he found the town deserted.

By the time President James Monroe paid a visit in 1819, all that remained was the outline of an earthen fort, presumed to have been built by the 1585 all-male colony. Digs near the earthwork in the 1890s and 1940s yielded little. The U.S. National Park Service (NPS) subsequently reconstructed the earthen mound, forming the centerpiece of today's Fort Raleigh National Historic Site.

Then in the 1990s, archaeologists led by Ivor Noël Hume of The Colonial Williamsburg Foundation in Virginia uncovered remains of what archaeologists agree was the workshop where Gans tested rocks for precious metals and Harriot studied plants with medicinal properties, such as tobacco. Crucibles and pharmaceutical jars littered the floor, along with bits of brick from a special furnace. The layout closely resembled those in 16th century woodcuts of German alchemical workshops.

In later digs Noël Hume determined that the ditch alongside the earthwork cuts across the workshop—suggesting the fort was built after the lab and possibly wasn't even Elizabethan. NPS refused to publish these controversial results, and Noël Hume died in 2017. But the foundation intends to publish his paper in coming months.

The foundation is also gearing up for a series of new digs. In September, archaeologists will re-excavate parts of the workshop, seeking clues to its size and precise design. In October, foundation and NPS archaeologists will excavate along nearby bluffs that are rapidly eroding. They are applying new dating methods to sand around a post hole near the shoreline. And after a century of work, they know which areas to rule out, such as by the fort, Klingelhofer says. He's confident the extensive new excavations will be more successful, and is eyeing more sites for 2019 digs.

But geologists think the settlement has vanished. Recent studies suggest that shifting currents and rising waters inundated the site in the past couple of centuries, says geologist J. P. Walsh of the University of North Carolina in nearby Wanchese. On a recent research trip into Albemarle Sound off Roanoke to collect cores, he pointed to a depth finder that revealed perilously shallow water. "This was all land back then," he shouted over the engine. He estimates the island's north end has lost about 750 meters in the past 4 centuries, and that strong currents and hurricanes buried any artifacts.

Klingelhofer rejects that idea, saying the loss of land "is more likely to have come since the last ice age" rather than after 1585. Guy Prentice, an archaeologist from NPS's Southeast Archeological Center in Tallahassee, agrees. "If you look at the maps from the 1700s, the island's geography has not changed much. … I just don't buy that a couple of thousand yards are gone." They both note that the Jamestown settlement in Virginia, founded a couple of decades after Roanoke, was long thought to have eroded away. But archaeologists discovered it in the 1990s and have gathered a wealth of artifacts.

All the scientists, however, concur that today's rising seas are swiftly wearing away Roanoke's northern end. Klingelhofer feels urgency to locate the town "before coastal erosion removes all traces." But if history has anything to teach, it is that Roanoke will not readily reveal its secrets. ■

ASTRONOMY

Middleweight black holes found at last

Discovery could explain the origin of million-solar-mass monsters at galactic centers

By **Daniel Clery**

How did giant black holes grow so big? Astronomers have long had evidence of baby black holes with masses of no more than tens of suns, and of million- or billion-solar-mass behemoths lurking at the centers of galaxies. But middle-size ones, with thousands or tens of thousands of solar masses, seemed to be missing. Their absence forced theorists to propose that supermassive black holes didn't grow gradually by slowly consuming matter, but somehow emerged as ready-made giants.

Now, astronomers appear to have located some missing middleweights. An international team has scoured an archive of galaxy spectra and found more than 300 small galaxies that have the signature of intermediate mass black holes (IMBHs) in their cores. The team confirmed that 10 of those candidates really do have middleweight black holes by consulting other data sets, raising their confidence that the original list "must include at least a few dozen genuine [IMBHs]," says team leader Igor Chilingarian of the Smithsonian Astrophysical Observatory in Cambridge, Massachusetts.

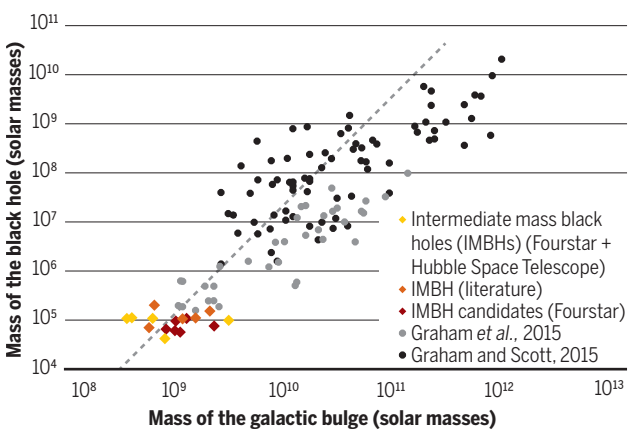
Ezequiel Treister of the Pontifical Catholic University of Chile in Santiago salutes the work. "Black hole measurements are really hard; we've been trying to do this for many years." Volker Bromm of the University of Texas in Austin says the team's technique is "pretty original" and calls their work "careful and responsible." The findings, researchers say, could begin to unravel the mystery of supermassive black hole formation.

Black holes of any size are hard to find because they don't emit light of their own. They can reveal themselves by sucking in nearby gas and dust, heating it so fiercely along the way that it emits x-rays. X-rays pouring from the centers of many galaxies betray the presence of supermassive black holes, known as active galactic nuclei (AGNs). But x-rays from an IMBH would be much fainter, and existing x-ray satellites are geared toward detailed observations of distant sources, not wide-ranging surveys of multiple galaxies. The archive from NASA's Chandra X-ray Observatory, for example, only covers 2.5% of the sky.

So Chilingarian's team searched for an alternative, visible-light signal in a catalog of 930,000 galaxy spectra from the Sloan Digital Sky Survey (SDSS). The x-rays produced by an AGN ionize clouds of hydrogen gas in the galactic bulge around it, setting them aglow at particular frequencies that produce distinct peaks in the galaxy's spectrum. The clouds closest to the black hole swirl around it at high speed, which shifts the frequencies via the Doppler effect and smears out each peak. Gas clouds farther out move more slowly, so the peaks remain sharp. To identify galaxies that have a small AGN, the team looked for spectral peaks that were sharp at the top but smeared out around the base.

Black hole growth chart

Black holes, including the newly discovered middleweights (color), have masses that correlate with the size of their host galaxy.



GRAPH: I. CHILINGARIAN *ET AL.*, ARXIV 1805.01467V1 (2018). ADAPTED BY A. CUADRA/*SCIENCE*

The search of the SDSS catalog yielded 305 candidates. Because other short-lived phenomena could mimic the key signal, the team checked other surveys, gathered at different times, to be sure the candidate galaxies showed the same smeared peaks. They also inspected some of the galaxies with the Giant Magellan Telescope in Chile. But the real clincher of an AGN is the telltale x-ray signal, so the team searched archived observations from NASA's Chandra and Swift satellites and Europe's X-ray Multi-Mirror Mission to see whether, by chance, they had observed any of those candidates. The result was a final short list of 10, the team reports in a paper posted on arXiv and submitted to *The Astrophysical Journal*.

That is enough to challenge recent thinking about how supermassive black holes formed. Theorists needed alternatives to gradual growth not just because IMBHs were missing, but also because astronomers have identified giant quasars—very luminous AGNs—shining when the universe was less than a billion years old. "How could very massive black holes grow so big [so early]?" asks theorist Avi Loeb of Harvard University. He and others proposed that in the early universe, huge gas clouds collapsed directly into black holes of between 100,000 and 1 million solar masses, which formed the seeds of the early quasars. That scenario would explain the rapid formation and the lack of IMBHs.

The new result suggests, however, that at least some giant black holes grew from smaller seeds. The 10 confirmed IMBHs fit a pattern of gradual growth: Their masses correlate with the size of galactic bulge around them, suggesting each is growing in step with its host (see diagram, left). "There is not one scenario: both happen," Treister says. "Now, the question is, which is more common?"

Bromm says direct collapse might only have been possible in the very early universe. The big bang generated only hydrogen and helium; heavier elements, forged by early stars, did not join the mix for millions of years. They helped cool the primordial gas when they became ionized, causing it to glow and shed heat. A cooler gas cloud is more likely to fragment into many stars—the seeds of small black holes. In contrast, a hot primordial cloud could collapse into a single giant object, with help from the gravitational pull of dark matter, Bromm suggests. It would be a rare occurrence, he admits, something of a "cosmic miracle."

The only way to find out which scenario dominates is to find more middleweights. The hurdle, Chilingarian says, is the lack of x-ray surveys. A German-built x-ray survey telescope called eROSITA is expected to be launched later this year or early next year on board the Russian observatory Spektr-RG. "It'll produce a very nice data set," Chilingarian says. He's betting it will yield hundreds more confirmed IMBHs—and shed more light on black hole origins. ■

FEATURES

BEAVERS, REBOOTED

Artificial beaver dams are a hot restoration strategy, but the projects aren't always welcome

By **Ben Goldfarb**, *in the Scott Valley, California*

In 1836, an explorer named Stephen Meek wandered down the piney slopes of Northern California's Klamath Mountains and ended up here, in the finest fur trapping ground he'd ever encountered. This swampy basin would ultimately become known as the Scott Valley, but Meek's men named it Beaver Valley after its most salient resource: the rodents whose dams shaped its ponds, marshes, and meadows. Meek's crew caught 1800 beavers here in 1850 alone, shipping their pelts to Europe to be felted into waterproof hats. More trappers followed, and in 1929 one killed and skinned the valley's last known beaver.

The massacre spelled disaster not only for the beavers, but also for the Scott River's salmon, which once sheltered in beaver-built ponds and channels. As old beaver dams collapsed and washed away, wetlands dried up and streams carved into their beds. Gold mining destroyed more habitat. Today, the Scott resembles a postindustrial sacrifice zone, its once lush floodplain buried under heaps of mine tailings. "This is what we call 'completely hosed,'" sighed Charnna Gilmore, executive director of the Scott River Watershed Council in Etna, California, as she crunched over the rubble on a sweltering June morning last year.

All is not lost, however. Beyond one slag heap, a tributary called Sugar Creek has been transformed into a shimmering pond, broad as several tennis courts and fringed with willow and alder. Gilmore tugged up her shorts and waded into the basin, sandals sinking deep into chocolatey mud. Schools of salmon fry flowed like mercury around her ankles. It was as if she had stepped into a time machine and been transported back to the Scott's fecund past.

This oasis, Gilmore explained, is the fruit of a seemingly quixotic effort to re-beaver Beaver Valley. At the downstream end of the pond stood the structure that made the resurrection possible: a rodent-human collaboration known as a beaver dam analog (BDA). Human hands felled and peeled Douglas fir logs, pounded them upright into the stream bed, and wove a lattice of willow sticks through the posts. A few beavers that had recently returned to the valley promptly took over, gnawing down nearby trees and reinforcing the dam with branches and mud.

"It's fantastic to see beavers working on this," Gilmore said as she bent to examine a chewed stick. "They do a much better job than we do." The result is a bit too orderly to be a beaver dam, a touch too messy to have been created solely by humans.

Gilmore's group is just one of many now deploying BDAs, perhaps the fastest-growing stream restoration technique in the U.S. West. Federal agencies such as the U.S. Forest Service, nonprofits such as The Nature Conservancy, and even private

PHOTOS: (LEFT TO RIGHT) SARAH KOENIGSBERG; COURTESY OF THE BEAVER BELIEVERS

The loss of beavers, as well as damage from overgrazing and other activities, has caused many streams to erode deep into their beds (opposite page). Humans are now trying to reverse the damage, and attract beavers back, by building artificial dams.

ranchers have installed the structures to return life to deeply eroded streams and, in some cases, to help re-establish beavers in long-abandoned territories. In Wyoming, BDAs are creating wet meadows for a vulnerable bird. In Oregon, they're rebuilding salmon streams. In Utah, they're helping irrigate pastures for cattle.

Part of the allure is that BDAs are cheap compared with other restoration techniques. "Instead of spending $1 million per stream mile, maybe you spend $10,000," says Joe Wheaton, a geomorphologist at Utah State University (USU) in Logan who's among the leading proponents of beaver-based restoration. "Relying on the labor of a rodent helps a ton."

The BDA craze is experiencing growing pains, however. Regulators unfamiliar with the approach are sometimes skeptical, and some landowners and government agencies are loath to aid a rodent infamous for felling valuable trees, flooding property, and clogging road culverts. Last year alone, the U.S. Department of Agriculture (USDA) killed more than 23,000 beavers deemed to be nuisances.

Beavers might be vaunted ecosystem architects, says Joe Cannon, an ecologist at The Lands Council in Spokane, Washington, a group that has installed BDAs and relocated beavers in the eastern part of that state. "But we've got greater protection on tree squirrels."

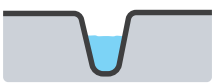
FROM OUR 21ST CENTURY vantage, it's hard to conceive how profoundly beavers shaped the landscape. Indeed, North America might better be termed Beaverland. Surveying the Missouri River Basin in 1805, the explorers Meriwether Lewis and William Clark encountered beaver dams "extending as far up those streams as [we] could discover them." Scientists calculate that up to 250 million beaver ponds once puddled the continent—impounding enough water to submerge Washington, Oregon, and California. *Castor canadensis* even paved the way for agriculture: By trapping sediment in their ponds, beavers "produced the rich farm land … of the northern half of North America," paleontologist Rudolf Ruedemann wrote in *Science* in 1938.

But Beaverland could not withstand the fur trappers who arrived in New England in the 17th century and quickly spread west. By 1843, naturalist John James Audubon found the Missouri Basin "quite destitute." At the outset of the 20th century, researchers estimate, just 100,000 beavers survived—less than 1% of historic numbers.

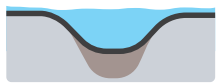
The slaughter transfigured North America's waterways. In a healthy, beaver-rich

A stream comes back to life

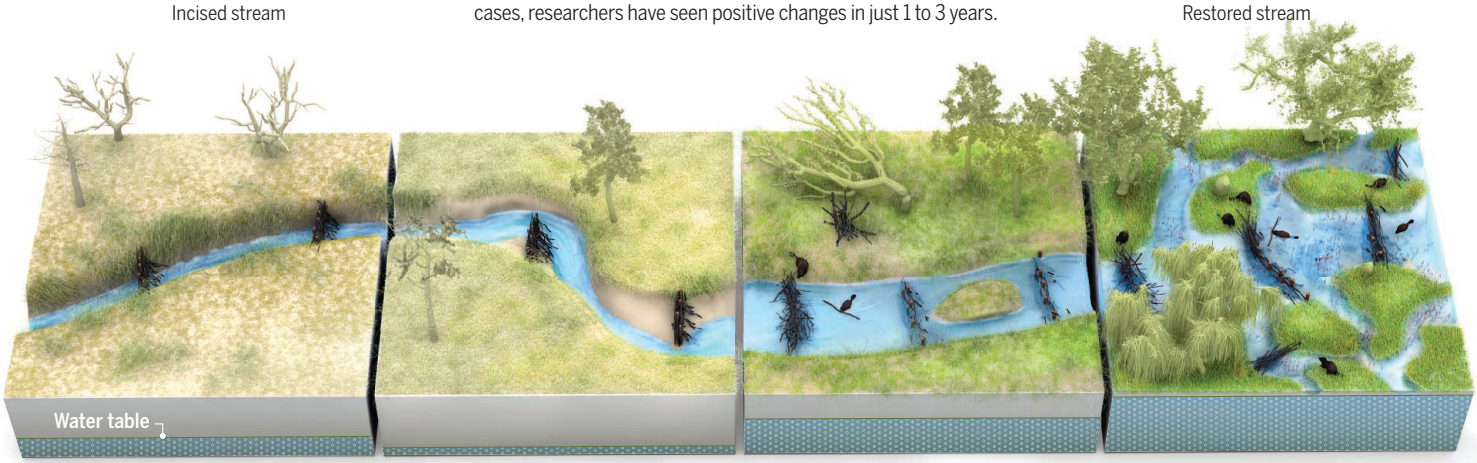
Across the U.S. West, scientists and land managers are using beaver dam analogs (BDAs) to heal damaged streams, re-establish beaver populations, and aid wildlife. In some cases, researchers have seen positive changes in just 1 to 3 years.



Incised stream



Restored stream



Adding dams
Beaver trapping and overgrazing have caused countless creeks to cut deep trenches and water tables to drop, drying floodplains. Installing BDAs can help.

Widening the trench
BDAs divert flows, causing streams to cut into banks, widening the incised channel, and creating a supply of sediment that helps raise the stream bed.

Beavers return
As BDAs trap sediment, the stream bed rebuilds and forces water onto the floodplain, recharging groundwater. Slower flows allow beavers to recolonize.

A complex haven
Re-established beavers raise water tables, irrigate new stands of willow and alder, and create a maze of pools and side channels for fish and wildlife.

creek, dams slow water flows, capture sediment, and counteract erosion. But after beavers and their speed bumps disappeared, streams eroded into their beds, cutting deep gullies in a process called incision. These steep-sided, straitjacketed streams lost the ability to spill onto their floodplains and recharge aquifers. Some groundwater-fed streams dried up altogether.

This tragic history played out along central Oregon's Bridge Creek, a 45-kilometer-long waterway that is the site of the country's most extensive BDA experiment. In the 1820s, U.K. operatives deliberately exterminated the region's beavers to dissuade U.S. trappers from invading the Oregon Territory, which was claimed at the time by both the United Kingdom and the United States. The gambit failed, but the beavers' destruction, combined with unchecked cattle grazing, left an enduring legacy. Bridge Creek devolved into a narrow trench bordered by desiccated pastures. "It was kind of a godforsaken place," says Michael Pollock, an ecosystems analyst with the National Oceanic and Atmospheric Administration's Northwest Fisheries Science Center in Seattle, Washington.

Despite its grim appearance, Bridge Creek wasn't barren. The stream hosted a lingering population of endangered steelhead—rainbow trout that, like salmon, migrate to the ocean and back. A skeleton crew of beavers had also survived, although any dams they built across the sluicelike channel tended to wash away. But Pollock, who had studied the connection between beavers and salmon in Alaska, suspected that, if given a chance, the rodents could capture enough sediment to elevate Bridge Creek's bed, reconnect it with the floodplain, and inundate side channels and backwaters in which juvenile steelhead could thrive. In a 2007 study he found that even relatively short-lived beaver dams trapped significant amounts of sediment.

If a few collapsing dams were good, Pollock figured more stable ones would be better. So he decided to add some beaver-like structures of his own. To many salmon biologists, the experiment seemed the height of insanity: The dams, they warned, would bury key habitat in silt and expose still waters to the sun, making ponds too hot for young fish. "Nobody really understood it," Pollock recalls with a laugh.

The scheme also posed logistical headaches. How could people, wielding tools instead of teeth, mimic nature's most talented builders? When Pollock and a colleague, USU ecologist Nick Bouwes, asked firms for artificial beaver dam designs, the prices came back at $50,000 per structure. "I was appalled," Bouwes remembers. "I'd just gotten done building a log home for that much."

Bouwes combed the internet and found a thriftier alternative: a hydraulic post pounder, a machine that resembles a cross between a jackhammer and a bazooka. In 2009, the pair used their new toy to build 76 BDAs, fashioned from upright posts with willow branches woven between them, on 3.4 kilometers of Bridge Creek. They added 45 more between 2010 and 2012. "My back still hurts," Pollock says.

The crew experimented with size and function. Some BDAs were meant to capture sediment, others to widen the channel by redirecting flows. The overarching goal was to convert a drastically simplified stream into a complex one.

Beavers soon lent a paw. "Wherever we put structures, beavers came and set up shop," recalls Nick Weber, the project's coordinator who is based in Bend, Oregon. By 2013, beavers had fortified nearly 60 BDAs and built 115 new dams, monitoring studies found. All told, Bridge Creek's beaver activity increased eightfold. Some dams captured so much sediment that they became interred in muck. And, like a plant seeking sunlight, the stream bed began climbing out of its trench, spilling water onto floodplains. The creek's submerged area tripled and side channels grew by more than 1200%. "Habitat changes that we thought would take a decade happened in 1 to 3 years," Bouwes says.

Steelhead soon took advantage. Bridge Creek produced nearly three times more fish than a nearby control stream, and its young steelhead were 52% more likely to survive, the researchers reported in 2016 in *Scientific Reports*. Other studies found the dams and ponds actually helped blunt water temperature spikes, perhaps by allowing water to percolate underground and cool.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

As word spread about Bridge Creek, observers realized BDAs might help more than fish. Jeremy Maestas, an ecologist with USDA's Natural Resources Conservation Service in Portland, Oregon, visited the site in 2015 and recognized potential benefits for the greater sage grouse, a ground-nesting bird that is the focus of a major conservation effort. It relies on wetlands and wet meadows for summer forage. Maestas became a BDA evangelist, leading workshops across the western United States. "We're gaining traction all over," he says.

Some ranchers have also embraced beaver-based restoration. Jay Wilde, a rancher based in Mink Creek, Idaho, spent years trying to restore perennial flow to Birch Creek, a seasonal stream on his land. But it wasn't until he hit on beavers that he saw results. In 2015, he invited scientists at USU to build 19 BDAs on the creek and release five beavers nearby; the following summer the stream stayed wet 2 months longer than usual, helping irrigate grazing meadows. Although Wilde, a gruff, tobacco-chewing cattleman, didn't grow up a beaver lover, he's become a staunch advocate, even lecturing at local universities about the project's success. "Now, I'll put in an earring and grow a ponytail if that's what it takes to get the message out."

FOR ALL THAT MOMENTUM, however, BDAs continue to hit snags. Building a structure in a stream typically requires a federal or state permit, but many regulators simply don't know what to make of structures that are neither natural nor entirely human-built. John Coffman, manager of The Nature Conservancy's Red Canyon Ranch near Lander, Wyoming, learned that the hard way when he asked to install 10 BDAs along the Little Popo Agie River in 2017. The project stalled for a year after state officials required him to obtain the legal right to use the water that would be stored behind the BDAs—despite the fact that the semipermeable dams were designed only to delay, not stop, the water from flowing downstream to other users. Although Coffman eventually secured his water rights and built his BDAs, the state forbade structures that exceeded the stream's width or bank height, diminishing their ability to spread water onto the floodplain.

In some places, BDA skepticism has deep historical roots. Some river restoration engineers, for instance, fear the structures are the second coming of the so-called check dams that the U.S. Forest Service once built by the thousands to help curb erosion. Many of the rock dams ended up failing and doing more harm than good by encouraging problematic erosion and littering stream beds with debris.

On Oregon's Bridge Creek, researchers create an artificial beaver dam by pounding logs into the stream bed (top), then weaving a lattice of sticks (bottom).

BDA proponents downplay such concerns. Check dams were intended to be permanent, they note, whereas BDAs are inherently ephemeral. At Bridge Creek, for example, many structures have rapidly fallen into disrepair—which is fine. "It's not about how long the structures last," Wheaton says. "It's about getting beavers back in the system and letting them do the work."

But he concedes that the chaos beavers breed—dams can flood roads, for instance—is not easily reconciled with civilization. In Utah, regulators have denied permits for BDAs amid fears that the structures would alter streams too radically—which, of course, is the whole point. "Beavers seem like a slam dunk," says USU geomorphologist Wally Macfarlane, "but we're getting our shots blocked all the time."

Beaver skepticism has even undermined the pioneering Bridge Creek experiment. In 2017, the federal Bonneville Power Administration pulled its funding from the project after at least one member of the agency's council questioned whether documenting "the value of beavers" was worth the cost of monitoring. The move has cast doubt on the future of the site's research program. "There's so much about this system we're just starting to understand," one researcher laments.

PERHAPS NO PLACE is as ambivalent about beavers as California, where the Department of Fish and Wildlife once claimed—despite ample evidence to the contrary—that the animals were not native to much of the state. Although officials now acknowledge that beavers belong, they've been reluctant to encourage an animal notorious for meddling with the irrigation infrastructure that supports California's agricultural economy. That apprehension has, at times, confounded restoration efforts. In the Scott Valley, for example, the watershed council originally proposed building 36 BDAs, but regulators permitted just six.

Even here, however, the rodent revolution is gaining allies. Last year, state officials showed signs of warming to BDAs after the council invited them to a workshop. And once-suspicious local ranchers have shifted their views, persuaded in part by water tables that have risen by as much as a meter, helping improve water supplies and reduce irrigation costs.

Even 5 years ago, says Gilmore, her colleagues "were like closet beaver people," so fearful of antibeaver sentiment that they wouldn't so much as wear T-shirts decorated with the rodent's portrait. Her group even dubbed BDAs "post-assisted wood structures" to avoid associations with the controversial animal. Today? "We have a lot of landowners that would love for us to put [BDAs] up," she says. "Now, people see me in town and they're like: 'Oh, you're the beaver gal!'" ■

Ben Goldfarb is a journalist and the author of Eager: The Surprising, Secret Life of Beavers and Why They Matter.

PHOTOS: (TOP TO BOTTOM) SARAH KOENIGSBERG; COURTESY OF THE BEAVER BELIEVERS

SUMMER BOOKS

Summer reading, science style

Fearing career repercussions, a theoretical physicist calls out a growing crisis. Setting aside DNA for the filing cabinet, a historian unearths a radically new history of human genetics. Pondering the anatomical correlates of human intelligence, a neurologist searches for Einstein's brain. From an eye-opening tour of bioluminescence to an idiosyncratic history of energy, this year's summer reading picks are chock full of thoughtful research and passionately argued perspectives. Savor a round-the-world search for the origins of wine. Dive below the ocean's waves, where fish—some familiar, some outrageous, and some imaginary—shimmer and shoal. Time travel through the Solar System or tag along on a quest to find long-missing shipwrecks. —**Valerie Thompson**

Energy: A Human History

Reviewed by **Joshua McGuffie**[1]

Motivated by the climate change crisis, Richard Rhodes's *Energy: A Human History* sets out on a historical tour of how humans have manipulated nature to lift, transport, heat, and illuminate things over the past four centuries. Rhodes brings the same storytelling finesse to this work that he brought to his 1986 Pulitzer Prize–winning *The Making of the Atomic Bomb*, as well as the conviction that nuclear power is the solution for moving humankind away from fossil fuels. Accordingly, *Energy* is both a work of history and a passionately written moral tale.

Rhodes excels at exploring lesser-known tales, technologies, and cultural connections that do not often find places in traditional stories about energy, the environment, and climate change. He starts the book's first section, "Power," for example, by explaining that a great shortage of wood in late-16th-century England compelled Shakespeare and his colleagues to steal the materials used to build the Globe Theatre.

From the Bard, Rhodes moves to industrious English landowners who began to exploit exposed coal faces for energy in the early 17th century. Soon, wagonways with wooden rails sprang up to transport the bulky fuel. As miners followed coal seams deep underground, inventors experimented with coal-fired engines to pump water out of flooded shafts. Rhodes spends time with major figures like Thomas Newcomen and James Watt but also considers key tinkerers, such as Richard Trevithick, who helped make steam engines useful for transportation.

In the book's second section, "Light," Rhodes considers an array of 19th-century fuels and technologies that banished nighttime darkness. He turns to William "Uncle Billy" Smith, who engineered the first purpose-drilled oil well in the United States at Oil Creek, Pennsylvania, in 1859. Like coal before it, oil required new techniques for transport. Barrels from breweries and distilleries loaded aboard barges did the job.

Oil took its place alongside coal and, eventually, so did electricity. Westinghouse's station at Niagara Falls began generating hydroelectric power in 1895. Although water produced clean energy, Rhodes concludes his discussion of the 19th century with a dark turn, describing the smoke and caustic pollution that fouled the era's cities.

In the final section, "New fires," Rhodes introduces newer technologies, such as the internal combustion engine. Here, he describes how Thomas Midgley Jr. first added lead to gasoline in 1921. The additive improved engine performance but proved deadly to workers who synthesized it at Standard Oil and DuPont. Another energy source, another pollutant.

Rhodes closes the section with new problems (e.g., photochemical smog and greenhouse gases) and new potential paths forward (e.g., nuclear, wind, and solar technologies). Only nuclear energy, he argues, can provide a workable and sustainable replacement for fossil fuels. Wind and solar energy aren't practical because they simply won't produce enough energy for the increasingly large and wealthy global population. But his cavalier treatment of nuclear disasters and the radioactive waste problem fails to commend the atom as a green energy source.

Nevertheless, Rhodes's hope that a critical look at past energy technologies will benefit those of the future is heartening. May this come to pass.

Energy: A Human History, *Richard Rhodes,* Simon & Schuster, 2018. 479 pp.

Eye of the Shoal

Reviewed by **Taryn S. Murray**[2]

The first time Helen Scales watched fish in the wild, she wasn't expecting to be impressed. Fifteen years old and on a family holiday in California, she was more concerned with spotting a sea otter. Peering from a high bluff south of Monterey Bay, however, she was captivated by what she saw: fish of all shapes and sizes, mesmerizing, complex, and beautiful.

Today an author, marine biologist, and avid scuba diver, Scales's latest book, *Eye of the Shoal*, takes readers on a discovery of the aquatic realm and its incredible ichthyological inhabitants. "How do shoaling fish avoid bumping into each other?" she asks and then sets out to answer. "How do they avoid the jaws of fast-thinking predators? ... How do thousands of fish species get along when they live in crowded places, like the Great Lakes of Africa and the Amazon basin? ... What do fish do when their water dries up?"

We journey from the depths of the oceans to shallow pools in the middle of America's Death Valley to the frozen Antarctic oceans, all the while discussing fishes weird and wonderful. These include mighty whale sharks reaching up to 20 meters in length and tiny tiddlers measuring a mere 8 millimeters; icefish, which make their own antifreeze, and anglerfish that bioluminesce; the recently rediscovered coelacanth and the quickly evolving cichlids of the African Great Lakes.

To understand what it means to be a fish, one has to understand where they came from. In a chapter entitled "A view from the deep – introducing the fish," Scales breaks down the fish evolutionary tree, simplifying a sometimes complex and difficult-to-understand concept in a way that is accessible to the layman, all while maintaining factual integrity.

Scales's genuine appreciation and awe for fish are contagious. She continually entices the reader by introducing exciting aspects of fish in each chapter. "Outrageous acts of colour" discusses the multitude of color adaptations that fishes have evolved, whereas "Illuminations" dives into the depths of the oceans, discussing bioluminescence, a trait among vertebrates that is unique to fish.

Interspersed throughout are traditional tales of fishy folklore from all over the world, highlighting the deep-rooted and sometimes conflicting feelings people have toward these ocean occupants. In 16th-century Iceland, for example, the rare vatnagedda, a flaming golden flounder, was thought to protect against evil spirits and powerful ghosts and could only be caught using gold as bait while wearing a pair of gloves made of human skin. The Inuit people venerate and fear "Sedna," a being with the body of a woman and the tail of a fish. Should the Inuit people need more animals to eat, a shaman must transform himself into a fish and swim down to Sedna so he can comb the tangles from her hair. In return, she releases more animals the people can hunt.

After reading this book, you might be inspired to immerse yourself in the nearest watery realm. While there, you'll likely find that you are well prepared to appreciate the ichthyological inhabitants for what they are—weird, wonderful, and whimsical.

Eye of the Shoal: A Fishwatcher's Guide to Life, the Ocean and Everything, *Helen Scales,* Bloomsbury Sigma, 2018. 320 pp.

[1]*The reviewer is at the Department of History, University of California, Los Angeles, Los Angeles, CA 90095, USA. Email: jmcguffi@ucla.edu* [2]*The reviewer is at the South African Institute for Aquatic Biodiversity, Grahamstown, 6139, South Africa. Email: t.murray@saiab.ac.za*

A diver investigates the remains of a World War II battleship in Palau, Micronesia.

The Shipwreck Hunter

Reviewed by **Alanna Casey**[3]

In his memoir *The Shipwreck Hunter*, David Mearns invites readers to travel along on seven of the most exciting and meaningful investigations of his 21 (and counting) career major shipwreck finds.

In chapters bearing their names, Mearns's thoughtful and detailed account chronologically traces each shipwreck, sharing vivid stories of every vessel from design to demise. The rich historical details and singular characters offer at least one chapter for everyone: from the murderous plot aboard MV *Lucona* to the World War II battles of HMAS *Sydney*; from the haunting rescue efforts that followed the sinking of TSS *Athenia* to the piratical actions aboard *Esmeralda*. Amid these tales, Mearns conveys equal zeal for the competitive bidding process that pushes deep-sea recovery technologies forward and the suspenseful technological glitches that can affect the ability to gather video footage of the wrecks.

Mearns makes the acoustic and robotic technologies used to detect and explore deep-sea wrecks accessible by describing the systems in use. Newcomers and experienced sonar operators alike will feel the frustration of willing the SM-30 sonar to work and the excitement of seeing the outer edge of a debris field, where items—once a part of a vessel and voyage—are today harbingers of a shipwreck location. His sonar and remotely operated vehicle (ROV) demonstrations are interspersed with raw recollections of shipwreck survivors, introspective moments, and personal anecdotes.

A repeated theme in the text is the meticulous work necessary to refine high-probability search areas. Whether decoding secret notations in a German-English dictionary, determining how an oil slick would spread from a sinking vessel, or interviewing scores of witnesses to an offshore submarine attack, the confined bibliography of the memoir belies the depth and breadth of archival, navigational, weather, and witness-testimony research and data revealed in the narrative.

The book links each discovery to the broader importance that investigating shipwrecks has today: prosecuting murderers with MV *Lucona*; improving safety for seamen with MV *Derbyshire*; creating memorials for grieving survivors and families of HMS *Hood*, KTB *Bismarck*, HMAS *Sydney*, HSK *Kormoran*, and AHS *Centaur*; and the documentation of maritime history with Portuguese nau *Esmeralda*.

Mearns closes with the prediction that history will remember the current era as "a golden age of shipwreck hunting," a time when technology has risen to meet ambition and curiosity. Indeed, of the two key wrecks in Mearns's wish list, included in the book's final chapter, USS *Indianapolis* was relocated in August 2017, within months of *The Shipwreck Hunter*'s release in the United Kingdom and Australia, and, beginning in January 2019, the Weddell Sea Expedition will depart for Antarctica to seek Shackleton's *Endurance*.

Excepting the first few pages of the introduction and afterword, which are playfully framed as advice for job seekers, the remainder of the book is equal parts a suspenseful story, a lesson in hard work, and a compelling argument for the modern importance of discovering and documenting shipwrecks.

The Shipwreck Hunter: A Lifetime of Extraordinary Discoveries on the Ocean Floor
David L. Mearns, Pegasus, 2018. 416 pp.

Lost in Math

Reviewed by **Djuna Lize Croon**[4]

Lost in Math is the debut book by Sabine Hossenfelder, a theoretical physicist known to many from her blog, "Backreaction," which is one of the most well-read of its kind by practitioners of theoretical high-energy physics. Hossenfelder has gained some notoriety for her strong opposition to common arguments that physicists make when formulating new theories.

"[Hossenfelder and other bloggers] just spout platitudes without offering any real solutions or making a concrete contribution," complained a fellow researcher on his Facebook page the day I started *Lost in Math*. My colleague, I'm afraid, will not be happy with this book.

Hossenfelder herself seems resigned to a dismal reception, predicting in October 2017, "This isn't a nice book and sadly it's foreseeable most of my colleagues will hate it. By writing it, I waived my hopes of ever getting tenure (*1*)." Although sure to be unpopular, her critical assessment of the field is appropriately timed. In recent decades, high-energy physicists have increasingly relied on theoretical guiding principles to develop new models of nature and to motivate new experiments. But these principles are losing validity, as the Large Hadron Collider has failed to verify many of their predictions. It is a real crisis: Just as the stakes surrounding experimental tests have risen (many experiments have become so costly that they need funding from several governments), our theoretical criteria are starting to fall apart.

Lost in Math paints a very bleak picture of the state of affairs, with Hossenfelder serving as the iconoclast. Our theoretical guiding principles, she insists, are more aesthetic

[3]*The reviewer is at the Department of Marine Affairs, University of Rhode Island, Kingston, RI 02881, USA. Email: alannacasey@uri.edu* [4]*The reviewer is at the Department of Physics and Astronomy, Dartmouth College, Hanover, NH 03755, USA. Email: djuna.lize.croon@dartmouth.edu* [5]*The reviewer is at the University of California Santa Cruz, Santa Cruz, CA 95064, USA, and the Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA. Email: dts@ucsc.edu* [6]*The reviewer is at the Department of History, Princeton University, Princeton, NJ, USA. Email: emilymk@princeton.edu*

than scientific. Although they may have influenced some historical successes, most of these successes were "postdictions" rather than predictions and should therefore not be counted as evidence. The theoretical physics community, she argues, is falling victim to group thinking and cognitive bias.

The book relies heavily on interviews with important stakeholders in the physics community, including several Nobel laureates and other well-known physicists. Hossenfelder interlaces direct quotes from the interviewees with her own interpretations of what they mean (and, often, why she thinks they are wrong). The interviews are the book's main source of nuance, but her heavy-handed contextualization spoils them.

Hossenfelder's book is not the first exposition on the state of theoretical physics for a general audience, but it is more sweeping in its scope. An academic dialogue might have been more appropriate, however; the choice to write for a lay audience is limiting. Although good analogies are found for some technical concepts, most readers will be left with only a low-level understanding of the arguments under discussion. Different concepts are conflated throughout the book (for example, technical naturalness, which has a statistical meaning, and mathematical elegance) and are somewhat mockingly referred to collectively as "beauty." Even with these simplifications, however, the writing level will likely be challenging for nonphysicists.

Of course, all this might be forgiven if Hossenfelder offered a convincing alternative vision for the future of the field. Instead, the book's last chapter includes a half-hearted argument in favor of more collaborations with philosophers (recognizing the philosophical nature of many of our theoretical guiding principles). But the reader is left to imagine what such a synergy would look like and what it would give rise to.

REFERENCE

1. Backreaction blog; http://backreaction.blogspot.com/2017/10/book-update.html.

Lost in Math: How Beauty Leads Physics Astray, *Sabine Hossenfelder,* Basic Books, 2018. 304 pp.

Luminous Creatures

Reviewed by **Darrin Schultz**[5]

Although scholars have been documenting and studying the production of light by living things since the time of the ancient Greeks, the word "bioluminescence" still elicits a sense of mystery and wonderment for the natural world. Despite its title, Michel Anctil's book *Luminous Creatures* does not focus solely on bioluminescent organisms but instead reveals these creatures through colorful stories of the men and women who have studied them.

The book carefully leads the reader back to antiquity and shows how human knowledge of bioluminescence was intertwined with the development of the scientific method and technologies over the millennia. For example, bioluminescent animals and fungi were the test subjects that revealed that oxygen was the essential gas for biological processes in experiments conducted in the 17th and 18th centuries. The invention of the bathysphere, a spherical deep-sea submersible developed in the 1920s, was driven by a researcher's desire to observe bioluminescence in situ. The earliest use of research submersibles and remotely operated underwater vehicles (ROVs) was to observe and collect bioluminescent animals.

John Woodland Hastings reportedly paid children a penny apiece to collect fireflies for his research.

Even more important, *Luminous Creatures* provides an unparalleled holistic narrative of the development of oceanography as a scientific practice, the characters that drove the efforts, and the animals that they studied. Records from oceanographic expeditions in the 19th century are full of stories about living specimens bioluminescing aboard ships, for example, and Anctil explains how these organisms fueled many questions about evolution. "Why is it that both luminous and non-luminous species belong to the same genus?," wondered the French zoologist Henri Gadeau de Kerville, for example, and "Why is it that the number of non-luminous organisms far exceeds the number of luminous organisms and that marine luminous species far outnumber those of terrestrial ones?"

Bioluminescence research expanded after World War II as the field of biochemistry matured. Accordingly, *Luminous Creatures* breaks from a chronological format in its discussion of the 20th century and beyond, relying on character sketches and brief, single-topic passages to explore more recent achievements in bioluminescence.

One weakness of Anctil's account is that it is heavily focused on Western science. True, he devotes a chapter to Yata Haneda, an early-20th-century scientist who is considered the "grandfather" of bioluminescence research in Japan, but the contributions of contemporary Japanese scientists such as Nobuyoshi Ohba, Yoshihiro Ohmiya, Yuichi Oba, and the Nobel-winning Osamu Shimomura are skipped or only briefly mentioned.

In addition, the book omits much of the important research conducted by researchers such as Brazilian biochemist Vadim Viviani and American biochemist Bruce Branchini, who helped to determine the molecular structures of luciferins and the biochemical mechanism of bioluminescence. For a more comprehensive contemporary view of bioluminescence research in the 20th century, readers may wish to check out the more technical *Bioluminescence: Chemical Principles and Methods*.

Despite these shortcomings, *Luminous Creatures* is profoundly well researched, is scientifically accurate (except for the upside-down ctenophore on the cover page), and provides a unique view into the lives of bioluminescence scientists through the ages. Take time, if you can, to step away from the bright laboratory lights and dive into the enchanting world of living light.

Luminous Creatures: The History and Science of Light Production in Living Organisms, *Michel Anctil,* McGill-Queen's University Press, 2018. 488 pp.

Genetics in the Madhouse

Reviewed by **Emily M. Kern**[6]

Decades before Gregor Mendel studied pea plants or Thomas Hunt Morgan cultivated fruitflies, an isolated but vital international community gathered enormous bodies of data on hereditary traits. As Theodore Porter describes in his fascinating and original *Genetics in the Madhouse*, physicians and state officials tasked with overseeing insane asylums throughout the 19th century attempted to understand the origins and nature of madness and, in so doing, laid the foundations for human genetics research today.

PHOTO: PHIL DEGGINGER/ALAMY STOCK PHOTO

Between 1789 and 1900, populations of the insane or "feeble-minded" grew explosively in industrialized Europe and North America. Some 19th-century observers argued that it was a sociological artifact produced by new medical-legal systems and better diagnosis, whereas others put the blame on the social upheaval of industrialization. As patient numbers increased and governments demanded demonstrations of the efficacy of asylum "cures" to justify footing the exponentially growing bills, doctors and administrators turned to new methods of recordkeeping and data organization.

Handwritten narratives of patients' personal histories of madness were captured in account books and then in a succession of preprinted forms, cards, and charts, which placed familial relationships at the forefront of asylum diagnosis. Standard forms simplified data-sharing, allowing administrators to plot correlations between patterns of familial inheritance and specific types of madness or to conduct national censuses of hereditary insanity. By following the technologies of paperwork and data collection, Porter has unearthed a radically new history of human genetics, one that evokes not the double helix but the humble filing cabinet.

The eugenics movement of the early 20th century did not emerge abruptly from an attempt to apply newly rediscovered Mendelian principles to human beings, Porter reveals, but rather as an extension of the legal and scientific technologies that had been practiced in state-run asylums, prisons, and special schools for the better part of a century. Key early figures of eugenics such as Francis Galton, Karl Pearson, and Charles Davenport were not the precursors of a new science, he argues, but the inheritors of long-running medical-statistical tradition. After the Nazis, eugenics may have been repudiated by scientists and state officials, but the practices of pedigree charting and obsessive data-gathering inherited from the 19th century remained a part of the new human genetics and have been carried forward to the present day.

Genetics in the Madhouse is the result of detailed, painstaking work on the data collection practices of many far-flung and forgotten asylum physicians and state statisticians. Yet, sometimes, the reader begins to feel as if they, too, are being sucked into an endless sea of filing cabinets, statistical tables, and handwritten medical reports. However, part of Porter's argument is that the narrative of genetic science has not been clean or straightforward, either in its 19th-century origins or in its contemporary incarnations. Genetics is inescapably intertwined with messy, multivalent subjects like health and race, law and education, poverty and warfare.

As a data science, human heredity has a long history, one that current researchers, physicians, policy-makers, and engaged citizens would be well served to keep in mind.

Genetics in the Madhouse: The Unknown History of Human Heredity, *Theodore M. Porter,* Princeton University Press, 2018. 461 pp.

Tasting the Past

Reviewed by **Larry A. Lerno**[7]

In *Tasting the Past*, journalist Kevin Begos takes readers along on a journey to find the historical origins of wine. During these travels, he introduces researchers in the fields of grape and wine science, presenting the people behind the science as just as important as the science itself. A diverse world of grapes, wine, and winemakers (all of whom have their own stories to tell) complements the scientific story.

A woman harvests Mourvèdre grapes at the Carmel vineyard in southern Israel.

PHOTO: DAVID SILVERMAN/GETTY IMAGES

Begos's search for the origins of wine begins with a chance encounter with a wine made from a little-known grape variety in Amman, Jordan. On a reporting assignment in the Middle East, Begos discovers a bottle of alluring red wine produced by Cremisan Cellars in Bethlehem with "a spicy flavor" and "a hint of earthy terroir." Little did he know that uncorking this bottle was the beginning of what would become a 10-year quest to understand this wine and the people who made it.

Throughout his journey, Begos travels to a number of countries that Western societies do not often associate with wine, including Israel, Georgia, and Cyprus. Here, he finds winemakers using methods of wine production unchanged for hundreds of years. They eschew modern, sleek stainless-steel winemaking equipment for fermentation vessels and presses made of stone, ceramic, or concrete, maintaining a cultural link to the past and preserving native wine.

Begos explores the heritage of modern wine grapes through the work of researchers, including José Vouillamoz and Carole Meredith (Swiss and American geneticists, respectively), who use DNA analysis to determine the genealogical relationships between different varieties. Meredith has shown, for example, that Gouais blanc—a much-maligned grape that winemakers tend to consider a subpar variety—is actually one of the parent varieties of Chardonnay. "We realized that a limited number of varieties are responsible for most of the diversity that we observe today," Vouillamoz tells Begos.

The second part of the book concerns the spread of winemaking. Emphasis is given to native wine grapes and to the struggle winemakers face in keeping old vineyards in production. Winemakers using native grapes often face market pressure to replace native vines with ones having strong global dominance, such as Cabernet Sauvignon, reveals Begos. Often these popular varieties do not grow well in environments where native grapes do. As Olivier Bourdet-Pees, director of the French winery Plaimont, states, "'If I want to drink a wine from Romania, I don't want to drink Merlot. … The climate [there] is not so good for Merlot.'" This frustration is echoed by University of California–Davis researcher Andy Walker: "'… it's a marketing scam that we ended up with ten varieties that are destined to be the best. …'" "All good wine grapes match a particular environmental niche."

Begos provides tasting notes at the ends of the chapters, including wineries and purchasing options for each of the wines discussed. His story unfolds in a manner similar to the growth of ancient grapevines; rooted in a strong central narrative, side stories grow like tendrils, wrapping around and supporting each other, while clusters of vividly described wines emerge like ripe grapes. Anyone who is interested in wine history, viniculture, or just enjoying a glass of wine will likely find *Tasting the Past* a pleasurable read.

Tasting the Past: The Science of Flavor & the Search for the Origins of Wine, *Kevin Begos,* Algonquin Books, 2018. 285 pp.

Catching Stardust

Reviewed by **Charlotte Götz**[8]

Comets, with their glistening dust tails, have fascinated humanity from our earliest days. The comet Halley, for example, is depicted on the 11th-century Bayeux Tapestry, where it

Meteors and other small astronomical bodies offer insight into the origins of the Solar System.

represents a bad omen for an upcoming battle. But should these tiny Solar System objects be feared? Natalie Starkey's answer is a resounding no. In her book *Catching Stardust*, comets—and their rockier counterparts, asteroids—are instead revered because they are key to understanding how the Solar System and its planets formed, evolved, and ultimately led to life on Earth. "Comets and asteroids can be viewed as visitors from a distant place, not only in space but also in time," she writes. "[T]hey bring with them material collected up from the very beginning of the Solar System."

Using her background in geology, Starkey breaks down the categories of small Solar System bodies. Meteorites, she explains, are the debris left over from asteroids after they have traveled through our atmosphere. This makes them the easiest asteroids to study, but easy access comes with a trade-off: Meteorites are heated and deformed during their descent; thus, drawing conclusions about some of an original asteroid's properties (e.g., its "fluffiness"—the density of the comet compared to its volume—or its ice content) is not possible.

Having described the limitations of studying meteorites, Starkey seamlessly transitions to a discussion of the European Rosetta mission, which spent 2 years orbiting the comet 67P/Churyumov-Gerasimenko and was able to capture data about some of these hard-to-measure properties.

Comets, in particular, which have spent most of their existence far from the reaches of the Sun, offer insight into the birth of planets and the shifts in their orbits over time. After the Solar System had outgrown its infancy, comets may have altered Earth in another important way: bringing water or even the building blocks of life to our planet.

Although the book could have benefited from a more detailed look into the results from the astonishing Rosetta mission, the overall story of comets and asteroids is well presented and bridges many gaps between different observational methods.

But why should we care how fluffy cometary dust is or how much metal is in an asteroid? In the first chapter, we learn that comets and asteroids are the perfect "space lab" to study the conditions that prevailed in the solar nebula. "Without the invention of a time-travel machine," she writes, "this is our best chance of understanding a crucial time in our history."

Catching Stardust: Comets, Asteroids and the Birth of the Solar System, *Natalie Starkey,* Bloomsbury Sigma, 2018. 264 pp.

Finding Einstein's Brain

Reviewed by **Katia Andrade**[9]

On the day of Albert Einstein's death, an April morning in 1955, the pathologist Thomas Harvey performed an autopsy and, controversially, took possession of the physicist's brain. Days later, Harvey convinced Einstein's closest relatives of his purpose: to retain the brain for scientific research. Three decades passed, however, until the first work on Einstein's brain was disclosed and, to date, only a few studies, whether histological or anatomical, have been published in peer-reviewed journals.

With intellectual rigor and a quite intimate tone, Frederick E. Lepore meticulously tracks Einstein's brain in space-time for the more than 60 years that have elapsed since the eminent physicist's death. But *Finding Einstein's Brain* is not only the biography of a genius's brain. That specific quest is meshed into an instructive perspective that encompasses several decades of scientific landmarks in physics and neuroscience.

In addition, the book extensively explores recent advances in neurotechnology and the impact of these advances on our understanding of human cognitive and behavioral capabilities. These technologies could shed light on the brain functioning "of a future Einstein," argues Lepore, claiming that "the royal road for studying the next Einstein's brain will be functional neuroimaging and not dissection."

Einstein, few will be surprised to learn, had an "exceptional" brain. In particular, the book presents recent anatomical evidence showing two main findings. First, Einstein's corpus callosum—the white matter bundle that connects the left and right hemispheres—was larger than that of control subjects, which might suggest that "Einstein had greater neural interconnectivity" than an average human. Second, every lobe of Einstein's brain presented differences relative to standard atlases of brain anatomy. For instance, Einstein's right frontal lobe had four gyri, one more than typically found in humans.

Anecdotally, Einstein's brain weighed "a little *less* than expected for a seventy-six-year-old man." This observation is actually unsurprising, given the current agreement that the correlation between our intelligence, however measured, and our brain size (normalized to body size and age) is vanishingly small. In any case, it will never be possible to know what Einstein's brain looked like 50 years earlier, when he wrote the fundamental energy equation.

Lepore discusses the philosophical and medical implications of brain examination, focusing on the brain-mind dilemma. Indeed, although the brain has long been considered the primary organ of the mind, it seems clear that the relationship between the brain and the mind is far from understood.

A professor of neurology and clinical researcher, Lepore has an impressive knowledge of the history of science. With this original book, he gives Einstein's brain a second life and offers the reader a rare opportunity to discover the distinctive features of a genius's brain, while insisting on the explanatory gap that still exists between brain and mind.

Finding Einstein's Brain, *Frederick E. Lepore,* Rutgers University Press, 2018. 315 pp.

10.1126/science.aau1264

[7]*The reviewer is at the Food Safety and Measurement Facility, University of California, Davis, Davis, CA 95616, USA. Email: lalerno@ucdavis.edu* [8]*The reviewer is at the Institut für Geophysik und Extraterrestrische Physik, Technische Universität Braunschweig, 38106 Braunschweig, Germany. Email: c.goetz@tu-bs.de* [9]*The reviewer is at Laboratoire Plasticité du Cerveau, Ecole Supérieure de Physique et Chimie Industrielles (ESPCI Paris); Paris Sciences et Lettres (PSL) University; Laboratoire de Neurosciences Cognitives, Ecole Normale Supérieure (ENS); and Institut de la Mémoire et de la Maladie d'Alzheimer (IM2A), Hôpital de la Pitié-Salpêtrière, Paris, France. Email: katia.santosandrade@gmail.com*

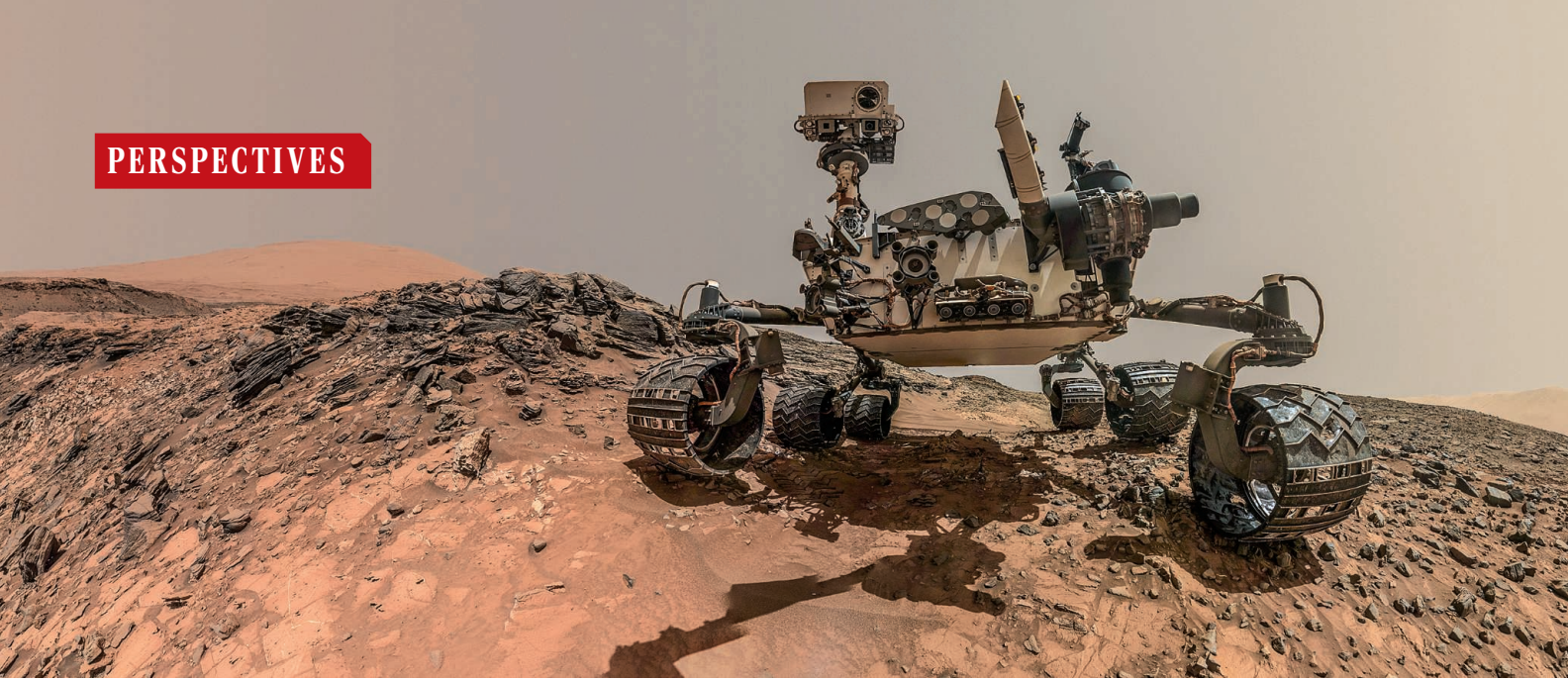


PHOTO: NASA/JPL-CALTECH/MSSS

PLANETARY SCIENCE

Organic molecules on Mars

Data from the Curiosity rover provide evidence for organic molecules in ancient martian rocks and in the atmosphere

By **Inge Loes ten Kate**

On 6 August 2012, the Sample Analysis at Mars (SAM) instrument suite (*1*) arrived on Mars onboard the Curiosity rover. SAM's main aim was to search for organic molecules on the martian surface. On page 1096 of this issue, Eigenbrode *et al.* (*2*) report SAM data that provide conclusive evidence for the presence of organic compounds—thiophenic, aromatic, and aliphatic compounds—in drill samples from Mars' Gale crater. In a related paper on page 1093, Webster *et al.* (*3*) report a strong seasonal variation in atmospheric methane, the simplest organic molecule, in the martian atmosphere. Both these finding are breakthroughs in astrobiology.

To appreciate the importance of these detections, we must go back to NASA's 1976 Viking mission and its search for life on Mars. Viking 1 and 2 were two stationary landers that studied the atmosphere and surface of their local environment with a range of instruments, including a gas chromatograph mass spectrometer (GCMS) dedicated to the detection of organic compounds. However, neither signs of life nor organic compounds were detected in the regolith samples analyzed during this mission (*4*). It is arguable whether not finding signs of life was surprising, but finding no evidence for organic molecules was unexpected. What makes organic compounds so special that we are still searching for them on Mars, more than 40 years later?

Nearly all molecules containing carbon are organic compounds, apart from a few such as CO and CO_2. Many organic molecules are not produced by living organisms. Organic molecules on Mars may have been formed abiotically on the martian surface, delivered from space, or produced by past or present martian life. Space missions to Mars are carefully cleaned to prevent accidental delivery of terrestrial organic molecules to the planet (*5*).

Throughout the Universe, organic compounds are produced abiotically (*6*) and delivered to planetary surfaces through impacts of comets, asteroids, meteorites, and interplanetary dust particles (*7*). They are therefore expected to exist on the martian surface. More speculative is the possibility of past or even present life on Mars. Life on Earth uses and produces four major types of organic compounds: carbohydrates, lipids, proteins, and nucleic acids. Each of these types is constructed from smaller organic molecules, such as sugars, amino acids, and nucleobases. Based on the assumption that hypothetical martian life would not greatly differ from terrestrial life, the search for martian life focuses on these building blocks.

The current influx of abiotically produced organic molecules to the martian surface, estimated from scaling the terrestrial influx to a martian scenario, combined with measurements of the martian atmosphere (*8, 9*), is 100 to 300 metric tons per year. Most of the martian surface is billions of years old (*10*), so organic molecules should be abundant. Why did the Viking landers not detect any? Are organic molecules degraded on the martian surface, particularly by ultraviolet radiation (which has much shorter wavelength on Mars than on Earth) (*11*), ionizing radiation (*12*), or oxidizing compounds (*13*)? All such processes may eradicate organic molecules from the upper few centimeters or even meters of the surface (*12*).

But even if all abiotic organic molecules on the martian surface were degraded, their degradation products should still be detectable. Moreover, minerals such as sulfates and clay minerals that are present on Mars may store organic molecules in their crystal structure, protecting them from the destructive environment (*13*). It thus remained unclear how representative the regolith samples, and the analyses Viking performed on them, were of the organic inventory of Mars.

Clearly, there were still ample reasons to justify a second mission to Mars with an instrument dedicated to the search for organic molecules. SAM was inspired by Viking's GCMS. In the instrument, martian regolith samples are heated so that gases trapped in the samples, organic compounds adsorbed onto the samples, and compounds released by thermal breakdown of minerals are released. The gases are analyzed on a GCMS and a tunable laser spectrometer (*1*).

In 2015, the first analyses of SAM hinted at the presence of organic molecules on

A low-angle self-portrait of NASA's Curiosity Mars rover. SAM is safely hidden inside the rover, ready to analyze when samples are delivered from the top.

Department of Earth Sciences, Utrecht University, Utrecht, Netherlands. Email: i.l.tenkate@uu.nl

Mars (*14*), but those measurements were hampered by the presence of perchlorate salts. These salts, present in martian regolith, break down upon heating within the SAM instruments to temperatures of 200°C. The oxygen and chlorine hereby released react with organic molecules. Leakage of reactive agents presented another challenge. Eigenbrode *et al.* overcame both challenges by only analyzing the gases released above 400°C. They can be certain that these gases are not a result of leaking reagent or reaction with perchlorate. The authors meticulously show all data obtained on Mars by the SAM instrument since its first measurements in 2013 and have thoroughly analyzed all potential contaminants and other signals that might have influenced the actual measurements. They thereby carefully avoid any bias toward hypotheses developed over the past decades. The results convincingly show the long-awaited detection of organic compounds on Mars.

As Webster *et al.* show, methane has also been conclusively detected in the martian atmosphere (*3*). During 5 years of analysis, SAM has found not only a stable methane background, but also local seasonal peaks. It may be that the gas is released from a large subsurface reservoir, but neither the source of that methane nor the driving force of its release is understood. Although many geological processes produce methane, its possible link with biological processes warrants further study to fully understand the martian methane cycle.

The detection of organic molecules and methane on Mars has far-ranging implications in light of potential past life on Mars. Curiosity has shown that Gale crater was habitable around 3.5 billion years ago (*15*), with conditions comparable to those on the early Earth, where life evolved around that time. The question of whether life might have originated or existed on Mars is a lot more opportune now that we know that organic molecules were present on its surface at that time. ■

REFERENCES

1. P. R. Mahaffy *et al.*, *Space Sci. Rev.* **170**, 401 (2012).
2. J. L. Eigenbrode *et al.*, *Science* **360**, 1096 (2018).
3. C. R. Webster *et al.*, *Science* **360**, 1093 (2018).
4. K. Biemann, J. M. Lavoie Jr., *J. Geophys. Res.* **84**, 8385 (1979).
5. I. L. ten Kate *et al.*, *Astrobiology* **8**, 571 (2008).
6. P. Ehrenfreund *et al.*, *Philos. Trans. R. Soc A* **369**, 538 (2011).
7. C. Chyba, C. Sagan, *Nature* **355**, 125 (1992).
8. M. M. J. Crismani *et al.*, *Nat. Geosci.* **10**, 401 (2017).
9. K. Frantseva *et al.*, *Icarus* **309**, 125 (2018).
10. M. H. Carr, J. W. Head III, *Earth Planet. Sci. Lett.* **294**, 185 (2010).
11. I. L. ten Kate, *Astrobiology* **10**, 589 (2010).
12. A. A. Pavlov *et al.*, *Geophys. Res. Lett.* **39**, L13202 (2012).
13. R. G. Keil, L. M. Mayer, in *Treatise on Geochemistry*, H. Holland, K. Turekian, Eds. (Elsevier, ed. 2, 2014), vol. 12, pp. 337–359.
14. C. Freissinet *et al.*, *J. Geophys. Res. Planet* **120**, 495 (2015).
15. J. P. Grotzinger *et al.*, *Science* **343**, 1242777 (2014).

10.1126/science.aat2662

ORGANISMAL BIOLOGY

Honey bees zero in on the empty set

Honey bees join a select number of animals shown to understand the concept of nothing

By **Andreas Nieder**

The number zero is central to contemporary mathematics and to our scientifically and technologically advanced culture (*1*). Yet, it is a difficult number to understand. Children grasp the symbolic number zero long after they start to understand, at around the age of 4 years, that "nothing" can be a numerical quantity—the empty set—that is smaller than one (*2*). Scientists therefore assumed that the concept of "nothing" as a numerical quantity was beyond the reach of any animal. Recent studies on cognitively advanced vertebrates challenge this view, however. Monkeys and birds can not only distinguish numerical quantities (*3*) but also grasp the empty set as the smallest quantity on the mental number line (*4*, *5*). On page 1124 of this issue, Howard *et al.* (*6*) show that the honey bee, a small insect on a branch very remote from humans on the animal tree of life, also belongs to the elite club of animals that comprehend the empty set as the conceptual precursor of the number zero.

Honey bees have a reputation as smart insects. They possess elaborate short-term memory to consider upcoming decisions (*7*), understand abstract concepts such as sameness and difference (*8*), and learn intricate skills from other bees (*9*). Bees can also estimate the number of up to four objects (*10*, *11*). But Howard *et al.* demonstrate even more astonishing number skills in these insects. The researchers report that honey bees can not only rank numerical quantities according to the rules "greater than" and "less than" but they can also extrapolate the less-than rule to place empty sets next to the number one at the lower end of the mental number line.

For their experiments, the authors lured free-flying honey bees from maintained hives to their testing apparatus (see the figure) and marked the insects with color for identification. They rewarded the bees for discriminating displays on a screen that showed different numbers (numerosities) of items. The researchers controlled for systematic changes in the appearance of the numerosity displays that occur when the number of items is changed. They thus ensured that the bees were discriminating between different numbers, rather than responding to low-level visual cues.

First, the researchers trained the bees to rank two numerosity displays at a time. Over the course of training, they changed the numbers presented to encourage rule learning. Bees from one group were rewarded with a sugar solution whenever they flew to the display showing more items, thereby following a greater-than rule. The other group of bees was trained on the less-than rule and rewarded for landing at the display that presented fewer items. The bees learned to master this task with displays consisting of one to four items; they were able to do so not only for familiar numerosity displays but also for new displays.

Next, the researchers occasionally inserted displays containing no item. Would the bees understand that empty displays could be ranked with countable numerosities? Indeed, the bees obeying the less-than rule spontaneously landed on displays showing no item, that is, an empty set (see the figure). In doing so, bees understood that the empty set was numerically smaller than sets of one, two, or more items. Further experiments confirmed that this behavior was related to quantity estimation and not a product of the learning history.

The bees' accuracy in performance improved as the magnitude of difference between two respective numerosities increased. They found it hard to judge whether the empty set was smaller than one but were progressively better when they had to compare two, three, or larger numbers with an empty set. With this behavior, the bees demonstrated the numerical-distance effect with empty sets, a hallmark of number discrimination. The series of experiments (*6*) therefore demonstrates that bees grasp the empty set as a quantitative concept.

The findings are all the more exciting when considering the phylogenetic remoteness of insects and vertebrates. Their last common ancestor, a humble creature that barely had a brain at all, lived more than 600 million years

Animal Physiology Unit, Institute of Neurobiology, University of Tübingen, Auf der Morgenstelle 28, 72076 Tübingen, Germany. Email: andreas.nieder@uni-tuebingen.de

ago (*12*), an eternity in evolutionary terms. As they evolved, separating this time, the building plans of vertebrate and insect bodies developed quite differently and independently from one another. This includes their notably different brains. For instance, a bee's brain has fewer than 1 million neurons, compared to 86,000 million neurons that make up a human brain. But such differences do not prevent bees from knowing how to understand numbers, including zero. It constitutes a fascinating case of convergent evolution of numerical competence.

What may have been the selective pressures that gave rise to numerical competence in such diverse and independently evolved animal groups? Studies examining animals in their ecological environments suggest that numerical competence is beneficial for animals by enhancing their ability to reproduce, navigate, exploit food sources, hunt prey, avoid predation, and engage in social interactions (*3*). Numerical competence is of adaptive value because it helps animals survive and pass on genes to the next generation. This can explain why numerical competence is so widespread in the animal kingdom.

The advanced numerical skills of bees and other animals raise the question of how their brains transform "nothing" into an abstract concept of zero. The neurophysiological basis of number competence is still unknown in insects. Studies on nonhuman primates and corvid songbirds, however, show that "number neurons" in associative endbrain areas give rise to numerical cognition (*13*). Such number neurons respond preferentially to a specific number of elements in a set, irrespective of stimulus appearance. The representation of abstract numerical quantities is demanding, but conceiving of "nothing" as a quantity is even more challenging for the brain. After all, brains have evolved to process stimuli, which represent "something." Without light, a visual neuron does not signal optic information; without sound, an auditory neuron carries no acoustic information.

But this is only part of the story, as "nothing" can be informative. A study in trained monkeys showed that, beyond the sensory input at higher processing stages of the brain, cortical neurons actively represent empty sets as conveying a quantitative null value (*14*, *15*). Such neurons could arise from reinforcement learning if "nothing" constitutes a behaviorally relevant category. It stands to reason that such neurons also emerge in children that learn to understand numerical symbols and the number zero. We have only just begun to zoom in on "nothing" as a relevant quantitative concept for the brain. ■

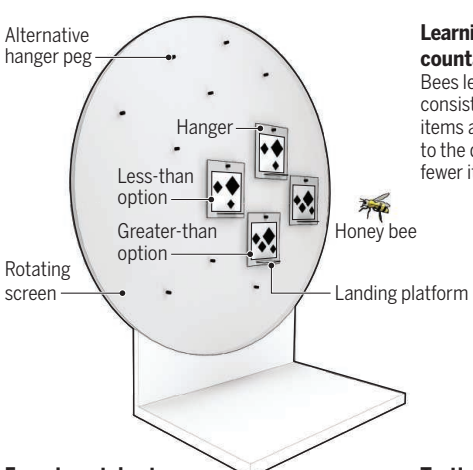
REFERENCES

1. A. Nieder, *Trends Cogn. Sci.* **20**, 830 (2016).
2. D. J. Merritt, E. M. Brannon, *Behav. Proc.* **93**, 91 (2013).
3. A. Nieder, *Philos. Trans. R. Soc. B* **373**, 1740 (2017).
4. I. M. Pepperberg, J. D. Gordon, *J. Comp. Psychol.* **119**, 197 (2005).
5. D. J. Merritt *et al.*, *J. Exp. Psychol. Gen.* **138**, 258 (2009).
6. S. R. Howard *et al.*, *Science* **360**, 1124 (2018).
7. S. W. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 5250 (2005).
8. M. Giurfa *et al.*, *Nature* **410**, 930 (2001).
9. O. J. Loukola *et al.*, *Science* **355**, 833 (2017).
10. M. Dacke, M. V. Srinivasan, *Anim. Cogn.* **11**, 683 (2008).
11. H. J. Gross *et al.*, *PLOS ONE* **4**, e4263 (2009).
12. G. A. Wray, *Philos. Trans. R. Soc. B* **370**, 1684 (2015).
13. A. Nieder, *Nat. Rev. Neurosci.* **17**, 366 (2016).
14. S. Okuyama *et al.*, *Sci. Rep.* **5**, 10059 (2015).
15. A. Ramirez-Cardenas *et al.*, *Curr. Biol.* **26**, 1285 (2016).

10.1126/science.aat8958

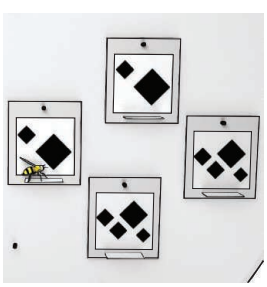
Testing bees on number discrimination

One group of bees was rewarded with sugar solution whenever they flew to the display showing more items (greater-than rule); the other group was rewarded whenever they chose fewer items (less-than rule).

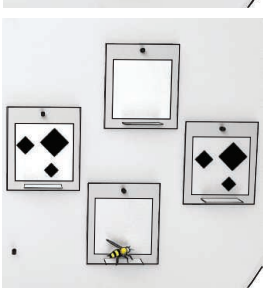


Experimental setup
On the rotating screen, four stimuli (two identical correct stimuli and two identical incorrect stimuli) were presented simultaneously above landing platforms on the hangers.

Learning "less than" with countable numerosities
Bees learn to rank displays consisting of one to four items and always fly to the displays showing fewer items.



Testing with empty sets
Bees obeying the less-than rule spontaneously land on displays showing no items (empty set).



PLANT IMMUNOLOGY

Targeting microbial pathogens

Plants secrete extracellular vesicles to prevent fungal infection

By **Bart P. H. J. Thomma**[1] *and* **David E. Cook**[2]

One of the most important challenges in agricultural production is to safeguard crops from pathogen infection. Uncovering the molecular mechanisms governing plant-microbe interactions can provide new strategies by which to sustainably intensify agriculture and can additionally contribute to our broader understanding of interspecies interactions (*1*). On page 1126 of this issue, Cai *et al.* (*2*) report that plant hosts secrete extracellular vesicles containing small RNA (sRNA), which are taken up by, and lead to silencing of, fungal virulence–related genes during infection (see the figure). These findings address the previously unknown phenomenon by which host sRNA can alter gene expression in other organisms, highlighting the role of extracellular vesicle–mediated transport as a key element of cross-kingdom RNA interference (RNAi). This could be exploited in the development of RNAi-based pathogen control strategies to protect crops.

Since Harold Flor's gene-for-gene theory in the 1940s describing host-pathogen interactions, which postulated that a matching gene pair in the host and pathogen dictate disease development, it is now apparent that hosts and pathogens are involved in an ongoing coevolutionary arms race in which the pathogen aims to continue the symbiosis that the host tries to interrupt (*1*). Considerable attention has been paid to proteinaceous molecules used by both interaction partners, including pathogen-secreted effectors that perturb host immunity and host-secreted antimicrobial proteins. Recently, cross-kingdom RNAi was identified as an additional pathogen-host interaction mechanism in which sRNAs are used in an attempt

[1]*Laboratory of Phytopathology, Wageningen University and Research, Droevendaalsesteeg 1, 6708 PB Wageningen, Netherlands.* [2]*Department of Plant Pathology, Kansas State University, Manhattan, KS 66506, USA. Email: bart.thomma@wur.nl*

GRAPHIC: A. KITTERMAN/*SCIENCE*

to induce gene silencing in the other organism. For instance, the gray mold fungus *Botrytis cinerea*, an aggressive fungal pathogen that infects hundreds of plant hosts, secretes sRNAs that are taken up by host cells and hijack the host RNAi machinery to selectively silence genes involved in immunity (*3*). Conversely, plants can secrete sRNAs that hijack the fungal pathogen RNAi machinery to attenuate virulence mechanisms (*4*). More recently, parasitic plants were shown to use sRNA to facilitate host infection (*5*). Although it remains unclear how sRNAs that target RNAi in another organism evolved, these findings suggest that bidirectional interorganismal, and even cross-kingdom, RNAi is an important determinant for host-pathogen interactions.

Typically, animal cells secrete three types of extracellular vesicles—exosomes, microvesicles, and apoptotic bodies—which can be characterized by their size and origin. Although initially considered waste trafficking vesicles, exosomes are increasingly recognized to function in intercellular communication, development of various diseases including cancers, and immune responses. Studies in mammalian systems have shown that exosomes can carry functional messenger RNAs (mRNAs) as well as noncoding sRNAs that play roles in pathogenic processes (*6*, *7*). Additionally, a gastrointestinal parasitic roundworm in mice was shown to modulate host immunity through RNAi by means of exosome-delivered sRNAs, expanding known exosome function to interorganismal communication as well (*8*).

After the observation in 1967 that carrot cells secrete extracellular vesicles (*9*), research into the functions of such vesicles has developed more slowly in plants than in animal systems. A productive line of research regarding the accumulation of chemical and physical host defenses at the site of fungal invasion has uncovered the requirement of several cell-trafficking components and was speculated to involve exosomes (*10*). Subsequent isolation and characterization of pathogen-induced plant exosomes revealed that they are enriched in proteins with roles in stress responses, including immunity (*11*). Plant extracellular vesicles are also taken up by fungal pathogens, leading to growth inhibition (*12*). Together, these observations suggest that extracellular vesicles play an important role in plant immunity.

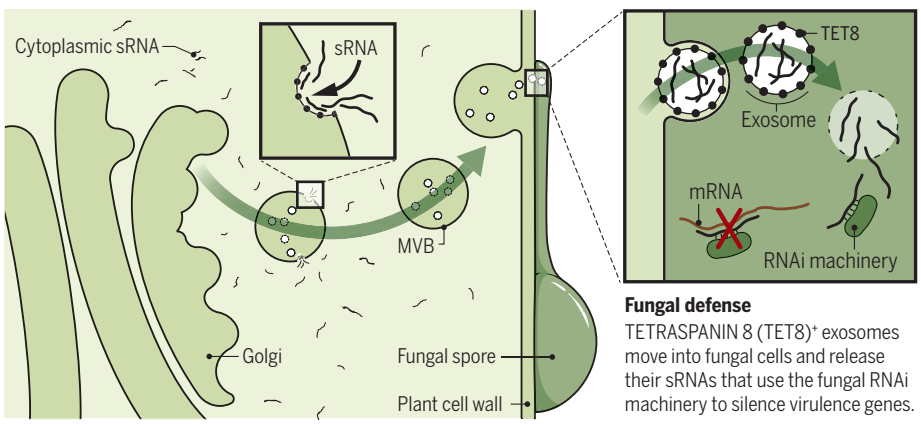
Cai *et al*. now bring together two lines of observation: the importance of RNAi in host-pathogen interactions and that host immunity involves exosomes that contain sRNAs with functional activity toward the pathogen. The authors digested *B. cinerea*–infected host tissues and isolated sRNAs from plant and fungal cells for sequencing. Host-derived sRNAs, contained within exosomes, were readily taken up by fungal cells. Intriguingly, many sRNAs found at low abundance in the plant were among the most abundant in the fungal cells, supporting the idea of a mechanism of selective sRNA transfer. Importantly, several of the transferred sRNAs were implicated in reducing fungal virulence by targeting genes that are normally important for infection. How sRNAs are selectively loaded into exosomes and how exosomes are targeted to, and fuse with, recipient cells are important future questions. It is additionally interesting to consider whether defense proteins and sRNAs are loaded into the same vesicles.

It is conceivable that pathogens similarly use exosomes or other types of extracellular vesicles to attack their host plants. If so, and if these vesicles are readily taken up by plant cells, this could also solve another enigma: It has been widely accepted that pathogens secrete effectors to subvert host defenses, and many of these effectors are presumed to enter host cells, although the mechanism through which uptake is established remains unclear (*13*, *14*). Future research will establish whether pathogens use exosomes not only for the delivery of sRNA effectors but also for the delivery of effectors.

The study of Cai *et al*. offers a powerful perspective on the role of exosomes for interorganismal communication and cross-kingdom RNAi. Establishing whether exosome-mediated delivery is widely used across diverse host-pathogen interactions, during other types of symbioses, and by nonplant hosts will be important. For example, it remains unknown whether similar mechanisms operate in animal immunity. Moreover, it is unclear how hosts direct vesicles to the site of infection once they sense invasion. Importantly, the cargo of host-secreted exosomes may reveal crucial pathogenicity genes and whether this mechanism bears specificity toward particular pathogens or attacks generic pathogen traits. Ultimately, results from such studies may be used for designing improved disease control. Host-induced gene silencing in pathogens has recently been identified as an effective method with which to provide crop protection (*15*). By transgenically expressing pathogen-specific sRNAs in host plants, it could be possible to control disease. However, because the use of genetically modified crops still meets with considerable public concern, it is unlikely that host-induced gene silencing will be widely implemented soon. Recently, it was shown that externally applied sRNAs and double-stranded RNAs are readily taken up by pathogens and can potentially be exploited as "RNA fungicide" in "spray-induced gene silencing," but efficacy in a large field setting remains unknown (*4*). Packaging sRNA into extracellular vesicles may increase the feasibility of developing a robust control strategy against an array of disease-causing agents in the future. ■

Exosome-delivered sRNA to a fungal pathogen

In plant cell defense to a fungal pathogen, exosomes containing sRNA are delivered to invading fungal cells. sRNAs are packed into exosomes that assemble into multivesicular bodies (MVBs) and are secreted.



Fungal defense
TETRASPANIN 8 (TET8)+ exosomes move into fungal cells and release their sRNAs that use the fungal RNAi machinery to silence virulence genes.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

REFERENCES

1. D. E. Cook *et al.*, *Annu. Rev. Phytopathol.* **53**, 541 (2015).
2. Q. Cai *et al.*, *Science* **360**, 1126 (2018).
3. A. Weiberg *et al.*, *Science* **342**, 118 (2013).
4. M. Wang *et al.*, *Nat. Plants* **2**, 16151 (2016).
5. S. Shahid *et al.*, *Nature* **553**, 82 (2018).
6. H. Valadi *et al.*, *Nat. Cell Biol.* **9**, 654 (2007).
7. M. Mittelbrunn *et al.*, *Nat. Commun.* **2**, 282 (2011).
8. A. H. Buck *et al.*, *Nat. Commun.* **5**, 5488 (2014).
9. W. Halperin *et al.*, *J. Ultrastruct. Res.* **18**, 428 (1967).
10. M. E. Nielsen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11443 (2012).
11. B. D. Rutter *et al.*, *Plant Physiol.* **173**, 728 (2017).
12. M. Regente *et al.*, *J. Exp. Bot.* **68**, 5485 (2017).
13. S. D. Kale *et al.*, *Cell* **142**, 284 (2010).
14. S. Wawra *et al.*, *Plant Cell* **29**, 1184 (2017).
15. D. Nowara *et al.*, *Plant Cell* **22**, 3130 (2010).

10.1162/science.aat9343

CLIMATE CHANGE

What precipitation is extreme?

How extreme precipitation is defined affects the conclusions drawn about the way it changes with warming

By **Angeline G. Pendergrass**

A warmer atmosphere has more water vapor. Scientists have been trying to predict what this means for precipitation, but this is more complex and harder to model than temperature. One explanation has been that the intensity of extreme precipitation events will increase at a rate proportional to the increase in atmospheric moisture. But recent findings show that this explanation is too simplistic. There are many ways to define extreme precipitation, and the choice of definition affects how it responds to warming. Researchers must choose their definition of extreme precipitation with care and articulate it clearly, and users should consider how extreme precipitation is defined when interpreting analyses of its change with warming.

One way to quantify extreme precipitation change is to determine how much more intense an event becomes in a warmer climate. The conventional wisdom (*1*, *2*) is that the intensity of the most extreme events rises by 6 to 7% per degree of warming, which is also the rate of moisture increase (*3*). That is, moving toward progressively more extreme precipitation events, the rate of increase of individual events in response to warming approaches the rate of moisture increase. If the change of extreme precipitation intensity did indeed converge to the rate of moisture increase, then the magnitude of the change in extreme precipitation intensity would be independent of how extreme precipitation is defined.

But there are indications that extreme precipitation does not always change at the rate of moisture increase. Take, for example, the record-breaking rainfall associated with Hurricane Harvey, which led to devastating flooding in Houston, Texas, last year. Three groups have published attribution studies of Hurricane Harvey (*4*–*6*) and compared their results against the increase in moisture (*7*). All three studies found that climate change increased the intensity of the precipitation during this very extreme event by more than the increase in moisture. Studies of large-scale changes in precipitation also suggest that not all extreme precipitation events intensify at the rate of moisture increase (*8*–*11*).

At least three mechanisms influence the change in precipitation intensity with warming. First, increasing moisture drives increases in the intensity of all precipitation events, if atmospheric circulation does not change. However, atmospheric circulation affects, and is in turn affected by, precipitation. This leads to two other mechanisms: Atmospheric stability increases with warming, weakening circulation and reducing the intensity of precipitation events. Also, latent heat release strengthens the storms that drive precipitation events in proportion to their intensity, amplifying the most extreme events by the largest amount.

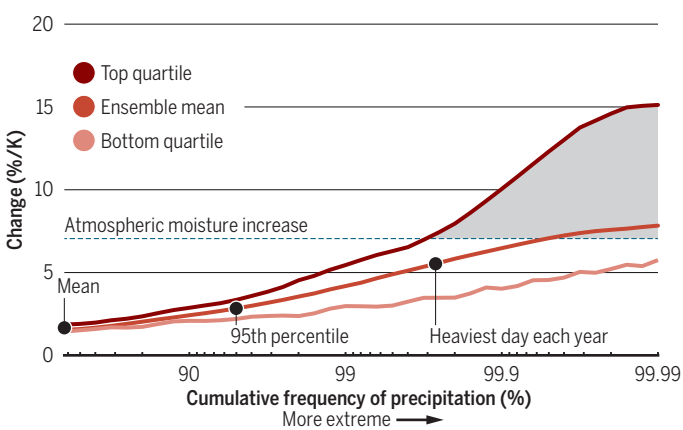
Overall, changes in atmospheric circulation drive larger increases in more extreme precipitation events compared with less extreme ones. The most extreme events could increase at or above the rate of moisture increase. In climate model simulations, these tendencies are seen both in the tropics (*8*) and in mid-latitudes (*12*). The result is that not all extreme events change in the same way: There is no convergence to the rate of atmospheric moistening.

The variation in the magnitude of extreme precipitation change can be illustrated by examining the response of extreme precipitation to warming in climate model simulations (see the figure). We begin with the most extreme precipitation events (see the figure, right) and then expand the definition of "extreme" to include ever more moderate events, until all precipitation events are included. Eventually, the no-longer extreme precipitation will increase at the same rate as total precipitation. But, total precipitation changes more slowly than moisture because of the role of precipitation in the planetary energy budget (*2*). In order to maintain energy balance in the atmosphere, precipitation can increase only as much as radiative and sensible heat exchange with the surface. We should thus expect that for some sufficiently moderate definition of "extreme," extreme precipitation will increase at a rate below that of atmospheric moisture.

More precipitation falls in events that can be considered extreme than is often appreciated. The table lists the fraction of total precipitation falling in events considered extreme by various

Extreme precipitation response to warming

The intensity of more extreme events (which are more intense and less frequent) increases more than the intensity of less extreme events in response to warming. The change in precipitation beyond levels of the cumulative frequency distribution of daily precipitation in the global mean in an ensemble of climate model simulations (*11*) is shown.



Definitions of extreme precipitation

The percentage of total precipitation classified as extreme depends strongly on the definition of extreme precipitation, as shown by this comparison of different definitions for two observing stations between 1978 and 2017. Also shown is the frequency beyond which 50% of precipitation falls. Data are from (*14*).

DEFINITION	SAN DIEGO, CA, US	TROMSØ, NORWAY
1 day	16%	3.0%
99th percentile	42%	9.5%
98.6 percentile	50%	
95th percentile	89%	32%
90.5 percentile		50%
90th percentile	99%	52%
Wet day 95th (≥1 mm/d)	23%	17%

National Center for Atmospheric Research, Boulder, CO 80307, USA.
Email: apgrass@ucar.edu

GRAPHIC: J. YOU/SCIENCE

metrics at two observing stations: San Diego, California, where most precipitation falls in a small number of events, and Tromsø, Norway, where it rains very frequently, and precipitation is thus spread over many events.

One common definition of extreme precipitation is the 95th percentile of the cumulative frequency distribution of daily precipitation (including days with no precipitation)—that is, the top 5% of days in terms of their accumulated precipitation. In San Diego, almost 90% of precipitation falls above the 95th percentile. Even in Tromsø, more than half of precipitation falls in events above the all-day 90th percentile, which is sometimes used as a metric for extreme precipitation. But in San Diego, the degree of extreme beyond which half of all precipitation falls is striking: the 98.6th percentile, equivalent to the top 5.1 days each year. In both locations, the day with the most precipitation each year makes up a much smaller fraction of total precipitation. We might thus expect this measure to change with warming at a higher rate than total precipitation. For extreme precipitation to evade energetic requirements and increase faster than total precipitation, it can only constitute a small fraction of total precipitation.

"...how we define extreme precipitation affects the conclusions we draw about the way it changes."

The point at which precipitation events considered extreme by some definitions increase at a rate close to the total precipitation arrives sooner than might be expected. Moderate events increase more slowly than the rate of moistening, whereas the most extreme events increase as much or possibly faster. Uncertainty about changes in extreme precipitation is largest for the most extreme events (see the figure), particularly in the tropics. Meanwhile, there is often some definition for which extreme precipitation does change at the same rate as atmospheric moisture. Some studies explicitly address how the definition of extreme precipitation affects its rate of change with atmospheric warming (*9*), but many do not. Scaling changes of extreme precipitation to the rate of atmospheric moisture increase remains the default null hypothesis, regardless of how extreme precipitation is defined.

There are caveats. The magnitude of extreme precipitation change will vary with location and season (*10*). It will also vary with the time scale considered, the mechanisms driving the event, and other factors. In some places, no definition of extreme precipitation will follow the rate of moisture increase; this is, for example, the case in places affected by systematic shifts in the location of precipitating systems (such as those that find themselves in the newly expanded subtropical dry zones). There are also yet more ways to define extreme precipitation—for example, the increase in frequency of events above a certain threshold, or with only wet days (or wet hours) rather than all days (or all hours) (*13*). These definitions will be affected differently by changes in moisture and circulation.

The key implication is that how we define extreme precipitation affects the conclusions we draw about the way it changes. We should expect moderate precipitation events to change more slowly than the rate of moisture increase in the atmosphere. On the other hand, the most extreme events might increase at or above the rate of atmospheric moisture increase, as studies have indicated in the case of Hurricane Harvey. The decision of what definition of extreme precipitation to use is important when climate change information is carried into other areas, such as engineering, socioeconomic impacts, adaptation, and policy. The rate of moisture increase provides a starting point for understanding extreme precipitation change, but the whole story is more nuanced because of the complex relationship among precipitation extremes, atmospheric circulation, moisture, and warming. ■

REFERENCES AND NOTES

1. K. E. Trenberth, *Clim. Change* **42**, 327 (1999).
2. M. R. Allen, W. J. Ingram, *Nature* **419**, 224 (2002).
3. P. Pall, M. R. Allen, D. A. Stone, *Clim. Dyn.* **28**, 351 (2007).
4. M. D. Risser, M. F. Wehner, *Geophys. Res. Lett.* 10.1002/2017GL075888 (2017).
5. G. J. van Oldenborgh *et al.*, *Environ. Res. Lett.* **12**, 124009 (2017).
6. S.-Y. S. Wang, L. Zhao, J.-H. Yoon, P. Klotzbach, R. Gillies, *Environ. Res. Lett.* 10.1088/1748-9326/aabb85 (2018).
7. K. E. Trenberth *et al.*, *Nat. Clim. Change* **5**, 725 (2015).
8. A. G. Pendergrass, E. P. Gerber, *J. Clim.* **29** 10.1175/JCLI-D-16-0097.1 (2016).
9. A. F. Prein *et al.*, *Nat. Clim. Change* **7**, 48 (2017).
10. S. Pfahl *et al.*, *Nat. Clim. Change* **7**, 423 (2017).
11. A. G. Pendergrass, D. L. Hartmann, *J. Clim.* **27**, 8372 (2014).
12. P. A. O'Gorman, T. M. Merlis, M. S. Singh, *Q. J. R. Meteorol. Soc.* 10.1002/qj.3195 (2017).
13. C. Schär *et al.*, *Clim. Change* **137**, 201 (2016).
14. M. J. Menne *et al.*, *NOAA Natl. Clim. Data Center* 10.7289/V5D21VHZ (2012).

ACKNOWLEDGMENTS

R. Knutti, F. Lehner, G. Meehl, P. O'Gorman, and N. Rosenbloom provided useful feedback on the manuscript. The climate model simulations were contributed by many modeling groups to the CMIP5 archive, hosted by the Program for Climate Model Diagnosis and Intercomparison at esgf-node.llnl.gov/projects/cmip5. GHCN-Daily data are made available by the National Oceanic and Atmospheric Administration National Climatic Data Center. The author was supported by the Regional and Global and Global Model Analysis Program (RGMA) of the U.S. Department of Energy's Office of Science (BER) cooperative agreement DE-FC02-97ER62402. The National Center for Atmospheric Research is sponsored by the National Science Foundation. The author declares no competing interests.

10.1126/science.aat1871

QUANTUM MATERIALS

The expanding materials multiverse

Heat capacity and Raman experiments point to fractionalized excitations in a dipole liquid

By **Ben J. Powell**

High-energy physicists are limited to studying a single vacuum and its excitations, the particles of the standard model. For condensed-matter physicists, every new phase of matter brings a new "vacuum." Remarkably, the low-energy excitations of these new vacua can be very different from the individual electrons, protons, and neutrons that constitute the material. The materials multiverse contains universes where the particle-like excitations carry only a fraction of the elementary electronic charge (*1*), are magnetic monopoles (*2*), or are their own antiparticles (*3*). None of these properties have ever been observed in the particles found in free space. Often, emergent gauge fields accompany these "fractionalized" particles (*2*, *4*, *5*), just as electromagnetic gauge fields accompany charged particles. On page 1101 of this issue, Hassan *et al.* (*6*) provide a glimpse of the emergent behaviors of a putative new phase of matter, the dipole liquid. What particles live in this universe, and what new physics is found in this and neighboring parts of the multiverse?

Liquids and gases look the same everywhere, but the periodic arrays of atoms in crystals break translational and rotational symmetries. This broken symmetry leads directly to the important differences between a crystal and a fluid—for example, the crystal's rigidity (*7*). The differences between gases and liquids are more subtle. Particles move freely in a gas, like cars on the open road. In a liquid, the motion of particles is correlated—i.e., it depends on what other particles are doing, like city driving (see the figure). Continuing this analogy, a glass is like a traffic jam, and a crystal resembles a parking lot with every particle neatly in place.

Some phases of matter are easier to detect than others. In ferromagnets, the spins

School of Mathematics and Physics, The University of Queensland, QLD 4072, Australia. Email: bjpowell@gmail.com

of unpaired electrons (radicals) align, creating a net magnetization (see the figure). This property allowed ancient civilizations to discover ferromagnetism. In antiferromagnets, neighboring spins point in opposite directions, leaving no net magnetic moment. Thus, antiferromagnetism was not observed until the 20th century. We still await conclusive experimental evidence for topological spin liquids—proposed phases where the spins lack long-range order but display long-range quantum correlations known as entanglement (*5*).

(Anti)ferroelectrics are phases in which electric dipoles align (anti)parallel to their nearest neighbors (see the figure). (Anti)ferromagnets and (anti)ferroelectrics could be described as spin crystals or dipole crystals, respectively, as a symmetry is broken. Simple paramagnets and paraelectrics, with little correlation between spins or dipoles, respectively, are essentially spin gases or dipole gases. Similarly, systems with strong correlations between spins or dipoles, but no broken symmetry, are known as spin liquids (*4*, *5*) and dipole liquids (*6*, *8*, *9*).

Hassan *et al.* exploited two special features of molecular crystals: their internal structure and the dipole inherent in dimer Mott insulators. They studied salts of bis(ethylenedithio)-tetrathiafulvalene (BEDT-TTF). In the dimer Mott phase, most $(BEDT\text{-}TTF)_2$ dimers carry a charge of +1. They are insulating because excitations that increase the charge of one dimer are bound to excitations that decrease the charge on a nearby dimer. Classically, the positive charge on a dimer must reside on one molecule or other, so there are two states that differ by the flip of an electric dipole. Quantum tunneling can occur between these states, as in the Creutz-Taube ion (*10*). If the tunneling is sufficiently rapid, the two states are in a quantum superposition, and a dipole gas forms.

Hassan *et al.* used Raman scattering to probe the two vibrational modes of the molecules: one mode that is sensitive to the charge on the molecule and another that is not. This difference allowed them to study dipolar fluctuations. $\kappa\text{-}(BEDT\text{-}TTF)_2Hg(SCN)_2Cl$ is a dipole solid at low temperature (T), spontaneously breaking the inversion symmetry of the $(BEDT\text{-}TTF)_2^+$ dimers. However, in $\kappa\text{-}(BEDT\text{-}TTF)_2Hg(SCN)_2Br$, the dipoles fluctuate rapidly, leading to a dipole fluid. The observation of a low-energy continuum of excitations in the Raman spectrum is consistent with the formation of a dipole liquid with fractionalized excitations.

The low-temperature heat capacity of the dipole solid in $\kappa\text{-}(BEDT\text{-}TTF)_2Hg(SCN)_2Cl$ is proportional to T^3, as one would expect from phonons or other bosonic excitations. However, in $\kappa\text{-}(BEDT\text{-}TTF)_2Hg(SCN)_2Br$, there is an additional term in the heat capacity proportional to T. Itinerant fermionic excitations would be a natural explanation of this term.

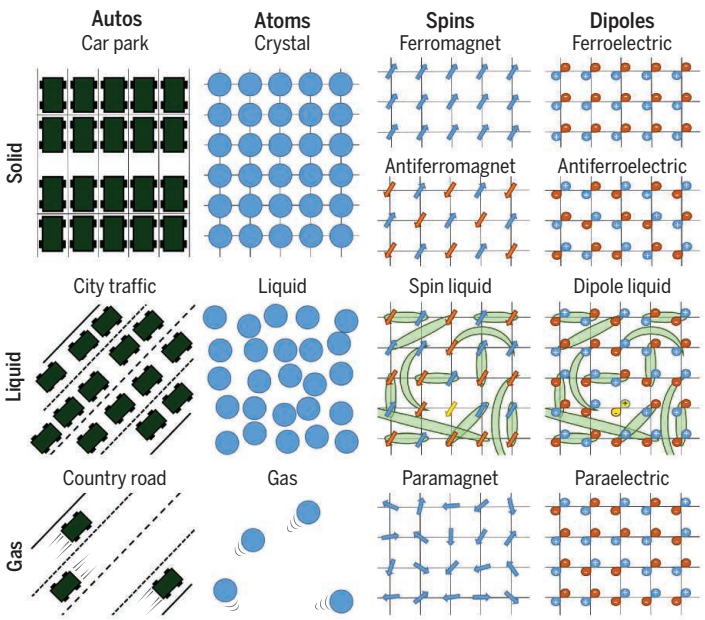
Hassan *et al.* argued that these two experiments are evidence for either collective excitations of the dipoles or hybrid spin-dipole excitations driven by interactions between the dipoles and the unpaired spins of the $(BEDT\text{-}TTF)_2^+$ radicals. In a broken symmetry phase, such as a dipole solid, the collective excitations are Goldstone bosons (*7*). However, in a dipole liquid, excitations might fractionalize into fermionic particles, as appears to happen in some spin liquids (*4*, *5*). A possible route to liquid phases is sketched in the figure. Pairs of spins (or dipoles) form singlets. Within each singlet the spins (dipoles) point in opposite directions, but both spins (dipoles) can flip simultaneously. This is an example of quantum entanglement. A quantum state composed of many different configurations of spin singlets can produce a spin liquid (*4*, *5*). Excitations in such a state, called spinons, involve unpaired spins moving about the crystal and can obey either Fermi or Bose statistics (*4*, *5*). Can something similar happen in a dipole liquid?

If the excitations have hybrid spin-dipole character, might spin-dipole interactions also drive the exotic physics of the BEDT-TTF salts? Little is known about the spin physics in $\kappa\text{-}(BEDT\text{-}TTF)_2Hg(SCN)_2Br$, so Hassan *et al.* turned to $\kappa\text{-}(BEDT\text{-}TTF)_2Cu_2(CN)_3$, which is believed to be a spin liquid (*4*). No dipole liquids or solids were seen in this material, so its physics likely arises solely from spins.

Like many of the most interesting experiments, the work of Hassan *et al.* contains more questions than answers. They have discovered an exciting new universe ripe for exploration. The conclusive demonstration that the excitations to its "vacuum" are itinerant fermions may require new experimental tools that can directly probe the particle-like excitations. BEDT-TTF salts often become superconducting under pressure (*4*), and many physicists believe that this superconductivity results from the material failing to become a spin liquid, despite its propensity to do so (*11*). Do failed dipole liquids also superconduct? How common are dipole liquids? There is evidence for one in $BaFe_{12}O_{19}$ (*8*) and proposals for building them from polar molecules (*9*). Might the exotic behaviors of ice X (*12*) be understood as a quantum dipole liquid? ■

A motor tour of the multiverse

Solid, liquid, and gas phases of particles, spins, and dipoles, along with their automotive analogs, are shown. The sketch for the dipole liquid is speculative. Hassan *et al.*'s work suggests that dipole liquids are similar to spin liquids and may support itinerant fermionic quasiparticles.



For the spin and dipole liquids, the green ribbons indicate quantum correlations between spins or dipoles pointing in opposite directions. Spinons, excitations in spin liquids that are quantum superpositions of unpaired spins, are highlighted in yellow.

REFERENCES

1. R. de-Picciotto *et al.*, *Nature* **389**, 162 (1997).
2. C. Castelnovo, R. Moessner, S. L. Sondhi, *Annu. Rev. Cond. Matter Phys.* **3**, 35 (2012).
3. V. Mourik *et al.*, *Science* **336**, 1003 (2012).
4. B. J. Powell, R. H. McKenzie, *Rep. Progr. Phys.* **74**, 056501 (2011).
5. L. Balents, *Nature* **464**, 199 (2010).
6. N. Hassan *et al.*, *Science* **360**, 1101 (2018).
7. P. W. Anderson, *Basic Notions of Condensed Matter Physics* (Addison-Wesley, Reading, MA, 1984).
8. S.-P. Shen *et al.*, *Nat. Commun.* **7**, 10569 (2016).
9. N. Y. Yao, M. P. Zaletel, D. M. Stamper-Kurn, A. Vishwanath, *Nat. Phys.* **14**, 405 (2018).
10. K. D. Demadis, C. M. Hartshorn, T. J. Meyer, *Chem. Rev.* **101**, 2655 (2001).
11. B. J. Powell, R. H. McKenzie, *Phys. Rev. Lett.* **94**, 047004 (2005).
12. M. Benoit, D. Marx, M. Parrinello, *Nature* **392**, 258 (1998).

10.1126/science.aat7282

GRAPHIC: ADAPTED BY N. CARY/*SCIENCE*

DEVELOPMENT

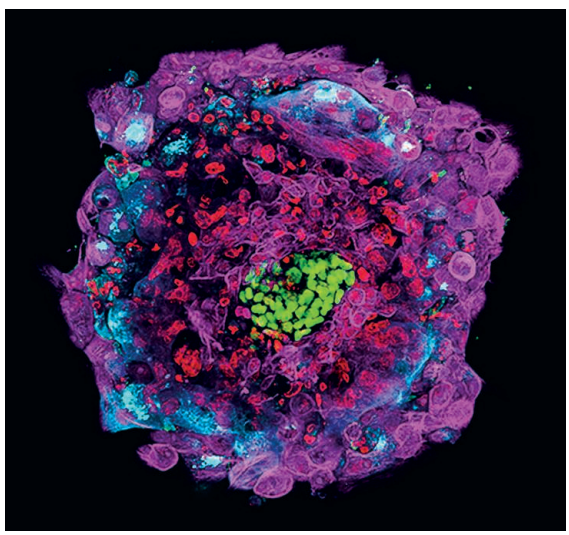
Exploring early human embryo development

Stem cell–derived models enable understanding of human embryogenesis

By **Janet Rossant**[1] *and* **Patrick P. L. Tam**[2]

Human in vitro fertilization (IVF) has been around for 40 years and yet we still do not fully understand how to ensure the generation of healthy embryos and to prevent unwanted genetic or epigenetic changes that may arise during in vitro culture. Once the embryo is transferred to the uterus to continue gestation, early implantation and the initial phases of embryogenesis and placenta formation are hidden from easy access for research. Errors in early development may lead to implantation problems, fetal defects, and placental insufficiencies, resulting in early pregnancy loss. Studying mice has provided clues about the major genetic and epigenetic events of early embryo development but, as is becoming increasingly evident, there are morphological and genetic differences between mice and humans that make cross-species comparisons problematic (*1*). Recent experimental approaches working directly with human embryos, or with embryo-derived stem cells and nonhuman primate embryos, have opened new avenues for studying the development of early human embryos.

Recent single-cell gene expression (RNA sequencing) analysis of early human embryos has provided a molecular time course of developmental progression (*2*). This has identified three distinct cell lineages in the expanded blastocyst [embryonic day 5 (E5) cavitated blastocyst]: the outer trophectoderm (TE), and the epiblast (EPI) and primitive endoderm (PrE) of the enclosed inner cell mass (ICM) (see the figure). Many key lineage specifiers, such as CDX2 (caudal-type homeobox protein 2), POU5F1 (POU domain, class 5, transcription factor 1), and SOX17 (SRY-box 17), show conserved expression with mouse embryo development, albeit with differences in timing of expression. In humans, these three cell lineages cannot be easily separated by transcriptional profiling prior to the expanded blastocyst stage, whereas in the mouse, the ICM and TE transcriptional profiles are distinct before the morphological events of cavitation of the blastocyst. The outer cells of the blastocyst are committed to a TE fate in the 32-cell-stage mouse embryo (*3*). By contrast, TE cells of the fully expanded human blastocyst are still capable of regenerating an entire blastocyst (*4*). These observations suggest that lineage segregation takes place after blastocyst formation in humans instead of progressively as in mice.



A human blastocyst cultured in vitro to 12 days postfertilization shows cell layers reminiscent of postimplantation development.

If true, the difference in lineage segregation between mice and humans has implications for the cross-species extrapolation of the upstream signaling events that drive lineage specification. For example, differences in Hippo signaling between the outer and inner cells of mouse embryos prior to blastocyst formation are important in specifying ICM versus TE fate. However, in humans, it is not clear whether the same pathway is involved. Additionally, fibroblast growth factor (FGF) signaling differences are key to the specification of EPI and PrE in the mouse ICM, but blocking FGF signaling in human embryos does not affect PrE formation (*1*), suggesting that other signaling pathways may be involved in this lineage decision, with implications for providing the critical signaling conditions in vitro to ensure correct development of IVF embryos.

Direct assessment of gene function in early embryos could provide insights into the importance of differences between mice and humans, but this was not considered feasible until the advent of CRISPR-Cas gene editing. Several jurisdictions have allowed experimental gene editing in early human embryos to explore gene function and the potential utility of germline correction of genetic diseases. Gene editing of *POU5F1* in human embryos has shown unexpectedly that mutant embryos in which this gene was ablated failed to reach the blastocyst stage, suggesting a role for *POU5F1* earlier in human than in mouse development (*5*).

Although it is feasible to study early development up to the blastocyst stage directly in human embryos cultured under IVF conditions, extension of such studies to implantation and the early postimplantation period requires new technical capabilities to maintain the embryo beyond the blastocyst stage. Recently, human blastocysts have been grown over the implantation period in a two-dimensional (2D) culture system and achieved the initiation of EPI, amnion, and yolk sac formation (*6, 7*). Although these cultured embryos are not properly organized and fail to progress beyond 10 to 12 days, advances in the fabrication of extracellular matrices and improved culture conditions could support extended human embryo development toward the next key milestone of development that occurs at around E14, gastrulation. Such experiments have not taken place, because there is an almost universal prohibition on growing intact human embryos beyond 14 days of development in culture. This time point marks the onset of nervous system development. Given the potential for such cultures to provide new insights into the key events of early development, there are emerging discussions about whether the 14-day rule should be revisited.

In view of the ethical concerns around the extended culture of human embryos and the inaccessibility of the early implantation stages in vivo, alternate models need

[1]*Peter Gilgan Centre for Research and Learning, Hospital for Sick Children, and Department of Molecular Genetics, University of Toronto, Toronto, ON M5G 0A4, Canada.*
[2]*Embryology Unit, Children's Medical Research Institute and School of Medical Sciences, Faculty of Medicine and Health, University of Sydney, Westmead, NSW 2145, Australia.*
Email: janet.rossant@sickkids.ca; ptam@cmri.org.au

to be considered for research on early human development. The similarities between the morphological features of human and nonhuman primate embryos have long been apparent (see the figure). A study on cynomolgus monkey postimplantation embryos has revealed details of the process of amnion formation, trophoblast development (which forms the placenta), germ cell formation and gastrulation, and the timing and spatial location of expression of some of the key genes underpinning these events (*8*). Although nonhuman primate embryos may be amenable for direct experimentation and, potentially, for genetic modification, such studies are technically challenging and come with concerns of animal ethics and care. There are thus still limitations on the extent to which nonhuman primate embryos can be considered a surrogate for human embryo experimentation.

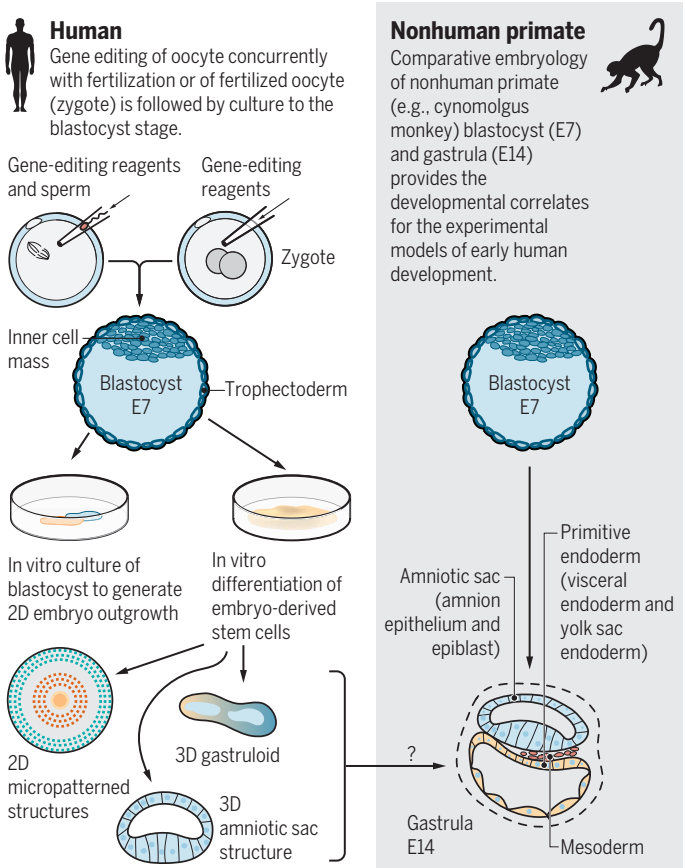
Recently, it has been shown that embryo-like structures (embryoids) can be generated in culture from human pluripotent embryonic stem (ES) cells. In micropatterned cultures of human pluripotent embryonic stem cells, organized 2D patterns of cell types and associated gene expression profiles, reminiscent of the ordered pattern in the germ layers of mouse embryos, can be discerned (*9*). These patterned constructs do not fully recapitulate the 3D organization that is the hallmark of the embryo at gastrulation. By contrast, cultures of pluripotent stem cells in 3D matrices can generate embryoid structures that appear to recapitulate development of the EPI and amnion formation (*10*).

In mice, ES cells grown as aggregates in 3D supporting matrices generate gastruloids (*11*). These entities not only mimic EPI organization but also display a localized primitive streak-like structure and anterior-posterior tissue patterning, which are the distinctive features of a gastrulation-stage embryo. This degree of organization in the absence of the extraembryonic tissues (such as the trophoblast and PrE), which normally provide localized signals to establish the body plan at gastrulation, is remarkable. Furthermore, combining ES cells and trophoblast stem (TS) cells can mimic some of these embryonic-extraembryonic interactions and enables the generation of blastoids and embryoids with architecture that resembles that of blastocysts and pregastrulation embryos, respectively (*12*, *13*). These findings raise the possibility that the combination of human ES cells and human TS cells (*14*) may provide similar experimental models for human development. At this juncture, the pressing question is whether the development of these embryo-like structures mirrors that of the human embryo in vivo closely enough that the outcomes are scientifically relevant to early human development. Further study directly on human embryos is needed to establish a paradigm of developmental correlates to guide the evaluation of the findings from these embryo-like entities.

The goal of research on human embryology is to gain scientific knowledge on the fundamentals of the cellular and molecular control of early human development, and its application to assisted reproductive technologies, gene editing, stem cell research, and the prevention of genetic birth defects. Although such knowledge would ideally be gleaned from studying the human embryo per se, the ethical, legal, and practical issues concerning the provenance of "excess" and "consented" human IVF embryos for research must be observed. The prohibition in many jurisdictions on creating embryos for research also precludes the feasibility of certain experimentation, such as germline gene editing prior to fertilization and zygote formation. In the future, stem cell–derived embryo-like structures could be the model systems, in place of human embryos, that offer the most useful insights into early human development.

As the technology for the generation of stem cell–derived embryo-like structures improves, these biological constructs will more closely mimic human embryos. This raises the question of whether the emergent properties of these entities should bring them under the same ethical limitations as human embryos. Should the 14-day limit of in vitro development be applied to these embryo-like cultures (*15*)? The embryological knowledge that informed the 14-day rule is gleaned from studies of mouse development and morphological studies of a rare archive of human embryonic materials, but not on experimental findings of the defining attributes of human embryos. The 14-day rule might lose its relevance as a meaningful and practical defining limit of the extent of development for the study of embryoids and gastruloids if they can advance beyond the nominal signposts of a 14-day human embryo. There is, therefore, a strong imperative to perform robustly controlled studies of human postblastocyst embryo development to test the scientific merit, and the ethics and legal practicality, of the 14-day rule before the consideration of whether to uphold or modify this rule for research on early human embryos and alternative embryological models. ■

Experimental studies of embryo development



GRAPHIC: N. DESAI/*SCIENCE*

REFERENCES

1. J. Rossant, P. P. L. Tam, *Cell Stem Cell* **20**, 18 (2017).
2. G. G. Stirparo *et al.*, *Development* **145**, dev158501 (2018).
3. E. Posfai *et al.*, *eLife* **6**, e22906 (2017).
4. C. De Paepe *et al.*, *Mol. Hum. Reprod.* **20**, 599 (2014).
5. N. M. E. Fogarty *et al.*, *Nature* **550**, 67 (2017).
6. A. Deglincerti *et al.*, *Nature* **533**, 251 (2016).
7. M. N. Shahbazi *et al.*, *Nat. Cell Biol.* **18**, 700 (2016).
8. T. Nakamura *et al.*, *Nature* **537**, 57 (2016).
9. A. Warmflash *et al.*, *Nat. Methods* **11**, 847 (2014).
10. Y. Shao *et al.*, *Nat. Commun.* **8**, 208 (2017).
11. D. A. Turner *et al.*, *Development* **144**, 3894 (2017).
12. N. C. Rivron *et al.*, *Nature* **557**, 106 (2018).
13. S. E. Harrison *et al.*, *Science* **356**, eaal1810 (2017).
14. H. Okae *et al.*, *Cell Stem Cell* **22**, 50 (2018).
15. J. Aach *et al.*, *eLife* **6**, e20674 (2017).

10.1126/science.aas9302

RETROSPECTIVE

Stanley Falkow (1934–2018)

Founder of molecular pathogenesis who discovered how antibiotic resistance spreads

By **Denise Monack**[1] *and* **Evelyn Strauss**[2]

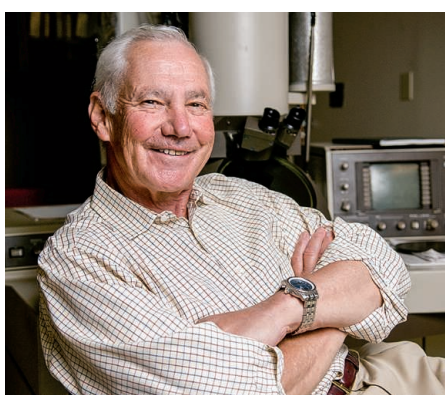
A self-described "terrible student," Stanley Falkow loved to read, though not what his teachers assigned. In the library at age 11, he discovered the book *Microbe Hunters* and was captivated by tales of tiny disease-causing critters and the pursuit to uncover their secrets. That experience ignited a lifelong passion that fueled an extraordinary career. He died on 5 May at his home in Portola Valley, California. Outside Stanley's former lab at Stanford University, we hear his voice and see his smile in our minds. We remember a visionary scientist and a devoted mentor.

Stanley spearheaded the development of the interrelated fields of microbial pathogenesis, molecular epidemiology, and cellular microbiology. As a graduate student in the 1960s, he unveiled a previously unknown mechanism by which unrelated bacteria share traits when he showed (with Julius Marmur of Brandeis University) that they can transfer—on extrachromosomal pieces of DNA—the ability to use lactose as food. He went on to demonstrate that similar genetic elements, now called plasmids, carry antibiotic resistance genes. These and other groundbreaking discoveries not only explained how antibiotic resistance spreads among bacteria, they also helped lay the foundation for the recombinant DNA enterprise. Soon, scientists would insert any gene of interest into a plasmid and place it in living cells to decipher the gene's function. Stanley immediately recognized this technique's potential for studying pathogenesis, and he used it to isolate, for the first time, a gene involved in bacterial virulence, one that encodes a toxin from a diarrhea-causing *Escherichia coli*.

Years later, he created what he termed "molecular Koch's postulates," a framework for discerning how particular genes and proteins contribute to infection and disease. The paradigm melded traditional microbiology with the molecular revolution he helped launch.

Stanley predicted the spread of clinical antibiotic resistance and, in the late 1970s, he urged the U.S. Food and Drug Administration to ban antibiotics in animal feed. The livestock industry pushed back and the attempt failed, but Stanley's opinions and recommendations on the risks of placing new genes in bacteria influenced guidelines for modern molecular biology research.

Stanley Falkow was born in Albany, New York, in 1934. After studying at the University of Maine and earning a doctorate from Brown University, his career path led him to Stanford University in 1981 where he remained on the faculty for the next 23 years and continued as an emeritus professor until his death. He mentored more than 120 graduate students, postdocs, and clinical fellows; his trainees are among the most prominent microbiologists in the world. His style was unusual in many ways, and his generosity shone. He frequently said that he learned more from his students than vice versa, and he prioritized their interests. His postdocs could take their projects with them when they set up their own labs, and he would focus on other topics to avoid competition. When I (E.S.) expressed a desire to learn how to study a pathogen "from the beginning," he suggested that I start an *Edwardsiella tarda* project. Initial studies on this obscure pathogen likely would not break new scientific ground or yield high-profile publications, yet he recognized the potential for teaching me what I wanted to learn.



Stanley's wit infused many lessons and stories, as did his irreverence. When I (D.M.) began working in his lab as a technician in the mid-1980s, I was fortunate enough to sit in on his lectures to medical students. He began by telling us about the normal flora in the "warm Miami" of the colon and informed us that "normal stool is about the size, shape, and consistency of an overripe banana." To convey modes of bacterial transmission, he used a marker on transparencies (those were the days of overhead projectors) to draw stick figures and arrows—from rats to humans, for instance. By the time he got to fecal-oral spread and sexual transmission, the students were in stitches; Stanley maintained a deadpan delivery. In informal situations, he'd illustrate the clinical symptoms caused by enteric bacterial pathogens such as *Shigella* by showing pictures of bloody diarrhea. He would get the expected groan or blanch and his eyes would twinkle. "Well, it may be shit to you," he'd say, "but it's my bread and butter."

When Stanley received the Lasker-Koshland Special Achievement Award in Medical Science in 2008, he titled an associated commentary, "I never met a microbe I didn't like." The range of bacterial pathogens he studied—*Yersinia*, *Helicobacter*, *Bordetella*, *Salmonella*, *Neisseria*, and *Edwardsiella*, to name a few—bore out this sentiment. Not only did he like them, but he identified with them. Knowing only Stanley's stature in the field, one of us (E.S.) probably benefited from this affinity. In my postdoc interview, he asked why a yeast biologist wanted to study bacterial pathogenesis. I told him that I held a long interest in microbial disease, and then I blurted out, "Also, I like pretending that I *am* a bacterium." I cringed and thought, "No way did that statement make sense to him." But Stanley got it. His proclivity to "see" things from a microbial point of view likely fueled some of his imaginative and insightful thinking. He frequently pointed out that bacteria have no malignant intent; their goal is not to make anyone sick. His dictum was, "a bacterium's goal is to become bacteria."

Stanley shaped not only the way that I (D.M.) approach science, but also how I think about my students and postdocs. Several decades ago, I had occasion to attend a meeting at a biotech company on whose scientific advisory board Stanley sat. Employees, scientific advisers, and executives were discussing finances and assets. At one point in the conversation, Stanley's face clouded. He was irritated by a comment someone had made. "Don't forget that your most important and valuable assets are sitting here at this table," he said as he looked around at the employees.

In the hallway outside his former lab at Stanford, a portrait of Stanley by a former student captures his facial asymmetry and his warm gaze. We look up and see a giant among microbe hunters, an inspiring teacher, and a beloved friend. ■

10.1126/science.aau2284

[1]*Department of Microbiology and Immunology, Stanford School of Medicine, Stanford, CA 94305, USA.* [2]*Science Communication Program, University of California, Santa Cruz, Santa Cruz, CA 95064, USA. Email: ejstrauss3@gmail.com; dmonack@stanford.edu*

PHOTO: LINDA A. CICERO/STANFORD NEWS SERVICE

POLICY FORUM

GENETICS AND PRIVACY

Genealogy databases and the future of criminal investigation

The police can access your online family-tree research—and use it to investigate your relatives

By **Natalie Ram,[1] Christi J. Guerrini,[2] Amy L. McGuire[2]**

The 24 April 2018 arrest of Joseph James DeAngelo as the alleged Golden State Killer, suspected of more than a dozen murders and 50 rapes in California, has raised serious societal questions related to personal privacy. The break in the case came when investigators compared DNA recovered from victims and crime scenes to other DNA profiles searchable in a free genealogical database called GEDmatch. This presents a different situation from the analysis of DNA of individuals arrested or convicted of certain crimes, which has been collected in the U.S. National DNA Index System (NDIS) for forensic purposes since 1989. The search of a nonforensic database for law enforcement purposes has caught public attention, with many wondering how common such searches are, whether they are legal, and what consumers can do to protect themselves and their families from prying police eyes. Investigators are already rushing to make similar searches of GEDmatch in other cases, making ethical and legal inquiry into such use urgent.

In the United States, every state, as well as the federal government, has enacted laws enumerating which convicted or arrested persons are subject to compulsory DNA sampling and inclusion in the NDIS database. The NDIS contains more than 12 million profiles, and it is regularly used to match DNA from crime scenes to identify potential suspects. It is only helpful, however, if the suspect—or a family member of the suspect—has been arrested or committed a crime and their DNA has been collected and stored.

The case of the Golden State Killer is not the first instance of investigators turning to nonforensic DNA databases to generate leads. This was not even the first time investigators used genealogical DNA matches to develop and pursue a suspect in the Golden State Killer case itself. A year before investigators zeroed in on DeAngelo, they subpoenaed another genetic testing company for the name and payment information of one of its users and obtained a warrant for the man's DNA. He was not a match. Similarly, in 2014, Michael Usry found himself the target of a police investigation stemming from a partial genetic match between his father's DNA, stored in an Ancestry.com database, and DNA left at a 1996 murder scene. On the basis of the partial match, police were able to obtain a court order requiring Ancestry.com to disclose the identity of the database DNA match. After mapping out several generations of Usry's father's family, investigators zeroed in on Usry, eventually securing a warrant for his DNA. Ultimately, Usry was cleared as a suspect when his DNA proved not to match the crime scene DNA.

But there have also been reported successes. In 2015, for example, Arizona police arrested and charged Bryan Patrick Miller in the Canal Killer murders based in part on a tip drawn from a genealogical database search (*1*). Searches like these, drawing on genetic information unrelated to the criminal justice system, may offer substantial benefits. Allowing police to conduct similar database searches in other cases is likely to lead to more solved crimes. Moreover, expanding law enforcement investigations to encompass genealogical databases may help to remedy the racial and ethnic disparities that plague traditional forensic searches. In accordance with state laws, official forensic databases are typically limited to individuals arrested or convicted of certain crimes. Racial and ethnic disparities throughout the criminal justice system are therefore reproduced in the racial and ethnic makeup of these forensic databases. Genealogical databases, by contrast, are biased toward different demographics. The 23andMe database, for instance, consists disproportionately of individuals of European descent. Including genealogical databases in forensic searches might thus begin to redress, in at least one respect, disparities in the criminal justice system.

There are few legal roadblocks to police use of genetic databases intended to help individuals explore their health or identify genetic relatives. The Fourth Amendment's protection against warrantless searches and seizures generally does not apply to material or data voluntarily shared with a third party, like a direct-to-consumer genetics testing or interpretation company or a genetic matching platform like GEDmatch. Once an individual has voluntarily shared her data with a third party, she typically cannot claim any expectation of privacy in those data—and so the government need not secure a warrant before searching it.

Beyond the Constitution, three federal laws protect some genetic data against certain disclosures, but these too are unlikely to provide an effective shield against law enforcement searches in nonforensic genetic databases. The Genetic Information Nondiscrimination Act (GINA) protects genetic data, but only against certain uses by employers and health insurers (*2*). GINA provides no protection against law enforcement searches. Similarly, most companies and websites offering DNA testing, interpretation, or matching services directly to individuals likely are not covered by the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, which governs the use and disclosure of identifiable health information. These providers are usually careful to explain that they are not engaged in health care or the manipulation or provision of health data (*3*). Finally, although certificates of confidentiality protect scientific researchers from disclosing data to law enforcement—even against a warrant (*4*)—they do not extend to scenarios in which law enforcement is just another contributor to and user of online genetic resources, such as public databases and matching tools. Certificates of confidentiality have faced few challenges in court, and so it is also uncertain whether the protection they purport to provide will hold up against a challenge by law enforcement seeking access.

Consistent with this legal landscape, companies and websites that generate, interpret, or match genetic data directly for individuals often do not promise complete protection. In terms of law enforcement, for instance, 23andMe states in its privacy policy, "23andMe will preserve and disclose any and all information to law enforcement agencies or others if required to do so by law or in the good faith belief that such preservation or disclosure is reasonably necessary to…comply with legal or regulatory process (such as a judicial proceeding, court order, or government inquiry)…" (*5*). Ancestry.com similarly

[1]University of Baltimore School of Law, Baltimore, MD, USA. [2]Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX, USA. Email: nram@ubalt.edu; amcguire@bcm.edu

discloses, "We may share your Personal Information if we believe it is reasonably necessary to: [c]omply with valid legal process (e.g., subpoenas, warrants)..." (*6*). And in the wake of the Golden State Killer arrest, GEDmatch has altered its terms of service to explicitly permit law enforcement use of its database to investigate homicides and sexual assault (*7*). Although these disclaimers are usually unambiguous, they are sometimes buried in terms of service or privacy policies that many individuals do not take care to read or fully understand.

Despite the lack of legal protection against law enforcement searches of nonforensic databases, such searches may run counter to core values of American law. The Fourth Amendment is a constitutional commitment to protect fundamental civil rights. Part of that is a commitment to protecting privacy or freedom from government surveillance. Police cannot search a house without suspecting a specific individual of particular acts—even if doing so would enable the police to solve many more crimes. Yet, database searches permit law enforcement to search the genetic data of each database member without any suspicion that a particular member is tied to a particular crime. Although the U.S. Supreme Court has approved suspicionless genetic searches for individuals with diminished expectations of privacy, like those arrested or convicted of crimes (*8*), ordinary members of the public are different. Familial searches, like those used in the Golden State Killer investigation, are an even further departure from the Supreme Court standard. Certainly, individuals who commit crimes and leave their DNA behind forfeit any expectation of privacy in that DNA. But a usable forensic identification requires two matching parts: a crime scene sample and a database profile that matches it. Suspects identified through familial searches cannot be said to have voluntarily shared their genetic profile in a database of known individuals, even if a genetic relative has.

The Supreme Court is poised to reconsider its broad rule that the voluntary sharing of data negates expectations of privacy—and thus negates Fourth Amendment protections against warrantless government searches. In *Carpenter* v. *United States*, the Supreme Court will determine whether police must obtain a warrant to justify access to historical cell phone records revealing the movements and location of a cell phone user over a long period of time (*9*). In the digital age, in which nearly all data are at least nominally shared with third parties like internet service providers, website hosts, and cell phone companies, the current rule means that the Fourth Amendment often does not apply. *Carpenter* may reshape this rule to account for the realities of a big-data world. A ruling in *Carpenter* that limits police use of historical cell phone data may substantially affect police practices surrounding genetic data as well, as merely sharing data with another might well be insufficient to permit its suspicionless search by the government for crime-detection purposes.

Even if the Supreme Court decision in *Carpenter* does not revamp Fourth Amendment rules governing police access to shared data, the setting of that case suggests another way to resolve concerns about police access to nonforensic genetic databases. In the Stored Communications Act, Congress provided substantial statutory protection for email and other digital information maintained on the internet. Under the act, a court may order disclosure of electronic records if the government "offers specific and articulable facts showing that there are reasonable grounds to believe" that the records sought "are relevant and material to an ongoing criminal investigation" (*10*). This standard is less onerous than the Fourth Amendment's warrant requirement, but it is notably more demanding than any protections the law currently provides.

Enacting similar protection for genetic data stored in nonforensic databases would ensure that the government cannot subject ordinary individuals to suspicionless genetic searches, while allowing investigators to access genetic data where there is reason to believe a particular individual may be tied to a particular crime. A Stored Genetics Act would likely render law enforcement searches of nonforensic genetic databases unlawful for crime-detection purposes, as there can be no "specific and articulable" connection between particular database records and a particular crime when investigators seek to use such a search to generate leads, not investigate them. Thus, although such an approach would preserve freedom from perpetual genetic surveillance by the government, it may well result in fewer solved cases.

Legislatures may understandably be loath to enact a total prohibition of such searches. At a minimum, however, policy-makers should delineate under what circumstances such searches are acceptable. For example, several states, including California, Colorado, and Texas, have identified prerequisites to the use of familial searches of the state's own forensic database, including that the crime to be investigated is serious and that traditional investigative techniques have been exhausted without success (*11*). Similar constraints could be placed on law enforcement searches of nonforensic databases. The challenge of this approach is that limitations on the scope of use can erode quickly. Thus, although Colorado's policy governing familial searches of the state's forensic database limits such searches to crimes with "significant public safety concerns," police in that state used a familial search to solve a car break-in where the perpetrator "left a drop of blood on a passenger seat when he broke a car window and stole $1.40 in change" (*11*). The erosion of limits on crime-solving technology may well be inevitable, and it threatens our collective civil liberties and opens the door to socially and politically unacceptable genetic surveillance.

Whatever legislative solution is adopted, it must at least take into account public perspectives to clearly delineate acceptable uses and balance the social benefit of solving cases with individuals' interests in avoiding unwarranted government scrutiny. ■

REFERENCES AND NOTES

1. M. Cassidy, "How forensic genealogy led to an arrest in the Phoenix 'Canal Killer' case," Arizona Republic, 10 November 2016; www.azcentral.com/story/news/local/phoenix/2016/11/30/how-forensic-genealogy-led-arrest-phoenix-canal-killer-case-bryan-patrick-miller-dna/94565410/.
2. 122 Statute 881.
3. J. Hsu, *I/S: A Journal of Law & Policy for the Information Society (Moritz College of Law)* **6**, 557 (2011).
4. 42 United States Code (U.S.C.) § 241 (d).
5. www.23andme.com/about/privacy/.
6. www.ancestry.com/cs/legal/privacystatement.
7. https://bit.ly/2IZKzGt.
8. *Maryland* v. *King*, 133 S. Ct. 1958 (2013).
9. *Carpenter* v. *United States*, No. 16-402 (argued 29 November 2017).
10. 18 U.S.C. § 2703 (d).
11. N. Ram, *Stanford Law Rev.* **63**, 751 (2011).

ACKNOWLEDGMENTS

Funding was provided by the National Institutes of Health, National Human Genome Research Institute (K01 HG009355 and R01 HG008918). We thank M. Majumder, D. Peterson, S. Pereira, R. Hsu, A. Gutierrez, and J. Robinson for contributions and assistance on this project.

10.1126/science.aau1083

LETTERS

Sled dogs at work in Greenland.

PHOTO: CARSTEN EGEVANG

Edited by **Jennifer Sills**

Greenland sled dogs at risk of extinction

Dog (*Canis familiaris*) and sled remains found together suggest that the first sled dogs arose approximately 10,000 years ago (*1*). The ancestors to current sled dog breeds were essential to the Inuit's conquest of the Arctic. Sled dogs provided the main transport and hunting platform (*2*) for various Inuit groups, including the Old Bering Sea culture and Punuk (*3*), dating back at least 2000 years (*4*). The endemic Greenland sled dog breed lives with local communities north of the Arctic Circle on both the west and east coasts of Greenland. Greenland sled dog numbers have declined substantially, from more than 25,000 in 2002 (*5*) to fewer than 15,000 in 2016 [p. 15 in (*6*)].

There are multiple reasons for this decline. First, because of climate change, sea-ice is retreating, which hampers traditional hunting and hinders the provision of basic food for humans and dogs alike (*7*). Second, a large number of dogs and entire subpopulations are being wiped out by epidemics of canine distemper and parvovirus (*8*). Third, a change in culture has led people to replace the sledge with the snowmobile [p. 23 in (*6*)]. Unlike most other endangered species, sled dogs are domestic animals that can potentially be bred if there are incentives to do so. However, these changes in the cultural legacy and the reduced need for the dogs to pull sledges have led to correspondingly reduced motivation to breed enough dogs to maintain population numbers.

This drastic population decline could lead to the extinction of this unique breed, which would substantially affect how the Greenland Inuit use their environment, and in turn could affect health and well-being. Furthermore, because Greenland sled dogs are widely distributed and are vulnerable to the same health risks as humans (such as zoonosis, environmental chemical contaminants, and climate change), the species is being used to monitor One Health—an initiative streamlining the health of humans, animals, and the environment—in the Arctic (*9*). To mitigate disease outbreaks and to halt population decline, we urgently need more research focused on these culturally and ecologically important Greenland sled dogs.

Christian Sonne,[1]* Rikke Langebæk,[2] Rune Dietz,[1] Emilie Andersen-Ranberg,[1,2] Geoff Houser,[2] Anders J. Hansen,[2] Mikkel-Holger S. Sinding,[2,3] Morten Tange Olsen,[2] Carsten Egevang,[3] M. Thomas P. Gilbert,[2] Morten Meldgaard[2,4]

[1]Aarhus University, Roskilde, Denmark. [2]University of Copenhagen, Copenhagen, Denmark. [3]Greenland Institute of Natural Resources, Nuuk, Greenland. [4]University of Greenland, Nuuk, Greenland.
*Corresponding author. Email: cs@bios.au.dk

REFERENCES

1. V.V. Pitulko *et al., J. Arch. Sci. Rep.* **13**, 491 (2017).
2. R. W. Park, *Arctic* **40**, 184 (1987).
3. O. K. Mason, *J. Anthrop. Arch.* **17**, 240 (1998).
4. M. Raghavan, *Science* **345**, 1255832 (2014).
5. Statistics Greenland, Table 1: Number of sled dogs in Greenland (www.stat.gl/dialog/main.asp?lang=da&sc=SA&subthemecode=t1&colcode=T&version=201308).
6. Greenland in Figures (2017); www.stat.gl/publ/en/GF/2017/pdf/Greenland%20in%20Figures%202017.pdf.
7. K. Hastrup *et al., Ambio* **47**, 162 (2018).
8. J. Bohm *et al., Arctic Med. Res.* **48**, 195 (1989).
9. C. Sonne *et al., Acta Vet. Scand.* **59**, 84 (2017).

10.1126/science.aat9578

The misunderstood sixth mass extinction

Scientific misunderstanding about the nature and consequences of the sixth mass extinction has led to confusion among policy-makers and the public. Scientists agree that there have been five mass extinctions in the past 600 million years (*1*). Although scientists also agree that Earth is now suffering the sixth mass extinction, they disagree about its consequences. Mass extinctions are defined as the loss of the majority of species in a relatively short geological time, caused by a catastrophic natural event (*2*). Some scientists argue that there is no reason for concern about the sixth mass extinction because extinction is normal, simply an inevitable consequence of the process of evolution (*3*, *4*). This misunderstanding ignores some critical issues. First, the rate of species extinction is now as much as 100 times that of the "normal rate" throughout geological time (*5*, *6*). Second, like the past mass extinctions, the current episode is not an inevitable consequence of the process of evolution. Rather, it is the result of a rare event changing the environment so quickly that many organisms cannot evolve in response to it.

In theory, evolution on Earth could proceed as long as conditions permitted with no mass extinction events. That has been the case for vast stretches of geological time between occasional encounters with unusual environmental circumstances. Extinctions did occur, but not suddenly and nearly universally, as is happening now (*7*, *8*). The rate and extent of current extinctions is similar to those of past mass extinctions, not the intervals between them (*9*, *10*). If past mass extinctions are any guide to the rate at which usual evolutionary diversification processes could restore a reasonable level of biodiversity and ecosystem services, the wait is likely to be millions, or even tens of millions of years (*8*, *9*).

At the time of the past mass extinctions, there was no industrialized human population of almost 8 billion people utterly dependent on the ecosystem services

biodiversity helps provide, such as pollination, pest control, and climate amelioration (*7*, *8*, *11*). Scientists who deny that the current mass extinction has dire consequences, and policy-makers who listen to them, fail to appreciate the penalties human civilization will suffer for continuing on society's business-as-usual course (*2*–*5*). Moreover, beyond the consequences to humans, exterminating most of the only known living things with which we share the universe is clearly wrong (*5*–*8*, *12*). The future of life on Earth, and human well-being, depends on the actions that we take to reduce the extinction of populations and species in the next two decades (*8*). It is irresponsible and unethical not to act despite the overwhelming scientific evidence indicating the severity of the current mass extinction event.

Gerardo Ceballos[1]* and Paul R. Ehrlich[2]
[1]Instituto de Ecologia, Universidad Nacional Autonoma de Mexico, Mexico City, DF 04510, Mexico. [2]Center for Conservation Biology, Department of Biology, Stanford University, Stanford, CA 94305, USA.
*Corresponding author.
Email: gceballo@ecologia.unam.mx

REFERENCES

1. W. J. Ripple *et al.*, *Bioscience* **67**, 197 (2017).
2. A. Hallam, P. B. Wignall, *Mass Extinctions and Their Aftermath* (Oxford University Press, UK. 1997).
3. S. Brand, "Rethinking extinction" (2015); https://aeon.co/essays/we-are-not-edging-up-to-a-mass-extinction.
4. C. D. Thomas, *Inheritors of the Earth* (Hachette, UK, 2017).
5. S. L. Pimm *et al.*, *Science*, **344**, 1246752 (2014).
6. G. Ceballos *et al.*, *Sci. Adv.* **1**, e1400253 (2015).
7. R. Dirzo *et al.*, *Science* **345**, 401 (2014).
8. G. Ceballos *et al.*, *The Annihilation of Nature: Human Extinction of Birds and Mammals* (JHU Press, 2015).
9. D. Jablonski, *Evol. Biol.* **44**, 451 (2017).
10. A. D. Barnosky *et al.*, *Nature* **471**, 51 (2011).
11. C. A. Hallmann, *PLOS One* **12**, e0185809 (2017).
12. P. R. Ehrlich, A. H. Ehrlich, *Proc. R. Soc. B* **280**, 20122845 (2013).

10.1126/science.aau0191

Injustices of foreign investment in coal

After remaining flat from 2014 to 2016, global greenhouse gas emissions in 2017 increased to a record high (*1*). One contributing factor is foreign investments in coal. Many countries are working to reduce their carbon footprints within their borders, but adding to emissions by investing in coal-based power projects elsewhere in the world. According to the Natural Resources Defense Council (NRDC), between 2007 and 2015, G20 countries financed US$76 billion worth of overseas coal projects (*2*). China, Japan, Germany, and South Korea were responsible for US$62 billion of the total (*2*).

After the 2015 UN Climate Change Conference (COP21), China, along with other nations, committed to cut greenhouse gas emissions. As promised, China has taken steps to substantially reduce coal use over the past few years (*3*). However, China's government has continued to finance coal-based power projects beyond its borders, especially in countries where environmental regulations and laws are weak. In 2016 alone, China financed coal projects worth US$6.3 billion in Egypt, Bangladesh, Pakistan, and Indonesia (*4*). In Pakistan, China plans to install coal-based power plants with a capacity of at least 7800 MW as part of the China-Pakistan Economic Corridor (CPEC) project (*5*). This investment includes excavation of coal-lignite from the Thar Desert of Pakistan (*6*). Coal-lignite emits about 1100 grams CO_2 per kilowatt-hour, compared with natural gas, which emits 150 to 430 grams per kilowatt-hour (*7*).

Between 1994 and 2012, Pakistan's greenhouse gas emissions increased at an annual rate of 4.1% (*8*), and the lignite-based power plans will likely increase the rate further. China's investments in coal-based energy plants, in the guise of producing energy, could have serious impacts on air and water quality and consequently serious repercussions to adjacent ecosystems, as well as to human health. The costs and consequences of such coal-based projects can stretch over decades, and they can trap developing nations in a system of carbon-intensive energy use. Pakistan, a country rated seventh most vulnerable to the impacts of climate change (*9*), must rethink its future in the energy sector.

Safdar Bashir,[1,2] Nabeel Khan Niazi,[2,3] Muhammad Arif Watto[4]*
[1]Institute of Soil and Environmental Sciences, University of Agriculture Faisalabad, Faisalabad 38040, Pakistan. [2]Sub-campus Depalpur, Okara, University of Agriculture Faisalabad, Pakistan. [3]Southern Cross GeoScience, Southern Cross University, Lismore, 2480, NSW, Australia. [4]Institute of Agricultural Extension and Rural Development, University of Agriculture Faisalabad, Faisalabad 38040, Pakistan.
*Corresponding author. Email: arifwattoo@ymail.com

REFERENCES

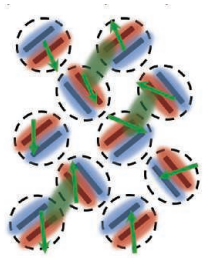
1. T. Buck, L. Hornby, "China recovery pushes greenhouse emissions to global record," *Financial Times* (2018); www.ft.com/content/ba4212b6-c63f-11e7-a1d2-6786f39ef675.
2. NRDC, "Carbon trap: How international coal finance undermines the Paris Agreement" (NRDC, 2016), p. 9.
3. C. Wang, F. Wang, *Science* **357**, 764 (2017).
4. BP Statistical Review of World Energy (2017); www.bp.com/content/dam/bp/en/corporate/pdf/energy-economics/statistical-review-2017/bp-statistical-review-of-world-energy-2017-full-report.pdf.
5. B. Walker, "China stokes global coal growth," chinadialogue (2016); www.chinadialogue.net/article/show/single/en/9264-China-stokes-global-coal-growth.
6. "Pakistan's biggest private-sector firm bets on a fabled coal mine," *The Economist* (2018); www.economist.com/business/2018/02/03/pakistans-biggest-private-sector-firm-bets-on-a-fabled-coal-mine.
7. F. Pearce, "On the road to green energy, Germany detours on dirty coal," *Yale Environment 360* (2014); https://e360.yale.edu/features/on_the_road_to_green_energy_germany_detours_on_dirty_coal.
8. K. A. Mir *et al.*, *Environ. Sci. Pollut. Res.* **24**, 35 (2017).
9. S. Kreft, D. Eckstein, I. Melchior, "Who suffers most from extreme weather events?" (Global Climate Risk Index, 2017).

10.1126/science.aat9852

RESEARCH

The makings of both a spin and a dipole liquid

Hassan et al., *p. 1101*



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

TETRAPOD EVOLUTION

Out of Antarctica

When we think of Devonian tetrapods, the ancestors of all modern vertebrates, we tend to picture amphibian-like creatures emerging from the water into a wet tropical forest or swamp. Indeed, all previously described specimens of this group have been recovered from the tropics. Gess and Ahlberg now describe two fossil tetrapods from Devonian Antarctica. Thus, the distribution of tetrapods may have been global, which encourages us to rethink the environments in which this important group was shaped. —SNV

Science, this issue p. 1120

***Tutusius mlambo*, a basal tetrapod from the Late Devonian, found in what was then the Antarctic Circle**

PLANETARY SCIENCE

Measuring martian organics and methane

The Curiosity rover has been sampling on Mars for the past 5 years (see the Perspective by ten Kate). Eigenbrode *et al*. used two instruments in the SAM (Sample Analysis at Mars) suite to catch traces of complex organics preserved in 3-billion-year-old sediments. Heating the sediments released an array of organics and volatiles reminiscent of organic-rich sedimentary rock found on Earth. Most methane on Earth is produced by biological sources, but numerous abiotic processes have been proposed to explain martian methane. Webster *et al*. report atmospheric measurements of methane covering 3 martian years and found that the background level varies with the local seasons. The seasonal variation provides an important clue for determining the origin of martian methane. —BG and KTS

Science, this issue p. 1096, p. 1093; see also p. 1068

APPLIED OPTICS

Making silicon shine bright

Silicon is the workhorse of the semiconductor electronics industry, but its lack of optical functionality is a barrier to developing a truly integrated silicon-based optoelectronics platform. Although there are several ways of exploiting nonlinear light-matter interactions to coax silicon into optical functionality, the effects tend to be weak. Otterstrom *et al*. used a suspended silicon waveguide racetrack structure to stimulate the stronger nonlinear effect of Brillouin scattering and achieve lasing from silicon. The ability to engineer the nonlinearity and tune the optical response through the design of the suspended cavity provides a powerful and flexible route for developing silicon-based optoelectronic circuits and devices. —ISO

Science, this issue p. 1113

SOCIAL SCIENCE

Tipping points in social convention

Once a population has converged on a consensus, how can a group with a minority viewpoint overturn it? Theoretical models have emphasized tipping points, whereby a sufficiently large minority can change the societal norm. Centola *et al*. devised a system to study this in controlled experiments. Groups of people who had achieved a consensus about the name of a person shown in a picture were individually exposed to a confederate who promoted a different name. The only incentive was to coordinate. When the number of confederates was roughly 25% of the group, the opinion of the majority could be tipped to that of the minority. —BJ

Science, this issue p. 1116

MEDICINE

Toward more predictable birthdays

Low-cost methods for monitoring fetal development

CREDITS: (TOP TO BOTTOM) N. HASSAN *ET AL.*; MAGGIE LAMBERT-NEWMAN

Published by AAAS

could improve prenatal care, especially in low-resource settings. By measuring the levels of certain placental RNA transcripts in maternal blood, Ngo *et al*. developed two noninvasive blood tests that provide a window into the progression of individual pregnancies. In a small proof-of-concept study, the first blood test predicted fetal age and delivery date with an accuracy comparable to that of ultrasound. The second blood test, also examined in a small pilot study, discriminated women at risk of preterm delivery from those who delivered at full term. The next step will be to assess the reliability of the tests in large, blinded clinical trials. —PAK

Science, this issue p. 1133

PLANT GENETICS

Sterility in rice via toxin and antidote

Crossing wild and domestic rice often results in hybrid sterility. Such genetic barriers can prevent the movement of potentially beneficial genes from wild rice into domestic varieties. To understand the barriers preventing gene flow, Yu *et al*. mapped a quantitative trait locus (QTL) that determines sterility between wild-type and domestic rice. This QTL encodes two open reading frames (ORFs) that are both expressed during gametogenesis. The ORFs encode a toxin, which affects the development of pollen, and an antidote, which is required for pollen viability. Thus, selfish genetic elements can underlie evolutionary strategies that facilitate reproductive isolation. —LMZ

Science, this issue p. 1130

Hybrid rice varieties produce sterile pollen.

PHOTOS: (LEFT TO RIGHT) X. YU *ET AL*.; KEVIN SCHAFER/ALAMY STOCK PHOTO

ENVIRONMENTAL STUDIES

Economic rationale for fishing the high seas

Economic evaluations of high-seas fishing have been lacking, in part owing to the scarcity of data on the costs and revenues of fleets that fish in these elusive waters. Sala *et al.* wanted to quantify high-seas fishing efforts globally and assess whether and when high-seas fishing makes economic sense. They used satellite data and machine learning to track the activity of more than 3600 fishing vessels in near real time. Patterns of fishing profitability varied widely between countries, types of fishing, and distance to port. As much as 54% of present high-seas fishing grounds would be unprofitable without large government subsidies, supporting recent calls for subsidy management reforms for the high seas. —PJB

Sci. Adv. 10.1126/sciadv.aat2504 (2018).

DNA REPAIR

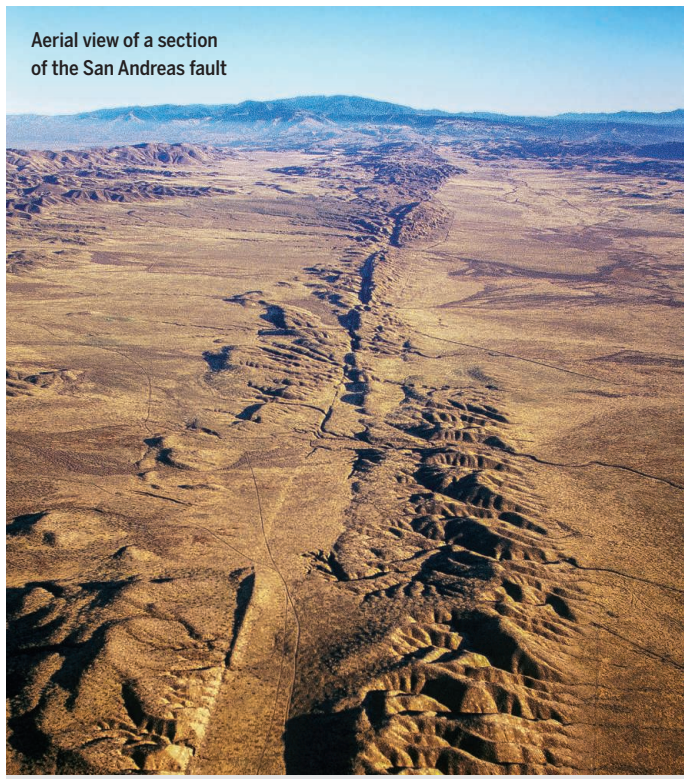
DNA-bound ubiquitin coordinates repair

Ubiquitylation is a posttranslational modification that reversibly alters various protein properties. Liu *et al*. discovered that Lys^{63}-linked polyubiquitin chains bound to the free ends of double-stranded DNA, bridged the broken ends of DNA, and recruited repair proteins. Ubiquitins with DNA-binding motif mutations were found in several types of tumors. When expressed in cultured cells, these mutant ubiquitins impaired the cellular response to DNA-damaging agents, suggesting that it might be possible to therapeutically exploit these mutations in some cancer patients. —LKF

Sci. Signal. **11**, eaar8133 (2018).

IN OTHER JOURNALS

Edited by **Sacha Vignieri** *and* **Jesse Smith**



Aerial view of a section of the San Andreas fault

GEOPHYSICS

The San Andreas creeps along the decade

The famous San Andreas fault in California is an excellent place to understand the behavior of faults. Khoshmanesh and Shirzaei used high-resolution satellite measurements to track surface deformation along the central portion of the fault over two decades. Accurate modeling of the deformation requires shifts in the time scale of the fault's aseismic creeping behavior from yearly to decadal. The different modes of creep are important for assessing seismic hazard and may provide some clues about fault rupture. —BG

Geophys. Res. Lett. 10.1002/2018GL077017 (2018).

MOLECULAR BIOLOGY

The long and short of RNA export

Circular RNAs (circRNAs) are back-spliced RNA products that have regulatory roles in gene expression, and most circRNAs are enriched in the cytoplasm. Huang *et al*. identified protein factors that are required to export circRNAs from the nucleus to the cytoplasm. An RNA interference screen targeting some proteins that are known to export various linear RNAs in *Drosophila* cells showed that a RNA helicase is required for the cytoplasmic accumulation of circRNAs that are longer than 800 nucleotides. The two human homologs of this *Drosophila* helicase play similar roles in human cells. Surprisingly, one controls long (>1300-nucleotide)

Human-induced changes in Lake Baikal have resulted in the spread of a previously rare filamentous alga.

ECOLOGICAL INVASIONS

An algal transformation in Lake Baikal

In central Siberia, Lake Baikal—the world's deepest freshwater lake—is undergoing ecological change. Volkova *et al.* report that species of the filamentous mat-forming alga *Spirogyra*, formerly present at low abundance in restricted areas of the lake, have proliferated and diversified in recent years. *Spirogyra* has invaded shallow-water zones of the lake, and several species have appeared in Baikal for the first time. These changes are likely to have been driven by a combination of factors, some of them anthropogenic, such as agricultural runoff and changing patterns of seasonal temperatures. The consequences may be hard to predict and control. —AMS

Phycologia **57**, 298 (2018).

and the other controls short (<350-nucleotide) circRNAs. Future investigation of pathways that measure and export circRNAs of different lengths will shed light on circRNA functions. —SYM

Genes Dev. 10.1101/gad.314856.118 (2018).

DNA METHYLATION

Tissue-specific DNA demethylation after birth

Because of changes in their environment, including the need to repair tissue, cells cannot remain static. They must do their job even if conditions around them vary. One way to accommodate change in gene expression is through DNA methylation. The vast majority of this modification takes place during mammalian embryogenesis. Initially, methyl groups are removed around implantation, and then de novo methylation occurs in specific tissues and at set times during cell differentiation. Using high-throughput and genetic analysis, Reizel *et al*. show that considerable postnatal demethylation also occurs. For example, hormone signaling triggers DNA demethylation at enhancer-like regions in the liver after birth in mice. These epigenetic changes give access to specific chromatin sites for proper hepatocyte gene expression and function. —BAP

Nat. Comm. 10.1038/s41467-018-04456-6 (2018).

PROTEIN STRUCTURE

Not just a LARK

Many proteins involved in neurodegenerative diseases contain low-complexity domains (LCDs) that frequently exhibit no secondary structure but are implicated in both functional reversible aggregation and pathological irreversible aggregation. Guenther *et al*. determined the structures of 10 peptide segments from the LCD domain of the RNA binding protein TDP-43. The LCD is implicated in the formation of stress granules, which disaggregate when stress is relieved, and of pathogenic amyloid fibrils. Six of the segments form structures characteristic of amyloid fibrils, whereas four form labile amyloid-like interactions, termed LARKs. Disease variants of TDP-43 convert the LARKs to irreversible aggregates. This raises the possibility that either mutagenesis or protein modifications such as phosphorylation may play a role in switching between functional and pathological aggregation. —VV

Nat. Struct. Mol. Biol. 10.1038/s41594-018-0064-2 (2018).

GENETICS

Genomic crowdsourcing with privacy

Privacy concerns can be an obstacle to large-scale donation of the genetic material that is essential for understanding how genetic variants are associated with disease. Cho *et al.* have developed an approach that has the potential to be scalable to as many as a million genomes. Their system is based on dividing genotype and phenotype data from each individual among multiple servers in such a way that no one server can infer the original data. An attacker would have to hack all of the servers to extract the data. Computational approaches that simplified corrections for misleading correlations and sped up the system were used to reproduce three published genome-wide association studies representing 23,000 genomes and could make the process cost-effective. —BJ

Nat. Biotechnol. 10.1038/nbt.4108 (2018).

CLIMATE CHANGE IMPACTS

Hurting the most vulnerable

The Paris Agreement set a goal of limiting global average anthropogenic climate warming to 1.5°C or less—cooler than the oft-cited target of 2°C that previously had been considered the threshold of dangerous consequences. Is the difference in impacts between the two goals meaningful, and who would be affected most by a failure to stay within the lower limit? King and Harrington find that exceeding the 1.5°C ceiling would affect tropical latitudes—which have a disproportionate share of poor populations—more than higher latitudes. Thus, if ways are not found to meet the global warming targets of the Paris Agreement, then the most vulnerable will experience the greatest consequences. —HJS

Geophys. Res. Lett. 10.1029/2018GL078430 (2018).

PHOTO: ANDREY NEKRASOV/ALAMY STOCK PHOTO

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

GREAT APE GENOMICS

A spotlight on great ape genomes

Most nonhuman primate genomes generated to date have been "humanized" owing to their many gaps and the reliance on guidance by the reference human genome. To remove this humanizing effect, Kronenberg *et al*. generated and assembled long-read genomes of a chimpanzee, an orangutan, and two humans and compared them with a previously generated gorilla genome. This analysis recognized genomic structural variation specific to humans and particular ape lineages. Comparisons between human and chimpanzee cerebral organoids showed down-regulation of the expression of specific genes in humans, relative to chimpanzees, related to noncoding variation identified in this analysis. —LMZ

Science, this issue p. 1085

IMMUNOLOGY

Finding a role for PNECs in asthma

Pulmonary neuroendocrine cells (PNECs) are a rare cell type located in airway and alveolar epithelia and are often in contact with sensory nerve fibers. They have a wide phylogenic distribution and are found even in the relatively primitive lungs of amphibia and reptiles, suggesting a critical function. Sui *et al*. found that mice lacking PNECs have suppressed type 2 (allergic) immune responses. PNECs were observed in close proximity to group 2 innate lymphoid cells (ILC2s) around airway branch points. The PNECs enhanced ILC2 activity by secreting CGRP (calcitonin gene-related peptide). They also induced goblet-cell hyperplasia via the neurotransmitter GABA (γ-aminobutyric acid). Interestingly, human asthma patients were found to have increased PNEC numbers, suggesting a potential therapeutic target for the treatment of asthma. —STS

Science, this issue p. 1086

METABOLISM

A missing link in cholesterol absorption

Cholesterol is important for general health, but too much can build up in artery walls and cause cardiovascular disease. Low-density lipoprotein cholesterol (LDL-C) is often referred to as "bad cholesterol"; keeping LDL-C within stringent limits is recommended to reduce the risk of heart attack and stroke. Zhang *et al*. discovered that some individuals have an inherited frameshift mutation in the *LIMA1* gene (also known as *EPLIN* or *SREBP3*). The gene has not been linked to lipid metabolism before, but altered *LIMA1* was found to maintain low plasma LDL-C by reducing the absorption of cholesterol through the intestine. Pharmacological targeting through the LIMA1 pathway might thus provide a strategy to improve heart health. —PNK

Science, this issue p. 1087

ORGANIC MATERIALS

Quantum dipoles go liquid

Quantum spin liquids do not achieve an ordered magnetic state, even at the lowest temperatures. Hassan *et al*. studied an organic compound that may be both a spin liquid and a dipole liquid (see the Perspective by Powell). In the layered material κ-$(BEDT\text{-}TTF)_2Hg(SCN)_2Br$, molecules form charged dimers whose sites are arranged on a triangular lattice. The extra charge associated with each dimer can "live" on one of the two molecules in the dimer, resulting in a nonzero electric dipole moment for the dimer. Raman spectroscopy and heat capacity measurements revealed that, like spins in a quantum spin liquid, these dimers remained disordered down to the lowest temperatures. —JS

Science, this issue p. 1101; see also p. 1073

PHOTOSYNTHESIS

Antenna switches partners in the shade

A cloudy day or an overshadowing tree causes fluctuations in light that can throw off the balance of energy flow in plant photosystems I and II (PSI and PSII). Pan *et al*. solved structures of PSI bound to two light-harvesting complexes (LHCs). One LHC is permanently associated with PSI. The other LHC delivers light energy to PSII under optimal conditions but can switch to a PSI-associated state after phosphorylation by a kinase that senses the redox environment of the chloroplast. The movement of LHCs between the photosystems helps maintain even energy flux. Two chlorophyll-containing subunits are visible in the structure that connect the PSI core to each LHC. —MAF

Science, this issue p. 1109

APPLIED OPTICS

Metasurfaces for molecular detection

Although mid-infrared (mid-IR) spectroscopy is a mainstay of molecular fingerprinting, its sensitivity is diminished somewhat when looking at small volumes of sample. Nanophotonics provides a platform to enhance the detection capability. Tittl *et al*. built a mid-IR nanophotonic sensor based on reflection from an all-dielectric metasurface array of specially designed scattering elements. The scattering elements could be tuned via geometry across a broad range of wavelengths in the mid-IR. The approach successfully detected and differentiated the absorption fingerprints of various molecules. The technique offers the prospect of on-chip molecular fingerprinting without the need for spectrometry, frequency scanning, or moving mechanical parts. —ISO

Science, this issue p. 1105

CLIMATE CHANGE

Predicting changes in "extreme" precipitation

As temperatures rise, Earth's atmosphere can hold more moisture. This rise in moisture content is expected to lead to a broadly comparable rise in the intensity of the most extreme precipitation events. In a Perspective, Pendergrass explains that whether this expectation is met depends on the definition of extreme precipitation. The intensity of the most extreme events may rise more than expected, as seen for Hurricane Harvey, whereas that of less extreme events may rise less than anticipated. Atmospheric circulation changes will also affect precipitation events in ways that differ from one place to another. Clear definitions of extreme precipitation are key to anticipating and preparing for future changes in extreme events. —JFU

Science, this issue p. 1072

PLANT SCIENCE

Defense cargo shuttles in vesicles

Plants can use small RNAs (sRNAs) to interfere with virulence factor gene expression in pathogens. Cai *et al*. show that the small mustard plant *Arabidopsis* shuttles defensive sRNAs into the necrotrophic fungus *Botrytis cinerea* via extracellular vesicles (see the Perspective by Thomma and Cook). The vesicles are associated with tetraspanin proteins, which can interact and form membrane microdomains. Several dozen different sRNAs targeting the pathogenic

process were transported from *Arabidopsis* to *B. cinerea* in a selective manner. —PJH

Science, this issue p. 1126; see also p. 1070

EVOLUTIONARY COGNITION

Understanding zero

It has been said that the development of an understanding of zero by society initiated a major intellectual advance in humans, and we have been thought to be unique in this understanding. Although recent research has shown that some other vertebrates understand the concept of the "empty set," Howard *et al*. now show that an understanding of this concept is present in untrained honey bees (see the Perspective by Nieder). This finding suggests that such an understanding has evolved independently in distantly related species that deal with complexity in their environments, and that it may be more widespread than previously appreciated. —SNV

Science, this issue p. 1124; see also p. 1069

ZIKA

Zika leaves a lasting impact on the brain

Perinatal Zika virus (ZIKV) infection has been associated with brain alterations in newborns. However, whether ZIKV exposure during development has long-term neurological consequences is not completely understood. Nem de Oliveira Souza *et al*. report that newborn mice infected with ZIKV developed acute brain abnormalities. During adulthood, perinatally infected mice showed persistent viral replication, neuropathological alterations, behavioral impairments, and altered brain excitability. Blocking tumor necrosis factor–α early after infection prevented this hyperexcitability in the mouse brain. Thus, anti-inflammatory treatments might help to prevent the persistent increase in neuronal excitability induced by ZIKV infection in brain tissue. —MM

Sci. Transl. Med. **10**, eaar2749 (2018).

EMBRYOGENESIS

Modeling embryogenesis

Understanding the molecular and cellular events of early embryogenesis is crucial to improve assisted reproductive technologies and prevent genetic birth defects. Although some clarity on this process in humans has come from comparative studies in mice, differences exist, so there is a need to better model embryogenesis. In a Perspective, Rossant and Tam discuss the possibility of using nonhuman primate embryos and human stem cell–derived models to better study and understand early development. —GKA

Science, this issue p. 1075

HIV

Taking residence to defend

In HIV$^+$ individuals receiving antiretroviral therapy, CD4$^+$ T cells home to lymphoid tissues (LTs) that are a key site of HIV persistence. Studying the immune response to HIV in LTs has been a challenge. Buggert *et al.* obtained LTs from HIV$^+$ individuals and carried out comprehensive transcriptional and epigenetic analyses on CD8$^+$ T cells found there. The CD8$^+$ T cells had a signature associated with resident memory T cells. The frequency of these HIV-responsive LT-resident CD8$^+$ T cells was considerably increased in so-called elite controllers—people able to restrain their HIV infections. —AB

Sci. Immunol. **3**, eaar4526 (2018).

RESEARCH ARTICLE SUMMARY

GREAT APE GENOMICS

High-resolution comparative analysis of great ape genomes

Zev N. Kronenberg, Ian T. Fiddes*, David Gordon*, Shwetha Murali*, Stuart Cantsilieris*, Olivia S. Meyerson*, Jason G. Underwood*, Bradley J. Nelson*, Mark J. P. Chaisson, Max L. Dougherty, Katherine M. Munson, Alex R. Hastie, Mark Diekhans, Fereydoun Hormozdiari, Nicola Lorusso, Kendra Hoekzema, Ruolan Qiu, Karen Clark, Archana Raja, AnneMarie E. Welch, Melanie Sorensen, Carl Baker, Robert S. Fulton, Joel Armstrong, Tina A. Graves-Lindsay, Ahmet M. Denli, Emma R. Hoppe, PingHsun Hsieh, Christopher M. Hill, Andy Wing Chun Pang, Joyce Lee, Ernest T. Lam, Susan K. Dutcher, Fred H. Gage, Wesley C. Warren, Jay Shendure, David Haussler, Valerie A. Schneider, Han Cao, Mario Ventura, Richard K. Wilson, Benedict Paten, Alex Pollen, Evan E. Eichler†

INTRODUCTION: Understanding the genetic differences that make us human is a long-standing endeavor that requires the comprehensive discovery and comparison of all forms of genetic variation within great ape lineages.

RATIONALE: The varied quality and completeness of ape genomes have limited comparative genetic analyses. To eliminate this contiguity and quality disparity, we generated human and nonhuman ape genome assemblies without the guidance of the human reference genome. These new genome assemblies enable both coarse and fine-scale comparative genomic studies.

RESULTS: We sequenced and assembled two human, one chimpanzee, and one orangutan genome using high-coverage (>65x) single-molecule, real-time (SMRT) long-read sequencing technology. We also sequenced more than 500,000 full-length complementary DNA samples from induced pluripotent stem cells to construct de novo gene models, increasing our knowledge of transcript diversity in each ape lineage. The new nonhuman ape genome assemblies improve gene annotation and genomic contiguity (by 30- to 500-fold), resulting in the identification of larger synteny blocks (by 22- to 74-fold) when compared to earlier assemblies. Including the latest gorilla genome, we now estimate that 83% of the ape genomes can be compared in a multiple sequence alignment.

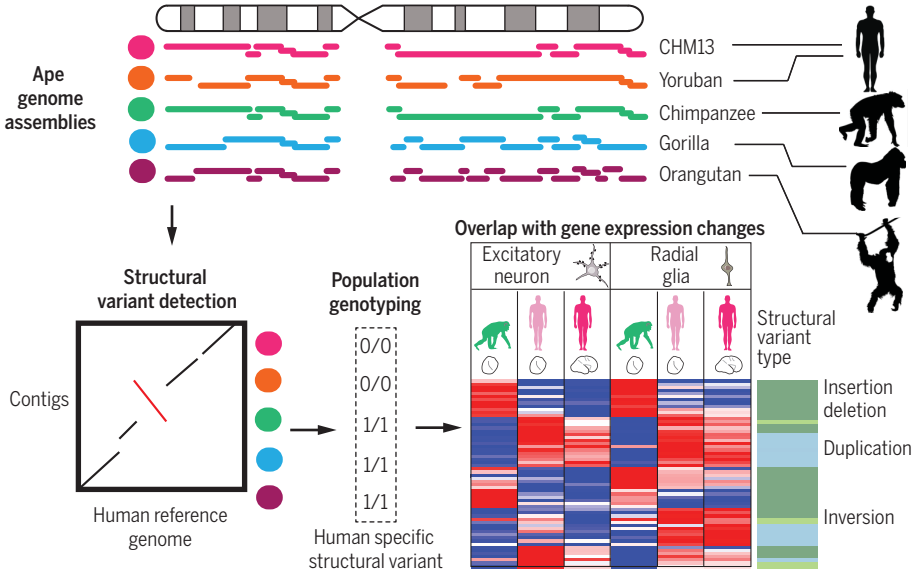
We observe a modest increase in single-nucleotide variant divergence compared to previous genome analyses and estimate that 36% of human autosomal DNA is subject to incomplete lineage sorting. We fully resolve most common repeat differences, including full-length retrotransposons such as the African ape-specific endogenous retroviral element PtERV1. We show that the spread of this element independently in the gorilla and chimpanzee lineage likely resulted from a founder element that failed to segregate to the human lineage because of incomplete lineage sorting.

The improved sequence contiguity allowed a more systematic discovery of structural variation (>50 base pairs in length) (see the figure). We detected 614,186 ape deletions, insertions, and inversions, assigning each to specific ape lineages. Unbiased genome scaffolding (optical maps, bacterial artificial chromosome sequencing, and fluorescence in situ hybridization) led to the discovery of large, unknown complex inversions in gene-rich regions. Of the 17,789 fixed human-specific insertions and deletions, we focus on those of potential functional effect. We identify 90 that are predicted to disrupt genes and an additional 643 that likely affect regulatory regions, more than doubling the number of human-specific deletions that remove regulatory sequence in the human lineage. We investigate the association of structural variation with changes in human-chimpanzee brain gene expression using cerebral organoids as a proxy for expression differences. Genes associated with fixed structural variants (SVs) show a pattern of down-regulation in human radial glial neural progenitors, whereas human-specific duplications are associated with up-regulated genes in human radial glial and excitatory neurons (see the figure).

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aar6343

CONCLUSION: The improved ape genome assemblies provide the most comprehensive view to date of intermediate-size structural variation and highlight several dozen genes associated with structural variation and brain-expression differences between humans and chimpanzees. These new references will provide a stepping stone for the completion of great ape genomes at a quality commensurate with the human reference genome and, ultimately, an understanding of the genetic differences that make us human. ■



SMRT assemblies and SV analyses. (Top) Contiguity of the de novo assemblies. (Bottom, left to right) For each ape, SV detection was done against the human reference genome as represented by a dot plot of an inversion). Human-specific SVs, identified by comparing ape SVs and population genotyping (0/0, homozygous reference), were compared to single-cell gene expression differences [range: low (dark blue) to high (dark red)] in primary and organoid tissues. Each heatmap row is a gene that intersects an insertion or deletion (green), duplication (cyan), or inversion (light green).

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: eee@gs.washington.edu
Cite this article as Z. N. Kronenberg *et al.*, *Science* **360**, eaar6343 (2018). DOI: 10.1126/science.aar6343

RESEARCH ARTICLE

GREAT APE GENOMICS

High-resolution comparative analysis of great ape genomes

Zev N. Kronenberg[1], Ian T. Fiddes[2]*, David Gordon[1,3]*, Shwetha Murali[1,3]*, Stuart Cantsilieris[1]*, Olivia S. Meyerson[4]*, Jason G. Underwood[1,5]*, Bradley J. Nelson[1]*, Mark J. P. Chaisson[1,6], Max L. Dougherty[1], Katherine M. Munson[1], Alex R. Hastie[7], Mark Diekhans[2], Fereydoun Hormozdiari[8], Nicola Lorusso[9], Kendra Hoekzema[1], Ruolan Qiu[1], Karen Clark[10], Archana Raja[1,3], AnneMarie E. Welch[1], Melanie Sorensen[1], Carl Baker[1], Robert S. Fulton[11], Joel Armstrong[2], Tina A. Graves-Lindsay[11], Ahmet M. Denli[12], Emma R. Hoppe[1], PingHsun Hsieh[1], Christopher M. Hill[1], Andy Wing Chun Pang[7], Joyce Lee[7], Ernest T. Lam[7], Susan K. Dutcher[11], Fred H. Gage[12], Wesley C. Warren[11], Jay Shendure[1,3], David Haussler[2,13], Valerie A. Schneider[10], Han Cao[7], Mario Ventura[9], Richard K. Wilson[11], Benedict Paten[2], Alex Pollen[4,14], Evan E. Eichler[1,3]†

Genetic studies of human evolution require high-quality contiguous ape genome assemblies that are not guided by the human reference. We coupled long-read sequence assembly and full-length complementary DNA sequencing with a multiplatform scaffolding approach to produce ab initio chimpanzee and orangutan genome assemblies. By comparing these with two long-read de novo human genome assemblies and a gorilla genome assembly, we characterized lineage-specific and shared great ape genetic variation ranging from single– to mega–base pair–sized variants. We identified ~17,000 fixed human-specific structural variants identifying genic and putative regulatory changes that have emerged in humans since divergence from nonhuman apes. Interestingly, these variants are enriched near genes that are down-regulated in human compared to chimpanzee cerebral organoids, particularly in cells analogous to radial glial neural progenitors.

Scientists have long been interested in the functional genetic differences that distinguish humans from other ape species (*1*). Human and chimpanzee protein-encoding changes and structural differences in regulatory DNA or in the copy number of gene families have all been implicated in adaptation (*2*, *3*). Indeed, several potentially high-impact regulatory changes (*4*, *5*) and human-specific genes (*6*–*9*) that are important in synapse density, neuronal count, and other morphological differences have been identified. Most of these genetic differences, however, were not initially recognized upon comparison of human and ape genomes because the genetic changes mapped to regions of rapid genomic structural change that were not resolved in draft genome assemblies.

Despite recent efforts to sequence and assemble ape genomes (*10*–*12*), our understanding of structural differences, and particularly those specific to the human lineage, remains far from complete. There are two fundamental problems. First, there is considerable heterogeneity in the contiguity of ape genome assemblies. The presence of tens to hundreds of thousands of gaps in ape genomes limits the proportion of the genome that can be compared in a multispecies sequence alignment. Therefore, a large fraction of human-specific insertions and deletions, including those that alter regulatory sequences, are not resolved. Second, the higher-quality human genome assemblies have often been used to guide the final stages of nonhuman genome projects, including the order and orientation of sequence contigs and, perhaps more importantly, the annotation of genes. This bias has effectively "humanized" other ape genome assemblies, minimizing potential structural and transcript differences observed between the species. Using long-read, long-range sequence and mapping technologies (*13*–*15*), we generated new great ape genome assemblies, along with full-length cDNA annotation, without guidance from the human genome. We also generated and analyzed an African genome and an effectively haploid human genome complement to distinguish fixed differences in the human ancestral lineage and to further mitigate human genome reference biases.

Results

Genome assembly

We sequenced two human, one chimpanzee, and one orangutan genome to high depth (>65-fold coverage) using single-molecule, real-time (SMRT; PacBio) sequence data and assembled each ab initio using the same underlying assembly algorithm (Table 1) (*16*). For each species, we generated assemblies ranging from 2.9 to 3.1 giga–base pairs (Gbp) in size, where most of the euchromatic DNA mapped to <1000 large contigs (Table 1). We error-corrected sequence contigs with Quiver (*17*) and Pilon (*18*), followed by a procedure that reduced the remaining 1- to 2-bp indels (insertions or deletions) specifically in regions with clustered single-nucleotide variants (SNVs) (*16*). We next scaffolded the chimpanzee and orangutan genomes without guidance from the human reference genome. In total, 93% (2.79 Gbp, excluding chromosome Y) of the chimpanzee-assembled bases and 92.7% (2.82 Gbp) of the orangutan-assembled bases were incorporated into chromosomal-level scaffolds (Table 1). We confirmed most large-scale chromosomal inversions among the great apes (*19*), some of which were absent from previous assemblies.

Sequence accuracy and quality assessment

More than 96% of our assembled sequence was concordant by length and orientation by different metrics (Table 1) (*16*). We conservatively estimate that these assemblies have improved contiguity for the chimpanzee and orangutan genomes by 32- and 533-fold, respectively (Fig. 1, A and B). Consistent with the gorilla genome (*20*), the application of long-read sequence data closed most of the genome gaps in earlier assemblies. The extent of the change varied, however, depending on the prior level of finishing. In the case of the chimpanzee, 52% of the remaining 27,797 gaps were closed. We added 6.9 Mbp of new sequence and removed at least 27.2 Mbp of duplicated or extraneous sequence, possibly artifacts of scaffolding and gap filling (*21*). In the case of the orangutan, we added 54.5 Mbp of sequence while removing 4.2 Mbp, closing an estimated 96.8% (305,069/315,124) of the remaining euchromatic gaps. We determined the sequence contigs to be highly accurate at the base-pair level (>99.9%) on the basis of comparisons of each genome to Sanger end-sequence data, completely sequenced clone inserts, and Illumina whole-genome sequencing data generated from the same source individuals (Table 1) (*16*).

[1]Department of Genome Sciences, University of Washington School of Medicine, Seattle, WA 98195, USA. [2]UC Santa Cruz Genomics Institute, University of California, Santa Cruz, Santa Cruz, CA 95064, USA. [3]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA. [4]Department of Neurology, University of California, San Francisco, San Francisco, CA 94158, USA. [5]Pacific Biosciences (PacBio) of California, Inc., Menlo Park, CA 94025, USA. [6]Computational Biology and Bioinformatics, University of Southern California, Los Angeles, CA 90089, USA. [7]Bionano Genomics, San Diego, CA 92121, USA. [8]Department of Biochemistry and Molecular Medicine, University of California, Davis, Davis, CA 95817, USA. [9]Department of Biology, University of Bari, Aldo Moro, Bari 70121, Italy. [10]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. [11]Departments of Medicine and Genetics, McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO 63108, USA. [12]The Salk Institute for Biological Studies, La Jolla, CA 92037, USA. [13]Howard Hughes Medical Institute, University of California, Santa Cruz, Santa Cruz, CA 95064, USA. [14]Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, San Francisco, San Francisco, CA 94143, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: eee@gs.washington.edu

Gene annotation

Nonhuman primate (NHP) genome assemblies have typically relied almost exclusively on the human reference to define gene models (table S1). To provide a less biased source of gene annotation, we generated long-read transcriptome sequencing data to produce an average of 658,546 full-length nonchimeric (FLNC) transcripts from induced pluripotent stem cells (iPSCs) derived from each of the three nonhuman ape lineages (*16*). We selected iPSC material to maximize transcript diversity and enrich for early developmental genes. We next annotated the genomes of chimpanzee, gorilla, and orangutan, using FLNC transcripts along with short-read RNA-sequencing (RNA-seq) to guide gene and previously undescribed isoform predictions (*22*).

The number of genes and most gene models (coding and noncoding, including long noncoding RNA) are consistent among the different ape genomes (Table 2). However, we saw differential mapping of FLNC transcripts that favored the SMRT assemblies, especially in repeat-rich transcripts (Fig. 1C). Concordantly, human transcript models (GENCODE V27) aligned better to SMRT assemblies (Fig. 1D). For chimpanzee, 17,744 human protein-coding transcript models showed an increase of mapping coverage, which averaged 5.6%. This pattern was more pronounced in orangutan, where 28,033 of the 91,578 protein-coding transcript annotations showed an average improvement of 5.7% in mapping coverage. Overall, human protein-coding transcript models mapped to chimpanzee and orangutan SMRT assemblies with 99.1 and 98.8% average coverage, respectively—1.5 and 2.5% improvements. These improvements stemmed largely from gap closures, which rescue missing exons and recover more full-length transcripts, including untranslated regions (UTRs).

We identified a small fraction (~1.5%) of putative protein-encoding genes present among NHPs that were absent in human annotations (GENCODE V27). In addition, a larger fraction (3.1 to 3.8%) of transcripts exhibited RNA-seq– or isoform sequencing (Iso-Seq)–supported splice junctions present in NHPs but not in human transcripts. Finally, we evaluated the NHP annotations, identifying full exons that affect coding sequences, which have been gained or lost between humans and other great apes (table S1).

Comparative sequence analyses

We constructed a five-way genome-wide multiple sequence alignment (MSA) of the ab initio assembled genomes (Table 1) by identifying syntenic (20 kbp) blocks against the human reference genome. In total, 83% of the ape genome was represented in MSAs. This allowed us to identify a comprehensive set of SNVs, indels, and structural variants (SVs); calculate divergence; and perform genome-wide phylogenetic analyses (Fig. 2). We observed a modest elevation in SNV divergence compared to previous genome comparisons (Fig. 2A and table S2) and estimated that 35.6% of the human genome is subject to incomplete lineage sorting among the African apes (Fig. 2B). Human and chimpanzee branch lengths are remarkably similar within coding regions (0.026% difference in branch length); however, we observed a 3.5% slowdown of the human mutation rate in noncoding regions (*23*, *24*) (Fig. 2C). Human and chimpanzee branch lengths were considerably shorter compared to the other apes, consistent with the hominid slowdown hypothesis (*25*).

Repeat comparisons

Although the general repeat content of primate genomes has been well established (*16*), the longest and most complex repetitive regions have been more difficult to assay. Because long-read sequence data resolve most microsatellites and high-copy interspersed repeats (*20*, *26*), we focused on comparative analysis of short tandem repeats (STRs) and endogenous retrovirus elements. Previous studies have suggested differential expansion of STR sequences between humans and other NHPs (*27*, *28*). However, these studies suffer from ascertainment bias owing to methodological differences in genome sequencing or STR enrichment, differential access to GC-rich regions, and discovery bias in the human reference genome.

We analyzed each genome independently and, after clustering STRs that mapped within 25 bp, identified a consistent number of STRs per ape genome (344,354 to 358,622 STR regions; table S3). Because STRs often map within or adjacent to other classes of repetitive DNA, we restricted our analysis to the subset where orthology and STR lengths were clearly defined (12,694 to 16,138 STRs; fig. S28 and table S4). The average length difference between human and chimpanzee STR loci is 0.02 bp, with only a slight difference in distributions [$P = 0.015$, Kolmogorov-Smirnov (KS) test; table S5 and Fig. 2D]. Other ape comparisons show a modest increase in overall STR length (for example, a 1.2-bp average increase in gorilla versus chimpanzee; $P = 8.76 \times 10^{12}$, KS test). We found no significant difference between human and chimpanzee STR length in coding sequences ($n = 2199$, $P = 0.28$, KS test) or UTRs of genes ($n = 2794$, $P = 0.16$, KS test), although we identified 4920 loci preferentially

Table 1. Assembly statistics for the great ape genomes. QV, quality value score; AGP, a golden path assembly; ND, no data.

Statistic	Ape assembly				
	CHM13_HSAv1* (human)	YRI_HSAv1 (human)	Clint_PTRv1 (chimpanzee)	GSMRT3.2 (gorilla)	Susie_PABv1 (orangutan)
Estimated depth†	72	116	117	86.3	94.9
Subread length N50 (kbp)‡	16.2	13.4	17.4	18.6	16.6
Contigs, initial and final counts§	1,923 and 1,916	3,645 and 3,642	4,912 and 5,037	15,997	5,771 and 5,813
Assembly size (Gbp)	2.88	2.88	2.99	3.08	3.04
Contig length >3 Mbp (Gbp)	2.65	2.27	2.45	2.42	2.48
Initial and final contig N50 lengths (Mbp)§	29.26	6.60	12.76 and 12.42	10.02	11.27 and 11.07
Scaffold N50 (Mbp)	83.02	ND	53.1	ND	98.47
Longest contig (Mbp)‖	81	27	80	36	53
BAC concordance (%)	97.11	97.73	99.13	96.85	96.75
Sequence accuracy (QV)	36	31	33 to 38	30 to 38	28 to 33
Iso-Seq transcripts	710,974	ND	565,691	881,801	528,145
Contigs in AGP	ND	ND	685	794	544
Contigs aligned to GRCh38¶ (length in Gbp)	407 (2.8)	1,167 (2.8)	656 (2.8)	907 (2.8)	524 (2.8)

*Haploid genome assembly derived from a complete hydatidiform mole. †Estimated coverage in raw SMRT subreads based on 3.5-Gbp (gorilla) or 3.2-Gbp (all others) estimated genome size. ‡N50 subread lengths of raw input data. §Initial and final contigs before and after resolving chimeras by optical map comparison; note that optical maps are not available for each assembly. These stats do not consider the National Center for Biotechnology Information (NCBI) minimum–contig length filter. ‖Longest contig without gross assembly error. ¶Contigs with less than 95% of sequence aligning to GRCh38, depth-of-coverage greater than two SDs above the mean, or no coverage were excluded.

expanded in the human lineage (table S6), including loci associated with genomic instability and disease.

Endogenous retroelements are among the longest retrotransposons within mammalian genomes (up to 10 kbp) and are frequently misassembled because of their copy number and sequence identity. The chimpanzee and gorilla lineages carry an endogenous retrovirus, PtERV1, that is absent in orangutan and human genomes (*29*, *30*). None of the PtERV1 integrations between chimpanzees and gorillas appear orthologous, suggesting either that independent retroviral integrations occurred in these two lineages (*29*, *30*) or that humans and orangutans contain extrinsic factors that differentially restricted propagation (*31*). A high-quality map of 540 PtERV1 elements [both full-length and solo long terminal repeat (LTR)] in chimpanzee and gorilla (table S7) (*16*) shows that their integration events are nonorthologous (99.8%), biased against genes, and integrated in the antisense orientation (figs. S30 and S31), consistent with the action of purifying selection.

Using the more complete ape genomes, we identified only one chimpanzee-gorilla orthologous PtERV1 element, not present in modern humans, that was lost through incomplete lineage sorting and integrated roughly 4.7 million years ago [95% highest posterior density: 1.9, 7.2 million years ago; Fig. 2E]. We named this element the "source PtERV1," as it was present in the common ancestor of all African apes and was likely the progenitor for independent expansions to nonorthologous loci in the chimpanzee and gorilla genomes. The source PtERV1

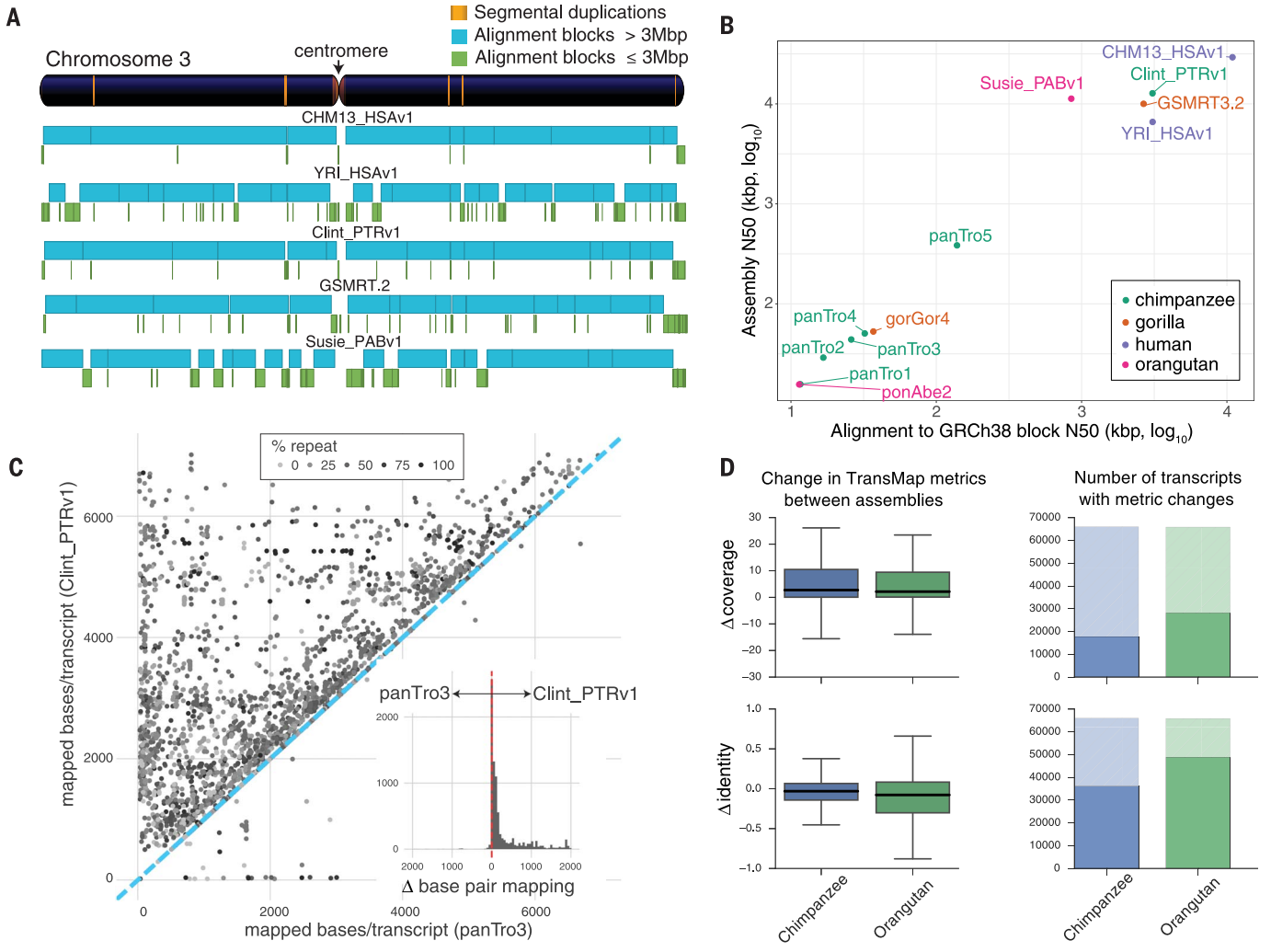


Fig. 1. Assembly and annotation of great ape genomes. (**A**) Comparison of genome sequence contiguity. Chromosome 3 contiguity is compared among the great ape genome assemblies by alignment to human reference genome sequence GRCh38. Contigs larger than (blue) and smaller than (green) 3 Mbp are compared with the position of SDs (SDs >50 kbp in size, orange) shown in the reference ideogram. (**B**) Scatterplot of syntenic-alignment block lengths (*x* axis) against GRCh38 versus FALCON assembly contig N50 length (*y* axis) of the great ape assemblies. The SMRT assemblies are Clint_PTRv1, Susie_PABv1, GSMRT3.2, CHM13_HSAv1, and YRI_HSAv1. The previous reference genomes are ponAbe2 (GCF_000001545.3), gorGor4 (GCA_000151905.3), panTro2 (GCF_000001515.2), panTro3 (GCA_000001515.3), panTro4 (GCA_000001515.4), and panTro5 (GCA_000001515.5). (**C**) Full-length assembled transcripts mapped to Clint_PTRv1 and panTro3. Each point denotes the number of bases per transcript matching the two assemblies. Repeat content is indicated by gray shading of the points. Although most of the transcripts map well to both assemblies (Pearson's correlation = 0.95), the subset of differentially mapped transcripts (12,724; 60% of 21,118) aligns better to Clint_PTRv1 (dots above the blue dashed line). The histogram inset shows the effect, per transcript, with a total of 4.8 million more bases aligned to Clint_PTRv1. Δ, difference in mapped bases per transcript. (**D**) Comparative Annotation Toolkit was used to project transcripts from GRCh38 to Clint_PTRv1, panTro3, Susie_PABv1, and ponAbe2. Alignment coverage and identity were compared for orthologous transcripts found in each assembly pair. The boxplots (left) summarize TransMap differences between the short-read and SMRT assemblies, in terms of coverage and identity. The solid-shaded portions of the bar plots (right) represent alignments, which had identical coverage or identity in both assemblies.

was likely missed in earlier genomic studies of draft genomes because the locus (sharing orthology with human chromosome 19) (*16*) is repeat rich and the integration site is an ancient LTR element.

Structural variation analyses

We focused on identifying all SVs >50 bp in size within ape genomes because these are the least well-characterized differences and are more likely to affect gene function than SNVs (*32*). SVs were identified by mapping each assembly back to the human reference genome, by using the two newly assembled human genomes as a control for reference effects and fixed human differences (CHM13_HSAv1 and YRI_HSAv1). We detected 614,186 ape deletions, insertions, and inversions, with the number of SVs increasing as a function of evolutionary distance from human (Fig. 3 and Table 3). We confirmed 92% of 61 events (from 2.7 to 95 kbp in size) by bacterial artificial chromosome (BAC) sequencing (table S8) (three of the remaining events were polymorphic among the great apes, suggesting a validation rate of >95%). We assigned SVs as shared or lineage-specific and genotyped each at the population level, with a panel of 86 great apes (*33*) (Fig. 3A). We identified 17,789 fixed human-specific structural variants (fhSVs), including 11,897 fixed human-specific insertions (fhINSs) and 5892 fixed human-specific deletions (fhDELs) (Fig. 3A and table S9). Projecting these onto the human genome identifies potential hotspots of structural variation (Fig. 3B).

We annotated fhSVs against chimpanzee and human gene models (table S10). The Variant Effect Predictor annotated the loss of 13 start codons, 16 stop codons, and 61 exonic deletions in the human lineage. By contrast, we estimate that fhSVs disrupt 643 regulatory regions near 479 genes (for example, Fig. 3, C to E). Interestingly, 139 of the fhSVs intersect with regions recently classified as super-enhancers (*34*). A comparison with a previous analysis of human-conserved deletions (hCONDELs) from earlier

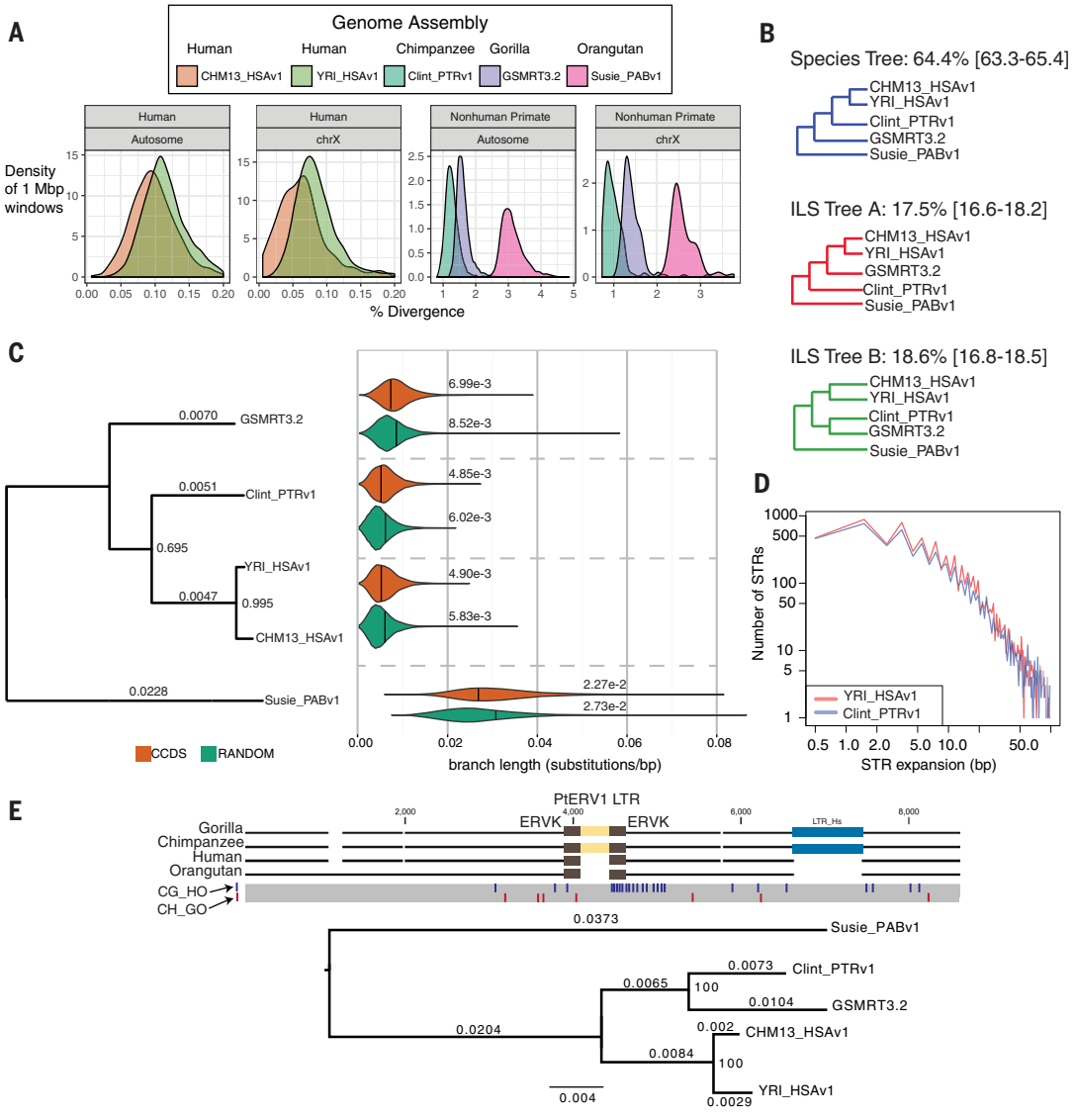


Fig. 2. Ape genetic diversity and lineage sorting. (**A**) SNV divergence between each primate assembly and GRCh38 was calculated in 1-Mbp nonoverlapping windows across all autosomes and chromosome X (excluding X and Y homologous regions). Mean autosomal divergence is 1.27 ± 0.20% (human-chimpanzee), 1.61 ± 0.21% (human-gorilla), and 3.12 ± 0.33% (human-orangutan). The African genome (YRI_HSAv1) shows a 17% increase in SNV diversity. (**B**) Proportion of phylogenetic trees supporting standard species topology and incomplete lineage sorting (ILS). The mean and 95% confidence intervals (in brackets) are based on 100 genome-wide permutations. (**C**) A phylogenetic tree (maximum clade credibility consensus tree) comparing genic regions [~9000 consensus coding sequences (CCDSs) and 1000-bp flanking sequence (orange)] to a randomly genome-shuffled set matched to coding-sequence lengths (green). The analysis excludes regions of SDs, SVs, and large tandem repeats. Branch lengths (noted above the lines) and proportion of trees supporting each bifurcation (internal nodes) are shown. Violin plots summarize the distribution and mean divergence (substitutions per base pair) for a subset of trees consistent with the species tree. YRI_HSAv1 is the representative human in the violin plots. (**D**) A comparison of the expanded STR sequences (n = 16,138 loci) between human (African) and chimpanzee ab initio genome assemblies shows little to no species bias (0.02 bp). (**E**) Top, a MSA of ape genomes (gorilla BAC CH277-16N20 and chimpanzee CH251-550G17) identifies an orthologous 379-bp PtERV1 element nested within another LTR and shared between gorilla and chimpanzee. Bottom, a maximum likelihood phylogenetic tree (GTR+Gamma) built from 12,108 bp that supports ILS. Single-nucleotide polymorphisms that support chimpanzee-gorilla sorting (CG_HO) and the species tree (CH_GO) are shown as blue and red lines, respectively. Branch lengths (substitutions per site) are shown above the lineages, and internal nodes are labeled with bootstrap support (proportion of replicates supporting split; 1000 replicates).

versions of the human, chimpanzee, and macaque genomes (*5*) confirms that 77% (451/583) of the hCONDELs intersect the fhDELs, with the remainder corresponding primarily to polymorphic events in the human population (Fig. 3F). We also predicted an additional 694 hCONDELs (table S11). A comparison of the SMRT gorilla assembly to the human reference genome identified an hCONDEL sequence previously reported as affecting an androgen receptor (AR) enhancer and associated with the loss of penile spines in humans. In gorilla, this fhDEL involves a complex SV, including an inversion, that may independently influence *AR* gene expression in the gorilla lineage (Fig. 3G) (*35*).

The spectrum of structural variation ranges from simple insertion and deletion events to larger events of increasing complexity (Fig. 4). We identified 46 fhSV deletions that putatively disrupt the orthologous chimpanzee gene, of which only six were previously reported (*5*). Seven of the 46 fhSV deletions can also be seen in the transcript data (Iso-Seq). The largest previously unidentified fhSV deletion is 61,265 bp in size. It contains almost all of the caspase recruitment domain family member 8 (*CARD8*) gene and removes 13 exons that are transcribed into full-length cDNA in the chimpanzee (Fig. 4A). We also resolve a 65-kbp human-specific deletion in *FADS1* and *FADS2*, genes involved in fatty acid biosynthesis that have been the target of positive selection (*36*) and potential dietary changes in human evolution (*37*, *38*). The deletion brings the promoters of *FADS1* and *FADS2* (major isoform) in closer proximity and shortens the first intron of the other two *FADS2* isoforms (Fig. 4B). The fhDEL might alter the relative abundance of the *FADS2* isoforms, as supported by quantifying the number of splice junction–containing reads specific to each isoform (*16*). The relative abundance of the minor *FADS2* isoforms is significantly increased in humans (χ^2 = 165.65, df = 1, $P < 2.2 \times 10^{-16}$). These minor isoforms differ only in their N terminus, and, of the two, one (NM_001281502.1) shows evidence of encoding a signal peptide (*39*), potentially altering the protein's subcellular location. Because great ape diets range from herbivorous to omnivorous, genic and structural changes related to diet metabolism may be of particular relevance for the evolution of ape species.

We further discovered two fhDELs in *WEE1* (Fig. 4C) and *CDC25C* (Fig. 4D), two highly conserved cell-cycle genes that act as ultrasensitive antagonists during the interphase to mitotic transition, G_2/M (*40*). *WEE1* encodes a serine-threonine protein kinase that delays mitosis by phosphorylating cyclin-dependent kinase 1 (CDK1), whereas *CDC25C* is a member of the phosphatase gene family and encodes a protein that dephosphorylates CDK1, triggering entry into mitosis. Expression of these genes in radial glia is particularly interesting because additional cell divisions are thought to have played a role in increasing the number of cortical neurons in human evolution (*41*). These cell-cycle regulators that display different protein sequence or differential expression between chimpanzee and human are, thus, candidates for future investigation to explain neocortical expansion in the human lineage.

We also identified several larger, subcytogenetic structural differences using optical (Bionano) (*42*, *43*) and BAC end-sequence mapping data that were not detected or sequence-resolved in previous genome assemblies. We validated large inversions and more-complex SV events by integrating fluorescence in situ hybridization (FISH) and large-insert clone sequencing at the breakpoints (table S12). We identified 29 human-chimpanzee-orangutan inversions (16 in chimpanzee, 10 in orangutan, and 3 shared between chimpanzee and orangutan) ranging from 100 kbp to 5 Mbp in size, of which 55% (16/29) have not been previously described (table S12 and Fig. 5) (*44*–*48*). More than 93% of inversions are flanked by large complex segmental-duplication (SD) blocks, 38% of which show evidence of other structural and copy-number variation at the boundaries of the inversion (Fig. 5).

Interestingly, ~28% (8/29) of these ape-human inversions are also polymorphic among humans (*49*, *50*), some in regions previously shown to be hotspots of recurrent rearrangement and disease (*48*, *51*). Notably, these regions of genomic instability also associate with expression differences in radial glial and excitatory neurons between the species. For example, among the 18 chimpanzee-human inversions (table S12), we identified 18 differentially expressed brain genes between chimpanzee and human (10 radial glia, 11 excitatory neurons, and 3 common to both sets), of which 78% resided in SD regions. Three of these genes (*GLG1*, *ST3GAL2*, and *EXOSC6*) were significantly up-regulated in human and associated with a 5-Mbp human-specific inversion on chromosome 16q22 (Fig. 5D). *ST3GAL2* encodes the main mammalian sialyltransferase for GD1a and GT1b ganglioside biosynthesis in the brain (*52*).

Table 2. Great ape gene and transcript annotation summary. TPM, transcripts per million; NA, not applicable to this genome; ND, no data.

Statistic	Ape assembly		
	Clint_PTRv1	GSMRT3.2	Susie_PABv1
Genes	55,894	55,985	55,522
Orthologs in human	55,594 (95.4*)	55,570 (95.4*)	54,900 (94.2*)
Isoforms	192,725	192,734	190,716
Coding genes	19,153	19,311	19,043
Previously undescribed predicted genes†	300	415	322
Coding isoforms	92,610	92,713	91,578
Transcript predictions with previously undescribed splice junctions‡	2,809	2,902	2,333
Transcripts with TPM > 0.1 (%)	66.3	67.3	50.6
Transcripts supported by Iso-Seq reads (%)	66.5	46.5	63.4
Previously unannotated exons identified	29	16	16
Putative exons gained in human	57	NA	ND
Putative exons lost in human	13	NA	NA

*Percent of GENCODE V27 represented. †Previously undescribed predicted genes based on GENCODE V27 annotation. ‡Previously undescribed splice junctions compared to liftover annotation set from the human reference genome, where splice junction is supported by NHP RNA-seq.

Radial glial neural progenitor expression differences and human-specific SVs

Over the course of human evolution, human brain volume has nearly tripled compared to that of chimpanzees (*53*), likely owing to differential expression of genes during brain development (*6*, *8*, *54*). We investigated the association of structural variation with changes in human-chimpanzee brain gene expression using cerebral organoids as a proxy for brain expression differences (*55*). Importantly, because great ape brain tissue is largely inaccessible, these organoid models provide a realistic window into developmental cell behavior and gene expression differences between human and ape radial glia and other early developmental cell types (*56*). We processed several single-cell RNA-seq brain data sets from primary human cortex and from human and chimpanzee cortical organoids, focusing on cortical excitatory neurons and radial glia (*55*–*57*). Using the new chimpanzee SMRT assembly and genome annotations increases the sensitivity of gene expression analyses; our data set reveals 2625 additional chimpanzee genes with expression in the brain relative to previous studies (*58*). After performing unsupervised clustering, we analyzed 52,875 orthologous genes in 320 primary neurons, 176 human organoid cells, and 210 chimpanzee organoid cells expressing cortical radial glia and excitatory neuron genes.

Our analysis identified 383 and 219 genes up-regulated in human radial glial and excitatory neurons, respectively, when compared to chimpanzee (table S13) (*16*). Conversely, we defined a set of 285 and 165 genes down-regulated in human radial glia and excitatory neurons (Fig. 6), respectively; most of these changes have not been identified previously (*56*, *59*). Because SVs

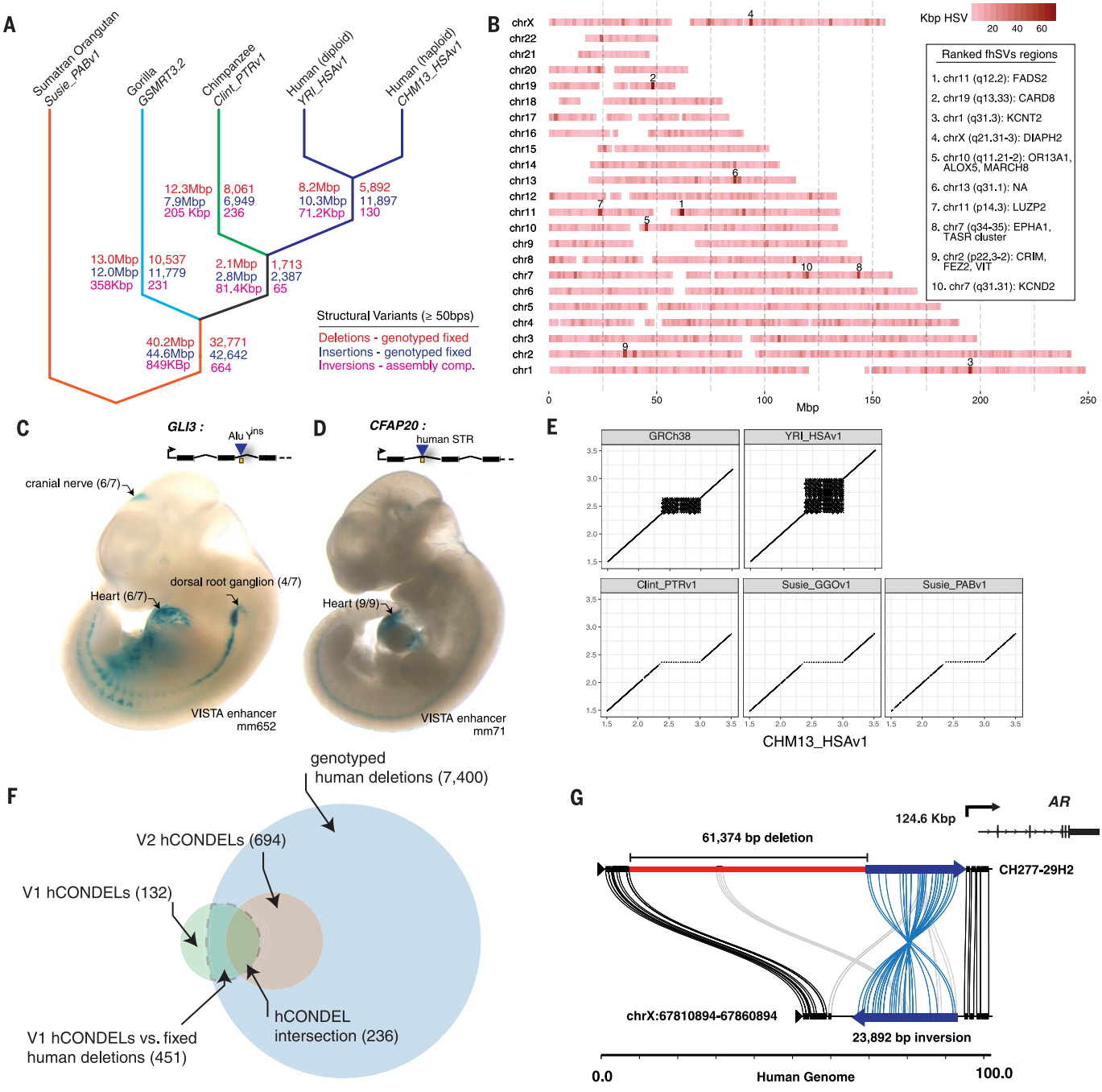


Fig. 3. Fixed structural variation and regulatory mutation. (**A**) The great ape cladogram with fixed structural variation assigned to lineages on the basis of assembly comparison, genotyping, and stratification (except for inversions). The total amount of sequence is shown on the left side of the branches, and the number of SVs is shown on the right for deletions (blue), insertions (red), and inversions (magenta). Inversions were assigned to branches on the basis of the comparison of our five assemblies because genotyping was less reliable. The cladogram was rooted against Susie_PABv1, meaning that the assignment of SVs to the orangutan or the common ancestor of human, chimpanzee, and gorilla is arbitrary. (**B**) A map of fhSVs. The color denotes the number of fhSVs bases (kbp), within a 1-Mbp sliding window (0.5-Mbp step). Each chromosome is labeled on the *y* axis. Key regions are annotated with genes. (**C**) The cell specificity for a mouse enhancer element (mm652, represented as a yellow box) that shares orthology in chimpanzee. In human, an AluY element has been inserted directly into the mm652 enhancer. *GLI3*, a GLI family zinc finger 3 gene. (**D**) A human-specific STR interrupts a mouse heart-specific enhancer shared with chimpanzee (yellow box). The STR is contained within a *CFAP20* intron (*CFAP20* encodes cilia and flagella associated protein 20). (**E**) Dot-plots of the human-specific STR expansion. The *x* axis corresponds to the CHM13_HSAv1 sequence (0.1-kbp units); the other sequences are on the *y* axis. The two human assemblies, CHM13_HSAv1 and YRI_HSAv1, show additional STR expansion relative to GRCh38, suggesting that the reference is collapsed. (**F**) A comparison of the hCONDEL set reported by McLean *et al*. (*5*) (V1) versus the hCONDELs reported here (V2). The current hCONDELs are from conservation (25-bp MSA windows) between chimpanzee, macaque, and mouse. The area enclosed by the dashed gray line shows the overlap between all fixed human deletions and all V1 hCONDELs. (**G**) A Miropeats diagram of the gorilla complex SV (inversion and deletion) upstream of the *AR* locus; the human reference genome is shown on the bottom (100-kbp window).

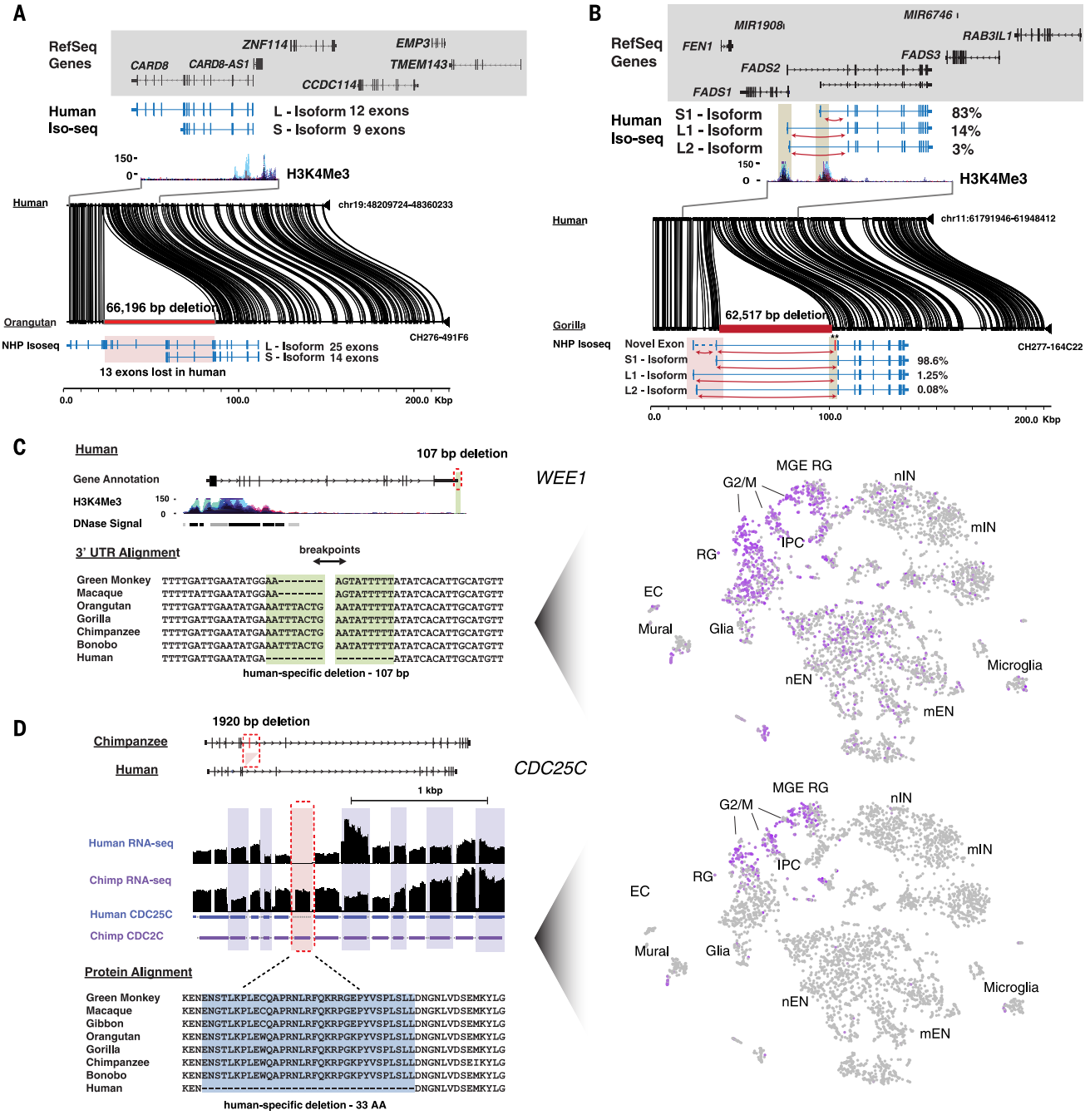


Fig. 4. Examples of intragenic human-specific structural variation. Shown are annotated MSAs between the human reference (GRCh38) and NHPs generated with Multiple Alignment using Fast Fourier Transform or visualized with Miropeats against sequenced large-insert primate clones. Single-cell gene expression for select genes is highlighted across 4261 cells from developing human telencephalon plotted using t-distributed stochastic neighbor embedding (tSNE) (*66*). (**A**) A 66.2-kbp intragenic deletion of *CARD8* removes 13 putative coding exons in human. Iso-Seq data from chimpanzee and human iPSCs identifies isoforms with and without the deleted exons, respectively. L, long; S, short; H3K4Me3, trimethylated histone H3 lysine 4. (**B**) A 62.5-kbp intergenic deletion of *FADS2* is found in humans, along with an altered isoform ratio: The relative abundance of the long isoforms is increased in humans relative to chimpanzee, as seen in the counts of junction-spanning short reads specific to each isoform. Additionally, a previously undescribed, rare (<5%) 75-bp exon is observed in chimpanzee and gorilla but absent in human, likely resulting from a human-specific splice-site mutation. (**C**) A 107-bp deletion in the 3′UTR of *WEE1* (red dashed box) reduces AU-rich sequence content in the mRNA. The tSNE plot illustrates that *WEE1* is highly expressed in cortical radial glia (RG), intermediate progenitor cells (IPCs), and medial ganglionic eminence progenitors (MGE RG) but shows limited expression in newborn and maturing inhibitory and excitatory neurons (nIN, mIN, nEN, and mEN, respectively), microglia, endothelial cells (ECs), and glia. DNase, deoxyribonuclease. (**D**) A 1920-bp deletion of cell-cycle regulator *CDC25C* (red dashed box) removes a 99-bp constitutive exon conserved in mouse, resulting in a 33–amino acid deletion and shorter N-terminal regulatory domain in humans. The tSNE plot illustrates that *CDC25C* shows restricted expression to telencephalon progenitors in the G_2/M cell-cycle phase. Human and chimpanzee RNA-seq data were aligned directly to the exonic regions of *CDC25C*.

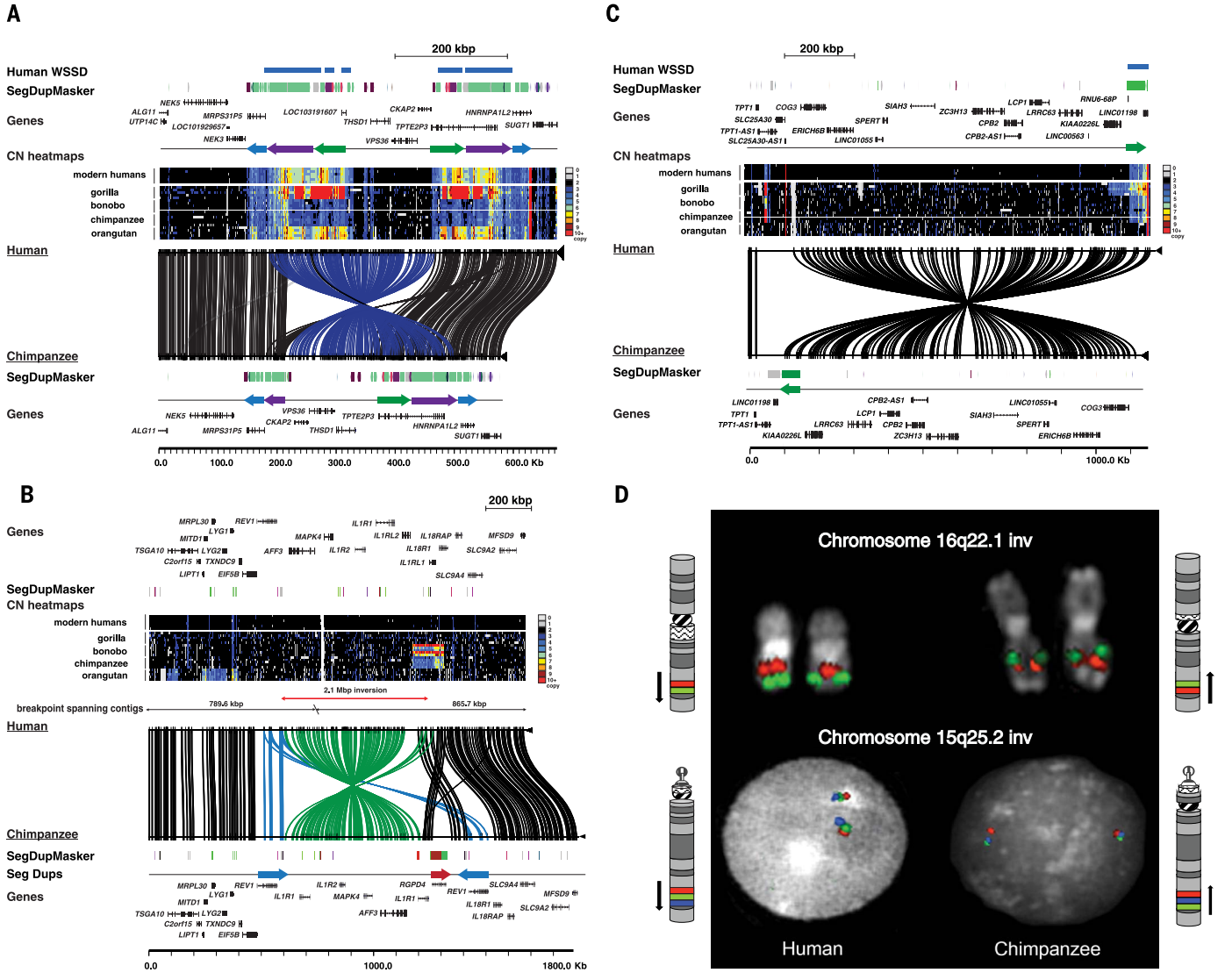


Fig. 5. Complex structural variation. Large-scale inversions between human and chimpanzee are depicted. The human reference genome sequence (GRCh38) with gene annotation is compared to large-insert clone-based assemblies from the chimpanzee BAC library CH251 using Miropeats. Connecting lines identify homologous regions of high sequence identity. SD organization is depicted with colored arrows, as defined by whole-genome shotgun sequence detection (WSSD) and DupMasker. Heatmap indicates copy number (CN) estimated by read depth from ape genome sequence. (**A**) A ~265-kbp inversion on chromosome 13q14.3 detected by optical mapping in chimpanzee (annotated blue lines). The inverted region is flanked by large ~180-kbp inverted SD blocks that vary with respect to copy number among great apes. (**B**) A 2.7-Mbp inversion on chromosome 2q12 to 13 detected by BAC end sequencing in chimpanzee (annotated green lines). The inverted region is flanked by duplication blocks containing lineage-specific expansions of the interleukins, an inverted duplication of *REV1*, and an additional copy of the *RGPD4* core duplicon. (**C**) A ~1.1-Mbp inversion at chromosome 13q14.13 identified by optical mapping in chimpanzee encompassing 15 genes. On the telomeric side of the inversion lies a ~60-kbp duplication block that demonstrates lineage-specific duplications in great apes. (**D**) Chromosome inversions, originally detected by optical mapping and BAC end sequencing, confirmed by metaphase analysis and interphase FISH experiments. A human-specific inversion of the chromosome 16q22.1 region was confirmed with orangutan clones CH276-89P20 (red) and CH276-192M7 (green) (top), and the 15q25.2 inversion was confirmed using chimpanzee clones CH251-321P13 (red), CH251-511D5 (green), and CH251-66E11 (blue) (bottom).

are more likely (*32*) to affect gene expression, we considered fhSV overlap on the basis of variant effect predictor annotations (including GRCh38 and Clint_PTRv1 annotation sets), which correlate both coding and noncoding variation to genes (Fig. 6A). Of the differentially expressed genes, 252 radial glia genes ($P = 9.78 \times 10^{-8}$, χ^2 test; 252/668) and 123 excitatory neuron genes ($P = 0.27$, χ^2 test; 123/360) had annotated fhSVs associated with them. To test if this observation was an artifact of gene size, we shuffled fhSVs and counted the number of fhSVs that mapped within 50 kbp of a differentially expressed gene.

Overall, genes down-regulated in humans remain enriched for fhSVs, compared to the null distribution, whereas up-regulated genes did not show a significant overlap. In particular, genes down-regulated in human radial glial neural progenitors showed significant enrichment for structural variation ($P = 0.02$; 10^4 permutations) (Fig. 6B). Although we observe the same trend in excitatory neurons, the effect did not reach significance. As a control, we repeated the same analysis for genes mapping to human-specific SDs (*54*), a form of structural variation not accessed in this study. Genes mapping to human SDs were up-regulated in radial glial and excitatory neurons

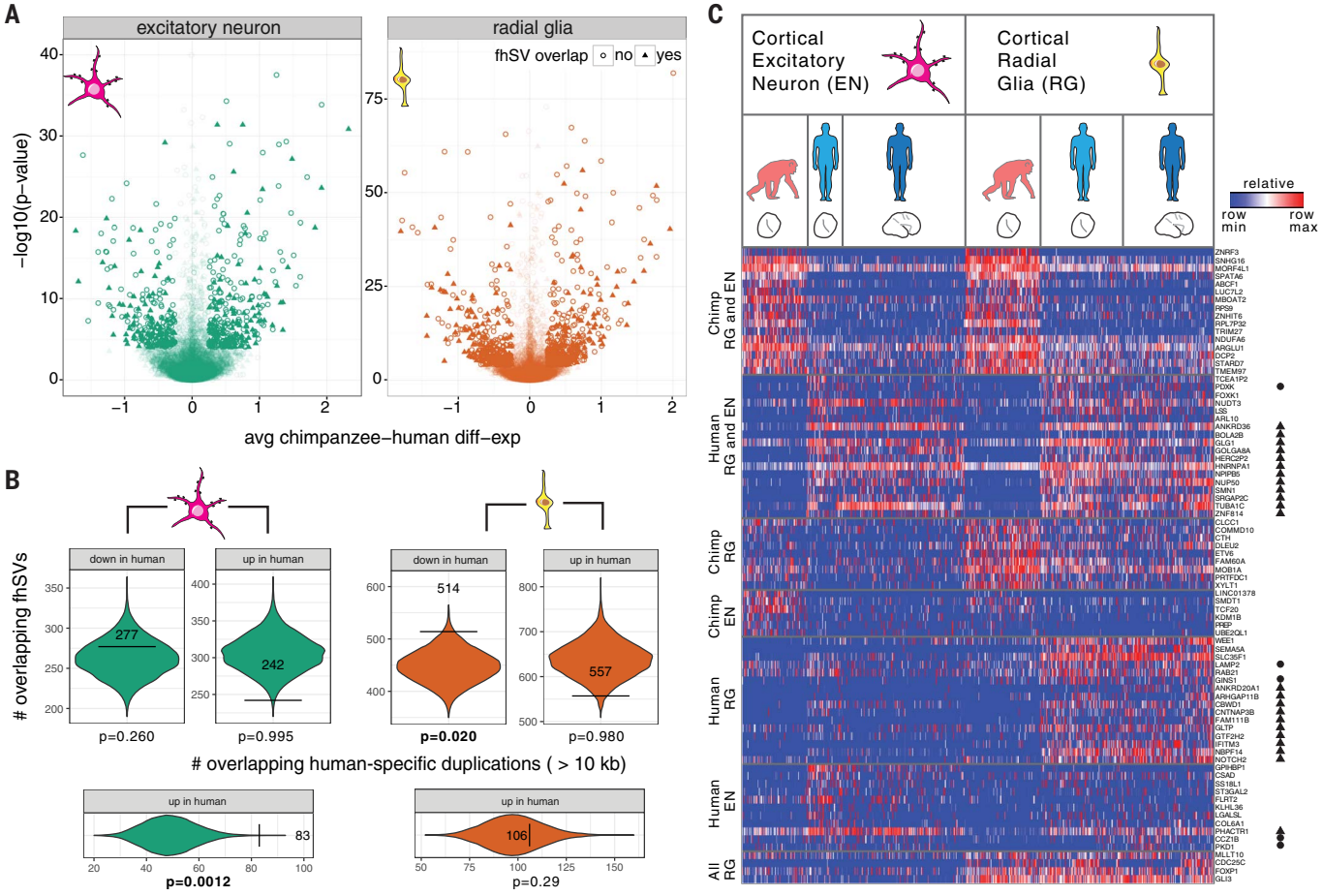


Fig. 6. Structural variation and neural progenitor expression differences between human and chimpanzee. (**A**) Volcano plots for chimpanzee-human gene expression in excitatory neuron (left) and radial glia (right) organoid single-cell data. Each point represents a gene, with sufficient data to assess significance between human and chimpanzee organoid cells. Genes with fhSVs within 50 kbp of their start or end are indicated with a triangle. The data points are shaded by significance (lighter shade indicates less significance). (**B**) Spatial permutation test for overlap between fhSVs and differentially expressed genes. Each violin plot shows the null distribution of human-specific SV overlap (±50 kbp of transcript start or end) with genes that are significantly differentially down- or up-regulated, relative to chimpanzee. The horizontal bars and observed counts are overlaid on the null distribution. (**C**) Heatmap illustrating the percentile gene expression of differentially expressed genes near fhSVs (rows) across single cells (columns), including genes near the start or end of inversions (circles) and duplicated regions (WSSD) (triangles). Cells include 333 excitatory neurons (97 chimpanzee organoid, 53 human organoid, and 183 human primary cells) and 373 radial glia (113 chimpanzee organoid, 123 human organoid, and 137 human primary cells) (*56*, *57*). Expression patterns include concerted changes between chimpanzee and human cells across radial glia and excitatory neurons (chimp RG and EN and human RG and EN), cell-type-specific changes (human EN and human RG), and conserved radial glia expression (all RG).

Table 3. Summary of great ape genome structural variation. SV events (>50 bp in size) called against the human reference genome (GRCh38) using smartie-sv.

Statistic	Ape assembly				
	CHM13_HSAv1	YRI_HSAv1	Clint_PTRv1	GSMRT3.2	Susie_PABv1
Deletion count	9,126	11,747	63,634	73,681	136,980
Insertion count	14,962	14,528	68,589	76,230	142,631
Inversion count	74	55	446	533	969
Deletion size (Mbp)	4.76	4.85	41.88	45.48	84.76
Insertion size (Mbp)	6.85	7.17	40.34	47.53	120.35
Average deletion size (bp)	552	413	658	617	618
Average insertion size (bp)	458	493	588	623	843
Size of largest variation (kbp) [type]	84 [deletion]	124 [insertion]	133 [insertion]	90 [insertion]	123 [insertion]

when compared to chimpanzee (Fig. 6). This association identifies dozens of putative candidates for functional investigation, including some of the most differentially expressed genes between humans and chimpanzees in neural progenitor cells (Fig. 6 and table S14).

Discussion

Our great ape genome assemblies improve sequence contiguity by orders of magnitude (*20*, *60*), leading to a more comprehensive understanding of the evolution of structural variation. Coupling this effort with full-length cDNA sequencing improved gene annotation, especially for the discovery of transcripts and isoforms that have recently diverged between closely related species. Because genomes of species may be sequenced and assembled using the same platforms and experimental designs, we minimized biases introduced by ascertainment or an uneven sequencing quality between genomes.

These improved genomes yield a comprehensive view of intermediate-size structural variation among apes. As we focused on SVs that potentially disrupt genes or regulatory sequences, we began to address potential functional effect. Differential gene expression, especially in cortical radial glia, has been hypothesized to be a critical effector of brain size and a likely selective target of human brain evolution (*41*). Nearly 4.1% of the genes down-regulated in human radial glia, when compared to chimpanzee radial glial analogs from cerebral organoids, associate with an fhSV and most often as a deletion or a retroposon insertion. These findings are consistent with the "less-is-more" hypothesis (*61*), which argues that the loss of functional elements underlies critical aspects of human evolution. By contrast, human-specific gene duplications associate with up-regulated expression in both neural progenitors and excitatory neurons, although the effect is stronger for the latter. This finding is consistent with recent studies evidencing that human-specific SDs contribute to cortical differences between humans and chimpanzees (*6*–*8*). It is intriguing that the repeat-rich nature of ape genomes and, in particular, the expansion of SDs in the common ancestral lineage of the African ape lineage (*62*) may have made great ape genomes particularly prone to both deletion and duplication events, accelerating the rate of structural changes and large-effect mutations during the evolution of these species.

Despite this more comprehensive assessment of structural variation, not all SV types have been fully resolved among the great apes. In particular, we are still missing many larger, more complex events, including inversions and SDs that have differentially evolved between the lineages. For example, we recovered only one of five ape inversions identified by comparative BAC-based sequencing of a 2-Mbp region of chromosome 16p11.2 (*63*), although optical mapping techniques did identify four of the events. In this case, all inversions are flanked by large blocks of SDs (>200 kbp) that cannot be currently assembled by long-read whole-genome sequencing. We predict that such large, multi–mega–base pair inversions represent a common uncharacterized source of human-ape genetic variation that has been underestimated. Long-range sequencing and mapping technologies, such as Strand-seq (*49*), BAC-based sequencing (*63*), optical mapping (table S12), and longer-read sequencing (*64*) will be necessary to sequence resolve such large, more complex SVs.

Materials and Methods

We sequenced and assembled four genomes [chimpanzee (Clint), Sumatran orangutan (Susie), CHM13 (human), and YRI19240 (human)] using long-read PacBio RS II sequencing chemistry and the Falcon genome assembler. Sequence contigs were error-corrected using Quiver (*17*), Pilon (*18*), and a FreeBayes-based (*65*) indel correction pipeline. A chromosomal-level AGP was generated using optical maps (Bionano Genomics Saphyr platform) for scaffold building and bicolor FISH of ~700 large-insert clones. The Comparative Annotation Toolkit (CAT) (*22*) was used to annotate all of the great ape genomes using the human GENCODE V27 as reference with a combination of RNA-seq obtained from SRA as well as Iso-Seq data specifically from NHP iPSCs. STRs were defined using RepeatMasker v4.0.1 and Tandem Repeats multiple sequence Finder v4.07b. Syntenic regions and MSAs were constructed with MUSCLE (v3.8.31), phylogenetic analyses were performed using a general time-reversible model ("GTR+GAMMA") under a maximum likelihood RAxML (8.2.3) framework, and phylogenetic trees were generated using DendroP. A BLASR-based computational pipeline, smartie-sv, was developed to align, compare, and call insertions, deletions, and inversions (https://github.com/zeeev/smartie-sv). Insertions and deletions were genotyped against a panel of 45 ape genomes using SVTyper (paired-end) and WSSD (read depth). FISH and BAC clone sequencing was used to estimate sequence accuracy and validate the breakpoints of complex rearrangements. We compared SV locations with genes showing differential expression during human and chimpanzee cortical development using single-cell gene expression data from cerebral organoid models and from primary cortex.

REFERENCES AND NOTES

1. A. Varki, D. H. Geschwind, E. E. Eichler, Human uniqueness: Genome interactions with environment, behaviour and culture. *Nat. Rev. Genet.* **9**, 749–763 (2008). doi: 10.1038/nrg2428; pmid: 18802414
2. M. C. King, A. C. Wilson, Evolution at two levels in humans and chimpanzees. *Science* **188**, 107–116 (1975). doi: 10.1126/science.1090005; pmid: 1090005
3. A. Fortna *et al.*, Lineage-specific gene duplication and loss in human and great ape evolution. *PLOS Biol.* **2**, e207 (2004). doi: 10.1371/journal.pbio.0020207; pmid: 15252450
4. J. L. Boyd *et al.*, Human-chimpanzee differences in a FZD8 enhancer alter cell-cycle dynamics in the developing neocortex. *Curr. Biol.* **25**, 772–779 (2015). doi: 10.1016/j.cub.2015.01.041; pmid: 25702574
5. C. Y. McLean *et al.*, Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature* **471**, 216–219 (2011). doi: 10.1038/nature09774; pmid: 21390129
6. M. Y. Dennis *et al.*, Evolution of human-specific neural SRGAP2 genes by incomplete segmental duplication. *Cell* **149**, 912–922 (2012). doi: 10.1016/j.cell.2012.03.033; pmid: 22559943
7. C. Charrier *et al.*, Inhibition of SRGAP2 function by its human-specific paralogs induces neoteny during spine maturation. *Cell* **149**, 923–935 (2012). doi: 10.1016/j.cell.2012.03.034; pmid: 22559944
8. M. Florio *et al.*, Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science* **347**, 1465–1470 (2015). doi: 10.1126/science.aaa1975; pmid: 25721503
9. X.-C. Ju *et al.*, The hominoid-specific gene TBC1D3 promotes generation of basal neural progenitors and induces cortical folding in mice. *eLife* **5**, 206 (2016). doi: 10.7554/eLife.18197; pmid: 27504805
10. A. Scally *et al.*, Insights into hominid evolution from the gorilla genome sequence. *Nature* **483**, 169–175 (2012). doi: 10.1038/nature10842; pmid: 22398555
11. D. P. Locke *et al.*, Comparative and demographic analysis of orangutan genomes. *Nature* **469**, 529–533 (2011). doi: 10.1038/nature09687; pmid: 21270892
12. Chimpanzee Sequencing and Analysis Consortium, Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* **437**, 69–87 (2005). doi: 10.1038/nature04072; pmid: 16136131
13. E. T. Lam *et al.*, Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol.* **30**, 771–776 (2012). doi: 10.1038/nbt.2303; pmid: 22797562
14. J. N. Burton *et al.*, Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013). doi: 10.1038/nbt.2727; pmid: 24185095
15. J. Eid *et al.*, Real-time DNA sequencing from single polymerase molecules. *Science* **323**, 133–138 (2009). doi: 10.1126/science.1162986; pmid: 19023044
16. Materials and methods are available as supplementary materials.
17. C.-S. Chin *et al.*, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* **10**, 563–569 (2013). doi: 10.1038/nmeth.2474; pmid: 23644548
18. B. J. Walker *et al.*, Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLOS ONE* **9**, e112963 (2014). doi: 10.1371/journal.pone.0112963; pmid: 25409509
19. J. J. Yunis, O. Prakash, The origin of man: A chromosomal pictorial legacy. *Science* **215**, 1525–1530 (1982). doi: 10.1126/science.7063861; pmid: 7063861
20. D. Gordon *et al.*, Long-read sequence assembly of the gorilla genome. *Science* **352**, aae0344 (2016). doi: 10.1126/science.aae0344; pmid: 27034376
21. L. F. K. Kuderna *et al.*, A 3-way hybrid approach to generate a new high-quality chimpanzee reference genome (Pan_tro_3.0). *Gigascience* **6**, 1–6 (2017). doi: 10.1093/gigascience/gix098; pmid: 29092041
22. I. T. Fiddes, J. Armstrong, M. Diekhans, S. Nachtweide, Z. N. Kronenberg, J. G. Underwood, D. Gordon, D. Earl, T. Keane, E. E. Eichler, D. Haussler, M. Stanke, B. Paten, Comparative Annotation Toolkit (CAT)–simultaneous clade and personal genome annotation. bioRxiv 231118 [Preprint]. 8 December 2017. doi: 10.1101/231118
23. N. Elango, J. W. Thomas, S. V. Yi; NISC Comparative Sequencing Program, Variable molecular clocks in hominoids. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 1370–1375 (2006). doi: 10.1073/pnas.0510716103; pmid: 16432233
24. P. Moorjani, C. E. G. Amorim, P. F. Arndt, M. Przeworski, Variation in the molecular clock of primates. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 10607–10612 (2016). doi: 10.1073/pnas.1600374113; pmid: 27601674
25. W. H. Li, M. Tanimura, P. M. Sharp, An evaluation of the molecular clock hypothesis using mammalian DNA sequences. *J. Mol. Evol.* **25**, 330–342 (1987). doi: 10.1007/BF02603118; pmid: 3118047
26. D. M. Bickhart *et al.*, Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat. Genet.* **49**, 643–650 (2017). doi: 10.1038/ng.3802; pmid: 28263316
27. D. C. Rubinsztein *et al.*, Microsatellite evolution—evidence for directionality and variation in rate between species. *Nat. Genet.* **10**, 337–343 (1995). doi: 10.1038/ng0795-337; pmid: 7670473
28. M. T. Webster, N. G. C. Smith, H. Ellegren, Microsatellite evolution inferred from human-chimpanzee genomic

sequence alignments. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8748–8753 (2002). doi: 10.1073/pnas.122067599; pmid: 12070344
29. C. T. Yohn *et al*., Lineage-specific expansions of retroviral insertions within the genomes of African great apes but not humans and orangutans. *PLOS Biol.* **3**, e110 (2005). doi: 10.1371/journal.pbio.0030110; pmid: 15737067
30. N. Polavarapu, N. J. Bowen, J. F. McDonald, Identification, characterization and comparative genomics of chimpanzee endogenous retroviruses. *Genome Biol.* **7**, R51 (2006). doi: 10.1186/gb-2006-7-6-r51; pmid: 16805923
31. S. M. Kaiser, H. S. Malik, M. Emerman, Restriction of an extinct retrovirus by the human TRIM5alpha antiviral protein. *Science* **316**, 1756–1758 (2007). doi: 10.1126/science.1140579; pmid: 17588933
32. P. H. Sudmant *et al*., An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75–81 (2015). doi: 10.1038/nature15394; pmid: 26432246
33. J. Prado-Martinez *et al*., Great ape genetic diversity and population history. *Nature* **499**, 471–475 (2013). doi: 10.1038/nature12228; pmid: 23823723
34. Y. A. Pérez-Rico *et al*., Comparative analyses of super-enhancers reveal conserved elements in vertebrate genomes. *Genome Res.* **27**, 259–268 (2017). doi: 10.1101/gr.203679.115; pmid: 27965291
35. P. L. Reno *et al*., A penile spine/vibrissa enhancer sequence is missing in modern and extinct humans but is retained in multiple primates with penile spines and sensory vibrissae. *PLOS ONE* **8**, e84258 (2013). doi: 10.1371/journal.pone.0084258; pmid: 24367647
36. A. Ameur *et al*., Genetic adaptation of fatty-acid metabolism: A human-specific haplotype increasing the biosynthesis of long-chain omega-3 and omega-6 fatty acids. *Am. J. Hum. Genet.* **90**, 809–820 (2012). doi: 10.1016/j.ajhg.2012.03.014; pmid: 22503634
37. K. Ye, F. Gao, D. Wang, O. Bar-Yosef, A. Keinan, Dietary adaptation of FADS genes in Europe varied across time and geography. *Nat. Ecol. Evol.* **1**, 0167 (2017). doi: 10.1038/s41559-017-0167; pmid: 29094686
38. M. T. Buckley *et al*., Selection in Europeans on fatty acid desaturases associated with dietary changes. *Mol. Biol. Evol.* **34**, 1307–1318 (2017). doi: 10.1093/molbev/msx103; pmid: 28333262
39. T. N. Petersen, S. Brunak, G. von Heijne, H. Nielsen, SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **8**, 785–786 (2011). doi: 10.1038/nmeth.1701; pmid: 21959131
40. N. B. Trunnell, A. C. Poon, S. Y. Kim, J. E. Ferrell Jr., Ultrasensitivity in the Regulation of Cdc25C by Cdk1. *Mol. Cell* **41**, 263–274 (2011). doi: 10.1016/j.molcel.2011.01.012; pmid: 21292159
41. P. Rakic, A small step for the cell, a giant leap for mankind: A hypothesis of neocortical expansion during evolution. *Trends Neurosci.* **18**, 383–388 (1995). doi: 10.1016/0166-2236(95)93934-P; pmid: 7482803
42. M. Pendleton *et al*., Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat. Methods* **12**, 780–786 (2015). doi: 10.1038/nmeth.3454; pmid: 26121404
43. A. C. Y. Mak *et al*., Genome-wide structural variation detection by genome mapping on nanochannel arrays. *Genetics* **202**, 351–362 (2016). doi: 10.1534/genetics.115.183483; pmid: 26510793
44. L. Feuk *et al*., Discovery of human inversion polymorphisms by comparative analysis of human and chimpanzee DNA sequence assemblies. *PLOS Genet.* **1**, e56 (2005). doi: 10.1371/journal.pgen.0010056; pmid: 16254605
45. T. L. Newman *et al*., A genome-wide survey of structural variation between human and chimpanzee. *Genome Res.* **15**, 1344–1356 (2005). doi: 10.1101/gr.4338005; pmid: 16169929
46. J. M. Szamalek *et al*., Polymorphic micro-inversions contribute to the genomic variability of humans and chimpanzees. *Hum. Genet.* **119**, 103–112 (2006). doi: 10.1007/s00439-005-0117-6; pmid: 16362346
47. M. F. Cardone *et al*., Hominoid chromosomal rearrangements on 17q map to complex regions of segmental duplication. *Genome Biol.* **9**, R28 (2008). doi: 10.1186/gb-2008-9-2-r28; pmid: 18257913
48. M. C. Zody *et al*., Evolutionary toggling of the MAPT 17q21.31 inversion region. *Nat. Genet.* **40**, 1076–1083 (2008). doi: 10.1038/ng.193; pmid: 19165922
49. A. D. Sanders *et al*., Characterizing polymorphic inversions in human genomes by single-cell sequencing. *Genome Res.* **26**, 1575–1587 (2016). doi: 10.1101/gr.201160.115; pmid: 27472961
50. M. J. P. Chaisson, A. D. Sanders, X. Zhao, A. Malhotra, D. Porubsky, T. Rausch, E. J. Gardner, O. Rodriguez, L. Guo, R. L. Collins, X. Fan, J. Wen, R. E. Handsaker, S. Fairley, Z. N. Kronenberg, X. Kong, F. Hormozdiari, D. Lee, A. M. Wenger, A. Hastie, D. Antaki, P. Audano, H. Brand, S. Cantsilieris, H. Cao, E. Cerveira, C. Chen, X. Chen, C.-S. Chin, Z. Chong, N. T. Chuang, D. M. Church, L. Clarke, A. Farrell, J. Flores, T. Galeev, G. David, M. Gujral, V. Guryev, W. Haynes-Heaton, J. Korlach, S. Kumar, J. Y. Kwon, J. E. Lee, J. Lee, W.-P. Lee, S. P. Lee, P. Marks, K. Valud-Martinez, S. Meiers, K. M. Munson, F. Navarro, B. J. Nelson, C. Nodzak, A. Noor, S. Kyriazopoulou-Panagiotopoulou, A. Pang, Y. Qiu, G. Rosanio, M. Ryan, A. Stutz, D. C. J. Spierings, A. Ward, A. E. Welsch, M. Xiao, W. Xu, C. Zhang, Q. Zhu, X. Zheng-Bradley, G. Jun, L. Ding, C. L. Koh, B. Ren, P. Flicek, K. Chen, M. B. Gerstein, P.-Y. Kwok, P. M. Lansdorp, G. Marth, J. Sebat, X. Shi, A. Bashir, K. Ye, S. E. Devine, M. Talkowski, R. E. Mills, T. Marschall, J. Korbel, E. E. Eichler, C. Lee, Multi-platform discovery of haplotype-resolved structural variation in human genomes. bioRxiv 193144 [Preprint]. 23 September 2017. doi: 10.1101/193144
51. B. P. Coe *et al*., Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat. Genet.* **46**, 1063–1071 (2014). doi: 10.1038/ng.3092; pmid: 25217958
52. E. R. Sturgill *et al*., Biosynthesis of the major brain gangliosides GD1a and GT1b. *Glycobiology* **22**, 1289–1301 (2012). doi: 10.1093/glycob/cws103; pmid: 22735313
53. S. Herculano-Houzel, The human brain in numbers: A linearly scaled-up primate brain. *Front. Hum. Neurosci.* **3**, 31 (2009). doi: 10.3389/neuro.09.031.2009; pmid: 19915731
54. M. Y. Dennis *et al*., The evolution and population diversity of human-specific segmental duplications. *Nat. Ecol. Evol.* **1**, 0069 (2017). doi: 10.1038/s41559-016-0069; pmid: 28580430
55. J. G. Camp *et al*., Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15672–15677 (2015). pmid: 26644564
56. F. Mora-Bermúdez *et al*., Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. *eLife* **5**, 166 (2016). doi: 10.7554/eLife.18683; pmid: 27669147
57. A. A. Pollen *et al*., Molecular identity of human outer radial glia during cortical development. *Cell* **163**, 55–67 (2015). doi: 10.1016/j.cell.2015.09.004; pmid: 26406371
58. Z. He *et al*., Comprehensive transcriptome analysis of neocortical layers in humans, chimpanzees and macaques. *Nat. Neurosci.* **20**, 886–895 (2017). doi: 10.1038/nn.4548; pmid: 28414332
59. M. C. N. Marchetto *et al*., Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature* **503**, 525–529 (2013). doi: 10.1038/nature12686; pmid: 24153179
60. J. Korlach *et al*., De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *Gigascience* **6**, 1–16 (2017). doi: 10.1093/gigascience/gix085; pmid: 29020750
61. M. V. Olson, When less is more: Gene loss as an engine of evolutionary change. *Am. J. Hum. Genet.* **64**, 18–23 (1999). doi: 10.1086/302219; pmid: 9915938
62. T. Marques-Bonet *et al*., A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* **457**, 877–881 (2009). doi: 10.1038/nature07744; pmid: 19212409
63. X. Nuttle *et al*., Emergence of a *Homo sapiens*-specific gene family and chromosome 16p11.2 CNV susceptibility. *Nature* **536**, 205–209 (2016). doi: 10.1038/nature19075; pmid: 27487209
64. M. Jain, H. E. Olsen, B. Paten, M. Akeson, The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biol.* **17**, 239 (2016). doi: 10.1186/s13059-016-1103-0; pmid: 27887629
65. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907 [q-bio.GN] (17 July 2012).
66. T. J. Nowakowski *et al*., Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science* **358**, 1318–1323 (2017). doi: 10.1126/science.aap8809; pmid: 29217575

ACKNOWLEDGMENTS

We thank C. Lee and A. Lewis for technical assistance and quality control in generating sequencing data and G. I. Saunders from EBI for help with submitting our structural variation data to the EBI database. The authors thank C. Dunn and J. Chin for assistance with troubleshooting newer releases of the Falcon assembler and T. Brown and M. Lynn Gage for assistance in editing this manuscript. We thank L. W. Hillier for her insight on the chimpanzee assembly. Z.N.K. would like to thank E.E.E. and E. Kronenberg for their steady guidance. **Funding:** This work was supported, in part, by grants from the NIH (HG002385 to E.E.E.; HG007635 to R.K.W. and E.E.E.; HG003079 to R.K.W.; HG007990 to D.H., M.D., and B.P.; HG006283 to J.S.; HG009081 to S.K.D. and E.E.E.; HG007234 to D.H., B.P., J.A., and M.D.; HG008742 to B.P.; and HG009478 to M.L.D.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. This work was also partially supported by RBFR103CE3 by Ministero dell'Università e Ricerca (MIUR), Italy to M.V. and GC1R-06673-A by California Institute for Regenerative Medicine (CIRM) to I.T.F. and D.H. S.C. was supported by a National Health and Medical Research Council (NHMRC) CJ Martin Biomedical Fellowship (no. 1073726). The work carried out by NCBI was supported by the Intramural Research Program of the NIH, National Library of Medicine. E.E.E., J.S., and D.H. are investigators of the Howard Hughes Medical Institute. **Author contributions:** Z.N.K., I.T.F., D.G., S.C., O.S.M., M.J.P.C., J.G.U., A.P., F.H., S.K.D., H.C., and E.E.E. designed and planned experiments. K.M.M., J.G.U., R.Q., A.E.W., M.S., K.H., C.B., R.S.F., and M.L.D. prepared libraries and generated sequencing data. D.G., B.J.N., M.J.P.C., C.M.H., Z.N.K., M.L.D., S.M., E.R.H., O.S.M., P.H., and A.R. performed bioinformatics analyses. S.M., M.V., A.R.H., Z.N.K., and N.L. constructed the AGP. A.R.H., A.W.C.P., J.L., E.T.L., and H.C. generated Bionano Genomics de novo assembled optical maps, hybrid scaffolding with sequence assemblies, structural variation detection, and cross-platform data comparison. I.T.F., J.A., M.D., B.P., and D.H. annotated genome assembly and assessed gene models and accuracy. R.K.W., W.C.W., and T.A.G.-L. provided human genome assembly sequence data. J.S. helped with Hi-C data evaluation. F.H.G. and A.M.D. provided iPSC material. M.D., I.T.F., V.A.S., and K.C. deposited sequencing data and performed NCBI-UCSC annotation. K.H., S.C., M.V., N.L., A.R.H., and K.M.M. validated SV and indel events. E.E.E., M.L.D., J.G.U., S.M., B.J.N., I.T.F., S.C., M.J.P.C., A.P., D.G., and Z.N.K. wrote the manuscript. **Competing interests:** E.E.E. is on the scientific advisory board of DNAnexus, Inc. A.R.H., A.W.C.P., J.L., E.T.L., and H.C. are employees of Bionano Genomics, Inc. J.G.U. is an employee of Pacific Biosciences, Inc. **Data and materials availability:** The underlying PacBio sequence data, Illumina sequencing, assembled contigs, and assemblies for each of the ape species have been deposited in NCBI under the project accession numbers PRJNA369439 (chimpanzee, orangutan, CHM13, and NA19240) and PRJEB10880 (gorilla) (table S56). Clone sequences have been deposited in GenBank under umbrella BioProject ID PRJNA369439 (table S15). Transcriptional data was deposited in NCBI (table S57). The SVs were deposited in the Database of Genomic Variants archive under accession number estd235. The genome assemblies have different names and aliases depending on the institution hosting these genomes (table S58).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/eaar6343/suppl/DC1
Materials and Methods
Figs. S1 to S53
Tables S1 to S58
References (*67*–*104*)

7 December 2017; accepted 2 April 2018
10.1126/science.aar6343

RESEARCH ARTICLE SUMMARY

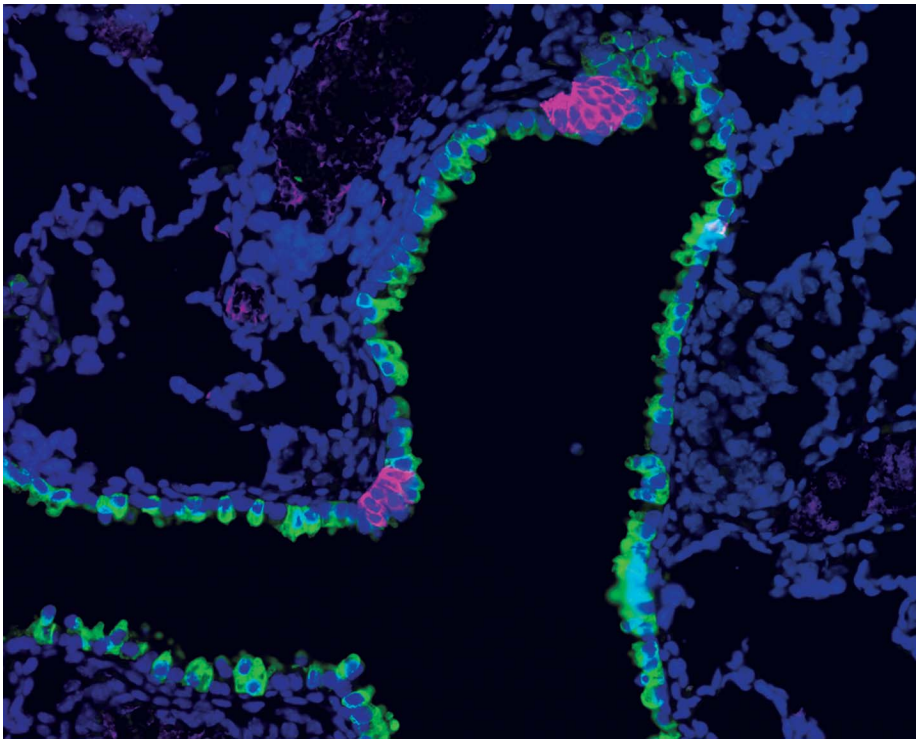
IMMUNOLOGY

Pulmonary neuroendocrine cells amplify allergic asthma responses

Pengfei Sui, Darin L. Wiesner, Jinhao Xu, Yan Zhang, Jinwoo Lee, Steven Van Dyken, Amber Lashua, Chuyue Yu, Bruce S. Klein, Richard M. Locksley, Gail Deutsch, Xin Sun*

INTRODUCTION: The lung, with its vast surface area, senses and responds to signals in inhaled air. Aberrant interactions between the lung and the environment underlie many diseases, including asthma. In vitro data show that pulmonary neuroendocrine cells (PNECs), a rare airway epithelial cell population, can act as chemosensors. Once stimulated in culture, they release dense core vesicles rich in neuropeptides, amines, and neurotransmitters. These bioactive molecules are capable of eliciting immune and physiological responses. A recent in vivo study by our group revealed that the proper development of PNECs into self-clustering units called neuroepithelial bodies is essential for restricting the number of immune cells in the naïve lung. However, whether PNECs can function in vivo to translate exogenous airway signals such as allergens into the cascade of downstream responses is unknown.

RATIONALE: To test the hypothesis that PNECs act as sensors in the lung, we generated mouse mutants that lack PNECs by inactivating *Ascl1* in the airway epithelium—i.e., mutants that were depleted of PNECs starting at development. We exposed these mutants to either ovalbumin or house dust mites, following regimes of existing asthma models. We determined whether the mutants showed different asthmatic responses than controls. We elucidated the underlying mechanisms by identifying molecular effectors and cellular targets of PNECs. To complement the functional tests in mice, we investigated whether human asthma patients showed pathological changes in their PNECs.



PNECs are preferentially localized at branch points. A mouse airway stained by antibody against CGRP, to label PNECs (magenta) and antibody against SCGB1A1 to label club cells (green) (200× magnification). PNECs often cluster into neuroepithelial bodies and are preferentially localized at branch points.

RESULTS: Although normal at baseline, *Ascl1*-mutant mice exhibited severely reduced goblet cell hyperplasia and immune cell numbers compared with controls after allergen challenge. In investigating possible molecular effectors, we found that several PNEC products were decreased in mutants relative to controls after allergen challenge, including calcitonin gene-related peptide (CGRP) and γ-aminobutyric acid (GABA). In exploring possible cellular targets, we found that innate lymphoid group 2 cells (ILC2s) were enriched at airway branch points, similar to PNECs. The PNEC product CGRP stimulated ILC2 production of interleukin-5 in culture. Conversely, inactivation of the CGRP receptor gene *Calcrl* in ILC2s led to dampened immune responses to allergens. In contrast to CGRP, GABA did not increase ILC2 cytokine secretion. Rather, inactivation of GABA biogenesis led to defective goblet cell hyperplasia after allergen challenge, suggesting that GABA is required for this response in the airway epithelium. The instillation of a mixture of CGRP and GABA in *Ascl1* mutants restored both immune cell increases and goblet cell hyperplasia after allergen challenge, indicating that these products are the primary molecular effectors of PNECs in vivo. Consistent with these results from mice, we found increased PNEC numbers and cluster sizes in human asthma patients, which may underlie the heightened response to allergens in these individuals.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aan8546

CONCLUSION: Our results demonstrate that PNECs, despite being a rare population of cells in the airway, are critical for amplifying the airway allergen signal into mucosal type 2 responses. Specifically, PNECs act through their product GABA to stimulate airway epithelial mucus production. In parallel, PNECs act through another product, CGRP, to stimulate ILC2 production of cytokines, which in turn recruit downstream immune cells. PNECs and ILC2s form neuroimmunological modules at the airway branch points, which are also the sites where airway particles are enriched. Our findings indicate that the PNEC-ILC2 axis functions to sense inhaled inputs, such as allergens, and amplify them into lung outputs, such as the allergic asthma response. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: xinsun@ucsd.edu
Cite this article as P. Sui *et al.*, *Science* **360**, eaan8546 (2018). DOI: 10.1126/science.aan8546

RESEARCH ARTICLE

IMMUNOLOGY

Pulmonary neuroendocrine cells amplify allergic asthma responses

Pengfei Sui[1,2], Darin L. Wiesner[3], Jinhao Xu[1,2], Yan Zhang[1,2], Jinwoo Lee[4], Steven Van Dyken[4], Amber Lashua[2], Chuyue Yu[5], Bruce S. Klein[3], Richard M. Locksley[4], Gail Deutsch[6], Xin Sun[1,2]*

Pulmonary neuroendocrine cells (PNECs) are rare airway epithelial cells whose function is poorly understood. Here we show that *Ascl1*-mutant mice that have no PNECs exhibit severely blunted mucosal type 2 response in models of allergic asthma. PNECs reside in close proximity to group 2 innate lymphoid cells (ILC2s) near airway branch points. PNECs act through calcitonin gene-related peptide (CGRP) to stimulate ILC2s and elicit downstream immune responses. In addition, PNECs act through the neurotransmitter γ-aminobutyric acid (GABA) to induce goblet cell hyperplasia. The instillation of a mixture of CGRP and GABA in *Ascl1*-mutant airways restores both immune and goblet cell responses. In accordance, lungs from human asthmatics show increased PNECs. These findings demonstrate that the PNEC-ILC2 neuroimmunological modules function at airway branch points to amplify allergic asthma responses.

A report by the National Center for Health Statistics of the U.S. Centers for Disease Control and Prevention estimates that 24.6 million people in the United States have asthma (*1*). Although much attention has been placed on immune cells as the source of symptoms, in recent years, the concept that nonhematopoietic tissues are a critical source of cytokines has gained support (*2*, *3*). For example, lung epithelial cells such as alveolar type 2 cells produce interleukin-33 (IL-33), which in turn activates immune cells, including residential group 2 innate lymphoid cells (ILC2s) (*4*, *5*). ILC2s secrete type 2 cytokines, including IL-5 and IL-13, which promote smooth muscle contraction, eosinophil infiltration, and goblet cell hyperplasia—key features of an asthmatic response (*6*, *7*).

PNECs are a distinctive and poorly understood cell population in the lung. They are rare, endoderm-derived epithelial cells that constitute ~1% of the airway cell population (*8*, *9*). They are the earliest specified lung epithelial cells, expressing *Ascl1*, their defining marker, starting at embryonic day 12.5. In rodents, a majority of PNECs form clusters, which are called neuroepithelial bodies (NEBs). NEBs are highly innervated by both afferent and efferent neurons (*10*). PNECs contain dense core vesicles filled with neuropeptides, amines, and neurotransmitters. In vitro, PNECs can be stimulated by oxygen, mechanical stretch, and chemical stimuli and release their vesicular content (*11*). Excess PNECs have been reported in lungs with various diseases, including chronic obstructive pulmonary disease, sudden infant death syndrome, bronchopulmonary dysplasia, and small cell lung cancer (*12*–*16*). Recently, we showed that proper clustering of PNECs is essential for restricting immune cell number in the naïve neonatal lung (*17*). However, the precise in vivo role of PNECs in lung pathogenesis remains unknown.

[1]Department of Pediatrics, University of California, San Diego, San Diego, CA 92093, USA. [2]Laboratory of Genetics, University of Wisconsin–Madison, Madison, WI 53706, USA. [3]Department of Pediatrics, University of Wisconsin–Madison, Madison, WI 53706, USA. [4]Department of Medicine, Howard Hughes Medical Institute, University of California, San Francisco, San Francisco, CA 94143, USA. [5]Zhiyuan College, Shanghai JiaoTong University, Shanghai, China. [6]Department of Laboratories, Seattle Children's Hospital, University of Washington, Seattle, WA 98105, USA.
***Corresponding author. Email: xinsun@ucsd.edu**

PNECs are required for goblet cell hyperplasia in asthma models

To generate a mouse mutant lacking PNECs, we inactivated *Ascl1*, which encodes a transcription factor that is essential for their formation, as shown in *Ascl1* global null mutants (*18*, *19*). To bypass the perinatal lethality of the global nulls, we inactivated *Ascl1* in precursors of PNECs in the airway epithelium by using *Shh*^{cre} (a cre insertion into the *sonic hedgehog* gene) at the onset of lung development (hereafter, *Ascl1CKO* for conditional knockout) (fig. S1) (*20*). This early inactivation led to a complete absence of PNECs, as indicated by the lack of calcitonin gene-related peptide (CGRP)– or synaptophysin-positive cells in the airway epithelium (Fig. 1). Synaptophysin also labels the nerves that normally innervate PNECs. In the mutant, these nerves remained present subjacent to the airway, even though they no longer intercalated into the epithelium (Fig. 1, C, D, C', and D'). In contrast to the absent PNECs, the intrinsic neurons, which are also dependent on *Ascl1*, remained, as indicated by TUJ1 (tubulin β3) expression—presumably because *Shh*^{cre} is not active in these cells to delete *Ascl1* (fig. S2).

Ascl1CKO mutants were viable at birth, which is counter to the postulation that PNECs are required for the transition from the intrauterine environment to breathing air. The mutant lungs showed normal branching morphogenesis and alveolar morphology (fig. S3). Using immunofluorescent staining and quantitative reverse transcription polymerase chain reaction (qRT-PCR) of key cell type markers, we found no apparent defects in airway club and ciliated cells or in type I and II alveolar cells. Like in the airways of control mice, no goblet cells were detected in the naïve mutant airways. These data indicate that, despite PNECs being the first differentiated cell type in the airway, their absence has no gross effects on the development of other cell types in the lung at baseline.

To test whether disruption of PNECs affects lung responses to environmental cues, we used an established model to induce allergic asthma-like responses by sensitizing early postnatal mice to ovalbumin (OVA) injected in combination with alum. Mice were then challenged with OVA inhalation (Fig. 2A) (*21*). As predicted, control mice showed robust goblet cell hyperplasia along the airways (Fig. 2B). All MUC5AC-positive goblet cells were also positive for SCGB1A1, a club cell marker, suggesting that club cells were partially converted into mucus-producing cells (fig. S4). In contrast, *Ascl1CKO* mice showed substantially reduced goblet cell hyperplasia, which registered at about 10% of the control response on the basis of a histological mucin index (Fig. 2, C to F). This observation was corroborated by dramatically reduced expression of *Muc5ac* and goblet cell–promoting factors including *Spdef* and *Foxa3* (Fig. 2G). Markers for other airway epithelial cell types—club cells and ciliated cells—remained normal in *Ascl1CKO* mice after OVA challenge (fig. S4).

To determine whether PNECs are also required for allergen-induced goblet cell hyperplasia in the adult lung, we exposed mutant and control mice to house dust mites (HDMs), a common allergen that also causes asthma in humans (*22*). By periodic acid–Schiff (PAS) staining and quantification of *Muc5ac* transcripts, we found that goblet cell hyperplasia response was reduced in the *Ascl1CKO* mutants compared with controls in the HDM model (fig. S5). Thus, PNECs amplify airway goblet cell hyperplasia in the asthma models tested.

PNECs are required for type 2 immune responses in asthma models

A cardinal feature of asthma is eosinophilic infiltration and type 2 T helper (T_H2) cell priming in the lung (*7*). To further explore the role of PNECs in this aspect of the asthmatic response, we compared immune cell frequencies in neonatal OVA and alum–treated *Ascl1CKO* and control mice. The *Ascl1CKO* mice exhibited lower numbers of ILC2s, eosinophils, and T_H2 cells than control mice after OVA challenge (Fig. 2, H to J, and fig. S6). Further, *Ascl1CKO* mice showed

reduced *Il5* and *Il13* expression (Fig. 2K). These differences were not due to baseline immune cell differences, given that phosphate-buffered saline (PBS)–treated *Ascl1CKO* and control mice had similar levels of immune cells (Fig. 2, H to J, and fig. S6).

To determine whether PNECs are required for allergen-induced immune response in the adult lung, we returned to the adult HDM model. In this adult model, *Ascl1CKO* mice showed decreased immune cell infiltration compared with controls (fig. S7). We also depleted PNECs in the postnatal stage by using a different genetic model. We used Shh^{cre} to activate diphtheria toxin receptor (DTR) expression under the transcriptional control of the PNEC product CGRP (encoded by *Calca*) in a $Calca^{loxP\text{-}GFP\text{-}loxP\text{-}DTR}$ strain (GFP, green fluorescent protein) (*23*). The introduction of diphtheria toxin into the lung resulted in a ~70% reduction in PNECs. When challenged with HDMs, these mice also showed an immune response that was dampened, although to a lesser extent than in the *Ascl1CKO* mutant, likely owing to residual PNECs in the DTR model (fig. S8). Thus, PNECs amplify allergen-induced type 2 immune responses in the asthma models tested.

Absence of PNECs leads to the reduction of key neuropeptides and the neurotransmitter GABA

In a normal lung, type 2 immune responses are led by changes in cytokines such as IL-25, IL-33, and TSLP (thymic stromal lymphopoietin) (*24*–*26*). In *Ascl1CKO* lungs, these cytokines were detected at similar levels to those in controls after each of the OVA challenges (fig. S9). These results suggest that these epithelium-derived cytokines may not explain why type 2 immune responses are different when PNECs are absent.

A key feature of PNECs is their production and secretion of neuropeptides and neurotransmitters (*11*). To test the hypothesis that PNECs control type 2 responses through some of these molecules, we assayed the expression of genes encoding several neuropeptides that have been implicated in inflammatory diseases. These included *Vip* (encoding vasoactive intestinal peptide), *Calca* (CGRP), *Chga* (chromogranin A), and *Npy* (neuropeptide Y) (*11*, *27*–*29*). In control lungs, *Calca*, *Chga*, *Npy*, and *Vip* levels were greater after OVA challenge than after administration of PBS, with *Calca* showing the largest fold increase. In *Ascl1CKO* lungs, *Calca*, *Chga*, and *Npy*, but not *Vip*, were significantly lower than in lungs of control mice after OVA challenge (Fig. 3A). The relative fold changes in *Calca* and *Chga* were larger than that in *Npy*, possibly owing to the concentrated expression of *Calca* and *Chga*, but not *Npy*, in PNECs (*11*, *27*).

To further investigate the source of the neuropeptide changes, we focused on CGRP as an example. CGRP is expressed in the PNECs and in lung sensory nerves with cell bodies located in the vagal ganglia. By qRT-PCR, we found that *Calca* expression in the vagal ganglia was similar in wild-type mice treated with OVA and PBS controls under the regime that we used (fig. S10A). *Calca* expression was also similar in the vagal ganglia of OVA-treated *Ascl1CKO* mice and genotype controls (fig. S10B). Within the lung epithelium, we found by immunofluorescent staining that CGRP expression remained restricted to PNECs after challenge (fig. S11). PNEC proliferation did not increase (fig. S11). Rather, the CGRP signal was detected at a notably higher intensity, suggesting that this PNEC-specific increase contributes to the overall increase of CGRP levels in the lung (fig. S11).

In addition to neuropeptides, we also assayed for the levels of γ-aminobutyric acid (GABA), a PNEC-produced neurotransmitter that can stimulate goblet cell hyperplasia (*30*, *31*). GABA levels were reduced in *Ascl1CKO* mutants, possibly owing to loss of GAD1, a rate-limiting enzyme in GABA biosynthesis, and VGAT, the vesicular GABA transporter, which are both specifically expressed in PNECs in the lung (Fig. 3B) (*31*). Together, these differences in neuropeptide and neurotransmitter levels suggest that they may be responsible for the reduced immune cell infiltration and goblet cell hyperplasia in the *Ascl1CKO* mutant.

PNECs reside in proximity to ILC2s and can stimulate ILC2 cytokine production

To identify possible direct targets of PNEC signaling, we made use of a distinct feature of these cells: their preferential localization to airway branch points (*8*, *9*). We searched for immune cells residing in close proximity to these junctions. ILC2s are a population of tissue-resident immune cells that have been shown to play a central role in type 2 immune responses (*32*, *33*). We investigated the localization of ILC2s by using a mouse reporter line with *tdTomato* knocked into the *Il5* locus (*28*). In naïve mice, in addition to confirming the previous finding that ILC2s are localized near the basement membrane subjacent to the airway epithelium (*11*, *27*–*29*), we found that a majority of ILC2s (178 of 217 total cells counted, or 82%) reside within 70 μm of airway branch points , also called nodal points (Fig. 3, C to F) (*9*). By labeling PNECs with antibody against CGRP (anti-CGRP antibody) in the *Il5-tdTomato* background, we found that ILC2s were preferentially localized within 70 μm of PNECs (132 of 217 total cells counted, or 61%) (Fig. 3G). This proximity suggests ILC2s as a candidate direct target of PNEC signaling.

To test whether ILC2s can perceive PNEC signals, we first addressed whether they express the appropriate receptors. Existing RNA-sequencing data indicate that ILC2s express the CGRP co-receptor genes *Calcrl* and *Ramp1* and the GABA receptor gene *Gabrr1*, but not the CHGA receptor gene (*34*). We confirmed the expression of CGRP and GABA receptor genes by qPCR assay in primary ILC2s sorted from lungs of naïve wild-type mice (fig. S12).

To test whether ILC2s could be stimulated by PNEC signals, we cultured primary lung ILC2s from naïve lungs and addressed their response to CGRP or GABA. ILC2s respond to cytokines such as IL-33 (*35*). Consistent with this, after in vivo OVA challenge, there was an increase in IL-33

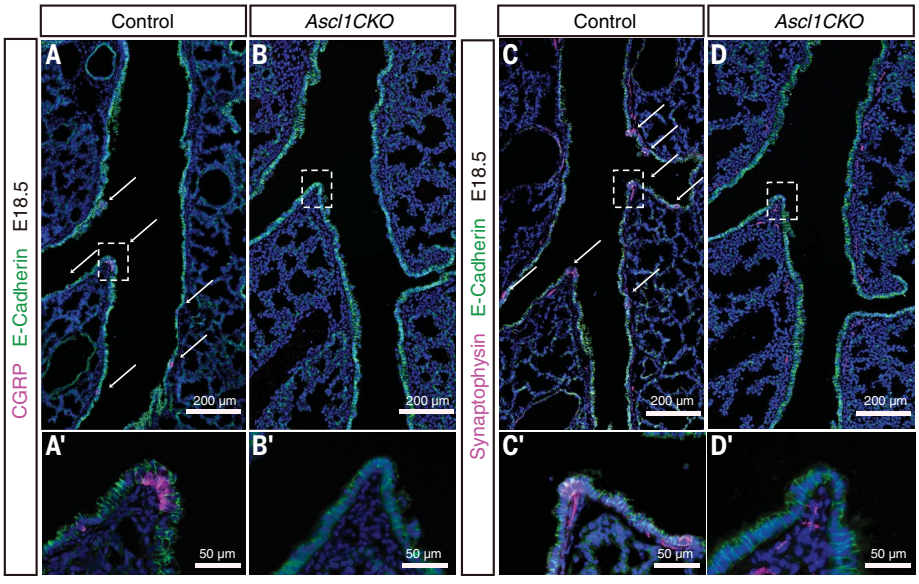


Fig. 1. Inactivation of *Ascl1* by Shh^{cre} in lung epithelium prevents PNEC formation, whereas nerves remain subjacent to airways. (**A** to **D** and **A'** to **D'**) At embryonic day 18.5 (E18.5), CGRP staining outlines PNECs in E-cadherin–positive epithelium. Synaptophysin staining outlines PNECs and nerves, including those that innervate PNECs. Arrows indicate PNECs in the control airway. Boxed areas in the upper panels are magnified in the corresponding lower panels. No CGRP or synaptophysin staining is evident in the epithelium in *Ascl1CKO* mice, indicating a complete absence of PNECs. In the mutant, although the nerves no longer innervate epithelial cells [compare (C') and (D')], the nerve tracks are still apparent nearby. Data are representative of sections from $n = 3$ mice of each genotype.

receptor gene *St2*, but no change in the more general IL-1 family signal mediator gene *Il1rap*, in sorted primary ILC2s (fig. S13). CGRP increased IL-5 production by ILC2s when cultured in the presence of IL-33. However, GABA had a negligible effect (Fig. 3H). In addition, CGRP enhanced IL-5 production in the presence of both IL-25 and IL-33 (Fig. 3I) (*35*). CGRP stimulation did not occur in the absence of IL-25 and IL-33 (Fig. 3J). To complement data from IL-5 enzyme-linked immunosorbent assays (ELISAs), we also assayed an array of candidate cytokines using the LEGENDplex T_H2 system (Biolegend). We confirmed that IL-5 and IL-6 were increased when ILC2s were cultured in the presence of CGRP (fig. S14).

Using CFSE (carboxyfluorescein succinimidyl ester) staining, we observed little change in ILC2

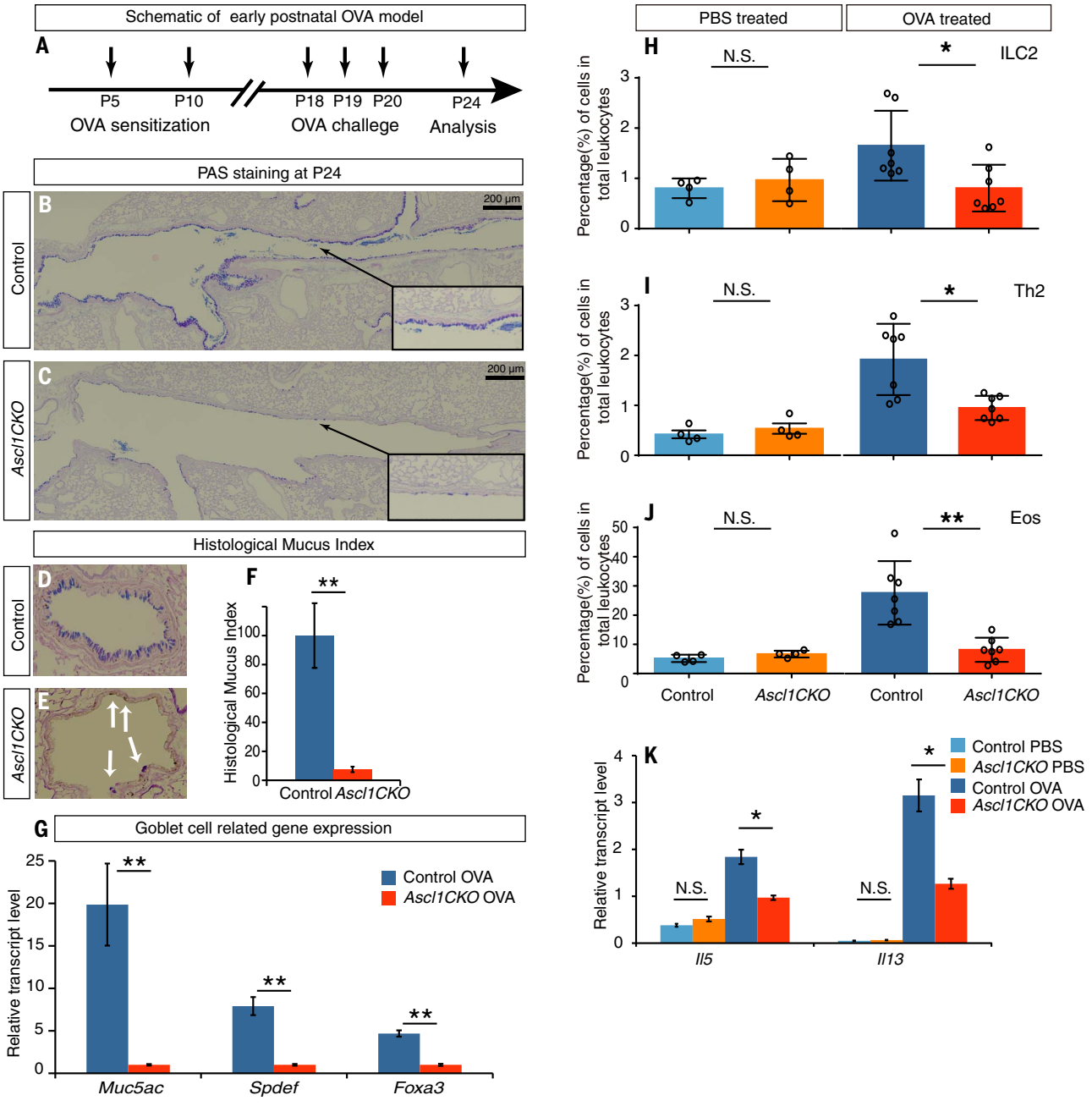


Fig. 2. PNECs are required for goblet cell hyperplasia and immune cell infiltration in a model of asthma. (**A**) Regime of the early postnatal OVA model, indicating postnatal day (P) time points for sensitization, challenge, and analysis. (**B** and **C**) Representative PAS staining of P24 longitudinal airway sections of OVA-challenged control and *Ascl1CKO* mice. Arrows from insets point to the areas that are magnified. (**D** to **F**) Representative PAS staining of transverse sections, from which histological mucus index was calculated by the linear percent of epithelium positive for PAS-stained mucus (n = 3 for each). Student's t test. Arrows in (E) indicate residual PAS staining. (**G**) Goblet cell–related gene expression as assayed by qRT-PCR of P24 whole lungs after OVA challenge (n = 3 for each). Student's t test. (**H** to **J**) Flow cytometry analysis of immune cells from whole lungs, after either PBS (control) or OVA challenge as labeled. Mann-Whitney U test (n = 4 for PBS groups and 7 for OVA groups). (**K**) Cytokine gene expression as assayed by qRT-PCR of P24 whole lungs (n = 3 each). Student's t test. N.S. (not significant), $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$. Data are representative of three experiments. Error bars represent means ± SEM.

division, indicating that CGRP promotes ILC2 cytokine production but not proliferation (fig. S15). IL-33 did not appear to alter CGRP receptor gene expression level in ILC2s, suggesting that CGRP and IL-33 may work in parallel to stimulate ILC2s (table S1). We also tested CGRP and GABA function on sorted T_H2 cells but observed no change in IL-5 production (fig. S16). Furthermore, the addition of CGRP to sorted eosinophils led to only a slight increase in leukotriene C4 production, a minor change compared with the effect of CGRP on ILC2s at the same concentration (fig. S17). Thus, CGRP can directly act on ILC2s to promote their maturation and production of cytokines such as IL-5. Activated ILC2s, in turn, have been shown to recruit eosinophils and trigger a cascade of T_H2 responses (*32*, *33*).

To further validate the PNEC-CGRP-ILC2 axis in vivo, we deleted *Calcrl* in ILC2 cells by using *Il5*^{cre}. When treated with HDMs, mutants showed reduced immune cell infiltration compared with controls, consistent with the notion that CGRP signaling through ILC2s is required for a full T_H2 immune response (Fig. 3, K to M, and fig. S18).

Inactivation of GABA synthesis or transport disrupts goblet cell hyperplasia without affecting immune responses

Increased expression of GABA receptors in lung epithelial cells from asthmatic patients has been reported, and intranasal administration of GABA inhibitor has been shown to suppress OVA-induced goblet cell hyperplasia in mice (*30*). Thus, even though GABA had no effect on ILC2 stimulation, we tested the hypothesis that GABA signaling from PNECs is required for goblet cell hyperplasia in vivo. We disrupted GABA signaling by inactivating *Gad1*, which encodes a GABA synthesis protein, or *Vgat*, which encodes GABA transporter. Because PNECs are the only cells in lung that express these genes, inactivation by *Shh*^{cre} in these animals (hereafter, *Gad1CKO* or *VgatCKO* mice) led to the specific disruption of GABA production from PNECs (*31*). Airway epithelial cell development remained normal in these mutants (fig. S19). However, when treated with OVA as neonates, both mutants showed a striking deficit in goblet cell hyperplasia (Fig. 4, A to D), confirming recent findings (*36*). To test whether GABA is also required in adults for the goblet cell hyperplasia response, we subjected adult *Gad1CKO* mice to the adult OVA model and found dampened goblet cell production of mucus (fig. S20). To further validate that disrupting GABA production in the lung was responsible for reduced goblet cell hyperplasia, we inactivated *Gad1* in the lung epithelium by using *Nkx2-1*^{creERT2} [generating *Nkx2-1*^{creERT2} *Gad1*$^{fl/fl}$ (*Nkx2-1creER;Gad1*) mutant]. *Nkx2-1creER;Gad1* mutants also showed a deficit in goblet cell hyperplasia (fig. S21). These results suggest that GABA production from the PNECs is essential for the goblet cell hyperplasia response.

Next, we addressed whether the goblet cell phenotype in the GABA-signaling mutants was associated with a disruption of immune cell infiltration and cytokine production in vivo. We found that the *Gad1CKO* mice had similar levels of ILC2s, T_H2 cells, and eosinophils compared with controls upon OVA challenge, unlike in the case of the *Ascl1CKO* mutant (Fig. 4, E to G, and fig. S22). *Il5* and *Il13* levels were not significantly different. Thus, consistent with in vitro data (Fig. 3H), GABA signaling from PNECs is essential for goblet cell hyperplasia, but not type 2 immune cell recruitment, in the OVA asthma model.

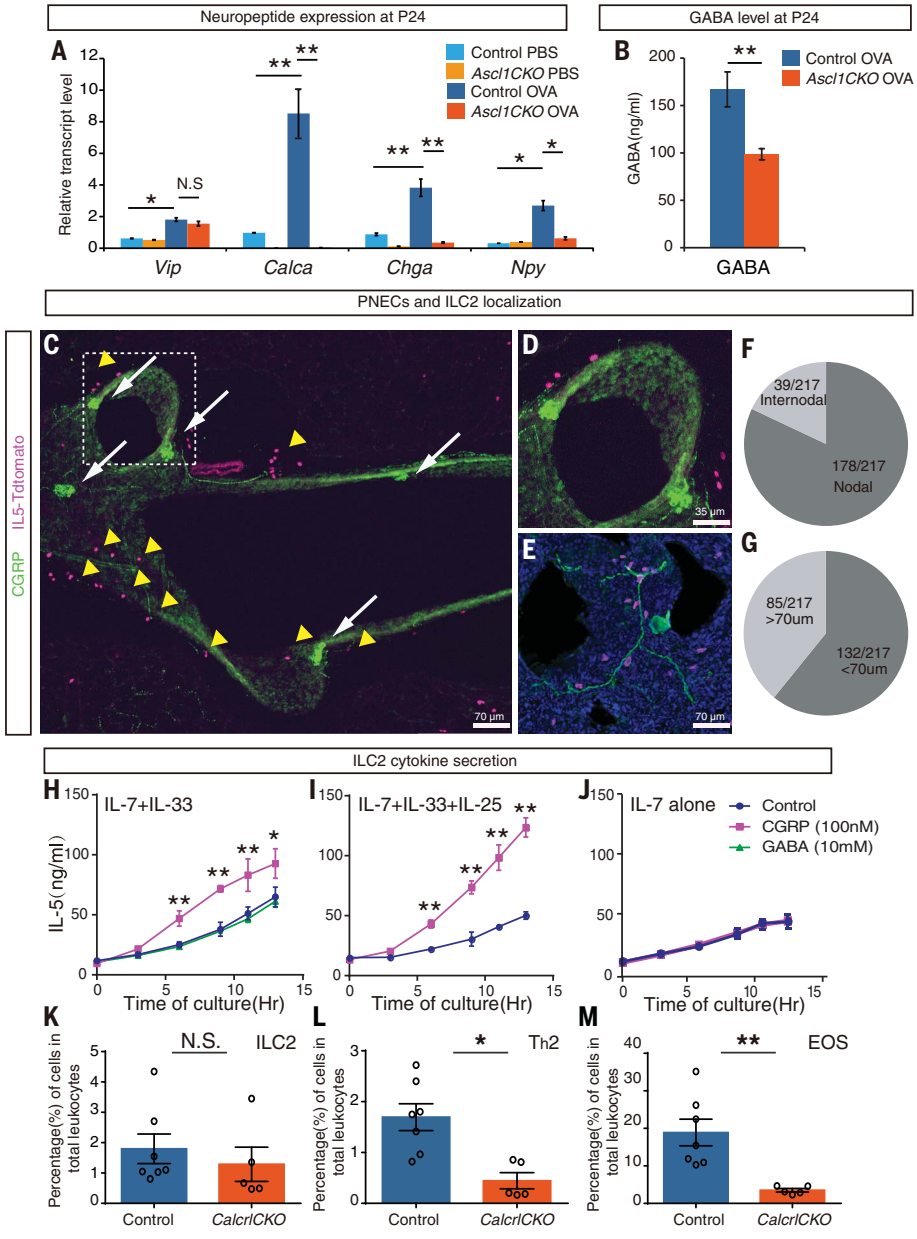


Fig. 3. PNECs reside in proximity to ILC2s and stimulate their cytokine production through CGRP. (**A**) Neuropeptide gene expression as assayed by qRT-PCR in P24 whole lungs after PBS or OVA challenge (n = 3 each). Student's t test. (**B**) GABA level as assayed by ELISA in P24 whole lungs after OVA challenge (n = 3 each). Student's t test. (**C** to **E**) From naïve mice, representative images of PNEC (green, anti-CGRP antibody, indicated by arrows) and ILC2 (magenta, *Il5-tdTomato* reporter, indicated by arrowheads) localization near branch points in longitudinal [(C) and (D)] or transverse (E) vibratome sections of airways. The boxed area in (C) is magnified in (D). Anti-CGRP antibody also labels sensory nerve (E). (**F**) Of the 217 total ILC2s analyzed, 178 (82%) were localized within 70 μm of airway branch (nodal) points. (**G**) Of the 217 total ILC2s analyzed, 132 (61%) of them were localized within 70 μm of PNECs. (**H** to **J**) Effects of CGRP or GABA on IL-5 production by sorted primary ILC2s in culture, in the presence of cytokines indicated at the top left. CGRP, but not GABA, can increase IL-5 secretion as assayed by ELISA (n = 3 each). Student's t test. (**K** to **M**) Data from flow cytometry analysis of immune cells from whole lungs of HDM-treated adult control and mutant mice as labeled. Mann-Whitney U test (n = 7 for each group). N.S., $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$. Error bars represent means ± SEM.

Instillation of CGRP and GABA in the *Ascl1CKO* mutant reverses defective goblet cell hyperplasia and immune cell infiltration

To address whether CGRP and GABA are sufficient to reverse the defective asthma-like response in the absence of PNECs, we introduced a mixture of CGRP and GABA via intratracheal administration immediately after each OVA challenge to mimic the in vivo role of PNECs. On the basis of previously published doses, we used a 40-µl mixture of 100 nM CGRP and 10 mM GABA (*37*, *38*). We found that the exogenous administration of CGRP and GABA was sufficient to restore asthma-like phenotypes in *Ascl1CKO* mice, including goblet cell hyperplasia, immune cell infiltration, and type 2 cytokine expression (Fig. 5 and fig. S23). Other airway cell types were not perturbed (fig. S24). As a control, we also administered the same mixture to wild-type mice that were sensitized but not challenged with OVA. GABA and CGRP did not induce goblet cell hyperplasia in these mice, indicating that these PNEC products collaborate with immune cells to mount goblet cell hyperplasia triggered by allergen challenge (fig. S25). Together, these findings offer an in vivo demonstration that CGRP and GABA are key bioactive products from PNECs that can drive lung allergic responses in the absence of these cells.

PNECs are increased in the airways of human asthma patients

Although PNEC pathology has been documented in a number of lung diseases, whether they change in human asthma has not been directly addressed. To investigate this, we stained lung sections from asthmatics and age-matched controls by using anti-bombesin antibody, which is commonly used to outline all human PNECs. We also used anti-CGRP antibody, which, like in mice, outlines the subset of CGRP-producing PNECs (Fig. 6, A to D, and table S1 for patient characteristics, treatment regime, and primary autopsy findings). We quantified total PNECs, PNECs in the proximal bronchiole and distal respiratory bronchiole, and PNEC cluster size, and we found that all of these were increased in asthmatic samples compared with controls (Fig. 6, E to L). The increase in the CGRP-positive subset of total PNECs was more significant than that in the total bombesin-positive PNECs (P = 0.0052 and 0.0485, respectively). These findings in humans are consistent with our findings in mice and suggest that an increase in PNECs, particularly CGRP-expressing PNECs, may contribute to allergic asthma.

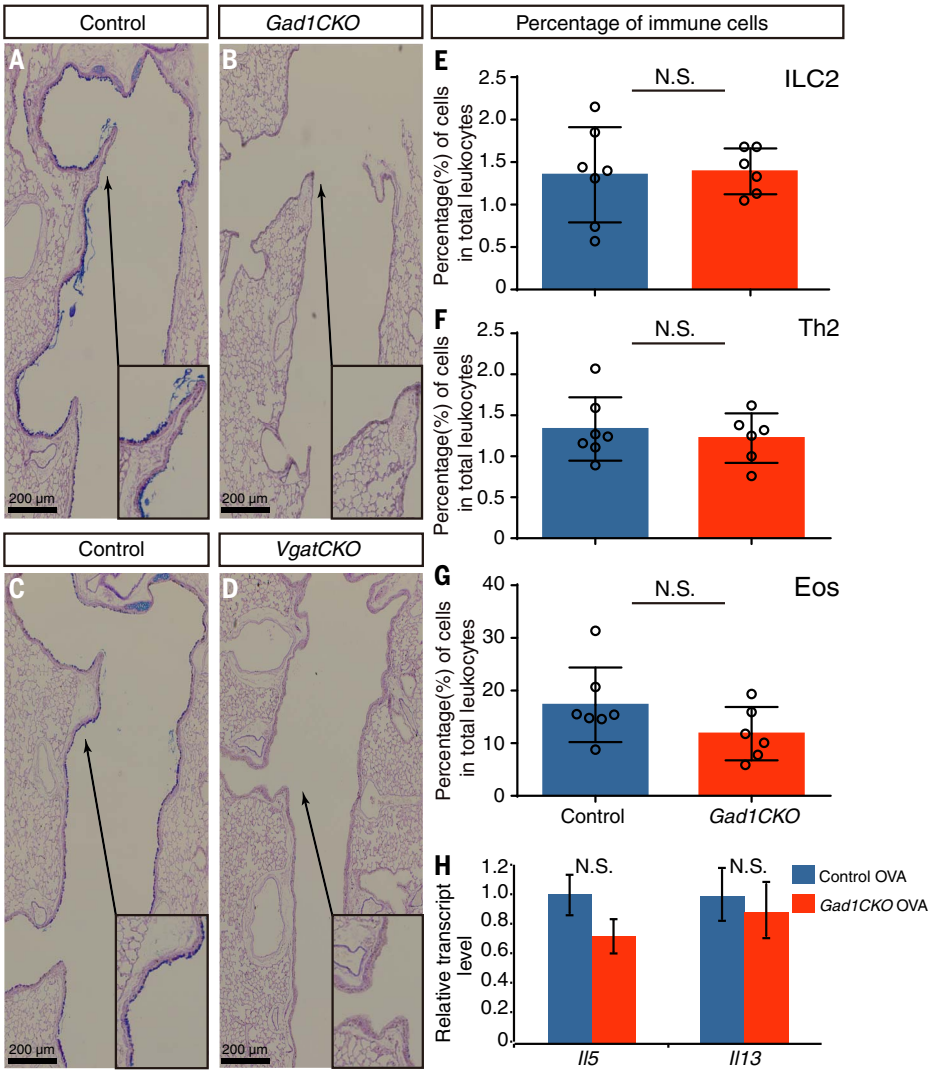


Fig. 4. GABA from PNECs is essential for goblet cell hyperplasia. (**A** to **D**) Representative PAS staining of OVA-challenged airways at P24. Arrows from insets point to the areas that are magnified. (**E** to **G**) Percentage of immune cells in whole lungs after OVA challenge (n = 7 for control and 6 for mutant). Mann-Whitney U test. (**H**) Cytokine gene expression as assayed by qRT-PCR in whole lungs after OVA challenge. Student's t test (n = 3 for each group). N.S., $P \geq 0.05$. Error bars represent means ± SEM.

Discussion

The precise in vivo function of PNECs has been a long-standing question in lung biology. Here we show that PNECs are critical for mounting type 2 immune responses in mouse models of asthma. They act through secreted neuropeptides and neurotransmitters, including CGRP and GABA (Fig. 7A). PNECs signal directly to ILC2s, and together they form a neuroimmunological module to sense and respond to environmental stimuli that enter the airway (Fig. 7B).

The finding that the collaboration between PNECs and ILCs occurs preferentially at airway branch points is intriguing because computational modeling of particle dynamics shows that over time, particles that enter the airway concentrate at branch points (*39*). Thus, cells that reside at these sites are at a prime location to sample diverse inputs, including chemosignals (Fig. 7B). This preferential positioning of PNECs is established in utero, poising these cells to respond to the environment at first breath. In line with growing evidence supporting the impact of early life exposures on the immune system (*40*), we propose that in the early postnatal environment, with every stimulation, the PNEC-ILC2 neuroimmunological modules mature into signaling centers that are critical in airway defense against pathogens and tissue damage.

Our results show that PNECs act through distinct secreted products to control different aspects of the responses to allergens (Fig. 7A). Although GABA synthesis and secretion are essential for goblet cell hyperplasia (*36*), GABA regulates neither cytokine levels nor immune cell infiltration. In contrast, CGRP can stimulate residential ILC2s at close range, inducing them to release IL-5, which can in turn recruit eosinophils and elicit a cascade of immune responses (*32*, *33*). In the absence of PNECs and consequently of the CGRP from these cells, the ILC2 response is severely dampened even though IL-33 is still present. Consistent with this finding, global inactivation of either *Calcrl* or *Ramp1*—genes encoding co-receptors for CGRP—results in diminished lung inflammation upon OVA challenge

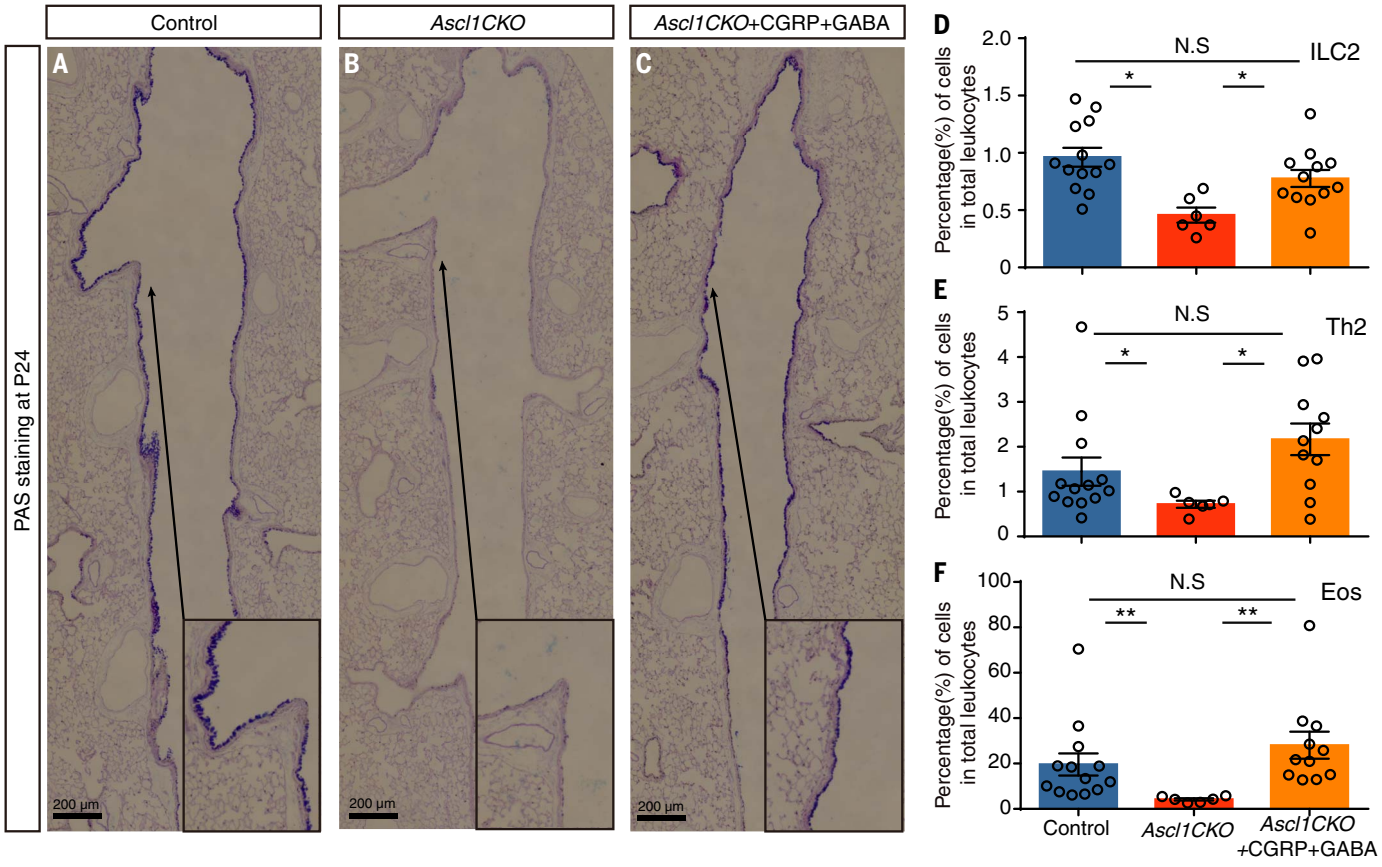


Fig. 5. In vivo restoration of OVA responses in *Ascl1CKO* mutants with intratracheal instillation of CGRP and GABA. (**A** to **C**) PAS staining after OVA challenge. Arrows from insets point to the areas that are magnified. (**D** to **F**) Immune cell numbers from whole lungs after OVA challenge (n = 13 for control, 6 for *Ascl1CKO*, and 11 for *Ascl1CKO* + CGRP + GABA). Mann-Whitney U test. (**G**) Gene expression as assayed by qRT-PCR in whole lungs after OVA challenge. Student's t test (n = 3 for each group). N.S., $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$. Error bars represent means ± SEM.

(*29*). Furthermore, our data show that *Il5*^{cre}-mediated inactivation of *Calcrl* also results in a deficit in the immune response. These results demonstrate that signals from PNECs are important triggers of ILC2 activation in the two allergic asthma models that we used.

It remains unclear how PNECs are activated after allergen challenge. Our data suggest that the increase in CGRP production in the lung is not due to an increase in PNEC cells, but rather to an increase in the levels of CGRP per PNEC. Recent data indicate that innervation of PNECs is required for the increased production of GABA, suggesting that stimulating efferent neurons that innervate PNECs may be a mechanism that can activate PNECs (*36*). It is also possible that allergen-induced upstream immune cells and signals may activate PNECs.

In addition to PNECs, neurons are also a source of neuropeptides in the lung. Although genetic ablation of TRPV1-expressing sensory neurons did not disrupt an OVA-induced immune response, ablation of a larger pool of NAV1.8 nociceptor–expressing sensory neurons blunted eosinophil and T cell responses (*41*, *42*). Several recent studies also showed that neuron-derived neuromedin can directly interact with ILC2s and regulate immune responses in both the intestine and lung (*43*–*45*). Another study showed that TRPV1 neurons, acting through CGRP, regulate protective immunity against lethal *Staphylococcus aureus* pneumonia (*46*). With the increasing awareness that neuropeptides are important regulators of lung immune responses, our findings here establish PNECs as a crucial source of signals.

PNECs are the first specialized cell type that forms in the lung epithelium, suggesting that early establishment of this sensory population is of prime importance to building a functional organ that is responsive to environmental inputs. Our finding that PNECs are increased in asthmatic patient lungs raises the possibility that this increase may contribute to disease. Given the evolutionary conservation of these cells, it is unlikely that PNECs act as the lung "appendix," a site of pathogenesis with no apparent role in homeostasis. Rather, PNECs are likely essential for lung defense, in addition to conferring responses to allergens as demonstrated here. Elucidating the full capacity of PNECs is important to understanding how blocking their function, or the function of their secreted products, can be used safely and effectively to treat allergic lung diseases

such as asthma. Cells similar to PNECs are found in other tissues—for example, enteric endocrine cells in the intestine. These neuroendocrine cells may prove to play critical roles in translating organ-specific environmental inputs into appropriate tissue responses.

Materials and methods

Animals

All mice were housed and all experimental procedures were carried out in American Association for Accreditation of Laboratory Animal Care accredited facilities and labs at either the University of Wisconsin–Madison or University of California, San Diego. This study followed the Guide for the Care and Use of Laboratory Animals. *Shh*^{cre}, *Ascl1*^{fl}, *Vgat*^{fl}, *Gad1*^{fl}, *Calcrl*^{fl}, *Il5*$^{tdTomato\text{-}cre}$ *Nkx2-1*^{creERT2}, and *Calca*$^{lox\text{-}Gfp\text{-}lox\text{-}Dta}$ mice have all been described previously (*20*, *23*, *28*, *47*–*49*).

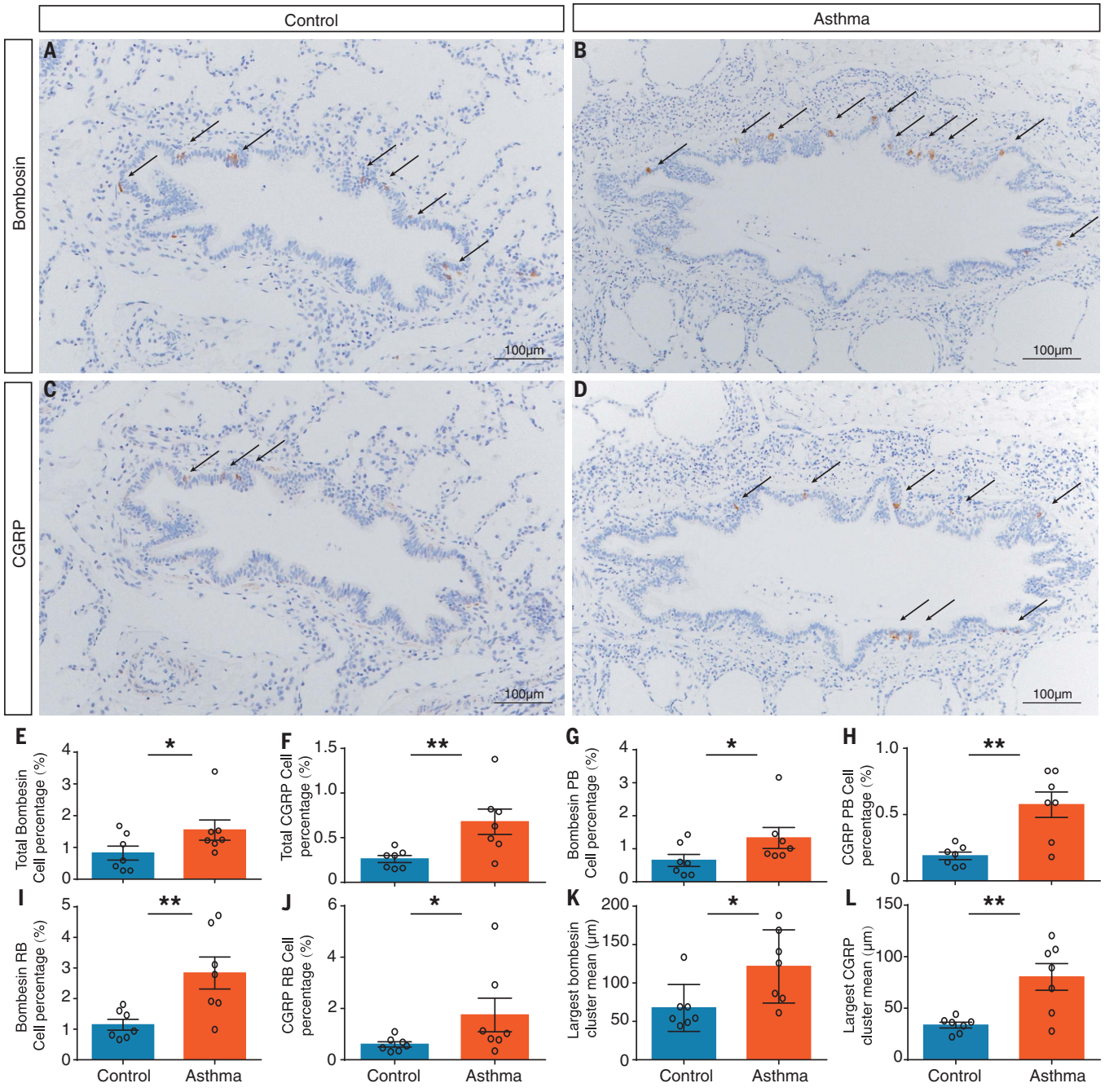


Fig. 6. The number of PNECs is higher in asthmatic lungs. (**A** to **D**) Representative images of control and asthma autopsy lung sections stained with anti-bombesin antibody or anti-CGRP antibody to delineate PNECs expressing each marker. Arrows indicate labeled PNECs. (**E** to **J**) Quantification of the percentage of bombesin- and CGRP-positive epithelial area [total, in the proximal bronchiole (PB), and in the distal respiratory bronchiole (RB)]. Asthmatic samples have higher percentages in all cases. Bombesin: total, $P = 0.0485$; PB, $P = 0.0274$; and RB, $P = 0.0037$. CGRP: total, $P = 0.0052$; PB, $P = 0.0076$; and RB, $P = 0.0274$. (**K** and **L**) Quantification of the mean large bombesin- or CGRP-positive cluster size. Asthmatic samples have larger cluster sizes in both cases. Bombesin, $P = 0.010$; CGRP, $P = 0.0075$. Mann-Whitney *U* test ($n = 7$ controls and 7 asthmatics). Ages, 16 months to 13 years (table S2 shows age-match and other clinical data). $^{*}P < 0.05$; $^{**}P < 0.01$. Error bars represent means ± SEM.

Il5$^{tdTomato\text{-}cre}$, *Gad1*^{fl}, *Calcrl*^{fl}, and *Ascl1*^{fl} mice were on the B6 background. *Shh*^{cre}, *Nkx2-1*^{creERT2}, *Calca*$^{fl\text{-}Gfp\text{-}fl\text{-}Dta}$, and *Vgat*^{fl} mice were back-crossed to the B6 background for at least three generations starting from a mixed background. Littermates were used as controls to minimize potential genetic background effects.

Allergen-induced asthma model

For the early postnatal OVA asthma model, pups were sensitized by intraperitoneal injections of 10 μg of ovalbumin in 10 μl of PBS (OVA, A5503; Sigma) with 10 μl Imject alum (ThermoScientific) on postnatal day (P) 5 and P10, followed by three 15 min challenges with 3% aerosolized OVA solution on P18, P19, and P20. For controls in this protocol, all pups were injected on P5 and P10 with OVA and alum, and challenged on P18, P19, and P20 with PBS. Mice were sacrificed on P24 for analysis.

For the adult OVA asthma model, both mutant and control mice were sensitized by intraperitoneal injections of 50 μg of ovalbumin (OVA, A5503; Sigma) with 10 μl Imject alum (ThermoScientific) at 6 weeks of age on day 0 and day 7, followed by five 15 min challenges with 3% aerosolized OVA solution on days 15, 16, 17, 18, and 19. Mice were sacrificed on day 23 for analysis.

For the adult HDM model, 50 μg of HDM extract (*Dermatophagoides pteronyssinus*, Greer Labs) was introduced intranasally at 4 weeks of age on days 0, 7, 17, and 21. For controls in this protocol, PBS was administered intranasally on the same schedule instead of HDM. Mice were sacrificed on day 24 for analysis.

In *Shh*^{cre};*Calca*$^{lox\text{-}Gfp\text{-}lox\text{-}DTR}$ mutant and corresponding controls, diphtheria toxin (List labs) at 10 ng/g body weight was administered intranasally at P18 and P22. Treated mice were rested for at least 4 days before HDM treatment.

In pregnant females used to generate *Nkx2-1*^{creERT2};*Gad1*$^{fl/fl}$ mutant and corresponding controls, tamoxifen was injected at embryonic days (E) 11 and 12 to induce cre for the inactivation of *Gad1*.

Histology, mean linear intercept, and histological mucus index analysis

Mice were euthanized using CO_2. Lungs were inflated with 4% PFA (paraformaldehyde) at 35 cmH_2O airway pressure, and then fixed overnight. Lungs were then prepared for paraffin (8 μm), cryo (10 μm), or vibratome (100 μm) sectioning. Goblet cells were stained using a periodic acid–Schiff (PAS) staining kit (Sigma). The histological mucus index (the percentage of PAS-positive cells in the bronchial epithelium) was determined by using Image-J software (NIH). To quantify mean linear intercept (MLI), 20× H & E images were used. For each genotype, three mice, three sections per mouse and three independent fields per section were analyzed. Samples were compared using Student's *t* test, by MLI ± SEM with statistical significance set at $P < 0.05$.

Immunostaining

The following primary antibodies were used at the indicated final concentrations for immunofluorescence or immunohistological staining: rabbit anti-Synaptophysin polyclonal antibody [5 μg/ml] (Fisher), rabbit anti-bombesin polyclonal antibody [5 μg/ml] (ImmunoStar), rabbit anti-Calcitonin gene-related peptide/CGRP polyclonal antibody [5 μg/ml] (LifeSpan BioSciences), rabbit anti-CGRP polyclonal antibody [2 μg/ml] (Sigma), rat anti-E-Cadherin polyclonal antibody [5 μg/ml] (Abcam), rabbit anti-SCGB1A1 polyclonal antibody [5 μg/ml] (Seven Hills Bioreagents), rabbit anti-SPC polyclonal antibody [5 μg/ml] (Seven Hills Bioreagents), mouse anti-MUC5AC monoclonal antibody [5 μg/ml] (MRQ-9, Sigma), and syrian hamster anti-T1alpha (PDPN) polyclonal antibody [5 μg/ml] (Developmental Studies Hybridoma Bank). The following secondary antibodies were used: Cy3-conjugated goat anti-rabbit IgG [2 μg/ml], FITC-conjugated goat anti-rabbit IgG [2 μg/ml], FITC-conjugated goat anti-rabbit IgG [2 μg/ml], goat anti-rat FITC [2 μg/ml] (all from Jackson Immunoresearch), HRP-conjugated goat anti-rabbit IgG [2 μg/ml] (Vector Laboratories), and ImmPACT DAB peroxidase (HRP) substrate following manufacturer's protocol (Vector Laboratories). Images were acquired either by ZEISS LSM 880 with Airyscan (Fig. 3, C and D, or by ZEISS Axio Imager 2 (the rest of mouse images), or Nikon DSRi1 digital camera mounted on a Nikon Eclipse 80i microscope (human sample images).

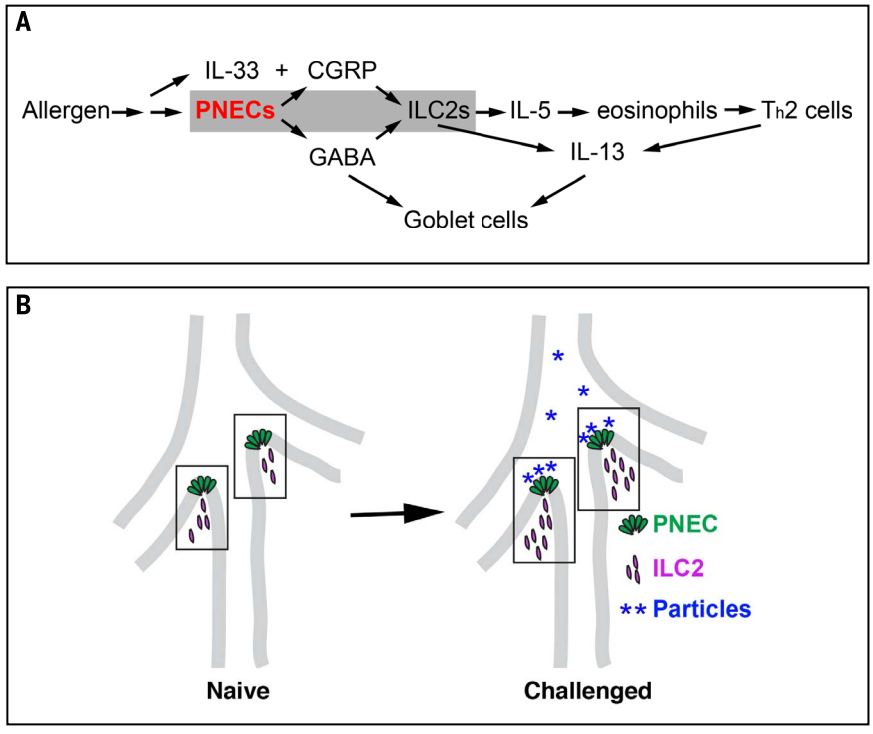


Fig. 7. The PNEC-ILC2 axis functions at airway branch points to amplify the allergic asthma response. (**A**) A model for PNEC function in asthmatic responses. PNECs act through CGRP to stimulate ILC2s and through GABA to stimulate goblet cell hyperplasia. (**B**) A diagram illustrating PNEC-ILC2 interactions in neuroimmunological modules (boxed areas) at airway branch points, where particles carrying allergens or other antigens congregate.

EdU analysis of cell proliferation

For EdU analysis of cell proliferation, ~300 μl of 400 μM EdU solution (ThermoScientific) was intraperitoneally injected. Mice were sacrificed 1 hour after EdU injection. Lungs were fixed in 4% PFA overnight and prepared for cryo sectioning (20 μm). EdU was detected using the Click-iT EdU Kit with Alexa Fluor 488 (Invitrogen).

Tissue processing for cell sorting and flow cytometry

After euthanasia and before tissue harvest, mice were transcardially perfused with 5 ml of cold PBS. Whole lungs were digested by gentle shaking in 5 ml of HBSS with 0.1 Wünsch units (WU) ml of Liberase (Roche) and 25 mg DNase I (Roche) for 60 min at 37°C, and then mechanically dissociated using GentleMACS C tubes (Miltenyi Biotec) followed by straining through a 70-μm filter. The cells were then re-suspended in 40% Percoll, underlined with 66% Percoll and centrifuged. Hematopoietic cells were isolated from the interphase for sorting and analysis.

Cell sorting and flow cytometry

The single-cell suspensions from above were pelleted, aliquoted at ~1 × 10^6 cells per tube, and stained with Fc blocking antibody (5 μg/ml, BD) and Live/Dead Fixable Dead Cell Stain Kit (Invitrogen) at room temperature for 10 min. The cells were washed and then incubated with surface marker antibody cocktail for 45 min at 4°C. For sorting ILC2s and T_H2 cells, the following antibodies were used (all antibodies are from Biolegend, used at 2 μg/ml, unless otherwise specified): BV785-conjugated anti-CD90.2

(30/H12), AF700-conjugated anti-B220 (RA36B2), AF700-conjugated anti-CD11b (M1/70), AF700-conjugated anti-CD11c (N418), AF700-conjugated anti-Nk1.1 (PK136), BV510-conjugated anti-CD4 (GK1.5), PE-Cy7-conjugated anti-TCR beta (H57-597), FITC-conjugated anti-TCR delta (GL3), and PE-conjugated anti-ST2 (DIH9). For sorting eosinophils, the following antibodies were used: PE-conjugated anti-CD64 (X54-5/7.1), PE-Cy7-conjugated anti-CD11c (N418), and BB515 anti-Siglec F (2 μg/ml, E50-2440, BD). Samples were analyzed on an LSR II (BD Biosciences) with four lasers (405 nm, 488 nm, 561 nm, and 635 nm). Data were analyzed with FlowJo software (Treestar).

Cell culture

Freshly sorted ILC2s were cultured in RPMI culture media (Sigma) with 10% (vol/vol) heat-inactivated FBS (from Japan BioSerum), 100 U/ml penicillin/streptomycin, 10 mM HEPES buffer solution, 1× MEM nonessential amino acids, 1 mM sodium pyruvate, 50 μM 2-mercaptoethanol and 50 μg/ml gentamycin sulfate. Cells were cultured with IL-7 (10 ng/ml) and additional indicated cytokines (each at 10 ng/ml) for 12 hours. Then either CGRP (Sigma, 100 nM) or GABA (Sigma, 10 mM) were added. Freshly sorted T_H2 cells were cultured with immobilized anti-CD3 antibody (Biolegend, 145-2C11, 10 μg/mL) with IL-33 (R&D Systems, 10 ng/ml) and IL-7 (R&D Systems, 10 ng/ml), and either CGRP (100 nM) or GABA (100mM). Freshly sorted eosinophils were cultured with IL-5 (R&D Systems, 10 ng/ml), and with or without CGRP (100 nM). Cytokine production were determined by using IL-5 ELISA kit from eBioscience or the Biolegend LEGENDplex T_H2 system. Leukotriene C4 production were measure by using the LTC4 ELISA kit (Cayman Chemical).

Human asthma study population

After Institutional Review Board approval, the pathology database at Seattle Children's Hospital was queried from 1960–2016 to identify autopsies in which asthma was documented in the final report. An age-matched control was obtained from the same year as each asthma case.

Human sample morphometric analysis

All slides were stained, imaged, and quantified together. For each subject, immunostaining for CGRP and bombesin was quantified in up to 12 random peripheral (non–cartilage-containing) airways. Airways were classified as either proximal (comprising membranous and terminal bronchioles with complete muscular walls) or respiratory bronchioles (air passages composed of part muscular bronchial wall and part alveoli). Airways were visualized and captured with a Nikon DSRi1 digital camera mounted on a Nikon Eclipse 80i microscope and analyzed using NIS-Elements Advanced Research Software v4.13 (Nikon Instruments, Melville, NY). All cases were photographed and analyzed using the same magnification, exposure time, lamp intensity, and camera gain. No image processing was carried out prior to intensity analysis. In each airway, the immunostained area was expressed as a percentage of the total airway epithelial area (cell %). The largest immune-positive cluster size was documented for each airway and divided by total airways counted to obtain mean large cluster size.

Statistical methods

As indicated in the figure legends, Student's *t* test was used for parametric data and the Mann-Whitney *U* test was used for nonparametric data. All statistical analyses were performed using Prism 6 (GraphPad).

Primers

All primers are listed in table S3.

REFERENCES AND NOTES

1. L. J. Akinbami, J. E. Moorman, X. Liu, "Asthma prevalence, health care use, and mortality: United States, 2005–2009" (National Health Statistics Report no. 32, National Center for Health Statistics, 2011). pmid: 21355352
2. C. S. Hardman, V. Panova, A. N. McKenzie, IL-33 citrine reporter mice reveal the temporal and spatial expression of IL-33 during allergic lung inflammation. *Eur. J. Immunol.* **43**, 488–498 (2013). doi: 10.1002/eji.201242863; pmid: 23169007
3. S. F. Ziegler, D. Artis, Sensing the outside world: TSLP regulates barrier immunity. *Nat. Immunol.* **11**, 289–293 (2010). doi: 10.1038/ni.1852; pmid: 20300138
4. A. B. Molofsky, A. K. Savage, R. M. Locksley, Interleukin-33 in tissue homeostasis, injury, and inflammation. *Immunity* **42**, 1005–1019 (2015). doi: 10.1016/j.immuni.2015.06.006; pmid: 26084021
5. D. E. Byers *et al.*, Long-term IL-33-producing epithelial progenitor cells in chronic obstructive lung disease. *J. Clin. Invest.* **123**, 3967–3982 (2013). doi: 10.1172/JCI65570; pmid: 23945235
6. R. G. J. Klein Wolterink *et al.*, Pulmonary innate lymphoid cells are major producers of IL-5 and IL-13 in murine models of allergic asthma. *Eur. J. Immunol.* **42**, 1106–1116 (2012). doi: 10.1002/eji.201142018; pmid: 22539286
7. B. N. Lambrecht, H. Hammad, The immunology of asthma. *Nat. Immunol.* **16**, 45–56 (2015). doi: 10.1038/ni.3049; pmid: 25521684
8. C. S. Kuo, M. A. Krasnow, Formation of a neurosensory organ by epithelial cell slithering. *Cell* **163**, 394–405 (2015). doi: 10.1016/j.cell.2015.09.021; pmid: 26435104
9. M. Noguchi, K. Sumiyama, M. Morimoto, Directed migration of pulmonary neuroendocrine cells toward airway branches organizes the stereotypic location of neuroepithelial bodies. *Cell Rep.* **13**, 2679–2686 (2015). doi: 10.1016/j.celrep.2015.11.058; pmid: 26711336
10. I. Brouns *et al.*, Neurochemical pattern of the complex innervation of neuroepithelial bodies in mouse lungs. *Histochem. Cell Biol.* **131**, 55–74 (2009). doi: 10.1007/s00418-008-0495-7; pmid: 18762965
11. E. Cutz, J. Pan, H. Yeger, N. J. Domnik, J. T. Fisher, Recent advances and contraversies on the role of pulmonary neuroepithelial bodies as airway sensors. *Semin. Cell Dev. Biol.* **24**, 40–50 (2013). doi: 10.1016/j.semcdb.2012.09.003; pmid: 23022441
12. X. Gu *et al.*, Chemosensory functions for pulmonary neuroendocrine cells. *Am. J. Respir. Cell Mol. Biol.* **50**, 637–646 (2014). doi: 10.1165/rcmb.2013-0199OC; pmid: 24134460
13. E. Cutz, D. G. Perrin, J. Pan, E. A. Haas, H. F. Krous, Pulmonary neuroendocrine cells and neuroepithelial bodies in sudden infant death syndrome: Potential markers of airway chemoreceptor dysfunction. *Pediatr. Dev. Pathol.* **10**, 106–116 (2007). doi: 10.2350/06-06-0113.1; pmid: 17378691
14. J. E. Gillan, E. Cutz, Abnormal pulmonary bombesin immunoreactive cells in Wilson-Mikity syndrome (pulmonary dysmaturity) and bronchopulmonary dysplasia. *Pediatr. Pathol.* **13**, 165–180 (1993). doi: 10.3109/15513819309048204; pmid: 8464778
15. L. R. Young *et al.*, Neuroendocrine cell distribution and frequency distinguish neuroendocrine cell hyperplasia of infancy from other pulmonary disorders. *Chest* **139**, 1060–1071 (2011). doi: 10.1378/chest.10-1304; pmid: 20884725
16. J. P. van Meerbeeck, D. A. Fennell, D. K. M. De Ruysscher, Small-cell lung cancer. *Lancet* **378**, 1741–1755 (2011). doi: 10.1016/S0140-6736(11)60165-7; pmid: 21565397
17. K. Branchfield *et al.*, Pulmonary neuroendocrine cells function as airway sensors to control lung immune response. *Science* **351**, 707–710 (2016). doi: 10.1126/science.aad7969; pmid: 26743624
18. T. Ito *et al.*, Basic helix-loop-helix transcription factors regulate the neuroendocrine differentiation of fetal mouse pulmonary epithelium. *Development* **127**, 3913–3921 (2000). pmid: 10952889
19. M. Borges *et al.*, An achaete-scute homologue essential for neuroendocrine differentiation in the lung. *Nature* **386**, 852–855 (1997). doi: 10.1038/386852a0; pmid: 9126746
20. K. S. Harris-Johnson, E. T. Domyan, C. M. Vezina, X. Sun, β-Catenin promotes respiratory progenitor identity in mouse foregut. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16287–16292 (2009). doi: 10.1073/pnas.0902274106; pmid: 19805295
21. L. Aven *et al.*, An NT4/TrkB-dependent increase in innervation links early-life allergen exposure to persistent airway hyperreactivity. *FASEB J.* **28**, 897–907 (2014). doi: 10.1096/fj.13-238212; pmid: 24221086
22. M. A. Calderón *et al.*, Respiratory allergy caused by house dust mites: What do we really know? *J. Allergy Clin. Immunol.* **136**, 38–48 (2015). doi: 10.1016/j.jaci.2014.10.012; pmid: 25457152
23. E. S. McCoy *et al.*, Peptidergic CGRPα primary sensory neurons encode heat and itch and tonically suppress sensitivity to cold. *Neuron* **78**, 138–151 (2013). doi: 10.1016/j.neuron.2013.01.030; pmid: 23523592
24. M. Kurowska-Stolarska *et al.*, IL-33 induces antigen-specific $IL\text{-}5^+$ T cells and promotes allergic-induced airway inflammation independent of IL-4. *J. Immunol.* **181**, 4780–4790 (2008). doi: 10.4049/jimmunol.181.7.4780; pmid: 18802081
25. B. Zhou *et al.*, Thymic stromal lymphopoietin as a key initiator of allergic airway inflammation in mice. *Nat. Immunol.* **6**, 1047–1053 (2005). doi: 10.1038/ni1247; pmid: 16142237
26. S. J. Ballantyne *et al.*, Blocking IL-25 prevents airway hyperresponsiveness in allergic asthma. *J. Allergy Clin. Immunol.* **120**, 1324–1331 (2007). doi: 10.1016/j.jaci.2007.07.051; pmid: 17889290
27. S. Li *et al.*, Epithelium-generated neuropeptide Y induces smooth muscle contraction to promote airway hyperresponsiveness. *J. Clin. Invest.* **126**, 1978–1982 (2016). doi: 10.1172/JCI81389; pmid: 27088802
28. J. C. Nussbaum *et al.*, Type 2 innate lymphoid cells control eosinophil homeostasis. *Nature* **502**, 245–248 (2013). doi: 10.1038/nature12526; pmid: 24037376
29. M. Li *et al.*, Deficiency of RAMP1 attenuates antigen-induced airway hyperresponsiveness in mice. *PLOS ONE* **9**, e102356 (2014). doi: 10.1371/journal.pone.0102356; pmid: 25010197
30. Y. Y. Xiang *et al.*, A GABAergic system in airway epithelium is essential for mucus overproduction in asthma. *Nat. Med.* **13**, 862–867 (2007). doi: 10.1038/nm1604; pmid: 17589520
31. K. Schnorbusch *et al.*, GABAergic signaling in the pulmonary neuroepithelial body microenvironment: Functional imaging in GAD67-GFP mice. *Histochem. Cell Biol.* **140**, 549–566 (2013). doi: 10.1007/s00418-013-1093-x; pmid: 23568330
32. T. Y. F. Halim *et al.*, Group 2 innate lymphoid cells are critical for the initiation of adaptive T helper 2 cell-mediated allergic lung inflammation. *Immunity* **40**, 425–435 (2014). doi: 10.1016/j.immuni.2014.01.011; pmid: 24613091
33. K. R. Bartemes *et al.*, IL-33-responsive $lineage^-CD25^+CD44^{hi}$ lymphoid cells mediate innate type 2 immunity and allergic inflammation in the lungs. *J. Immunol.* **188**, 1503–1513 (2012). doi: 10.4049/jimmunol.1102832; pmid: 22198948
34. K. Moro *et al.*, Interferon and IL-27 antagonize the function of group 2 innate lymphoid cells and type 2 innate immune responses. *Nat. Immunol.* **17**, 76–86 (2016). doi: 10.1038/ni.3309; pmid: 26595888
35. K. Moro *et al.*, Innate production of T_H2 cytokines by adipose tissue-associated $c\text{-}Kit^+Sca\text{-}1^+$ lymphoid cells. *Nature* **463**, 540–544 (2010). doi: 10.1038/nature08636; pmid: 20023630
36. J. Barrios *et al.*, Early life allergen-induced mucus overproduction requires augmented neural stimulation of pulmonary neuroendocrine cell secretion. *FASEB J.* **31**, 4117–4128 (2017). doi: 10.1096/fj.201700115R; pmid: 28566470
37. A. B. Young, D. Chu, Distribution of $GABA_A$ and $GABA_B$ receptors in mammalian brain: Potential targets for drug development. *Drug Dev. Res.* **21**, 161–167 (1990). doi: 10.1002/ddr.430210303
38. B. N. Mason *et al.*, Induction of migraine-like photophobic behavior in mice by both peripheral and central CGRP

mechanisms. *J. Neurosci.* **37**, 204–216 (2017). doi: 10.1523/JNEUROSCI.2967-16.2016; pmid: 28053042
39. M. J. Hoegger *et al*., Impaired mucus detachment disrupts mucociliary transport in a piglet model of cystic fibrosis. *Science* **345**, 818–822 (2014). doi: 10.1126/science.1255825; pmid: 25124441
40. E. S. Gollwitzer, B. J. Marsland, Impact of early-life exposures on immune maturation and susceptibility to disease. *Trends Immunol.* **36**, 684–696 (2015). doi: 10.1016/j.it.2015.09.009; pmid: 26497259
41. D. Tränkner, N. Hahne, K. Sugino, M. A. Hoon, C. Zuker, Population of sensory neurons essential for asthmatic hyperreactivity of inflamed airways. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 11515–11520 (2014). doi: 10.1073/pnas.1411032111; pmid: 25049382
42. S. Talbot *et al*., Silencing nociceptor neurons reduces allergic airway inflammation. *Neuron* **87**, 341–354 (2015). doi: 10.1016/j.neuron.2015.06.007; pmid: 26119026
43. A. Wallrapp *et al*., The neuropeptide NMU amplifies ILC2-driven allergic lung inflammation. *Nature* **549**, 351–356 (2017). doi: 10.1038/nature24029; pmid: 28902842
44. C. S. N. Klose *et al*., The neuropeptide neuromedin U stimulates innate lymphoid cells and type 2 inflammation. *Nature* **549**, 282–286 (2017). doi: 10.1038/nature23676; pmid: 28869965
45. V. Cardoso *et al*., Neuronal regulation of type 2 innate lymphoid cells via neuromedin U. *Nature* **549**, 277–281 (2017). doi: 10.1038/nature23469; pmid: 28869974
46. P. Baral *et al*., Nociceptor sensory neurons suppress neutrophil and γδ T cell responses in bacterial lung infections and lethal pneumonia. *Nat. Med.* **24**, 417–426 (2018). doi: 10.1038/nm.4501; pmid: 29505031
47. E. Pacary *et al*., Proneural transcription factors regulate different steps of cortical neuron migration through Rnd-mediated inhibition of RhoA signaling. *Neuron* **69**, 1069–1084 (2011). doi: 10.1016/j.neuron.2011.02.018; pmid: 21435554
48. Q. Tong, C. P. Ye, J. E. Jones, J. K. Elmquist, B. B. Lowell, Synaptic release of GABA by AgRP neurons is required for normal regulation of energy balance. *Nat. Neurosci.* **11**, 998–1000 (2008). doi: 10.1038/nn.2167; pmid: 19160495
49. B. Chattopadhyaya *et al*., GAD67-mediated GABA synthesis and signaling regulate inhibitory synaptic innervation in the visual cortex. *Neuron* **54**, 889–903 (2007). doi: 10.1016/j.neuron.2007.05.015; pmid: 17582330

ACKNOWLEDGMENTS

We thank Q. Wu, K. Caron, and S. Han for sharing mouse strains; J. Whitsett for providing antibody; V. Nizet for the use of a flow cytometer; D. Broide for assistance with the HDM model; L. Nantie for help with imaging; S. Gong and Y. Wang for consultation on the delivery of peptides; and H. Hoffman, G. Haddad, and M. Kronenberg for advice. We thank the University of Wisconsin Carbone Cancer Center Flow Cytometry Laboratory for assistance with the use of flow cytometers. **Funding:** This work was supported by NIH grants OT2OD023857, R01HL113870, R01HL119946, and R01HL122406 and by Wisconsin Partnership Program grant 2897 (to X.S.). **Author contributions:** P.S., D.L.W., J.X., Y.Z., J.L., S.V.D., A.L., C.Y., R.M.L., B.S.K., G.D., and X.S. designed and performed the research; P.S. and D.L.W. analyzed the data; and P.S. and X.S. wrote the manuscript with input from other authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** *Ascl1*^{fl} mice were obtained from F. Guillemot under a material transfer agreement with the Francis Crick Institute, Mill Hill Laboratory. All data are presented either in the body of the paper or in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/eaan8546/suppl/DC1
Figs. S1 to S25
Tables S1 to S3

28 May 2017; resubmitted 11 February 2018
Accepted 21 March 2018
Published online 29 March 2018
10.1126/science.aan8546

RESEARCH ARTICLE

METABOLISM

A *LIMA1* variant promotes low plasma LDL cholesterol and decreases intestinal cholesterol absorption

Ying-Yu Zhang[1,2]*, Zhen-Yan Fu[3]*, Jian Wei[2]*, Wei Qi[4], Gulinaer Baituola[3], Jie Luo[2], Ya-Jie Meng[3], Shu-Yuan Guo[4,5], Huiyong Yin[4,5], Shi-You Jiang[2], Yun-Feng Li[2], Hong-Hua Miao[1], Yong Liu[2], Yan Wang[2], Bo-Liang Li[1], Yi-Tong Ma[3]†, Bao-Liang Song[2]†

A high concentration of low-density lipoprotein cholesterol (LDL-C) is a major risk factor for cardiovascular disease. Although LDL-C levels vary among humans and are heritable, the genetic factors affecting LDL-C are not fully characterized. We identified a rare frameshift variant in the *LIMA1* (also known as *EPLIN* or *SREBP3*) gene from a Chinese family of Kazakh ethnicity with inherited low LDL-C and reduced cholesterol absorption. In a mouse model, LIMA1 was mainly expressed in the small intestine and localized on the brush border membrane. LIMA1 bridged NPC1L1, an essential protein for cholesterol absorption, to a transportation complex containing myosin Vb and facilitated cholesterol uptake. Similar to the human phenotype, *Lima1*-deficient mice displayed reduced cholesterol absorption and were resistant to diet-induced hypercholesterolemia. Through our study of both mice and humans, we identify LIMA1 as a key protein regulating intestinal cholesterol absorption.

Cardiovascular disease (CVD) is the leading cause of death worldwide (*1*), and a high concentration of plasma low-density lipoprotein cholesterol (LDL-C) is a major risk factor (*2*). LDL-C concentration is a complex trait that is influenced by environmental and genetic factors. About 40 to 50% of the phenotypic variances of LDL-C are due to genetic factors (*3*), including mutations in the LDL receptor (*LDLR*) (*4*), autosomal recessive hypercholesterolemia (*ARH*) (*5*), proprotein convertase subtilisin/kexin type *9* (*PCSK9*) (*6*), and Niemann-Pick C1-like 1 (*NPC1L1*) (*7*) genes. However, a recent large-scale analysis showed that only 2.5% of subjects with severely high LDL-C harbored the known genetic variants identified from familial hypercholesterolemia (*8*). In addition, only about 10 to 20% of the total variances in LDL-C can be attributed to the common single-nucleotide variants (SNVs) identified by genome-wide association studies (GWASs). Together, these analyses suggest that the genetic factors influencing LDL-C have not been fully characterized.

Because the majority of human genetic variants are rare and vary among populations with different demographic histories (*9*), examination of a diverged population may help to identify additional susceptible variants. In our search for LDL-C–associated mutations, we focused on Chinese Kazakhs, one of the major ethnic groups in western China that has not (to the best of our knowledge) been included in lipid GWASs to date. The Kazakhs mainly descend from the Turkic and medieval Mongol peoples (*10*) and exhibit marked differences in the SNVs across their genomes. The Kazakhs live in isolated regions and usually marry within their own ethnic group. These features make the Chinese Kazakhs a resourceful population to characterize the ethnic-specific variants associated with LDL-C.

Identification of a *LIMA1* variant associated with lower plasma LDL-C

During the Cardiovascular Risk Survey in western China (*11*), we found a Chinese Kazakh family (named Family 1) with inherited low levels of LDL-C (Fig. 1A). To identify the causal SNV(s), the samples from three subjects exhibiting low LDL-C and one exhibiting normal LDL-C were analyzed by whole-exome sequencing. By using a dominant model, filtering against common variants, and considering functional prediction for the mutations, we narrowed potential candidates to seven SNVs, including a heterozygous frameshift (fs) deletion in the LIM domain and actin binding 1 (*LIMA1*) gene (LIMA1: NM_001113546:exon7:c.916_923del:p.K306fs) on chromosome 12 (fig. S1). LIMA1 is also named as epithelial protein lost in neoplasm (EPLIN) or sterol regulatory element binding protein 3 (SREBP3). LIMA1 shows little sequence similarity to SREBP-1 or -2 but is a homolog of an unknown gene that fused with the N-terminal domain of SREBP2 in SRD-2 cells (*12*). The function of LIMA1 in lipid metabolism has not been described to date.

We validated the seven variants by Sanger sequencing and found *LIMA1*-K306fs SNV to be the only one that cosegregated with the low LDL-C phenotype within this family (Fig. 1, A and B). Genome-wide linkage analysis on nine members of Family 1 revealed that *LIMA1*, but not the other six candidate SNVs, was located in the chromosomal regions with a logarithm of odds (LOD) score > 1.5 (fig. S2). Together, these analyses suggest the *LIMA1*-K306fs SNV as the responsible variant for low LDL-C. The genomic structure of *LIMA1* is shown in Fig. 1C.

Detailed information about Family 1 is listed in table S1. The plasma total cholesterol (TC) and LDL-C levels of *LIMA1*-K306fs carriers (+/K306fs) were significantly lower than those of wild-type (WT) (+/+) individuals (Fig. 1, D and E). However, the plasma levels of triglyceride (TG), high-density lipoprotein cholesterol (HDL-C), and glucose were similar in both *LIMA1*-K306fs and WT individuals (Fig. 1, F to H). Given that LDL-C is influenced by both endogenous cholesterol biosynthesis and intestinal cholesterol absorption, we used gas chromatography–mass spectrometry (GC-MS) to measure the campesterol:lathosterol ratio (Ca:L ratio) (*13*) to estimate relative cholesterol absorption in the plasma of Family 1 members. We found that the *LIMA1*-K306fs mutation was associated with a significantly lower Ca:L ratio than that of the WT members (Fig. 1I), suggesting that *LIMA1*-K306fs carriers have reduced intestinal cholesterol absorption.

To our knowledge, the *LIMA1*-K306fs is a previously unknown mutation and has not been reported in any published databases, including the 121,370 allele–containing Exome Aggregation Consortium (ExAC) database. We further sequenced the coding regions of *LIMA1* in 510 Chinese Kazakh individuals with relatively normal LDL-C concentrations (2.7 to 3.36 mmol/liter) and 509 Chinese Kazakh individuals with low LDL-C concentrations (0.29 to 2.42 mmol/liter). No K306fs variants were found for individuals with normal LDL-C, but an additional K306fs heterozygote was detected in participants with low LDL-C (fig. S3A). In addition, no K306fs variant was found in a total of ~9400 individuals from the Dallas Heart Study and the Biobank Study (fig. S3A). Together, these results demonstrate that K306fs is a rare mutation in different populations, including Chinese Kazakhs (fig. S3A). The K306fs created a premature stop codon and caused a 60% truncation of the LIMA1 protein

[1]The State Key Laboratory of Molecular Biology, Institute of Biochemistry and Cell Biology, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai 200031, China. [2]Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Institute for Advanced Studies, Wuhan University, Wuhan 430072, China. [3]State Key Laboratory of Pathogenesis, Prevention and Treatment of High Incidence Diseases in Central Asia, Heart Center, First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, Xinjiang, China. [4]School of Life Science and Technology, ShanghaiTech University, Shanghai 200031, China. [5]Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China.
*These authors contributed equally to this work.
†Corresponding author. Email: myt-xj@163.com (Y.-T.M.); blsong@whu.edu.cn (B.-L.S.)

(Fig. 1J). This mutation did not change the mRNA stability, and the truncated protein could be detected in humans (fig. S1, C to E).

Targeted sequencing of *LIMA1* in ~1000 Chinese Kazakhs revealed three additional families, with low LDL-C levels, carrying the L25I variant (LIMA1:NM_001113546:exon2:c.73C>A:p.L25I) in *LIMA1* (fig. S3, A and B) (see table S1 for details about the additional families). The *LIMA1*-L25I (Leu^{25}→Ile) mutation cosegregated with the low LDL-C phenotype within the families. The L25I carriers (+/L25I) exhibited lower plasma TC and LDL-C concentrations and a lower Ca:L ratio than WT individuals (fig. S3, C to E). When expressed in cells, the L25I mutation did not affect RNA stability or translational efficiency but resulted in accelerated turnover rate (fig. S3, G to K). The LDL-C levels of the +/L25I individuals were not as low as those of the +/K306fs individuals (compare Fig. 1A with fig. S3B), suggesting that L25I may partially impair LIMA1 function by destabilizing the protein.

Lima1-deficient mice display lower dietary cholesterol absorption

The human genetic study suggested that LIMA1 may have a role in cholesterol metabolism. To investigate its function, we first examined the expression pattern of LIMA1 in mice. LIMA1 was highly expressed in the small intestine, including the duodenum, jejunum, and ileum. It was modestly expressed in the liver and was minimally detectable in the heart, spleen, lung, brain, and pancreas (Fig. 2A). The intestine accounts for ~50% of cholesterol input daily (*14*) and is the major tissue for cholesterol absorption, which is mediated by the key transmembrane protein NPC1L1 (*15*). It is known that NPC1L1 facilitates intestinal cholesterol uptake through a vesicular transport mechanism (*16*). Next, we generated intestine-specific *Lima1*–deficient (I-$Lima1^{-/-}$) mice to investigate the function of LIMA1 (fig. S4A) (*17*). LIMA1 was specifically depleted from the mouse intestine without affecting the level of NPC1L1 (Fig. 2B). Notably, LIMA1 mainly localized on the brush border membrane of the small intestine (Fig. 2C). The I-$Lima1^{-/-}$ mice appeared normal without obvious morphological change in the small intestine, as revealed by immunostaining with antibody to Villin (fig. S4B). By orally administering radiolabeled cholesterol, we observed significantly lower cholesterol uptake in I-$Lima1^{+/-}$ (35.5%) and I-$Lima1^{-/-}$ (28.6%) mice than in WT littermates (51.6%) (Fig. 2D). Two hours after administration, the amount of ^{3}H-cholesterol in the liver was lower in I-$Lima1^{+/-}$ and I-$Lima1^{-/-}$ mice (34.0 and 59.1%, respectively) than in WT mice (Fig. 2E). The amount of plasma ^{3}H-cholesterol in I-$Lima1^{+/-}$ and I-$Lima1^{-/-}$ mice was 29.6 and 54.5% less than in WT mice, respectively (Fig. 2F). The plasma dual-isotope ratio method also showed that cholesterol absorption was reduced by ~40% in *Lima1*-deficient mice (Fig. 2G) (*18*). Together, these results demonstrate that ablation of *Lima1* in the small intestine impairs dietary cholesterol absorption in a gene dosage–dependent manner.

To rule out the possibility that dietary cholesterol might be trapped in the enterocytes following uptake, we performed filipin staining on mouse intestine. A robust cholesterol signal was observed in the enterocytes of WT mice gavaged with cholesterol. However, the cholesterol fluorescence was substantially reduced in *Lima1*- or *Npc1l1*-deficient mice (Fig. 2H). We also measured the cholesterol contents in intestinal epithelial cells after cholesterol gavage and found

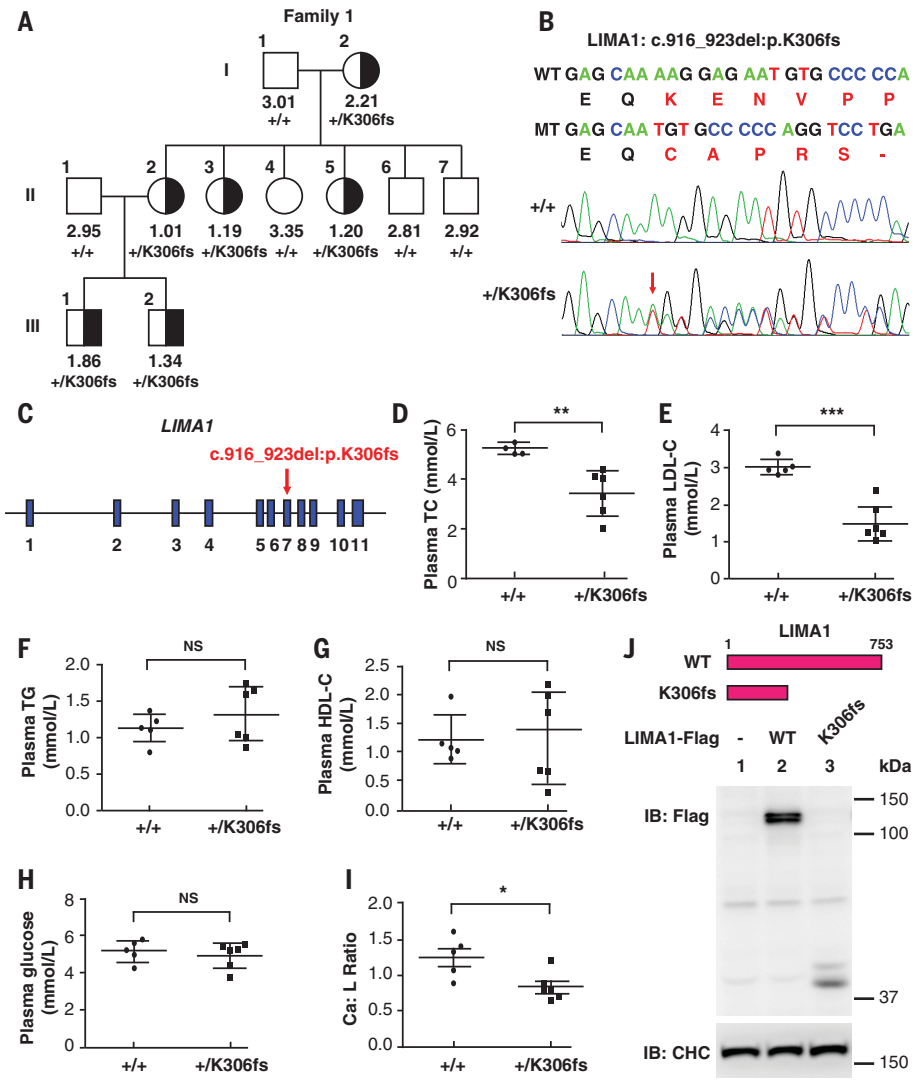


Fig. 1. Identification of the K306fs mutation in the *LIMA1* gene associated with lower plasma LDL-C in a Chinese Kazakh family. (**A**) Pedigree of the Kazakh family with lower levels of plasma LDL-C. Squares and circles indicate males and females, respectively; Roman numerals indicate generations; and Arabic numerals indicate individual family members. Each member's LDL-C concentration (millimoles per liter) and *LIMA1* genotype are shown below the squares and circles. The normal range of LDL-C concentration is 2.7 to 3.1 mmol/liter (*37*). Half-filled squares and circles represent individuals (+/K306fs) with lower LDL-C levels. (**B**) DNA sequencing data of an unaffected man (II: 4) and an affected man (II: 2) with the heterozygous frameshift deletion in *LIMA1*. WT, wild type; MT, mutant; E, Glu; Q, Gln; K, Lys; N, Asn; V, Val; P, Pro; C, Cys; A, Ala; R, Arg; S, Ser. (**C**) Genomic structure of the human *LIMA1* gene. The *LIMA1*-K306fs mutation identified in Family 1 (A) is indicated in red. (**D** to **H**) Plasma TC (D), LDL-C (E), TG (F), HDL-C (G), and glucose (H) levels of the members of Family 1 (A). Data are expressed as mean ± SD. Statistical analyses, unpaired *t* test. $^{**}P < 0.01$; $^{***}P < 0.001$; NS, not statistically significant. (**I**) Plasma Ca:L ratio from family members in (A), measured by GC-MS. Ca, campesterol; L, lathosterol. Data are expressed as mean ± SD. Statistical analyses, unpaired *t* test. $^{*}P < 0.05$. (**J**) Immunoblots (IB) showing the protein levels of LIMA1 variants. Plasmids encoding FLAG-tagged LIMA1(WT) and LIMA1(K306fs) were transfected into CRL1601 cells, and cells were harvested for Western blotting analysis 48 hours later. Immunoblots represent at least three independent experiments. Clathrin heavy chain (CHC) was used as a loading control.

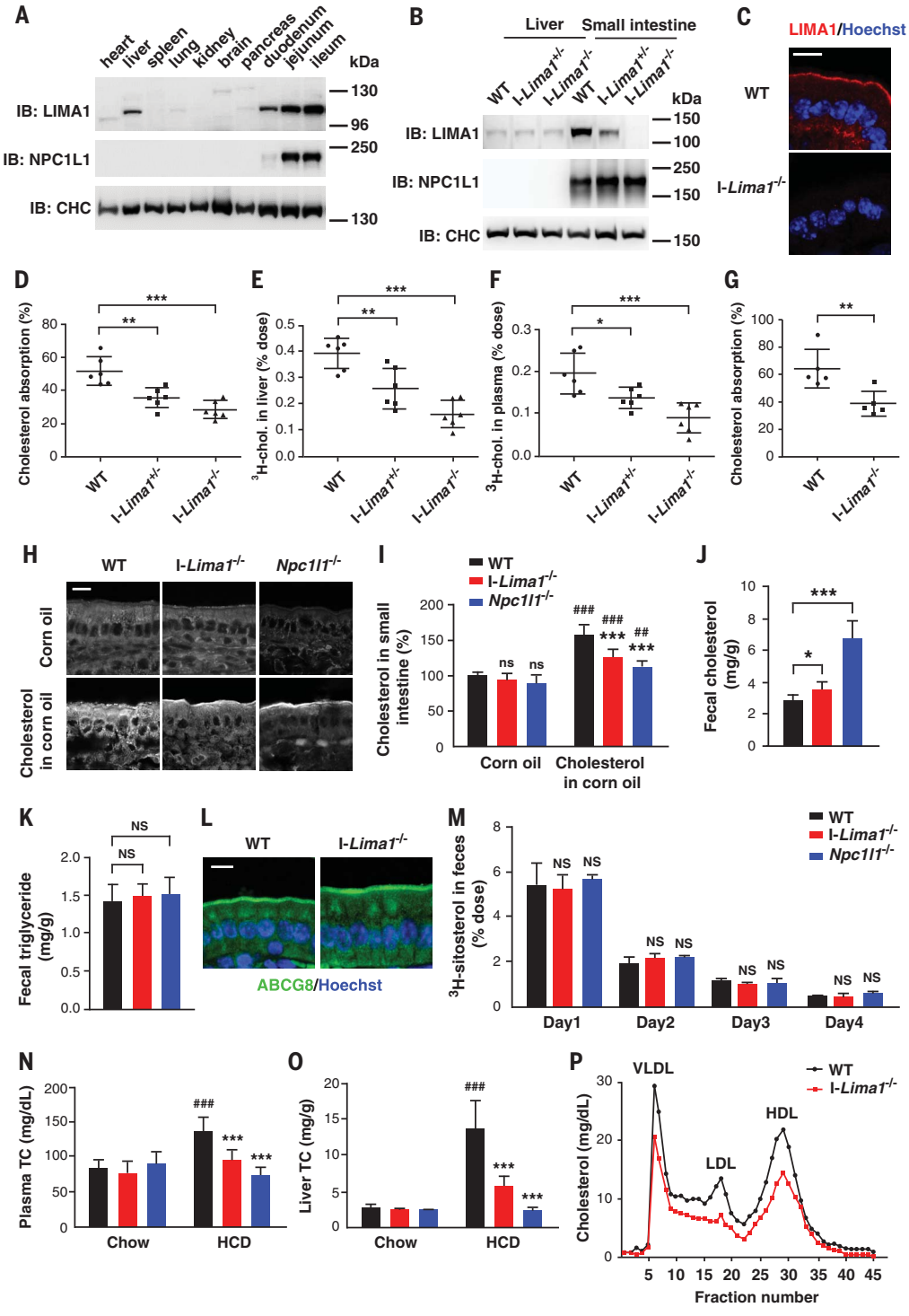


Fig. 2. Metabolic characteristics of intestine-specific *Lima1* knockout mice. (**A**) Expression profile of mouse LIMA1 and NPC1L1 proteins. Tissues taken from 8-week-old male C57BL/6 mice were immediately homogenized and subjected to SDS–polyacrylamide gel electrophoresis followed by immunoblotting with anti-LIMA1, anti-NPC1L1, or anti-CHC. Immunoblots are representative of at least three independent experiments. (**B**) Expression of LIMA1 and NPC1L1 proteins in liver and small intestine from WT, l-*Lima1*$^{+/-}$, and l-*Lima1*$^{-/-}$ mice. Immunoblots are representative of at least three independent experiments. CHC was used as a loading control. (**C**) Representative images showing localization of LIMA1 in the small intestine from 8-week-old male mice (WT or l-*Lima1*$^{-/-}$). Scale bar, 10 μm. (**D**) Cholesterol absorption from mice (8 to 12 weeks, $n = 6$), as measured using the fecal dual-isotope ratio method after oral gavage with ^{14}C-cholesterol and ^{3}H-sitosterol for 3 days. Data are expressed as mean ± SD. Statistical analyses, one-way ANOVA (analysis of variance). $^{**}P < 0.01$; $^{***}P < 0.001$. (**E** and **F**) Amount of ^{3}H-cholesterol in the liver (E) and plasma (F) of mice ($n = 6$) after oral gavage with ^{3}H-cholesterol for 2 hours. Data are mean ± SD. Statistical analyses, one-way ANOVA. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. (**G**) Cholesterol absorption in mice (8 to 12 weeks, $n = 5$), as measured using the plasma dual-isotope ratio method after injection of ^{3}H-cholesterol intravenously and oral gavage with ^{14}C-cholesterol for 3 days. Data are mean ± SD. Statistical analyses, unpaired *t* test. $^{**}P < 0.01$. (**H** and **I**) Analysis of free cholesterol from intestinal samples of l-*Lima1*$^{-/-}$, *Npc1l1*$^{-/-}$, and WT male mice (8 weeks, $n = 6$). (H) Representative images of filipin staining from at least three independent experiments. Scale bar, 10 μm. (I) The levels of cholesterol and phospholipids in the small intestine were measured enzymatically. The relative amount of cholesterol was normalized to that of phospholipids. Data are expressed as mean ± SD. Statistical analyses, two-way ANOVA. $^{\#\#}P < 0.01$ and $^{\#\#\#}P < 0.001$ as compared with mice of the same genotype gavaged with corn oil; $^{***}P < 0.001$ as compared with WT mice in the same group; ns, not statistically significant as compared with WT mice in the same group. (**J** and **K**) Fecal cholesterol and triglyceride levels in WT, l-*Lima1*$^{-/-}$, and *Npc1l1*$^{-/-}$ mice. Feces were collected for 3 days to measure fecal cholesterol (J) and triglycerides (K). Data are expressed as mean ± SD. Statistical analyses, one-way ANOVA. $^{*}P < 0.05$; $^{***}P < 0.001$; NS, not statistically significant. (**L**) Distribution of ABCG8 protein in the small intestine of l-*Lima1*$^{-/-}$ or WT mice (8-week-old mice). Images are representative of at least three independent experiments. Scale bar, 10 μm. (**M**) Fecal excretion of sitosterol in l-*Lima1*$^{-/-}$, *Npc1l1*$^{-/-}$, and WT mice. Data show amount of ^{3}H-sitosterol in the feces and are expressed as mean ± SD for 8- to 12-week-old mice ($n = 6$). Statistical analyses, two-way ANOVA. NS, not statistically significant. (**N** and **O**) l-*Lima1*$^{-/-}$, *Npc1l1*$^{-/-}$, and WT male mice (8 to 12 weeks, $n = 6$) were fed a chow diet or a HCD (1% cholesterol) for 7 days. Plasma TC (N) and liver TC (O) levels were measured. Data are expressed as mean ± SD. Statistical analyses, two-way ANOVA. $^{\#\#\#}P < 0.001$ as compared with WT mice fed the chow diet; $^{***}P < 0.001$ as compared with WT mice fed the HCD. (**P**) Representative cholesterol concentration in FPLC fractions from the serum of l-*Lima1*$^{-/-}$ and WT mice fed the HCD for 7 days.

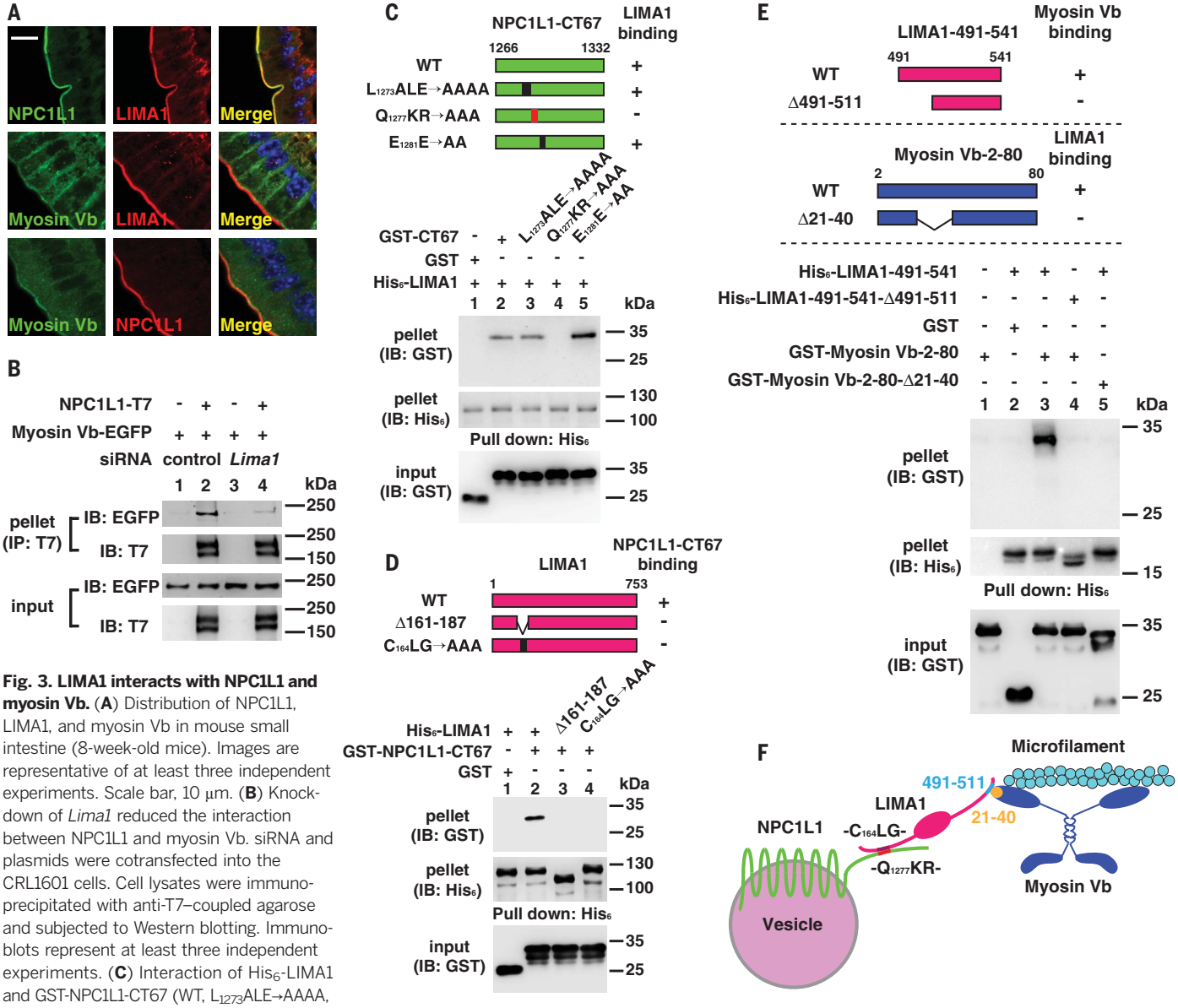


Fig. 3. LIMA1 interacts with NPC1L1 and myosin Vb. (**A**) Distribution of NPC1L1, LIMA1, and myosin Vb in mouse small intestine (8-week-old mice). Images are representative of at least three independent experiments. Scale bar, 10 μm. (**B**) Knockdown of *Lima1* reduced the interaction between NPC1L1 and myosin Vb. siRNA and plasmids were cotransfected into the CRL1601 cells. Cell lysates were immunoprecipitated with anti-T7–coupled agarose and subjected to Western blotting. Immunoblots represent at least three independent experiments. (**C**) Interaction of His_6-LIMA1 and GST-NPC1L1-CT67 (WT, L_{1273}ALE→AAAA, Q_{1277}KR→AAA, or E_{1281}E→AA) proteins analyzed by in vitro pull-down assay. (**D**) Interaction of His_6-LIMA1 (WT, Δ161-187, or C_{164}LG→AAA) with GST-NPC1L1-CT67 proteins analyzed by in vitro pull-down assay. Immunoblots represent at least three independent experiments. (**E**) Interaction of His_6-LIMA1(491-541) (WT or Δ491-511) and GST-myosin Vb(2-80) (WT or Δ21-40) proteins analyzed by in vitro pull-down assay. (**F**) Schematic model of NPC1L1–LIMA1–myosin Vb interaction. The Q_{1277}KR residues of NPC1L1 and the C_{164}LG residues of LIMA1 are required for NPC1L1-LIMA1 interaction. The region (residues 491 to 511) of LIMA1 and the region (residues 21 to 40) of myosin Vb are required for LIMA1–myosin Vb association.

lower levels in *Lima1*- or *Npc1l1*-deficient mice than in WT mice (Fig. 2I). In contrast, fecal cholesterol was higher in I-*Lima1*$^{-/-}$ and *Npc1l1*$^{-/-}$ mice than in WT mice (Fig. 2J), although the fecal triglyceride level was similar (Fig. 2K). These results demonstrate that intestinal cholesterol absorption was reduced in I-*Lima1*$^{-/-}$ mice.

ABCG5 and ABCG8 are known to form an essential protein complex to excrete cholesterol and plant sterols into the intestinal lumen (*19*). We examined whether ablation of *Lima1* affected the localization and function of ABCG5-ABCG8, which would then alter cholesterol homeostasis. The brush border membrane localization of endogenous ABCG8 protein appeared normal in I-*Lima1*$^{-/-}$ mice, as revealed by anti-ABCG8 immunostaining (Fig. 2L). Similarly, knockdown of *Lima1* did not impair the plasma membrane localization of the transfected ABCG5-ABCG8 protein complex in cultured cells (fig. S4C). Functional studies showed that the kinetics of fecal excretion of sitosterol was comparable in I-*Lima1*$^{-/-}$ mice and WT littermates (Fig. 2M), which suggests that *Lima1* ablation affects neither the function of ABCG5-ABCG8 nor the excretion of sterols.

We next evaluated the role of LIMA1 in diet-induced hypercholesterolemia. When fed a chow diet, all mice showed similar TC levels in the plasma and liver (Fig. 2, N and O). In comparison, mice that consumed a high-cholesterol diet (HCD) had 1.63-fold and 4-fold increases in plasma and liver TC levels, respectively. The plasma and liver TC contents of I-*Lima1*$^{-/-}$ were, respectively, 28.8 and 58.3% lower than those of WT mice fed the HCD (Fig. 2, N and O). All mice showed similar plasma and liver TG levels (fig. S4, D and E). Fast protein liquid chromatography (FPLC) analysis showed that cholesterol contents in very-low-density lipoprotein (VLDL), LDL, and HDL were all markedly lower in I-*Lima1*$^{-/-}$ animals than in WT mice fed the HCD (Fig. 2P).

Relative to levels of mice on the chow diet, protein levels of HMG-CoA reductase (HMGCR) decreased in the small intestine and liver of WT mice fed the HCD. However, the HMGCR levels in I-*Lima1*$^{-/-}$ and *Npc1l1*$^{-/-}$ animals were higher than in WT mice on the HCD (fig. S4F). We

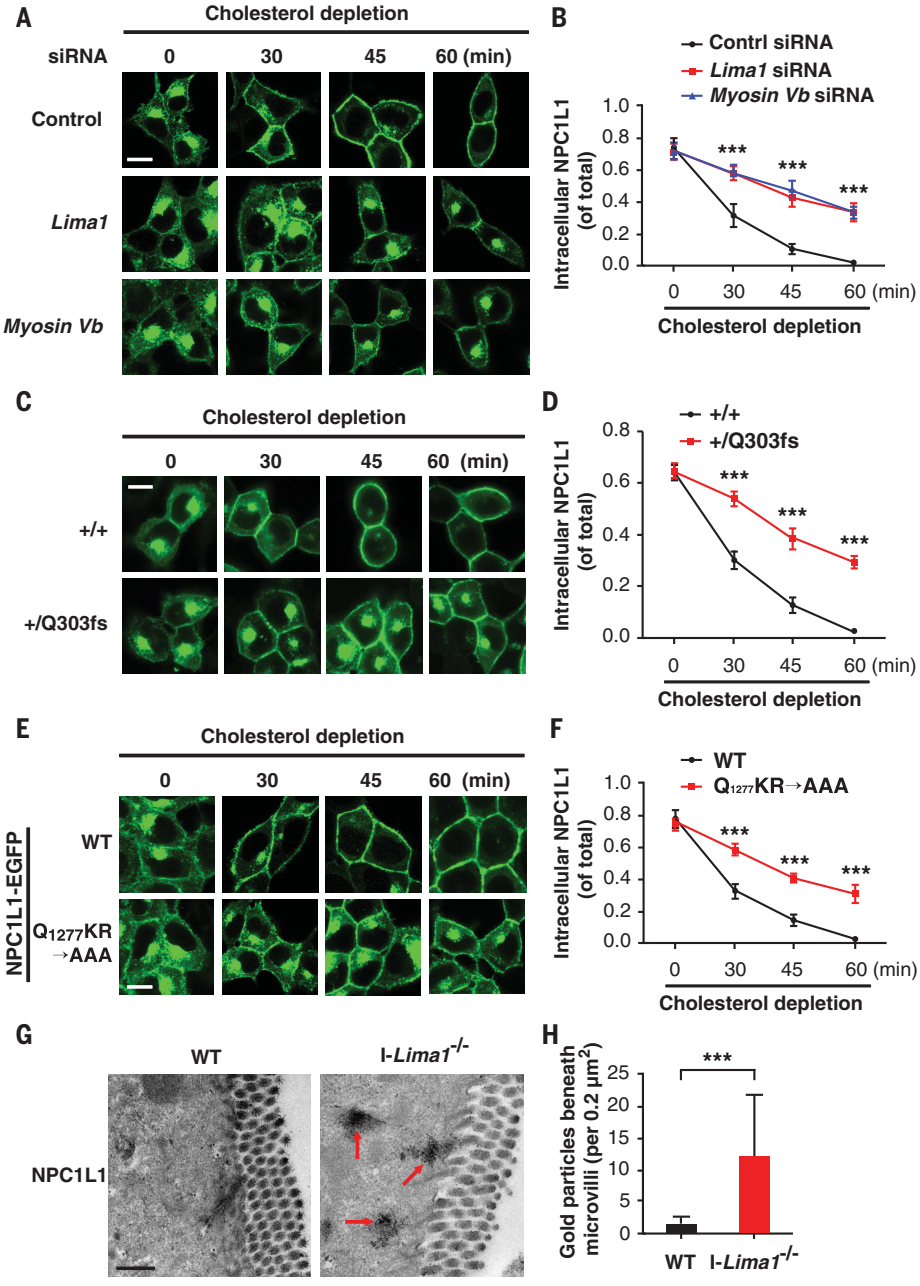


Fig. 4. LIMA1 regulates the NPC1L1 transportation. (**A**) Knockdown of *Lima1* or *myosin Vb* slows transport of NPC1L1 from the ERC to the PM. Control siRNA or siRNA targeting rat *Lima1* or *myosin Vb* was transfected into CRL1601 cells stably expressing NPC1L1-EGFP individually. Forty-eight hours later, the cells were depleted of cholesterol for various time durations. Scale bar, 10 μm. Confocal microscopy images are representative of at least three independent experiments. (**B**) Quantification of the intracellular localization of NPC1L1-EGFP shown in (A). Data are expressed as mean ± SD of three independent experiments. Statistical analyses, two-way ANOVA. ***$P < 0.001$. (**C**) WT (+/+) and heterozygous *Lima1* frameshift deletion (+/Q303fs) CRL1601 cells stably expressing NPC1L1-EGFP were depleted of cholesterol for the indicated times. Confocal microscopy images are representative of at least three independent experiments. Scale bar, 10 μm. (**D**) Quantification of the intracellular localization of NPC1L1-EGFP shown in (C). Data are expressed as mean ± SD of three independent experiments. Statistical analyses, two-way ANOVA. ***$P < 0.001$. (**E**) NPC1L1-EGFP(WT) and NPC1L1-EGFP (Q_{1277}KR→AAA) variants were expressed in CRL1601 cells and depleted of cholesterol for the indicated times. Confocal microscopy images represent at least three independent experiments. Scale bar, 10 μm. (**F**) Quantification of the intracellular localization of NPC1L1-EGFP shown in (E). Data are expressed as mean ± SD of three independent experiments. Statistical analyses, two-way ANOVA. ***$P < 0.001$. (**G**) Representative immunoelectron microscopy images of NPC1L1 in the jejunum of I-*Lima1*$^{-/-}$ and WT mice from three independent experiments. Arrows indicate NPC1L1-positive particles. Scale bar, 500 nm. (**H**) Quantification from (G) of the density of gold particles indicating NPC1L1 protein beneath microvilli. Data are expressed as mean ± SD of three independent experiments. Statistical analyses, unpaired *t* test. ***$P < 0.001$.

observed similar trends in the mRNA levels of *Hmgcr*, *HMG-CoA synthase* 1 (*Hmgcs1*), and *Ldlr* (fig. S4, G to J), indicating that I-*Lima1*$^{-/-}$ mice take up less cholesterol from diet and compensatorily express more cholesterol synthesis– and cholesterol uptake–related genes. Similar phenotypes were detected in *Npc1l1*$^{-/-}$ mice (*15*).

We next generated whole-body *Lima1* knockout mice (fig. S5A). The heterozygotes appeared normal but displayed less dietary cholesterol absorption and lower plasma TC levels as compared with WT mice (fig. S5, B to H), similar to human heterozygotes with the *LIMA1* loss-of-function mutation (Fig. 1A). The homozygotes also appeared normal and displayed reduced dietary cholesterol absorption (fig. S5, I and J). Because ezetimibe is a potent inhibitor of cholesterol absorption through targeting of NPC1L1 (*15*, *20*), we used this drug to pharmacologically inactivate NPC1L1 to address whether LIMA1 is epistatic to NPC1L1. After treatment, cholesterol absorption efficiency in WT mice was reduced to a similar level as in *Npc1l1*-deficient mice (fig. S5K). Ablation of *Lima1* decreased cholesterol absorption, but to a lesser extent than in *Npc1l1* knockout mice. Ezetimibe treatment in I-*Lima1*$^{-/-}$ mice further decreased cholesterol absorption to the same level as in *Npc1l1*-deficient mice (fig. S5K). These data indicate that LIMA1 functions upstream of NPC1L1 in affecting cholesterol absorption.

LIMA1 interacts with NPCIL1 and myosin Vb

To reveal the mechanism by which LIMA1 mediates dietary cholesterol absorption, we immunoprecipitated the LIMA1-containing complex from WT mouse intestinal epithelial cells and identified its binding proteins by tandem mass spectrometry (fig. S6A). The specificity of the antibody to LIMA1 (anti-LIMA1) was validated by Western blotting and immunostaining (Fig. 2, B and C). The anti-LIMA1 immunoprecipitation (IP) from *Lima1*$^{-/-}$ intestinal epithelial cells and the anti-EGFP IP from WT intestinal epithelial cells were similarly performed, and the identified proteins served as nonspecific binders (fig. S6A and table S2). The LIMA1-binding candidates specifically identified from anti-LIMA1 IP of WT samples are listed in fig. S6B and table S2. Cadherin and β-catenin are known LIMA1-binding proteins (*21*), thus validating our purification results. Notably, NPC1L1 and myosin Vb, both required for efficient intestinal cholesterol absorption (*16*, *22*–*30*), were

among the LIMA1-binding candidates, and we found that LIMA1 bound NPC1L1 and myosin Vb (fig. S7A). LIMA1 was mainly present on the brush border membrane of mouse small intestine and colocalized with NPC1L1 and myosin Vb (Fig. 3A). These results suggest that LIMA1 may work together with NPC1L1 and myosin Vb to facilitate cholesterol absorption.

The association between NPC1L1 and myosin Vb was reduced by *Lima1* depletion (Fig. 3B), indicating that LIMA1 may bridge myosin Vb to NPC1L1. Through a series of truncations and mutations followed by co-IP and in vitro pull-down assays, we identified that the Q_{1277}KR residues (Q, Gln; K, Lys; R, Arg) of NPC1L1 and the C_{164}LG residues (C, Cys; G, Gly) of LIMA1 were critical for NPC1L1-LIMA1 interaction (Fig. 3, C and D, and figs. S7 and S8) and that the amino acid 21 to 40 region of myosin Vb and the amino acid 491 to 511 region of LIMA1 mediated myosin Vb–LIMA1 interaction (Fig. 3E and figs. S9 and S10). The binding relationship of LIMA1, NPC1L1, and myosin Vb is shown in Fig. 3F.

LIMA1 regulates NPC1L1 trafficking by recruiting myosin Vb to NPC1L1

We next tested whether LIMA1 played a role in NPC1L1 translocation. Knockdown of *Lima1* or *myosin Vb* blocked transportation of NPC1L1 from the endocytic recycling compartment (ERC) to the plasma membrane (PM) (Fig. 4, A and B). Expression of small interfering RNA (siRNA)–resistant LIMA1 effectively rescued the impaired NPC1L1 translocation in *Lima1*-knockdown cells (fig. S11, A and B), but expression of the truncated protein LIMA1(1-306) failed to do so (fig. S11, C and D). Overexpression of LIMA1(1-306) outcompeted the LIMA1-NPC1L1 interaction (fig. S12A) and delayed the transportation of NPC1L1 from the ERC to the PM (fig. S12, B and C). We also generated a heterozygous *Lima1* frameshift deletion cell line (+/Q303fs) by using CRISPR-Cas9 to mimic the human *LIMA1*-K306fs mutation (fig. S12D). The +/Q303fs cells showed significantly attenuated NPC1L1 transportation to the PM compared with WT cells (Fig. 4, C and D), suggesting a dominant inhibitory function. Given that LIMA1(1-306) was expressed in human carriers (fig. S1, C to E), the +/K306fs heterozygotes may harbor a low Ca:L ratio due to the dominant inhibitory effect of the K306fs allele (Fig. 1I), which also contributes to the low LDL-C concentration in this family (Fig. 1, D and E).

Replacement of Q_{1277}KR with AAA (A, Ala) in NPC1L1, which abolished the NPC1L1-LIMA1 interaction, substantially decreased the transportation rate of NPC1L1 from the ERC to the PM (Fig. 4, E and F). In mouse small intestine, a considerable number of NPC1L1-positive particles accumulated in the apical cytoplasm beneath the microvilli of I-*Lima1*$^{-/-}$ mice compared with WT animals (Fig. 4, G and H). Together, these results suggest that LIMA1 acts as a scaffold protein regulating NPC1L1 transportation to the PM by forming the NPC1L1–LIMA1–myosin Vb triplex (Fig. 3F).

The LIMA1-NPC1L1 interaction is required for cholesterol absorption

Because NPC1L1-LIMA1 interaction is critical for NPC1L1 recycling, ablating this interaction by peptide competition would impair NPC1L1 trafficking and therefore decrease cholesterol absorption. To test this hypothesis, we fused the LIMA1(161-187) region with the transferrin receptor (TrfR) for membrane anchoring (fig. S13A) (*31*). LIMA1(161-187) outcompeted the LIMA1-NPC1L1 interaction, whereas the mutated peptide LIMA1(161-187)-C_{164}LG→AAA [LIMA1(161-187)-mu] failed to do so (fig. S13B). Consistently, LIMA1 (161-187), but not LIMA1(161-187)-mu, substantially decreased the cellular transport of NPC1L1 to the PM (fig. S13, C and D).

We next delivered the fusion protein into mouse liver using adenovirus and investigated the function of the NPC1L1-LIMA1 complex in vivo. In contrast to the human liver, which expresses high levels of NPC1L1 for cholesterol reabsorption from bile, the mouse liver barely expresses NPC1L1. Forced expression of NPC1L1 in mouse liver highly resembled expression in the human setting (*26*, *27*, *32*). Consistent with in vitro results (fig. S13B), LIMA1(161-187)-TrfR fusion, but not LIMA1(161-187)-mu, disrupted the interaction between NPC1L1 and LIMA1 in mouse liver (fig. S13E). Compared with TrfR-expressing mice, an increase in the biliary cholesterol concentration and a decrease in the liver and plasma TC levels were detected in mice expressing the LIMA1(161-187)-TrfR fusion protein (fig. S13F). However, compared with the control, expression of LIMA1(161-187)-mu had no effect on biliary cholesterol, liver TC, or plasma TC concentrations (fig. S13F), indicating that LIMA1(161-187) effectively abrogated NPC1L1-mediated cholesterol reabsorption from bile. Plasma levels of aspartate aminotransferase (AST) were similarly low among different groups (fig. S13F, bottom panel), suggesting no obvious liver injury.

Discussion

Our study identifies that LIMA1 influences plasma cholesterol levels through regulation of intestinal cholesterol absorption in both humans and mice. A growing body of evidence has indicated that lower LDL-C levels are associated with reduced risk of CVD. LDL-C can be decreased by statin drugs, which inhibit the rate-limiting enzyme HMGCR in the cholesterol biosynthesis pathway (*33*), or by ezetimibe, which blocks NPC1L1-mediated intestinal cholesterol absorption (*15*). The PCSK9 inhibitors, which act by elevating LDLR levels, are a third class of drugs that decrease plasma LDL-C and have demonstrated clinical benefits (*34*). Despite the availability of these therapeutics, the prevalence of CVD continues to rise, and many individuals are intolerant to statins and/or ezetimibe or are unable to reach their target LDL-C levels using these strategies (*35*, *36*). Thus, there is still a need to identify targets and therapeutics that provide alternative ways to lower LDL-C and treat CVD, and inhibition of LIMA1 may provide a new direction for treating hypercholesterolemia.

REFERENCES AND NOTES

1. R. Lozano *et al.*, *Lancet* **380**, 2095–2128 (2012).
2. W. B. Kannel, T. R. Dawber, A. Kagan, N. Revotskie, J. Stokes III, *Ann. Intern. Med.* **55**, 33–50 (1961).
3. G. Pilia *et al.*, *PLOS Genet.* **2**, e132 (2006).
4. H. H. Hobbs, M. S. Brown, J. L. Goldstein, *Hum. Mutat.* **1**, 445–466 (1992).
5. C. K. Garcia *et al.*, *Science* **292**, 1394–1398 (2001).
6. M. Abifadel *et al.*, *Nat. Genet.* **34**, 154–156 (2003).
7. Myocardial Infarction Genetics Consortium Investigators, *N. Engl. J. Med.* **371**, 2072–2082 (2014).
8. N. S. Abul-Husn *et al.*, *Science* **354**, aaf7000 (2016).
9. S. Gravel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11983–11988 (2011).
10. H. B. Paksoy, Z. V. Togan, *Centr. Asian Surv.* **11**, 83–100 (1992).
11. Y. N. Yang *et al.*, *PLOS ONE* **7**, e35270 (2012).
12. J. Yang, M. S. Brown, Y. K. Ho, J. L. Goldstein, *J. Biol. Chem.* **270**, 12152–12161 (1995).
13. T. A. Miettinen, R. S. Tilvis, Y. A. Kesäniemi, *Am. J. Epidemiol.* **131**, 20–31 (1990).
14. J. J. Repa, D. J. Mangelsdorf, *Annu. Rev. Cell Dev. Biol.* **16**, 459–481 (2000).
15. S. W. Altmann *et al.*, *Science* **303**, 1201–1204 (2004).
16. L. Ge *et al.*, *Cell Metab.* **7**, 508–519 (2008).
17. F. El Marjou *et al.*, *Genesis* **39**, 186–193 (2004).
18. D. B. Zilversmit, *Proc. Soc. Exp. Biol. Med.* **140**, 862–865 (1972).
19. J. Wang *et al.*, *J. Lipid Res.* **56**, 319–330 (2015).
20. A. B. Weinglass *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 11140–11145 (2008).
21. K. Abe, M. Takeichi, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 13–19 (2008).
22. B. B. Chu *et al.*, *J. Biol. Chem.* **284**, 22481–22490 (2009).
23. J. Wang *et al.*, *J. Lipid Res.* **50**, 1653–1662 (2009).
24. L. Ge *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 551–556 (2011).
25. L. J. Wang *et al.*, *J. Biol. Chem.* **286**, 7397–7408 (2011).
26. C. Xie, N. Li, Z. J. Chen, B. L. Li, B. L. Song, *J. Biol. Chem.* **286**, 35933–35942 (2011).
27. J. H. Zhang *et al.*, *J. Biol. Chem.* **286**, 25088–25097 (2011).
28. C. Xie *et al.*, *J. Lipid Res.* **53**, 2092–2101 (2012).
29. P. S. Li *et al.*, *Nat. Med.* **20**, 80–86 (2014).
30. P. Xie *et al.*, *Atherosclerosis* **237**, 609–617 (2014).
31. D. J. Yamashiro, B. Tycko, S. R. Fluss, F. R. Maxfield, *Cell* **37**, 789–800 (1984).
32. R. E. Temel *et al.*, *J. Clin. Invest.* **117**, 1968–1978 (2007).
33. J. L. Goldstein, M. S. Brown, *Arterioscler. Thromb. Vasc. Biol.* **29**, 431–438 (2009).
34. E. A. Stein *et al.*, *N. Engl. J. Med.* **366**, 1108–1118 (2012).
35. D. Lloyd-Jones *et al.*, *Circulation* **121**, 948–954 (2010).
36. B. A. Golomb, M. A. Evans, *Am. J. Cardiovasc. Drugs* **8**, 373–418 (2008).
37. Joint Committee for the Revision of Guideline for the Prevention and Treatment of Dyslipidemia in Chinese Adults, *Chin. Circ. J.* **31**, 10 (2016).

ACKNOWLEDGMENTS

We thank H. Hobbs and J. Cohen at the UT Southwestern Medical Center for help with human genetic studies, Y.-X. Qu and J. Xu for technical assistance, and S. Robine (Institut Curie) for the villin-Cre ERT2 transgenic mice. **Funding:** This work was supported by grants from the MOST of China (2016YFA0500100), NNSF of China (31690102, 31430044, 81260041, 31771568, 31701030, 91754102, U1403221, and 31600651), Xinjiang Key Research and Development Project (2016B03053), Science and Technology Support Project of Xinjiang (2017E0269), 111 Project of Ministry of Education of China (B16036), and the Natural Science Foundation of Hubei Province (2016CFA012). **Author contributions:** B.-L.S. conceived of the project. Y.-Y.Z., Z.-Y.F., J.W., Y.-T.M., and B.-L.S. designed the experiments. Y.-Y.Z., Z.-Y.F., J.W., W.Q., G.B., Y.-J.M., S.-Y.G., S.-Y.J., Y.-F.L., and H.-H.M. performed the experiments. Y.-Y.Z., Z.-Y.F., J.W., W.Q., J.L., H.Y., Y.L., Y.W., B.-L.L, Y.-T.M., and B.-L.S. analyzed the data. Y.-Y.Z., W.Q., and B.-L.S. wrote the paper with input from all authors. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Whole-exome sequencing datasets were deposited in Sequence Read Archive with accession number SRP139972 (www.ncbi.nlm.nih.gov/sra/).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1087/suppl/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 and S2
References (*38*–*44*)

13 August 2017; resubmitted 7 February 2018
Accepted 17 April 2018
10.1126/science.aao6575

REPORT

PLANETARY SCIENCE

Background levels of methane in Mars' atmosphere show strong seasonal variations

Christopher R. Webster[1]*, Paul R. Mahaffy[2], Sushil K. Atreya[3], John E. Moores[4], Gregory J. Flesch[1], Charles Malespin[2], Christopher P. McKay[5], German Martinez[3], Christina L. Smith[4], Javier Martin-Torres[6,7], Javier Gomez-Elvira[8], Maria-Paz Zorzano[6,8], Michael H. Wong[3], Melissa G. Trainer[2], Andrew Steele[9], Doug Archer Jr.[10], Brad Sutter[10], Patrice J. Coll[11], Caroline Freissinet[12], Pierre-Yves Meslin[13], Raina V. Gough[14], Christopher H. House[15], Alexander Pavlov[2], Jennifer L. Eigenbrode[2], Daniel P. Glavin[2], John C. Pearson[1], Didier Keymeulen[1], Lance E. Christensen[1], Susanne P. Schwenzer[16], Rafael Navarro-Gonzalez[17], Jorge Pla-García[8,18], Scot C. R. Rafkin[19], Álvaro Vicente-Retortillo[3], Henrik Kahanpää[20], Daniel Viudez-Moreiras[8], Michael D. Smith[2], Ari-Matti Harri[20], Maria Genzer[20], Donald M. Hassler[19], Mark Lemmon[21], Joy Crisp[1], Stanley P. Sander[1], Richard W. Zurek[1], Ashwin R. Vasavada[1]

Variable levels of methane in the martian atmosphere have eluded explanation partly because the measurements are not repeatable in time or location. We report in situ measurements at Gale crater made over a 5-year period by the Tunable Laser Spectrometer on the Curiosity rover. The background levels of methane have a mean value 0.41 ± 0.16 parts per billion by volume (ppbv) (95% confidence interval) and exhibit a strong, repeatable seasonal variation (0.24 to 0.65 ppbv). This variation is greater than that predicted from either ultraviolet degradation of impact-delivered organics on the surface or from the annual surface pressure cycle. The large seasonal variation in the background and occurrences of higher temporary spikes (~7 ppbv) are consistent with small localized sources of methane released from martian surface or subsurface reservoirs.

Methane is produced in Earth's atmosphere predominantly through biological processes (*1*, *2*). Its existence in an oxidizing atmosphere such as Mars' is recognized as a potential biosignature whose putative sources could include methanogenic bacteria (*1*, *3*, *4*). Alternative nonbiological methane production mechanisms include geological processes such as serpentinization of olivine or pyroxene (*5*), ultraviolet (UV) degradation of indigenous or meteoritically delivered organics (*6*, *7*), formation by the impact of comets (*8*), release from subsurface clathrates (*9*) or gas absorbed in the regolith (*10*, *11*), erosion of basalt with methane inclusions (*12*), or geothermal processes (*13*). As a strong greenhouse gas, methane bursts on early Mars may have been responsible for intermittent lake-forming climates—a process which could be ongoing today (*14*).

There have been numerous reports of methane in Mars' atmosphere by Earth-based remote sensing and from Mars orbit since 2004. None of those observations show a repeatable seasonal, temporal, or spatial trend. From Earth telescopes, a global average value of 10 ± 3 parts per billion by volume (ppbv) was observed in 1999 (*3*), whereas observations in 2003 showed plumes of methane from discrete sources with a summertime maximum of ~45 ppbv near the equator (*4*), but only an upper limit of 7.8 ppbv 3 years later in January 2006 (*15*). Further data taken in February 2006 yielded a detection of 10 ppbv at 45°S to 7°N over the Valles Marineris region, but an upper limit of 3 ppbv outside that region, and no detection (<8 ppbv) in December 2009 (*16*). From Mars orbit, the Planetary Fourier Spectrometer (PFS) on the Mars Express spacecraft measured a global average abundance of 15 ± 5 ppbv from 2004 to 2010 (*17*, *18*). The Thermal Emission Spectrometer on the Mars Global Surveyor spacecraft measured abundances ranging from 5 to 60 ppbv (*19*), although the claims of local variations were later withdrawn (*20*). Published maps of PFS data (*18*) at Curiosity's landing site in Gale crater (4.5°S, 137°E) show an increase over 1 year from ~15 ppbv in Autumn to ~30 ppbv in winter. In situ measurements of Mars methane began soon after the August 2012 landing of the Mars Science Laboratory (MSL) Curiosity rover at Gale crater. The Tunable Laser Spectrometer (TLS) of the Sample Analysis at Mars (SAM) instrument suite initially reported (*21*) a low methane abundance at Gale crater with an upper limit of 1.3 ppbv [95% confidence interval (CI)]. Subsequent observations over a 20-month period showed (*22*) a background level of 0.69 ± 0.25 ppbv and revealed unexpected occasional spikes to 7.2 ± 2.1 ppbv (95% CI).

Existing models including atmospheric transport and circulation (*23*–*26*) are unable to reproduce the reported high concentrations of methane and its spatial and temporal variability, even when including possible clathrate release (*9*), surface/regolith adsorption/desorption (*10*), seasonally variable production from UV breakdown of surface organics (*6*, *7*), or proposed mechanisms of rapid loss (*27*, *28*). Analysis of all methane measurements up to 2016 (*29*) provides little evidence for any correlation between meteor streams and methane plumes as previously suggested (*30*).

The TLS-SAM instrument (*31*) is a two-channel tunable laser spectrometer that uses both direct and second harmonic detection of absorbed infrared (IR) laser light. One channel uses a near-IR tunable diode laser at 2.78 μm to measure carbon, oxygen, and hydrogen isotopic ratios in both the Mars atmosphere and from gases evolved from rock pyrolysis (*32*, *33*). The high spectral resolution provides high sensitivity to methane by resolving its distinct fingerprint spectral pattern of three adjacent R(3) $^{12}CH_4$ lines in the 3.3-μm band (*32*). An interband cascade semiconductor laser source reflects 81 times between two spherical mirrors in a 20-cm-long sample cell of the Herriott design. The sample cell is fitted with high-vacuum microvalves to control evacuation by a SAM turbomolecular pump for empty cell scans or is filled to Mars

[1]NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. [2]NASA Goddard Space Flight Center, Greenbelt, MD, USA. [3]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA. [4]Department of Earth and Space Science and Engineering, York University, Toronto, ON, Canada. [5]NASA Ames Research Center, Moffett Field, CA, USA. [6]Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Luleå, Sweden. [7]Instituto Andaluz de Ciencias de la Tierra, Granada, Spain. [8]Centro de Astrobiología, Instituto National de Tecnica Aerospacial, Madrid, Spain. [9]Geophysical Laboratory, Carnegie Institution for Science, Washington, DC, USA. [10]Jacobs Technology, NASA Johnson Space Center, Houston, TX, USA. [11]Laboratoire Interuniversitaire des Systèmes Atmosphériques, Creteil Cedex, France. [12]Laboratoire Atmosphères, Milieux, Observations Spatiales, Paris Cedex 05, France. [13]Institut de Recherche en Astrophysique et Planétologie, Toulouse, France. [14]Department of Chemistry and Biochemistry and Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, CO, USA. [15]Department of Geosciences, Pennsylvania State University, University Park, PA, USA. [16]School of Environment, Earth and Ecosystem Sciences, Open University, Milton Keynes, UK. [17]Instituto de Ciencias Nucleares, Universidad Nacional Autonoma de Mexico, Ciudad Universitaria, Mexico City, Mexico. [18]Space Science Institute, Boulder, CO, USA. [19]Department of Space Studies, Southwest Research Institute Boulder, Boulder, CO, USA. [20]Finnish Meteorological Institute, Helsinki, Finland. [21]Department of Atmospheric Sciences, Texas A&M University, College Station, TX, USA.
***Corresponding author. Email: chris.r.webster@jpl.nasa.gov**

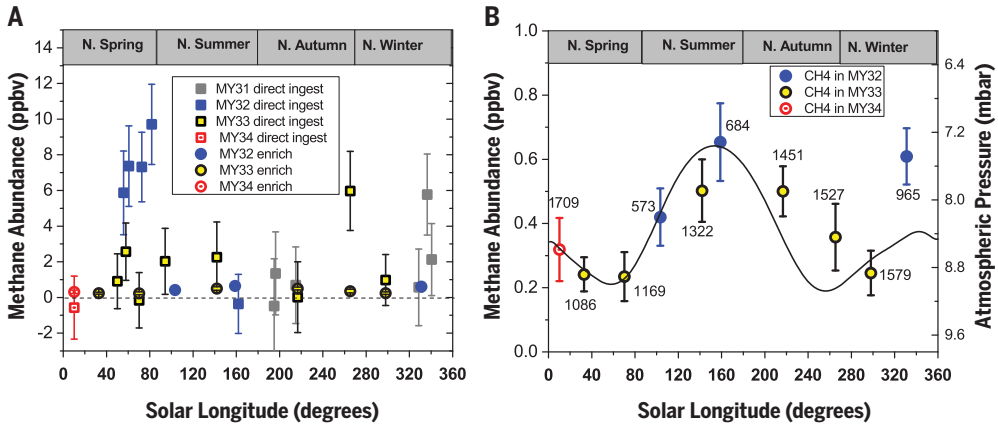


Fig. 1. The TLS-SAM methane measurements versus martian solar longitude. All plotted values are listed in table S2, have error bars of ± 1 SEM, and are corrected to global mean annual values. MY, Mars year. (**A**) All measurements up to 27 May 2017, including those from direct ingests (squares) and enrichment ingests (circles with smaller error bars). (**B**) Background measurements from enrichment ingests show strong seasonal variation. The atmospheric pressure (inverted scale at right) from REMS is plotted for comparison, the solid line representing the mean values over the 3 Mars years.

ambient pressure (~7 mbar) for full cell runs. Because the instrument's foreoptics chamber contains residual terrestrial methane gas, an atmospheric determination is made by subtracting an empty cell measurement from each full cell run (*34*). We used two methods of atmospheric ingestion: The first is a direct ingest method, in which gas is fed into the evacuated sample cell through an inlet port located on the side of the Curiosity rover, taking ~20 min to fill to ~7 mbar and producing uncertainties of ~2 ppbv for each measurement (*34*). The second is an enrichment method that ingests atmospheric gas through a second inlet port, which is passed over a CO_2 scrubber to fill the sample cell more slowly (~2 hours) to the same pressure of ~7 mbar. The latter method efficiently removes incoming CO_2 and H_2O but not methane, effectively enriching its abundance by a factor of 25 ± 4 (*34*). This reduces the overall uncertainty correspondingly, allowing more precise determination of low background levels. Along with other monitoring data, the Herriott cell pressure is recorded every 3 s during the entire ingestion process and data collection.

All our measurements to 27 May 2017 (values listed in table S2), over a period of 55 (Earth) months spanning 3 martian years, are shown in Fig. 1A. We consider values above 3 ppbv to be high spikes of methane; these were occasionally observed in the direct ingest mode. For the high methane spike seen in direct ingest on the four sols (martian days since rover landing) 467, 475, 505, and 525, our improved analysis (*34*) produces a mean value of 7.6 ± 1.6 ppbv (95% CI), which is slightly higher than the 7.2 ppbv previously reported (*22*). All high-precision measurements in the enrichment mode are below 0.7 ppbv and identify the background levels and their associated seasonal variation. These individual background level measurements are given in Table 1, with a mean value of 0.41 ± 0.16 (95% CI) and a variation that ranges from 0.24 to 0.65 ppbv, an increase of nearly a factor of 3 from its lowest value.

As shown in Fig. 1B, the background methane levels have a strong seasonal cycle, peaking near the end of the northern summer/southern winter (Gale crater at 4.5°S, 137.4°E is near the equator). There is no large variation in the mean background level from year to year over this period. The direct ingest measurements with values below 3 ppbv produce a mean value of 0.59 ± 0.54 (95% CI), which is consistent with the higher precision enrichment mean value. This rules out some potential contamination sources because direct ingest uses a different inlet and plumbing from the enrichment runs. Indeed, for the many occasions (table S2) when a direct ingest measurement was run soon after an enrichment run (~4 hours between sample midpoints), the two results agree within the uncertainty of the direct ingest measurement, with only one exception on sol 1527 [solar longitude (L_s) = 265.3°]. On this sol, a 20-min direct atmospheric ingest started 3 hours after the end of a 2-hour atmospheric enrichment ingest. The direct sample contained 5.98 ppbv CH_4, whereas the enriched sample contained 0.27 ppbv CH_4. We attribute this to the arrival at Curiosity of a high spike from a location that could have been up to tens of kilometers away, according to model wind fields (*34*). The enrichment value at L_s = 331° of 0.61 ppbv appears higher than expected for a single modal seasonal variation shape, so we cannot rule out that this may be due to the atmosphere in recovery from a higher spike sometime before that measurement.

An earlier report (*22*) ruled out several mechanisms that may have caused false TLS readings—namely, methane left over from evolved-gas pyrolysis of rock samples, incomplete pumping of the Herriott cell, reactive coatings inside the Herriott cell, wheel degradation or rock-crushing release during transit, and varying surface material under the rover. We argue against the possibility (*35*) that the rover itself is a source of methane because we cannot identify any source large enough to produce even an instantaneous cloud of ~7 ppbv methane in a 10-m-diameter sphere around the rover, which would require ~10^{18} methane molecules (*34*). With typical Mars wind speeds of ~3 to 5 m/s (*36*) replenishing the air around the rover, a supply of ~10^{24} methane molecules would be needed over the 2-month duration of the highest spike period. Although the TLS-SAM fore-optics chamber contains some terrestrial methane [~10^{15} molecules (*34*)], this is too small an amount to be considered as a bulk source for later ingestion even if somehow contained within the rover instrumentation. By monitoring the fore-optics chamber pressure and methane content over the 5-year period on Mars, we see no evidence of gross leakage from the foreoptics chamber.

Correlations of the high methane spikes with other measurements were investigated in an earlier publication (*22*). We compare our background values with the same parameters measured by Curiosity's instruments (*34*): pressure, surface temperature, relative humidity, inferred water vapor abundance, and surface UV from the Rover Environmental Monitoring Station (REMS) instrument; dust opacity from REMS and the Mastcam instrument; radiation flux from the Radiation Assessment Detector (RAD) instrument; and argon measurements from SAM's Quadrupole Mass Spectrometer (QMS).

The rough seasonal trend of a maximum near the end of northern summer is seen in several parameters—including the atmospheric pressure, surface UV, surface temperature, and argon abundance—but comparing these small seasonal changes with our methane background measurements produces no significant bivariate linear correlation (figs. S3 to S38) (*34*).

As with all Solar System bodies, Mars is expected to receive material exogenously (from outside the planet) as infalling dust, micrometeorites, and cometary sources containing organic materials that can partially survive (or be reexposed during atmospheric entry) on the surface, later releasing methane from UV photolytic processes either directly (*7*) or through secondary photochemical reaction (*37*). One such UV/CH_4 model (*7*) predicts that production is carbon-limited and over very long time periods can produce 2.2 ppbv methane in the martian atmosphere. We considered models of methane arising from exogenous material but found that they are inconsistent with the observed background value, its disproportionality with the UV flux, and the size of its seasonal variation (*34*).

The Mars atmosphere has a seasonal surface pressure cycle due to a combination of the

Table 1. Curiosity TLS-SAM methane enrichment measurements at Gale crater (4.5°S, 137.4°E) over a 38-month period. SEM, standard error of the mean; EF, enrichment factor; L_s, solar longitude; CI, confidence interval. The global pressure multiplier is derived from in situ REMS pressure measurements; it is the number by which the original measured in situ values of methane (given in table S3) were multiplied to correct the results to the global mean annual mixing ratio given in the right two columns. Earth dates refer to the time when the gas ingest was started. The decimal portion of the sol is used so that sol 573.08 represents local time 01:57.

Martian sol after landing on 6 August 2012	Earth date	L_s (degrees)	Global pressure multiplier	CH_4 (ppbv)	Error ± 1 SEM (ppbv)
573.08	17 March 2014	103.48	0.970	0.419	0.089
684.06	9 July 2014	158.61	0.877	0.653	0.121
965.99	25 April 2015	331.57	1.003	0.609	0.088
1086.06	26 August 2015	32.81	1.050	0.241	0.053
1169.02	19 November 2015	70.57	1.062	0.235	0.076
1322.00	24 April 2016	142.46	0.881	0.502	0.097
1451.06	4 September 2016	216.58	1.007	0.500	0.078
1527.06	21 November 2016	265.78	1.076	0.357	0.104
1579.00	13 January 2017	298.76	1.036	0.246	0.069
1709.00	27 May 2017	10.84	1.020	0.319	0.098
				Mean value ± 1 SEM (68% CI) = 0.408 ± 0.049 ppbv	
				Mean value including EF error ± 1 SEM (68% CI) = 0.408 ± 0.082 ppbv	
				Mean value ± 2 SEM (95% CI) = 0.41 ± 0.16 ppbv	

condensation and sublimation of carbon dioxide in the polar caps, an effect arising from the difference in mean surface altitude of the two hemispheres, and a dynamical effect resulting from the balance between mass and wind field (*38*). At equatorial locations like Gale crater, the pressure cycle is dominated by the polar cap contribution, although departures from regular seasonal trends can be present locally, sometimes because of dust storms (*39*). In situ measurements from REMS show pressures that range from 7.3 to 9.1 mbar (Fig. 1B), with a mean value of 8.4 mbar. When reporting gas mixing ratios, it is customary to correct in situ values to produce global mean annual mixing ratios, as we have done for the TLS-SAM data (Table 1 and Fig. 1). The observations of spatial and temporal variations (such as the spikes) indicate that normalization to a global mean value may not be appropriate, although this correction is at most only ~15% of the observed amplitude. To explain the large amplitude of the background methane observations, we considered whether large quantities of poorly mixed subliming CO_2 could reach the low latitudes of Gale crater during the higher pressures and thereby result in low mixing ratios locally; this scenario is not borne out by modeling (*36*) nor by the QMS-SAM in situ measurements of argon (*34*), which can be considered a long-lived tracer of atmospheric transport and mixing.

The Mars Regional Atmospheric Modeling System (MRAMS) (*34*) shows that both horizontal mixing and vertical transport play a role in the transport of air into and out of Gale crater. For a small (~2° latitude/longitude area) short (~1 hour) methane release inside Gale crater, all methane is gone (reduced by an order of magnitude) within 6 to 8 hours; when released outside the crater, methane is diluted by several orders of magnitude in similar time, regardless of the season. Mixing time scales are ~1 sol regardless of season, which is much faster than previously thought (*34*). For steady-state release in a small to medium area (~2° to 10° latitude/longitude area) mimicking expectations for release inside or outside the crater, MRAMS shows daily variations of an order of magnitude occurring because of nighttime/daytime differences in flows between crater rim and floor.

With ancient atmospheric pressures of several hundred millibars (*40*), large amounts of methane may be stored in the cold martian subsurface as clathrates in a stability zone several times thicker than that of Earth (*41*–*43*). Although the seasonal signature of the TLS-SAM measurements is not consistent with direct clathrate release, clathrates may provide a source of surface microseepage (diffuse exhalations without any specific morphological structure that may vent from outcropping of rocks or river or lake beds) (*43*–*45*). On Mars, such seepage would occur preferentially through permeable pathways, such as faults, fractures, or in breaches in sealing lithologies; this would not require identifiable geomorphological structures on the surface. Weak microseepage exhalations could explain background and plume methane anomalies observed on Mars (*43*), perhaps near the dichotomy boundary and at Gale crater, where there is fractured sedimentary rock. Microseepage flux may vary over time, depending on variations of gas pressures along the subsurface migration pathway or on seasonal changes in the soil, or even where microbial activity may consume methane.

Regardless of the subsurface origin, methane that finds its way to surface layers over long time periods (*42*, *43*) may be expected to show seasonal variation. We consider a process that retains methane at the surface temporarily before releasing it through a process linked to the surface temperature. That process could be adsorption on a surface with a high surface area–to-volume ratio, such as dust or soil. Although mineral dust cannot serve as a methane sink, it can moderate the release (*11*, *12*). Adopting an energy barrier of ~20 to 35 kJ/mol—which is somewhat higher than that reported for the physical adsorption of methane into clays (*46*), zeolites (*47*), and Mars analog soil (*12*)—we found that large seasonal variations are expected (fig. S41). Plausible correlations of the background methane values with atmospheric water vapor and with surface temperatures point to physical or chemical surface (or dust) processes, or microseepage release. The amplitude of the seasonal cycle indicates that there remain unknown atmospheric or surface processes occurring in present-day Mars.

REFERENCES AND NOTES

1. S. K. Atreya, P. R. Mahaffy, A. S. Wong, *Planet. Space Sci.* **55**, 358–369 (2007).
2. Core Writing Team, Intergovernmental Panel for Climate Change (IPCC), *Fifth Assessment Report 2014*, R. K. Pachauri, L. A. Meyer, Eds. (IPCC, 2014); www.ipcc.ch.
3. V. A. Krasnopolsky, J. P. Maillard, T. C. Owen, *Icarus* **172**, 537–547 (2004).
4. M. J. Mumma *et al.*, *Science* **323**, 1041–1045 (2009).
5. C. Oze, M. Sharma, *Geophys. Res. Lett.* **32**, L10203 (2005).
6. F. Keppler *et al.*, *Nature* **486**, 93–96 (2012).
7. A. Schuerger, J. E. Moores, C. A. Clausen, N. G. Barlow, D. T. Britt, *J. Geophys. Res.* **117**, E08007 (2012).
8. V. A. Krasnopolsky, *Icarus* **180**, 359–367 (2006).
9. E. Chassefière, *Icarus* **204**, 137–144 (2009).
10. P.-Y. Meslin, R. Gough, L. Lefevre, F. Forget, *Planet. Space Sci.* **59**, 247–258 (2010).
11. R. V. Gough, M. A. Tolbert, C. P. McKay, O. B. Toon, *Icarus* **207**, 165–174 (2010).
12. S. McMahon, J. Parnell, N. J. F. Blamey, *Int. J. Astrobiol.* **12**, 113–122 (2013).
13. G. Etiope, D. Z. Oehler, C. C. Allen, *Planet. Space Sci.* **59**, 182–195 (2011).
14. E. S. Kite *et al.*, *Nat. Geosci.* **10**, 737–740 (2017).
15. G. L. Villanueva *et al.*, *Icarus* **223**, 11–27 (2013).
16. V. A. Krasnopolsky, *Icarus* **217**, 144–152 (2012).
17. V. Formisano, S. Atreya, T. Encrenaz, N. Ignatiev, M. Giuranna, *Science* **306**, 1758–1761 (2004).

18. A. Geminale, V. Formisano, G. Sindoni, *Planet. Space Sci.* **59**, 137–148 (2011).
19. S. Fonti, G. A. Marzo, *Astron. Astrophys.* **512**, A51 (2010).
20. S. Fonti *et al.*, *Astron. Astrophys.* **581**, A136 (2015).
21. C. R. Webster, P. R. Mahaffy, S. K. Atreya, G. J. Flesch, K. A. Farley, MSL Science Team, *Science* **342**, 355–357 (2013).
22. C. R. Webster *et al.*, *Science* **347**, 415–417 (2015).
23. F. Lefèvre, F. Forget, *Nature* **460**, 720–723 (2009).
24. S. K. Atreya *et al.*, *Planet. Space Sci.* **59**, 133–136 (2011).
25. K. J. Zahnle, R. S. Freedman, D. C. Catling, *Icarus* **212**, 493–503 (2011).
26. M. A. Mischna, M. Allen, M. I. Richardson, C. E. Newman, A. D. Toigo, *Planet. Space Sci.* **59**, 227–237 (2011).
27. S. K. Atreya *et al.*, *Astrobiology* **6**, 439–450 (2006).
28. W. M. Farrell, G. T. Delory, S. K. Atreya, *J. Geophys. Res.* **33**, L21203 (2006).
29. M. Roos-Serote, S. K. Atreya, C. R. Webster, P. R. Mahaffy, *J. Geophys. Res.* **121**, 2108–2119 (2016).
30. M. Fries *et al.*, *Geochem. Perspect. Lett.* **2**, 10–23 (2016).
31. P. R. Mahaffy *et al.*, *Space Sci. Rev.* **170**, 401–478 (2012).
32. C. R. Webster, P. R. Mahaffy, *Planet. Space Sci.* **59**, 271–283 (2011).
33. C. R. Webster *et al.*, *Science* **341**, 260–263 (2013).
34. Materials and methods are available as supplementary materials.
35. K. Zahnle, *Science* **347**, 370–371 (2015).
36. C. E. Newman *et al.*, *Icarus* **291**, 203–231 (2017).
37. J. R. C. Garry, I. L. ten Kate, Z. Martins, P. Nørnberg, P. Ehrenfreund, *Meteorit. Planet. Sci.* **41**, 391–405 (2006).
38. F. Hourdin, P. Le Van, F. Forget, O. Talagrand, *J. Atmos. Sci.* **50**, 3625–3640 (1993).
39. G. M. Martínez *et al.*, *Space Sci. Rev.* **212**, 295–338 (2017).
40. B. M. Jakosky, R. J. Phillips, *Nature* **412**, 237–244 (2001).
41. B. K. Chastain, V. Chevrier, *Planet. Space Sci.* **55**, 1246–1256 (2007).
42. O. Mousis *et al.*, *Icarus* **278**, 1–6 (2016).
43. D. Z. Oehler, G. Etiope, *Astrobiology* **17**, 1233–1264 (2017).
44. G. Etiope, *Natural Gas Seepage: The Earth's Hydrocarbon Degassing* (Springer, 2015).
45. M. G. Trainer, M. A. Tolbert, C. P. McKay, O. B. Toon, *Icarus* **208**, 192–197 (2010).
46. P. R. Pereira, J. Pires, M. Brotas de Carvalho, *Separ. Purif. Tech.* **21**, 237–246 (2001).
47. S. Y. Zhang, O. Talu, D. T. Hayhurst, *J. Phys. Chem.* **95**, 1722–1726 (1991).

ACKNOWLEDGMENTS

The authors thank the reviewers for constructive comments that greatly improved the manuscript. The research described here was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration (NASA). **Funding:** Funding from NASA's Planetary Science Division is acknowledged by authors C.R.W., P.R.M., S.K.A., G.J.F., C.M., C.P.M., M.H.W., M.G.T., A.S., D.A., C.H.H., R.V.G., A.P., J.L.E., D.P.G., J.C.P., D.K., L.E.C., J.P.-G., S.C.R.R., M.D.S., D.M.H., M.L., J.C., R.W.Z., and A.R.V. R.N.-G. acknowledges funding from the National Autonomous University of Mexico and Consejo Nacional de Ciencia y Tecnología. J.E.M. and C.L.S. acknowledge funding from the Canadian Space Agency MSL participating scientist program. S.P.Sc. acknowledges funding from the UK Space Agency. A.-M.H. acknowledges funding from the Finnish Academy under grant 310509. J.P.-G. acknowledges funding from the Spanish Ministry of Economy and Competitiveness under contract ESP2016-79612-C3-1-R. **Author contributions:** C.R.W. and P.R.M. performed TLS-SAM instrument design, build, and testing (IDBT); surface operations (SO); test-bed activities (TBA); data analysis (DA); data correlations (DC); and science interpretation (SI). G.J.F. and C.M. performed IDBT, SO, TBA, and DA. S.K.A., J.E.M., C.P.M., C.L.S., A.S., D.A., B.S., P.J.C., C.F., P.-Y.M., R.V.G., C.H.H., A.P., J.L.E., D.P.G., S.P.Sa., and R.W.Z. performed SI. J.C. and A.R.V. performed SO. J.C.P., D.K., and L.E.C. performed IDBT. G.M., J.M.-T., J.G.-E., M.-P.Z., M.G.T., S.P.Sc., R.N.-G., A.V.-R., H.K., D.V.-M., M.D.S., A.-M.H., M.G., D.M.H., and M.L. performed DC. J.P.-G. and S.C.R.R. performed DC and SI. **Competing interests:** No potential conflicts of interest exist for any of the listed authors. **Data and materials availability:** The data described in this paper are publicly available from NASA's Planetary Data System (PDS) under an arrangement with the Mars Science Laboratory (MSL) project at http://pds-geosciences.wustl.edu/missions/msl/sam.htm, under the run numbers given in table S2.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1093/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S44
Tables S1 and S2
References (*48*–*62*)

20 September 2017; accepted 4 May 2018
10.1126/science.aaq0131

PLANETARY SCIENCE

Organic matter preserved in 3-billion-year-old mudstones at Gale crater, Mars

Jennifer L. Eigenbrode[1]*, Roger E. Summons[2], Andrew Steele[3], Caroline Freissinet[1]†, Maëva Millan[4]‡, Rafael Navarro-González[5], Brad Sutter[6,7], Amy C. McAdam[1], Heather B. Franz[1], Daniel P. Glavin[1], Paul D. Archer Jr.[6], Paul R. Mahaffy[1], Pamela G. Conrad[1]§, Joel A. Hurowitz[8], John P. Grotzinger[9], Sanjeev Gupta[10], Doug W. Ming[7], Dawn Y. Sumner[11], Cyril Szopa[4], Charles Malespin[1], Arnaud Buch[12], Patrice Coll[13]

Establishing the presence and state of organic matter, including its possible biosignatures, in martian materials has been an elusive quest, despite limited reports of the existence of organic matter on Mars. We report the in situ detection of organic matter preserved in lacustrine mudstones at the base of the ~3.5-billion-year-old Murray formation at Pahrump Hills, Gale crater, by the Sample Analysis at Mars instrument suite onboard the Curiosity rover. Diverse pyrolysis products, including thiophenic, aromatic, and aliphatic compounds released at high temperatures (500° to 820°C), were directly detected by evolved gas analysis. Thiophenes were also observed by gas chromatography–mass spectrometry. Their presence suggests that sulfurization aided organic matter preservation.
At least 50 nanomoles of organic carbon persists, probably as macromolecules containing 5% carbon as organic sulfur molecules.

Organic matter preservation is central to understanding biological potential on Mars through time. Whether it holds a record of ancient life, is the food for extant life, or has existed in the absence of life, organic matter in martian materials holds chemical clues to planetary conditions and processes.

Prior reports of organic matter indigenous to martian sediments include 150 to 300 parts per billion (ppb) of chlorobenzene, with lesser amounts of C_2 to C_4 dichloroalkanes, detected in Sheepbed mudstone upon heating to <400°C by the Sample Analysis at Mars (SAM) instrument suite of the Mars Science Laboratory (MSL) mission (*1*). Martian organic carbon may have also contributed to CO and CO_2 pyrolysis and combustion products evolved below 700°C from various sediments by SAM (*2*, *3*). Chloromethanes detected in Viking's pyrolysis gas chromatography–mass spectrometry (GC-MS) analysis of martian regolith have been reinterpreted as having a martian origin (*4*). Additionally, indigenous organic matter in martian meteorites [e.g., (*5*)] is related to a deep igneous domain excavated by impact events. Although these reports indicate the presence of martian organic matter, they do not constrain our understanding of ancient organic matter in sediments.

Exploration of the lowermost exposed sedimentary rocks at the base of Aeolis Mons in Gale crater by the Curiosity rover has led to the discovery of a finely laminated mudstone succession, the Murray formation, that is interpreted to record deposition in a long-lived ancient circumneutral to alkaline lake fed by a fluviodeltaic sedimentary system (*6*–*8*). Mudstones are composed of basaltic minerals mixed with phyllosilicate, sulfate, iron oxide, and x-ray amorphous components (*7*). The ~3.5-billion-year-old Gale lake environment(s) are expected to have been ideal settings for concentrating and preserving organic matter (*9*).

Drilled samples delivered to SAM were heated at ~35°C/min to ~860°C. Evolved gas analysis (EGA) continuously and directly measured bulk gas composition during heating with the mass spectrometer. A portion of the evolved gas from the same sample was trapped and analyzed by GC-MS for molecular identifications (*10*, *11*). Here we report the analysis of gases evolved at high temperatures (>500°C) from the Mojave and Confidence Hills sites. We also make comparisons with Sheepbed and other Murray mudstones.

Figure 1, A and B, shows mass-to-charge ratio (*m/z*) profiles for the release of organic sulfur compounds from Mojave samples. The profiles, which are indicative of particular compounds or fragments of similar structures, reached their peak between 500° and 820°C, consistent with the presence of thiophene (C_4H_4S), 2- and 3-methylthiophenes (C_5H_6S), methanethiol (CH_4S), and dimethylsulfide (C_2H_6S). The presence of benzothiophene (C_8H_6S), a bicyclic thiophene that usually co-occurs with thiophenes, is also suggested by a weak peak in both Mojave (Fig. 1A) and Confidence Hills (fig. S1F) EGA data. Other volatiles—carbonyl sulfide (COS), CS_2, H_2S, SO_2, O_2, CO, and CO_2 but not H_2—evolved concurrently (Fig. 1, C to E). A similar release of organic sulfur compounds and related volatiles was observed for Confidence Hills (fig. S1).

Other, nonthiophenic, aromatic compounds were also detected in EGA of Mojave (Fig. 2B) and Confidence Hills (fig. S2F) samples. The temperatures of peak release and the corresponding representative molecular groups are ~800°C for benzene (C_6H_6) and ~750°C for toluene (or tropylium ion $C_7H_7^+$), with two peaks at 625° and 790° to 820°C for alkylbenzenes (C_8H_9 or benzoate ion $C_7H_5O^-$) and possible chlorobenzene (C_6H_5Cl) (supplementary text). A peak near detection limits in the Mojave data suggests that naphthalene ($C_{10}H_8$) may also be present, assuming that the Confidence Hills blank (fig. S2E) is representative of noise levels.

EGA mass profiles for *m/z* 15, 26, 27, 29c [where "c" indicates a profile correction to remove contributions from other volatiles (*11*)], 30, 41 to 43, 55 to 57, and 69 to 71 from Mojave (Fig. 2A) and Confidence Hills (fig. S2B) samples are consistent with a 550° to 820°C evolution of aliphatic compound products composed of C_1 to C_5 chains or their branched moieties. Correlation and relative intensities of peaks suggest molecular structures that differ by single carbon additions (–CH–, $–CH_2–$, and $–CH_3$ additions), which are characteristic of an array of aliphatic fragments from larger molecules (*12*) and commonly observed in high-temperature pyrolysis products of terrestrial kerogens and coals (*13*) and carbonaceous chondrites (*14*). Some structures may contain N- and O-bearing groups, such as amides or carboxyl or carbonyl groups, but these cannot be clearly identified in EGA because mass spectra are not resolvable in EGA and other molecules share the diagnostic *m/z* values.

Three temperatures or ranges characterize the peaks of the aliphatic compound signals in the Mojave data (Fig. 2A): 625°C (square), 750°C (circles), and 790° to 820°C (triangles). Immediately preceding the 750°C peak set is a notable O_2 release from sulfate decomposition (*3*), with an increase in CO_2 (Fig. 1E) suggesting that combustion limited to the most ignitable volatiles (*12*) occurred in parallel with pyrolysis. It is also possible that portions of the CO_2 and CO (Fig. 2A) were derived from the decarboxylation (*2*, *3*) and decarbonylation of larger organic compounds, which have been observed for Murchison macromolecular isolates (*15*). The same three peaks are present but less discernable in Confidence Hills data, where the 750°C O_2 peak is lower, suggesting that combustion was less influential on hydrocarbon evolution (fig. S1 and S2).

Abundance estimates for thiophenes, C_1 and C_2 sulfur, and aromatic and aliphatic compounds in Mojave and Confidence Hills samples are in the nanomolar C range (Table 1) (*11*). The total organic carbon (TOC) abundance in these samples (10 to 100 nmol of C) reflects only the portion detected in EGA and does not account for any char remaining in the samples after heating. Calculated TOC abundances are considered lower limits for the actual amounts of organic carbon in the samples.

Organic sulfur and aromatic and aliphatic components evolve between 550° and 820°C and share peak maxima (noted by symbols in Figs. 1 and 2 and figs. S1 and S2). This matching indicates the presence of diverse molecular structures in organic matter of the Mojave and Confidence Hills samples. Peaks for aromatic or thiophenic volatiles in the Sheepbed and other

Murray mudstones (fig. S1 to S6) are weaker and less defined and have inconsistent temperatures, despite a clear indication of aliphatic compound presence. Thiophenic and aromatic compound abundances for these mudstones are less than 50%, as observed for Mojave and Confidence Hills samples (table S1). In the Cumberland sample of the Sheepbed mudstone, thiophene abundances are equivalent to or less than values for blanks, indicating their absence.

GC-MS analysis of the 226° to 860°C cut of gases released from Mojave (Fig. 3) confirms the presence of thiophene, 2-methylthiophene, and 3-methylthiophene. Dimethylsulfide was observed in all GC-MS analyses (table S2). The abundance of thiophenes detected in Mojave by GC-MS is 20 (±5) pmol of molecules (table S2), which is equivalent to ~100 pmol of C, indicating 10 times less thiophenic C than in EGA. The difference reflects a combination of contributions from other unknown molecules [e.g., for the *m/z* 84c profile (Fig. 1A), other thiophenic or sulfur aromatic compounds and C_6H_{12} fragments from aliphatic chains with fewer than six carbons, and for the *m/z* 97 profile, C_7H_{13} fragments or HSO_4^- cleaved from aliphatic and aromatic sulfates] (*12*, *16*) in EGA, leading to inflation of calculated EGA abundances; incomplete hydrocarbon trapping by the SAM hydrocarbon and injection traps of the GC system because of interference from other volatiles; and incomplete release from flash heating of the injection trap.

GC-MS confirmations of molecular identities assigned to EGA aromatic signals are limited. Benzene, alkylbenzenes, naphthalene, and chlorobenzene are observed in GC-MS data for Mojave and Confidence Hills, but molecules from the GC instrument background contribute to these signals, and the GC detections likely include contributions evolved at low temperatures (<500°C) because of the broad GC cut (table S2). Sample-to-sample carryover is also known to affect low GC-MS signals (*1*). Together, these issues make distinction among small amounts of high-temperature, sample-derived aromatic molecules difficult. Moreover, because peaks in EGA profiles reflect the sum of hydrocarbon fragments with a characteristic structure that are contributed by numerous, low-abundance pyrolysis products, signals of individual molecules can be significantly lower in GC-MS. Such diverse molecular contributions are consistent with the complex chemistry of meteoritic and geological organic matter and the interactions that occur during pyrolysis of sediments (*12*). The Mojave and Confidence Hills GC-MS data do not and are not expected to provide unequivocal molecular identifications for the small amounts of aromatic compounds indicated by the EGA data.

Aliphatic compound identifications in GC-MS data for Mojave and Confidence Hills do not reflect those observed in EGA. Generally, the strongest *m/z* values in mass spectra of aliphatic compounds are from the C_2 to C_5 fragments of parent molecules with larger carbon structures. Thus, the disparity between EGA and GC-MS data is largely attributable to C_2 to C_5 fragment contributions from a large molecular pool in EGA, which renders identification of any one molecule below detection limits by GC-MS, as previously described for aromatic compounds.

The SAM EGA instrument background is not the source of the molecular diversity observed at high temperatures (supplementary text). Sample signals are significantly greater than blanks. Laboratory tests demonstrate that in the presence of silicates and perchlorates, instrument background has little effect on signals above 550°C (fig. S7). Other possible but unlikely contamination sources cannot account for the temperature breadth and the molecular diversity observed. Lastly, C_5 structures suggested by *m/z* profiles cannot be explained by SAM's instrument background, which is limited to C_1 to C_4 backbones.

The diversity, composition, and temperatures of coevolving volatiles observed in the Mojave and Confidence Hills analyses above 500°C are consistent with the pyrolysis of geologically refractory organic macromolecules that are typically found in carbonaceous chondrites (*14*, *15*), kerogens (*17*), and coals (*18*, *19*). The more stable these macromolecules are, the higher the temperature needed to thermally cleave fragments from them. Pyrolysis of organic matter-laden sediments with co-occurring inorganic materials that also decompose or act as catalysts can result in a complex array of chemical reactions during heating, including sulfurization, addition, cyclization, and condensation (Diels-Alder

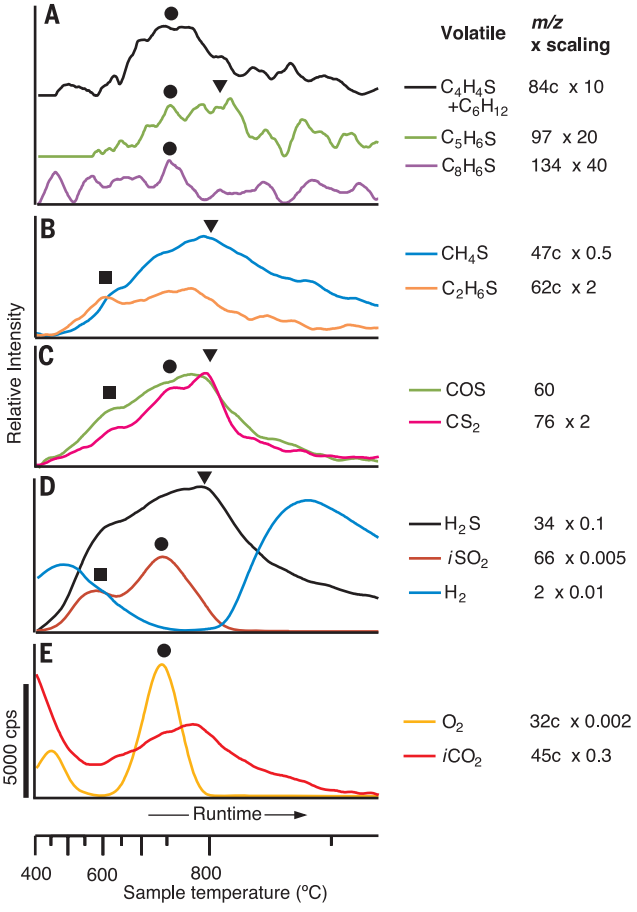


Fig. 1. Mojave EGA profiles for volatiles. Profiles for thiophenes (**A**), thiols and sulfides (**B** and **C**), other volatiles (**D**), and O_2 and CO_2 (**E**) are shown. The identity of the volatile, the *m/z*, and the scaling factor are listed for each profile. Profiles within panels are multiplied by scaling factors. Symbols mark correlations between panels in peak maxima within an error of ±25°C due to signal smoothing: squares, 625°C; circles, 750°C; and triangles, 790 to 820°C. Axes and the placement of symbols relative to the temperature are the same in Fig. 2 and figs. S1 to S6. The *x* axis is scaled linearly relative to the run time, and the corresponding sample temperature is shown. The *y*-axis scale bar in counts per second (cps) is for all panels. Profiles in (A) are shifted along the *y* axis to show peaks clearly. In *m/z* values, "c" indicates corrections to profiles to remove contributions from other volatiles (*11*). *i*, isotopologue.

[1]Solar System Exploration Division, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. [2]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Geophysical Laboratory, Carnegie Institution of Washington, Washington, DC 20015, USA. [4]Laboratoire Atmosphères, Milieux, Observations Spatiales, Université Pierre et Marie Curie, Université Versailles Saint-Quentin and CNRS, Paris, France. [5]Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, Ciudad Universitaria, Cuidad de México 04510, México. [6]Jacobs Technology, Houston, TX 77058, USA. [7]Astromaterials Research and Exploration Science Division, NASA Johnson Space Center, Houston, TX 77058, USA. [8]Department of Geosciences, Stony Brook University, Stony Brook, NY 11794, USA. [9]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA. [10]Department of Earth Science and Engineering, Imperial College London, London SW7 2AZ, UK. [11]Department of Earth and Planetary Sciences, University of California, Davis, CA 95616, USA. [12]Laboratoire Génie des Procédés et Matériaux, Centrale Supélec, Gif-sur-Yvette, France. [13]Laboratoire Inter-Universitaire des Systèmes Atmosphériques, Université Paris-Est Créteil, Université Paris Diderot and CNRS, Créteil, France.
***Corresponding author. Email: jennifer.l.eigenbrode@nasa.gov** †Present address: Laboratoire Atmosphères, Milieux, Observations Spatiales, Institut Pierre Simon Laplace, CNRS, Guyancourt, France. ‡Present address: Department of Biology, Georgetown University, Washington, DC 20057, USA. §Present address: Geophysical Laboratory, Carnegie Institution of Washington, Washington, DC 20015, USA.

type) reactions that may produce cyclic structures (thiophenes and other aromatics). Sufficient He flow mitigates these secondary reactions. In spite of this complexity, larger molecular fragments can maintain structural information regarding the parent organic matter (*12*, *13*). In Mojave, CO, COS, CS_2, CH_4S, and C_2H_6S likely reflect cleavage directly from the precursor organic structures and reaction products. Similarly, CO_2, H_2S, H_2, and H_2O may be derived partly from organic matter in addition to mineral decomposition (*3*, *20*). However, under SAM oven conditions at high temperatures and low pressure under flowing He, entropic factors govern reactions (*12*, *21*, *22*), and the formation of thiophenic and other aromatic volatiles in the oven is not favored. This conclusion is supported by EGA blank analyses that indicate the absence of cyclic structures (fig. S2) despite available reactants from the background organic molecules (fig. S7) and by the absence of thiophenes in EGA data for the sulfide-bearing Cumberland sample of the Sheepbed mudstone, indicating that no cyclization occurred (table S1). Thus, the thiophenic and aromatic volatiles likely reflect compounds directly released from organic matter in the Murray mudstones. It is very plausible that aliphatics and some portion of the C_1 and C_2 sulfur compounds also reflect the in situ sample chemistry. Even if some portion of these compounds are not direct pyrolysis products from sediments, their carbon must largely be derived from organic matter indigenous to martian sediments.

The weaker and less diverse organic signals of the Sheepbed and other Murray mudstones indicate that the sediments had less organic input at the time of deposition or that the organic inputs were more substantially degraded over geological time. Organic materials in the lacustrine mudstones at Yellowknife Bay and in the lower Mount Sharp group strata have survived multiple aqueous diagenetic events (*6*–*8*), though the number and extent of these are difficult to constrain. Further, direct and indirect reactions induced by ionizing cosmic rays degrade organic matter (*23*), and SAM measurements of noble gas isotopes in the Sheepbed mudstones indicate that the sediments have been irradiated for ~80 million years, implying substantial degradation (*24*). It is possible that the Murray mudstones experienced less exposure.

Within the lower Murray mudstones at Pahrump Hills, sulfide minerals were likely altered to jarosite (1 to 3% in Mojave and Confidence Hills) by acidic diagenetic fluids (pH 2 to 6) ~2.1 billion years ago (*25*), though it is unclear whether these fluids were localized to sediment grains (*8*) or were more pervasive in the strata, leaching mafic minerals of metals and increasing in pH as they passed downward through the section (*7*). In either case, the exposure must have been limited in time and space, as pH-sensitive minerals such as apatite and olivine persist. Acidic fluids can effectively oxidize exposed organics; however, acidic diagenesis in the lower Murray may have had a small effect on organics, as it did on pH-sensitive minerals. If these fluids moved downward through the Pahrump Hills section as proposed by Rampe *et al.* (*7*), Confidence Hills and Mojave would have been exposed to only mildly acidic fluids (pH 6), resulting in milder organic degradation than that of overlying Murray mudstones. Alternatively, variations in organic matter abundance and composition in Murray mudstones may reflect geological inputs from transported detritus that was already in a refractory state, which would support the survival of organic detritus exposed to varied lake redox conditions (*8*). Ultimately, the fate of organic matter is determined by both degradation and preservation mechanisms. Both are likely important to the mudstones in Gale crater. By what preservation mechanisms did the martian organic matter survive?

Macromolecules on their own are self-preserving because surface organics shield interior organics from oxidation and stabilize the bulk organic mass (*26*). Ancient biomacromolecules deposited in lake sediments may have been transformed into geomacromolecules (kerogen) over time. Other possible sources of recalcitrant macromolecules include interplanetary dust particles (IDPs) (*27*) and abiotic organic materials from igneous rocks (*5*). In an effort to better constrain the composition and possible origin of organic matter in the mudstones, we conducted SAM test bed EGA of the Murchison meteorite, a proxy for IDP composition, and laboratory EGA of the Tissint martian meteorite, which hosts igneous rock–related organics. Both show the evolution of C_1-C_2 sulfur volatiles and aliphatic, aromatic, and thiophenic pyrolysis products during EGA above 500°C (fig. S9 and S10), but the profiles related to these molecular groups are distinctive for each sample type. These results are consistent with the known

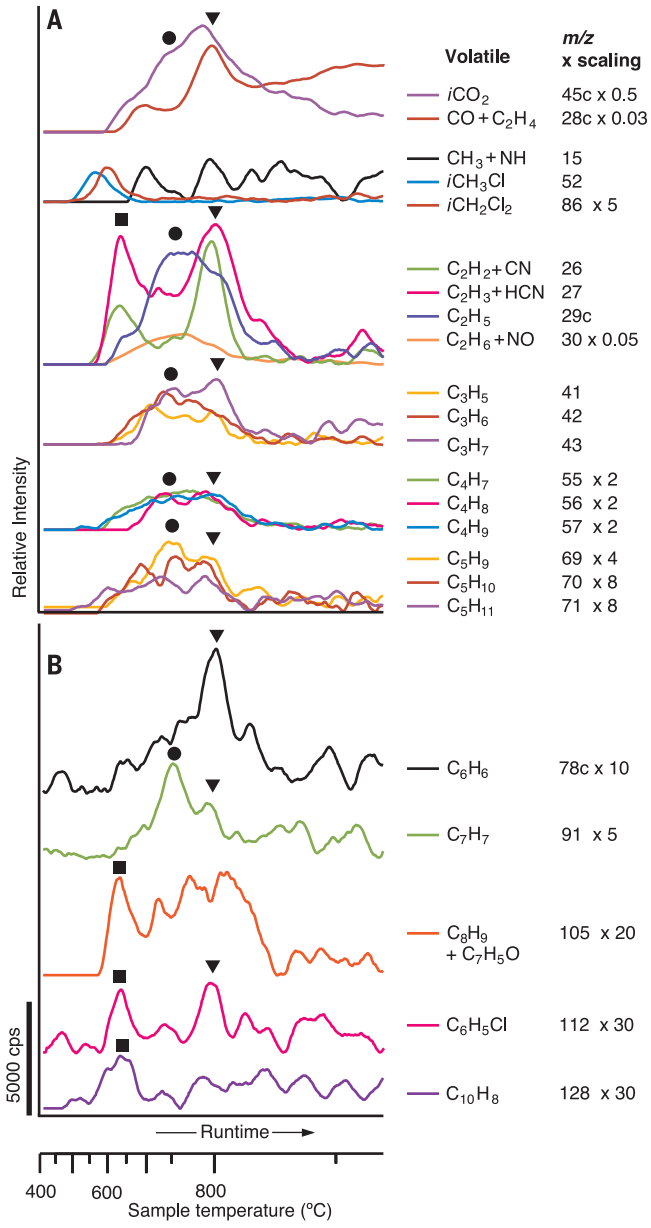


Fig. 2. Mojave EGA profiles for aliphatic and aromatic compounds. CO and CO_2 profiles are included with aliphatic profiles in (**A**). Profiles in (A) are grouped by carbon number and shifted along the *y* axis for clarity. Profiles for aromatic compounds in (**B**) are similarly shifted. Plotting details are as described in the legend to Fig. 1 (*11*).

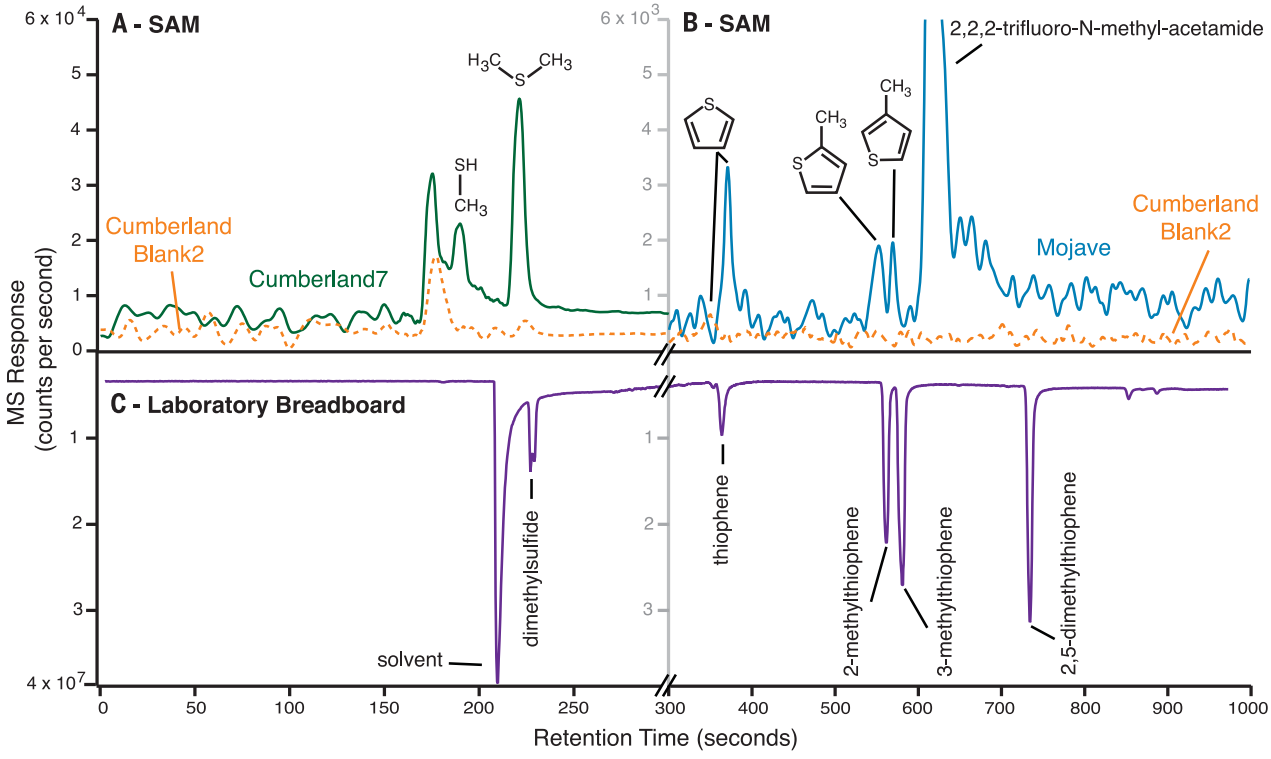


Fig. 3. Example of SAM GC-MS identification of S-containing pyrolysis products compared with results from a high-fidelity SAM-like GC run in the same manner but in the laboratory. (**A**) SAM GC-MS chromatograms summing *m/z* 47 and 62 from 0 to 300 s for Cumberland7 of the Sheepbed mudstone and a blank showing methanethiol and dimethylsulfide. (**B**) SAM GC-MS chromatograms summing *m/z* 84, 97, and 98 from 300 to 1000 s for Mojave and a blank showing the presence of thiophene, 2-methylthiophene, and 3-methylthiophene. 2,5-Dimethylthiophene was not identifiable. (**C**) Chromatograms from GC run in the laboratory. Chromatograms were smoothed, and the off-nominal GC run from Mojave resulted in the 20-s offset for the thiophene retention time as observed in (B) (*11*). The 2,2,2-trifluoro-*N*-methyl-acetamide peak in (B) is a derivatization reaction product that is part of the instrument background (*11*). GC-MS identifications are based on both SAM and laboratory retention times (table S3) compared to standards and mass spectra in a reference database (*33*). Axis breaks denote a change in the *x*-axis scale.

presence of refractory organic matter in each meteorite but indicate differences in the organic chemistry and mineral associations of the meteorites. The supplemental EGAs do not provide constraints on organic matter origin. However, they do support the interpretation that the mudstones host refractory organic matter that is most likely macromolecular.

Minerals can further aid preservation by several mechanisms: occlusion by minerals, organic-mineral interactions (e.g., organic binding to phyllosilicate and Al-Fe oxyhydroxide surfaces), and the establishment of chemically reducing microenvironments that host organics (*28*). If organic matter entering sediments was labile, such as autochthonous biomolecules (e.g., carbohydrates, amino acids, and fatty acids), then reactions with mineral surfaces or sedimentary chemicals would have been favored. Phyllosilicate, iron oxyhydroxide, and amorphous materials are prevalent in all of the studied mudstones (*7*, *8*). Further, iron sulfides detected in the Sheepbed mudstone (*29*) and suspected in the original Murray mudstone detritus (*7*, *8*) may have aided organic matter preservation by providing an additional oxygen sink during diagenesis.

Reduced permeability limits the exposure of organics to migrating fluids and gas. Permeability is generally reduced by abundant fine-grained sediments and precipitates (e.g., sulfate cement, vein fill, and nodules). All mudstones studied exhibit indications of groundwater alteration (*6*–*8*, *30*); however, the timing of cementation and postdepositional aqueous alteration is not well constrained, so the extent of organic exposure to these fluids is unknown.

Macromolecules, mineral interactions, and permeability factors were all likely contributors to organic matter preservation in the Murray mudstone, but sulfurization of organic molecules was probably the principal preservation mechanism responsible for the distinct record in Mojave and Confidence Hills given the presence of 3 to 10 times as much thiophenic and total organic sulfur in these samples as in the other mudstones (tables S1 and S4). Natural vulcanization results in an enhanced refractory state for organic materials. The addition of sulfur structurally links the organic components into a macromolecular form and provides an additional oxidative sink for degradation reactions. On Earth, sulfurization enhances initial preservation while also imparting long-term recalcitrance to structural transformations and oxidation, such as during acidic diagenesis. Sulfurization probably occurred during early diagenesis in the presence of reduced sulfur (HS^- or H_2S) gas (*31*) more than 3 billion years ago. The large sulfur isotopic fractionation observed in the SO_2 evolved via EGA from the mudstones indicates that sulfide was transported via hydrothermal groundwater to the Gale lake basin (*32*). Alternatively, organic sulfur was native to the detritus deposited in the lake.

SAM's molecular observations do not clearly reveal the source of the organic matter in the Murray formation. Biological, geological, and meteoritic sources are all possible. Certainly, if ancient life was the organic source, then despite sulfur incorporation, the material has been altered sufficiently, such as by diagenesis or ionizing radiation (*23*), to obscure original molecular features more consistent with life (e.g., a greater diversity of molecules or patterns of limited structural variation within compound classes, such as hydrocarbon chains), or an insufficient amount of organic matter was deposited to allow detection by pyrolysis–GC-MS.

Past habitability interpreted for the Sheepbed lacustrine mudstones focused on chemolithoautotrophy (*8*, *30*), but observations of geologically refractory organic matter in Murray lacustrine mudstones opens the door for past and present habitability for heterotrophy as well. Organic matter can directly or indirectly fuel both energy and carbon metabolisms and in doing so can support carbon cycling at the microbial community level.

Table 1. Organic carbon abundance estimates for EGA signals above 500°C. See table S1 for abundances of individual molecules and abundances in other mudstones. Errors are propagated from integration uncertainty (30%, ±1σ SD, $n \geq 3$ analyses), smoothing error (5%, ±1σ SD, $n \geq 3$ analyses), and ionization cross-section uncertainties (reported in the literature). The total for thiophenic compounds is the sum of thiophene and methyl thiophene abundances. The total for other aromatic compounds is the sum of abundances of benzene, toluene, and benzoic acid (a proxy for benzoate ion or alkylbenzene contributions). The total for aliphatic compounds is the sum of C_1 to C_3 alkanes and C_2 to C_5 alkenes determined from modeling. The total for C_1 and C_2 sulfur compounds is the sum of methanethiol, dimethylsulfide, carbonyl sulfide, and carbon disulfide abundances (*11*).

Sample	Organic carbon abundance (nmol of C) in compound class(es)					Total organic carbon (nmol of C)
	Thiophenic	Aromatic	Aliphatic	Thiophenic, aromatic, and aliphatic	C_1 and C_2 sulfur	
Mojave	2.20 ± 0.93	6.99 ± 1.99	38.4 ± 5.63	47.61 ± 6.04	43.8 ± 11.6	91.4 ± 13.1
Confidence Hills	2.02 ± 0.84	8.03 ± 2.11	20.8 ± 2.79	30.8 ± 3.60	29.3 ± 7.61	60.2 ± 8.42
Confidence Hills blank	0.49 ± 0.18	3.04 ± 0.86	5.35 ± 0.83	8.88 ± 1.21	1.49 ± 0.42	10.4 ± 1.28

Our results suggest that it is likely that organic matter from various sources may be widely distributed in the martian rock record. Even if life was not a key contributor, meteoritic and igneous or hydrothermal sources have a strong potential to be broadly emplaced. Our detection of organic matter at the martian surface, where ionizing and oxidizing conditions are extreme, suggests that better-preserved molecular records may be present below the surface, where the effects of radiation are small, or in materials exposed in the last several thousand years.

REFERENCES AND NOTES

1. C. Freissinet *et al.*, *J. Geophys. Res. Planets* **120**, 495–514 (2015).
2. D. W. Ming *et al.*, *Science* **343**, 1245267 (2014).
3. B. Sutter *et al.*, *J. Geophys. Res. Planets* **122**, 2574–2609 (2017).
4. R. Navarro-González, E. Vargas, J. de la Rosa, A. C. Raga, C. P. McKay, *J. Geophys. Res. Solid Earth* **115** (E12), E12010 (2010).
5. A. Steele, F. M. McCubbin, M. D. Fries, *Meteorit. Planet. Sci.* **51**, 2203–2225 (2016).
6. J. P. Grotzinger *et al.*, *Science* **350**, aac7575 (2015).
7. E. Rampe *et al.*, *Earth Planet. Sci. Lett.* **471**, 172–185 (2017).
8. J. A. Hurowitz *et al.*, *Science* **356**, eaah6849 (2017).
9. R. E. Summons *et al.*, *Astrobiology* **11**, 157–181 (2011).
10. P. R. Mahaffy *et al.*, *Space Sci. Rev.* **170**, 401–478 (2012).
11. Materials and methods are available as supplementary materials.
12. S. C. Moldoveanu, *Pyrolysis of Organic Molecules with Applications to Health and Environmental Issues*. S. C. Moldoveanu, Ed., Techniques and Instrumentation in Analytical Chemistry (Elsevier, New York, 2010), vol. 28, pp. 724.
13. S. C. Moldoveanu, *Analytical Pyrolysis of Natural Organic Polymers*. S. C. Moldoveanu, Ed., Techniques and instrumentation in analytical chemistry (Elsevier, New York, 1998), vol. 20, pp. 496.
14. F. Okumura, K. Mimura, *Geochim. Cosmochim. Acta* **75**, 7063–7080 (2011).
15. L. Remusat, L. S. Derenne, F. Robert, H. Knicker, *Geochim. Cosmochim. Acta* **69**, 3919–3932 (2005).
16. F. W. McLafferty, F. Tureek, *Interpretation of Mass Spectra* (University Science Books, 1993).
17. J. S. Sinninghe Damsté, T. I. Eglinton, J. W. De Leeuw, P. A. Schenck, *Geochim. Cosmochim. Acta* **53**, 873–889 (1989).
18. B. P. Baruah, P. Khare, *Energy Fuels* **21**, 3346–3352 (2007).
19. L. Xu, J. Yang, Y. Li, Z. Liu, *Fuel Process. Technol.* **85**, 1013–1024 (2004).
20. A. C. McAdam *et al.*, *J. Geophys. Res. Planets* **119**, 373–393 (2014).
21. R. F. C. Brown, *Pyrolytic Methods in Organic Chemistry: Application of Flow and Flash Vacuum Pyrolytic Techniques* (Academic Press, 1980).
22. M. Bajus, *Sulfur Rep.* **9**, 25–66 (1989).
23. D. M. Hassler *et al.*, *Science* **343**, 1244797 (2014).
24. K. A. Farley *et al.*, *Science* **343**, 1247166 (2014).
25. P. E. Martin *et al.*, *J. Geophys. Res. Planets* **122**, 2803–2818 (2017).
26. L. Mayer, *Mar. Chem.* **92**, 135–140 (2004).
27. G. J. Flynn, L. R. Nittler, C. Engrand, *Elements* **12**, 177–183 (2016).
28. R. G. Keil, L. M. Mayer, *Org. Geochem.* **12**, 337 (2014).
29. D. T. Vaniman *et al.*, *Science* **343**, 1243480 (2014).
30. J. P. Grotzinger *et al.*, *Science* **343**, 1242777 (2014).
31. Y. Hebting *et al.*, *Science* **312**, 1627–1631 (2006).
32. H. B. Franz *et al.*, *Nat. Geosci.* **10**, 658–662 (2017).
33. P. J. Linstrom, W. G. Mallard, Eds., NIST Chemistry WebBook, NIST Standard Reference Database Number 69 (National Institute of Standards and Technology); https://webbook.nist.gov/chemistry/.

ACKNOWLEDGMENTS

We thank reviewers for their constructive comments, the MSL and SAM teams for successful operations of SAM on Mars and thoughtful science discussions, and K. Irikura for calculating ionization cross sections. **Funding:** This work was funded by NASA's Mars Exploration Program. NASA's MSL Participating Scientist program supported J.L.E., R.E.S., and D.P.G. for this effort. S.G. acknowledges funding from the UK Space Agency. R.N.-G. was funded by the Universidad Nacional Autónoma de México and the Consejo Nacional de Ciencia y Tecnología de México. **Author contributions:** J.L.E. developed data processing methods, calculated and interpreted EGA data, and wrote most of the manuscript and supplementary materials. R.E.S, A.S, B.S., and P.R.M. wrote sections of text on pyrolysis, meteorites, geological context, and methods, respectively. C.F., M.M., D.P.G., A.S., and C.S. calculated and interpreted GC-MS data. M.M. performed SAM GC breadboard tests. H.B.F., P.D.A., and B.S. contributed to analysis of EGA data. R.N-G. performed laboratory analyses necessary to understand the SAM background. All authors participated in data acquisition, discussion of results, and/or editing of the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Reduced data records (RDRs) from SAM flight experiments are archived in the Planetary Data System (https://pds.nasa.gov) and are identifiable by the sol or test identification (TID) numbers listed in table S6. All processed data are available in the text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1096/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S11
Tables S1 to S10
References (*34*–*58*)

4 January 2018; accepted 29 March 2018
10.1126/science.aas9185

ORGANIC MATERIALS

Evidence for a quantum dipole liquid state in an organic quasi-two-dimensional material

Nora Hassan[1], Streit Cunningham[1], Martin Mourigal[2], Elena I. Zhilyaeva[3], Svetlana A. Torunova[3], Rimma N. Lyubovskaya[3], John A. Schlueter[4,5], Natalia Drichko[1]*

Mott insulators are commonly pictured with electrons localized on lattice sites, with their low-energy degrees of freedom involving spins only. Here, we observe emergent charge degrees of freedom in a molecule-based Mott insulator κ-(BEDT-TTF)$_2$Hg(SCN)$_2$Br, resulting in a quantum dipole liquid state. Electrons localized on molecular dimer lattice sites form electric dipoles that do not order at low temperatures and fluctuate with frequency detected experimentally in our Raman spectroscopy experiments. The heat capacity and Raman scattering response are consistent with a scenario in which the composite spin and electric dipole degrees of freedom remain fluctuating down to the lowest measured temperatures.

Fluctuating dipolar degrees of freedom have been predicted to appear in molecular-based Mott insulators (*1*, *2*) and optical lattices of dipolar molecules (*3*) and to lead to a spin-liquid state in the presence of charge-spin coupling. In quantum paraelectrics, fluctuations of electric dipoles are observed in the vicinity of a ferroelectric transition (*4*). A quantum dipole liquid in an antiferroelectric on a triangular lattice was recently reported for $BaFe_{12}O_{19}$ (*5*), but as a band insulator this compound is nonmagnetic. A quantum dipole liquid in a Mott insulator is a paradigm for quantum states of matter that brings together quantum paraelectrics and spin liquids. Here, we discuss experimental evidence for the quantum dipole liquid state in a layered organic Mott insulator κ-(BEDT-TTF)$_2$Hg(SCN)$_2$Br [here, BEDT-TTF stands for a molecule bis(ethylenedithio) tetrathiafulvalene]. In the presence of charge-spin coupling, it may result in a spin-liquid state.

Electronic and magnetic phenomena observed in this class of materials are determined by the properties of the molecular-based cation layers (Fig. 1A). In a Mott insulator of this kind, electrons are localized on the lattice sites of dimers $(BEDT-TTF)_2^{1+}$ with spin $S = 1/2$ per site. These dimer sites form layers (Fig. 1B) that can be represented by a two-dimensional anisotropic triangular lattice (Fig. 1C). In most such compounds, the dimers have an inversion center and thus zero electric dipole moment. Frustration of the lattice, competing electronic correlations, and magnetic interactions can lead to charges being distributed nonsymmetrically between the two molecules in a dimer (*1*, *6*), producing a dipole. This can lead to a broken symmetry ground state, so called "paired electron crystal" (*6*) or "quantum dipole solid" (*1*), where the dipoles order forming a ferroelectric state (Fig. 1D) (*6*). In contrast to a displacive ferroelectric, a change in the underlying lattice is not necessary in this case. It was suggested (*1*) that the dipoles can fluctuate in a quantum dipole liquid (Fig. 1E), providing an explanation of the spin-liquid state observed in a triangular lattice κ-(BEDT-TTF)$_2$Cu$_2$(CN)$_3$. However, the evidence for fluctuating quantum dipoles in this material remains elusive (*7*, *8*).

Here, we elucidate the properties of the quantum dipole liquid state in the triangular lattice Mott insulator (*9*, *10*) κ-(BEDT-TTF)$_2$Hg(SCN)$_2$Br (κ-Hg-Br) [T_{MI} = 80 K (*11*, *12*)] by comparing them to those of an isostructural compound κ-(BEDT-TTF)$_2$Hg(SCN)$_2$Cl (κ-Hg-Cl), which shows signatures of a quantum dipole solid below 30 K (*10*, *11*). We first use Raman molecular vibrational spectroscopy to follow the evolution of the distribution of charge on the lattice of these systems through the metal-insulator transition. The on-molecule charge is probed by measuring the frequency of the central C=C molecular bond vibration (v_2) (Fig. 2E), which changes by about –140 cm^{-1} when the charge state changes from $(BEDT-TTF)^0$ to $(BEDT-TTF)^{1+}$ (*13*, *14*). This is a result of a lengthening of the central C=C bond of the molecule when more charge occupies the highest occupied molecular orbital (HOMO). To exclude effects other than a change of charge distribution on the lattice, we compare the temperature dependence of the parameters of the charge-sensitive v_2 band to that of the v_3 mode at 1470 cm^{-1}. The latter involves the stretch of the off-center C=C bonds (Fig. 2E) and is not sensitive to charge on the molecule but would be affected by structural disorder or symmetry breaking. The ν_3 band stays a narrow single line throughout the measured temperature range for both materials. For κ-Hg-Cl, a single ν_2 band at about 1490 cm^{-1} is observed in the high-temperature metallic state, whereas in the insulating state below T_{CO} = 30 K, ν_2 is split into two bands at 1475 and 1507 cm^{-1} (Fig. 2A). The difference in frequencies of the two modes is much higher than expected for a structural phase transition but corresponds to charges redistributed within the $(BEDT-TTF)_2^{1+}$ dimer as +0.4*e* on one molecule and +0.6*e* on the other (*10*, *14*); this breaks the inversion symmetry in a $(BEDT-TTF)_2^{1+}$ dimer, creating an electric dipole. This result confirms that the ground state of κ-Hg-Cl is an order of electric dipoles localized on $(BEDT-TTF)_2^{1+}$ dimer sites (*10*), the so-called dipole solid (*1*).

In contrast, for κ-Hg-Br, a single v_2 mode observed in the whole studied temperature range (Fig. 2C) suggests a symmetric dimer with both molecules carrying half a hole $(BEDT-TTF)^{0.5+}$ on average. However, the width of the v_2 band shows abnormal behavior on cooling. Line widths of phonons are determined by decay mechanism (*15*), disorder, and dynamics of the lattice and charge systems. For example, the width of the v_3 vibrational band of BEDT-TTF molecule is determined by decay processes into lower-frequency modes and decreases down to about 5 cm^{-1} at 10 K (see Fig. 2D). In contrast, the width of v_2 goes through a minimum of 16 cm^{-1} at around 80 K and increases again up to 20 cm^{-1} at 10 K (Fig. 2, C and D). Abnormal temperature dependence of the width observed only for v_2 rules out the possibility that structural changes or structural disorder are at its origin. Another possible reason for an increased line width is charge fluctuations (*16*, *17*).

We estimate the effects of charge fluctuations on the shape of the v_2 vibration using a "two-site jump" model (*16*, *17*) (see eq. S1). In this model, we consider +0.6*e* and +0.4*e* charged molecules, where *e* is electron charge, observed in the ordered state of κ-Hg-Cl as two static species. They are characterized by frequencies of v_2 vibrations $v_2[BEDT-TTF^{0.4+}]$ = 1507 cm^{-1} and $v_2[BEDT-TTF^{0.6+}]$ = 1475 cm^{-1}. Their natural width Γ depends on temperature through the lifetime of the measured excited state and is expected to be the same as that of v_3. The system can jump between these two states with a frequency of $\omega_{EX} = 1/\tau$, where τ is the lifetime of each state defined by the exchange. As the exchange rate ω_{EX} between these two states increases, the shape of the resulting spectra changes (Fig. 2B). The two original bands get wider and the difference in positions between them decreases, and at a high enough rate ω_{EX}, they merge into a single band. The calculated spectrum at ω_{EX} = 0 reproduces the doublet shape of v_2 for κ-Hg-Cl in the dipole solid state (Fig. 2, A and B). The spectra calculated for ω_{EX} = 40 and 30 cm^{-1} reproduce the shape of the v_2 band in κ-Hg-Br spectra, where the width of v_2 increases from 16 to 20 cm^{-1} on cooling below 80 K, and the band gains slight asymmetry (Fig. 2, B and C). These results suggest that in κ-Hg-Br charges fluctuate between two molecules in a dimer with frequency ω_{EX}

[1]The Institute for Quantum Matter and the Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD 21218, USA. [2]School of Physics, Georgia Institute of Technology, Atlanta, GA 30332, USA. [3]Institute of Problems of Chemical Physics, Chernogolovka, Russia. [4]Division of Materials Research, National Science Foundation, Alexandria, VA 22314, USA. [5]Materials Science Division, Argonne National Laboratory, Argonne, IL 60439, USA.
*Corresponding author. Email: drichko@jhu.edu

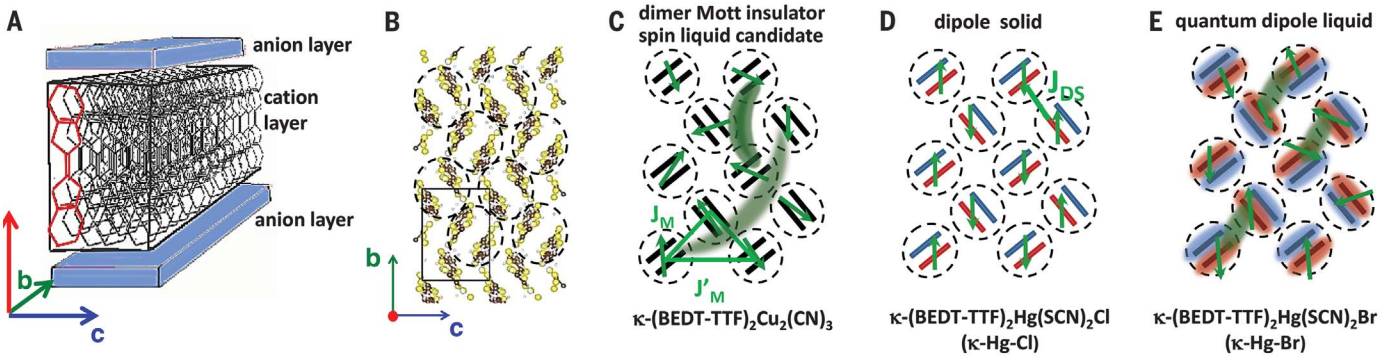


Fig. 1. Crystal structure and phases of BEDT-TTF–based crystals. (**A**) Schematic structure of a BEDT-TTF–based crystal; the molecule is highlighted in red. (**B**) Structure of a BEDT-TTF layer in the (*bc*) plane of the κ-Hg-Cl crystal as determined from x-ray diffraction (*10*). BEDT-TTF molecules are bound in dimers (circles). The dimer sites form an anisotropic triangular lattice. (**C**) Schematic structure of a BEDT-TTF layer in a dimer Mott insulator on an anisotropic triangular lattice formed by $(BEDT\text{-}TTF)_2^{1+}$ sites with $S = 1/2$ (spins depicted by green arrows) and magnetic exchange between sites J_M and J'_M. The model is relevant to the spin-liquid candidate $\kappa\text{-}(BEDT\text{-}TTF)_2Cu_2(CN)_3$, with $J_M/J'_M = 0.64$ (*37*). (**D**) Schematic structure of a BEDT-TTF layer in the case of a dipole solid [paired electron crystal (*6*)]. Within $(BEDT\text{-}TTF)_2^{1+}$ dimer sites, charge-rich and charge-poor molecules are denoted by red and blue, respectively. The dimer sites thus possess a dipole moment. J_{DC} is the magnetic interaction between spins (marked by green arrows) on neighboring charge-rich molecules. Spins of the nearest-neighbor charge-rich sites will form spin singlets (*6*). This situation is relevant to κ-Hg-Cl. (**E**) Schematic structure of a BEDT-TTF layer in a quantum dipole liquid. The charge is fluctuating between the molecules in $(BEDT\text{-}TTF)_2^{1+}$ dimers, as denoted by blurry red and blue ovals, leading to electric dipoles fluctuations. Associated spins also show fluctuations. This situation is relevant to κ-Hg-Br.

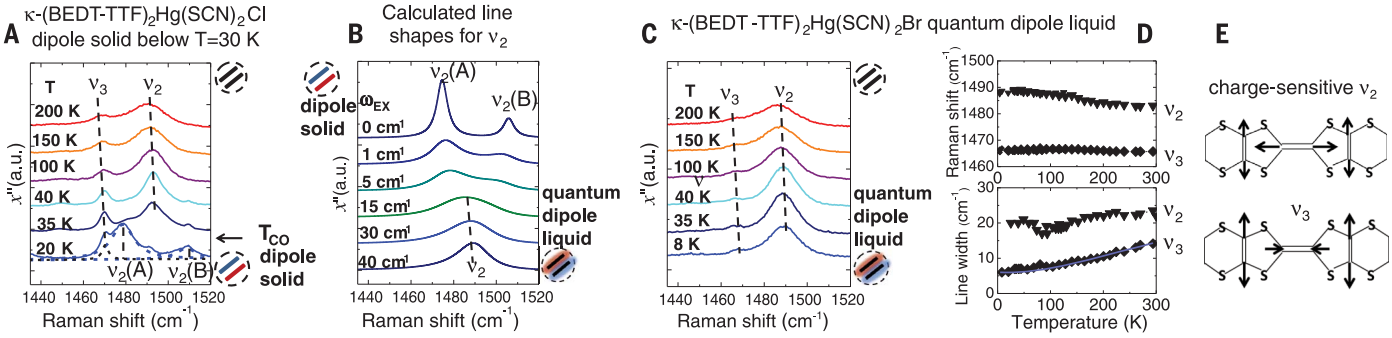


Fig. 2. Raman spectra in the region of C=C vibrations of BEDT-TTF. (**A**) Temperature dependence of the κ-Hg-Cl spectra in the region of ν_2 and ν_3 modes. Note the splitting of ν_2 mode in the dipole solid (charge ordered) state at 20 K with frequencies corresponding to $BEDT\text{-}TTF^{0.4+}$ and $BEDT\text{-}TTF^{0.6+}$. (**B**) Shape of ν_2 mode calculated from the two-sites jump model (see eq. S1). The upper spectrum is of a static system ($\omega_{EX} = 0$), with bands corresponding to $BEDT\text{-}TTF^{0.4+}$ and $BEDT\text{-}TTF^{0.6+}$ as in the dipole solid state of κ-Hg-Cl. On the increase of exchange frequency ω_{EX}, the bands widen and move close to each other. The lower two spectra at $\omega_{EX} = 30$ and 40 cm^{-1} reproduce the ν_2 shape of κ-Hg-Br at 8 and 35 K correspondingly, taking into account the natural width Γ for the relevant temperature. (**C**) Temperature dependence of the κ-Hg-Br spectra in the region of ν_2 and ν_3 modes. The ν_2 band does not split but shows some widening at the lowest temperature. (**D**) Temperature dependence of center frequency (top) and line width (bottom) for ν_2 (triangles) and ν_3 (diamonds) modes for κ-Hg-Br. The line width of ν_2 for κ-Hg-Br goes through a minimum at around 80 K, whereas that of ν_3 decreases continuously. (**E**) Illustration of the movements of atoms associated with ν_2 and ν_3-mode vibrations in a BEDT-TTF molecule.

that slightly decreases on cooling. In other words, electric dipoles in κ-Hg-Br fluctuate with this frequency, forming a quantum dipole liquid state.

Apart from the difference in the υ_2 band behavior, the phonon Raman spectra of κ-Hg-Br and κ-Hg-Cl are very similar (Fig. 3, A, B, and S1 A, B) because the two compounds have very similar crystal structures. Whereas the spectral region below about 200 cm^{-1} for κ-Hg-Cl shows only phonon bands (Fig. 3 B), in the A_{1g} scattering channel for κ-Hg-Br, the phonons are superimposed on a much wider feature with a maximum around 40 cm^{-1} (Fig. 3 A). In the B_{1g} scattering channel, we also observe this feature, with phonon bands showing broad asymmetric shapes, apparently as a result of electron-phonon coupling (Fig. 3A, and fig. S2B). This asymmetric feature with a maximum around 40 cm^{-1} (see spectra with phonons subtracted in Fig. 3C) with the width at half maximum of about 40 cm^{-1} gains intensity below 100 K, when κ-Hg-Br enters the insulating state, and shows weak softening at the lowest temperature. Apparently, this wide feature observed only in κ-Hg-Br spectra originates from a different scattering channel than phonons. Other potential scattering channels are electronic or magnetic excitations on a triangular lattice of $(BEDT\text{-}TTF)_2^{1+}$ dimers. On a triangular lattice, polarizations of electronic or magnetic excitations cannot be completely disentangled to elucidate the origin of the excitations, in contrast to a square lattice (*18*).

Magnetic excitations are expected in the Raman response of a Mott insulator with ordered spins or even spins developing short-range correlations (*18*–*21*). In Mott insulators based on $(BEDT\text{-}TTF)_2^{1+}$ dimers, magnetic excitations are observed both in Raman spectra of an antiferromagnetically ordered state on a square lattice (*19*) and in a spin-liquid candidate on triangular lattice $\kappa\text{-}(BEDT\text{-}TTF)_2Cu_2(CN)_3$ (Fig. 3E). The spectra of the latter show a continuum of magnetic excitations below 600 cm^{-1}. The position of the continuum is defined by the value of J and geometry of the lattice (*18*). For $\kappa\text{-}(BEDT\text{-}TTF)_2Cu_2(CN)_3$, it is in agreement with Hubbard-model–based calculations for the magnetic response of $S = 1/2$ on an anisotropic triangular lattice with $J_M = 250$ K (*20*).

It is clear at this point that magnetic interactions in a dipole solid, and possibly quantum

dipole liquid, would be renormalized in comparison with a simple $(BEDT\text{-}TTF)_2^{1+}$ dimer Mott insulator with charge symmetrically distributed on a dimer. Hotta (*1*) proposes a renormalization and a decrease of J in a quantum dipole liquid compared with a simple dimer Mott insulator, however, without estimating J values. A simple argument suggests that, in a dipole solid, magnetic interactions occur between charge-rich molecules of the neighboring dimers, whereas in a simple Mott insulator the interactions occur between dimer lattice sites (illustrations in Fig. 3E). An estimate provided by a tight-binding approximation as $J = \frac{4t^2}{U}$, where t is a transfer integral and U is on-molecule Coulomb repulsion, yields the value of about J_{DS} = 80 K for a dipole solid. This is considerably smaller than J_M = 250 K (*22*) for a simple dimer Mott insulator, where the on-dimer U defines magnetic interactions. Here, the Coulomb repulsion parameters, as well as transfer integrals, are estimated from the optical conductivity spectra (*23*), and the difference is produced mainly by a variation between the values of U in these two models. A lower J would result in a lower ordering temperature and a spectrum of magnetic excitation shifting to lower frequencies. However, the maximum of the observed background is about ~40 cm^{-1} and is found below the expected J_{DS} value, which is too low in frequency to be interpreted as purely magnetic excitations.

Another possibility is assigning this mode to a collective excitation associated with dipole fluctuations. Dipole fluctuations with a frequency of about ω_{EX} = 40 cm^{-1} are detected through the line-shape analysis of charge-sensitive vibrations. If these fluctuations are a collective phenomenon, we would expect a collective mode at about 40 cm^{-1}. The low-frequency mode observed in κ-Hg-Br thus is a good candidate for a collective response of dipole fluctuations. Optically detected collective modes associated with charge fluctuations are found in the metallic state close to a charge ordering metal-insulator transition in organic conductors (*24*, *25*) and in under-doped high-temperature cuprate superconductors (*26*). In an insulating state, the closest analogy would be a soft mode close to the transition into the ferroelectric state in displacive ferroelectrics such as $SrTiO_3$. A comparatively small width of the band, as well as its increase in intensity and its low-frequency shift below T_{MI} = 80 K distinguishes it from a boson peak observed in glasses (*27*) and supports an interpretation in terms of a fluctuating system of dipoles versus charge glass. An absence of glassy behavior is also supported by the low-frequency dielectric response of κ-Hg-Br (*12*).

Fluctuations of electric dipoles coupled to S = 1/2 spins on a triangular lattice of $(BEDT\text{-}TTF)_2^{1+}$ dimers have been suggested as a mechanism for spin-liquid behavior (*1*, *2*). Reference (*2*) discusses the coupling between the dipole and magnetic degrees of freedom within the Kugel-Khomskii model, showing an analogy between the fluctuating dipole liquid and orbital liquid (*28*, *29*). This model suggests that for certain values of frustration J'/J and spin-charge coupling K, spin order is destabilized and would produce mixed spin-charge excitations. To understand whether the collective mode observed in κ-Hg-Br Raman spectra originates purely in dipole fluctuations or in mixed charge-spin excitations, theoretical calculations of the excitation spectrum for such a system would be of great importance.

Our heat capacity data are consistent with the presence of itinerant excitations in κ-Hg-Br but not in κ-Hg-Cl. The heat capacity C_p of κ-Hg-Br and κ-Hg-Cl was measured in the temperature range between 40 K and 100 mK. The temperature dependencies of heat capacity for these

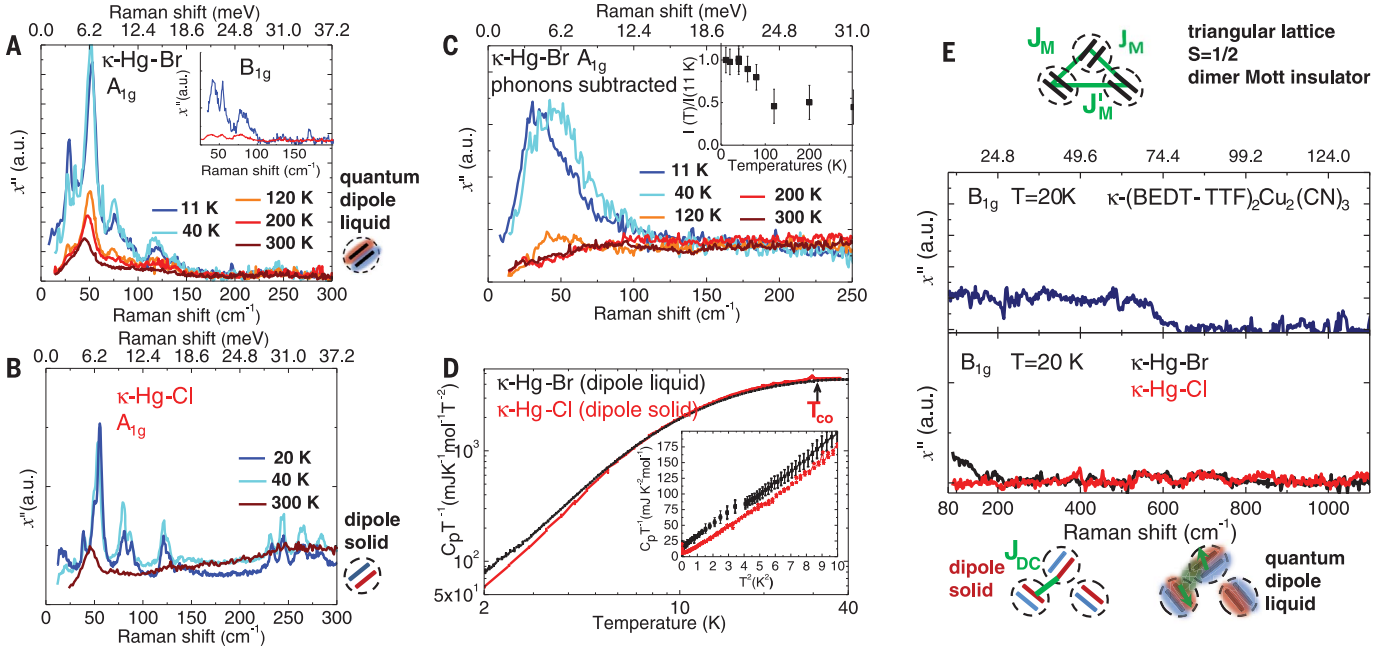


Fig. 3. Temperature dependence of Raman spectra. Shown are the spectra for (**A**) κ-Hg-Br and (**B**) κ-Hg-Cl in A_{1g} symmetry in the frequency range between 0 and 300 cm^{-1}. Phonons are found at similar frequencies for both compounds. In the spectra of κ-Hg-Br, a background develops at temperatures below 100 K. Spectra at 300 and 11 K for κ-Hg-Br for B_{1g} symmetry are shown in the inset to (A). In the B_{1g} scattering channel, the low-frequency background, interpreted as a collective mode, shows strong coupling to the phonons. (**C**) Temperature dependence of the collective mode in the A_{1g} scattering channel for κ-Hg-Br, determined by subtracting phonons from the full Raman spectrum (see the supplementary materials for details of the procedure. (Inset) Temperature dependence of the normalized intensity of the collective mode. (**D**) Temperature dependence of the heat capacity C_p for κ-Hg-Cl (red line) and κ-Hg-Br (black line) below 40 K. The two curves deviate from each other below ~6 K. (Inset) Low-temperature data with linear behavior of heat capacity for κ-Hg-Br. (**E**) Raman spectra in B_{1g} polarization at 20 K in the range between 800 and 1100 cm^{-1}, with phonons and luminescence background subtracted for κ-$(BEDT\text{-}TTF)_2Cu_2(CN)_3$ (top) and κ-Hg-Cl and κ-Hg-Br (bottom). See details on the subtraction procedure in the supplementary materials. Schematic pictures of the relevant models with different charge distribution are shown. The spectra of the dimer Mott insulator on the triangular lattice κ-$(BEDT\text{-}TTF)_2Cu_2(CN)_3$ (top) demonstrate magnetic excitations below ~600 cm^{-1}. This feature is absent in the spectra of both κ-Hg-Br (black) and κ-Hg-Cl (red). The increase of intensity in the spectra of κ-Hg-Br below 200 cm^{-1} is caused by the collective mode fully shown in (A).

compounds overlap within the error of the measurements in the temperature range above 6 K (Fig. 3D), excluding the feature at 30 K in the κ-Hg-Cl data indicating a charge order transition. The low-temperature heat capacity $C_p = \beta T^3 + \gamma T$ of both compounds shows basically the same bosonic contribution $\beta = 19.0 \pm 2.5$ mJ K^{-4} mol^{-1}. This is expected, as it is determined predominantly by phonons and vibrations of BEDT-TTF molecules, which are very similar for the studied compounds. The difference between the two materials appears below about 6 K, where for κ-Hg-Br, C_p shows a linear term $\gamma = 13.8 \pm 3.1$ mJK^{-2} mol^{-1} (inset in Fig. 3D). Spinon excitations can produce a linear term in heat capacity (*30*, *31*), suggesting a spin-liquid behavior of κ-Hg-Br. For κ-Hg-Cl, $\gamma = 0$ within the precision of our measurements.

An ordering of electric dipoles observed in κ-Hg-Cl does not necessarily imply magnetic order (*1*). However, a theoretical proposal for a dipole order in a "paired electron crystal" suggests magnetic interactions as a driving force for the charge order on a frustrated dimer lattice and a spin-singlet ground state (*6*). A single phase transition observed at 30 K can be evidence of simultaneous electric dipole ordering and singlet formation in κ-Hg-Cl. On the other hand, the temperature of spin ordering can be lower than that of the charge order, as is observed in one-dimensional materials and suggested by calculations (*32*). Because heat capacity is found to not be always sensitive to magnetic phase transitions in these two-dimensional materials (*33*), further studies, such as those with nuclear magnetic resonance (NMR), are necessary to identify the magnetic ground state of κ-Hg-Cl. Paired electron crystal proposed as a ground state of κ-Hg-Cl (*6*, *10*) can be regarded as a variation of a valence bond solid (*6*, *29*). In these terms, the quantum dipole liquid in κ-Hg-Br can be a realization of a resonant valence bond state (*34*).

The quantum dipole liquid was suggested as one possible explanation of the origin of the spin-liquid state in κ-(BEDT-TTF)$_2$Cu$_2$(CN)$_3$ (*1*); however, our work shows that this material does not demonstrate the signatures of this state. Its spectrum of magnetic excitations is well understood within a model of $S = 1/2$ on a triangular lattice with $J = 250$ K (Fig. 3) ((*20*). There is a recent suggestion (*35*) that κ-(BEDT-TTF)$_2$Cu$_2$(CN)$_3$ experiences a lowering of magnetic dimensionality owing to destructive interference of magnetic interactions in one of the directions. A necessary test would be a comparison of magnetic excitation spectra of this model to available experimental Raman scattering data on magnetic excitations in κ-(BEDT-TTF)$_2$Cu$_2$(CN)$_3$ and antiferromagnetically ordered BEDT-TTF–based material. At this point, it is clear that the spectrum of collective excitations in κ-Hg-Br is very different from that of κ-(BEDT-TTF)$_2$Cu$_2$(CN)$_3$. Based on that, we can suggest that κ-(BEDT-TTF)$_2$Cu$_2$(CN)$_3$ is a regular dimer Mott insulator with a homogeneous distribution of charge on $(\text{BEDT-TTF})_2^{1+}$ dimer on relevant time scales. If the quantum dipole liquid model is relevant to κ-(BEDT-TTF)$_2$Cu$_2$(CN)$_3$ at all, it puts this compound quite far from a quantum phase transition into a dipole solid state. According to the model in (*1*), tuning between a quantum dipole solid and a quantum dipole liquid can be accomplished by varying the t_b/t_d ratio, where t_b is an overlap integral between the dimers and t_d is an intradimer one. Indeed, t_d is found to be larger for κ-(BEDT-TTF)$_2$Cu$_2$(CN)$_3$ than for κ-Hg-Cl (*10*). The existing experimental data do not provide straightforward evidence of tuning from a dipole liquid to a dipole solid by hydrostatic or chemical pressure for κ-Hg-Br and κ-Hg-Cl family of materials. Although a charge ordered state in κ-Hg-Cl is suppressed by an external pressure of about 1 kbar (*36*), the unit cell of κ-Hg-Br is somewhat larger than that of κ-Hg-Cl. Calculations of the electronic structure of these materials and its change with pressure, as well as further explorations of magnetic properties, are necessary for further understanding of the phase diagram.

REFERENCES AND NOTES

1. C. Hotta, *Phys. Rev. B* **82**, 241104 (2010).
2. M. Naka, S. Ishihara, *Phys. Rev. B* **93**, 195114 (2016).
3. N. Y. Yao, M. P. Zaletel, D. M. Stamper-Kurn, A. Vishwanath, *Nat. Phys.* **14**, 405–410 (2018).
4. S. Rowley *et al.*, *Nat. Phys.* **10**, 367–372 (2014).
5. S.-P. Shen *et al.*, *Nat. Commun.* **7**, 10569 (2016).
6. S. Dayal, R. T. Clay, H. Li, S. Mazumdar, *Phys. Rev. B* **83**, 245106 (2011).
7. K. Yakushi, K. Yamamoto, T. Yamamoto, Y. Saito, A. Kawamoto, *J. Phys. Soc. Jpn.* **84**, 084711 (2015).
8. K. Sedlmeier *et al.*, *Phys. Rev. B* **86**, 245103 (2012).
9. M. Aldoshina *et al.*, *Synth. Met.* **56**, 1905–1909 (1993).
10. N. Drichko *et al.*, *Phys. Rev. B* **89**, 075133 (2014).
11. R. Lyubovskii, R. Lyubovskaya, O. Dyachenko, *J. Phys. I* **6**, 1609–1630 (1996).
12. T. Ivek *et al.*, *Phys. Rev. B* **96**, 085116 (2017).
13. M. Dressel, N. Drichko, *Chem. Rev.* **104**, 5689–5716 (2004).
14. T. Yamamoto *et al.*, *J. Phys. Chem. B* **109**, 15226–15235 (2005).
15. Y. Kim *et al.*, *Appl. Phys. Lett.* **100**, 071907 (2012).
16. K. Yakushi, *Crystals (Basel)* **2**, 1291–1346 (2012).
17. A. Girlando *et al.*, *Phys. Rev. B* **89**, 174503 (2014).
18. T. P. Devereaux, R. Hackl, *Rev. Mod. Phys.* **79**, 175–233 (2007).
19. N. Drichko, R. Hackl, J. A. Schlueter, *Phys. Rev. B* **92**, 161112 (2015).
20. Y. Nakamura *et al.*, *J. Phys. Soc. Jpn.* **83**, 074708 (2014).
21. Y. Nakamura, T. Hiramatsu, Y. Yoshida, G. Saito, H. Kishida, *J. Phys. Soc. Jpn.* **86**, 014710 (2016).
22. Y. Shimizu, K. Miyagawa, K. Kanoda, M. Maesato, G. Saito, *Phys. Rev. Lett.* **91**, 107001 (2003).
23. N. Drichko *et al.*, *Physica C* **460**, 125–128 (2007).
24. J. Merino, A. Greco, R. H. McKenzie, M. Calandra, *Phys. Rev. B* **68**, 245121 (2003).
25. M. Dressel, N. Drichko, J. Schlueter, J. Merino, *Phys. Rev. Lett.* **90**, 167002 (2003).
26. S. Caprara, C. Di Castro, M. Grilli, D. Suppa, *Phys. Rev. Lett.* **95**, 117004 (2005).
27. V. Malinovsky, A. Sokolov, *Solid State Commun.* **57**, 757–761 (1986).
28. G. Chen, L. Balents, A. P. Schnyder, *Phys. Rev. Lett.* **102**, 096406 (2009).
29. L. Balents, *Nature* **464**, 199–208 (2010).
30. S. Yamashita *et al.*, *Nat. Phys.* **4**, 459–462 (2008).
31. Y. Zhou, K. Kanoda, T.-K. Ng, *Rev. Mod. Phys.* **89**, 025003 (2017).
32. M. Naka, S. Ishihara, *Sci. Rep.* **6**, 20781 (2016).
33. Y. Nakazawa, S. Yamashita, *Crystals (Basel)* **2**, 741–761 (2012).
34. P. W. Anderson, G. Baskaran, Z. Zou, T. Hsu, *Phys. Rev. Lett.* **58**, 2790–2793 (1987).
35. B. J. Powell, E. P. Kenny, J. Merino, *Phys. Rev. Lett.* **119**, 087204 (2017).
36. A. Löhle *et al.*, *J. Phys. Condens. Matter* **29**, 055601 (2017).
37. H. C. Kandpal, I. Opahle, Y.-Z. Zhang, H. O. Jeschke, R. Valentí, *Phys. Rev. Lett.* **103**, 067004 (2009).

ACKNOWLEDGMENTS

Funding: The work at the Institute of Quantum Matter was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Material Sciences and Engineering under grant no. DE-FG02-08ER46544. The work in Chernogolovka was supported by FASO Russia, state registration number 0089-2014-0036. J.A.S acknowledges support from the Independent Research and Development program from the NSF while working at the Foundation and from the National High Magnetic Field Laboratory (NHMFL) User Collaboration Grants Program (UCGP). Work at ANL was supported by University of Chicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne") Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under contract no. DE-AC02-06CH11357. **Author contributions**: N.D. conceived and designed the experiments; N.H., S.C., and N. D. performed the experiments and data analysis; M.M. led heat capacity experiments; J.A.S. contributed characterized samples of κ-(BEDT-TTF)$_2$Cu$_2$(CN)$_3$; S.T., E.I.Z., and R.N.L. contributed characterized samples of κ-(BEDT-TTF)$_2$Hg(SCN)$_2$Br and κ-(BEDT-TTF)$_2$Hg(SCN)$_2$Cl. **Competing interests:** The authors declare that they have no competing financial interests. **Data availability:** Tables of data presented in the paper are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1101/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 and S2
Tables S1 and S2
References (*38*–*40*)
Data Files

19 May 2017; accepted 7 May 2018
10.1126/science.aan6286

APPLIED OPTICS

Imaging-based molecular barcoding with pixelated dielectric metasurfaces

Andreas Tittl[1], Aleksandrs Leitis[1], Mingkai Liu[2], Filiz Yesilkoy[1], Duk-Yong Choi[3], Dragomir N. Neshev[2], Yuri S. Kivshar[2], Hatice Altug[1]*

Metasurfaces provide opportunities for wavefront control, flat optics, and subwavelength light focusing. We developed an imaging-based nanophotonic method for detecting mid-infrared molecular fingerprints and implemented it for the chemical identification and compositional analysis of surface-bound analytes. Our technique features a two-dimensional pixelated dielectric metasurface with a range of ultrasharp resonances, each tuned to a discrete frequency; this enables molecular absorption signatures to be read out at multiple spectral points, and the resulting information is then translated into a barcode-like spatial absorption map for imaging. The signatures of biological, polymer, and pesticide molecules can be detected with high sensitivity, covering applications such as biosensing and environmental monitoring. Our chemically specific technique can resolve absorption fingerprints without the need for spectrometry, frequency scanning, or moving mechanical parts, thereby paving the way toward sensitive and versatile miniaturized mid-infrared spectroscopy devices.

The mid-infrared (mid-IR) spectrum is essential for sensing because of the presence of characteristic molecular absorption fingerprints originating from the intrinsic vibrational modes of chemical bonds. Mid-IR spectroscopy allows direct characterization of molecular structures with chemical specificity unique to this spectral range, and is therefore widely recognized as the gold standard for chemical analysis (*1*). It is also a powerful nondestructive and label-free technique for identifying biochemical building blocks, including proteins, lipids, and DNA. However, because of the mismatch between mid-IR wavelengths and dimensions of molecules, the sensitivity of mid-IR spectroscopy is limited when detecting signals from nanometer-scale samples (*2*), biological membranes (*3*), or low numbers of surface-bound molecules (*4*).

Nanophotonics can overcome this limitation by exploiting the strong near-field enhancement of subwavelength resonators. When the resonance is spectrally overlapped with the absorption fingerprints, the enhanced molecule-resonator coupling can lead to a change in either the frequency or the strength of the resonance, from which the molecular fingerprints can be extracted. This concept, surface-enhanced infrared absorption (SEIRA), has been realized using various plasmonic platforms (*5*–*7*); however, the achieved performance is still far from ideal because of the inherent limitation of low-*Q* (quality factor) resonances imposed by resistive loss. Nanostructured resonators based on high-index dielectric materials offer low intrinsic loss and complementary metal-oxide semiconductor (CMOS) compatibility and have recently emerged as a flexible toolkit for various metadevices, with capabilities for controlling the propagation and localization of light (*8*, *9*). Many applications, including generalized wavefront control (*10*, *11*), ultrathin optical elements (*12*, *13*), and antenna-based nanofocusing (*14*), have been shown experimentally. A key concept underlying the functionalities of many metasurface approaches is their use of constituent elements with spatially varying optical properties. The full potential of metasurface-based SEIRA with both spectral and spatial control over nanoscale field enhancement has yet to be realized.

We report a mid-IR nanophotonic sensor based on all-dielectric high-*Q* metasurface elements and demonstrate its capability for enhancing, detecting, and differentiating the absorption fingerprints of various molecules. Traditionally, high-*Q* resonances in metasurfaces are generated via the interference of superradiant and subradiant modes (*15*–*17*). In contrast, our design exploits the collective behavior of Mie resonances, which can be recognized as supercavity modes driven by the physics of bound states in the continuum (*18*). Furthermore, the high-*Q* resonances of our design are spectrally clean without additional resonance background, which is particularly attractive because it allows for the highly spectrally selective enhancement of spectroscopically rich molecular fingerprint information. Specifically, we implement a two-dimensional (2D) array of high-*Q* metasurface pixels, where the resonance positions of individual metapixels are linearly varied over a target mid-IR fingerprint range. This configuration allows us to assign each resonance position to a specific pixel of the metasurface, establishing a one-to-one mapping between spectral and spatial information (Fig. 1A). By comparing the imaging-based readout of this spatially encoded vibrational information before and after the coating of target analyte molecules, we demonstrate chemically specific molecular barcodes suitable for chemical identification and compositional analysis.

Individual metapixels contain a zigzag array of anisotropic hydrogenated amorphous silicon (a-Si:H) resonators, which provide high-*Q* resonances when excited with linearly polarized light and allow for straightforward resonance tuning via scaling of the unit cell geometry by a factor *S* (Fig. 1B). Numerically simulated reflectance spectra of an exemplary 5 × 5 metasurface pixel array with a scaling factor variation from $S = 1.0$ to $S = 1.3$ show sharp resonance peaks (average $Q > 200$) with near-unity reflectance intensity and linear tunability of the resonance positions covering the spectral range from 1350 to 1750 cm^{-1} (Fig. 1C). Our metapixel design also provides enhancement of the local electric near-field intensity confined to the resonator surface by more than three orders of magnitude (*19*), which is ideal for the sensitive detection of analyte molecular vibrations (Fig. 1D and fig. S1). The target spectral range from 1350 to 1750 cm^{-1} contains characteristic molecular stretching and bending vibrations found in hydrocarbons and amino acids, making it crucial for detecting and differentiating the absorption signatures of biomolecules, environmental pollutants, and polymeric species, among others. We first focus on a biosensing application by showing chemical-specific protein detection, where the distinct protein absorption fingerprint is governed by the amide I and II vibrational bands located near 1660 cm^{-1} and 1550 cm^{-1}, respectively.

A sub–5-nm conformal protein layer covering the pixelated metasurface causes a pronounced modulation of the individual metapixel reflectance spectra due to the coupling between the molecular vibrations and the enhanced electric near-fields around the dielectric resonators. This reflectance modulation manifests primarily as an attenuation and broadening of the metapixel resonance, which are correlated with the strength of the amide I and II molecular vibrations (Fig. 1E). The envelope of the metapixel reflectance spectra unambiguously reproduces the protein absorption signature, confirming efficient molecular fingerprint detection. The metapixel resonances provide linewidths much narrower than the spectral feature size of the individual amide I and II absorption bands near 60 cm^{-1}. This is in strong contrast to metal-based antennas used in plasmonic SEIRA approaches, which typically exhibit linewidths above 200 cm^{-1} limited by the intrinsic damping of the metal (*5*). This advantage allows us to read out the protein absorption signature at multiple discrete frequency points and to translate this spectrally resolved absorption information into a barcode-like spatial map of the individual metapixel absorption signals (Fig. 1F).

A pixelated dielectric metasurface design consisting of an array of 10 × 10 metapixels was fabricated using electron-beam lithography and

[1]Institute of BioEngineering, École Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. [2]Nonlinear Physics Centre, Research School of Physics and Engineering, Australian National University, Canberra, ACT 2601, Australia. [3]Laser Physics Centre, Research School of Physics and Engineering, Australian National University, Canberra, ACT 2601, Australia.
*Corresponding author. Email: hatice.altug@epfl.ch

reactive ion beam etching. Ellipse axes and unit cell periodicities were identical to the values given for the numerical simulations in Fig. 1, and the unit cell was linearly scaled between S = 1.00 and S = 1.34 in 100 steps. A fixed metapixel size of 100 μm × 100 μm was chosen to provide a trade-off between metapixel signal-to-noise ratio and number of pixels (Fig. 2A and fig. S2). Analysis of scanning electron microscopy (SEM) images captured for multiple metapixels confirmed the accurate reproduction of the resonator design as well as the linear scaling of the unit cell geometry over the entire metasurface area (Fig. 2B and fig. S3).

The metasurface was characterized in reflection using a quantum cascade laser-based mid-IR microscope equipped with a 480 × 480 pixel array-based imaging detector (Daylight Solutions, Spero). We used a refractive 4× objective with a 2-mm field of view to acquire the optical response of all metasurface pixels simultaneously (Fig. 2C). Reflectance images captured for different wavenumbers of the incident mid-IR radiation are shown in Fig. 2D. At each incident wavenumber, high reflectance intensity indicates the excitation of a metapixel with matching resonance frequency in a specific spatial location on the metasurface. Reflectance spectra are obtained by combining reflectance images for a range of incident wavenumbers and integrating the resulting spectrally resolved data over the individual metapixel areas (*19*). The spectra of 21 exemplary metapixels and extracted resonance positions of all 100 metasurface pixels are shown in Fig. 2, E and F. The fabricated pixelated metasurface delivers resonances with low average full width at half maximum (FWHM) of 13.7 cm^{-1} and uniform tuning of the resonance frequency over the amide band range from 1370 to 1770 cm^{-1}. This corresponds to a spectral resolution of 4 cm^{-1} and an average Q of 115 (fig. S4), which is an improvement of more than one order of magnitude over metallic antenna geometries (*20*, *21*). Note that our design can easily be extended to cover a larger spectral region by increasing the range of geometrical scaling parameters (fig. S5).

We demonstrated molecular fingerprint detection by interrogating a physisorbed monolayer of recombinant protein A/G. Metapixel reflectance spectra before and after the protein A/G physisorption are shown in Fig. 3, A and B, respectively. All spectra are normalized to the peak reflectance values of the reference measurement without analyte. The absorbance signal calculated from the peak reflectance envelopes before (R_0) and after physisorption (R_S) reveals the characteristic amide I and II absorption signature of the protein A/G molecules, which is in good agreement with an independent infrared reflection-absorption spectroscopy (IRRAS) measurement (Fig. 3C). Furthermore, the high absorbance signal of up to A = 140 mOD (milli–optical density units) extracted from a protein monolayer demonstrates the strong vibrational enhancement of our metasurface design, which exceeds the performance of widely used metal antenna geometries (*22*) by more than

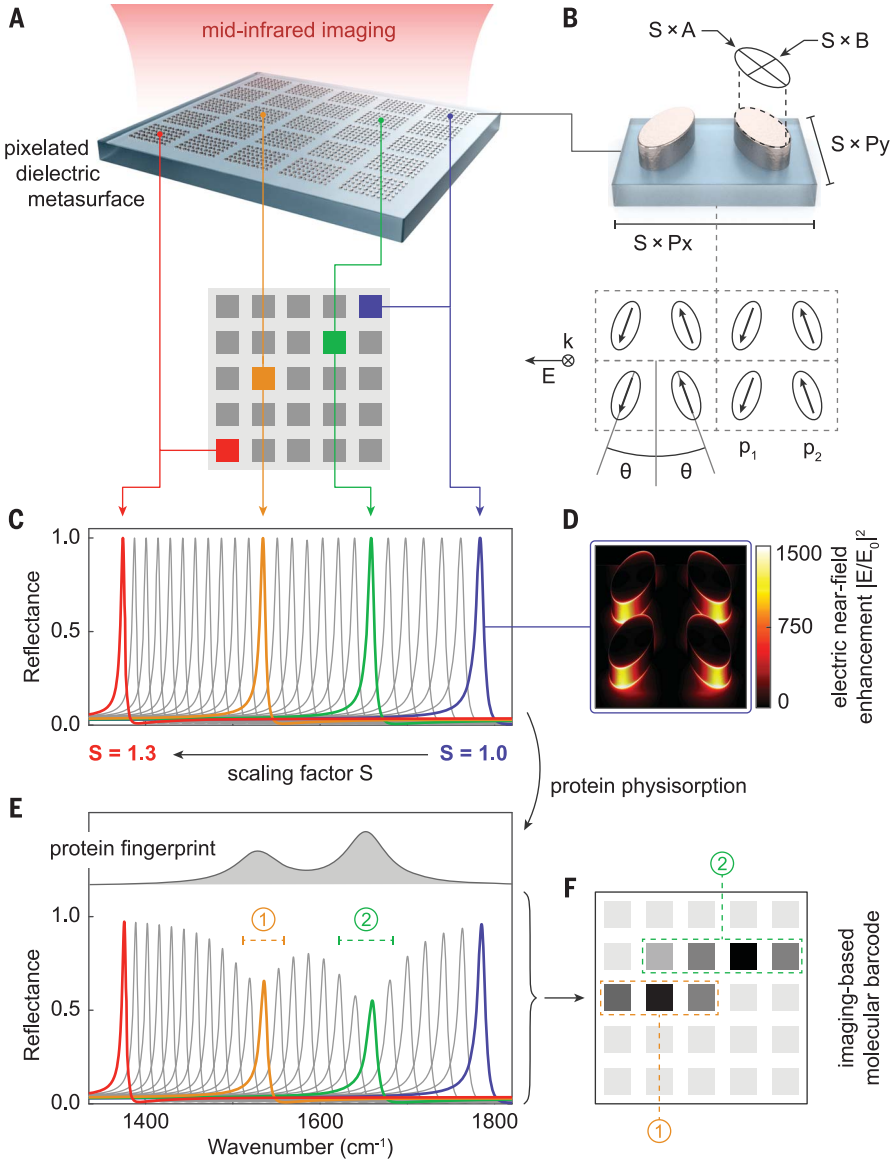


Fig. 1. Molecular fingerprint detection with pixelated dielectric metasurfaces. (**A**) Pixelated metasurface composed of a two-dimensional array of high-Q resonant metapixels with resonance frequencies tuned over a target molecular fingerprint range. (**B**) Spectrally clean high-Q (Q > 200) resonances are provided by zigzag arrays of anisotropic a-Si:H resonators. Resonance frequencies are controlled by scaling the unit cell lateral dimensions by a factor S. Geometrical parameters are A = 1.96 μm, B = 0.96 μm, Px = 3.92 μm, and Py = 2.26 μm, with a fixed structure height of H = 0.7 μm and an orientation angle of θ = 20°; p_1 and p_2 indicate the electric dipole moments of the individual resonators, and k represents the wave vector of the incident light. (**C**) Numerically simulated metapixel reflectance spectra for different values of the scaling parameter S, chosen to cover the amide band spectral region around 1600 cm^{-1}. (**D**) Simulated electric near-field intensity enhancement $|E/E_0|^2$ for S = 1, where $|E_0|$ denotes the incident field amplitude. (**E**) The envelope of metapixel reflectance amplitudes reproduces the absorption fingerprint of an adjacent model protein layer (top inset). (**F**) Conceptual sketch of a molecule-specific barcode produced by imaging-based readout of the metasurface's reflectance response. Image regions 1 and 2 indicate the spatially encoded vibrational information from the corresponding metapixel resonances in (E).

one order of magnitude (fig. S6). Combined with an experimental noise level of 1.8 mOD, this value corresponds to a detection limit of 2130 molecules/μm^2 (fig. S7).

Miniaturization of IR sensor devices has traditionally been challenging because of the need for scaled-down Fourier-transform IR (FTIR) spectrometers or frequency scanning laser sources (*23*). Recent sensor-on-chip approaches based on thin-film optical waveguides have made progress with monolithic integration, in which the light source, sensing element, and detector are fabricated on a single chip (*24*–*26*). In comparison, our technique enables a complementary integration scheme in which the pixelated metasurface sensor can be combined with an IR imaging detector such as a high-resolution microbolometer or a mercury cadmium telluride (MCT) focal plane array (fig. S8). When illuminated with an external broadband light source, this arrangement can overcome the need for a mid-IR spectrometer in a compact footprint. Furthermore, our scheme addresses several constraints of complete monolithic integration, such as the limited operating range of chip-based light sources and detectors (*27*), as well as sensitivity limitations caused by the low near-field enhancement factors of conventional waveguides (*26*).

We assessed the capability of our metasurface sensor for imaging-based spectrometerless fingerprint detection by calculating the integrated reflectance signal from the spectral data of each metapixel. These integrated signals are analogous to a readout of the metasurface's optical response with a broadband detector before (I_0) and after (I_S) addition of the protein layer (Fig. 3D) and are used to calculate metapixel absorbance signals via $A = -\log(I_S/I_0)$, which are then rescaled from 0 to 100 for presentation (Fig. 3E). The resulting barcode-like spatial absorption map of the protein A/G monolayer clearly shows the spectral location and relative intensity distribution of the characteristic amide I and II absorption bands as two distinct high-signal regions of the image, providing chemically specific fingerprint detection in a miniaturized design and without the need for spectrometry. This functionality is enabled by the spectrally clean high-Q resonances of the dielectric metapixels and cannot be achieved with metapixels based on conventional metal antennas as a result of linewidth limitations (fig. S9). In addition to protein A/G, we also tested a polymer mixture composed of polymethyl methacrylate (PMMA) and polyethylene (PE) as well as glyphosate pesticide to cover applications in fields as diverse as biosensing, materials science, and environmental monitoring. In all three cases, the molecular barcodes feature mutually distinct high-intensity image regions unique to the vibrational signature of the investigated analytes (Fig. 4A), underscoring the chemical identification capability.

Our barcoding technique offers the potential for identifying molecular species in arbitrary analyte compositions through pattern recognition based on a library of multiple molecular barcode signatures (fig. S10). To illustrate this approach, we detected a series of predefined mixtures of PMMA and PE polymer molecules deposited on the metasurface by thermal evaporation. Figure 4B shows molecular barcodes for pure PMMA and PE as well as PMMA/PE mixing ratios of 0.25, 0.50, and 0.75. The characteristic molecular signatures of PMMA and PE appear as distinct image features in the top and bottom halves of the barcode matrix, respectively. When increasing the relative amount of PE in the mixture, we observed a substantial increase of the PE signal versus mixing ratio combined with an associated decrease of the PMMA signal.

We carried out further image-based analysis by decomposing the barcode matrices of all mixing states Y_{mix} into a linear combination of the PMMA and PE molecular barcodes according to

$$Y_{mix} = (\beta_{PMMA} \cdot X_{PMMA}) + (\beta_{PE} \cdot X_{PE}) \quad (1)$$

where X_{PMMA} and X_{PE} are the input barcodes of the pure materials, and β_{PMMA} and β_{PE} are the output coefficients associated with the analyte

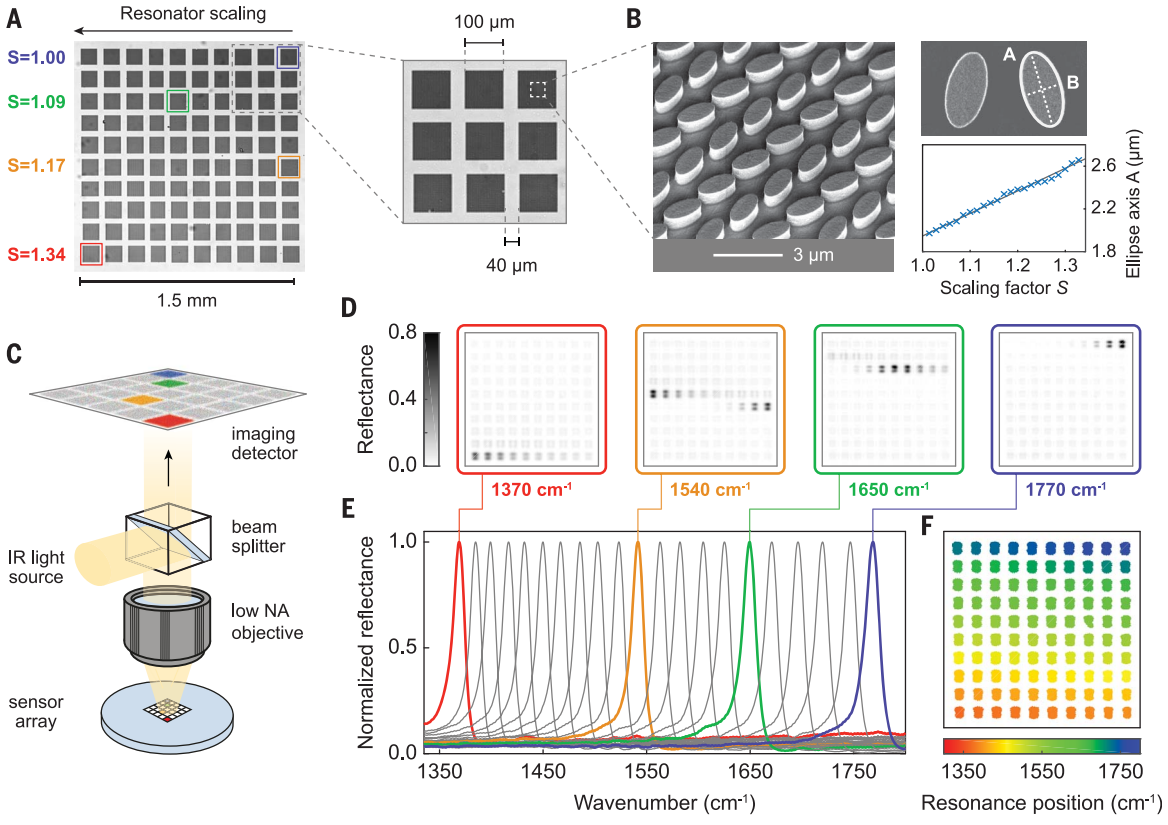


Fig. 2. Experimental realization of the pixelated metasurfaces. (**A**) Optical images of the fabricated 100-pixel metasurface. (**B**) SEM micrographs confirm the linear relationship between scaling factor and ellipse feature size. (**C**) Sketch of the imaging-based mid-IR microscopy system. (**D**) Reflectance images of the pixelated metasurface recorded at four specific wavenumbers in the mid-IR spectral range. (**E**) Normalized reflectance spectra for 21 of the 100 metapixels. Resonance positions of the colored curves correspond to the respective reflectance images in (D). (**F**) Extracted resonance positions for all metapixels.

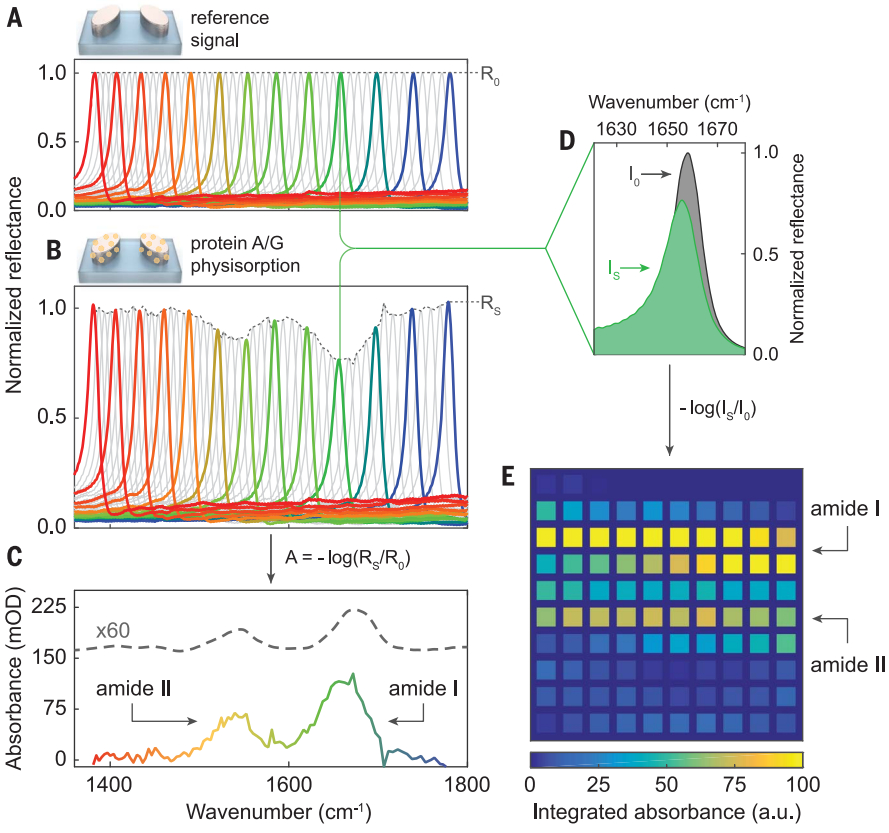


Fig. 3. Molecular fingerprint retrieval and spatial absorption mapping. (**A**) Normalized metapixel reflectance spectra before physisorption of a protein A/G monolayer. R_0 indicates the envelope of peak reflectance amplitudes (dashed line). (**B**) Normalized spectra after protein physisorption, including reflectance envelope R_S (dashed line). (**C**) Protein absorption fingerprint calculated from the reflectance envelopes R_0 and R_S compared to an independent IRRAS measurement (dashed line, scaled and offset for clarity). (**D**) Broadband spectrometerless operation of the metasurface can be emulated by integrating the reflectance signal of all pixels. (**E**) Spectral integration translates the absorption signature in (C) into a 2D spatial absorption map, which represents the molecular barcode of the protein (a.u., arbitrary units).

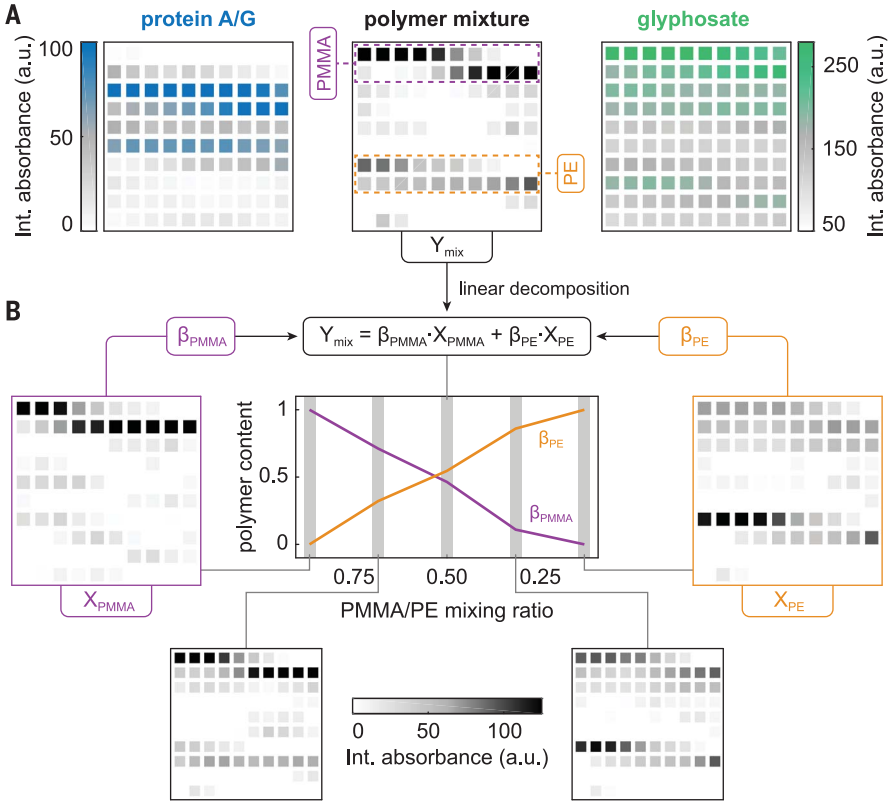


Fig. 4. Imaging-based chemical identification and compositional analysis. (**A**) Molecular barcodes of protein A/G, a mixture of PMMA and PE polymers, and glyphosate pesticide reveal the distinct absorption fingerprints of the analytes. (**B**) Barcode matrices for PMMA/PE polymer mixtures with several mixing ratios. Linear decomposition analysis of all mixing states Y_{mix} with respect to the pure PMMA and PE barcode matrices confirms accurate readout of the deposited polymer ratios.

content on the surface (Fig. 4B, center). The PMMA and PE polymer amounts obtained from our image decomposition analysis accurately captured the linear variation of the polymer composition, highlighting the rich chemical and compositional information available from such absorption maps. Although simple linear decomposition is implemented in this demonstration, nonlinear processes (such as those in biomolecular interactions involving multiple analytes and kinetics) are foreseen to make use of more sophisticated neural network–based image recognition methods and machine learning (*28*, *29*).

Our nanophotonic technique offers the prospect of IR absorption spectroscopy without the need for complex instrumentation. The Si-based pixelated metasurface is compatible with CMOS technology for low-cost wafer-scale sensor fabrication and can be combined with state-of-the-art surface functionalization techniques for operation in more complex bioassays (*30*). The sensitivity and *Q*-factor of our metasurface can be further improved by decreasing the resonator orientation angle, limited only by the inhomogeneity of the nanofabrication. Additionally, even stronger near-field enhancement could be achieved by using more sophisticated designs for meta-atoms. The molecular barcodes obtained with our method offer unique possibilities for advanced image analysis, paving the way toward versatile and sensitive miniaturized mid-IR spectroscopy devices.

REFERENCES AND NOTES

1. B. H. Stuart, *Infrared Spectroscopy: Fundamentals and Applications* (Wiley, 2005).
2. D. Dregely, F. Neubrech, H. Duan, R. Vogelgesang, H. Giessen, *Nat. Commun.* **4**, 2237 (2013).
3. O. Limaj *et al*., *Nano Lett.* **16**, 1502–1508 (2016).
4. C. Huck *et al*., *ACS Nano* **8**, 4908–4914 (2014).
5. F. Neubrech, C. Huck, K. Weber, A. Pucci, H. Giessen, *Chem. Rev.* **117**, 5110–5145 (2017).
6. L. Dong *et al*., *Nano Lett.* **17**, 5768–5774 (2017).
7. B. Cerjan, X. Yang, P. Nordlander, N. J. Halas, *ACS Photonics* **3**, 354–360 (2016).
8. A. I. Kuznetsov, A. E. Miroshnichenko, M. L. Brongersma, Y. S. Kivshar, B. Luk'yanchuk, *Science* **354**, aag2472 (2016).
9. A. Y. Zhu, A. I. Kuznetsov, B. Luk'yanchuk, N. Engheta, P. Genevet, *Nanophotonics* **6**, 1–20 (2017).
10. A. Arbabi, Y. Horie, M. Bagheri, A. Faraon, *Nat. Nanotechnol.* **10**, 937–943 (2015).
11. M. Decker *et al*., *Adv. Opt. Mater.* **3**, 813–820 (2015).
12. M. Khorasaninejad *et al*., *Science* **352**, 1190–1194 (2016).
13. D. Lin, P. Fan, E. Hasman, M. L. Brongersma, *Science* **345**, 298–302 (2014).
14. M. Caldarola *et al*., *Nat. Commun.* **6**, 7915 (2015).
15. M. F. Limonov, M. V. Rybin, A. N. Poddubny, Y. S. Kivshar, *Nat. Photonics* **11**, 543–554 (2017).
16. C. Wu *et al*., *Nat. Commun.* **5**, 3892 (2014).
17. S. Campione *et al*., *ACS Photonics* **3**, 2362–2367 (2016).
18. M. Rybin, Y. Kivshar, *Nature* **541**, 164–165 (2017).
19. See supplementary materials.
20. K. Chen, R. Adato, H. Altug, *ACS Nano* **6**, 7998–8006 (2012).
21. C. Wu *et al*., *Nat. Mater.* **11**, 69–75 (2011).
22. S. Bagheri *et al*., *ACS Photonics* **2**, 779–786 (2015).
23. F. K. Tittel, D. Richter, A. Fried, in *Solid-State Mid-Infrared Laser Sources*, I. T. Sorokina, K. L. Vodopyanov, Eds. (Springer, 2003), pp. 458–529.
24. B. Schwarz *et al*., *Nat. Commun.* **5**, 4085 (2014).
25. A. Harrer *et al*., *Sci. Rep.* **6**, 21795 (2016).
26. M. Sieger, B. Mizaikoff, *Anal. Chem.* **88**, 5562–5573 (2016).
27. H. Lin *et al*., *Nanophotonics* **7**, 393–420 (2017).
28. G. E. Hinton, R. R. Salakhutdinov, *Science* **313**, 504–507 (2006).
29. Y. LeCun, Y. Bengio, G. Hinton, *Nature* **521**, 436–444 (2015).
30. S. Jiang *et al*., *Nanoscale* **5**, 3127–3148 (2013).

ACKNOWLEDGMENTS

We thank R. Guo, E. Romero Arvelo, and D. Rodrigo for useful discussions and École Polytechnique Fédérale de Lausanne and Center of MicroNano Technology for nanofabrication. Sample fabrication was performed in part at the ACT node of the Australian National Fabrication Facility. **Funding:** The research leading to these results has received funding from the European Research Council under grant agreement no. 682167 VIBRANT-BIO and the European Union Horizon 2020 Framework Programme for Research and Innovation under grant agreements no. 665667 (call 2015), no. 777714 (NOCTURNO project), no. FETOPEN-737071 (ULTRACHIRAL project), and no. 644956 (RAIS project). The authors acknowledge the support of the Australian Research Council. **Author contributions:** A.T., D.N.N., Y.S.K., and H.A. conceived and designed the research; A.L., F.Y., and D.-Y.C. fabricated the dielectric metasurfaces; A.T. and A.L. carried out optical measurements and analyzed data; A.L. and M.L. carried out numerical simulations; and all authors contributed to writing the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1105/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
References (*31*–*34*)

24 January 2018; accepted 13 April 2018
10.1126/science.aas9768

PHOTOSYNTHESIS

Structure of the maize photosystem I supercomplex with light-harvesting complexes I and II

Xiaowei Pan[1]*, Jun Ma[1]*, Xiaodong Su[1]*, Peng Cao[1], Wenrui Chang[1,2], Zhenfeng Liu[1,2], Xinzheng Zhang[1,2,3]†, Mei Li[1]†

Plants regulate photosynthetic light harvesting to maintain balanced energy flux into photosystems I and II (PSI and PSII). Under light conditions favoring PSII excitation, the PSII antenna, light-harvesting complex II (LHCII), is phosphorylated and forms a supercomplex with PSI core and the PSI antenna, light-harvesting complex I (LHCI). Both LHCI and LHCII then transfer excitation energy to the PSI core. We report the structure of maize PSI-LHCI-LHCII solved by cryo–electron microscopy, revealing the recognition site between LHCII and PSI. The PSI subunits PsaN and PsaO are observed at the PSI-LHCI interface and the PSI-LHCII interface, respectively. Each subunit relays excitation to PSI core through a pair of chlorophyll molecules, thus revealing previously unseen paths for energy transfer between the antennas and the PSI core.

During oxygenic photosynthesis, photosystems II (PSII) and I (PSI) operate in series and are tightly coupled to allow efficient light-driven electron transport. Both photosystems are multisubunit supramolecular complexes containing a core complex and the peripheral antenna system (*1*, *2*). In plants, the peripheral antennas are composed of light-harvesting complexes (LHCs). LHCIs (containing the Lhca apoproteins) are associated with the PSI core, forming the PSI-LHCI complex, whereas LHCIIs (containing the Lhcb apoproteins) are mostly attached to the PSII core, constituting the PSII-LHCII complex. The antenna systems of PSI or PSII have different pigment compositions and hence different light absorption properties. Red and far-red light preferentially stimulate PSII and PSI, respectively, and fluctuating illumination can cause unequal excitation of the two photosystems. Balanced light harvesting is crucial for efficient photosynthesis; thus, plants have developed both short-term and long-term adaptation to the constantly changing light conditions in the natural environments. State transitions are a short-term response occurring on a time scale of minutes (*3*) and allow a balanced energy distribution between the two photosystems when light quality changes (*4*–*6*). During state transitions, the trimeric LHCII (composed of different combinations of Lhcb1-3) is reversibly phosphorylated and dephosphorylated, a process controlled by the redox state of plastoquinone (PQ) and regulated by a chloroplast kinase (STN7) and phosphatase (PPH1, also known as TAP38) in plants (*5*, *6*). In state 1, LHCIIs are mostly associated with PSII and transfer excitation energy to the PSII core. Under light conditions favoring PSII excitation, overexcitation of PSII leads to a reduction of the PQ pool, activation of STN7 kinase, and subsequent phosphorylation of the N-terminal region of LHCII. A portion of the phosphorylated LHCIIs (mobile LHCII) move laterally within the thylakoid membrane from PSII to PSI, forming the PSI-LHCI-LHCII supercomplex and resulting in a switch from state 1 to state 2. The mobile LHCII serves as a peripheral antenna of PSI in addition to LHCI, increasing energy transfer toward the PSI core. Under natural light conditions, state transitions are necessary for optimal plant growth and fitness (*7*–*9*).

A two-dimensional projection map of the PSI-LHCI-LHCII supercomplex at 16-Å resolution (*10*) revealed a single LHCII trimer associated with the PSI core on the opposite side of LHCI (*10*, *11*), yet the protein-protein and pigment-pigment interactions between LHCII and PSI remain unclear. Although crystal structures of plant LHCII had been solved previously (*12*, *13*), the N-terminal tail of LHCII containing the phosphorylation site was not observed in these structures. It remains an unresolved question as to how the phosphorylated LHCII enhances its interaction with PSI. Plant PSI-LHCI includes a core complex made up of 14 subunits (PsaA to L, PsaN, and PsaO) and a peripheral antenna system containing four LHCI proteins organized as two heterodimers (Lhca1-Lhca4 and Lhca2-Lhca3) (*2*). Recent crystal structures of pea PSI-LHCI (*14*, *15*) revealed the architecture and location of 16 subunits, but two plant-specific subunits, PsaN and PsaO, were not observed in these structures. An intact PSI-LHCI structure should allow

[1]National Laboratory of Biomacromolecules, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, P.R. China. [2]University of Chinese Academy of Sciences, Beijing 100049, P.R. China. [3]Center for Biological Imaging, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, P.R. China.
*These authors contributed equally to this work.
†Corresponding author. Email: xzzhang@ibp.ac.cn (X.Z.); meili@ibp.ac.cn (M.L.)

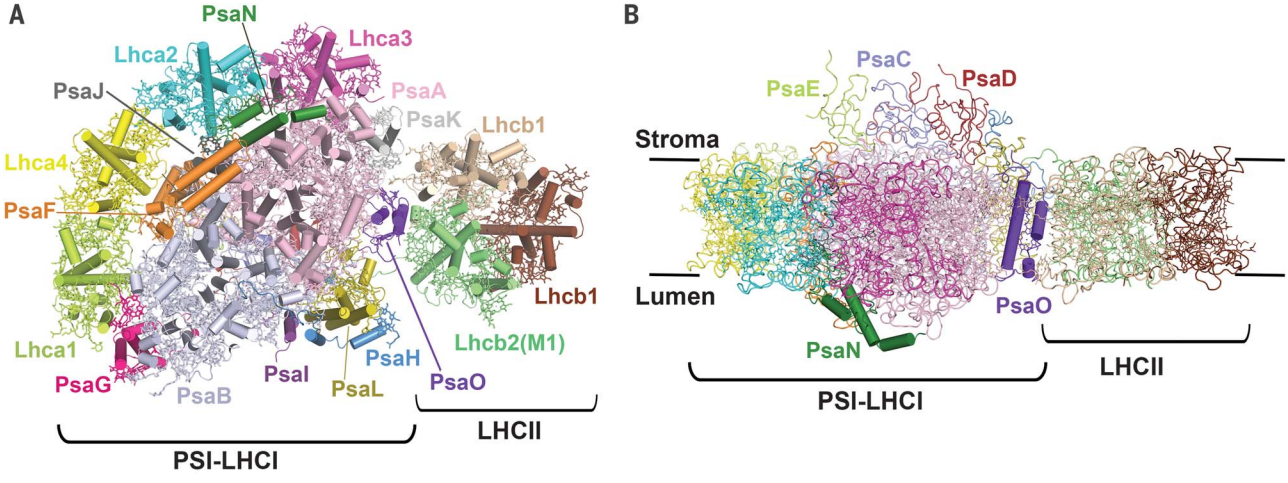


Fig. 1. The maize PSI-LHCI-LHCII supercomplex. (**A**) View from the luminal side. One LHCII monomer with its N-terminal region contacting PSI is labeled as M1. (**B**) Side view along membrane plane. PsaN and PsaO are highlighted in cartoon mode.

for a better understanding of the energy transfer within the PSI-LHCI complex.

We purified PSI-LHCI-LHCII from maize leaves treated with orange light, which induces the shift into state 2 (fig. S1). Using single-particle cryo–electron microscopy (cryo-EM), we solved the PSI-LHCI-LHCII structure at an overall resolution of 3.3 Å (fig. S2 and table S1). One LHCII trimer binds PSI at the PsaA side (Fig. 1). In addition to the 16 subunits present in the previously reported structures (*14*, *15*), we identified two areas of density corresponding to the subunits PsaN and PsaO (Figs. 1B and 2 and fig. S3). The PSI-LHCI moiety in the PSI-LHCI-LHCII supercomplex shows the most complete protein composition, with no known subunit missing (Fig. 1, fig. S4A, and table S2). LHCII exhibits lower resolution than the PSI-LHCI portion; however, its location and orientation are assigned unambiguously as the transmembrane helices are well resolved (fig. S3D). In our PSI-LHCI-LHCII structure, the membrane-spanning regions of PSI-LHCI and LHCII are not coplanar but instead form an angle of about 10° (fig. S3A). The curved structure of the PSI-LHCI-LHCII supercomplex may indicate a role for the membrane-embedded complexes in influencing the native membrane architecture, as previously suggested for the mitochondrial adenosine 5′-triphosphate (ATP) synthase (*16*). Superposition of the PSI-LHCI moiety in our maize supercomplex structure with the two previously reported pea PSI-LHCI crystal structures (*14*, *15*) reveals that, except for PsaN and PsaO, all other 16 PSI-LHCI subunits are similar to their counterparts from pea and exhibit small variations in their hydrophilic loop regions (fig. S4). On the basis of the cryo-EM density, we were able to build several subunits with more complete structures (fig. S4, B to D, and table S2) and assigned a few more chlorophyll molecules (fig. S4, E and F, and table S3).

PsaN exists exclusively in the eukaryotic PSI (*17*) and is the only membrane-extrinsic PSI subunit found at the luminal side (*18*). PsaN is easily detached from PSI by salt washing (*19*). We purified the supercomplex at low ion concentration through one-step isolation after mild solubilization of the thylakoid membranes to keep PsaN attached. We further confirmed the presence of PsaN in our PSI-LHCI-LHCII sample by immunoblotting (fig. S1E). PsaN is composed of three α helices and binds two chlorophylls (named a1001 and a1002) (Fig. 2A and fig. S5). PsaN is located near the luminal regions of PsaA, PsaF, and PsaJ, where it fills the empty space between Lhca2 and PsaA (Fig. 2B and fig. S6, A and B). This location is consistent with previous biochemical and genetic studies (*20*–*22*) and suggests a role in connecting these subcomplexes. In crystal structures of pea PSI-LHCI (*14*, *15*), a large gap is present between Lhca2 and the core complex at the luminal side. Because of the longer distances between interfacial chlorophylls, the excitation energy transfer (EET) from Lhca2 to the core was suggested to mainly flow via Lhca3, an inefficient pathway involving multiple steps. In our structure, two "linker chlorophylls" from PsaN create direct EET pathways between Lhca2 and the PSI core (Fig. 3, A and B, and table S4). Chl a1001 from PsaN is strongly connected with the red form Chl a603 from Lhca2 and is in close contact with Chl a808 from PsaA, thus establishing the major EET pathway from Lhca2 to the PSI core (Fig. 3C). In addition, both chlorophylls from PsaN can receive excitation energy from luminal chlorophylls of Lhca2, further facilitating the energy transfer between Lhca2 and the PSI core. The arrangement of chlorophylls in PSI-LHCI moiety in our structure is consistent with previous picosecond fluorescence spectroscopy showing that Lhca2 can efficiently transfer the absorbed energy to the core in a direct manner (*23*, *24*). PsaN is also necessary for the efficient interaction between PSI and plastocyanin (Pc) (*25*), a soluble protein that transfers electrons in the thylakoid lumen from cytochrome b_6f to PSI complex. The efficient binding of Pc to eukaryotic PSI requires the N-terminal extension domain of PsaF (*26*). In agreement, our structure shows that PsaN forms close contacts with the N-terminal

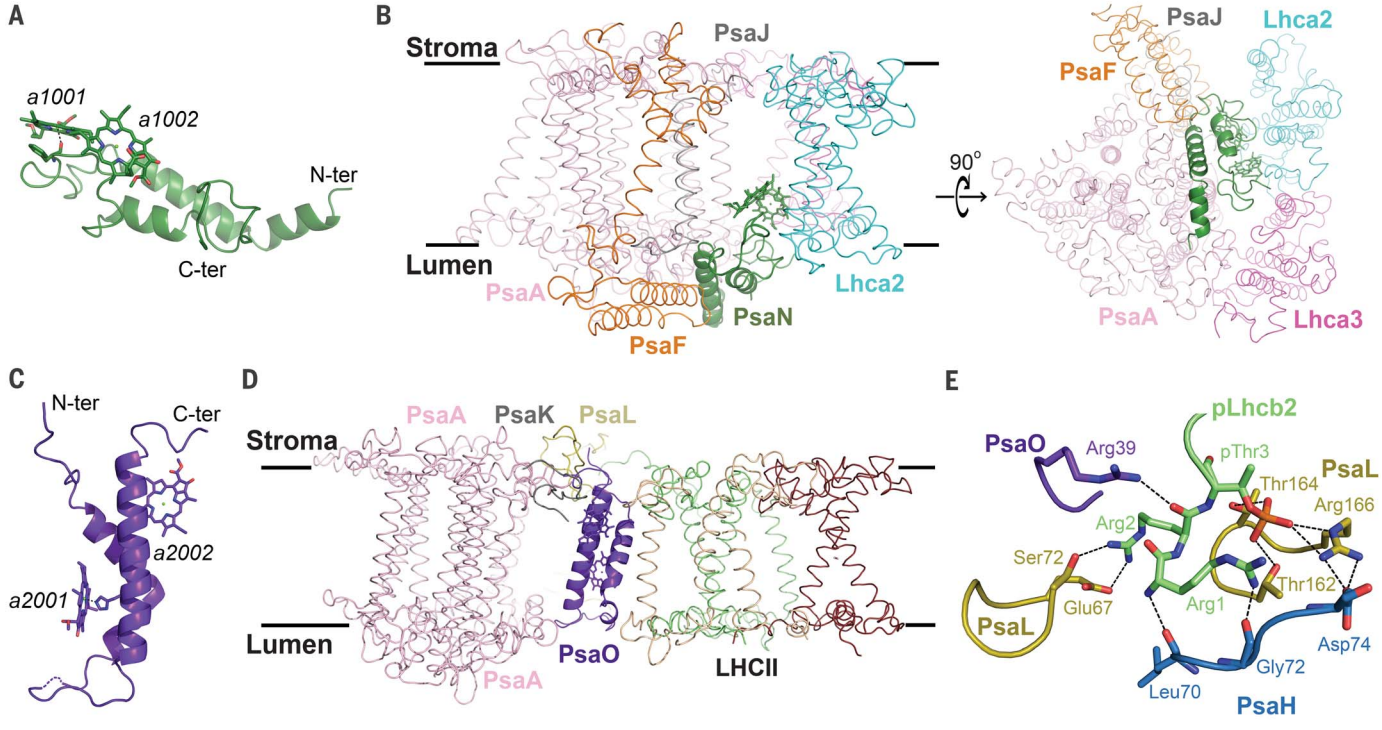


Fig. 2. Structures and interactions of individual subunits in the supercomplex. (**A**) PsaN binds two chlorophylls; Chl a1001 is coordinated by the main-chain carbonyl oxygen of Phe^{106}. (**B**) PsaN is located at the luminal side and interacts with PsaA, PsaF, PsaJ, Lhca2, and Lhca3. (**C**) PsaO is associated with two chlorophyll molecules; Chl a2001 is ligated by His^{103}. A fragment in the luminal loop region of PsaO is partly disordered in the structure and is shown as a dashed line. (**D**) PsaO bridges PsaA and LHCII through contacts within the membrane and on the stromal and luminal surface. (**E**) The interactions of the N-terminal region of pLhcb2 with PsaL, PsaH, and PsaO. Hydrogen bonds with distances between 2.6 and 3.4 Å are shown as black dashed lines.

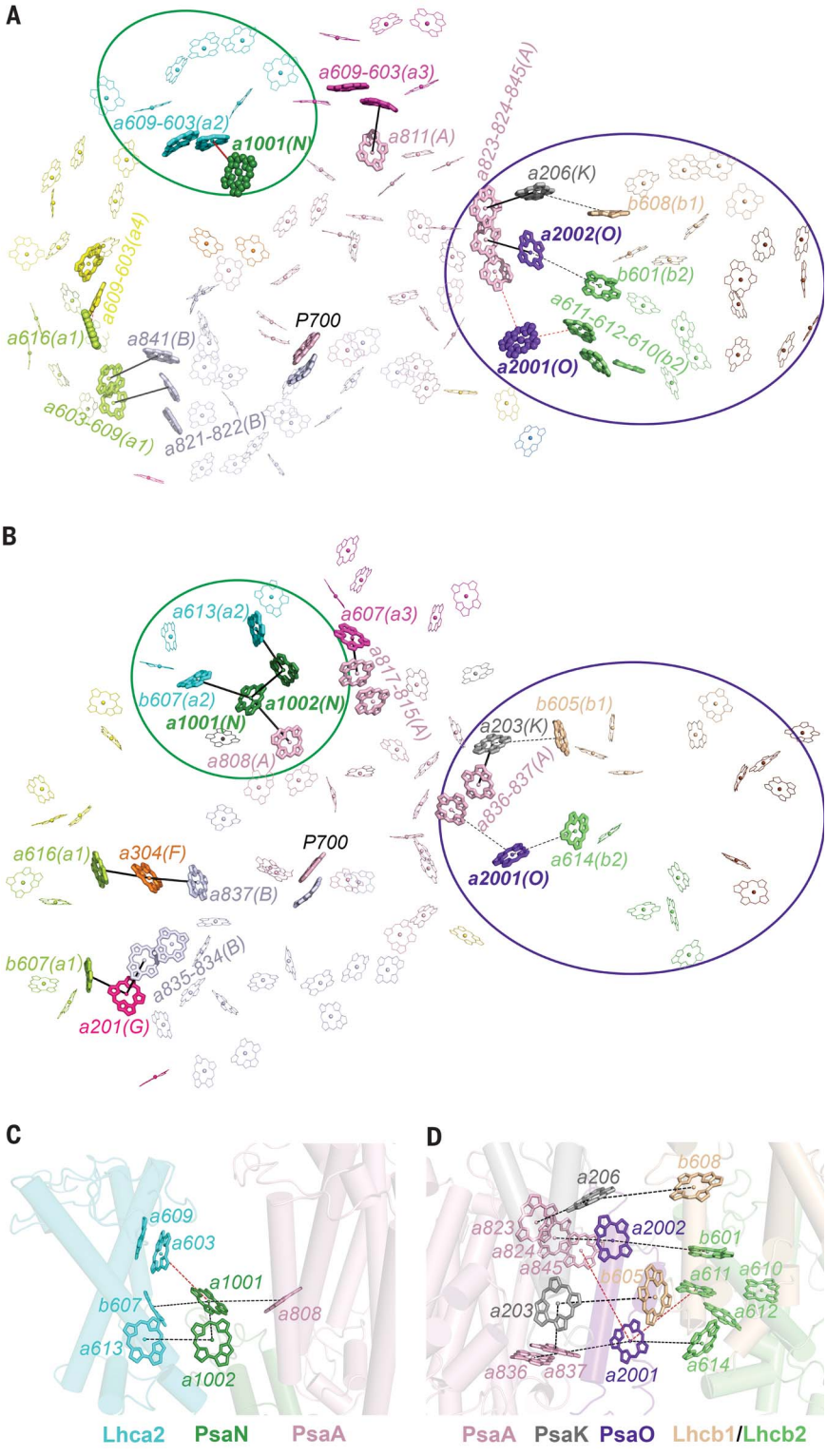


Fig. 3. Chlorophyll arrangement and proposed energy transfer pathways in the PSI-LHCI-LHCII supercomplex. (**A**) Distribution of chlorophylls within the PSI-LHCI-LHCII supercomplex at the stromal layer. Three chlorophylls (a1001 from PsaN, a2001 from PsaO, and a616 from Lhca1) located at the luminal side can receive energy from and transfer energy to chlorophylls at the stromal layer, and are shown in ball-and-stick model. The black and red lines indicate the pathways composed of chlorophyll molecules within the same layer and at different layers, respectively. (**B**) Chlorophylls at the luminal layer. In (A) and (B), the possible EET pathways from Lhca2 to PSI core (mediated by PsaN) and from LHCII to PSI core are highlighted by green and purple circles, respectively. The interfacial chlorophyll pairs with Mg-to-Mg distances shorter or longer than 17 Å are linked with solid or dashed lines, respectively. (**C**) The plausible EET pathways from Lhca2 to PSI core (shown as dashed lines). (**D**) The plausible EET pathways from LHCII to PSI core (shown as dashed lines). For clarity, the phytol chains of chlorophylls are omitted. The distances of interfacial chlorophyll pairs are summarized in table S4.

extension of PsaF (fig. S6C). Therefore, PsaN may be directly involved in the formation of the Pc binding site or assist the association of Pc by stabilizing the N-terminal domain of PsaF (fig. S6D).

We also identified density corresponding to PSI subunit PsaO, with two transmembrane helices and N- and C-termini loops along stromal surface. Two chlorophylls, named a2001 and a2002, are located near the luminal and stromal surfaces, respectively (Fig. 2C and fig. S7). An earlier report suggested that PsaO is involved in the formation of the docking site for LHCII (*27*), and state transitions were reduced by 50% when PsaO expression was abolished in *Arabidopsis* (*28*). Our structure shows a direct connection of PsaO with LHCII and several PSI core subunits (Fig. 2D), in agreement with cross-linking studies (*27*, *28*). PsaO interacts with PsaL, PsaA, and PsaK at the stromal surface, PsaA at the luminal side, and LHCII at their membrane-spanning interface. The N-terminal region of one LHCII monomer (M1) (Fig. 1A) forms close contacts with core subunits PsaL, PsaH, and PsaO at the stromal side (fig. S8). Recent reports showed that the mobile LHCII is mainly composed of $(Lhcb1)_2$ Lhcb2 trimer, and the phosphorylated Lhcb2 (pLhcb2) is responsible for the association of LHCII with PSI under state 2 conditions (*29*–*31*). The specific features of the cryo-EM density around the N-terminal tail of monomer M1 match with the Lhcb2 sequence (figs. S3D and S8C) but not with that of Lhcb1, which is five residues longer than Lhcb2 at the N-terminal region (fig. S9). We therefore assigned the monomer M1 as pLhcb2 and the other two monomers as Lhcb1 (Fig. 1A). The high-quality cryo-EM density (fig. S8C) allowed the construction of the complete N-terminal region of pLhcb2, including its phosphorylated Thr^3 ($pThr^3$) residue. The phosphate group of $pThr^3$ strongly interacts with several residues from PsaL (Fig. 2E). These residues are conserved in plant PsaL and located at an extended stromal loop, which is absent in cyanobacteria (fig. S10). The two basic residues (Arg^1 and Arg^2) preceding $pThr^3$ in Lhcb2 are also pivotal for PSI subunits recognizing LHCII. Arg^2 binds residues from PsaL and PsaO. Arg^1 interacts with the N-terminal regions of PsaH (Fig. 2E). The first three residues are completely conserved among Lhcb2

from different plants, and Lhcb1 always has a Lys residue at its second position (fig. S9), which may be less favorable for interacting with PsaL. The relatively shorter N-terminal region and the specific residue at the second position (Arg^2) of Lhcb2 are likely to account for the specificity of Lhcb2 interacting with the PSI core. In addition, our structure shows that the plant-specific subunits (PsaO and PsaH) and the extended stromal loop of PsaL are essential for PSI to interact with Lhcb2, explaining the crucial role of PsaH and PsaL in state transitions (*32*). The strong interactions at the stromal side between PSI and LHCII are probably responsible for the curved architecture of the supercomplex (fig. S3A).

In addition to its structural role, PsaO also mediates EET from LHCII to the PSI core through its two chlorophylls (Fig. 3, A and B, and table S4). PsaO may receive the excitation energy from Lhcb2 of LHCII at both stromal and luminal layers (Fig. 3D). Chl a2001 from PsaO is located close to Chl a611 from Lhcb2. The Chl a611/a612/a610 cluster in LHCII was proposed to be at the lowest-energy state (*33*), so the excitation energy equilibrated within the LHCII trimer may be focused on this Chl cluster in Lhcb2 and further transferred to PsaO. In addition, PsaK may be involved in alternative EET pathways by receiving the excitation energy from Lhcb1 at both stromal and luminal sides. The excitation energy collected by LHCII will be further transferred to the stromal cluster a823/a824/a845 and the luminal pair a836/a837 in PsaA. The network of chlorophylls within LHCII and PSI core observed in our structure agrees with previous spectroscopic results demonstrating that the LHCII trimer associated with PSI is a highly efficient light harvester (*11*, *34*, *35*). These well-connected interfacial pigments should provide multiple EET pathways between LHCII and the PSI core, similar to the connections mediated by PsaN between LHCI and the PSI core.

REFERENCES AND NOTES

1. J. P. Dekker, E. J. Boekema, *Biochim. Biophys. Acta* **1706**, 12–39 (2005).
2. N. Nelson, C. F. Yocum, *Annu. Rev. Plant Biol.* **57**, 521–565 (2006).
3. A. V. Ruban, M. P. Johnson, *Photosynth. Res.* **99**, 173–183 (2009).
4. J. Kargul, J. Barber, *FEBS J.* **275**, 1056–1068 (2008).
5. J. D. Rochaix *et al*., *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 3466–3474 (2012).
6. M. Goldschmidt-Clermont, R. Bassi, *Curr. Opin. Plant Biol.* **25**, 71–78 (2015).
7. S. Bellafiore, F. Barneche, G. Peltier, J. D. Rochaix, *Nature* **433**, 892–895 (2005).
8. M. Frenkel, S. Bellafiore, J. D. Rochaix, S. Jansson, *Physiol. Plant.* **129**, 455–459 (2007).
9. M. Tikkanen, M. Grieco, S. Kangasjärvi, E. M. Aro, *Plant Physiol.* **152**, 723–735 (2010).
10. R. Kouřil *et al*., *Biochemistry* **44**, 10935–10940 (2005).
11. P. Galka *et al*., *Plant Cell* **24**, 2963–2978 (2012).
12. J. Standfuss, A. C. Terwisscha van Scheltinga, M. Lamborghini, W. Kühlbrandt, *EMBO J.* **24**, 919–928 (2005).
13. Z. Liu *et al*., *Nature* **428**, 287–292 (2004).
14. Y. Mazor, A. Borovikova, I. Caspy, N. Nelson, *Nat. Plants* **3**, 17014 (2017).
15. X. Qin, M. Suga, T. Kuang, J. R. Shen, *Science* **348**, 989–995 (2015).
16. H. Guo, S. A. Bueler, J. L. Rubinstein, *Science* **358**, 936–940 (2017).
17. C. Vanselow, A. P. Weber, K. Krause, P. Fromme, *Biochim. Biophys. Acta* **1787**, 46–59 (2009).
18. V. S. Nielsen, A. Mant, J. Knoetzel, B. L. Møller, C. Robinson, *J. Biol. Chem.* **269**, 3762–3766 (1994).
19. W. Z. He, R. Malkin, *FEBS Lett.* **308**, 298–300 (1992).
20. S. Jansson, B. Andersen, H. V. Scheller, *Plant Physiol.* **112**, 409–420 (1996).
21. A. Haldrup, D. J. Simpson, H. V. Scheller, *J. Biol. Chem.* **275**, 31211–31218 (2000).
22. A. Hansson *et al*., *FEBS J.* **274**, 1734–1746 (2007).
23. E. Wientjes, I. H. van Stokkum, H. van Amerongen, R. Croce, *Biophys. J.* **101**, 745–754 (2011).
24. E. Wientjes, I. H. van Stokkum, H. van Amerongen, R. Croce, *Biophys. J.* **100**, 1372–1380 (2011).
25. A. Haldrup, H. Naver, H. V. Scheller, *Plant J.* **17**, 689–698 (1999).
26. M. Hippler, F. Drepper, W. Haehnel, J. D. Rochaix, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 7339–7344 (1998).
27. S. Zhang, H. V. Scheller, *J. Biol. Chem.* **279**, 3180–3187 (2004).
28. P. E. Jensen, A. Haldrup, S. Zhang, H. V. Scheller, *J. Biol. Chem.* **279**, 24212–24217 (2004).
29. M. Pietrzykowska *et al*., *Plant Cell* **26**, 3646–3660 (2014).
30. A. Crepin, S. Caffarri, *Biochim. Biophys. Acta* **1847**, 1539–1548 (2015).
31. P. Longoni, D. Douchi, F. Cariti, G. Fucile, M. Goldschmidt-Clermont, *Plant Physiol.* **169**, 2874–2883 (2015).
32. C. Lunde, P. E. Jensen, A. Haldrup, J. Knoetzel, H. V. Scheller, *Nature* **408**, 613–615 (2000).
33. V. I. Novoderezhkin, M. A. Palacios, H. van Amerongen, R. van Grondelle, *J. Phys. Chem. B* **109**, 10493–10504 (2005).
34. M. Bressan *et al*., *Nat. Plants* **2**, 16131 (2016).
35. E. Wientjes, H. van Amerongen, R. Croce, *Biochim. Biophys. Acta* **1827**, 420–426 (2013).

ACKNOWLEDGMENTS

We thank J. Zhang and X. Zhao for assistance in preparing thylakoid samples. Cryo-EM data collection was carried out at the Center for Biological Imaging, Core Facilities for Protein Science, at the Institute of Biophysics (IBP), Chinese Academy of Sciences (CAS). We thank X. Huang, G. Ji, Z. Guo, B. Zhu, D. Fan, T. Niu, F. Sun, and other staff members at the Center for Biological Imaging (IBP, CAS) for their support in data collection; L. Niu, X. Ding, F. Yang for mass spectrometry; J. Li for assistance in spectral measurement; and L. Kong for cryo-EM data storage and backup. **Funding:** The project was funded by the National Key R&D Program of China (2017YFA0503702, 2017YFA0504700, 2016YFA0502900), the Strategic Priority Research Program of CAS (XDB08020302, XDB08030204), the Key Research Program of Frontier Sciences of CAS (QYZDB-SSW-SMC005), and National Natural Science Foundation of China (31770778, 31700649, 31600609, 31570724). Z.L. and X.Z. received scholarships from the "National Thousand (Young) Talents Program" from the Office of Global Experts Recruitment in China. **Author contributions:** W.C., Z.L., M.L., and X.Z. conceived the project; X.S. performed the sample preparation, characterization, and data collection; J.M. and X.Z. processed the cryo-EM data and reconstructed the cryo-EM map; X.P. built and refined the structure model; P.C. assisted in the structural refinement; X.P., Z.L., X.Z., and M.L. analyzed the structure; M.L., Z.L., and X.Z. wrote the manuscript; all authors discussed and commented on the results and the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The atomic coordinate of the PSI-LHCI-LHCII supercomplex has been deposited in the Protein Data Bank with accession code 5ZJI. The cryo-EM map of the supercomplex has been deposited in the Electron Microscopy Data Bank with accession code EMD-6932.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1109/suppl/DC1
Materials and Methods
Figs. S1 to S10
Tables S1 to S4
References (*36*–*49*)

28 January 2018; accepted 16 April 2018
10.1126/science.aat1156

APPLIED OPTICS

A silicon Brillouin laser

Nils T. Otterstrom[1]*, Ryan O. Behunin[1,2], Eric A. Kittlaus[1], Zheng Wang[3], Peter T. Rakich[1]*

Brillouin laser oscillators offer powerful and flexible dynamics as the basis for mode-locked lasers, microwave oscillators, and optical gyroscopes in a variety of optical systems. However, Brillouin interactions are markedly weak in conventional silicon photonic waveguides, stifling progress toward silicon-based Brillouin lasers. The recent advent of hybrid photonic-phononic waveguides has revealed Brillouin interactions to be one of the strongest and most tailorable nonlinearities in silicon. In this study, we have harnessed these engineered nonlinearities to demonstrate Brillouin lasing in silicon. Moreover, we show that this silicon-based Brillouin laser enters a regime of dynamics in which optical self-oscillation produces phonon linewidth narrowing. Our results provide a platform to develop a range of applications for monolithic integration within silicon photonic circuits.

With the ability to control the optical and electronic properties of silicon, the field of silicon photonics has produced a variety of chip-scale optical devices (*1*–*3*) for applications ranging from high-bandwidth communications (*4*) to biosensing on a chip (*5*). The rapid proliferation of these technologies has spurred interest in strategies to reshape the spectral and coherence properties of light for a wide array of on-chip functionalities. One promising approach to customize on-chip light involves using the nonlinear optical properties of silicon to create optical laser oscillators (*3*). For example, Raman nonlinearities have been harnessed to create all-silicon Raman lasers (*6*, *7*). Brillouin interactions, produced by the coupling between light and sound, could offer a complementary set of behaviors and capabilities for laser technologies in silicon. By exploiting these nonlinearities in a variety of physical systems, Brillouin lasers have been designed to yield everything from frequency-tunable laser emission (*8*) and mode-locked pulsed lasers (*9*) to low-noise oscillators and optical gyroscopes (*10*–*13*).

Within an optical cavity, Brillouin lasing occurs when optical gain from stimulated Brillouin scattering (SBS) overcomes round-trip loss. This nonlinear light-sound coupling is typically strong, overtaking Kerr and Raman interactions in most transparent media. However, the same integrated silicon waveguides that enhance Raman and Kerr nonlinearities tend to produce minuscule Brillouin couplings because of substrate-induced acoustic dissipation (*14*). The recent advent of a class of suspended waveguides—which tightly confine both light and sound—has enabled appreciable nonlinearities through forward SBS (*14*–*17*). Though these suspended structures have produced large optical Brillouin gain (*15*, *16*) and net amplification (*17*, *18*), innovative strategies are needed to translate Brillouin interactions into silicon laser oscillators (*19*, *20*).

We demonstrate a Brillouin laser in silicon by leveraging a form of guided-wave forward Brillouin scattering, termed stimulated intermodal Brillouin scattering, which couples light fields guided in distinct optical spatial modes (*19*, *21*, *22*). Our silicon Brillouin laser system is fabricated from a single-crystal silicon-on-insulator wafer (supplementary materials section 6.2) (*23*). The laser is composed of a 4.6-cm-long racetrack resonator cavity with two extended Brillouin-active gain regions (Fig. 1A). Throughout the device, light is guided by total internal reflection with a ridge waveguide (Fig. 1E, i). This multimode waveguide provides low-loss guidance of both symmetric (red) and antisymmetric (blue) transverse electric (TE)–like spatial modes [with respective propagation constants $k_1(\omega)$ and $k_2(\omega)$, where ω is frequency], yielding two distinct sets of high–quality factor (Q) cavity modes with slightly different free spectral ranges (FSRs) ($Q_1 \cong 2.4 \times 10^6$, $FSR_1 \cong 1.614$ GHz, and $Q_2 \cong 4 \times 10^5$, $FSR_2 \cong 1.570$ GHz, respectively) (supplementary materials 3.7) (*23*). Simulated electric field profiles for these two optical spatial modes are shown in Fig. 1E, iii and iv. To access the cavity modes, we used a directional coupler that couples strongly to the antisymmetric mode and weakly to the symmetric mode, yielding a characteristic multimode transmission spectrum (Fig. 1B).

Optical gain is supplied by forward intermodal Brillouin scattering within the Brillouin-active segments (dark gray). These regions are created

[1]Department of Applied Physics, Yale University, New Haven, CT 06520, USA. [2]Department of Physics and Astronomy, Northern Arizona University, Flagstaff, AZ 86001, USA. [3]Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78758, USA.
*Corresponding author. Email: nils.otterstrom@yale.edu (N.T.O.); peter.rakich@yale.edu (P.T.R.)

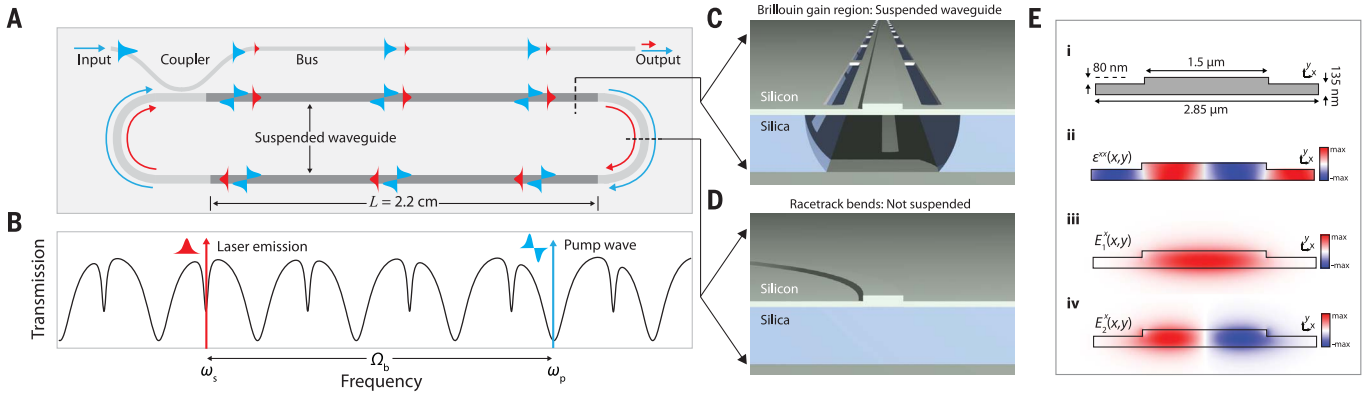


Fig. 1. Schematic of laser cavity and basic operation. (**A**) The Brillouin laser consists of a multimode racetrack cavity with two Brillouin-active regions (dark gray). Pump light (blue) is coupled into the antisymmetric spatial mode of the racetrack resonator. Intermodal Brillouin scattering mediates energy transfer from the pump wave (antisymmetric) to the Stokes wave (symmetric; red). L, length. (**B**) The idealized transmission spectrum for the racetrack cavity. Narrower and broader resonant features correspond to the symmetric and antisymmetric resonances, respectively. Brillouin lasing occurs when the resonance conditions for the pump (antisymmetric) and Stokes (symmetric) waves are simultaneously satisfied. (**C** and **D**) Cross sections of the suspended Brillouin-active region and the racetrack bend, respectively. (**E**) (i) Dimensions of the Brillouin-active waveguide. (ii) Strain profile $\varepsilon^{xx}(x, y)$ of the 6-GHz Lamb-like acoustic mode that mediates intermodal scattering. (iii and iv) x-directed electric field profiles (E_x) of the TE-like symmetric and antisymmetric optical modes, respectively. Red and blue represent the respective positive and negative values of the electric field and strain profiles.

by removing the oxide undercladding to yield a continuously suspended waveguide that produces large intermodal Brillouin gain (Fig. 1C). In addition to low-loss optical modes (Fig. 1E, iii and iv), this structure also supports guidance of a 6-GHz elastic wave (Fig. 1E, ii), which mediates efficient Brillouin coupling between symmetric and antisymmetric optical modes. By contrast, the fixed waveguide bends do not permit phononic guidance. The Brillouin-active waveguide structure is identical in design to that described in (*19*), which yields a peak intermodal Brillouin gain coefficient of $G_b \cong 470\ W^{-1}\ m^{-1}$ at a Brillouin frequency (Ω_b) of 6.03 GHz with a resonance bandwidth of 13 MHz [full width at half maximum (FWHM)]. For efficient nonlinear coupling, this scattering process requires that both energy conservation ($\omega_p = \omega_s + \Omega_b$) and phase-matching [$k_2(\omega_p) = k_1(\omega_s) + q(\Omega_b)$] conditions be satisfied. Here, ω_p and ω_s are the respective pump and Stokes frequencies, and $q(\Omega)$ is the wave vector of the acoustic wave. In intermodal Brillouin scattering, these conditions produce a form of phase-matched symmetry breaking that decouples the Stokes process from the anti-Stokes process, permitting single-sideband amplification (*19*).

Laser oscillation of the symmetric cavity mode occurs when Brillouin gain matches the round-trip loss, producing coherent laser emission at the Stokes frequency ω_s. These lasing requirements are met by injecting pump light (of power P_p) into an antisymmetric cavity mode that is separated in frequency from a symmetric cavity mode by the Brillouin frequency (Fig. 1B). Because the FSRs of the two sets of cavity modes differ by 3.1%, this resonance frequency condition is satisfied by symmetric and antisymmetric cavity mode pairs that occur frequently (every 0.40 nm) across the C band (from 1530 to 1565 nm). When this dual-resonance condition is satisfied and the pump power exceeds the threshold power ($P_p > P_{th}$), the Stokes field builds from thermal noise (produced by spontaneous Brillouin scattering) to yield appreciable line narrowing and coherent Stokes emission at frequency $\omega_s = \omega_p - \Omega_b$.

Many properties of this system could prove advantageous for scalable and robust integration of Brillouin lasers in complex silicon photonic circuits. Because this laser uses a forward scattering process, it alleviates the need for on-chip isolator and circulator technologies that would otherwise be necessary to integrate traditional Brillouin lasers (which use backward SBS). In addition, as this Brillouin nonlinearity is created through structural control, it is possible to independently engineer a range of characteristics, including Brillouin frequency, acoustic dissipation rate, and Brillouin gain, providing a flexible and robust laser design space. Moreover, the multimode properties of this system eliminate size constraints that are present in backward Brillouin lasers (FSRs that must correspond to Brillouin frequencies) and provide exceptional control over cascading dynamics (supplementary materials 5.4).

We investigated Brillouin lasing by injecting continuous-wave pump light into an antisymmetric cavity mode while analyzing the emission of Stokes light from a symmetric cavity mode. We characterized the power and coherence properties of the emitted laser light through high-resolution heterodyne spectral analysis (Fig. 2A) (*23*). The threshold and slope efficiency of this laser were quantified by measuring the total emitted Stokes power as a function of pump power (Fig. 2B). These data reveal a threshold on-chip pump power of 10.6 mW, corresponding to an intracavity power of 19 mW. This laser threshold agrees well with the condition for net amplification in Brillouin waveguides of this design (*19*). Further analysis of these data reveal an on-chip slope efficiency of 3% (supplementary materials 1.3.2 and 4.1) (*23*).

As the pump power increases, the emitted Stokes light exhibits spectral compression characteristic of laser oscillation. When the emitted Stokes spectrum is broader than the linewidth of the optical local oscillator (~13 kHz, derived from the same source as the pump wave), the heterodyne microwave spectrum provides an excellent representation of the emitted Stokes linewidth. Figure 2C compares the Stokes spectrum emitted by the laser (red) with the spontaneous Stokes spectrum emitted from an identical Brillouin-active waveguide segment (gray) in the absence of optical feedback. We see that optical feedback produces spectral compression by a factor of ~10^3; the relatively broad spontaneous Stokes spectrum (FWHM ≅ 13.1 MHz) is compressed to a resolution-limited value of 20 kHz.

We used heterodyne spectral analysis to measure the emitted Stokes linewidth below the laser threshold at various Stokes powers (Fig. 3A and red points in Fig. 3C). A complementary subcoherence self-heterodyne technique characterizes the laser coherence at higher powers (Fig. 3, B and C, and supplementary materials 2) (*23*). Above the threshold, the Stokes wave becomes exceptionally coherent with the incident pump field, with an excess phase noise linewidth ($\Delta\nu_b$) of less than 800 Hz (corresponding to a compression factor of 10^4). Because of the three-wave dynamics in this system, this phase noise corresponds directly to the phonon linewidth, revealing phonon linewidth narrowing far below that of the incident pump field. This behavior

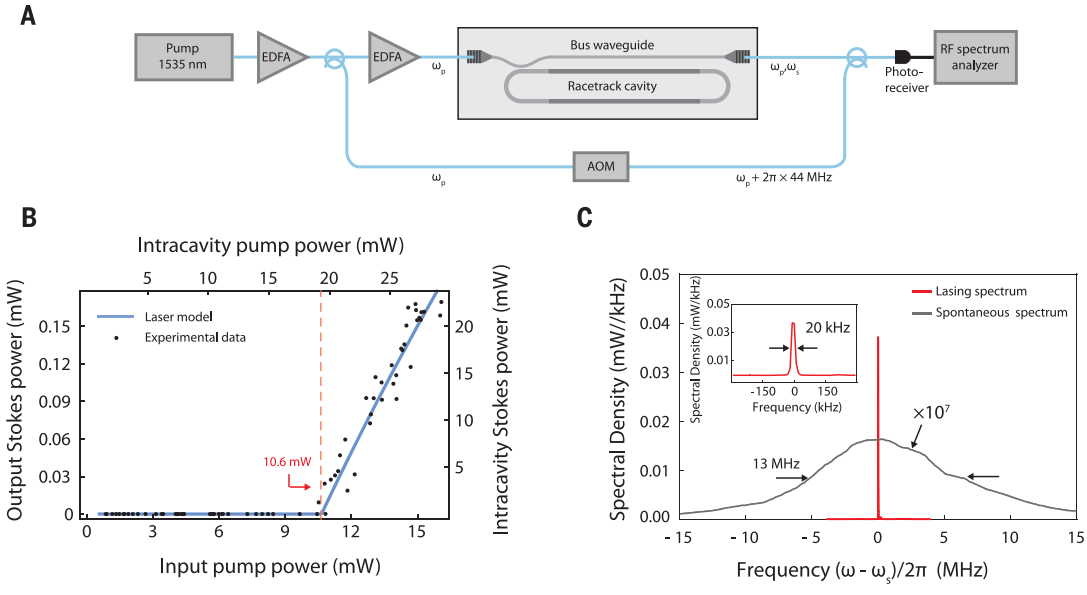


Fig. 2. Experimental apparatus and laser threshold behavior. (**A**) Apparatus used for heterodyne spectroscopy. Continuous-wave pump light (Agilent 81600B, linewidth = 13 kHz) used to initiate Brillouin lasing is amplified by an erbium-doped fiber amplifier (EDFA) and coupled on-chip by grating couplers. Laser light is frequency shifted (+44 MHz) by an acousto-optic modulator (AOM) in a reference arm and combined with the output Stokes light for heterodyne detection. RF, radio frequency. (**B**) Theory and experiment for the output laser power versus the input pump power. Intracavity pump powers are estimated by using the transmitted pump power and the detuning from resonance, and intracavity Stokes power is determined from the measured bus Stokes power and comparison with the theoretical model (supplementary materials 4.1) (*23*). (**C**) Heterodyne spectra of spontaneously scattered Stokes light from a linear waveguide (multiplied by 10^7) and the linewidth-narrowed intracavity laser spectrum above the laser threshold.

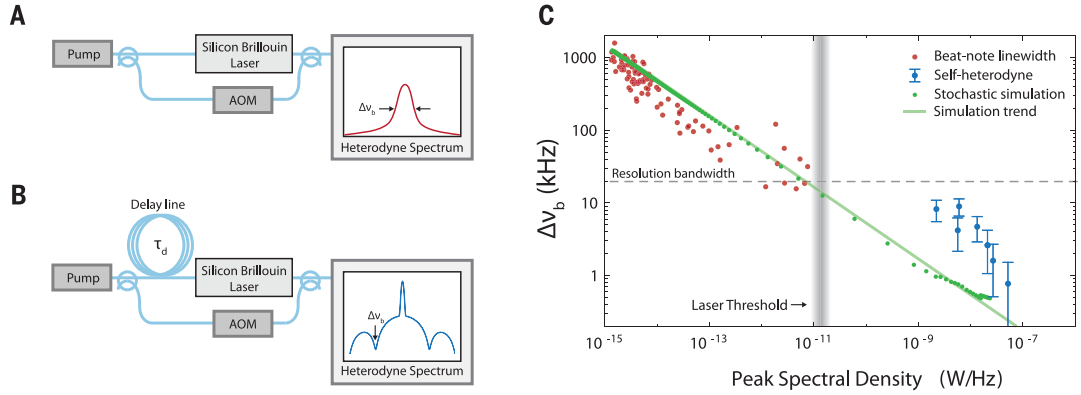


Fig. 3. Linewidth measurements. (**A**) Standard heterodyne spectroscopy apparatus to measure the pump-Stokes excess phase noise (or phonon) linewidth ($\Delta\nu_b$), given by the FWHM of the heterodyne spectrum. (**B**) Subcoherence self-heterodyne apparatus used to probe the phonon dynamics at higher-output Stokes powers (supplementary materials 2) (*23*). The phonon linewidth is determined by measuring the fringe contrast or coherence between output Stokes and pump waves (C). τ_d, delay line transit time. (**C**) Experimental and theoretical comparison of phonon linewidths ($\Delta\nu_b$) as a function of peak spectral density. Below the threshold, we use standard heterodyne spectroscopy (A) (red data points). At higher powers, this measurement becomes resolution bandwidth limited (supplementary materials 4.2) (*23*). For this reason, we use the subcoherence self-heterodyne technique (B), yielding the blue data points (error bars represent the 95% confidence interval of fits to data) (supplementary materials 4.2) (*23*).

represents a marked departure from the linewidth-narrowing dynamics conventionally exhibited by Brillouin lasers.

To understand our experimental observations, we derived simple analytical and numerical models that describe the basic spatial and temporal behavior of laser oscillation in this system (supplementary materials 2) (*23*). Steady-state analysis of the coupled envelope equations reveals that this silicon laser exhibits spatial dynamics (field evolution along the direction of propagation) that are characteristic of Brillouin lasers. Specifically, because the phonon field is spatially heavily damped and the only feedback mechanism is optical, this laser produces optical self-oscillation of the Stokes wave (supplementary materials 1.1) (*23*). Building on established treatments of Brillouin laser physics (*24*, *25*), we developed a simplified mean-field model to explore the salient features of the temporal dynamics (supplementary materials 1.2 to 1.5). This model incorporates parameters that are consistent with the measured resonator and nonlinear waveguide characteristics (supplementary materials 3.5 and 3.8) (*23*).

Well above the laser threshold, this model predicts Stokes emission that is highly coherent with the incident pump field, with an excess phase noise linewidth given by

$$\Delta\nu_b = \frac{\Gamma}{4\pi\beta^2}\left(n_b^{th} + n_s^{th} + 1\right) \quad (1)$$

Here, Γ is the intrinsic acoustic dissipation rate, β^2 is the coherently driven phonon occupation number, n_b^{th} is the thermal occupation number of the phonon field, n_s^{th} is the average thermal occupation number of the symmetric mode of the optical resonator ($n_b^{th} \approx 10^3$ and $n_s^{th} \approx 0$), and the +1 is due to vacuum fluctuations. As a result of the three-wave dynamics of this system, the pump-Stokes coherence provides a direct window into the spectrum of the distributed acoustic field (supplementary materials 1.3) (*23*), revealing that this regime of Brillouin lasing produces Schawlow-Townes linewidth narrowing of the coherent acoustic field. Although closed-form analytical expressions for phase noise are tractable well below and above the threshold, stochastic numerical simulations are necessary to model the noise characteristics in the vicinity of the laser threshold (Fig. 3C), revealing good qualitative agreement with our measurements.

These linewidth-narrowing dynamics are distinct from those typically produced in glass-based Brillouin lasers (*10*, *24*), which yield Schawlow-Townes optical linewidth narrowing (supplementary materials 1.6 and 1.7) (*23*). In this silicon system, phonon linewidth narrowing arises from an inverted dissipation hierarchy in which the phonon temporal dissipation rate is much lower than the optical dissipation rates for the pump and Stokes cavity modes ($\gamma_p, \gamma_s \gg \Gamma$), in contrast to the temporal dissipation hierarchy conventionally realized in Brillouin lasers ($\gamma_p, \gamma_s \ll \Gamma$) (*25*). As a result, this silicon Brillouin laser simultaneously operates where the spatial acoustic decay length (63 μm) is far smaller than the optical decay length (~0.1 to 1 m), whereas the intrinsic phonon lifetime (77 ns) exceeds that of the optical fields (~2 to 12 ns). This combination of spatial and temporal dynamics is made possible by the unusually large Brillouin coupling provided by this silicon system (~10^3 times that of silica fibers) and the disparate velocities of the interacting light and sound waves.

The observed phonon coherence is reminiscent of that produced in optomechanical self-oscillation (*22*, *26*, *27*) (phonon lasing). However, in contrast to phonon lasers, this Brillouin laser does not possess a phonon cavity that permits acoustic feedback necessary for phonon self-oscillation (supplementary materials 1.7.3 and 1.7.4) (*23*). Here we show that, despite a high degree of acoustic spatial damping and lack of phonon feedback (more than 1000 dB round-trip acoustic propagation loss), optical self-oscillation of the Stokes wave produces linewidth narrowing of the acoustic field, as long as the temporal acoustic dissipation rate is lower than that of the optical fields. In this way, this system is analogous to an extreme limit of singly resonant optical parametric oscillator physics, with a slow, ballistic, and long-lived idler wave (supplementary materials 1.7) (*23*). Other Brillouin laser systems may have operated near or in this temporal dissipation hierarchy (*28*–*30*); however, these dynamics were not identified. It is because of the stability of this monolithic silicon system that we are able to study this unusual combination of spatial and temporal dynamics.

REFERENCES AND NOTES

1. G. T. Reed, G. Mashanovich, F. Y. Gardes, D. J. Thomson, *Nat. Photonics* **4**, 518–526 (2010).
2. D. Liang, J. E. Bowers, *Nat. Photonics* **4**, 511–517 (2010).
3. J. Leuthold, C. Koos, W. Freude, *Nat. Photonics* **4**, 535–544 (2010).
4. A. Alduino, M. Paniccia, *Nat. Photonics* **1**, 153–155 (2007).
5. K. De Vos, I. Bartolozzi, E. Schacht, P. Bienstman, R. Baets, *Opt. Express* **15**, 7610–7615 (2007).
6. O. Boyraz, B. Jalali, *Opt. Express* **12**, 5269–5273 (2004).
7. H. Rong *et al.*, *Nature* **433**, 725–728 (2005).
8. H. Takuma, D. Jennings, *Appl. Phys. Lett.* **5**, 239–241 (1964).
9. V. Lecoeuche, D. J. Webb, C. N. Pannell, D. A. Jackson, *Opt. Commun.* **152**, 263–268 (1998).
10. S. P. Smith, F. Zarinetchi, S. Ezekiel, *Opt. Lett.* **16**, 393–395 (1991).
11. W. Loh *et al.*, *Optica* **2**, 225 (2015).
12. I. V. Kabakova *et al.*, *Opt. Lett.* **38**, 3208–3211 (2013).
13. J. Li, M.-G. Suh, K. Vahala, *Optica* **4**, 346 (2017).
14. P. T. Rakich, C. Reinke, R. Camacho, P. Davids, Z. Wang, *Phys. Rev. X* **2**, 011008 (2012).
15. H. Shin *et al.*, *Nat. Commun.* **4**, 1944 (2013).
16. R. Van Laer, B. Kuyken, D. Van Thourhout, R. Baets, *Nat. Photonics* **9**, 199–203 (2015).
17. E. A. Kittlaus, H. Shin, P. T. Rakich, *Nat. Photonics* **10**, 463–467 (2016).
18. R. V. Laer, A. Bazin, B. Kuyken, R. Baets, D. V. Thourhout, *New J. Phys.* **17**, 115005 (2015).
19. E. A. Kittlaus, N. T. Otterstrom, P. T. Rakich, *Nat. Commun.* **8**, 15819 (2017).
20. Y. A. Espinel, F. G. Santos, G. O. Luiz, T. P. Alegre, G. S. Wiederhecker, *Sci. Rep.* **7**, 43423 (2017).
21. M. S. Kang, A. Brenn, P. St. J. Russell, *Phys. Rev. Lett.* **105**, 153901 (2010).
22. G. Bahl, J. Zehnpfennig, M. Tomes, T. Carmon, *Nat. Commun.* **2**, 403 (2011).
23. Materials and methods are available as supplementary materials.
24. J. Li, H. Lee, T. Chen, K. J. Vahala, *Opt. Express* **20**, 20170–20180 (2012).

25. A. Debut, S. Randoux, J. Zemmouri, *Phys. Rev. A* **62**, 023803 (2000).
26. I. S. Grudinin, H. Lee, O. Painter, K. J. Vahala, *Phys. Rev. Lett.* **104**, 083901 (2010).
27. W. C. Jiang, X. Lu, J. Zhang, Q. Lin, *Opt. Express* **20**, 15991–15996 (2012).
28. K. Hill, B. Kawasaki, D. Johnson, *Appl. Phys. Lett.* **28**, 608–609 (1976).
29. K. S. Abedin, *Opt. Express* **14**, 4037–4042 (2006).
30. B. Morrison *et al.*, *Optica* **4**, 847 (2017).

ACKNOWLEDGMENTS

We thank A. D. Stone for discussions regarding the laser dynamics and M. Rooks for assistance with the fabrication. **Funding:** This work was supported through a seedling grant under the direction of D. Green at DARPA MTO and by the Packard Fellowship for Science and Engineering. N.T.O. acknowledges support from the NSF graduate research fellowship under grant DGE1122492. Z.W. acknowledges support from the NSF under grant 1641069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of DARPA or the NSF. **Author contributions:** P.T.R., N.T.O., and E.A.K. developed the device concept and design. N.T.O. and E.A.K. fabricated the devices and performed the experiments. R.O.B. and N.T.O. developed theory with assistance from P.T.R. and Z.W. All authors contributed to the preparation of the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1113/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S3
References (*31–42*)

1 December 2017; resubmitted 3 February 2018
Accepted 16 April 2018
10.1126/science.aar6113

SOCIAL SCIENCE

Experimental evidence for tipping points in social convention

Damon Centola[1,2]*, Joshua Becker[1], Devon Brackbill[1], Andrea Baronchelli[3]

Theoretical models of critical mass have shown how minority groups can initiate social change dynamics in the emergence of new social conventions. Here, we study an artificial system of social conventions in which human subjects interact to establish a new coordination equilibrium. The findings provide direct empirical demonstration of the existence of a tipping point in the dynamics of changing social conventions. When minority groups reached the critical mass—that is, the critical group size for initiating social change—they were consistently able to overturn the established behavior. The size of the required critical mass is expected to vary based on theoretically identifiable features of a social setting. Our results show that the theoretically predicted dynamics of critical mass do in fact emerge as expected within an empirical system of social coordination.

Observational accounts of rapid changes in social conventions have suggested that apparently stable societal norms can be effectively overturned by the efforts of small but committed minorities (*1*–*3*). From social expectations about gender roles in the workplace (*4*) to the popular acceptance of (or intolerance toward) tobacco use and marijuana use (*5*), accounts of changing social conventions have hypothesized that minority groups can trigger a shift in the conventions held by the majority of the population (*1*–*3*, *5*, *6*). Although this hypothesis presents a striking contrast to the expectations of classical equilibrium stability analysis from economic theory (*7*, *8*), it can nevertheless be well explained by the theory of critical mass as posited by evolutionary game theory (*9*–*11*). This theory argues that when a committed minority reaches a critical group size—commonly referred to as a "critical mass"—the social system crosses a tipping point. Once the tipping point is reached, the actions of a minority group trigger a cascade of behavior change that rapidly increases the acceptance of a minority view (*12*–*14*).

The simplest formulation of critical mass theory maintains that small groups of regular individuals—that is, with the same amount of social power and resources as everyone else—can successfully initiate a change in social conventions. According to this view, the power of small groups comes not from their authority or wealth but from their commitment to the cause (*14*, *15*).

Thus far, evidence for critical mass dynamics in changing social conventions has been limited to formal theoretical models and observations from qualitative studies. These studies have proposed a wide range of possible thresholds for the size of an effective critical mass, ranging from 10% of the population up to 40%. For instance, theoretical simulations of linguistic conventions have argued that a critical mass composed of 10% of the population is sufficient to overturn an established social equilibrium (*14*). By contrast, qualitative studies of gender conventions in corporate leadership roles have hypothesized that tipping points are only likely to emerge when a critical mass of 30% of the population is reached (*3*, *16*). Related observational work on gender conventions (*17*) has built on this line of research, speculating that effective critical mass sizes are likely to be even higher, approaching 40% of the population. Despite the broad practical (*18*, *19*) and scientific (*1*, *12*) importance of understanding the dynamics of critical mass in collective behavior, it has not been possible to identify whether there are in fact tipping points in empirical systems because such a test requires the ability to independently vary the size of minority groups within an evolving system of social coordination.

We addressed this problem by adopting an experimental approach to studying tipping-point dynamics within an artificially created system of evolving social conventions. Following the literature on social conventions (*9*, *20*, *21*), we study a system of coordination in which a minority group of actors attempt to disrupt an established equilibrium behavior. In both our theoretical framework and the empirical setting, we adopt the canonical approach of using coordination on a naming convention as a general model for conventional behavior (*21*–*24*). Our experimental approach is designed to test a broad range of theoretical predictions derived from the existing literature on critical mass dynamics in social conventions.

We first synthesized these diverse theoretical and observational accounts of tipping-point dynamics to derive theoretical predictions for the size of an effective critical mass (*25*). Based on earlier theoretical (*9*, *26*) and qualitative studies of social convention (*20*, *23*), we propose a simple model of strategic choice in which actors decide which social conventions to follow by choosing the option that yields the greatest expected individual reward given their history of social interactions (*9*). In this individual learning model, people coordinate with their peers so long as they benefit individually from coordinating. The

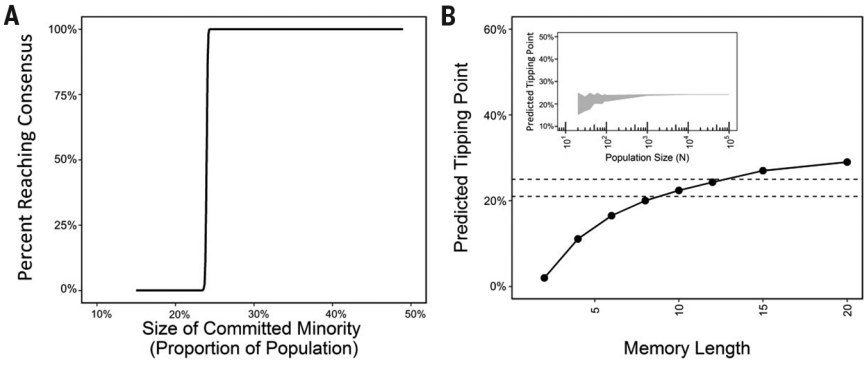


Fig. 1. Predicted tipping points in social stability. (**A**) Theoretical modeling of the proportion of outcomes in which the alternative behavior is adopted by 100% of the population. In this system, the number of agents (N) = 1000, the number of interactions (T) = 1000, the number of past interactions used in agent decisions (M) = 12. (**B**) The size of the predicted critical mass point is shown as a function of individuals' average memory length, M, where (N = 1000, T = 1000). The dashed lines indicate the range enclosed by our experimental trials, showing the largest unsuccessful minority (21%) and the smallest successful minority (25%). Although the expected size of the critical mass point increases with M, this relationship is concave, allowing the predicted tipping point to remain well below 50% as M gets large (>100). (Inset) Effect of increasing population size on the precision of the size of the committed minority (C) prediction (M = 12, T = 1000). For N < 1000, small variations in the predicted tipping point emerge due to stochastic variations in individual behavior. Shaded region indicates C sizes where success was observed frequently but without certainty. Above this region, for larger C sizes, the probability of success reaches 1; for C sizes below this region, the likelihood of success goes to 0.

[1]Annenberg School for Communication, University of Pennsylvania, Philadelphia, PA, USA. [2]School of Engineering, University of Pennsylvania, Philadelphia, PA, USA. [3]Department of Mathematics, City, University of London, London, UK.
*Corresponding author. Email: dcentola@asc.upenn.edu

model predicts a sharp transition in the collective dynamics of social convention as the size of the committed minority reaches a critical fraction of the population (Fig. 1). When the size of the committed minority is below this predicted tipping point, the dominant social convention is expected to remain stable, whereas above this size it is expected to change (*25*).

Our theoretical predictions for the size of the critical mass were determined by two parameters: individual memory length (M) and population size (N). Explorations of these parameters (Fig. 1) show that the predicted size of the tipping point changes with individuals' expected memory length (M). When participants have shorter memories ($M < 5$ interactions), the size of the critical mass is smaller. Even under the assumption that people have very long memories ($M >$ 100 interactions), the predicted critical mass size remains well below 50% of the population (*25*), indicating that critical mass dynamics may be possible even in systems with long histories. Variations in population size were explored computationally in the range $20 < N < 100{,}000$ and were not found to significantly affect the predicted critical mass size (*25*). Figure 1 shows that for populations in the range $20 < N < 1000$, stochastic fluctuations introduce a small uncertainty into the estimate of the critical mass size. However, for population sizes $N > 1000$, the predicted tipping point for social change is constant and independent of N [complete details in (*25*)].

We recruited 194 subjects from the World Wide Web and placed them into online communities where they participated in a social coordination process (*27*, *28*). Upon arrival to the study, participants were randomly assigned to participate in one of 10 independent online groups, which varied in size from 20 to 30 people. In a given round of the study, the members of each group were matched at random into pairs to interact with one another. Within each pair, both subjects simultaneously assigned names to a pictured object (i.e., a face), attempting to coordinate in the real-time exchange of linguistic alternatives (*20*, *25*). If the players entered the same name (i.e., coordinated), they were rewarded with a successful payment; if they entered different names (i.e., failed to coordinate), they were penalized. In each community, individuals interacted with each other over repeated rounds of randomly assigned pairings, with the goal of coordinating with one another (*25*). Participants were not incentivized to reach a "global" consensus but only to coordinate in a pairwise manner with their partner on each round. Participants were financially rewarded for coordinating and financially punished each time they failed to coordinate with each other (*25*). Once a convention was established for the entire population, the incentives strongly favored coordinating on the equilibrium behavior.

After each round, the participants could see only the choices that they and their partner had made, and their cumulative pay was updated accordingly. They were then randomly assigned to interact with a new member of their group, and a new round would begin. These dynamics reflect common types of online exchanges, in which community members directly interact with the other members of a large, often anonymous population—using, for instance, chat interfaces or messaging technologies—leading them to adopt linguistic and behavioral conventions that allow them to effectively coordinate their actions with other participants' expectations (*20*, *29 30*). Consistent with these types of settings, participants in the study did not have any information about the size of the population that was attempting to coordinate nor about the number of individuals to whom they were connected

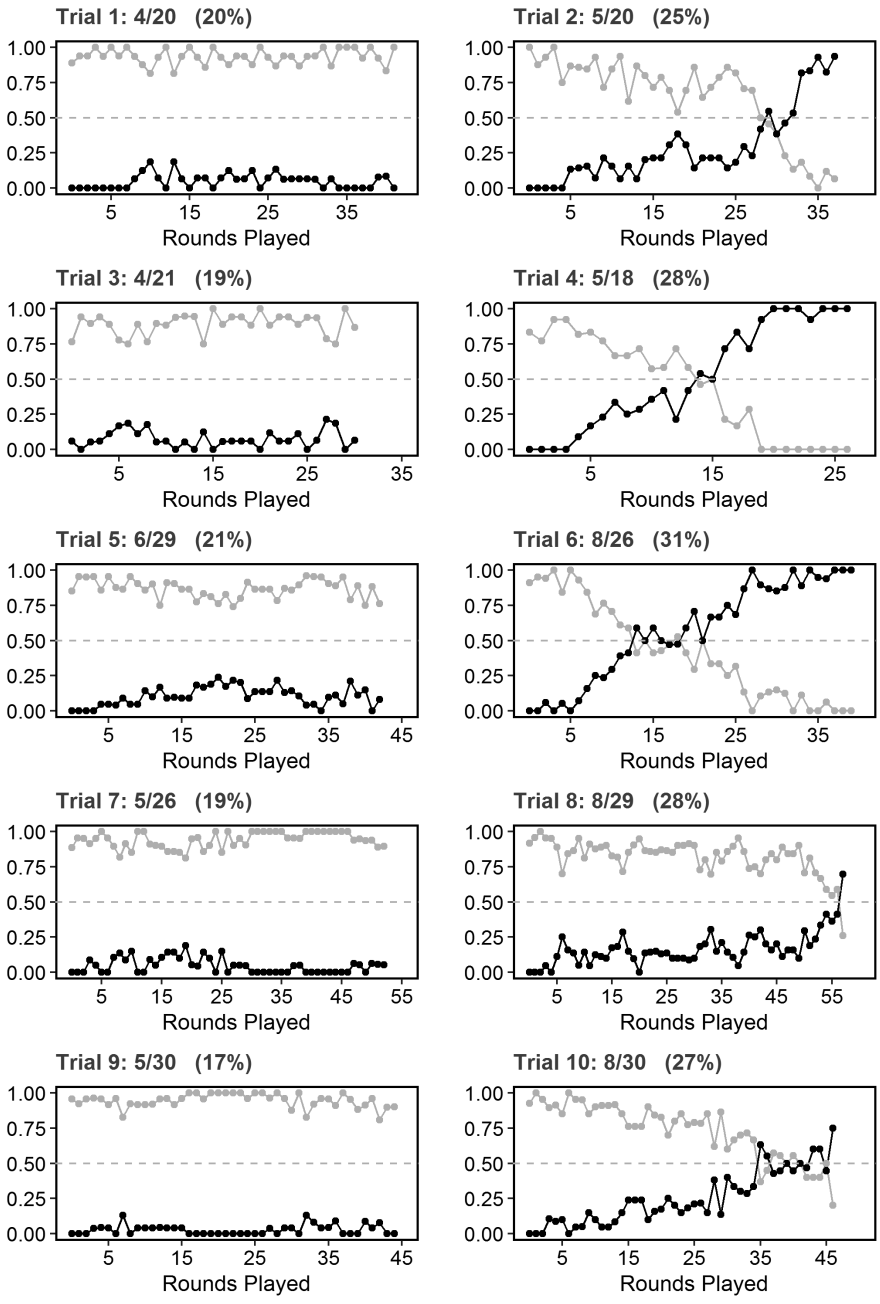


Fig. 2. Time series showing adoption of the alternative convention by noncommitted subjects (i.e., experimental subjects). Gray lines indicate the popularity of the established convention; black lines show the adoption of the alternative convention. Success was achieved when more than 50% of the noncommitted population adopted the new social convention. Trials in the left column show failed mobilization, whereas trials in the right column show successful mobilization. A transition in the collective dynamics happens when C reaches ~25% of the population. Each round is measured as $N/2$ pairwise interactions, such that each player has one interaction per round on average.

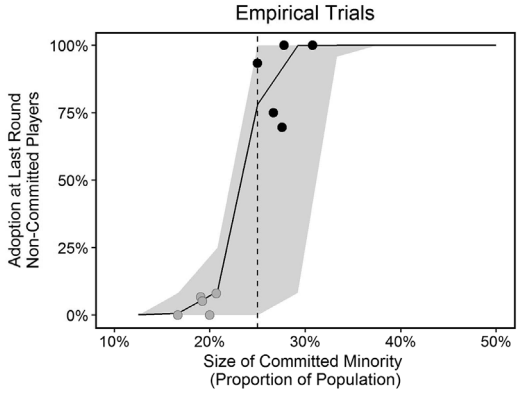


Fig. 3. Final success levels from all trials (gray points indicate trials with $C < 25\%$; black points indicate trials with $C \geq 25\%$. Also shown is the theoretically predicted critical mass point (solid line) with 95% confidence intervals ($N = 24$, $T = 45$, $M = 12$; gray area indicates 95% confidence intervals from 1000 replications). The dotted line indicates $C = 25\%$. The theoretical model of critical mass provides a good approximation of the empirical findings. For short time periods ($T < 100$ interactions), the critical mass prediction is not exact (ranging from 20% < C < 30% of the population); however, over longer time periods ($T > 1000$) the transition dynamics become more precise (solid line, *25*).

(*9*, *20*, *23*). In every group, this interaction process quickly led to the establishment of a group-wide social convention, in which all players in the network consistently coordinated on the same naming behavior (*20*, *25*). Once a convention was established among all experimental participants, we introduced a small number of confederates (that is, a "committed minority") into each group, who attempted to overturn the established convention by advancing a novel alternative (*25*).

Trials varied according to the size of the committed minority (C) that attempted to overturn the established convention. In total, we studied the dynamics of critical mass in 10 independent groups, each with a committed minority of a fixed size. Across all 10 groups, the sizes of the committed minorities were in the range (15% < C < 35%).

Figures 2 and 3 report tipping-point dynamics in the collective process of overturning an established equilibrium. Consistent with the expectations of our theoretical model (using empirically parameterized values of N and M), when the size of the committed minority reached ~25% of the population, a tipping point was triggered, and the minority group succeeded in changing the established social convention.

Five trials were conducted. Each trial was composed of two communities—one with the committed minority below the expected critical size ($C < 25\%$) and one with it equal to or above ($C \geq 25\%$). In every trial, the community with $C <$ 25% had only small numbers of converts to the minority view. Over the course of these trials, each of these converts eventually reverted to the dominant norm. Continuous interactions led to occasional switching by subjects throughout the study. However, over all the trials, in the condition where the minority group was smaller than 25% of the population, on average only 6% of the noncommitted population adopted the alternative behavior by the final round of the study.

For each of these unsuccessful trials, we conducted a corresponding trial using another population of the same size but with a larger committed minority ($25\% \leq C \leq 31\%$). In all these groups, the alternative norm reached the majority of the population within the experimental window of observation (Figs. 2 and 3). Over all trials, populations with $C \geq 25\%$ were significantly more likely to overturn the dominant convention than populations with a committed minority below 25% ($P = .01$, Wilcoxon rank sum). We found that in one case (Trial 1) this transition from failure to success was the result of increasing the size of the committed minority by only one person.

Figure 3 shows a summary of final adoption levels across all trials, along with expectations from our empirically parameterized theoretical model, with 95% confidence intervals. Populations with committed minorities ranging from $25\% \leq C \leq 27\%$ achieved uptake levels between 72 and 100% within the empirical observation window. At $C = 31\%$, the committed minority achieved consensus within the window of empirical observation. Figure 3 compares these observations to numerical simulations of the theoretical model using population sizes and observation windows (T rounds of interaction) comparable to the experimental study ($N = 24$, $T = 100$, $M = 12$). Memory length for these simulations was calibrated using subjects' empirical memory lengths in this study based on their observed behavior over all 10 groups. A memory length in the range $9 \leq M \leq 13$ provides a good approximation of subjects' observed behavior, correctly predicting 80% of subjects' choices across all trials (*25*). The theoretically predicted critical mass size from this model fit the experimental findings well (Fig 3). Numerical analyses indicate that with larger population sizes, the critical mass point becomes more exact (See Figs. 1 and 3), approaching 24.3% of the population.

Our experimental results do not show agreement with theoretical predictions from models of social convention that predict low critical mass thresholds, at 10% of the population. However, our findings show good agreement with qualitative studies of gender conventions within organizational settings (*3*), which hypothesized that a critical mass of ~30% could be sufficient to overturn established norms (*16*). Our results may suggest that in organizational contexts—where population boundaries are relatively well defined and there are clear expectations and rewards for social coordination among peers—the process of normative changes in social conventions may be well described by the dynamics of critical mass.

The design choices that aided our control of the study also put constraints on the behaviors that we could test. Our experimental design provided subjects with social and financial incentives that strongly favored coordinating on an established social convention (*25*). However, in the real world, individuals' emotional and psychological commitments to established behaviors can create additional resistance to behavior change (*31*). To further explore these expectations, supplementary analyses of our theoretical model (fig. S7) (*25*) extend our basic predictions to consider how the critical mass size may differ under conditions of greater social entrenchment. When actors are more conservative—exhibiting an explicit bias in favor of the established convention (based on a skewed best-response calculation favoring the equilibrium behavior)—tipping-point dynamics were still predicted to be achievable by committed minorities with only marginally larger group sizes.

In delimiting the scope of our findings, we emphasize that the critical mass value of 25% is not expected to be a universal value for changing social conventions. Our results demonstrate that within an endogenous system of social coordination, tipping-point dynamics emerged consistent with theoretical expectations. Further work is required to determine the applicability of our findings to specific social settings. In particular, alternative empirical parameterizations of our model can result in alternative predictions for the expected size of the critical mass. We expect that the findings from our study can be considerably expanded by future empirical work studying the dynamics of tipping points within other empirical systems of social convention.

For instance, an important setting in which these results might be usefully applied concerns the growing ability of organizations and governments to use confederate actors within online spaces to influence conventional behaviors and beliefs. Recent work on the 50 Cent Party in China (*32*, *33*) has argued that the Chinese government has incentivized small groups of motivated individuals to anonymously infiltrate social media communities such as Weibo with the intention of subtly shifting the tone of the collective dialogue to focus on topics that celebrate national pride and distract from collective grievances (*32*). We anticipate that social media spaces of this kind will be an increasingly important setting for extending the findings of our study to understand the role of committed minorities in shifting social conventions. Similarly, the results from our study may also be usefully applied to the dynamics of critical mass in other online settings, such as changing social expectations regarding (i) the standards of civility on Facebook and other online discussion forums (*19*, *34*), (ii) the acceptability of bullying behavior in adolescent chat groups (*35*), and

(iii) the appropriate kinds of content to share with strangers over social media (*36*), all of which have been suggested to exhibit susceptibility to shifts in conventional behavior as a result of the activity of a small fraction of the population (*19*, *34*, *36*).

REFERENCES AND NOTES

1. T. Kuran, *Am. J. Sociol.* **100**, 1528–1551 (1995).
2. K. D. Opp, C. Gern, *Am. Sociol. Rev.* **58**, 659–680 (1993).
3. R. M. Kanter, *Am. J. Sociol.* **82**, 965–990 (1977).
4. D. Dahlerup, L. Freidenvall, *Int. Fem. J. Polit.* **7**, 26–48 (2005).
5. S. Bikhchandani, D. Hirshleifer, I. Welch, *J. Polit. Econ.* **100**, 992–1026 (1992).
6. T. Kuran, *Private Truths, Public Lies: The Social Consequences of Preference Falsification* (Harvard Univ. Press, 1997).
7. J. C. Harsanyi, R. Selten, *A General Theory of Equilibrium Selection in Games* (MIT Press, Cambridge, 1988).
8. J. F. Nash, *Proc. Natl. Acad. Sci. U.S.A.* **36**, 48–49 (1950).
9. H. P. Young, *Econometrica J. Econom. Soc.* **61**, 57–84 (1993).
10. M. Kandori, G. J. Mailath, R. Rob, *Econometrica J. Econom. Soc.* **61**, 29–56 (1993).
11. G. Ellison, *Econometrica J. Econom. Soc.* **61**, 1047–1071 (1993).
12. T. Schelling, *Micromotives and Macrobehavior* (Norton, New York, 1978).
13. M. Granovetter, *Am. J. Sociol.* **83**, 1420–1443 (1978).
14. J. Xie *et al.*, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **84**, 011130 (2011).
15. D. M. Centola, *Ration. Soc.* **25**, 3–40 (2013).
16. D. Dahlerup, *Scand. Polit. Stud.* **11**, 275–298 (1988).
17. S. Grey, *Polit. Gend.* **2**, 492–502 (2006).
18. G. Marwell, P. Oliver, *The Critical Mass in Collective Action* (Cambridge Univ. Press, 1993).
19. K. Nyborg *et al.*, *Science* **354**, 42–43 (2016).
20. D. Centola, A. Baronchelli, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1989–1994 (2015).
21. D. Lewis, *Convention: A Philosophical Study* (Harvard Univ. Press, Cambridge, 1969).
22. C. Castellano, S. Fortunato, V. Loreto, *Rev. Mod. Phys.* **81**, 591 (2009).
23. S. Garrod, G. Doherty, *Cognition* **53**, 181–215 (1994).
24. L. Wittgenstein, *Philosophical Investigations* (John Wiley & Sons, 2009).
25. Materials and methods are available as supplementary materials.
26. A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, L. Steels, *J. Stat. Mech. Theory Exp.* **2006**, P06014 (2006).
27. D. Centola, *Science* **334**, 1269–1272 (2011).
28. D. Centola, *Science* **329**, 1194–1197 (2010).
29. L. Backstrom, D. Huttenlocher, J. Kleinberg, X. Lan, Group formation in large social networks: membership, growth, and evolution. Proc. 12th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 44–54 (2006).
30. F. Kooti, H. Yang, M. Cha, P. K. Gummadi, W. A. Mason, The Emergence of Conventions in Online Social Networks. Proc. Sixth Int. AAAI Conf. Weblogs Soc. Media ICWSM 2012 (2012).
31. C. Tilly, S. Tarrow, *Contentious Politics* (Oxford Univ. Press, New York, ed. 2, 2015).
32. G. King, J. Pan, M. E. Roberts, *Am. Polit. Sci. Rev.* **111**, 484–501 (2017).
33. G. King, J. Pan, M. E. Roberts, *Science* **345**, 1251722 (2014).
34. A. Antoci, A. Delfino, F. Paglieri, F. Panebianco, F. Sabatini, *PLOS ONE* **11**, e0164286 (2016).
35. E. L. Paluck, H. Shepherd, P. M. Aronow, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 566–571 (2016).
36. C. Shih, *The Facebook Era: Tapping Online Social Networks to Market, Sell, and Innovate* (Pearson Education, 2010).

ACKNOWLEDGMENTS

We gratefully acknowledge research assistance from S. Kim and N. Herbert, helpful suggestions from P. Starr and A. van de Rijt, and programming assistance from A. Wagner and R. Overbey. This research was approved by the Institutional Review Board at the University of Pennsylvania, where the study was conducted, and it included informed consent by all participants in the study. **Funding:** The authors received no external funding in support of this research. **Author contributions:** D.C. and A.B. designed the study. J.B. and D.B. collected the data. D.C., J.B., D.B., and A.B. analyzed the data. D.C. wrote the manuscript. All authors commented on and approved the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The datasets are available upon request and upon publication will be publicly available for download from the university Web site: http://ndg.asc.upenn.edu/experiments/creating-critical-mass. Upon publication, the data will be available at Dataverse, doi:10.7910/DVN/11HUAG. The code in R is available on GitHub at https://github.com/NetworkDynamicsGroup/BestResponseNameGame.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1116/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S7
References (*37*)

3 January 2018; accepted 20 April 2018
10.1126/science.aas8827

TETRAPOD EVOLUTION

A tetrapod fauna from within the Devonian Antarctic Circle

Robert Gess[1]* and Per Erik Ahlberg[2]

Until now, all known fossils of tetrapods (limbed vertebrates with digits) and near-tetrapods (such as *Elpistostege*, *Tiktaalik*, and *Panderichthys*) from the Devonian period have come from localities in tropical to subtropical paleolatitudes. Most are from Laurussia, a continent incorporating Europe, Greenland, and North America, with only one body fossil and one footprint locality from Australia representing the southern supercontinent Gondwana. Here we describe two previously unknown tetrapods from the Late Devonian (late Famennian) Gondwana locality of Waterloo Farm in South Africa, then located within the Antarctic Circle, which demonstrate that Devonian tetrapods were not restricted to warm environments and suggest that they may have been global in distribution.

The fossil locality at Waterloo Farm, near Grahamstown, South Africa (Fig. 1A), features an exceptionally preserved biota, including examples of soft-tissue preservation (*1*–*4*), deposited in the south polar region close to paleolatitude 70°S (Fig. 1B). In contrast, all previously known Devonian tetrapod and elpistostegid localities lie within about 30° of the palaeoequator (*5*). The Waterloo Farm fossils are metamorphosed and strongly flattened, with the bone tissue replaced by secondary metamorphic mica, partially altered to chlorite. Two tetrapods—*Tutusius umlambo* gen. et sp. nov. and *Umzantsia amazana* gen. et sp. nov., both represented by disarticulated material (Figs. 2 and 3 and figs. S1 and S2)—are present in the assemblage and are described here (formal taxonomic descriptions are in the supplementary materials). *Tutusius* is represented by a single cleithrum (Fig. 2, A and B) with a broad, flat, unornamented blade, resembling that of the early Famennian Russian genus *Jakubsonia* more than the slender cleithra of the late Famennian *Ichthyostega* and *Ventastega* (*6*–*8*) (Fig. 4).

Dermal bones of *Umzantsia* carry a distinctive ornament consisting of fine parallel ridges reminiscent of water ripples. This allows identification of a number of cranial bones and a cleithrum from one bedding plane as probably derived from a single individual, designated the holotype (Fig. 2, C to P). A lower jaw ramus from another bedding plane (Fig. 3) is also assigned to *Umzantsia*. Scaling the bones of the *Umzantsia* holotype to the skull reconstruction of *Ventastega* (*8*) suggests a head length of ~13 cm. The lower jaw is 17.9 cm long. The dermal ornament also covers much of the cleithrum of *Umzantsia*; this fishlike characteristic contrasts with the unornamented cleithra of other Devonian tetrapods, suggesting a phylogenetic position between those tetrapods and *Tiktaalik* (*9*) (Fig. 4).

The largest skull bone is a jugal (Fig. 2, E to G), identifiable from the presence of an orbital margin and characteristic set of sutural margins (*5*, *7*, *8*). The orbital margin is short [suggesting a triangular orbit with a ventral apex, similar in shape to that of *Anthracosaurus* (*10*), unless the eye was extremely small], the lacrimal is excluded from the orbital margin by a jugal-prefrontal contact, there is no distinct dorsal postorbital process, and the notch for the quadratojugal is deep. The preopercular (Fig. 2, H and I) is similar to that of *Ventastega*, with a rounded posterior margin that projects as a short process beyond the quadratojugal contact. The frontal resembles those of previously described Devonian tetrapods. A probable supratemporal is the only recovered skull table element.

The lower jaw is slender and gently curved. The splenial is the longest of the infradentaries, occupying about half the jaw length. The left lower jaw ramus has five infradentaries, instead of the normal four. This may be an autapomorphy of *Umzantsia*, but the associated infradentaries from the right ramus appear proportionately longer, raising the possibility that there were only four infradentaries on the right side and that this individual was asymmetrical. The infradentaries carry the typical tetrapod "starburst" ornament, grading dorsally into a ripple-like ornament. A series of short tooth-bearing

[1]Albany Museum and Geology Department, Rhodes University, Grahamstown, South Africa. [2]Department of Organismal Biology, Uppsala University, Uppsala, Sweden.
*Corresponding author. Email: robg@imaginet.co.za

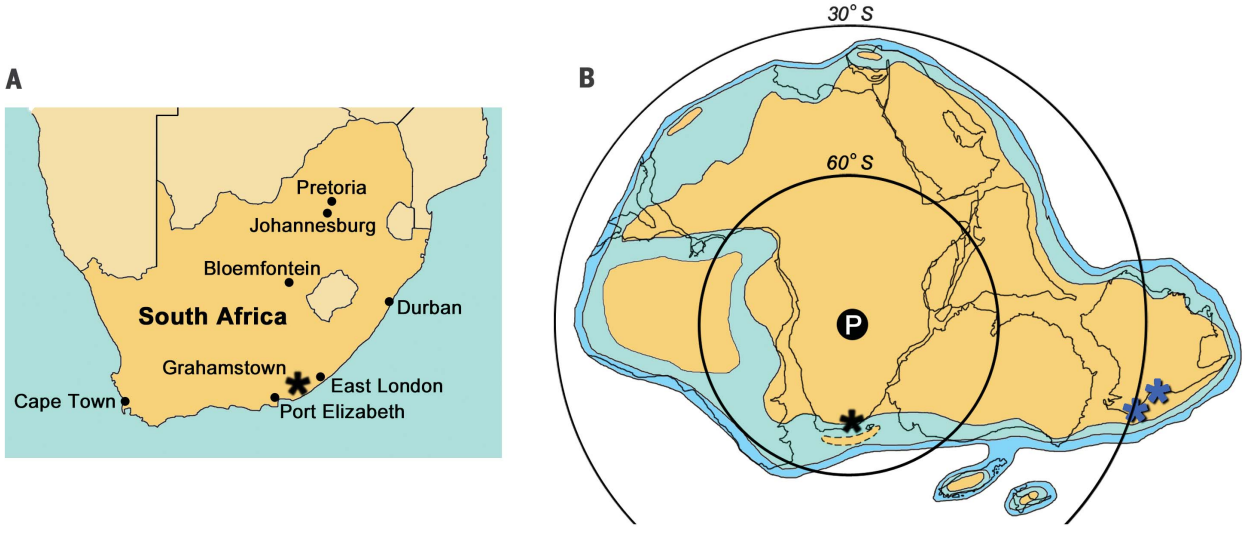


Fig. 1. Maps of the fossil locality. (**A**) Map of South Africa showing the Waterloo Farm fossil locality (black asterisk). (**B**) South-polar projection of Gondwana, modified from (*20*), showing Waterloo Farm (black asterisk) in relation to the reconstructed position of the South Pole 360 million years ago (*21*). Blue asterisks indicate the other two known Devonian tetrapod localities in Gondwana—Genoa River [left; footprints (*22*)] and Jemalong [right; *Metaxygnathus*, a single lower jaw ramus (*23*)]—both in Australia. The landmass with a dashed outline below Waterloo Farm is an emergent part of the Falklands Plateau, forming the outer margin of the semi-enclosed Agulhas Sea (*24*). Brown denotes land; pale blue, shallow shelf; blue, deep shelf.

ossifications appears to represent the coronoid series (Fig. 3C), implying that, of the three coronoids normally seen in tetrapods, at least the posterior one has been replaced by a chain of smaller elements in this taxon. An isolated element of this kind is also associated with the holotype (Fig. 2, N to P). The lateral line canals of the skull and lower jaw appear as a combination of continuous and discontinuous grooves, similar to the condition in other Devonian tetrapods (*5*, *8*), though poorly preserved.

Waterloo Farm demonstrates that the early evolution of tetrapods did not play out exclusively in tropical and subtropical environments. The late Famennian to Tournaisian witnessed a gradual transition from greenhouse to icehouse conditions, punctuated by an end-Famennian glaciation (*11*). Exact timing of this glaciation relative to Waterloo Farm deposition is uncertain, but late Famennian diamictites in South Africa are probably glaciogenic, and carbonates are

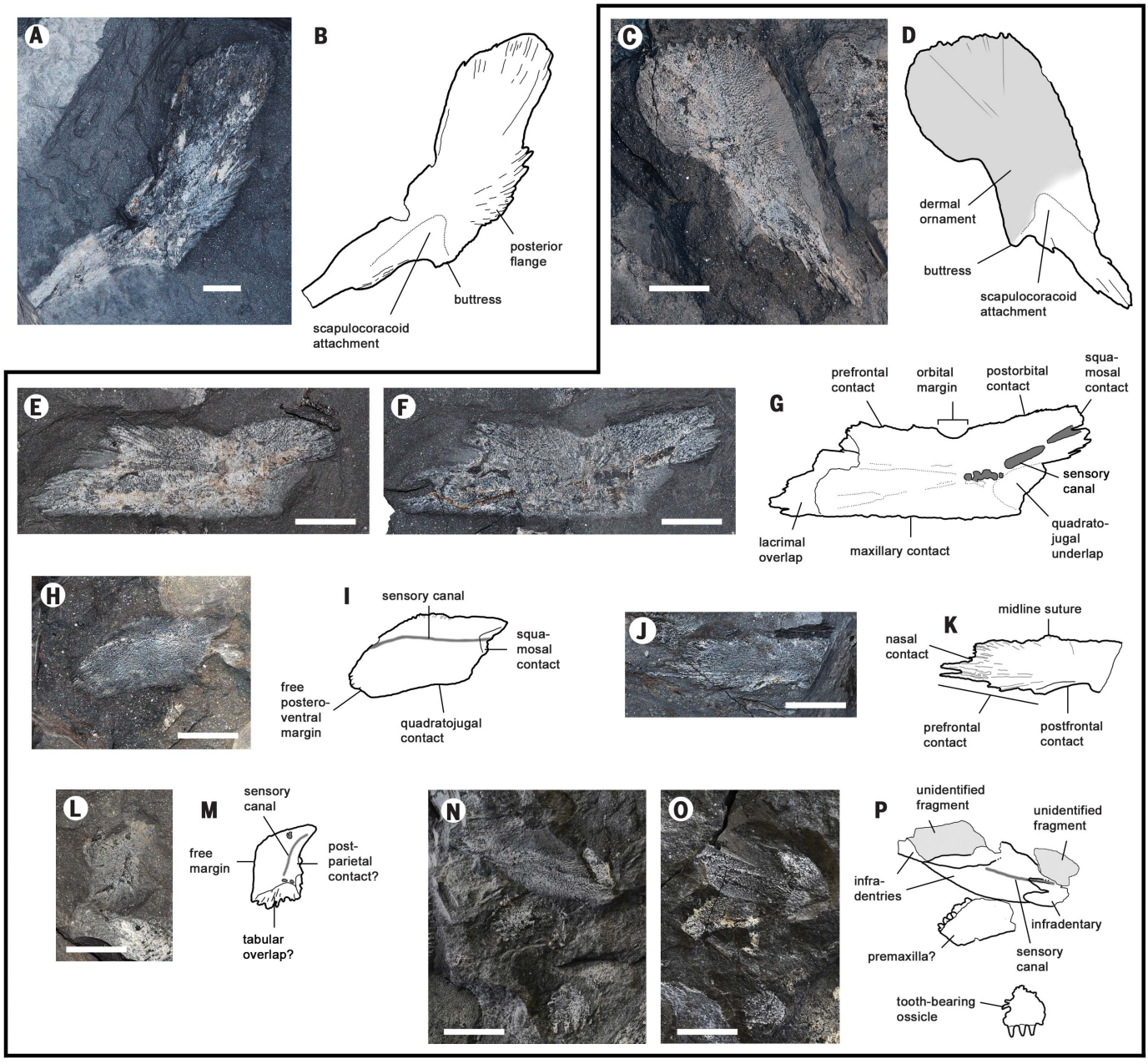


Fig. 2. Material of *Tutusius* and *Umzantsia*. (**A** and **B**) Photograph and line drawing of AM7527, a left cleithrum, the holotype and only known specimen of *Tutusius umlambo*. (**C** to **P**) AM7528a to -f, the bones of the holotype of *Umzantsia amazana*, believed to represent one individual. (C and D) AM7528a, right cleithrum (line drawing incorporates information from the counterpart); (E to G) AM7528b, left jugal, showing part, counterpart, and line drawing; (H and I) AM7528c, right preopercular; (J and K) AM7528d, incomplete left frontal; (L and M) probable left supratemporal; (N to P) AM7528e, a bone assemblage comprising a chain of two partial infradentaries, one near-complete infradentary, a probable premaxilla, and an unidentified tooth-bearing ossicle (see also Fig. 3). In all drawings, thick outlines denote true margins, and thin outlines denote broken or covered margins. In (D), gray shading indicates the dermal ornament. Anterior is to the left in (A), (B), (E) to (G), (J), and (K) and to the right in (C), (D), (H), and (I). All scale bars, 10 mm. (C) to (P) are shown to the same scale.

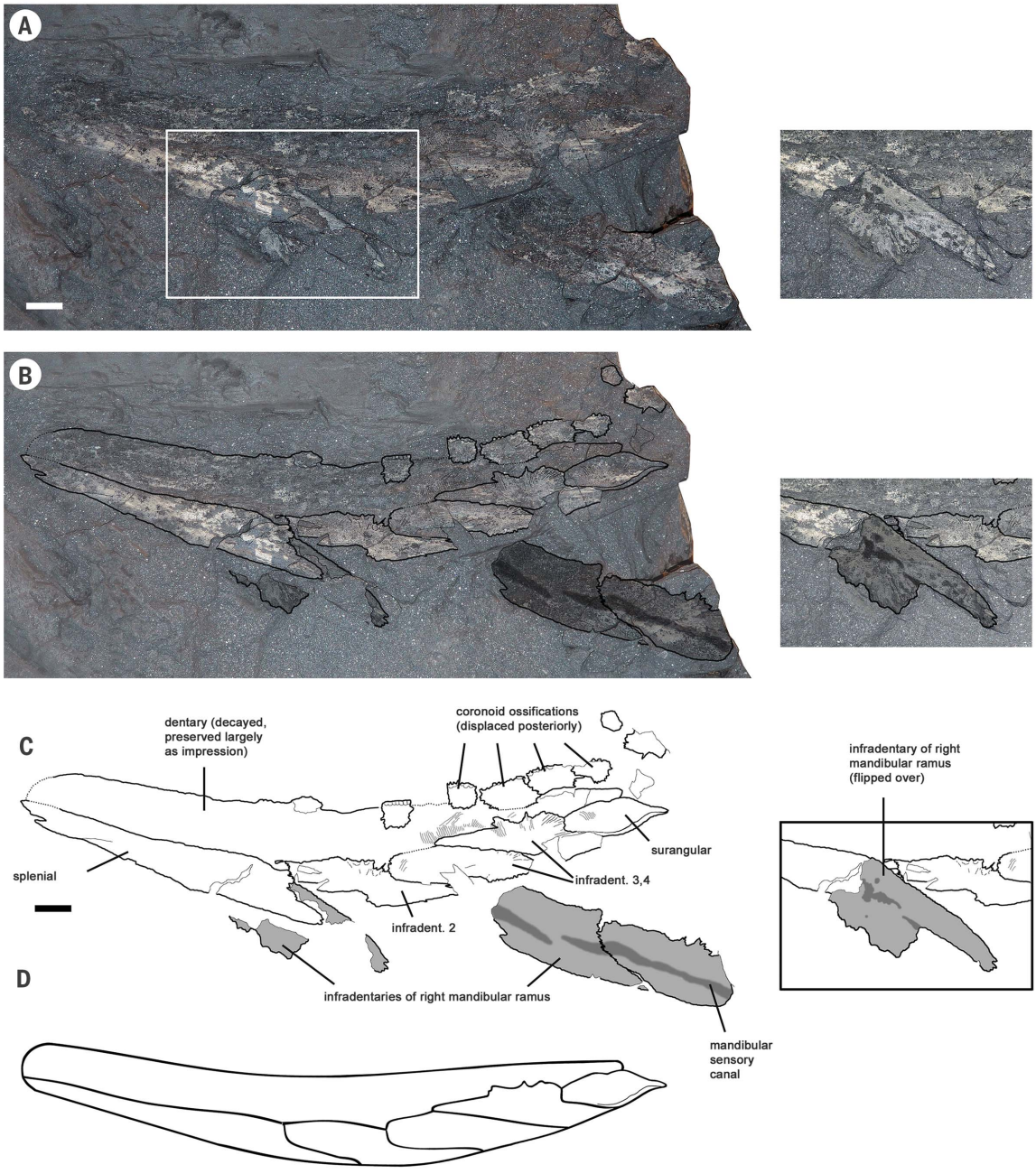


Fig. 3. The lower jaw of *Umzantsia*. (**A** to **C**) AM7529, left mandibular ramus and infradentaries of the right mandibular ramus of *Umzantsia amazana*. (A) Photograph of the specimen. The splenial of the left ramus partly overlies an infradentary of the right ramus; the area within the white box is shown on the left with the splenial in place and on the right as an excerpt box with the splenial removed. (B) Photograph overlaid with interpretative line drawing. (C) Interpretative line drawing. Light gray shading indicates infradentaries of the right mandibular ramus; dark gray, the sensory canals on these bones. Fine parallel gray lines on the left jaw ramus represent dermal ornament. (**D**) Sketch reconstruction of the left mandibular ramus (AM7529). Scale bars, 10 mm.

entirely absent from the region (*11*). Thus, even though Waterloo Farm yields a rich terrestrial flora that rules out a truly polar climate (*12*), it cannot have been very warm, and proximity to the pole implies several months of complete winter darkness.

The presence of tetrapods in such an environment raises the question of whether high-latitude environments played a distinctive role in the fish-tetrapod transition—for example, as drivers of innovation or as refuges for archaic taxa. The combination of autapomorphic and primitive characters in *Umzantsia* has bearing on this problem. All Devonian tetrapod cleithra described to date (*6–8*, *13–15*), including fragmentary late Frasnian material associated with *Elginerpeton* (*16*), lack dermal ornament (Fig. 4). This suggests that *Umzantsia* represents a deep but specialized branch of the tetrapod lineage that had been in existence since at least the Frasnian, a time interval of some 12 million years.

The Waterloo Farm tetrapod fossils and the Middle Devonian tetrapod trackways from Poland and Ireland (*17–19*) challenge the popular scenario of a tropical origin of tetrapods during the Late Devonian (*5*). Tetrapods originated no later than the Eifelian (early Middle Devonian), when they

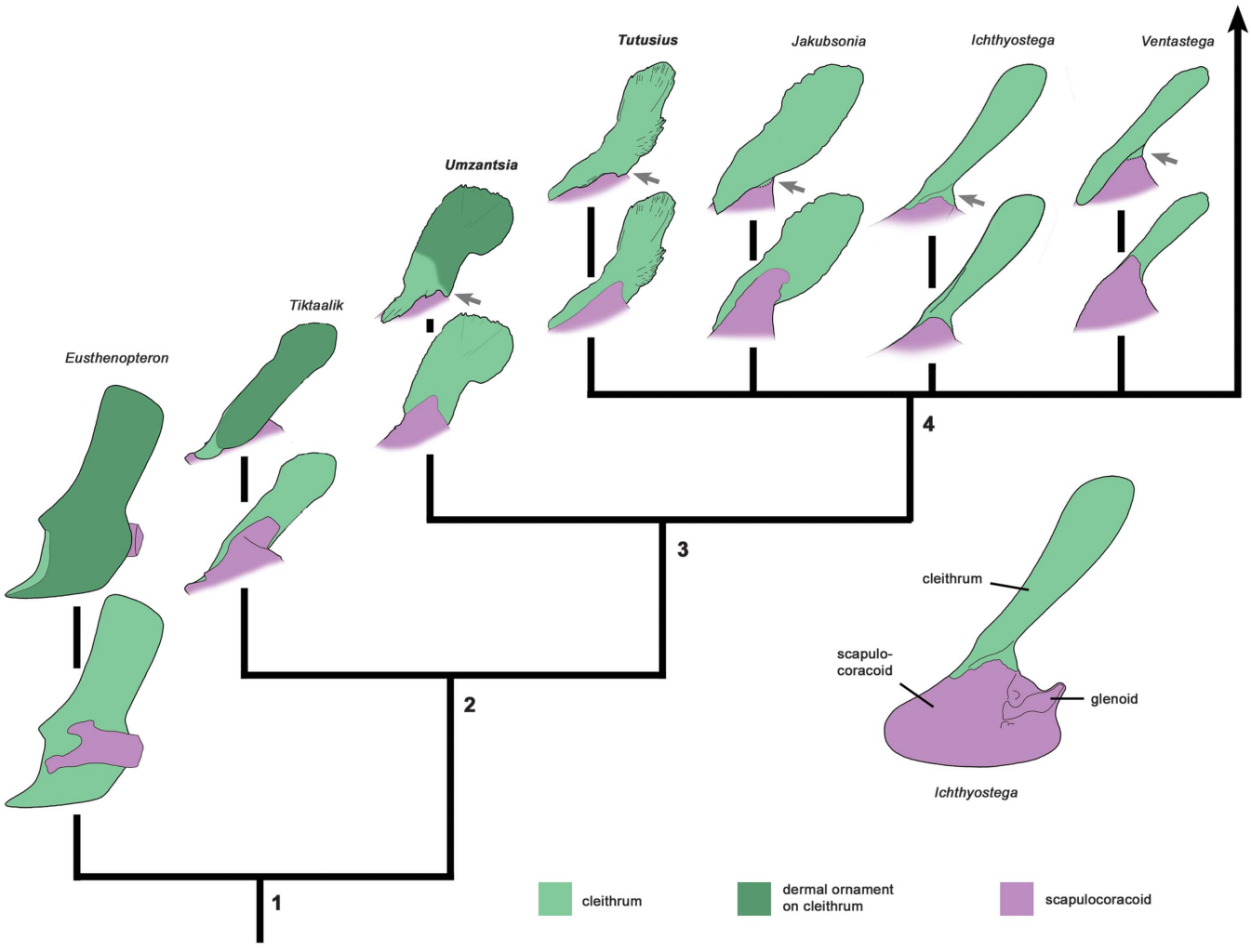


Fig. 4. Comparison of cleithra. A series of cleithra from the tetrapod stem group, spanning the fin-to-limb transition, placed on a simplified phylogeny that reflects recent analyses (*8*, *9*), illustrating the morphological transformation of the shoulder girdle and the tentative phylogenetic positions of *Tutusius* and *Umzantsia*. Not to scale. For each taxon, the upper image shows the cleithrum in external view, and the lower image shows it in internal view. Anterior is to the left in all cases. Only for *Eusthenopteron* (*25*) is the entire scapulocoracoid shown; in other taxa, the projecting ventral part has been cut off. For comparison, a complete scapulocoracoid plus cleithrum of *Ichthyostega* is shown (bottom right). The arrowhead branch of the phylogeny leads to the tetrapod crown group. Small gray arrows indicate the posteroventral buttress of the cleithrum. Character states at nodes: 1, the scapulocoracoid is small and concealed by the cleithrum in lateral view, and the cleithrum has a broad ventral lamina and is entirely covered with ornament (primitive condition, widely shared among Osteichthyes); 2, the cleithrum tapers to a point anteroventrally and attaches along the anterodorsal margin of the scapulocoracoid; 3, the dorsal end of the scapulocoracoid forms a v-shaped peak, and the cleithrum carries a posterodorsal buttress; 4, the cleithrum lacks ornament.

were present in southern Laurussia; by the late Famennian (latest Devonian), they ranged from the tropics to the south polar regions. This geographic pattern could still point to a tropical origin but may simply be a sampling artifact. Against this background, the continued investigation of nontropical localities such as Waterloo Farm must be a priority. Waterloo Farm is also the only known Devonian tetrapod locality to feature soft-tissue preservation, as exemplified by the earliest known lamprey, *Priscomyzon* (*1*). The locality thus has the potential not only to cast new light on early tetrapod biogeography and evolution, but also to illuminate unknown aspects of their morphology.

REFERENCES AND NOTES

1. R. W. Gess, M. I. Coates, B. S. Rubidge, *Nature* **443**, 981–984 (2006).
2. R. W. Gess, M. I. Coates, *Palaeont. Zeitschrift* **89**, 147–169 (2015).
3. R. W. Gess, M. I. Coates, *Zool. J. Linn. Soc.* **175**, 360–383 (2015).
4. R. W. Gess, K. M. Trinajstic, *PLOS ONE* **12**, e0173169 (2017).
5. J. A. Clack, *Gaining Ground: The Origin and Early Evolution of Tetrapods* (Indiana University Press, ed. 2, 2012).
6. O. A. Lebedev, *Acta Universitatis Latviensis* **679**, 79–98 (2004).
7. E. Jarvik, The Devonian tetrapod *Ichthyostega* (monograph no. 40, Fossils & Strata, Scandinavian Univ. Press, 1996).
8. P. E. Ahlberg, J. A. Clack, E. Luksevics, H. Blom, I. Zupiņs, *Nature* **453**, 1199–1204 (2008).
9. N. H. Shubin, E. B. Daeschler, F. A. Jenkins Jr., *Nature* **440**, 764–771 (2006).
10. J. A. Clack, *Palaeontology* **30**, 15–26 (1987).
11. J. A. Lakin, J. E. A. Marshall, I. Troth, I. C. Harding, *Geol. Soc. Lond. Spec. Publ.* **423**, 439–464 (2016).
12. R. W. Gess, N. Hiller, *Ann. Cape Provincial Museums* **19**, 225–304 (1995).
13. M. I. Coates, *Trans. R. Soc. Edinb. Earth Sci.* **87**, 363–421 (1996).
14. E. B. Daeschler, N. H. Shubin, K. S. Thomson, W. W. Amaral, *Science* **265**, 639–642 (1994).
15. O. A. Lebedev, M. I. Coates, *Zool. J. Linn. Soc.* **114**, 307–348 (1995).
16. P. E. Ahlberg, *Zool. J. Linn. Soc.* **122**, 99–141 (1998).
17. G. Niedźwiedzki, P. Szrek, K. Narkiewicz, M. Narkiewicz, P. E. Ahlberg, *Nature* **463**, 43–48 (2010).
18. I. Stössel, *J. Geol. Soc. Lond.* **152**, 407–413 (1995).
19. I. Stössel, E. A. Williams, K. T. Higgs, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **462**, 16–40 (2016).
20. T. H. Torsvik, L. R. M. Cocks, *Gondwana Res.* **24**, 999–1030 (2013).

21. T. H. Torsvik, L. R. M. Cocks, "The Palaeozoic palaeogeography of central Gondwana," in *The Formation and Evolution of Africa: A Synopsis of 3.8 Ga of Earth History*, D. J. J. Van Hinsbergen, S. J. H. Buiter, T. H. Torsvik, C. Gaina, S. J. Webb, Eds. (Special Publication 357, Geological Society of London, 2011), pp. 137–166.
22. J. W. Warren, N. A. Wakefield, *Nature* **238**, 469–470 (1972).
23. K. S. W. Campbell, M. W. Bell, *Alcheringa* **1**, 369–381 (1977).
24. T. McCarthy, B. S. Rubidge, *The Story of Earth & Life: A Southern African Perspective on a 4.6-Billion-Year Journey* (Struik Publishers, 2005).
25. E. Jarvik, *Basic Structure and Evolution of Vertebrates* (Academic Press, 1980), vol. 1.

ACKNOWLEDGMENTS

The South African National Roads Agency Limited supported rescue of shale during roadworks. R.G. acknowledges useful early discussions with M. Coates and P. Janvier regarding the morphology and confirming the identity of AM7527. B. Nosilela (Department of African Languages, Rhodes University) advised on taxonomic names. **Funding:** R.G. acknowledges funding from the South African Millennium Trust and the South African DST-NRF Centre of Excellence in Palaeosciences (CoE-Pal). P.E.A. acknowledges a Wallenberg Scholarship from the Knut and Alice Wallenberg Foundation. **Author contributions**: Fieldwork, collection and preparation of material, initial identification of tetrapod specimens (cleithra), and project conceptualization and design, R.G.; identification of additional specimens, R.G. and P.E.A.; manuscript writing and illustrations, P.E.A. and R.G. **Competing interests:** None declared. **Data and materials availability**: Formal taxonomy is presented in the supplementary materials. Specimens are accessioned at the Albany Museum, Grahamstown, South Africa, as AM7511 to AM7513.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1120/suppl/DC1
Materials and Methods
Systematic Paleontology
Figs. S1 and S2
Reference (*26*)

9 October 2017; accepted 25 April 2018
10.1126/science.aaq1645

EVOLUTIONARY COGNITION

Numerical ordering of zero in honey bees

Scarlett R. Howard[1], Aurore Avarguès-Weber[2]*, Jair E. Garcia[1]*, Andrew D. Greentree[3], Adrian G. Dyer[1,4]†

Some vertebrates demonstrate complex numerosity concepts—including addition, sequential ordering of numbers, or even the concept of zero—but whether an insect can develop an understanding for such concepts remains unknown. We trained individual honey bees to the numerical concepts of "greater than" or "less than" using stimuli containing one to six elemental features. Bees could subsequently extrapolate the concept of less than to order zero numerosity at the lower end of the numerical continuum. Bees demonstrated an understanding that parallels animals such as the African grey parrot, nonhuman primates, and even preschool children.

Four stages are used to describe the acquisition of understanding zero in human history, psychology, animal cognition, and neurophysiology (*1*). The first is the ability to define zero as nothing—the absence of a stimulus. The second is the categorical classification of zero as "nothing" versus "something." The third stage is understanding zero as a quantity at the low end of the positive integer numerical continuum. The fourth, and currently designated as the most advanced stage of understanding zero, is the symbolic representation of zero, as with an Arabic number and as used in modern mathematics and calculations (*1*).

Several ancient human civilizations lacked the full understanding and importance of zero, leading to constraints in their numeric systems (*1*). Interestingly, some vertebrate animals have recently demonstrated a capacity to acquire and understand this numerical concept. Rhesus monkeys learned that empty sets of objects occupy a position on a numerical continuum (*2*, *3*), vervet monkeys used subtraction-like reasoning to determine if food was present or absent (*4*), a chimpanzee reached near-perfect performance on zero-concept tasks with training (*5*), and an African grey parrot spontaneously labeled absent objects as "none" (*6*).

Honey bees have previously demonstrated the capacity to count and discriminate up to four objects (*7*–*10*) in experiments that use classic conditioning techniques. Recent advancements in conditioning protocols (*11*) reveal that bees can acquire rule-based relational concepts (*12*, *13*), thus enabling remarkable plasticity to acquire and apply seemingly advanced concepts such as size ordering (*14*). In this study, we tested the capacity of honey bees to extrapolate the acquired concepts of "greater than" and "less than," as shown in primates (*15*, *16*), and thus formally demonstrate that an invertebrate can understand the concept of zero numerosity.

We designed a set of experiments to test the extent to which honey bees may understand the concept of zero numerosity (*17*). In the first experiment, we trained bees to understand the concepts of less than and greater than using appetitive-aversive differential conditioning (*11*). Bees were trained to the respective concepts using white square stimuli containing one to four black elements (Fig. 1A, fig. S1, and table S1). After reaching a criterion of ≥80% accuracy, bees demonstrated in nonreinforced tests that they had learned the concept of "numerically less" [75.0 ± 4.1% (mean ± SEM); logistic regression with individual as random term tested differences between observed proportion of bee choices and chance level, $y = 0.5$, z score = 5.08, $P < 0.001$, $n = 10$] and "numerically greater" (75.5 ± 3.3%; z score = 6.556, $P < 0.001$, $n = 10$) when presented with novel stimuli of one to four elements. Furthermore, bees were able to apply these concepts to determine that five elements were greater than two or three elements (less-than group: 68.0 ± 5.0%, z score = 3.411, $P < 0.001$, $n = 10$; greater-than group: 75.0 ± 3.9%, z score = 5.333, $P < 0.001$, $n = 10$). Interestingly, bees demonstrated an understanding that zero numerosity lies at the lower end of the numerical continuum by choosing an "empty set" stimulus containing no elements if trained to less than (64.0 ± 5.4%; z score = 2.795, $P = 0.005$, $n = 10$; Fig. 1C) or by choosing unfamiliar stimuli containing elements if trained to greater than (74.5 ± 2.6%; z score = 6.609, $P < 0.001$, $n = 10$; Fig. 1C).

In the second experiment, we tested the extent to which bees may understand the quantitative concept of zero in comparison with other animals. As some animals find it challenging to differentiate between the numbers zero and one (*5*, *6*, *18*), we trained bees to less than using stimuli containing two to five elements and then tested their ability to differentiate between the unfamiliar numerosities of one and zero (Fig. 1B). After reaching a criterion of ≥80% accuracy, bees demonstrated the learned concept of numerically less when presented with the numbers two to five (73.8 ± 1.9%; z score = 10.18, $P < 0.001$). When presented with the unfamiliar numbers of one versus zero, bees chose the lower number of zero (63.0 ± 2.9%; z score = 4.23, $P < 0.001$; Fig. 1D), showing an understanding that an empty set is lower than one, which is challenging for some other animals (*5*, *6*, *18*).

When bees were presented with two conflicting pieces of information, two versus zero, where the two-element stimuli had always been rewarded in training and zero was the correct lower number, bees chose the empty set at a frequency level that was not significant from chance (56.2 ± 3.4%; z score = 1.64, $P = 0.101$; Fig. 1D); thus, bees perceived both plausible alternatives as consistent with their conditioning experience. These results demonstrate that bees were using both an associative mechanism for choosing two elements and a concept-based mechanism for choosing zero numerosity. This phenomenon was also observed in a dolphin trained to choose the numerically less option by using white dots on a black background. This result is explained in terms of an artifact of training-set conditioning causing a bias toward consistently rewarding stimuli (*19*).

To test if bees understood an empty set quantitatively along the numerical continuum, we evaluated numerical-distance effects, where accuracy of performance potentially improves as the difference in magnitude between two respective numbers increases (*1*). In the third experiment, we trained and tested bees on the less-than concept using the numbers zero to six. If bees considered zero numerosity as a number along the numerical continuum, we would expect accuracy of decisions to be the greatest with zero versus six and poorer for lower numbers versus zero numerosity (Fig. 2). After reaching a criterion of ≥80% accuracy during training, bees demonstrated in tests that they could discriminate an empty set from numbers one to six accurately (supplementary text S1 and Fig. 2B). Although bees could accurately discriminate all numbers from zero numerosity, there was a significant effect of numerical distance on accuracy (Fig. 2B). Bees were more accurate when numbers were numerically more distant (zero versus five and zero versus six) than when numerically closer (zero versus one), showing that bees are affected by number magnitude and thus exhibit numerical-distance effects.

An alternative explanation for our results could be that bees have a preference for the unfamiliar presentation of an empty-set stimulus. However, control experiments showed that the bees' understanding that zero belongs at the lower end of the numerical continuum was rule based and not driven by an unfamiliar preference (supplementary text S2 and fig. S2). The spatial

[1]Bio-inspired Digital Sensing (BIDS) Lab, School of Media and Communication, RMIT University, Melbourne, VIC, Australia. [2]Centre de Recherches sur la Cognition Animale, Centre de Biologie Intégrative (CBI), Université de Toulouse, CNRS, UPS, Toulouse, France. [3]ARC Centre of Excellence for Nanoscale BioPhotonics, School of Science, RMIT University, Melbourne, VIC, Australia. [4]Department of Physiology, Monash University, Clayton, VIC, Australia.
*These authors contributed equally to this work.
†Corresponding author. Email: adrian.dyer@rmit.edu.au

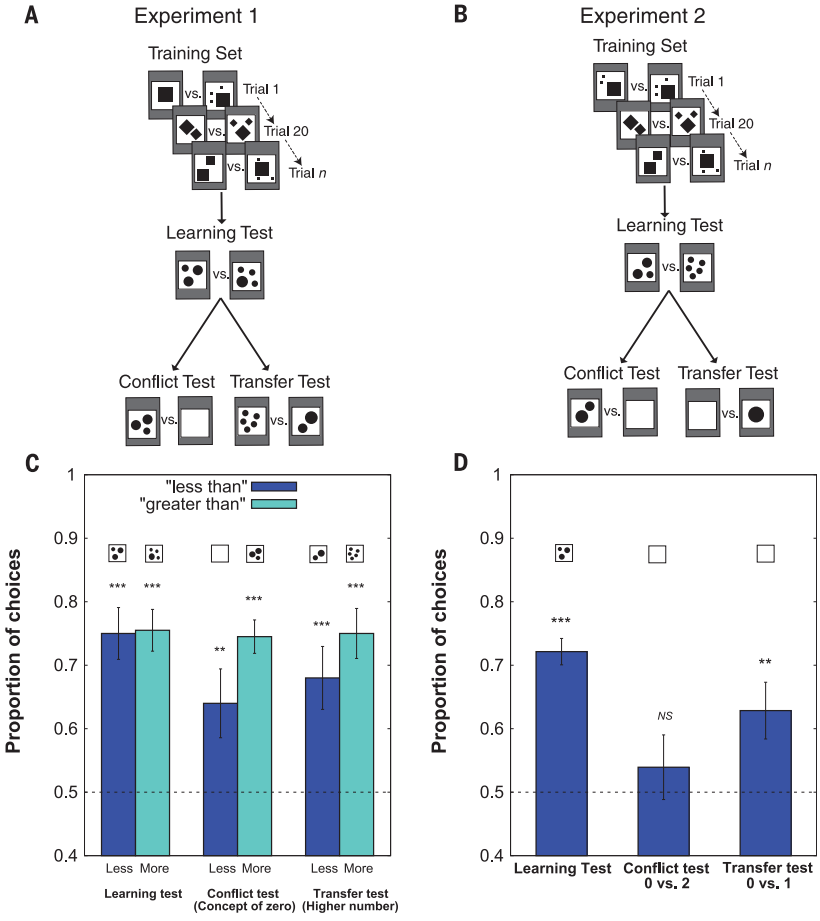


Fig. 1. Graphic representation of the method and results of experiments 1 and 2. (**A** and **B**) Examples of possible stimuli combinations during trials and tests in experiments 1 and 2. (**C** and **D**) Performance during the unreinforced testing phases during experiments 1 and 2. Data shown are means ± SEM for both treatment groups. Bees trained to less than are shown in dark blue; bees trained to greater than are shown in turquoise. Stimuli above the columns represent the choices for those stimuli in the data. In experiment 1, in the conflict test evaluating the bees' concept of zero, data shown for the less-than group (n = 10) are choices for zero, and data shown for the greater-than group (n = 10) are choices for stimuli containing elements. In the transfer test to a higher number, data shown for bees trained to less than are choices for a lower number, and data shown for bees trained to greater than are choices for the higher number of five. In experiment 2 (n = 25), the conflict and transfer tests show the bees' choices for zero. Dashed black line at 0.5 indicates chance-level performance. Significance from chance-level performance is indicated by ** $P \geq 0.01$ and *** $P \geq 0.001$. NS, not significant.

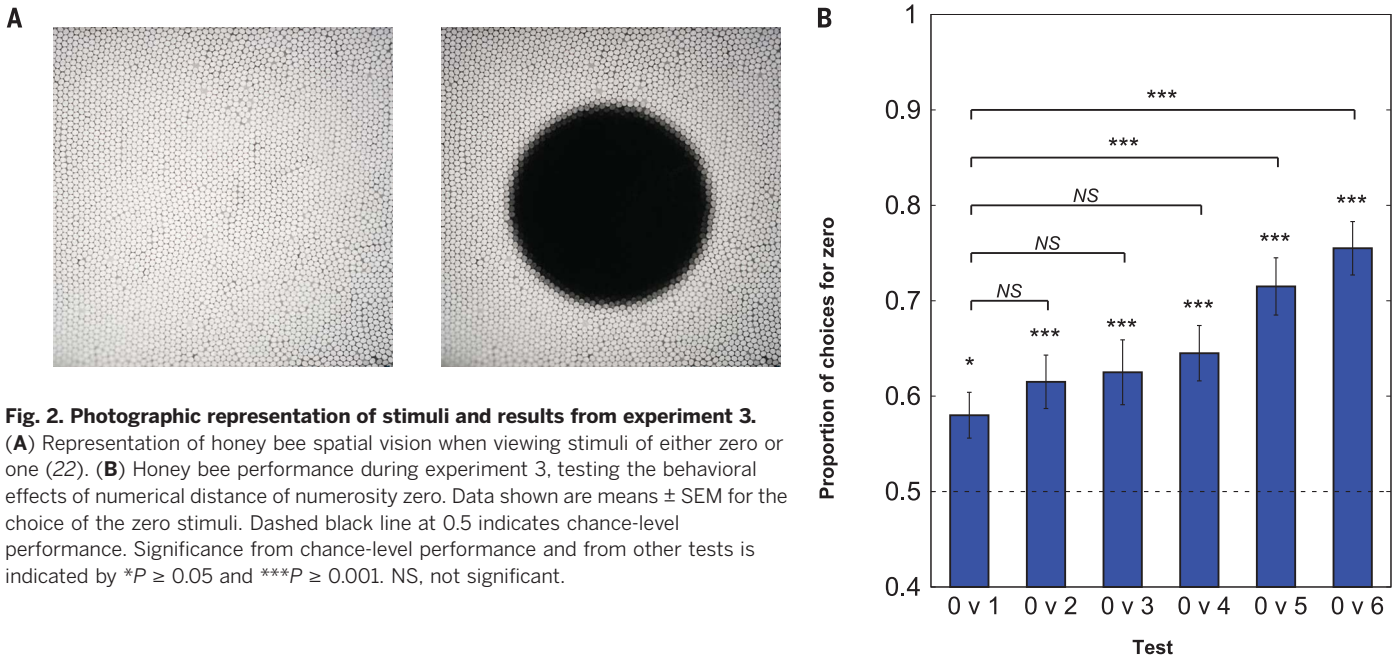


Fig. 2. Photographic representation of stimuli and results from experiment 3. (**A**) Representation of honey bee spatial vision when viewing stimuli of either zero or one (*22*). (**B**) Honey bee performance during experiment 3, testing the behavioral effects of numerical distance of numerosity zero. Data shown are means ± SEM for the choice of the zero stimuli. Dashed black line at 0.5 indicates chance-level performance. Significance from chance-level performance and from other tests is indicated by *$P \geq 0.05$ and ***$P \geq 0.001$. NS, not significant.

frequencies of stimuli are also ruled out as a potential explanation for results (supplementary text S1 and table S1). We additionally conducted further control experiments to exclude the possibility that bees learn to match pairs of numbers during training (supplementary text S3 and fig. S2).

Our findings show that honey bees can learn and apply the concepts of greater than and less than to interpret a blank stimulus as representing the conceptual number of zero and place zero in relation to other numerical values. Bees thus perform at a level consistent with that of nonhuman primates by understanding that zero is lower than one (*5*).

An open question remains as to whether such advanced numerical understandings may be widespread across many animals that deal with complexity in their environments or if our findings are the result of independent evolution in honey bees. Recent comparative studies of primate and crow brains found that similar levels of numeric processing are facilitated by very different brain structures, suggesting independent evolution of numeric processing (*20*, *21*). Because it can be demonstrated that an insect, with a different brain structure from primates and birds, can understand the concept of zero, it would be of high value to consider such capacities in other animals.

REFERENCES AND NOTES

1. A. Nieder, *Trends Cogn. Sci.* **20**, 830–842 (2016).
2. D. J. Merritt, R. Rugani, E. M. Brannon, *J. Exp. Psychol. Gen.* **138**, 258–269 (2009).
3. A. Ramirez-Cardenas, M. Moskaleva, A. Nieder, *Curr. Biol.* **26**, 1285–1294 (2016).
4. S. Tsutsumi, T. Ushitani, K. Fujita, *Int. J. Zool.* **2011**, 1–11 (2011).
5. D. Biro, T. Matsuzawa, *Anim. Cogn.* **4**, 193–199 (2001).
6. I. M. Pepperberg, J. D. Gordon, *J. Comp. Psychol.* **119**, 197–209 (2005).
7. M. Dacke, M. V. Srinivasan, *Anim. Cogn.* **11**, 683–689 (2008).
8. P. Skorupski, H. MaBouDi, H. S. Galpayage Dona, L. Chittka, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373**, 20160513 (2017).
9. L. Chittka, K. Geiger, *Anim. Behav.* **49**, 159–164 (1995).
10. H. J. Gross *et al.*, *PLOS ONE* **4**, e4263 (2009).
11. A. Avarguès-Weber, M. G. de Brito Sanchez, M. Giurfa, A. G. Dyer, *PLOS ONE* **5**, e15370 (2010).
12. A. Avarguès-Weber, A. G. Dyer, M. Giurfa, *Proc. Biol. Sci.* **278**, 898–905 (2011).
13. A. Avarguès-Weber, A. G. Dyer, M. Combe, M. Giurfa, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7481–7486 (2012).
14. S. R. Howard, A. Avarguès-Weber, J. Garcia, A. G. Dyer, *Anim. Cogn.* **20**, 627–638 (2017).
15. J. F. Cantlon, E. M. Brannon, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16507–16511 (2005).
16. S. Bongard, A. Nieder, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 2277–2282 (2010).
17. Materials and methods are available as supplementary materials.
18. I. M. Pepperberg, *J. Comp. Psychol.* **120**, 1–11 (2006).
19. K. Jaakkola, W. Fellner, L. Erb, M. Rodriguez, E. Guarino, *J. Comp. Psychol.* **119**, 296–303 (2005).
20. H. M. Ditz, A. Nieder, *J. Neurosci.* **36**, 12044–12052 (2016).
21. A. Nieder, *Curr. Opin. Behav. Sci.* **16**, 8–14 (2017).
22. A. G. Dyer, S. Williams, *Imaging Sci. J.* **53**, 209–213 (2005).

ACKNOWLEDGMENTS

We thank M. Giurfa, S. Zhang, and J. M. Devaud for suggestions. **Funding:** S.R.H. acknowledges the Australian Government Research Training Program (RTP) Scholarship and the Company of Biologists JEB Travelling Fellowship. A.D.G. is funded by the Australian Research Council through the Future Fellowship scheme, FT160100357. **Author contributions:** S.R.H., A.A.-W., J.E.G., A.D.G., and A.G.D. were involved in the design of the experiment, data interpretation, and drafting of the manuscript. S.R.H., J.E.G., A.D.G., and A.G.D. analyzed data. S.R.H. collected data. All authors gave final approval for submission. **Competing interests:** None declared. **Data and materials availability:** Additional data, including the individual choices of all bees, are available in the supplementary materials, and raw data are available in the Dryad Digital Repository at doi:10.5061/dryad.7187rf5.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1124/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S4
Table S1
References (*23*–*26*)

14 November 2017; accepted 25 April 2018
10.1126/science.aar4975

PLANT SCIENCE

Plants send small RNAs in extracellular vesicles to fungal pathogen to silence virulence genes

Qiang Cai[1], Lulu Qiao[1,2], Ming Wang[1], Baoye He[1], Feng-Mao Lin[3], Jared Palmquist[1], Sienna-Da Huang[3], Hailing Jin[1]*

Some pathogens and pests deliver small RNAs (sRNAs) into host cells to suppress host immunity. Conversely, hosts also transfer sRNAs into pathogens and pests to inhibit their virulence. Although sRNA trafficking has been observed in a wide variety of interactions, how sRNAs are transferred, especially from hosts to pathogens and pests, is still unknown. Here, we show that host *Arabidopsis* cells secrete exosome-like extracellular vesicles to deliver sRNAs into fungal pathogen *Botrytis cinerea*. These sRNA-containing vesicles accumulate at the infection sites and are taken up by the fungal cells. Transferred host sRNAs induce silencing of fungal genes critical for pathogenicity. Thus, *Arabidopsis* has adapted exosome-mediated cross-kingdom RNA interference as part of its immune responses during the evolutionary arms race with the pathogen.

Small RNAs (sRNAs) are short regulatory RNAs that silence genes with complementary sequences (*1*). Within an animal, sRNAs can be transported by extracellular vesicles, specific transmembrane proteins, high-density lipoprotein complexes, or gap junctions (*2*). Within a plant, sRNAs move from cell to cell, presumably through plasmodesmata, and travel systemically through vasculature (*3*). Remarkably, sRNAs also move between hosts and interacting organisms and induce gene silencing, a phenomenon called cross-kingdom/organism RNA interference (RNAi) (*4*–*10*). But how do mobile sRNAs travel across the boundaries between organisms of different taxonomic kingdoms? The gastrointestinal parasite *Heligmosomoides polygyrus* secretes exosomes, a class of extracellular vesicles, to transport microRNAs (miRNAs) into mammalian cells to suppress host immunity (*7*). By contrast, the mechanism by which sRNAs are transported from hosts to interacting pathogens and pests is totally unknown.

To investigate how host sRNAs move into interacting fungal cells and identify plant endogenous transferred sRNAs, we used the *Arabidopsis*–*Botrytis cinerea* pathosystem that displays bidirectional sRNA trafficking and cross-kingdom RNAi (*6*, *8*, *11*, *12*). Because the cell wall compositions of plants and fungi are different, we developed a sequential protoplast purification method to isolate pure fungal cells from infected tissues (fig. S1). sRNA profiling of purified *B. cinerea* protoplasts identified 42 *Arabidopsis* sRNAs in all three biological replicates, using 40 normalized reads per million total reads as a cutoff (table S1). We performed sRNA profiling on total RNAs of *Arabidopsis* leaves as a control. Although the more abundant sRNAs were more likely to be transported (table S2), there is a clear selection in transferred sRNAs. Among the 42 transferred *Arabidopsis* sRNAs, 25 were lowly abundant and not among the top 100 sRNAs in the total sRNA libraries (tables S2 and S3). Despite being generated from the same TAS1c mRNA precursor and belong to the top 20 most abundant sRNAs from the total sRNA libraries, trans-acting small interfering RNA locus 1C-derived siRNA483 (TAS1c-siR483) but not TAS1c-siR585 was enriched in the fungal protoplasts (tables S2 and S3). Similarly, although TAS2-siR710 had 30 times higher reads in the total sRNA libraries than TAS2-siR453 that derived from the same TAS2 precursor, only TAS2-siR453 was present in the fungal protoplasts. Furthermore, *Arabidopsis* heterochromatic sRNAs that derived from intergenic regions, such as IGN-siR1 but not IGN-siR107, were enriched in *B. cinerea* cells, although IGN-siR107 accumulated to a higher level in the total sRNA libraries (tables S2 and S3). These results were validated by means of sRNA reverse transcription polymerase chain reaction (RT-PCR) analysis (Fig. 1A). Thus, host cells transferring endogenous sRNAs into fungal cells are not simply through concentration-dependent diffusion but possibly through a more selective process.

Extracellular vesicles are implicated in systemic sRNA transport in animals (*13*). To test whether plants secrete vesicles to transfer sRNAs into fungal cells, we profiled vesicle sRNAs isolated from the extracellular apoplastic fluids of infected leaves. In all three biological replicates, TAS1c-siR483, TAS2-siR453, and IGN-siR1 accumulated to higher levels than TAS1c-siR585, TAS2-siR710, and IGN-siR107, respectively (Fig. 1B and tables S2 and S4). This was similar to fungal protoplast results. Of the 42 transferred host sRNAs, 31 (73.8%) were present in vesicle libraries (table S4). These results support a positive correlation between the sRNA profiles from extracellular vesicles and the fungal cells. Nuclease protection assay showed that transferred host sRNAs were protected from nuclease digestion unless vesicles were ruptured with Triton X-100 (Sigma, St. Louis, Missouri) (Fig. 1C) (*7*, *14*), confirming that transferred sRNAs are indeed inside the vesicles instead of bound to the surface. These

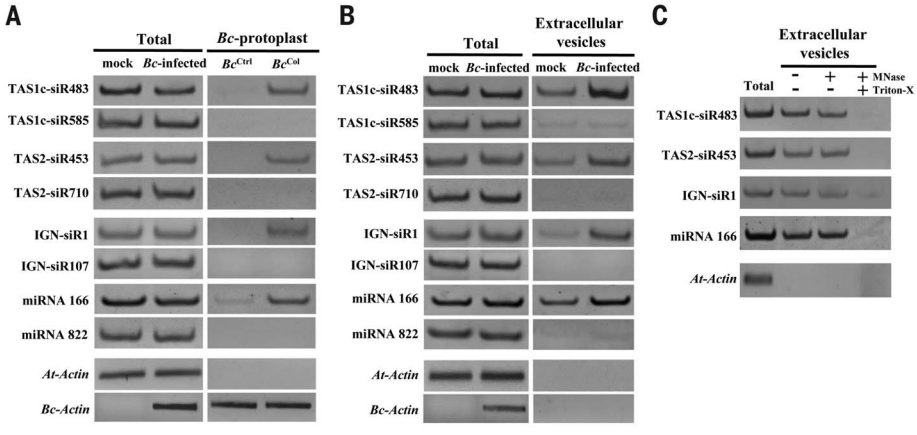


Fig. 1. Plant endogenous sRNAs are transferred into fungal cells via extracellular vesicles. (**A**) TAS1c-siR483, TAS2-siR453, IGN-siR1, and miRNA166 were detected by means of sRNA semiquantitative RT-PCR in *B. cinerea* protoplasts purified from *B. cinerea*–infected Col-0 *Arabidopsis* (Bc^{Col}). As a negative control, cultured *B. cinerea* mixed with uninfected leaves was subjected to the same procedure (Bc^{Ctrl}). (**B**) These host sRNAs were also in *Arabidopsis* extracellular vesicles. (**C**) *Arabidopsis* sRNAs were detected in micrococcal nuclease–treated extracellular vesicles. In (A) to (C), TAS1c-siR585, TAS2-siR710, IGN-siR107, miRNA822, and *Actin* genes were used as controls. The "total" lane indicates total RNAs from leaves. Similar results were obtained from two biological replicates.

[1]Department of Microbiology and Plant Pathology, Center for Plant Cell Biology, Institute for Integrative Genome Biology, University of California, Riverside, 900 University Avenue, Riverside, CA 92521, USA. [2]Department of Plant Protection, Nanjing Agriculture University, Nanjing, 210095, China. [3]Department of Biological Science and Technology, National Chiao Tung University, Hsin-Chu 300, Taiwan.
*Corresponding author. Email: hailingj@ucr.edu

findings indicate that plant cells secrete extracellular vesicles to transfer sRNAs into fungal cells.

Animal extracellular vesicles are classified into exosomes, shedding microvesicles and apoptotic bodies by their specific protein markers and origins (*13*), whereas plant extracellular vesicles remain undefined. Because exosomes are involved in transferring miRNAs between animal cells (*14*, *15*) and from the parasite *H. polygyrus* to its mammalian host (*7*), we sought to determine whether plants also use exosome-like vesicles to transfer sRNAs. Mammalian tetraspanins, such as CD63, are specific exosome markers (*16*). *Arabidopsis* has 17 *TETRASPANIN* (*TET*)–like genes (*17*), but only closely related *TET8* and *TET9* are induced through *B. cinerea* infection (fig. S2A) (*18*). TET8 and TET9 are structurally similar to mammalian CD63 (fig. S2B) (*17*).

In transgenic plants expressing *TET8-GFP* under its native promoter, TET8 accumulated at the fungal infection sites (Fig. 2A and fig. S3). Mammalian exosomes are derived from multivesicular bodies (MVBs) (*13*, *19*). We found that *Arabidopsis* MVB marker Rab5-like guanosine triphosphatase (GTPase) ARA6 (*20*) was enriched at infection sites (fig. S4) and partially colocalized with TET8 (fig. S5), suggesting that TET8-associated vesicles are likely derived from MVBs. By using transmission electron microscopy, we observed MVBs fusing with plasma membrane to release the extracellular vesicles at the infection sites (Fig. 2B and fig. S6A). Furthermore, TET8-associated vesicles are secreted because numerous green fluorescent protein (GFP)–labeled apoplastic vesicles were observed from *TET8-GFP* plants but none from *ARA6-GFP* plants (fig. S6B). TET8, but not ARA6, was detected in extracellular vesicles (fig. S6C) because ARA6 localizes to MVB outer

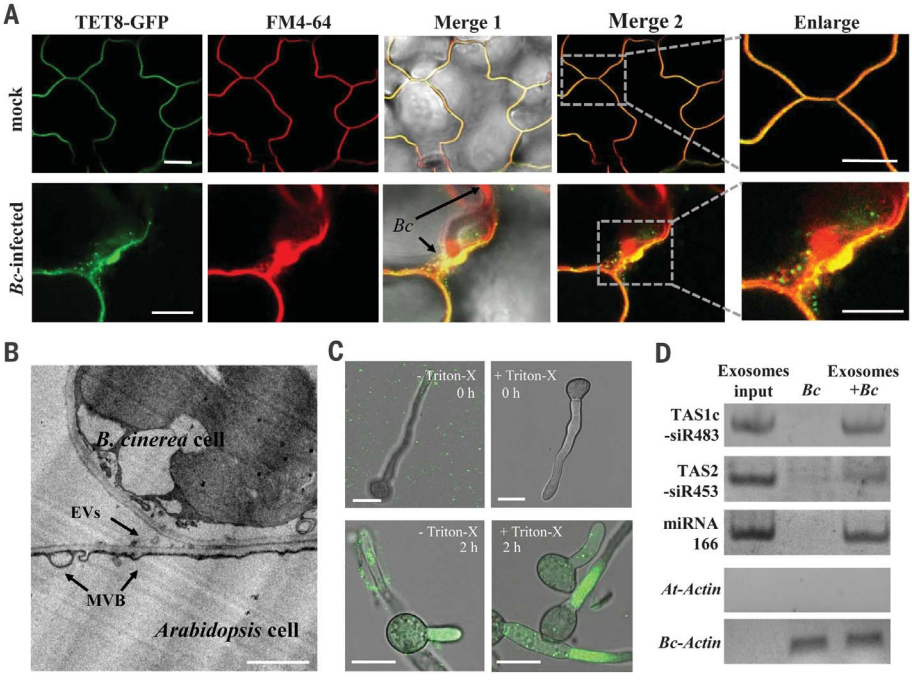


Fig. 2. Tetraspanin-associated exosomes are responsible for transferring plant sRNAs to fungal cells. (**A**) *B. cinerea* induces accumulation of TET8-GFP–labeled vesicles at the sites of infection. Short staining of FM4-64 shows extracellular membrane structures and plasma membranes. (**B**) Transmission electron microscopic image of *Arabidopsis* extracellular vesicles (EVs) near the *B. cinerea* infection sites. Scale bars, 1 μm. (**C**) TET8-GFP–labeled exosomes were taken up by *B. cinerea* within 2 hours of co-incubation. Scale bars, (A) and (C), 10 μm. (**D**) Plant sRNAs were detected in *B. cinerea* after uptake of the exosomes.

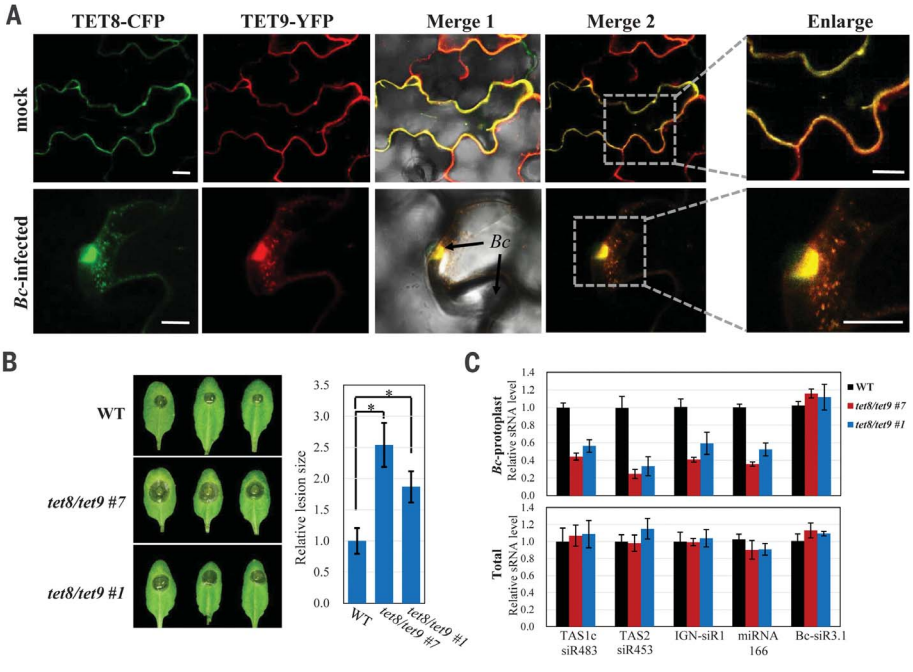


Fig. 3. TET8 and TET9 coordinately regulate sRNA secretion and host immunity. (**A**) TET8-CFP and TET9-YFP colocalized in vesicles that accumulated at the site of fungal infection. Scale bars, 10 μm. (**B**) Enhanced susceptibility to *B. cinerea* was observed in two independent *tet8tet9* double mutant lines. Relative lesion sizes were measured at 2 days after infection. Error bars indicate the SD of more than 10 leaves. The asterisks indicate significant difference [analysis of variance (ANOVA) Dunnett's multiple comparisons, $P < 0.01$]. (**C**) Accumulation of host transferred sRNAs was decreased in the purified fungal protoplasts isolated from the infected *tet8tet9* mutants as compared with that from the wild type. In fungal protoplasts and total RNA samples, *Bc-Actin* and *At-Actin* were used as internal controls, respectively. The *B. cinerea* sRNA Bc-siR3.1 serves as a control.

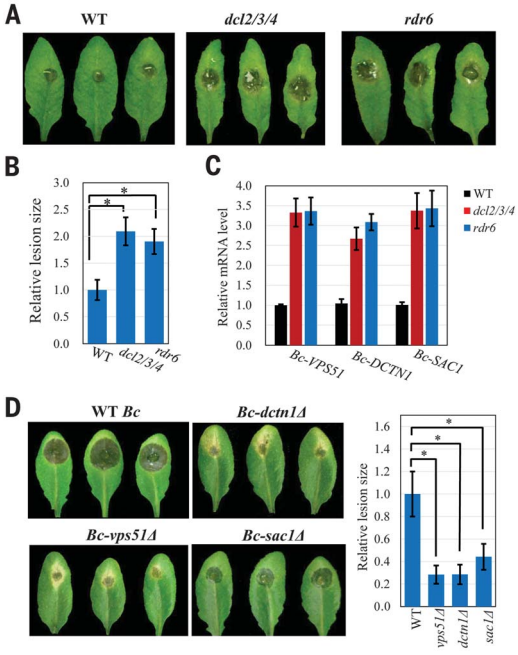


Fig. 4. Transferred host sRNAs silence fungal virulence genes and suppress fungal pathogenicity. (**A** and **B**) The *dcl2/3/4* and *rdr6* mutants exhibit enhanced susceptibility to *B. cinerea* as compared with the wild type. Relative lesion sizes were measured at 2 days after infection. (**C**) Fungal target genes of transferred sRNAs were derepressed in *B. cinerea* collected from the *dcl2/3/4* and *rdr6* mutants. (**D**) Mutant *B. cinerea* strains with deletions in the targets of transferred host sRNAs displayed reduced virulence. Relative lesion sizes were measured at 3 days after infection. In (B) and (D), error bars indicate the SD over 10 leaves. The asterisk indicates significant difference (ANOVA Dunnett's multiple comparisons; $P < 0.01$).

membranes that fused to and remained on plasma membranes. Thus, TET8-labeled extracellular vesicles can be considered plant exosomes.

To test whether plant exosomes can be taken up by fungal cells, we incubated *B. cinerea* cells with isolated TET8-GFP–labeled exosomes in vitro. TET8-GFP was detected in fungal cells within 2 hours and remained in the fungal cells after rupturing exosomes with Triton X-100 (Fig. 2C), indicating that fungal cells can efficiently take up plant exosomes. Furthermore, the exosome-carried sRNAs were detected in the fungal cells (Fig. 2D). These results support that plants secrete exosomes to transfer sRNAs into fungal cells.

Animal tetraspanin proteins often function together and form specific membrane microdomains (*21*). We found that TET8-CFP and TET9-YFP were colocalized to the infection sites (Fig. 3A and fig. S7, A and B). TET8-TET9 association was confirmed with reciprocal coimmunoprecipitation when transiently coexpressed in *Nicotiana benthamiana* (fig. S7C). To explore the physiological role of *TET8* and *TET9*, we examined their loss-of-function mutants. Both the *tet8* transferred DNA knockout mutant (*17*) and two independent *tet9* CRISPR frame-shift lines that we generated displayed weak but consistent enhanced susceptibility to *B. cinerea* as compared with the wild type and the *TET8* complementary lines (fig. S8). We generated the *tet8tet9* double mutants by knocking down *TET9* expression with artificial miRNA in the *tet8* background (fig. S9). Double mutants showed pronounced enhanced susceptibility (Fig. 3B), suggesting that TET8 and TET9 have partial functional redundancy. Furthermore, transferred host sRNA levels in the fungal cells isolated from the *tet8tet9* mutants were reduced (Fig. 3C), supporting that TET8- and TET9-associated exosomes contribute to plant immunity against fungal infection by transferring host sRNAs into fungal cells.

To determine whether transferred sRNAs affect fungal virulence, we infected the *Arabidopsis* triple mutant *dcl2/3/4* that impairs the biogenesis of trans-acting siRNAs (tasiRNAs) and heterochromatic siRNAs and the *rdr6* mutant that is compromised in tasiRNA formation (*22*). Both *dcl2/3/4* and *rdr6-15* mutants were more susceptible to *B. cinerea* as compared with wild type (Fig. 4, A and B), suggesting that transferred sRNAs inhibit *B. cinerea* infection. We hypothesize that transferred sRNAs suppress fungal pathogenicity by targeting fungal virulence genes. Of the transferred *Arabidopsis* sRNAs, 21 have predicted *B. cinerea* target genes (table S5), with a bias toward vesicle-trafficking pathways (7 out of 32 target genes) (fig. S10). We performed functional analysis on TAS1c-siR483 and TAS2-siR453, which showed selective accumulation in the fungal cells (Fig. 1). TAS1c-siR483 target two *B. cinerea* genes BC1G_10728 and BC1G_10508, and TAS2-siR453 target BC1G_08464, all of which are involved in vesicle-trafficking pathways. BC1G_10728 encodes vacuolar protein sorting 51 (Bc-Vps51) (*23*), and its homologous gene plays a key role in *Candida albicans* virulence (*24*). BC1G_10508 encodes the large subunit of the dynactin complex (Bc-DCTN1). DCTN binds to kinesin II and dynein and coordinates vesicle trafficking (*25*). BC1G_08464 encodes a suppressor of actin (SAC1)–like phosphoinositide phosphatase that regulates secretory membrane trafficking (*26*). As expected, these predicted target genes were down-regulated after infection (fig. S11A), but not in *B. cinerea* collected from the infected *dcl2/3/4* and *rdr6* mutants (Fig. 4C). The coexpression assay of host sRNAs and fungal targets in *N. benthamiana* confirmed that silencing of fungal genes was indeed triggered by host sRNAs, and silencing was abolished when sRNA target sites were mutated (fig. S11B). Fungal Argonaute proteins have ribonuclease activities that cleaves sRNA targets (*10*, *27*). The cleavage products of fungal targets guided by host sRNAs were detected in fungal cells isolated from the infected tissue (fig. S12), indicating that plant sRNAs silence fungal target genes through mRNA cleavage.

We further assessed whether these vesicle-trafficking target genes are important for fungal virulence by generating *vps51*Δ, *dctn1*Δ, and *sac1*Δ deletion mutant strains (fig. S13A). All mutant strains showed reduced virulence on *Arabidopsis* (Fig. 4D). The *sac1*Δ strain showed no reduction of growth on media (fig. S13B), suggesting that SAC1 has a direct role in fungal virulence and does not act by disrupting fungal growth. Thus, vesicle-trafficking pathways are important for fungal virulence. To further confirm that transferred host sRNAs promote host immunity, we generated transgenic *Arabidopsis* lines that either overexpress TAS1c-siR483 or TAS2-siR453 or knockdown both TAS1c-siR483 and TAS2-siR453 (figs. S14A and S15A). The overexpression lines of TAS1c-siR483 and TAS2-siR453, but not other unrelated sRNA, displayed reduced susceptibility to *B. cinerea* (fig. S14, B to D), and *B. cinerea* from infected tissue showed reduced target gene expression (fig. S14E). Consistently, the short tandem target mimic knockdown lines displayed increased susceptibility to *B. cinerea* (fig. S15B), and target gene expression was elevated (fig. S15C). These findings further support that transferred host sRNAs contribute to host immunity by silencing fungal genes.

In this study, we report that plant extracellular vesicles, especially exosomes, play an essential role in cross-kingdom sRNA trafficking between *Arabidopsis* and the fungal pathogen *B. cinerea*. *Arabidopsis* secretes exosomes to deliver host sRNAs into fungal cells to silence virulence-related genes. A parallel is seen in humans: Cross-kingdom trafficking of human miRNAs into the parasite *Plasmodium falciparum* inhibits its pathogenicity genes, which explains why sickle cell anemia patients, who have elevated transferred miRNAs, are more resistant to malaria (*28*). It is unclear whether human exosomes are also responsible for sRNA delivery to interacting organisms, and how ubiquitously such sRNA

trafficking-mediated defense mechanisms exist in animals and plants. Functional studies of host transferred sRNAs will help identify important virulence genes in interacting pathogens and pests. Furthermore, transgene-derived sRNAs are also delivered into fungal cells by extracellular vesicles (fig. S16) (*8*). Discovery of exosome function in cross-kingdom RNAi may help develop effective delivery methods of pathogen-targeting artificial RNAs with the goal of controlling plant diseases in various pre- and post-harvesting crops.

REFERENCES AND NOTES

1. D. Baulcombe, *Nature* **431**, 356–363 (2004).
2. M. Mittelbrunn, F. Sánchez-Madrid, *Nat. Rev. Mol. Cell Biol.* **13**, 328–335 (2012).
3. A. Molnar *et al*., *Science* **328**, 872–875 (2010).
4. A. Weiberg, M. Bellinger, H. Jin, *Curr. Opin. Biotechnol.* **32**, 207–215 (2015).
5. M. Wang, N. Thomas, H. Jin, *Curr. Opin. Plant Biol.* **38**, 133–141 (2017).
6. A. Weiberg *et al*., *Science* **342**, 118–123 (2013).
7. A. H. Buck *et al*., *Nat. Commun.* **5**, 5488 (2014).
8. M. Wang *et al*., *Nat. Plants* **2**, 16151 (2016).
9. S. Shahid *et al*., *Nature* **553**, 82–85 (2018).
10. T. Zhang *et al*., *Nat. Plants* **2**, 16153 (2016).
11. M. Wang, A. Weiberg, E. Dellota Jr., D. Yamane, H. Jin, *RNA Biol.* **14**, 421–428 (2017).
12. A. Weiberg, H. Jin, *Curr. Opin. Plant Biol.* **26**, 87–94 (2015).
13. M. Colombo, G. Raposo, C. Théry, *Annu. Rev. Cell Dev. Biol.* **30**, 255–289 (2014).
14. H. Valadi *et al*., *Nat. Cell Biol.* **9**, 654–659 (2007).
15. M. Mittelbrunn *et al*., *Nat. Commun.* **2**, 282 (2011).
16. S. Mathivanan, H. Ji, R. J. Simpson, *J. Proteomics* **73**, 1907–1920 (2010).
17. L. C. Boavida, P. Qin, M. Broz, J. D. Becker, S. McCormick, *Plant Physiol.* **163**, 696–712 (2013).
18. S. Ferrari *et al*., *Plant Physiol.* **144**, 367–379 (2007).
19. M. Ostrowski *et al*., *Nat. Cell Biol.* **12**, 19–30, 1–13 (2010).
20. K. Ebine *et al*., *Nat. Cell Biol.* **13**, 853–859 (2011).
21. S. Levy, T. Shoham, *Nat. Rev. Immunol.* **5**, 136–148 (2005).
22. Z. Xie, E. Allen, A. Wilken, J. C. Carrington, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 12984–12989 (2005).
23. J. S. Bonifacino, A. Hierro, *Trends Cell Biol.* **21**, 159–167 (2011).
24. Y. Liu, R. Mittal, N. V. Solis, N. V. Prasadarao, S. G. Filler, *PLOS Pathog.* **7**, e1002305 (2011).
25. M. O. Steinmetz, A. Akhmanova, *Trends Biochem. Sci.* **33**, 535–545 (2008).
26. Y. Liu, V. A. Bankaitis, *Prog. Lipid Res.* **49**, 201–217 (2010).
27. C. Catalanotto, G. Azzalin, G. Macino, C. Cogoni, *Genes Dev.* **16**, 790–795 (2002).
28. G. LaMonte *et al*., *Cell Host Microbe* **12**, 187–199 (2012).

ACKNOWLEDGMENTS

We thank N. Raikhel for the *ARA6-GFP* seeds and constructive suggestions, L. Boavida for *tet8* and *TET8-GFP* seeds, Z. Ma for the *B. cinerea* knockout vectors, G. Coaker for the CRISPR-Cas9 vector, and N. Thomas for editing the paper. **Funding:** This work was supported by grants from National Institute of Health (R01GM093008), National Science Foundation (IOS1557812), and AES-CE (PPA-7517H) to H.J. **Author contributions:** H.J. conceived the idea, designed the experiments, and supervised the project. Q.C. performed the majority of the experiments and analyzed data. L.Q., M.W., B.H., and J.P. contributed to the functional analysis of the fungal target genes. F.-M.L. and S.-D.H. performed bioinformatics analysis. Q.C. and H.J. wrote the manuscript. **Competing interests:** H.J. is the inventor on a patent application (US62/573,546) held/submitted by University of California, Riverside, covering control of fungal pathogens by using RNAi-based approaches. **Data and material availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. The sequencing data sets were deposited in the NCBI database (PRJNA431815).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1126/suppl/DC1
Materials and Methods
Supplementary Text
Fig. S1 to S16
References (*29*–*43*)
Tables S1 to S6

9 November 2017; accepted 12 April 2018
Published online 17 May 2018
10.1126/science.aar4142

PLANT GENETICS

A selfish genetic element confers non-Mendelian inheritance in rice

Xiaowen Yu[1]*, Zhigang Zhao[1]*, Xiaoming Zheng[2], Jiawu Zhou[3], Weiyi Kong[1], Peiran Wang[1], Wenting Bai[1], Hai Zheng[1], Huan Zhang[1], Jing Li[3], Jiafan Liu[2], Qiming Wang[1], Long Zhang[1], Kai Liu[1], Yang Yu[1], Xiuping Guo[2], Jiulin Wang[2], Qibing Lin[2], Fuqing Wu[2], Yulong Ren[2], Shanshan Zhu[2], Xin Zhang[2], Zhijun Cheng[2], Cailin Lei[2], Shijia Liu[1], Xi Liu[1], Yunlu Tian[1], Ling Jiang[1], Song Ge[4], Chuanyin Wu[2]†, Dayun Tao[3]†, Haiyang Wang[2]†, Jianmin Wan[2]†

Selfish genetic elements are pervasive in eukaryote genomes, but their role remains controversial. We show that *qHMS7*, a major quantitative genetic locus for hybrid male sterility between wild rice (*Oryza meridionalis*) and Asian cultivated rice (*O. sativa*), contains two tightly linked genes [*Open Reading Frame 2* (*ORF2*) and *ORF3*]. *ORF2* encodes a toxic genetic element that aborts pollen in a sporophytic manner, whereas *ORF3* encodes an antidote that protects pollen in a gametophytic manner. Pollens lacking *ORF3* are selectively eliminated, leading to segregation distortion in the progeny. Analysis of the genetic sequence suggests that *ORF3* arose first, followed by gradual functionalization of *ORF2*. Furthermore, this toxin-antidote system may have promoted the differentiation and/or maintained the genome stability of wild and cultivated rice.

Identifying so-called speciation genes that cause reproductive isolation is a central goal in evolutionary biology. Postzygotic reproductive isolation (PRI), embodied by hybrid sterility, inviability, or weakness, drives speciation and maintains species identity by restricting gene flow between populations (*1*, *2*). The Bateson-Dobzhansky-Muller model postulates that hybrid reproductive isolation results from deleterious interactions among at least two loci from evolutionarily divergent populations (*3*). Accumulating evidence suggests that selfish genetic elements (SGEs), DNA sequences that gain a transmission advantage relative to the rest of the genome, could drive genome evolution by causing hybrid incompatibilities and segregation distortion in different organisms (*4*–*10*); however, the role of SGEs in genome evolution and their underlying molecular mechanisms have remained obscure.

We performed quantitative trait locus (QTL) analysis of a backcross F_1 (BC_1F_1) population derived from the cross between two highly divergent rice species—wild rice, *O. meridionalis* accession 82031 (Mer), and *O. sativa* ssp. *japonica*, Dianjingyou1 (DJY1)—and detected four major QTLs controlling hybrid male sterility (fig. S1, A and B). Of these, *qHMS7* is located near several previously identified QTLs (fig. S1C).

To clone the causal gene(s) for *qHMS7*, we developed a near-isogenic line for *qHMS7* (NIL-*qHMS7*) (fig. S1D). Intriguingly, the NIL-*qHMS7* plant can only be maintained in a heterozygous status at the *qHMS7* locus in the DJY1 background. Examination of the self-pollinated progeny of NIL-*qHMS7* (BC_6F_2) revealed a bimodal distribution for pollen fertility (fig. S2). Genotyping revealed that the semi-sterile plants were of the *DJY1/Mer* genotype (*D/M*) and the fully fertile plants were of the *DJY1/DJY1* genotype (*D/D*) at the *qHMS7* locus. No *Mer/Mer* type (*M/M*) plant was detected. The seed-set rate of all 50 randomly selected plants was above 85% (fig. S3), indicating that this locus did not affect female gametes and seed setting. These results suggest that *qHMS7* acts as a single locus conferring male semi-sterility and that the Mer-type pollens were not transmissible to the progeny. This notion was further supported by a gametophytic transmission assay using three genetic populations of DJY1 and NIL-*qHMS7* (table S1). Histological and cellular examination of various stages of pollens revealed that Mer-type pollens aborted before the tricellular stage in NIL-*qHMS7* (Fig. 1, A to E, and figs. S4 to S7), possibly owing to defects in the second mitosis that produces tricellular mature pollen grains.

The gene(s) underlying *qHMS7* locus was delimited to a 31.6-kb genomic interval by the use of a map-based cloning strategy (fig. S8 and table S2). Sequencing analysis identified three genes in the *qHMS7* region in DJY1 (*ORF1*^{D}, *ORF2*^{D}, and *ORF3*), but only two genes in Mer (*ORF1*^{M} and *ORF2*^{M}) (Fig. 1F). *ORF1*^{D}, *ORF1*^{M}, and *ORF3* are predicted to encode homologous, grass family–specific proteins with a mitochondrial targeting signal at the N terminus, whereas *ORF2*^{D} and *ORF2*^{M} are predicted to encode a ribosome-inactivating protein (RIP) domain (*11*) containing protein conserved in monocots, with a putative bipartite nucleus localization signal and a nucleus exporting signal. The ORF2^M and ORF2^D proteins differ at 11 polymorphic sites (figs. S9 to S13).

Quantitative reverse transcription–polymerase chain reaction analysis revealed that expression of both *ORF2* and *ORF3* was significantly higher in stages 11 to 13 of anther development (fig. S14, A and B). *ORF2* expression was lower in NIL-*M/M* plants [near-isogenic line with *M/M* genotype at the *qHMS7* locus in DJY1 background (see below)]

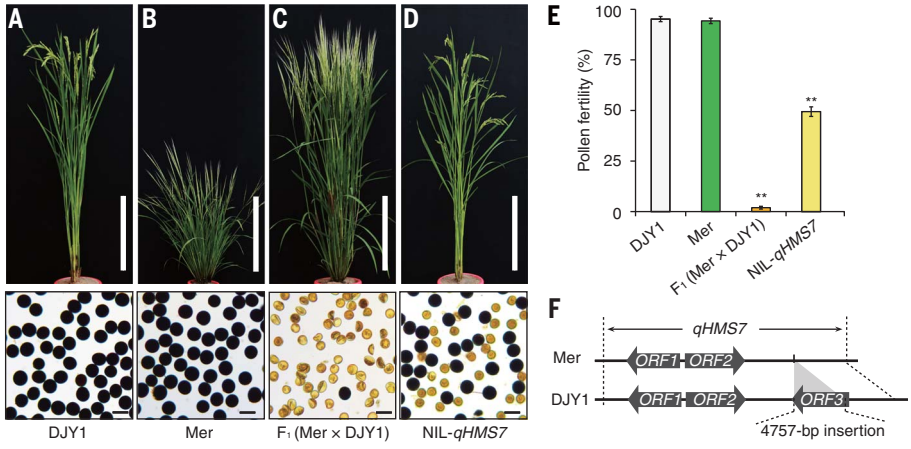


Fig. 1. Identification of *qHMS7*. (**A** to **D**) Morphology (upper panels) and stained pollen (lower panels) of DJY1, Mer, hybrid F_1 (Mer × DJY1), and NIL-*qHMS7*. Fertile pollen is stained dark and sterile pollen is not stained. Scale bars: 50 cm in the upper panels and 50 μm in the lower panels. (**E**) Pollen fertility of DJY1, Mer, hybrid F_1, and NIL-*qHMS7*, shown as mean ± SD (n = 5, 5, 7, and 10 plants examined, respectively). **$P < 0.01$ (versus DJY1, by Student's t test). (**F**) *ORF*s predicted in the fine-mapped region. The insertion of *ORF3* in DJY1 is indicated with a gray triangle. Arrows indicate the orientation of *ORF*s.

[1]National Key Laboratory for Crop Genetics and Germplasm Enhancement, Jiangsu Plant Gene Engineering Research Center, Nanjing Agricultural University, Nanjing 210095, China. [2]National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China. [3]Food Crops Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650200, China. [4]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China.
*These authors contributed equally to this work.
†Corresponding author. Email: wanjianmin@caas.cn (J.Wan); wanghaiyang@caas.cn (H.W.); taody12@aliyun.com (D.T.); wuchuanyin@caas.cn (C.W.)

at mature pollen stage, compared to that in DJY1 (fig. S14C). As expected, *ORF3* expression in DJY1 was about twice as high as that in NIL-*qHMS7* at mature pollen stage, and no expression of *ORF3* was detected in NIL-*M/M* plants (fig. S14D). Transient expression of green fluorescent protein fusions of $ORF2^D$, $ORF2^M$, and ORF3 showed that $ORF2^D$ and $ORF2^M$ are localized to both nucleus and cytoplasm (fig. S15), whereas ORF3 is localized to the mitochondria (fig. S16).

Hybrid sterility is often controlled by toxin-antidote or killer-protector systems that subvert Mendel's law of segregation (*12*). The selective abortion of Mer-type pollens in NIL-*qHMS7* but not in Mer suggests that *ORF3* may encode a protein that functions as an antidote to a pollen-killing toxin produced in NIL-*qHMS7*. To test this notion, we transformed an intact *ORF3* genomic fragment from DJY1 into *D/M*-type calli. As expected, primary (T_0) transgenic plants harboring a single-copy transgene of *ORF3* (genotype *D/M; ORF3/-*) showed a partial restoration of pollen fertility (~75%), compared to the transgene-negative plants (~50% fertility). No effect on pollen fertility was observed in the T_0 transformants of $ORF1^D$ or $ORF2^D$ (table S3). In addition, T_1 plants (derived from single-copy T_0 transformants of *ORF3*) of the genotype (*D/M; –/–*) were semi-sterile, whereas the T_1 plants of the genotype (*D/M; ORF3/-*) showed ~75% fertility, and the T_1 plants of the (*D/M; ORF3/ORF3*) genotype showed normal fertility (~95%). Notably, we recovered T_1 plants of the (*M/M; ORF3/-*) and (*M/M; ORF3/ORF3*) genotypes, and both were fully fertile (Fig. 2 and table S4). These results indicate that *ORF3* has a protective function for Mer-type pollen carrying it (thus acting in a gametophytic manner).

The selective abortion of Mer-type pollens in NIL-*qHMS7* but not in Mer also suggests that the pollen killer is most likely encoded by the *DJY1* allele(s) at the *qHMS7* locus. Consistent with this notion, the T_2 progeny derived from selfed (*M/M; ORF3/-*) type T_1 plants segregated with an expected ratio of 1:2:1 for the genotypes (*M/M; –/–*), (*M/M; ORF3/-*) and (*M/M; ORF3/ORF3*), and they were all normally fertile (fig. S17A and table S5). To identify the gene(s) encoding the predicted toxin, we individually transformed "$ORF1^D$," "$ORF2^D$," and "$ORF1^D$+$ORF2^D$" into calli derived from the seeds of the NIL-*M/M* plants. Transgenic plants harboring "$ORF2^D$" or "$ORF1^D$+$ORF2^D$" were completely male sterile, whereas the "$ORF1^D$" transgenic plants or transgene-negative plants were normally fertile (fig. S17, B and C, and table S6). These observations suggest that $ORF2^D$ has a pollen-killing function and acts in a sporophytic manner (i.e., the pollen-killing function is determined by the parental genotype). This notion was further supported by the observed normal pollen fertility in transgenic NIL-*qHMS7* plants in which $ORF2^D$ was knocked out via CRISPR-Cas9 (fig. S18).

To trace the evolutionary origins of *ORF2* and *ORF3*, we compared genomic sequences among a diverse group of AA-genome rice species (table S7). A single haplotype of *ORF3* was identified in 50.66% of *O. rufipogon* and 94.76% of Asian cultivated rice accessions, but not in other AA-genome rice species, including *O. meridionalis*, *O. longistaminata*, *O. barthii*, and *O. glaberrima* (Fig. 3A and table S7). In comparison, *ORF2* was found in all accessions that we examined. Besides $ORF2^M$ and $ORF2^D$, 25 additional haplotypes were identified (we collectively termed these haplotypes $ORF2^N$) (fig. S19). As *ORF3* is homologous to *ORF1* and *ORF1* was identified in all sequenced accessions, we speculated that *ORF3* might have derived from an ancient gene duplication event in a subpopulation of *O. rufipogon* earlier than the functional $ORF2^D$. As *O. meridionalis* is a basal lineage of AA-genome rice and $ORF2^M$ has the most amino acid substitutions compared to $ORF2^D$, $ORF2^M$ might represent the ancestral sequence from the common ancestor(s) of AA-genome wild rice. The observation that all haplotypes of $ORF2^N$ either lack the protective *ORF3* or contain polymorphisms suggests that they should be nonfunctional in pollen killing, similar to $ORF2^M$. This notion was supported by the detection of putative QTLs for *qHMS7* between DJY1 and different AA-genome species that carry $ORF2^M$ or $ORF2^N$ but lack *ORF3* (table S8). We further deduced that it is likely that $ORF2^M$ gradually evolved into $ORF2^N$ and later evolved into $ORF2^D$ in some subpopulations of *O. rufipogon* through a multistep process (Fig. 3A and fig. S19). It should be noted that although the above data support such a gradual evolution model, the possibility of nonfunctional forms derived from functional *ORF2* and *ORF3* preexisting in the ancestors of wild rice (a degenerative model) cannot be completely excluded yet.

Our combined results suggest that the toxin-antidote system encoded by *ORF2* and *ORF3* constitutes an SGE in rice that confers a transmission advantage to the *ORF3*-carrying male gametes (fig. S20). In the hybrids of different types of wild rice and cultivated rice, pollens of the genotype (*2M;* -) or (*2N;* -) are aborted owing to the lack of protective *ORF3* (table S8, Fig. 3B, and fig. S20). Our results suggest that *qHMS7* may play a role in promoting the differentiation and/or maintaining the genome stability of wild and cultivated rice by restricting gene flow between them. Further, the observed nearly complete male sterility of the *DJY1* × *Mer* hybrid suggests that the combined action of *qHMS7* and other hybrid incompatibility loci could effectively

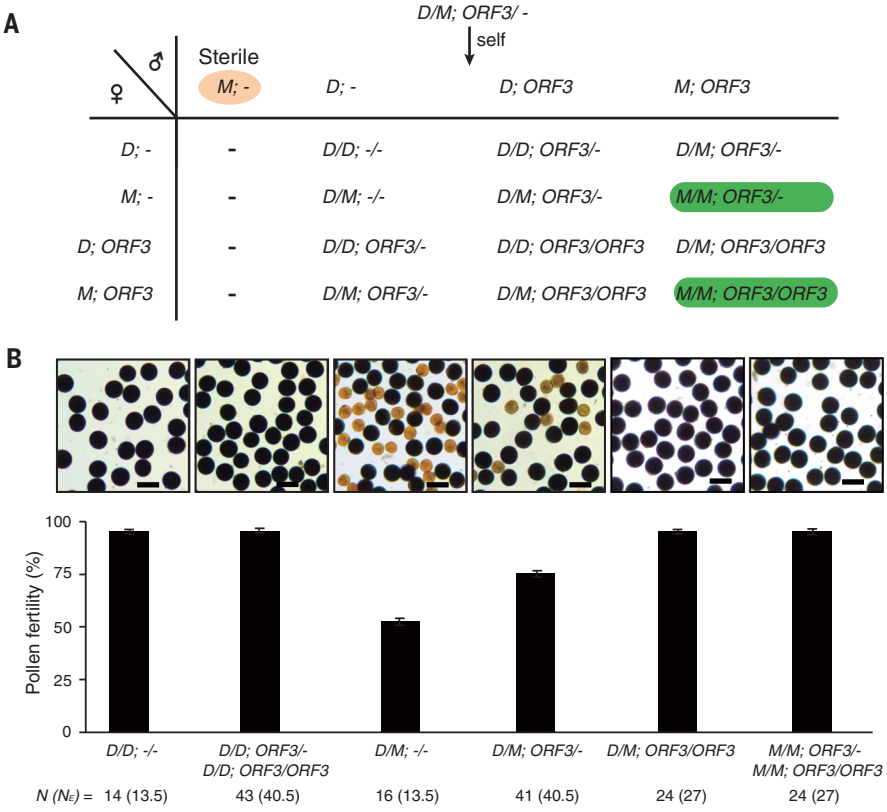


Fig. 2. *ORF3* performs a pollen-protection function in a gametophytic manner. (**A**) A schematic of expected genotypes of the gametes from the *ORF3* T_0 transgenic plants (genotype *D/M, ORF3/-*) and the resulting T_1 progeny from selfing. T_1 plants of the (*M/M; ORF3/-*) and (*M/M; ORF3/ORF3*) genotypes are highlighted in green ovals. (**B**) Genotype and pollen fertility of the T_1 plants from a selfed T_0 transgenic plant (genotype *D/M; ORF3/-*; line Q040). Scale bars, 50 μm. The observed and expected plant numbers of the corresponding genotype among 162 T_1 progeny are shown by *N* and (N_E), respectively. Pollen fertility of the T_1 plants of various genotypes is shown as mean ± SD. The genotype and pollen fertility of two additional T_1 families are shown in table S4.

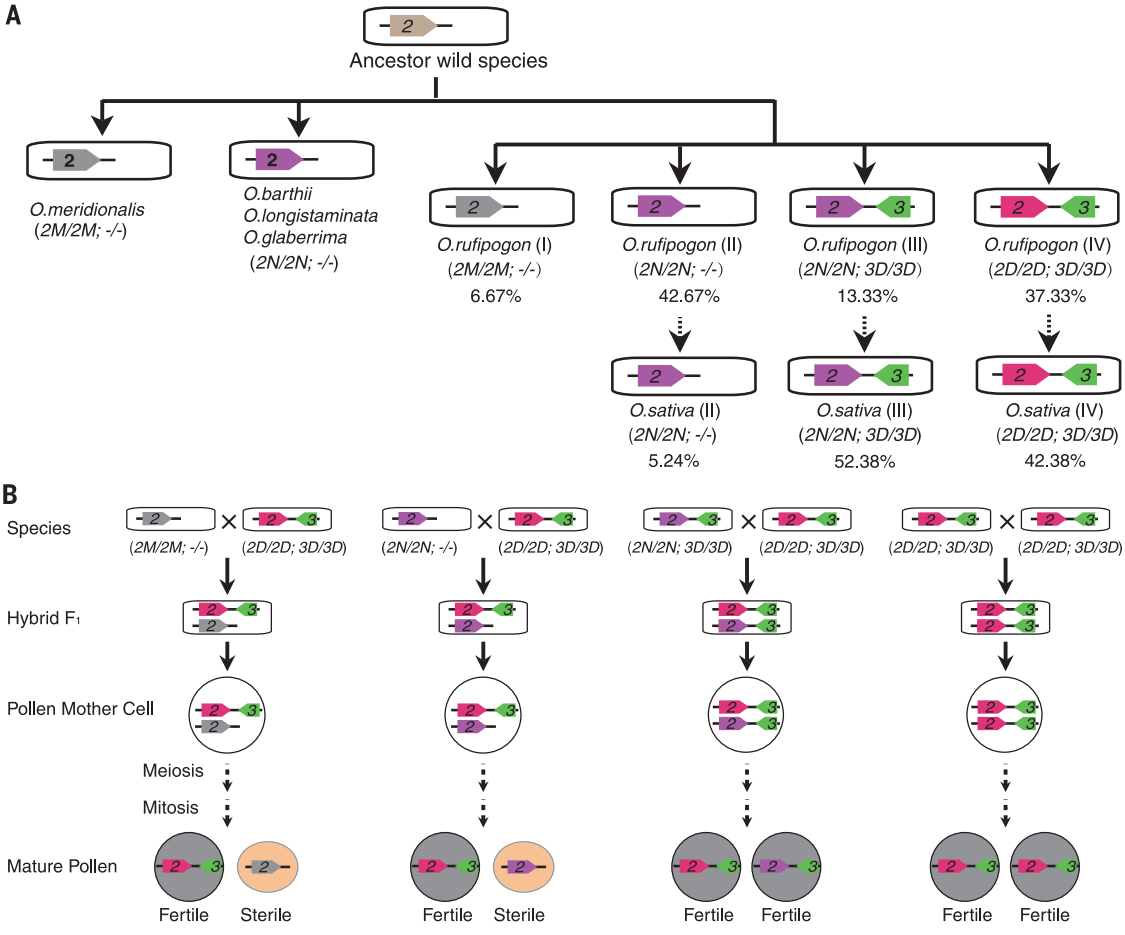


Fig. 3. The evolutionary origin and a genetic action model of *ORF2* and *ORF3*. (**A**) The ancestral wild rice has an ancestral genotype of *ORF2*, but lacks *ORF3*. On the basis of haplotype combinations of *ORF2* and *ORF3*, AA-genome wild rice and cultivated rice can be grouped into four types (I to IV). Their frequencies are shown under each type. $ORF2^{M}$, $ORF2^{N}$, and $ORF2^{D}$ are shown in gray, purple, and wine red, respectively, and *ORF3* is shown in green. (**B**) A diagram shows that in the hybrids of type IV *O. rufipogon* and Asian rice (genotype: *2D/2D; 3D/3D*) with other ancestral AA-genome wild rice or cultivated rice (genotypes *2M/2M; -/-* or *2N/2N; -/-*), pollen of the genotype *2M; -* or *2N; -* is aborted owing to the lack of protective *ORF3*.

prevent gene flow between rice populations (Fig. 1C and fig. S1B). These findings are consistent with and in support of the proposition that SGEs may affect the formation of new species (*13*, *14*).

RIPs are toxic RNA N-glycosidases that affect translation processes and have been implicated in apoptotic pathways in mammalian cells and antiviral, antifungal, and insecticidal activities in plants (*11*). Recent studies also demonstrated that RIPs can induce apoptosis through mitochondrial cascade independent of translation inhibition (*15*). The finding that *ORF2* encoding a RIP domain–containing protein localized to the nucleus and cytoplasm and that *ORF3* encoding an antidote localized to the mitochondria reaffirms a critical role of mitochondria and possibly energy supply for male sterility regulation in plants (*16*). Although the exact spatiotemporal interaction between the "toxin" and "antidote" remains to be elucidated, the identification of *qHMS7* as an SGE regulating PRI in rice adds new evidence supporting a general role of such elements in eukaryote genome evolution and speciation. Notably, several previously reported genes (such as *S5*, *Sa*, and *Sc*) regulating PRI in rice also embody characteristics of SGEs (*7*, *8*, *10*, *17*). As wild rice is an important germplasm resource for hybrid rice breeding, our findings may also offer approaches to overcome male sterility in wild rice–cultivated rice hybrids, thus facilitating utilization of the strong hybrid vigor. In addition, a toxin-antidote system might be modified to fend off infestations of weedy rice, which causes economic losses and poses a threat to ecosystems and biodiversity. Thus, continued efforts to identify SGEs and elucidate their mechanisms have important implications in agriculture.

REFERENCES AND NOTES

1. C. I. Wu, C. T. Ting, *Nat. Rev. Genet.* **5**, 114–122 (2004).
2. H. A. Orr, J. P. Masly, D. C. Presgraves, *Curr. Opin. Genet. Dev.* **14**, 675–679 (2004).
3. D. C. Presgraves, *Nat. Rev. Genet.* **11**, 175–180 (2010).
4. E. Ben-David, A. Burga, L. Kruglyak, *Science* **356**, 1051–1055 (2017).
5. N. L. Nuckolls *et al.*, *eLife* **6**, e26033 (2017).
6. W. Hu *et al.*, *eLife* **6**, e26057 (2017).
7. Y. Long *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18871–18876 (2008).
8. J. Yang *et al.*, *Science* **337**, 1336–1340 (2012).
9. T. Kubo, T. Takashi, M. Ashikari, A. Yoshimura, N. Kurata, *Mol. Plant* **9**, 221–232 (2016).
10. R. Shen *et al.*, *Nat. Commun.* **8**, 1310 (2017).
11. K. Nielsen, R. S. Boston, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**, 785–816 (2001).
12. A. Burt, R. Trivers, *Genes in Conflict: The Biology of Selfish Genetic Elements* (Belknap Press of Harvard Univ. Press, 2006).
13. G. D. Hurst, J. H. Werren, *Nat. Rev. Genet.* **2**, 597–606 (2001).
14. N. Phadnis, *Science* **356**, 1013 (2017).
15. J. M. Lord, M. R. Hartley, *Toxic plant proteins* (Springer Science & Business Media, 2010), vol. 18.
16. L. Chen, Y. G. Liu, *Annu. Rev. Plant Biol.* **65**, 579–606 (2014).
17. Y. Xie *et al.*, *J. Integr. Plant Biol.* **59**, 669–679 (2017).

ACKNOWLEDGMENTS

We thank Q. Yang and L. Han for providing rice germplasm and K. Olsen for discussion. **Funding:** Supported by National Key Research and Development Program of China (2016YFD0100301), National Natural Science Foundation of China (U1502265), and the Agricultural Science and Technology Innovation Program of the Chinese Academy of Agricultural Sciences. **Author contributions:** X.Y. and Z.Z. performed all experiments and analyzed data. X. Zheng, J.Z., W.K., P.W., W.B., H. Zheng, H. Zhang, J. Li, J. Liu, Q.W., L.Z., K.L., Y.Y., X.G., J. Wang, Q.L, F.W., Y.R., S.Z., X. Zhang, Z.C., C.L., S.L., X.L., Y.T., L.J., and S.G. performed some of the experiments. J. Wan, H.W., C.W., and D.T. supervised the project and wrote the paper. **Competing interests:** J. Wan, X.Y., Z.Z., X. Zhang, L.J., S.Z., X.L., S.L., and Y.T. are inventors on patent application (no. 201810164610.3) filed in the State Intellectual Property Office of China by Nanjing Agricultural University that covers ORF3 and ORF2 sequences of the *qHMS7* locus and their applications. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. All of the DNA sequences obtained in this study have been deposited in GenBank (accession numbers MG865759 to MG865763).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1130/suppl/DC1
Materials and Methods
Figs. S1 to S20
Tables S1 to S9
References (*18*–*32*)

7 November 2017; accepted 25 April 2018
10.1126/science.aar4279

MEDICINE

Noninvasive blood tests for fetal development predict gestational age and preterm delivery

Thuy T. M. Ngo[1]*†, Mira N. Moufarrej[1]*, Marie-Louise H. Rasmussen[2], Joan Camunas-Soler[1], Wenying Pan[1], Jennifer Okamoto[1], Norma F. Neff[1], Keli Liu[3], Ronald J. Wong[4], Katheryne Downes[5], Robert Tibshirani[3,6], Gary M. Shaw[4], Line Skotte[2], David K. Stevenson[4], Joseph R. Biggio[7], Michal A. Elovitz[5], Mads Melbye[2,8]‡§, Stephen R. Quake[1]‡§

Noninvasive blood tests that provide information about fetal development and gestational age could potentially improve prenatal care. Ultrasound, the current gold standard, is not always affordable in low-resource settings and does not predict spontaneous preterm birth, a leading cause of infant death. In a pilot study of 31 healthy pregnant women, we found that measurement of nine cell-free RNA (cfRNA) transcripts in maternal blood predicted gestational age with comparable accuracy to ultrasound but at substantially lower cost. In a related study of 38 women (25 full-term and 13 preterm deliveries), all at elevated risk of delivering preterm, we identified seven cfRNA transcripts that accurately classified women who delivered preterm up to 2 months in advance of labor. These tests hold promise for prenatal care in both the developed and developing worlds, although they require validation in larger, blinded clinical trials.

Understanding the timing and programming of pregnancy has been a topic of interest for thousands of years. The ancient Greeks had surprisingly detailed knowledge of the different stages of fetal development; they proposed mathematical theories to account for the timing of important landmarks of pregnancy, including delivery of the baby (*1*–*3*). Although biologists now have detailed cellular and molecular portraits of both fetal and placental development, this knowledge has not yet translated into molecular tests that reliably predict gestational age for individual pregnancies. Blood levels of human chorionic gonadotropin (HCG) and α-fetoprotein are used to detect conception and fetal complications, respectively; however, neither molecule (either individually or in conjunction) establishes gestational age (*4*, *5*).

Ultrasound imaging and/or the patient's estimate of her last menstrual period are typically used to estimate gestational age, but the former can be expensive and the latter can be imprecise. Inaccurate dating sometimes leads to unnecessary induction of labor and Cesarean sections, extended postnatal care, and/or increased medical expenses (*6*–*9*). Current methods to estimate delivery date generally assume normal development and do not account for premature birth, which affects approximately 15 million neonates every year worldwide (*10*). In the United States (*11*), premature birth is the leading cause of neonatal death and complications later in life. Two-thirds of these occur spontaneously, and it would be beneficial to be able to identify which pregnancies are at risk (*12*, *13*). Efforts to identify genetic causes and risk factors have had limited success (*11*, *14*–*17*), and clinically, transvaginal sonographic cervical length (CL) and cervicovaginal fetal fibronectin (fFN) measurements have low positive predictive value (21% for CL and 17% for fFN) and specificity (52% for CL) (*18*).

In previous work, we showed that longitudinal phenotypic changes in both the mother and the fetus could be monitored by noninvasively measuring cell-free RNA (cfRNA) transcripts from fetal tissues in maternal blood (*19*). Here, we investigated whether this methodology can be developed into blood tests that establish gestational age and estimate the risk of preterm birth. In our initial study, we recruited 31 pregnant women from Denmark who agreed to donate a blood sample during each week of their pregnancy, resulting in a total of 521 samples (Fig. 1). All women delivered at full term, defined as gestational age at delivery of ≥37 weeks, and their medical records showed no unusual health changes during pregnancy (table S1). Each sample was analyzed by highly multiplexed real-time polymerase chain reaction (PCR) using a panel of genes with expression specific to the placenta or to the immune system, or highly enriched in the fetal liver (table S2).

We observed that cfRNA measurements over the course of pregnancy demonstrated differing time courses according to tissue of origin (Fig. 2A and fig. S1). As expected, the levels of cfRNA corresponding to genes specific to the placenta and enriched in the fetal liver increased throughout the course of pregnancy, with the exception of cfRNA corresponding to chorionic gonadotropin β subunit (CGB), which decreased from a peak found in the first trimester. Placental cfRNAs and several fetal liver cfRNAs were not detected above the noise floor after delivery, which supports their pregnancy-derived origin; some fetal liver transcripts were also expressed in the adult liver, and we observed a small maternal baseline for this subset. cfRNA measurements corresponding to immune system–related genes increased during gestation and showed a return to measurable baselines after delivery, which supports their predominantly maternal origin. The body mass index of the mother did not significantly affect cfRNA levels (see supplementary text). Using estimates of cfRNA concentrations in blood across all genes and all pregnancies (Fig. 2B and fig. S2), we discovered that genes within each set (i.e., placental, immune, and fetal) were highly correlated with each other [Median Pearson correlation r = 0.79 (placental), 0.79 (immune), 0.74 (fetal), $P < 10^{-14}$] and that placental and fetal cfRNA were weakly cross-correlated (r = 0.47, $P < 10^{-15}$). These findings suggested that cfRNA corresponding to placental genes might provide an accurate estimate of fetal development and gestational age throughout pregnancy.

We then built a random forest model to predict time from sample collection until delivery, using cfRNA measurements as the primary features. We trained and validated this model using data from the Danish cohort from 21 women (n = 306 blood samples) for training, and from 10 women (n = 215 blood samples) for validation. During training, we applied best-subset selection with 10-fold cross-validation repeated 10 times (see supplementary materials) to identify nine cfRNAs that are specific to the placenta (*CGA*, *CAPN6*, *CGB*, *ALPP*, *CSHL1*, *PLAC4*, *PSG7*, *PAPPA*, and *LGALS14*) and together provided equivalent predictive power to the full panel of 51 measured cfRNAs (fig. S3). Our model's predictions agreed with observed values for both training (Pearson correlation r = 0.91, $P < 10^{-15}$) (Fig. 2C) and validation sets (r = 0.89, $P < 10^{-15}$) (Fig. 2D). We also found that model performance improved significantly over the course of pregnancy, as measured by root mean squared error (RMSE) for both training [RMSE = 6.0 (first trimester, T1), 3.9 (second trimester, T2), 3.3 (third trimester, T3), 3.7 (postpartum, PP) weeks] (Fig. 2C) and validation sets [RMSE = 5.4 (T1), 4.2 (T2), 3.8 (T3), 2.6 (PP) weeks] (Fig. 2D). Although distinct subsets of the nine cfRNAs listed above

[1]Departments of Bioengineering and Applied Physics, Stanford University and Chan Zuckerberg Biohub, Stanford, CA 94305, USA. [2]Department of Epidemiology Research, Statens Serum Institute, Copenhagen 2300, Denmark. [3]Department of Statistics, Stanford University, Stanford, CA 94305, USA. [4]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305, USA. [5]Maternal and Child Health Research Center, Department of Obstetrics and Gynecology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA. [6]Department of Biomedical Data Sciences, Stanford University School of Medicine, Stanford, CA 94305, USA. [7]Center for Women's Reproductive Health, Department of Obstetrics and Gynecology, University of Alabama, Birmingham, AL 35294, USA. [8]Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, USA.
*These authors contributed equally to this work. †Present address: Cancer Early Detection Advanced Research Center, Knight Cancer Institute and Department of Molecular and Medical Genetics, Oregon Health and Science University, Portland, OR 97239, USA. ‡These authors contributed equally to this work.
§Corresponding author. Email: quake@stanford.edu (S.R.Q.); mmelbye@stanford.edu (M.M.)

were sufficient to predict time until delivery for subpopulations of women (i.e., nulliparous or multiparous women), we found that all nine genes identified were necessary for accurate prediction across subgroups (see supplementary text).

The model's two most important features, *CGA* and *CGB*, encoding chorionic gonadotropin α and β3 subunits of HCG, are known contributors to pregnancy initiation (*20*) and behaved consistently with what is known from HCG levels during pregnancy (*21*). Other genes included in the model, such as *PAPPA* (pregnancy-associated plasma protein A), are associated with pregnancy risks such as preterm birth (*22*).

We next compared our model to other established tools used to predict gestational age (Fig. 2E). In previous studies, ultrasound and last menstrual period estimates of gestational age, which assume delivery at 40 weeks gestation, fell within 14 days of the observed gestational age at delivery with 57.8% and 48.1% accuracy, respectively (*7*). In this study, for all 31 Danish women, cfRNA estimates of gestational age averaged over a given trimester fell within 14 days of the observed gestational age at delivery with 32% (T2), 23% (T3), and 45% (T2 and T3) accuracy, as compared to 48% (T1) for ultrasound (Table 1). Our results are thus generally comparable to ultrasound measurements, can be performed throughout pregnancy, and do not require a priori physiological knowledge such as the woman's last menstrual period.

Although the first-generation random forest model predicted time until delivery for full-term pregnancies, we were also interested in testing its performance to predict spontaneous preterm delivery (defined as spontaneous delivery earlier than 37 weeks; see supplementary materials). To explore this question, we studied two separate cohorts, one recruited by the University of Pennsylvania (n = 15) and the other by the University of Alabama at Birmingham (n = 23). All of the women in both of these cohorts were already known to be at elevated risk of preterm delivery because they had premature contractions (Pennsylvania) or had a prior spontaneous preterm delivery (Alabama) (Fig. 1, table S1, and supplementary materials). All women in the Alabama cohort and three women in the Pennsylvania cohort received progesterone injections because of a prior spontaneous preterm delivery. All women went into labor spontaneously.

We discovered that although the model validated performance for full-term pregnancies (n = 23, RMSE = 4.3 weeks) in these cohorts, it generally failed to predict time until delivery for preterm deliveries (n = 13, RMSE = 11.4 weeks) (fig. S4). This suggests that the model's content may not account for the various outlier physiological events that may lead to preterm birth. This conclusion is supported by the observation that pharmacological agents designed to stop or slow uterine contractions prevent only a small number of preterm deliveries (*23*, *24*).

To identify cfRNA transcripts that might be able to discriminate a spontaneous preterm delivery from a full-term delivery, we performed unblinded RNA sequencing (RNA-seq) on plasma-derived cfRNA collected from women who delivered at full term (n = 7) and preterm (n = 8) in a preterm-enriched cohort (Pennsylvania) (Fig. 1 and table S1). Analysis of RNA-seq data indicated that 38 genes could separate full-term from preterm births with statistical significance (P < 0.001; see supplementary materials) (Fig. 3A). We then created a PCR panel to measure the 38 cfRNAs identified by RNA-seq and other immune and placental genes (table S2). We confirmed that the differential expression observed using RNA-seq was also observed with quantitative reverse transcription PCR (qRT-PCR) (r = 0.72, $P < 10^{-15}$) (fig. S5).

We then developed a classifier to identify women who are at risk of preterm delivery and found that using the top seven cfRNAs from the panel (*CLCN3*, *DAPP1*, *PPBP*, *MAP3K7CL*, *MOB1B*, *RAB27B*, and *RGS18*) [false discovery rate (FDR) ≤ 5%, Hedges' $g \geq 0.8$] (Fig. 3B) in unique combinations of three (table S3 and supplementary materials) accurately classified 6 of 8 preterm samples (75%) and misclassified only 1 of 26 full-term samples (4%) from the Pennsylvania and Denmark cohorts, with a mean area under the curve (AUC) of 0.86 (Fig. 3C). In validation using a preterm-enriched independent cohort (the Alabama cohort), the test accurately classified 4 of 5 preterm samples (80%) and misclassified 3 of 18 full-term samples (17%) (Fig. 1), with a mean AUC of 0.81 (Fig. 3C), using samples collected up to 2 months in advance of labor. Several of the cfRNAs used to predict spontaneous preterm delivery were also individually elevated in women who delivered preterm (FDR ≤ 5%, Hedges' $g \geq 0.8$), demonstrating the robustness of their effect (Fig. 3B). Note that the differences in cfRNA levels cannot be accounted

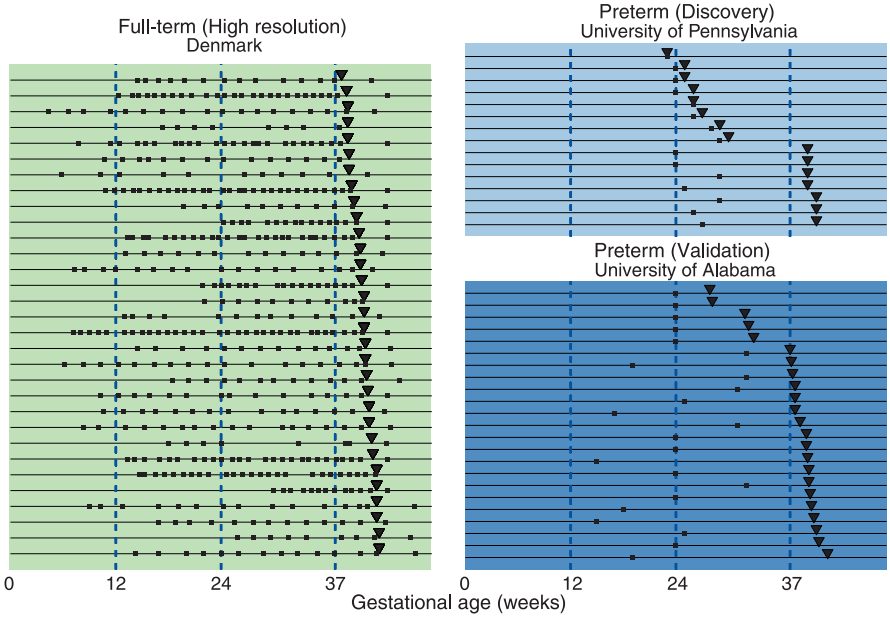


Fig. 1. Sample collection timelines from the Denmark, University of Pennsylvania, and University of Alabama cohorts. Squares, inverted triangles, and lines indicate sample collection times, delivery dates, and individual women, respectively.

Table 1. Comparison of gestational age estimates using cfRNA and ultrasound. Distribution of difference between estimates of gestational age, which assume delivery at 40 weeks gestation, and observed gestational age at delivery listed for four distinct methods, where n indicates the number of women included. Gestational age was estimated using cfRNA measurements from the second (T2), third (T3), or both (T2 and T3) trimesters and ultrasound measurements from the first trimester (T1).

Method	Δ [observed − expected delivery date (weeks)] (%)				
	< −2	−1 to −2	±1	+1 to +2	> +2
cfRNA (T2, n = 28)	50	18	32	0	0
cfRNA (T3, n = 31)	0	6	23	29	42
cfRNA (T2 and T3, n = 31)	19	6	45	10	20
Ultrasound (T1, n = 31)	0	26	48	23	3

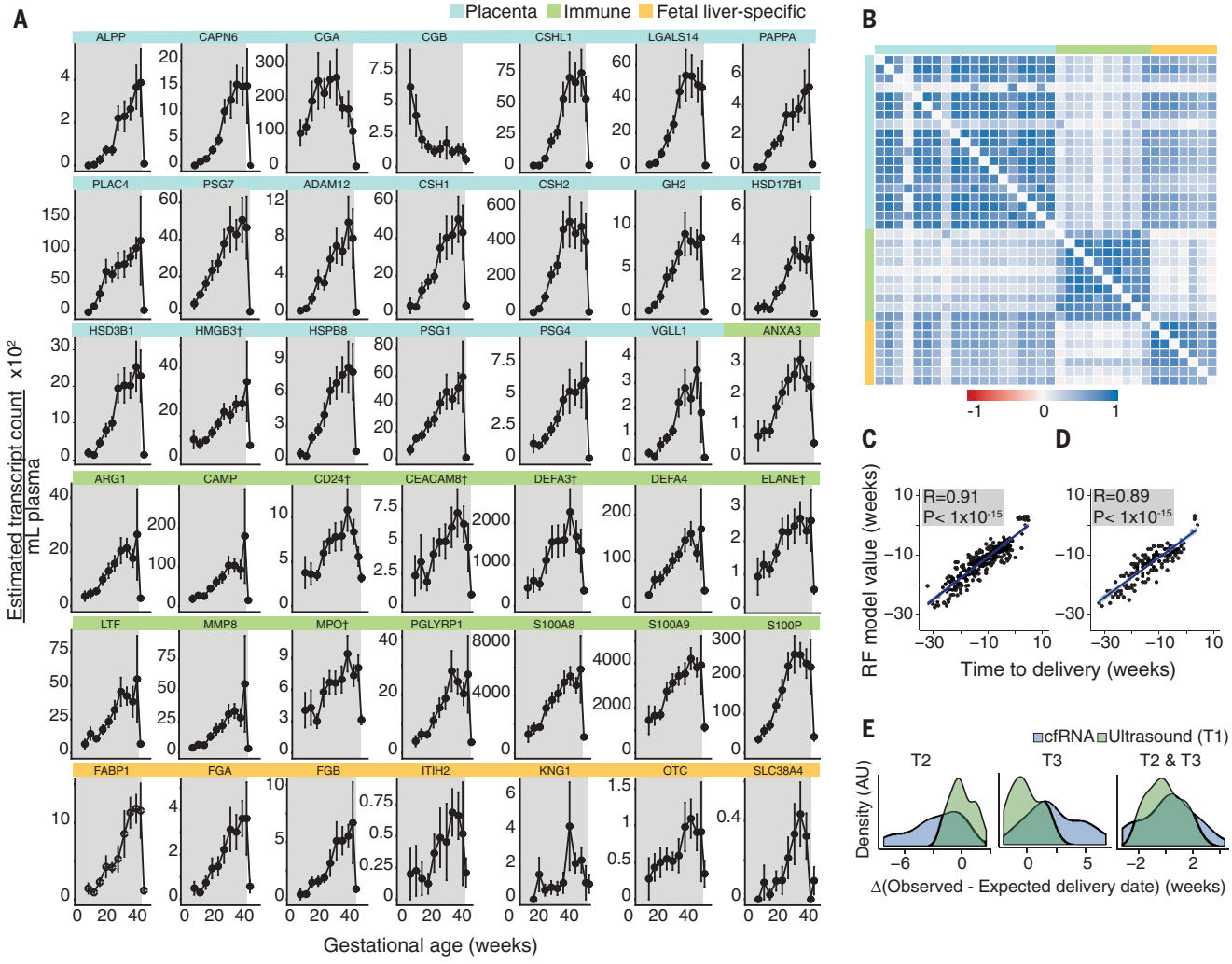


Fig. 2. Application of cfRNA measurements to predict gestational age. (**A**) For each gene, cfRNA transcript count measurements are shown over the course of gestation. Each point represents the mean cfRNA value ± SEM for either 31 women or 21 women (the latter denoted by †). The antepartum period is highlighted in gray. Placental, immune, and fetal liver–specific genes are highlighted in blue, green, and orange, respectively. (**B**) Heat map of the Pearson correlation coefficient for each gene pair shows that placental, immune, and fetal liver–specific cfRNA [same group colors as (A)] measurements are highly correlated with each other [median Pearson correlation r = 0.79 (placenta), 0.79 (immune), 0.74 (fetal liver); $P < 10^{-14}$]. Placental and fetal liver–specific genes also show a weak degree of cross-correlation (r = 0.47, $P < 10^{-15}$). Gene order matches order shown in (A), omitting genes denoted by † in (A). (**C**) Cross-validated random forest (RF) model predicts time to delivery from sampling time point (R = 0.91, $P < 10^{-15}$, n = 21) for training cohort. (**D**) Cross-validated random forest model predicts time to delivery from sampling time point (R = 0.89, $P < 10^{-15}$, n = 10) for validation cohort. (**E**) Distribution of difference in weeks between observed and predicted gestational age at delivery using cfRNA measurements from the second (T2), third (T3), or both (T2 & T3) trimesters (left to right) versus using ultrasound measurements from the first trimester (T1). AU, arbitrary units.

for entirely by progesterone injections, because every woman in the Alabama cohort received injections and the same differences between groups were observed.

Further investigation of the seven genes corresponding to the transcripts identified above revealed that most are ubiquitously expressed, with the exception of a member of the *RAS* oncogene family (*RAB27B*), which encodes a protein that regulates placental development (*25*) and the gene encoding pro-platelet basic protein (*PPBP*). Other protein products encoded by *DAPP1* (dual adaptor of phosphotyrosine 3-phosphoinositides 1) (*26*), *RGS18* (regulator of G protein signaling 18) (*27*), *CLCN3* (chloride voltage-gated channel 3) (*28*, *29*), and *MOB1B* (MOB kinase activator 1B) (*30*) are indirectly implicated in pregnancy through inflammation (*DAPP1*, *RGS18*), labor (*CLCN3*), and development (*MOB1B*).

The cfRNA results can be compared to efforts to estimate preterm risk using mass spectroscopic measurements of the ratio of two proteins in blood [SHBG (sex hormone binding globulin) and IBP4 (insulin-like growth factor binding protein 4)] (*31*) or CL and fFN measurements (*18*). In this pilot study, our blood test yielded higher mean accuracy than the mass spectroscopic approach for comparable sample sizes in the validation cohorts [AUC = 0.81 (cfRNA), AUC = 0.67 (IBP4/SHBG)]. When compared to CL and fFN measurements for symptomatic high-risk women experiencing preterm labor, the PCR-based test had a higher positive predictive value [17% (CL), 21% (fFN), 75% (cfRNA, discovery), 80% (cfRNA, validation)] (*18*).

Our study has important limitations. Before a diagnostic or screening test based on this work can be used in the clinic, a blinded clinical trial with a larger sample size and diverse ethnicities is essential. Our pilot studies included one Caucasian cohort and two African-American cohorts; data from other ethnic groups would

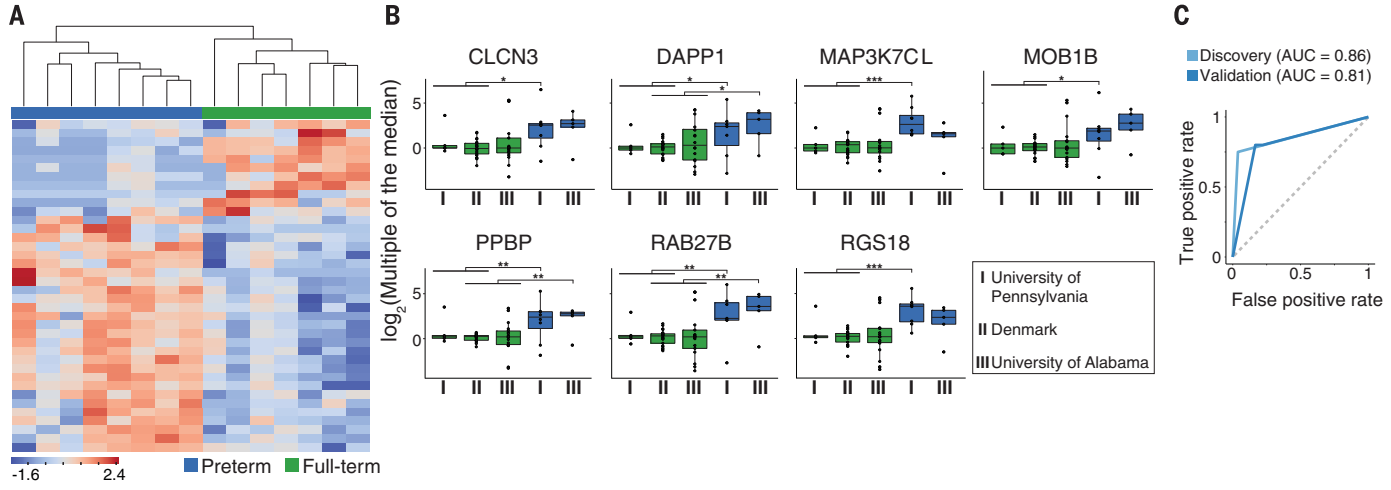


Fig. 3. Application of cfRNA measurements to predict risk of spontaneous preterm delivery. (**A**) Heat map of the *z*-scores for 38 differentially expressed genes identified using cfRNA-seq ($P < 0.001$, exact test, likelihood ratio test, and quasi-likelihood *F* test) shows that genes distinguish women who delivered spontaneously preterm from women who delivered at full term. The two groups of women were separated using hierarchical clustering. (**B**) Means ± SD for differentially expressed genes validated using qRT-PCR in the discovery [University of Pennsylvania (I) and Denmark (II)] and validation [University of Alabama (III)] cohorts. $*P < 0.05$, $**P < 0.01$, $***P < 0.0005$ (Fisher exact test). (**C**) Receiver operating characteristic curves for classifier designed to separate women who deliver spontaneously preterm from women who deliver at full term for both the discovery cohort (University of Pennsylvania and Denmark, AUC = 0.86) and the validation cohort (University of Alabama, AUC = 0.81).

be valuable. Another limitation is that the preterm risk cohorts were all recruited on the basis of elevated risk for preterm birth; it will be important to investigate the performance of the blood test in a broader, unselected population.

Our pilot studies have shown that noninvasive blood tests were able to predict gestational age and identify women at risk of preterm delivery from the same blood sample. These cfRNA PCR-based tests have two advantages over alternatives: broader applicability and lower cost. They can be applied across the globe as a complement to or substitute for ultrasound, which can be expensive and inaccurate during the second and third trimesters (*32*). Conceivably, similar approaches will prove to be useful for identifying and monitoring fetuses with congenital defects that can be treated in utero—a rapidly growing area of fetal medicine.

REFERENCES AND NOTES

1. A. E. Hanson, *Clio Med.* **27**, 291–307 (1995).
2. A. E. Hanson, *Bull. Hist. Med.* **61**, 589–602 (1987).
3. H. N. Parker, *Class. Q.* **49**, 515–534 (1999).
4. L. Dugoff *et al.*, *Obstet. Gynecol.* **106**, 260–267 (2005).
5. E. Yefet, O. Kuzmin, N. Schwartz, F. Basson, Z. Nachum, *Fetal Diagn. Ther.* **42**, 285–293 (2017).
6. W. B. Barr, C. C. Pecci, *Int. J. Gynaecol. Obstet.* **87**, 38–39 (2004).
7. D. A. Savitz *et al.*, *Am. J. Obstet. Gynecol.* **187**, 1660–1666 (2002).
8. K. A. Bennett *et al.*, *Am. J. Obstet. Gynecol.* **190**, 1077–1081 (2004).
9. M. Whitworth, L. Bricker, C. Mullan, *Cochrane Database Syst. Rev.* **2015**, CD007058 (2015).
10. H. Blencowe *et al.*, *Reprod. Health* **10** (suppl. 1), S2 (2013).
11. Institute of Medicine (US) Committee on Understanding Premature Birth and Assuring Healthy Outcomes, *Preterm Birth: Causes, Consequences, and Prevention* (National Academies Press, 2007).
12. C. J. Murray *et al.*, *Lancet* **380**, 2197–2223 (2012).
13. L. Liu *et al.*, *Lancet* **379**, 2151–2161 (2012).
14. K. Ward, V. Argyle, M. Meade, L. Nelson, *Obstet. Gynecol.* **106**, 1235–1239 (2005).
15. T. P. York, J. F. Strauss 3rd, M. C. Neale, L. J. Eaves, *Twin Res. Hum. Genet.* **12**, 333–342 (2009).
16. G. Zhang *et al.*, *N. Engl. J. Med.* **377**, 1156–1167 (2017).
17. L. J. Muglia, M. Katz, *N. Engl. J. Med.* **362**, 529–535 (2010).
18. R. Arisoy, M. Yayla, *J. Pregnancy* **2012**, 201628 (2012).
19. W. Koh *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 7361–7366 (2014).
20. R. B. Jaffe, P. A. Lee, A. R. Midgley Jr., *J. Clin. Endocrinol. Metab.* **29**, 1281–1283 (1969).
21. M. Cocquebert *et al.*, *Am. J. Physiol. Endocrinol. Metab.* **303**, E950–E958 (2012).
22. K. Spencer, N. J. Cowans, F. Molina, K. O. Kagan, K. H. Nicolaides, *Ultrasound Obstet. Gynecol.* **31**, 147–152 (2008).
23. R. Romero, S. K. Dey, S. J. Fisher, *Science* **345**, 760–765 (2014).
24. A. Conde-Agudelo, R. Romero, *Am. J. Obstet. Gynecol.* **214**, 235–242 (2016).
25. J. Davila *et al.*, *PLOS Genet.* **11**, e1005458 (2015).
26. S. Hou, S. D. Pauls, P. Liu, A. J. Marshall, *Mol. Immunol.* **48**, 89–97 (2010).
27. Z. Xie, E. C. Chan, K. M. Druey, *AAPS J.* **18**, 294–304 (2016).
28. S. Yamamoto-Mizuma *et al.*, *J. Physiol.* **557**, 439–456 (2004).
29. M. J. Berridge, *J. Physiol.* **586**, 5047–5061 (2008).
30. H. Sasaki *et al.*, *Clin. Lung Cancer* **8**, 273–276 (2007).
31. G. R. Saade *et al.*, *Am. J. Obstet. Gynecol.* **214**, 633.e1–633.e24 (2016).
32. D. G. Altman, L. S. Chitty, *Ultrasound Obstet. Gynecol.* **10**, 174–191 (1997).

ACKNOWLEDGMENTS

We thank D. Croote and M. Kowarsky for helpful discussions. **Funding:** Supported by the Bill and Melinda Gates Foundation, the March of Dimes Prematurity Research Center at Stanford University, the March of Dimes Prematurity Initiative Grant at the University of Pennsylvania, and the Chan Zuckerberg Biohub. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material. **Author contributions:** M.-L.H.R., R.J.W., K.D., G.M.S., L.S., D.K.S., J.R.B., M.A.E., and M.M. contributed to sample collection; M.M. and S.R.Q. developed the idea and conceptualized the study in collaboration with T.T.M.N., M.N.M., J.C.-S., G.M.S., and D.K.S.; T.T.M.N., J.C.-S., W.P., J.O., and N.F.N. processed and analyzed samples and developed experimental protocols; M.N.M., T.T.M.N., J.C.-S., K.L., R.T., and S.R.Q. developed statistical tools and analyzed the data; and all authors contributed to the writing and editing of the manuscript. **Competing interests:** M.N.M., S.R.Q., M.M., T.T.M.N., and J.C.-S. are inventors on a patent application (number 62/578,360) submitted by the Chan Zuckerberg Biohub that covers noninvasive estimates of gestational age, delivery, and preterm birth. The authors declare no other competing interests. **Data and materials availability:** cfRNA sequencing data have been deposited in the Sequence Read Archive (SRA) under study accession number SRP130149. qRT-PCR raw and processed data and code used in this manuscript are available at https://github.com/miramou/pregnancy_cfRNA.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/360/6393/1133/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S5
Tables S1 to S3
References (*33*–*40*)

1 November 2017; accepted 20 April 2018
10.1126/science.aar3819

Celebrating 90 years
Research milestones from SIMIT


Old campus, built in 1933

In the early 20th century, China's modern intellectual elites formed the "New Culture Movement" and began incorporating a Western-style research paradigm. In June 1928, under the leadership of famed educator Cai Yuanpei, they assembled in Shanghai and together founded the Academia Sinica. The Shanghai Institute of Microsystem and Information Technology (SIMIT), at the Chinese Academy of Sciences (CAS), was one of the first engineering research institutes of Academia Sinica and was also founded that same year.

Over the past 90 years, the institute has consistently followed the same principle: to better apply S&T in order to meet national strategic priorities and economic growth demands. Over the last 20 years, its research philosophy has shifted from pure scientific research to more applied, mission-oriented research. By combining the objectives and interests of the institute with the broader needs of Chinese society, SIMIT is devoted to driving China's technology innovation and adding to its rich tradition and heritage.

Research at SIMIT

SIMIT has three national-level key laboratories (the State Key Laboratory of Transducer Technology, the State Key Laboratory of Functional Materials for Informatics, and the National Key Laboratory of Microsystem Technology) and two CAS key laboratories (the CAS Key Laboratory of Wireless Sensor Networks and Communications, and the CAS Key Laboratory of Terahertz Solid-State Technology). Moreover, it includes the Center for Excellence in Superconducting Electronics (CENSE), a CAS-affiliated center aiming to be a global competitor in superconductor technology.

The mission of scientific research at SIMIT is to carry out strategic, innovative, and forward-looking studies, promote breakthroughs in key technologies and integrated innovation, provide systematic solutions, and consequently contribute to the development of information technology, advanced materials, and micro-/nanotechnologies. SIMIT fully exploits its advantages in two disciplines—electronics science and technology, and information and communications technology—to create a development strategy consisting of "three major breakthroughs and five top priorities."

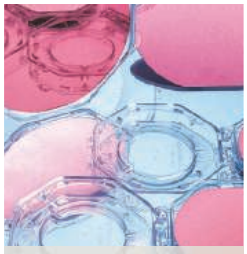
Three major breakthroughs

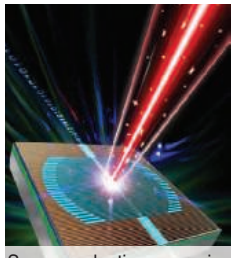
Intelligent sensing microsystems (ISMs)

As China's pioneer in wireless sensor networks and the Internet of Things (IoT), SIMIT is devoted to providing comprehensive solutions for public security, smart city planning, and industry IoT. The institute boasts a wide range of competencies, from sensor development and equipment to telecommunications, data processing, networking, and much more, and uses many powerful tools, such as big data and artificial intelligence, to meet its research goals.

Superconducting quantum devices and circuits

SIMIT is devoted to fundamental scientific research on novel quantum materials, such as superconducting heterostructures, sensors, detectors, and digital circuits. It is focused on developing China's proprietary advanced superconducting core electronic devices, and producing integrated systems with superior performance advantages compared to conventional technologies.


Silicon-on-insulator (SOI) wafers


Superconducting nanowire single-photon detector


The forestry sensors and IoT

Advanced silicon-based materials and applications

SIMIT is devoted to the research, key technology development, and industrialization of 12-in. silicon wafers and engineered substrates, such as fully depleted silicon-on-insulators (SOI), radio frequency (RF) SOI, power SOI, microelectromechanical system (MEMS) SOI, radiation-hardened SOI, gallium nitride (GaN) on silicon, etc., for China's semiconductor and microelectronics industries. The institute is also working intensively on the development of electronic and optoelectronic circuits based on SOI materials.

Five top priorities

Special broadband wireless communication technologies and equipment

SIMIT gives priority to advanced broadband wireless communication technologies, such as concurrent transmission using orbital angular momentum and signal energy, for emergency management, high-speed railways, and smart power grids.

Micro- and nanotechnologies for sensors and transducers

SIMIT gives priority to research on state-of-the-art microelectromechanical-system/nanoelectromechanical-system (MEMS/NEMS) technologies for physical, chemical, biochemical, implantable, and combo sensors, as well as hyperspectrum detection microsystems.

Phase-change random-access memory (PCRAM) and applications

SIMIT gives priority to researching phase-change materials, engineering phase-change random-access memory (PCRAM) chips, and understanding the underlying mechanisms responsible for the low power, high speed, and high reliability of these products.

Terahertz (THz) solid-state technologies

SIMIT gives priority to research on THz quantum cascade lasers, quantum-well photodetectors, interactions between THz waves and various substances, millimeter-wave monolithic integrated circuits (ICs), and inspection imaging equipment with terahertz/millimeter waves.

Brain-inspired chips and bionic vision

SIMIT gives priority to R&D on brain-inspired chips and bionic vision systems, including binocular image signal processing (BISP) chips, sensors that mimic our five sense organs, and robotic limbs.

New campus, built in 2014

Graduates

At modern China's oldest national institute, a mission unchanged since 1928, and a deep commitment to S&T development.

Achievements

SIMIT has been the recipient of 48 national science and technology awards in the past 90 years, including a special National Science and Technology Progress Award for "Fabrication Technology for Type A Separation Membrane," and two First Class National Science and Technology Progress Awards for "High-Speed and Ultra-High-Speed Bipolar Digital Integrated Circuits" and "Research and Industrialization of Silicon-On-Insulator (SOI) Materials." Additionally, SIMIT won the Second Prize of the National Natural Science Award for developing the Lei-Ding Equation, an equilibrium theory of carrier transport in semiconductors.

Because of its advances in wireless sensing and broadband communications, SIMIT made significant contributions to relief efforts after the tragic earthquakes in Wenchuan and Yushu counties. Other achievements include intrusion-prevention systems at the Shanghai World Expo and Pudong International Airport, a security system for the central line South-to-North Water Diversion Project, and other IoT application demonstration projects.

Within the last decade, SIMIT has developed a series of high-performance superconducting nanowire single-photon detectors (SNSPDs), including the recent success of a compact cooler–based SNSPD with a detection efficiency of over 90%. Superior performance has been demonstrated in many applications, such as the record 404-km, measurement-device-independent quantum key distribution achieved in cooperation with our partners at the University of Science and Technology in Hefei, China.

SIMIT has conducted research into SOI technology for more than 30 years. It has incubated China's sole SOI manufacturing base, Shanghai Simgui Technology Co., Ltd., and developed the country's first SOI wafer. Through international collaboration, Simgui products have expanded to include most SOI application areas, including power SOI, RF SOI, and radiation-hardened SOI materials, among others.

Collaborative innovation

SIMIT has established the "Three-in-One" Collaborative Innovation System, with itself as the core body of advanced research, the Shanghai Industrial μTechnology Research Institute (SITRI) as the technology transfer and pilot platform, and SIMIC Holdings as the industrialization platform focused on capital operation.

SITRI was established in 2013 as a new type of international innovation center committed to accelerating the innovation and commercialization of More-than-Moore (MtM) technologies to power IoT systems. SITRI served as the technology transfer and pilot platform for partners and startups, and was approved as the primary body for the creation of the Shanghai Technology Innovation Center by the Chinese State Council.

SIMIC Holdings is the industrialization and asset management platform of SIMIT, and consists of technology R&D (SIMIC R&D), technology incubation (SIMIC Incubator), venture capital (SIMIC Capital), and an industrial cluster (SIMIC Industry) to support the rapid development of innovative high-tech enterprises.

Talent

To meet the strategic goals of the 13th Five-Year Plan, many changes have been made—both macro changes at the institute level and microadjustments at the laboratory level—to recruit more talent to join the SIMIT family. Over the years, SIMIT has built a dedicated, highly educated team of professionals who have made important contributions in their respective research fields—including two CAS academicians, one foreign associate of the U.S. National Academy of Sciences, and other experts recognized by various national talent programs.

As of the end of 2017, the institute had 1,304 full-time staff members, including 610 full-time research employees (of whom 87 were professors and 125 were associate professors), as well as 644 graduate students and 50 postdoctoral researchers.

At SIMIT, we guarantee that the talents of staff members, students, and postdoctoral researchers will be proudly supported. We will encourage mutual understanding, trust, tolerance, respect, and appreciation among everyone, so that a positive culture will be established at the institute.

International cooperation

SIMIT continually seeks international collaboration and exchanges. We have exchange programs and collaboration projects with research institutes worldwide—in countries including the United States, Japan, France, and Canada, among others—in areas such as advanced silicon-based materials, micro-/nanosensor technology, bioelectronics, terahertz solid-state technology, PCRAM technology, superconductivity, etc. For example, our collaboration with Forschungszentrum Jülich GmbH, an affiliate of the Helmholtz Association, has led to the establishment of the Virtual Joint Institute on Functional Materials and Electronics.

Future

Looking to the future with an open-minded attitude, SIMIT aims to leverage its strengths in the disciplines of electronics science and technology, and information and communications technology, to achieve further success in research, development, and industrialization. By addressing key challenges in the areas of intelligent sensing microsystems, superconducting devices and circuits, and advanced silicon-based materials, we hope to push the boundaries of innovation and the wider adoption of cutting-edge technologies. Our goal is to become an internationally recognized center for research excellence. Meanwhile, we wish to make a significant contribution to the development of information technology, network technology, and communications technology in China.

CONTACT INFORMATION
www.sim.ac.cn
Tel: +86-21-62511070
Fax: +86-21-62524192

NEW ENGLAND BioLabs® Inc.

be INSPIRED
drive DISCOVERY
stay GENUINE

We've spread our wings.

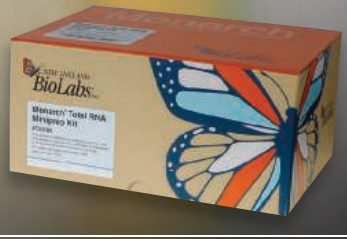
Monarch® Nucleic Acid Purification Kits Now available for DNA & RNA

Designed with sustainability in mind, Monarch® Nucleic Acid Purification Kits are the perfect complement to many molecular biology workflows. Available for DNA & RNA purification, with buffers and columns available separately, Monarch kits are optimized for excellent performance, convenience and value. Quickly and easily recover highly pure, intact DNA and RNA in minutes. Available kits include:

- Monarch Plasmid Miniprep Kit
- Monarch DNA Gel Extraction Kit
- Monarch PCR & DNA Cleanup Kit (5 µg)
- **NEW** MONARCH TOTAL RNA MINIPREP KIT – optimized for use with a variety of sample types, including cells, tissues, blood, and more

Make the change and migrate to Monarch today.

Interested in trying a sample of our new Monarch Total RNA Miniprep Kit?



Learn more at
NEBMonarch.com

"Global Ocean Summit (GOS)" is a high-level forum initiated by Qingdao National Laboratory for Marine Science and Technology (QNLM). Focused on the theme of "Building Partnerships on Ocean Observation and Research", the Summit explores solutions to challenges in marine science and technology resulted from economic development and global environmental change. It advocates for setting up an internationally open network, strengthening collaboration in marine research and observation to inform its use and sustainability, ensuring equal access to information on ocean preservation and utilization, promoting the sharing of observation and education resources and advancing the sharing of infrastructure and facilities.

Following the 1st GOS held in Qingdao, China during 26-28 September 2016, as the 2nd GOS, GOS2018 will be scheduled at Qingdao, China during 3-5 July 2018, and co-organized by QNLM, Science/AAAS and Department of Science & Technology of Shandong Province and supported by Qingdao People's Association for Friendship with Foreign Countries.

QNLM, as the initiator of GOS and the unique national laboratory in China, is launched in 2015 with Professor Lixin WU, Fellow of the Chinese Academy of Sciences, as its director. Being committed to build a world-class comprehensive research center of marine science and technology as well as an open platform for collaborative innovation, besides constructing the functional labs, joint labs, open studios, and centers for facilities and infrastructure collaborative with research institutions and enterprise domestic and overseas, QNLM jointly constructs the international Laboratories and centers with world-class institutions, including the Centre for Southern Hemisphere Oceans Research (CSHOR), the International Laboratory for High Resolution Earth System Prediction (iHESP), and the being constructed international joint center with Woods Hole Oceanographic Institute. More information about QNLM can be referred to www.qnlm.ac.

GOS2018 will inherit and develop the spirits of the "**2016 Summit Declaration**". Focusing on the global ocean observation and prediction, deep-sea research, polar seas research, and ocean sustainability, the event will be an opportunity to improve the sharing of technology and experience, promote the sharing of marine research infrastructure and facilities, and facilitate the initiating of international program, and finally contribute to a community of common destiny.

We sincerely expect the leaders and representatives to join this big event with us and share your experience. The participants are encouraged to give your poster and/or video exhibition in the meeting on your insight relative to the four topics mentioned above. The size requirement for the poster is 93cm x 192cm, and the ratio for the video is 16:9 with MP4 format. It would be very appreciated if you could send us the message for your poster and video exhibition to the email address jwzheng@qnlm.ac before 20 June, 2018.

Participants are encouraged to register for the conference as early as possible to facilitate the conference organizing committee in arranging the accommodation and to ensure that your reservation requirements are met.

Freezer

The CryoCube F740 series by Eppendorf makes ultralow temperature sample storage future-proof. An increased capacity of 740 L/576 boxes, improved energy efficiency without sacrificing temperature accuracy, and considerable noise reduction create a safe storage environment. Options such as air-cooling or water-cooling, right-handed or left-handed doors, and three or five interior compartments set new benchmarks for user demands. The CryoCube F740 is network-compatible and can be monitored through Eppendorf's smart VisioNize system—directly at the device's touchscreen or flexibly at a computer or tablet in the office.

Eppendorf
For info: 800-645-3050
www.eppendorf.com

Fluorescent Staining Kits for *C. elegans*

NemaMetrix has a series of RediStain fluorescent kits for improved visualization of *C. elegans* structure and function in phenotyping studies. Used with the company's ScreenChip System, the new staining kits provide a more efficient way to study nematode phenotypes and their directly related underlying genotypes, which would be impossible with nonfluorescent techniques. The ScreenChip System allows researchers to rapidly generate and collect precise whole-animal data via quick, reliable assessment of neuronal and physiological responses to genetic mutations, drug effects, and other environmental changes. It provides researchers an alternative to help them better understand human diseases and explore potential treatments for high-impact disorders. The ScreenChip System is currently being used in aging, neurodegenerative, and cardiac disease studies, as well as drug development and developmental biology.

NemaMetrix
For info: 844-663-8749
nemametrix.com/c-elegans-fluorescent-kits

Custom Antibody–Dye Combinations

Complete your artfully crafted panel with BD OptiBuild custom reagents. Procuring custom antibody–dye combinations is as easy and convenient as ordering a standard catalog item. Whether you want to minimize compensation or add new markers to complex experiments, BD OptiBuild reagents provide flexibility to evaluate new colors and simplify your panel design. The reagents are available in small, 50-µg vial sizes, and ship within 72 hours of your order. We've just added over 2,800 high-quality antibodies to our current offerings, expanding your options for multicolor panel design by 4,500 new reagents.

BD Biosciences
For info: 877-232-8995
www.bdbiosciences.com

Lentivirus Titer Kit

Origene offers a One-Wash Lentivirus Titer Kit, for HIV-1 p24 ELISA assays. Lentivirus is a powerful gene delivery tool. To ensure efficient gene expression/knockdown, the optimal MOI (multiplicity of infection) for specific cells needs to be tested; therefore lentiviral titer needs to be measured. Most lentiviral vectors are derived from HIV-1, and HIV-1 p24 is the viral capsid protein. p24 ELISA assays are developed to measure the amount of p24 in the lentivirus samples, which is proportional to the amount of lentiviral particles. OriGene's One-Wash titer kit has a simplified workflow; it uses single washing versus 4X washing and takes 3 hours versus overnight. The kit is meant for room-temperature handling—there is no need for 37°C incubation.

OriGene
For info: 888-267-4436
www.origene.com

Apoptosis Primary Antibodies

Rockland offers a comprehensive variety of antibodies and antibody-based tools for the numerous methods needed in apoptosis and cell-survival measurement. Its products are designed and tested to perform in a variety of assays, including flow cytometry, immunohistochemistry, Western blot, and immunoprecipitation. Each step of the antibody development process is completed within Rockland's in-house, controlled environment, delivering unsurpassed repeatability in antibody construction. Apoptosis is a genetically programmed process for the elimination of damaged or redundant cells by activation of caspases (aspartate-specific cysteine proteases). It is the balance between the pro-apoptotic and pro-survival signals that eventually determines whether cells will undergo apoptosis, survive, or proliferate. Apoptosis is a vital component of normal cell turnover, proper development and functioning of the immune system, hormone-dependent atrophy, embryonic development, and chemical-induced cell death.

Rockland Immunochemicals
For info: 800-656-7625
www.rockland-inc.com

Cell-Line Authentication Service

Now more than ever, it is vital to authenticate your cell lines prior to publication. At ATCC, we have the knowledge to help you do that with our easy-to-use Short Tandem Repeat (STR) Profiling Service. Cell-line authentication is a set of tests intended to demonstrate that a cell line is what it is reported to be. STR profiling is a form of intraspecies cell-line authentication that looks for the variation of tandem repetitive sequences in your cell line; it is rapid, reproducible, and uses a standardized PCR-based method. ATCC-trained experts amplify 17 STR loci plus amelogenin. A comprehensive analysis interprets both karyotypically normal and abnormal cell lines. A no-fuss sample mailer packet is provided so that cells can be spotted onto FTA paper and mailed back to ATCC using the preaddressed envelope. ATCC's STR report meets all requirements for funding, publication, and quality control.

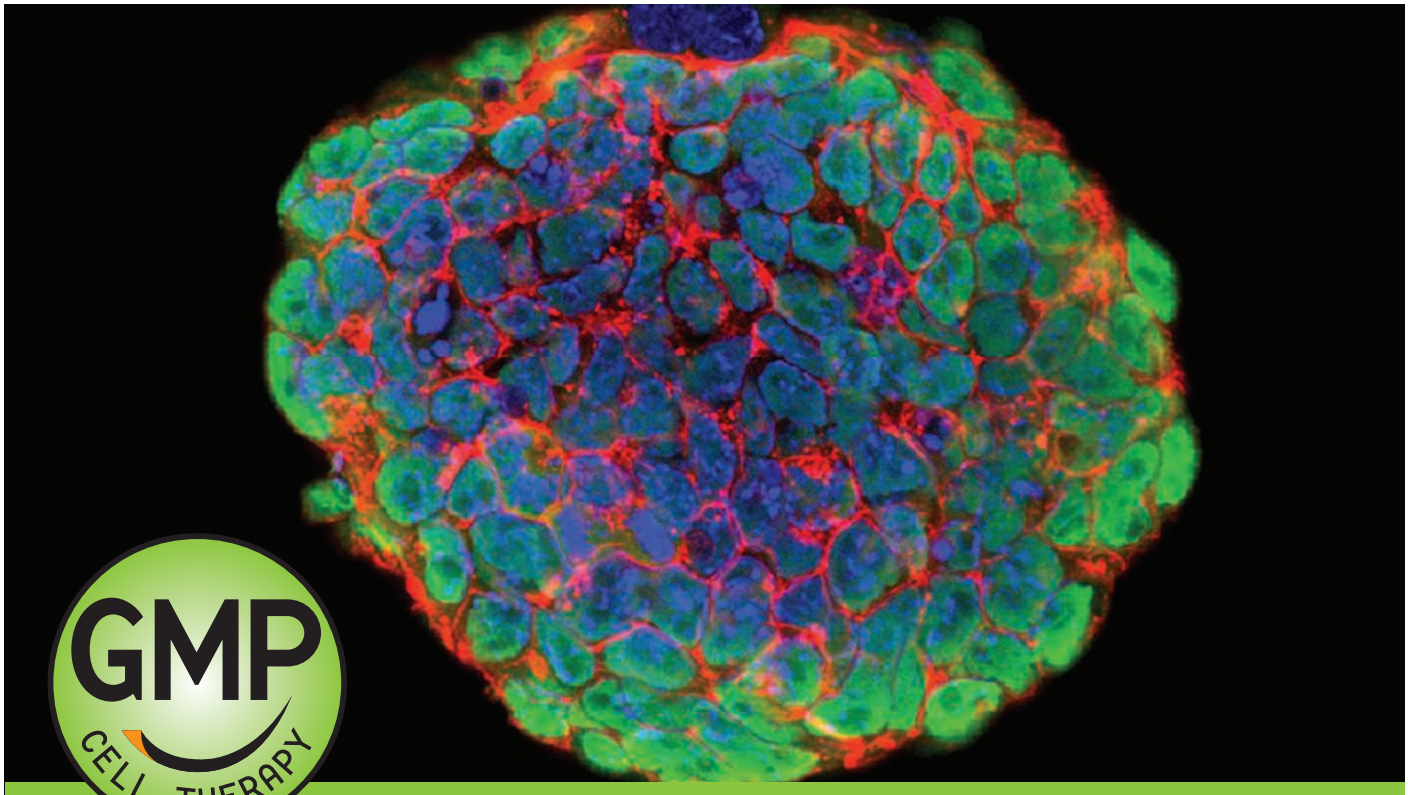
ATCC
For info: 800-638-6597
www.atcc.org

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

R&D SYSTEMS
a bio-techne brand

GMP-Grade Growth Factors



Dedicated to Consistent Proteins

The Widest Selection for

- Cell Manufacturing
- Immunotherapy
- Regenerative Medicine
- Stem Cell Therapy

Learn more | rndsystems.com/gmp

bio-techne®

Global bio-techne.com info@bio-techne.com TEL +1 612 379 2956 **North America** TEL 800 343 7475
Europe | Middle East | Africa TEL +44 (0)1235 529449 **China** info.cn@bio-techne.com TEL +86 (21) 52380373

THE GEORGE WASHINGTON UNIVERSITY

School of Medicine & Health Sciences

Chair
Department of Microbiology, Immunology, & Tropical Medicine

The George Washington University School of Medicine and Health Sciences (SMHS) invites applications and nominations for the position of chair of the Department of Microbiology, Immunology, & Tropical Medicine (MITM).

The Department of MITM (https://smhs.gwu.edu/microbiology/) has an international reputation for advancing knowledge and developing treatments for some of the most socioeconomically devastating infectious diseases in the world. These efforts are supported by our Research Center for Neglected Diseases of Poverty and Research Center for the Cure and Eradication of HIV. Our faculty conduct basic and translational research focused on HIV/AIDS; neglected tropical parasitic diseases, such as hookworm infection, schistosomiasis, and toxoplasmosis; immunology and vaccine development; and infection-induced cancers.

The Department is currently home to 13 tenured/tenure-track faculty members and six faculty members with research appointments. Their research is conducted in a recently renovated, 25,000 square foot state-of-the-art facility developed with support from a $15.6 million NIH infrastructure grant. Faculty have access to a Biosafety Level 2+ facility and core facilities that support research involving flow cytometry; microscopy and image analysis; genomics, proteomics, and bioinformatics; NIH-supported HIV/AIDS and cancer tissue collections; and transgenic animals. MITM faculty collaborations are supported by a city-wide NIH-funded Center for AIDS Research and our partnership with a NIH-funded Clinical and Translational Science Institute at Children's National Health System.

Research-active faculty participate in the School's interdisciplinary Ph.D. program in the biomedical sciences. Department faculty also teach in the Physician Assistant and M.D. programs.

The Chair's responsibilities include:
maintaining an active, externally-funded research program; providing management and leadership to the department; overseeing the department's research centers and collaborative research partnerships; participating in and coordinating the educational programs of the department; and engaging in service to the School and University.

Minimum Qualifications: The successful candidate will be an internationally-recognized scientist with an exemplary record of extramurally funded research in a research area aligned with the focus of the department. A Ph.D. in an appropriate field and/or a M.D. degree is required. Candidates should also possess the background and experience that have prepared them to lead the department in fulfilling its research agenda and educational responsibilities.

To Apply: To be considered, please complete an online faculty application at https://www.gwu.jobs/postings/51785. There you will be asked to upload a curriculum vitae and a cover letter and to provide the names of professional references. Review of applications will begin on June 29, 2018 and will continue until the position is filled. Only complete applications will be considered. Successful completion of a background screening will be required as a condition of hire.

The School of Medicine & Health Sciences is committed to building a culturally diverse faculty and strongly encourages applications from candidates from backgrounds underrepresented in science.

The George Washington University is an Equal Employment Opportunity/Affirmative Action Employer that does not unlawfully discriminate in any of its programs or activities on the basis of race, color, religion, sex, national origin, age, disability, veteran status, sexual orientation, gender identity or expression, or on any other basis prohibited by applicable law.

By **Asia L. Dowtin** *and* **Delphis F. Levia**

The power of persistence

After months of intense planning, I was finally ready to set up my dissertation research equipment in a patch of urban forest. It was such a perfect summer day that it was hard to imagine anything going wrong. But while installing the first of my stemflow collars—odd-looking devices that I was mounting on tree trunks to investigate part of the water cycle—I was startled by a stern voice demanding my attention. It was a law enforcement officer. I was bewildered. I hadn't called the police, and I could not fathom who would have. Then it hit me: I was an African-American woman in one of the city's most affluent, predominantly white neighborhoods, installing what could be perceived as an unsightly contraption in their prized space.

An intense round of questioning followed, laced with not so subtle suggestions that I should relocate my research, even though I repeatedly assured the officer that I'd obtained permits to do my work there. All of this was during the nationwide coverage of Sandra Bland, the African-American woman who died in jail after being arrested during a traffic stop. So, in addition to feeling defensive, offended, and unwelcome, I was terrified. I needed an advocate.

"Advocacy must be flexible and dynamic to surmount obstacles."

DELPHIS: When Asia told me what had happened, I was shocked and dismayed. As her adviser, I was initially unsure how to handle the situation, which quickly escalated into threats from regional officials to revoke her research permit. I was at a crossroads: I could steer the research in a safer but less impactful direction, possibly by shifting the emphasis to other field sites, which would appeal to my risk-averse side. Or I could maintain my allegiance to our vision of Asia's project.

ASIA: When I started my Ph.D., I wanted to conduct fieldwork in the community in part to help address the feelings of isolation that came with being the only black person in the department. Experiencing hostile behavior from the community I was aiming to connect with was incredibly discouraging. I knew that the only way I'd be able to move forward would be if I had Del's full support.

DELPHIS: I had started working with Asia after she considered leaving our graduate program because I knew that she was capable of great research. She had struggled at times with feeling dismissed in academia, and I had promised her that I would do everything I could to support her. As a first-generation college student from a working-class community, I have seen people in my own life squander their potential after facing adversity, and I was committed to helping Asia—and all my students—avoid the same fate.

So the choice quickly became clear: We would push ahead with the planned research. After a tense exchange, the officials honored the permit. And I learned an invaluable lesson as an adviser: To ensure student success, advocacy must be flexible and dynamic to surmount obstacles, both foreseen and unforeseen.

ASIA: Nearly as soon as the permit situation was resolved, my research sites were vandalized. For more than a year, my equipment was tampered with and stolen, jeopardizing my data quality and potentially my ability to graduate. But each time something went wrong, Del was like a boxing coach, wiping me off and helping me get back into the ring. His unwavering high expectations for me kept me motivated and helped me continue to push against these and the myriad other obstacles that presented themselves along the way to my degree.

Rather than knock me out, each jab taught me a little more about how to prepare for and bounce back from the next. With the completion of my Ph.D. now in sight, I know this victory would not have been possible without the persistence that these challenges taught me and the help of a committed coach throughout the fight. ■

Asia L. Dowtin is a doctoral candidate and Delphis F. Levia is a professor at the University of Delaware in Newark. *They thank Estella Atekwana and Carol Henderson for their initial insights. Do you have an interesting career story? Send it to SciCareerEditor@aaas.org.*

ILLUSTRATION: ROBERT NEUBECKER

A Sponsored Supplement to *Science*

90 years of scientific advances at the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences

Sponsored by

中国科学院
CHINESE ACADEMY OF SCIENCES

Produced by the *Science*/AAAS Custom Publishing Office

Science
AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Career Development Center

Career Development Center

Meet career challenges head-on with online courses designed for scientists

- Public engagement
- R&D funding analysis
- Effective communication
- Proposal writing
- Career development
- Science policy and advocacy

REGISTER TODAY
CareerDevelopment.aaas.org

AAAS

90 years of scientific advances

at the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences

Cover: The teapot and cups—forming the "90" that highlights SIMIT's 90th anniversary—represent research and resources flowing into three units created by SIMIT: SIMIC Holdings for industrialization, the Shanghai Industrial μTechnology Research Institute (SITRI) for technology transfer, and the Center for Excellence in Superconducting Electronics (CENSE) for superconducting quantum devices and circuits research.

Photo credits: front and back outside covers, WANG Zhaoqing; inside back cover: LIN Luoluo.

This supplement was produced by the *Science*/AAAS Custom Publishing Office and sponsored by the Shanghai Institute of Microsystem and Information Technology, CAS.

Editor: Sean Sanders, Ph.D.
Proofreader/Copyeditor: Bob French
Designer: Amy Hardcastle

Materials that appear in this supplement have not been peer-reviewed nor have they been assessed by *Science*. Articles can be cited using the following format: [AUTHOR NAME(S)] [CHAPTER TITLE] in 90 years of scientific advances at the Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences. (*Science*/AAAS, Washington, DC, 2018), p. [xx-xx].

Xiaoying Chu
Director, Global Collaboration and Business Development, Asia
xchu@aaas.org
+86-131-6136-3212

Danny Zhao
Regional Sales Manager, Asia
dzhao@aaas.org
+86-131-4114-0012

8 June 2018

Introductions

2 **90 years young**
Sean Sanders, Ph.D.
Science/AAAS

3 **From the Director**
Professor WANG Xi, Ph.D.
Director of SIMIT, CAS

Chapters

4 **Chapter 1: Overview of SIMIT**

6 **Chapter 2: Scientific research at SIMIT**

23 **Chapter 3: Research platforms at SIMIT**

28 **Chapter 4: "Three-in-One" Collaborative Innovation System**

30 **Chapter 5: International cooperation**

32 **Chapter 6: Research talent at SIMIT**

36 **Chapter 7: SIMIT Talent Cultivation and Recruitment Program**

90 years young

2018 marks the 90th anniversary of SIMIT's founding, making it modern China's oldest national institute.

From its humble beginnings in 1928 as the Institute of Engineering, Academia Sinica, the Shanghai Institute of Microsystem and Information Technology (SIMIT)—part of the Chinese Academy of Sciences (CAS) since 1950—has grown to become one of China's premier information technology research institutes.

2018 marks the 90th anniversary of SIMIT's founding, making it modern China's oldest national institute. To help celebrate its achievements over these past nine decades, we present this supplement to *Science*, in which the reader will find a wealth of information about SIMIT's past and present. Chapter one provides a brief background, including SIMIT's history and scientific achievements. The second chapter gives a detailed look at current research at the institute, outlining its primary areas of endeavor, namely intelligent sensing microsystems, superconducting quantum devices, advanced silicon-based materials, broadband wireless communication technologies, micro- and nanotechnologies for sensors and transducers, phase-change random access memory and its applications, terahertz solid-state technologies, and brain-inspired chips and bionic vision. Together, these research areas advance the field of information technology in China while also satisfying its domestic economic and research priorities, and addressing national security priorities.

Chapter 3 walks the reader through the five major research platforms established by SIMIT. These range from a production line for microelectromechanical systems (MEMS) and other "More-than-Moore" (MtM) technologies, to a platform for the microstructural characterization of functional materials for informatics applications. Also described is the Superconducting ELectronics Facility (SELF), one of the world's most advanced R&D platforms for the fabrication and testing of superconducting devices and circuits. It is an open, shared platform, available to domestic and international researchers as a service, or through collaborative projects.

SIMIT's unique "Three-in-One" Collaborative Innovation System is described and explained in chapter 4. This system brings together research know-how from SIMIT, technology transfer expertise from the Shanghai Industrial μTechnology Research Institute (SITRI), and industry knowledge from SIMIC Holdings, SIMIT's capitalization and investment arm. This model was born out of CAS's guiding principle of driving innovation that supports national needs.

The final two chapters, 5 and 6, highlight SIMIT's international outreach and collaboration, and introduce some of its globally recognized staff. SIMIT is proud of its collaborative approach to basic research, having established cooperative agreements with research centers throughout the world, including Germany and Finland. Its staff numbers over 600, including 80 foreign researchers as well as many winners of prestigious domestic awards.

At 90, SIMIT is still a young and growing institute. As it approaches its 100th anniversary, it is not resting on its laurels. It continues to expand and attract exceptional talent from within China and around the world. It welcomes top applicants (see chapter 7) who wish to be involved in cutting-edge electronic science and technology research, and information and communication technology development.

We wish the institute and its staff congratulations on reaching this milestone and look forward to seeing it move from strength to strength over the next decade and beyond.

Sean Sanders, Ph.D.
Science/AAAS Custom Publishing Office

PHOTO: ALEX LEE

From the Director

Our primary focus at SIMIT is to better apply scientific and technological breakthroughs to the harmonious development of society and the sustainable growth of the Chinese economy.

On the 90th anniversary of the Shanghai Institute of Microsystem and Information Technology (SIMIT) at the Chinese Academy of Sciences (CAS), we are proud to sponsor this supplement produced by the *Science*/AAAS Custom Publishing Office and present it to international readers and researchers. This is a new and unique opportunity for SIMIT to provide a broad description of its research, scientific platforms, collaborative innovation, international cooperation, talent, achievements, ongoing projects, and future visions to the global research community.

SIMIT is a respected national institute with a long and rich scientific research history, being one of the first national research institutes in China. Since the establishment of the Institute of Engineering, Academia Sinica, in 1928, multiple generations of scientists have made significant advances in answering critical scientific questions. Their dedication, passion, and achievements laid a solid foundation upon which modern-day SIMIT has been built.

Over the past 90 years, despite numerous changes in research focus, SIMIT has consistently followed the unchanging principle that it must continually adjust and adapt to meet the demands of China's strategic and economic growth. In the last 20 years in particular, research and development at the institute has focused on basic science and technology, as well as strategic advanced technologies related to emerging industry development. By aligning the objectives and interests of the institute with the needs of society, SIMIT aims to catalyze innovation for the advancement of the country.

Today, science and technology are advancing at an unprecedented pace. Our primary focus at SIMIT is to better apply scientific and technological breakthroughs to the harmonious development of society and the sustainable growth of the Chinese economy. As we enter a new era of Chinese-styled socialism, the nation is moving toward a historic moment of great rejuvenation. The national goal of strengthening the country's system of innovation and enhancing the power of China's scientific research is motivating SIMIT to take advantage of this unique opportunity to play a leading role in the nation's technological development.

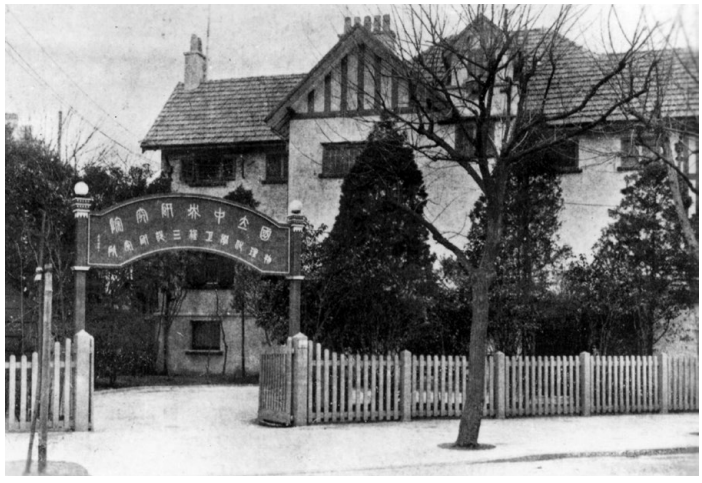
Looking to the future, SIMIT will retain its focus on remaining at the forefront of science, national priorities, and economic progress, while improving its capacity for innovation and enhancing its proficiency in information and communication technology. At the same time, we will also attach importance to solving key scientific and technical problems in the areas of intelligent sensing microsystems, superconducting quantum devices and circuits, and advanced silicon-based materials. SIMIT will continue to innovate, striving to maintain its position as an irreplaceable national scientific research institution.

I wish to express my appreciation to my colleagues, whose efforts and dedication made this supplement possible. I sincerely hope it will encourage future collaborations between SIMIT and scientists in many diverse fields from around the world.

Professor WANG Xi, Ph.D.
Director of SIMIT, CAS

Overview of SIMIT

The Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences (SIMIT, CAS)—formerly known as the Institute of Engineering, Academia Sinica—was founded in 1928 and is China's oldest engineering research institution. In 1950, it became part of the newly established Chinese Academy of Sciences, and was renamed the Engineering Laboratory, CAS. Its name changed multiple times over the next five decades, finally becoming SIMIT in May 2001. The institute is proud of its long legacy as the oldest national institute in modern China.



SIMIT's first location in 1928: 899 Xiafei Road (now 1337 Middle Huaihai Road)

Historical achievements

In its 90 years of development, SIMIT has been awarded 48 national science and technology awards, with some projects winning prestigious honors, including a Special National Science and Technology Progress Award for "Fabrication Technology for A-Separation Membrane," and two First Class National Science and Technology Progress Awards for "High-Speed and Ultrahigh-Speed Bipolar Digital Integrated Circuits" and "Research and Industrialization of Silicon-On-Insulator (SOI) Materials."

In the 1950s, SIMIT was the first institute to successfully develop spheroidal graphite cast iron domestically and the first to establish an ultrapure metal and III–V compound semiconductor materials research base. In the 1970s, SIMIT developed microelectronics and integrated circuit (IC) manufacturing technology for China. In collaboration with industry partners, SIMIT developed industrial PN junction–based isolation ICs, emitter-coupled logic (ECL) high-speed circuits, and 8-bit and 16-bit microprocessors, all firsts for the country. In the 1980s, SIMIT led the research on microelectromechanical systems (MEMS) and silicon-on-insulator (SOI) development in China.

From 1998, SIMIT broadened its research scope from materials science to information and communication. The institute created a strategic plan summarized as "systems drive devices, devices drive materials," emphasizing electronics science and technology; information and communication engineering including wireless information systems and networks, microsystem technologies, and functional materials and devices for information technology; and micro-type and new energy sources. SIMIT successfully launched the Chuangxin No. 01 and 02 satellites and the Shenzhou VII companion satellite. It quickly became an important base for R&D of microsatellites in China. SIMIT implemented the Shanghai World Expo Intrusion Prevention Sensor Network, the Pudong Airport Intrusion Prevention System, the Taihu Lake Water Quality Monitoring Sensor Network, the Security System for the central line South-to-North Water Diversion Project, and other Internet of Things (IoT) application pilot projects. The institute led the development of IoT core technology and promoted it as a national strategic emerging industry. Additionally, SIMIT developed broadband wireless communication systems with independent intellectual property rights. These systems were deployed for the Wenchuan and Yushu earthquake relief efforts, the Shanghai World Expo, and other major events in China. The broadband wireless communication system played an invaluable role in the earthquake relief work, providing social stability and enhancing public safety and security. SIMIT established Shanghai Simgui Technology Co., Ltd., the first and only world-class advanced SOI materials manufacturer in China. Simgui's products addressed the gaps in domestic SOI material needs and have been used in key state projects.

During the 12th Five-Year Plan period (2011–2015), as a leading technical institute within CAS, SIMIT made breakthroughs in multifunctional compound sensors and related key technologies. The institute worked on providing special broadband wireless sensor network system solutions and related large-scale devices, and on transitioning the country's wireless communication system from narrowband to broadband. SIMIT successfully developed the first domestic large-scale, high-reliability, dedicated application-specific integrated circuit (ASIC) chip based on the 0.13-μm SOI process. The chip was used in the Beidou navigational satellite and other key national projects. This helped to lay a solid foundation for the self-sustained design and manufacture of core spacecraft components in China. With the support of the superconducting nanowire single-photon detector (SNSPD) from SIMIT, Academician Pan Jianwei of the University of Science and Technology of China has globally pioneered the implementation of a hacker-proof, 200-km measurement-device-independent quantum encryption key distribution system. This project was selected as one of the top 10 scientific and technological achievements of 2014 by academicians from CAS and the Chinese Academy of Engineering. Another development, an ultra-high-speed gravity acceleration sensor, was a critical component of certain pieces of strategic equipment in China. The ability to develop such a sensor lessens the effect of potential international embargoes. SIMIT also launched a MEMS technology industrialization platform known as the Shanghai Industrial μTechnology Research Institute (SITRI), now regarded by the Shanghai government as one of the "beams and pillars" of the globally influential Shanghai Technology Innovation Center.

SIMIT's Xingfo Building

Jiading Campus of SIMIT

Research strength

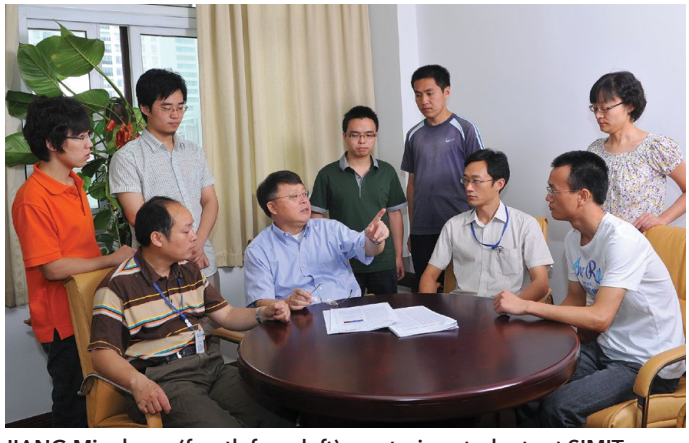
SIMIT includes three national-level key laboratories: the State Key Laboratory of Transducer Technology, the State Key Laboratory of Functional Materials for Informatics, and the National Key Laboratory of Microsystem Technology. Two CAS key laboratories are also part of the institute: the CAS Key Laboratory of Wireless Sensor Networks and Communications and the CAS Key Laboratory of Terahertz Solid-State Technology, as well as the Center for Excellence in Superconducting Electronics (CENSE). Currently, SIMIT has seven research departments: the Department of Micro/Nano Technologies for Sensors and Transducers, the Department of Functional Materials for Informatics, the Department of Terahertz Solid-State Technology, the Department of Wireless Sensor Networks, the Department of Broadband Wireless Technology, the Department of Superconductors, and the Department of Bionic Vision Systems. Six affiliated groups scattered around China, in Nanjing, Hangzhou, Jiaxing, Fuzhou, Chengdu, and Nantong, are also part of the institute. SIMIT also has several joint partnerships, including laboratories with the Juelich Research Center (Forschungszentrum Jülich GmbH) of the Helmholtz Association, the Collaborative Innovation Center for High-Reliability Components with the CAS Shanghai Engineering Center for Microsatellites, and the Joint Laboratory on Superconducting Quantum Devices and Quantum Information with the University of Science and Technology of China.

Talent

SIMIT plays a key role in cultivating talent and supporting nationwide science and technology development by attracting world-class faculty members. Currently, the institute employs 610 full-time staff members, of whom 87 are full professors. Nearly 700 students also work at SIMIT. Over the years, SIMIT has built a highly educated and ambitious team that has won several CAS funding awards, including four National Outstanding Young Scientist Fund awards, seven National Thousand Talents Plan awards, two National Thousand Young Talents Plan awards, and 22 Hundred Talents Plan awards. CAS academicians as well as international associates of the U.S. National Academy of Sciences, together with the rest of the SIMIT staff, are dedicated to nurturing exceptional talent in addition to producing world-class scientific research.

Innovation

Since SIMIT was established, it has displayed a pioneering spirit of innovation and entrepreneurship. It has developed state-of-the-art technologies across a broad range of fields, and has instituted the "Three-in-One" Collaborative Innovation System to promote technology transfer and industrialization of scientific and technological achievements. This system consists of SIMIT leading the principal technology R&D, SITRI as the technology transfer platform, and SIMIC Holdings as the industrialization partner, combining technology and finance.



JIANG Mianheng (fourth from left) mentoring students at SIMIT

Future vision

According to the new CAS operating principle of "an orientation toward the forefront of international science and technology, major national demands, and primary challenges of the national economy," SIMIT will fully utilize its advantages in electronic science, technology, information, and communication engineering to solve key scientific and technical problems through the design and application of intelligent sensing microsystems, superconducting quantum devices and circuits, and advanced silicon-based materials. SIMIT also aims to promote the application of major research achievements, in addition to becoming a "Four First-Class" institution and an irreplaceable national research treasure in the information and communication technology field.

Scientific research at SIMIT

To fulfill its mission of meeting national needs in China, SIMIT has been conducting strategic, innovative, and forward-looking research with the aim of promoting breakthroughs in key technologies and multidisciplinary innovation, providing systematic solutions, and contributing to the development of information technology, advanced materials, and micro-/nanotechnologies. Under the Chinese Academy of Sciences (CAS) innovation plan of "focusing on international cutting-edge science and technology, major national demands, and the main challenges of the national economy," SIMIT has established a research strategy consisting of three major breakthroughs and five top priorities that mirror its strengths in electronic science and technology (S&T), information systems, and communication engineering.

The three major breakthroughs are as follows:

Intelligent sensing microsystems (ISMs). SIMIT aims to fulfill national needs by enhancing wireless sensor networks, expanding the industrial Internet of Things (IoT), and creating an information database for national security and industry applications.

Superconducting quantum devices and circuits. SIMIT focuses on basic scientific research on quantum materials and devices, with the broad goal of developing proprietary advanced superconducting core electronic devices and integrated systems.

Advanced silicon-based materials and applications. SIMIT is devoted to the research and industrialization of advanced silicon-on-insulator (SOI) materials and 12-in. silicon wafers for China's semiconductor and microelectronics industries.

The five top priorities are as follows:

1. *Special broadband wireless communication technologies and equipment*. Research on key wireless communication technology that enables concurrent transmission using orbital angular momentum and signal energy is being prioritized, together with the development of broadband wireless communication technology dedicated to high-speed rail and smart power grids.

2. *Micro- and nanotechnologies for sensors and transducers*. Research on state-of-the-art microelectromechanical or nanoelectromechanical systems (MEMS/NEMS) technologies for physical, chemical, biochemical, implantable, and compound combination sensors, as well as hyperspectrum detection microsystems.

3. *Phase-change random-access memory (PCRAM) and applications*. Research on the electrical, thermal, and crystallographic principles of low-power, high-speed, reliable phase-change memory units to drive the engineering of PCRAM chips.

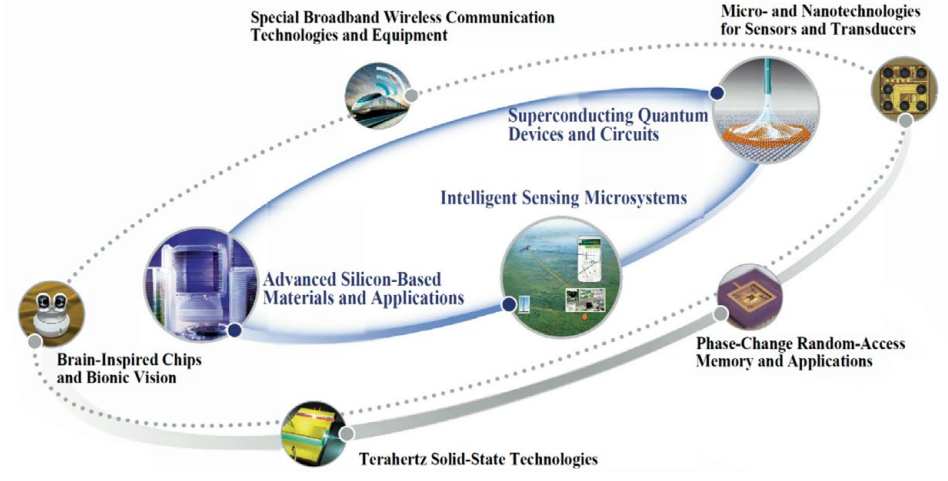
4. *Terahertz (THz) solid-state technologies*. Research on THz radiation sources, the detection principles of THz quantum cascade lasers, the interaction between THz waves and various materials, millimeter-wave monolithic integrated circuits (ICs), and millimeter-wavelength inspection and imaging equipment.

5. *Brain-inspired chips and bionic vision*. R&D of brain-inspired chips and bionic vision systems, including binocular image signal processing chips, sensors that replicate the five human senses, and robotic limbs.

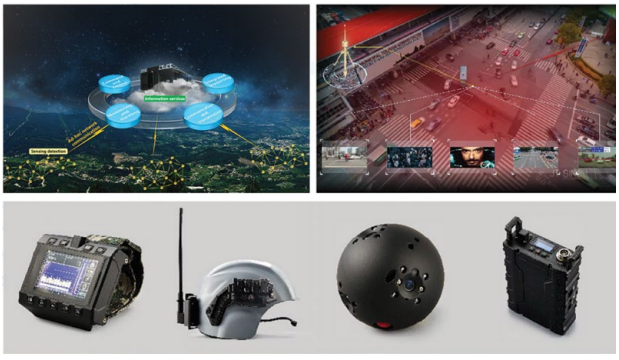
Intelligent sensing microsystems

The development of ISMs involves cross-disciplinary research encompassing traditional sensors as well as emerging IoT technology. Domestically, both wireless microsensor networks and intelligent terminal equipment have been developed in China for national security and industrial applications. More importantly, SIMIT holds core intellectual property assets related to hybrid-target cooperative detection, recognition algorithm architectures, equipment miniaturization, low power consumption, large-scale random networking, and heterogeneous system fusion. Dozens of ISM standards proposed by SIMIT have been accepted by standardization organizations in China and Europe.

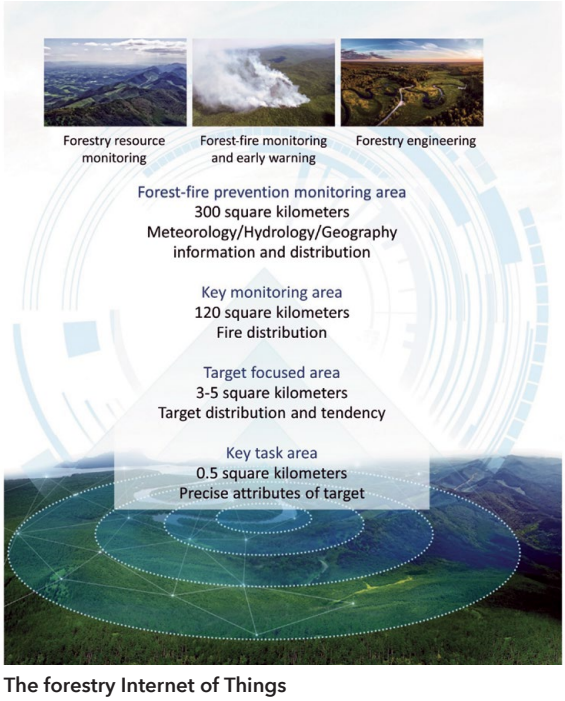
Based on a proposal by CAS in 1999, SIMIT established a research team for wireless sensor networks that became the first group to study both sensors and IoT in China. SIMIT has become the central department of the Shanghai IoT League, the core and leading institute for sensors and IoT R&D at CAS, and the central and technical lead unit in the field of sensing and reconnaissance equipment in China. Subsequently, SIMIT has also become the secretariat of the standardization working group for the Chinese sensor network and made numerous proposals for setting international standards. To further consolidate R&D teams and promote the innovation and development of sensors and IoT technology, SIMIT set up the Wireless Sensor Network Department in May 2015. The ISM team, working within this department, received a Shanghai S&T Progress Award (first prize) in 2008 as well as the S&T Progress Award for the Chinese People's Liberation Army in 2010. Yuan Xiaobing, the



Representation of research directions at SIMIT under the 13th Five-Year Plan (2016–2020). The three breakthroughs are shown in the center ring, surrounded by the five priorities.



Applications and examples of intelligent sensing microsystems (ISMs). Top row: Applications of ISMs including emergency communication systems. **Bottom row:** Devices containing ISMs, including a wrist-mounted terminal, a communications helmet, a reconnaissance ball, and a communication replay device.



The forestry Internet of Things

Forestry fireman using a personal intelligent sensing microsystem

team leader and deputy director of SIMIT, won the Advanced Individual Award from the General Equipment Department and the Ministry of Industry and Information Technology in 2012 and was honored by the State Council of the People's Republic of China.

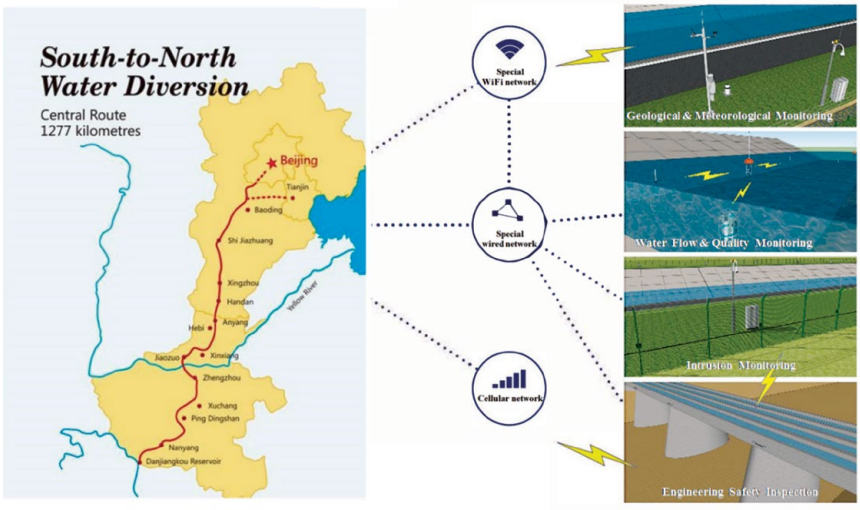
ISMs developed at SIMIT have attained worldwide prominence because of their technical advantages in two major areas. First, SIMIT constructed an innovative sensing network architecture, including sensing equipment, a sensing index system, and an application model. An ISM consists of integrated sensors and communication devices. Through systematic integration, functional integration, and complementary morphology, the success of various sensing applications has been demonstrated. The latest, third-generation sensing equipment architecture incorporates upgrades that provide significant improvements in the scale, functionality, and networking capabilities of sensing equipment.

Second, SIMIT used key advantages of ISMs, such as multimodal heterogeneity, multitasking ability, and stochastic dynamics, to create a special collaborative sensing network that addressed a number of technical challenges including task-oriented target recognition in complex environments, antienvironmental interference, network adaptation and self-reconfiguration, and multisource information fusion.

SIMIT found that accuracy in the detection of target distance using the latest ISMs could be increased by over 30%. Moreover, the accuracy of single-sensor target recognition could be increased to more than 95%, and the accuracy of multisensory fusion target recognition could be improved to more than 98%. Certain performance indexes exceeded those of similar equipment produced internationally. When SIMIT first researched randomly distributed networking and covert communication in special sensor networks, it made advances in dynamic cross-layer protocol optimization, double cluster topology control, and dynamic multiple access control (MAC), among other areas. A sensing network platform with a multidimensional reconnaissance system, multisource acquisition, and sensor information fusion has been constructed. It significantly improves ISM network self-expansion, self-reconfiguration, anti-jamming ability, and mobile survivability, and helps to simultaneously expand the capabilities of various reconnaissance sensors in the ISM from single sensors to network groups. Furthermore, SIMIT's research demonstrates that the use of broadband spread-spectrum communication technology, multipath rake receiver technology, fast frequency synchronization technology, and fast code chip synchronization technology can effectively improve the performance of anti-jamming, anti-capture, and information shielding systems in ISMs.

Recently, certain sensors employed in MEMS technology have been co-opted for ISMs. These sensors now feature integrated circuits, high-speed-algorithm hardware, complex-logic software, the MAC protocol, and an energy-efficient

mixed-measurement routing algorithm. This enables their further miniaturization and concomitant reduced power consumption. Extensive experiments have shown that these new sensors can greatly improve reconnaissance equipment performance. Compared with the previous generation of equipment, both the weight and power consumption of the new MEMS-enhanced equipment can be reduced by 70%. Instruments such as the ISM tactical terminal have enabled the creation of an integrated information network that allows the seamless linkage of equipment, personnel, and information systems as well as interconnections between all communication terminals and from terminals to the central communications center.



Large-scale sensor network development for the security system constructed for the central line South-to-North Water Diversion Project

A personal communication system for individual forestry firemen as well as a forestry IoT were developed at SIMIT utilizing ISMs. The ISM links personnel and transportation and integrates "fire situation, self-situation, and environment" in an innovative forest-fire prevention command system. In a Chinese national fire drill, it was highly praised by the director of the State Forestry Administration. They agreed that a system with civil–military integration, advanced technology, and strong practicability was essential for forest-fire prevention. Additionally, a public security system based on ISM has been popularized and applied in Shanghai Pudong Airport, Hongqiao Airport, Shanghai Expo Park, the central line South-to-North Water Diversion Project, the Shanghai Public Security Bureau, and the Zhejiang Maritime Safety Administration, among others. These applications have resulted in significant economic and social benefits.

Since wireless communication plays an important role in ISMs, SIMIT set up the Key Laboratory of Wireless Sensor Networks and Communications, CAS, which is dedicated to major national strategic and application requirements. The laboratory is committed to fundamental research, application development, systems integration, and the development of testing and evaluation technologies for wireless sensor networks and broadband wireless mobile communication. It aims to establish industry-leading wireless sensing networks and mobile communication systems.

The main research directions of the Key Laboratory of Wireless Sensor Networks and Communications include: (1) intelligent networking technology using wireless sensor networks, (2) efficient wireless transmission technology, (3) integration solutions and information processing technology related to wireless sensor networks, and (4) technology for testing and evaluating wireless systems. In addition, a large-scale application demonstration system will be set up to promote the standardization of wireless sensor networks and next-generation broadband mobile communication.

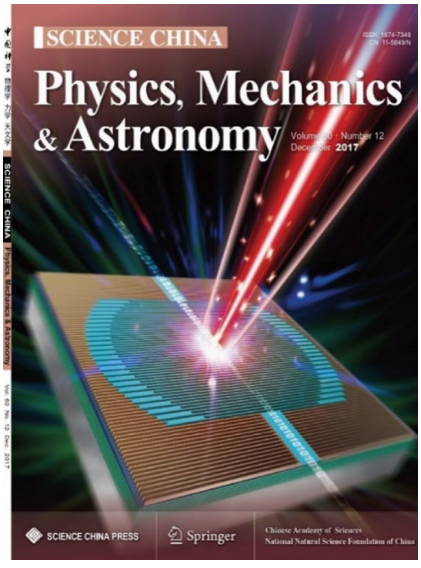
In the future, ISMs will focus on target detection technology with microaperture arrays, adaptive technology for sensing under complex conditions, deep intelligent front-end sensing technology, and ultra-large-scale dynamic intelligent networking technology. Furthermore, new wearable information terminals and next-generation wireless sensing systems will be developed in the future. The team will strive to meet the demands for national defense, national security, and public safety.

Superconducting quantum devices and circuits

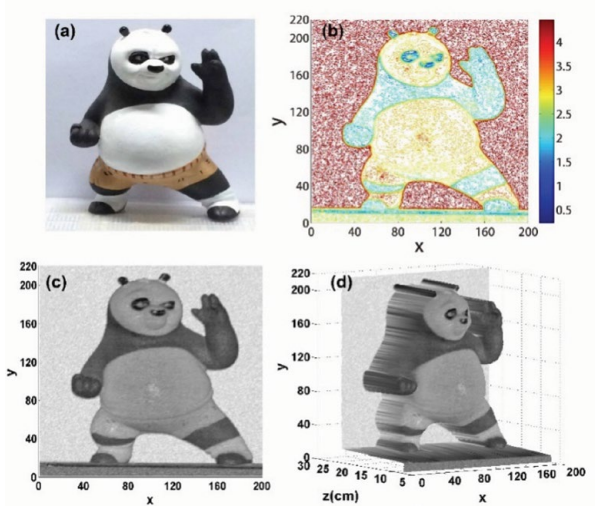
Superconducting electronics is an interdisciplinary research area involving superconducting physics, materials science, and electronic technologies. The scope of superconducting electronics includes superconducting passive devices, sensors, detectors, digital electronics, and quantum bit circuits. Superconducting devices and circuits have high performance and hence play indispensable roles in ultrasensitive detection, high-precision measurements, quantum information processing, quantum metrology, and many other important research areas.

SIMIT has historically formed the backbone of research on superconductivity in China. It has achieved many successes, including the development of China's first low-temperature superconducting wire and the creation of a high-temperature superconductor with the world's highest critical current density for melt texturing. In 2005, the major focus of superconductivity research at SIMIT changed from superconducting materials to superconducting electronics. After more than 10 years of effort, SIMIT deployed the development strategy for a superconducting electronics research facility and a world-class fabrication platform for superconducting quantum devices and circuits. Many breakthroughs, such as a superconducting nanowire single-photon detector (SNSPD) and superconducting quantum interference device (SQUID), have been realized and validated in a variety of applications.

In 2015, the CAS Center for Excellence in Superconducting Electronics (CENSE) was formally founded. CENSE focuses on applied fundamental research in superconducting electronics. Committed to meeting national strategic requirements and remaining at the forefront of the superconducting electronics field, it is dedicated to top-level research on quantum materials science, superconducting sensors/detectors and their applica-



Cover Image for *Science China: Physics, Mechanics & Astronomy* issue (December 2017), showing a schematic of a superconducting nanowire single-photon detector (SNSPD) (*1*).



Few-photon imaging based on time-of-flight laser ranging using an SNSPD (*2*).

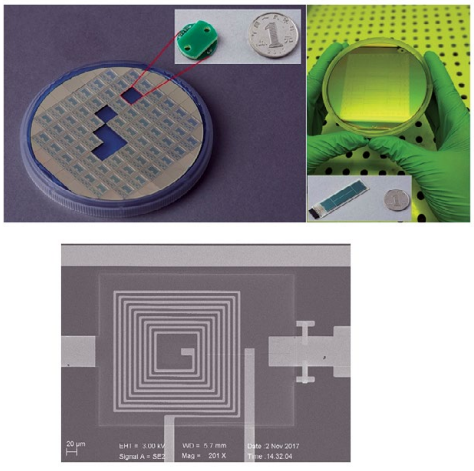
tions, superconducting ICs, and superconducting computers. By undertaking several major research programs, CENSE aims to build an interactive research environment for materials, devices, and applications, and to construct a multidisciplinary platform for both scientific achievement and talent development. Its goal is to build a world-class research center in superconducting electronics and to develop independent intellectual property related to superconducting chips, devices, and systems.

CENSE is working on some key detectors, among which are SNSPDs. CENSE focuses on the fabrication and function of SNSPDs based on ultrathin niobium nitride (NbN) films. Related studies include the growth of ultrathin superconducting films, detection mechanisms, design and fabrication of SNSPDs, and key parameters of SNSPDs. A correlation was found between the factors that determine both the absorption and intrinsic detection efficiency of SNSPDs. By optimizing the fabrication parameters, an NbN SNSPD was demonstrated using a Gifford–McMahon (GM) cryocooler, having a system detection efficiency (SDE) of over 90% at 1550 nm. By understanding the origin of dark counts, SIMIT developed on-chip and fiber-endface band-pass filter technologies to suppress the dark count rate (DCR) to less than 1 Hz while maintaining an SDE as high as 80%. The contribution of each part to the timing jitter (TJ) of the system was quantitatively analyzed, and a system with a TJ of 14 pico-seconds (ps) was obtained. The wavelength range of the SNSPD was extended from the near-infrared to visible wavelengths. The typical SDE is approximately 80% for all wavelengths. Furthermore, CENSE proposed and fabricated novel SNSPDs such as microfiber-coupled SNSPDs and SNSPDs with a high polarization sensitivity.

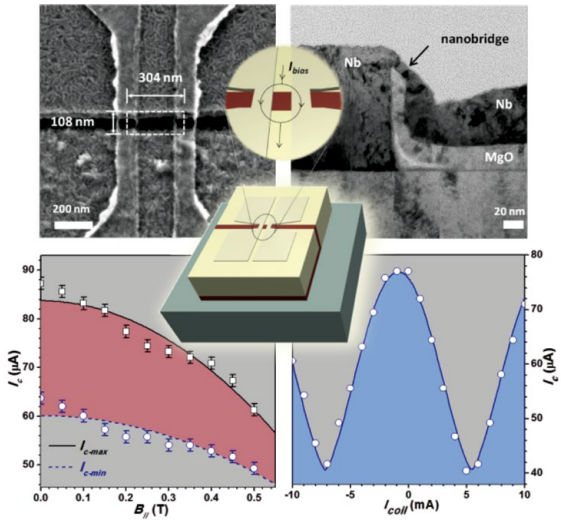
CENSE collaborates with many Chinese institutions in various research fields, such as quantum information processing and satellite laser ranging, to explore advanced applications for SNSPDs. More than 20 SNSPD-based systems were installed for various applications at 10 universities and institutes. These collaborations have produced a series of important results that have enabled significant progress in the field of superconducting electronics.

Homemade SNSPD systems have been applied to quantum key distribution (QKD) experiments and have achieved several world records, including a record distance of 404 km for fiber-based QKD. An SNSPD at a wavelength of 532 nm was applied to a ground station for satellite laser ranging at a distance of approximately 3,000 km, and a new method for time-of-flight few-photon imaging was developed. Among other achievements, CENSE observed quantum fingerprinting, exceeded the classical information transmission limit by using a low-DCR SNSPD, demonstrated time-bin-encoded boson sampling with a fast SNSPD, and realized quantum teleportation over a network by using a high-SDE SNSPD.

In the field of SQUIDs, CENSE proposed and verified novel SQUID chips with electronics, focusing primarily on simplicity, robustness, and targeting real-world applications while maintaining an acceptable noise level. All-NbN SQUIDs were designed and developed with high-quality epitaxial NbN/aluminum nitride (AlN)/NbN Josephson junctions on single-crystal magnesium oxide substrates. Their performance is comparable to that of Nb-based SQUIDs at liquid-helium temperatures; temperature fluctuations do not lead to an appreciable degradation in their performance. Verification of these results has demonstrated the high performance and good temperature stability of the all-NbN SQUIDs. Moreover, CENSE has realized the fabrication of a SQUID magnetometer and gradiometer on a 4-in. silicon wafer based on Nb/Al technology. These technologies show outstanding noise performance—3.5 femtotesla (fT)/$Hz^{1/2}$ for the magnetometer and 0.65 fT/(cm·$Hz^{1/2}$) for the gradiometer—and are beneficial for multichannel application systems. These SQUIDs have been successfully employed in biomagnetism and geophysical applications. Moreover, a new type of Nb nanoSQUID composed of three-dimensional (3D) nanobridge junctions has been developed. Recently, research on nanoSQUIDs has received increased attention because of their applications in the readout of solid-state spin-based qubits and high-resolution scanning SQUID microscopy. However, existing Nb nanoSQUIDs with nanobridge junctions usually exhibit a low magnetic-flux modulation depth (approximately 15%). The improvement of the



Superconducting quantum interference devices (SQUIDs) (*3, 4*). A SQUID magnetometer wafer and packing module (**top left**), and gradiometer wafer and packing module (**top right**). A scanning electron microscope photograph of a SQUID (**bottom**).



Structural imaging and characterization of a 3D niobium (Nb) nanoSQUID (*5*).

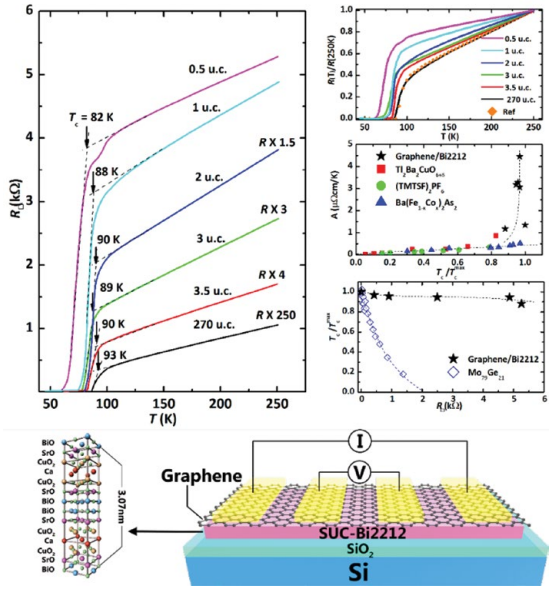
junction-current–phase relation by CENSE led to a significant increase in the magnetic-flux modulation depth (up to 73%). The magnetic flux noise of the device was measured to be as low as 230 $n\Phi_0/Hz^{1/2}$. Furthermore, it can withstand a parallel magnetic field as high as 0.5 tesla, which corresponds to a field in the X band (10 GHz) for electron spin resonance requirements. The 3D Nb nanoSQUID constitutes an important step toward the inductive detection of a single electron spin.

An airborne full-tensor magnetic gradient system was developed by CENSE, and several flight tests were carried out. The system's performance approaches that of the world's best. A superconducting system based on the transient electromagnetic method was also developed, and several profile surveys were undertaken. The results of the profile surveys were supported by the drilling data. The inversion depth exceeded 2,000 m. Several types of gradiometer modules with different baselines and gradient orders were fabricated in an unshielded environment and in a magnetically shielded room.

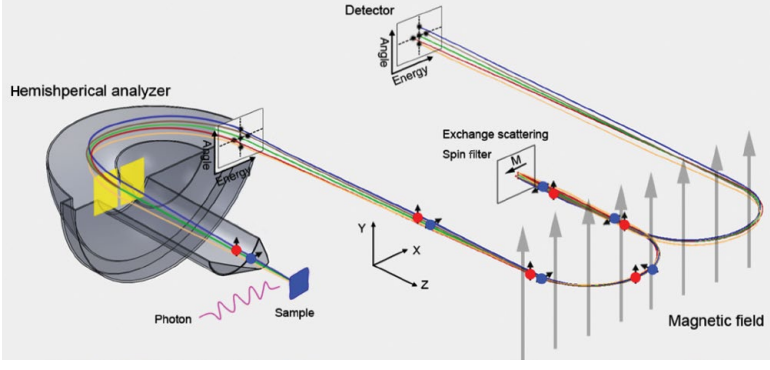
Additionally, CENSE piloted adult magnetocardiography (MCG), fetal MCG (fMCG), and magnetoencephalography signal detection, and set up the first 36-channel MCG system and 4-channel fMCG system in a magnetically shielded room in China. CENSE endeavors to improve system performance and introduce MCG systems for clinical applications. CENSE built the first unshielded ultralow field (ULF) magnetic resonance imaging (MRI) system working in a harsh urban radiation environment. The ULF-MRI group is now developing a low-cost MRI system for biological tissue imaging. They are also working on the design of novel imaging contrast agents, optimization of the prepolarization technique, and exploration of medical applications for the ULF-MRI system.

The mission of quantum materials research at CENSE is to carry out advanced research on the synthesis/fabrication, structure, and physical properties of materials as well as their complex interrelationships, using materials science and condensed-matter physics. The research covers novel functional materials, such as conventional and unconventional superconductors, two-dimensional (2D) atomic crystals, topological quantum materials, and related heterostructures. By studying their atomic/electronic structures and transport and regulation properties, CENSE aims to clarify the underlying mechanisms and correlations between material parameters and device performance, in order to discover new materials with superior properties for device applications. One of CENSE's goals is to provide scientific grounds for the development of new materials, new devices, and new techniques with the aim of generating collaborative innovation in materials science, physics, device engineering, and applications research. Recent breakthroughs include high-temperature superconductivity in ultrathin (half-unit cell) bismuth (Bi)-cuprate superconductors, the interfacial superconductivity of oxide heterostructures, wafer-size single-crystal graphene growth, and the edge chirality control of graphene nanoribbons. CENSE's advanced spectroscopy methods and equipment as well as its micro-/nanoprocessing and characterization facilities greatly extend the scope and depth of its materials research. In combination with theoretical simulations, quantum materials research at CENSE forms a complete innovation chain along with the R&D and application of superconducting devices and circuits.

The novel properties of materials originate from their electronic states and the interactions between electrons. Research on the physical properties of materials at the microscale is based on the characterization of the multiple degrees of freedom of electrons, including their energy, momentum, orbit, and spin. The research area of in situ characterization of electronic structure is based on the Shanghai Integrated Platform for Materials of Energy and Environment (SiP·ME^2) founded by the National Natural Science Foundation of China. The platform aims to develop advanced synchrotron-based characterization methodologies and apply them to a broad range of fundamental and application fields such as superconductivity, magnetism, energy storage, catalysis, and the electronic structure of new functional materials. Currently, CENSE's research area covers in situ X-ray spectroscopy technology, spintronics, novel quantum materials for informatics, strongly correlated electronic systems, surface catalytic reactions, and energy storage materials.



High-temperature superconductivity in graphene/ultrathin $Bi_2Sr_2CaCu_2O_{8+x}$ heterostructure down to a half-unit cell (*6*).



High-efficiency multichannel spin- and angle-resolved photoelectron spectroscopy apparatus measuring simultaneous parallel detection of 6,786 channels (78 x 87) of electron energies, angular momenta, and spins (*7*).

In the future, CENSE, in association with SIMIT, will strengthen R&D into superconducting materials, sensors, and detectors, and promote research on the application and industrialization of existing superconducting quantum device technologies such as SNSPDs and SQUIDs. At the same time, the design and fabrication of superconducting quantum devices and circuits will be further improved. In addition, the development of superconducting ICs and a superconducting computer will soon begin, which will raise the level of R&D of superconducting electronics.

Advanced silicon-based materials and applications

Both SOI wafers and large-scale silicon wafers form the foundation of IC manufacturing. SOI technology in particular is renowned as the "silicon IC technology of the 21st century," owing to its applications in high-speed, low-power-consumption circuits, high-temperature-resistant circuits, micromechanical sensors, and photoelectric integration processes. Moreover, SOI technology is important because of its radiation hardness and its strategic significance in national security and aerospace R&D.

In the 1980s, SIMIT was the first Chinese institute to conduct research on SOI materials and devices, and it continues to take a leading role in this field. SIMIT has gradually built a foundation for SOI technology through technological innovation. In 2001, it incubated the sole SOI manufacturing base in China, Shanghai Simgui Technology Co., Ltd., enabling the industrialization of SOI technology. In 2005, SIMIT realized the key technological breakthrough of bond and etch-back SOI (BESOI) fabrication, and ramped up production at Simgui. Meanwhile, SIMIT also merged BESOI technology with separation by ion implantation of oxygen (SIMOX) technology, and developed a world-class independent intellectual property, Simbond. Applying the Simbond technology, SIMIT produced an SOI material that includes a top silicon layer with a continuously tunable thickness and a buried insulator layer. The mass production of Simbond technology has now been achieved. In 2006, after years of R&D and industrialization, SIMIT was awarded a First Class National Science and Technology Progress Award for "Research and Industrialization of Silicon-On-Insulator (SOI) Materials." In 2007, the SOI research group at SIMIT received the Distinguished Scientific Achievement Award from CAS. Through the development of SOI technology, Simgui has manufactured products that cover most SOI application areas, and their quality and technical capabilities are recognized by prominent international companies. Currently, more than 90% of Simgui products are sold to North America, Japan, Europe, Russia, South Korea, Taiwan, and Singapore, and Simgui has become a supplier for NXP Semiconductors N.V., Toshiba, GlobalFoundries, Vanguard, Hitachi, CSMC Technologies Corporation, and Mellanox Technologies. In 2014, Simgui signed a strategic cooperation agreement with Soitec and initiated international industrial cooperation. In 2015, in collaboration with Soitec, Simgui successfully prepared the first 8-in. SOI wafers using hydrogen-implanted layer transfer technology. Currently, 8-in. SOI wafers are being produced at high volume for radio frequency (RF), power, and silicon photonics applications.

In view of the rapid development of ICs, it was apparent that a smaller feature size would stimulate a greater demand for SOI wafers. Aiming at meeting the demand for fully depleted SOI (FDSOI) materials, SIMIT developed Sim-split technology as an independent intellectual property and successfully produced FDSOI wafers with a controllable thickness. In order to address the issue of reduced carrier mobility in the device caused by shrinking channel length, SIMIT explored a solution to improve carrier mobility in three ways (strain engineering, new channel materials, and hybrid crystal orientation) and successfully developed three high-mobility SOI materials (strained SOI,

Examples of silicon-on-insulator (SOI) wafers fabricated by SIMIT

germanium-on-insulator, and hybrid-orientation SOI). Moreover, aiming at developing SOI materials from 3D to 2D, SIMIT successfully synthesized single-crystal graphene wafers on germanium, building the foundation for developing 2D SOI wafers.

In the future, microelectronic and optoelectronic devices, as well as intelligent microsystems assimilating technology trends such as miniaturization and integration, will continue to be developed. The functions of microsystem chips will become more complex, more diversified, and more easily integrated into other systems. These development trends have spurred a huge demand for heterogeneous integration technology. Heterogeneous integration paves the way for the development of microelectronics in the post-Moore era. SIMIT has conducted R&D of XOI technology (X-on-insulator, where X = a compound or functional thin film) for fabricating silicon-based heterogeneous integrated substrates, and has developed the "lattice cutting" and "heterogeneous interface fusion" processes for integrating dissimilar materials with a silicon (Si) substrate. SIMIT has achieved fabrication of different types of single-crystal thin films with heterogeneous integration, including silicon-based compound semiconductors (III–V, wide-bandgap semiconductors) and silicon-based functional films (piezoelectric and ferroelectric film).

SOI devices have an enhanced immunity to single-event effects and transient radiation effects compared to their bulk Si counterparts; however, the introduction of buried oxide leads to severe total ionizing dose (TID) effects. To increase radiation tolerance to TID effects, an excellent method for the fabrication of highly crystalline, high-quality radiation-hardened SOI wafers has been proposed, with independent intellectual property rights. These modified SOI materials show superior performance, and their technical specifications are comparable with those of commercial 8-in. SOI wafers. The total dose radiation tolerance of hardened SOI is 500 krad(Si), which can be substantially improved to 1 million rad (Mrad)(Si) with further modification. This technique brings the advantages of a strong process and product expansibility by adjusting implantation parameters. The successful mass production of radiation-hardened SOI wafers has provided a solid foundation for developing radiation-hardened SOI ICs for space applications.

Faced with the urgent demand for a new generation of

Signing of a strategic cooperation agreement between Shanghai Simgui Technology Co., Ltd., and Soitec

The first independently developed 8-in. radiation-hardened silicon-on-insulator (SOI) wafer, manufactured by SIMIT

Successful launch of an SOI-based "SIMIC" security chip (inset) onboard the Beidou Navigation Satellite

information technology for high-speed optical communication and optical switching in silicon ICs, silicon-based photonics—which could substitute optical interconnects for metal interconnects and use photonics as a medium for transmitting data between chips and devices—is inevitable in the development of computer systems and communication equipment characterized by broadband, high-speed, large-capacity, and large-scale parallel processing. SOI materials are an ideal platform for manufacturing complementary metal-oxide semiconductor (CMOS)-compatible silicon-based photonics devices, and they are a solid foundation for the rapid development and industrialization of silicon-based photonics. Together with Hua Hong Semiconductor Limited and Shanghai Industrial μTechnology Research Institute (SITRI),

Shanghai Zing Semiconductor Technology Co., Ltd., successfully produced the first 300-mm single-crystal ingot in China.

SIMIT established 0.13-μm silicon photonic devices and an optoelectronic integrated technology platform compatible with CMOS processes. The research group developed silicon nanowire optical waveguides, the transmission losses of which to the transverse electric and transverse magnetic modes are 2.4 + 0.2 dB/cm and 0.59 + 0.32 dB/cm, respectively. Furthermore, SIMIT developed a variety of high-performance silicon-based optoelectronic devices such as grating couplers, low-loss filters, multimode interference splitters/combiners, and other passive devices as well as adjustable optical attenuators, high-speed silicon-based photoelectric modulators and detectors, and other active devices.

In terms of photonics integration, the research team realized a monolithic, integrated, arbitrary microwave waveform-generation function underpinned by silicon-based integrated photonics technology, in order to overcome the shortcomings of the conventional microwave photonics system built with discrete devices—such as its large volume, high power consumption, and impact on the external environment. Signal modulation, an adjustable delay, a signal channel, and other functions are integrated on the same chip to create a lightweight, low-power-consumption, multichannel microwave photonic chip. To overcome the diffraction limitation of current photonics integration and propose new principles and schemes for realizing high-density, ultra-large-scale photonics integration, the research team developed a new optical manipulation system based on a silicon nanopillar structure, which broke the optical diffraction limit. In addition, an all-silicon, single-row cylindrical array with a high depth-to-width ratio and high surface quality has been developed.

Silicon wafers are the cornerstone of IC manufacturing and an important raw material that dictates the performance of ICs. In 2014, SIMIT participated in the establishment of Shanghai Zing Semiconductor Technology Co., Ltd. (ZingSemi), which was committed to providing a complete set of 12-in. silicon-wafer mass production technologies for the 40 nm–28 nm process, including single-crystal growth, wafer processing, epitaxial wafer growth, wafer analysis, and wafer characterization. The company was committed to setting up a 12-in. semiconductor silicon-wafer production base in China to fully meet the urgent need for 12-in. silicon wafers in China's IC industry. Thus far, ZingSemi has applied for 480 patents, of which 60 have been issued. In 2016, it successfully produced the first 300-mm single-crystal ingot in China. ZingSemi is committed to manufacturing and supplying key materials for the IC industry in China, thus securing the supply chain and bringing its technology up to global standards.

Going forward, SIMIT plans to develop 12-in. silicon wafers and 12-in. FDSOI wafers, and also to explore wafer-scale high-mobility SOI materials, all with the aim of meeting future demand for microelectronics technology. It will lead the development of advanced microelectronics processing in China and undertake the application of FDSOI technology to IoT and vehicle electronics. Moreover, SIMIT will develop 8-in. RF SOI wafers and 8-in. power SOI wafers, and explore technologies for integrating nonsilicon and wide-bandgap semiconductors such as SiC and gallium nitride (GaN) with silicon and SOI architectures. Simultaneously, SIMIT will develop special techniques to create radiation-hardened SOI wafers and manufacture highly reliable IC chips to meet the demands of IC chip applications in special, high-radiation environments. Additionally, SIMIT will build a first-class, radiation-hardened R&D center in China; realize the coordinated development of radiation-hardened SOI materials, production processes, devices, chips, and system applications; and develop radiation-hardened chips for aerospace applications. Furthermore, SIMIT will integrate silicon-based photonic devices with circuits in a single chip, using mature CMOS manufacturing techniques. With silicon photonic chips, digital information can be transmitted using optical signals; consequently, the bottlenecks related to the interconnections in high-performance computers and high-end computer processing units can be overcome. Through collaborative innovation, SIMIT is striving to be the world's leading large-silicon wafer producer.

Special broadband wireless communication technologies and equipment

Wireless communication technology is one of the most important of China's national core research areas. It can enhance the competitiveness of the country's S&T research and development, and greatly promote sustainable economic growth. Along with the rapid spread of broadband wireless communication technology, SIMIT is also focused on building highly specialized broadband wireless communication networks and related equipment that meets special demands, such as those of national security and the railway, electricity, and transportation industries.

Special broadband wireless communication equipment

In 2008, SIMIT set up the Broadband Wireless Technologies Laboratory (BWTL). As a leading special communication network solutions provider, BWTL focuses on the challenges of mission-critical communication scenarios, such as the absence of infrastructure, reliable high-speed transmission, massive node access, ultra-high-speed mobility, self-managing and self-healing, flexibility and scalability, and high security. Traditional

Testing a special broadband communication system during a disaster relief effort in Wenchuan City, Sichuan Province

Retrieving data from the Shenzhou XI spacecraft recently returned to Earth

Celebrating the return of the Shenzhou XI spacecraft

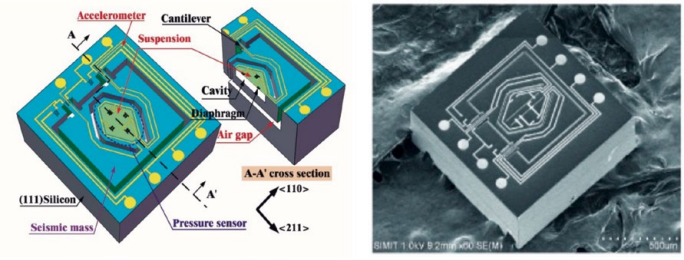
civil or commercial broadband communication technology focuses on optimization for low mobility and short-range transmission performance, and cannot meet the deployment needs of special industries that require complex, multipath high-speed/high-mobility capabilities and can only operate on authorized frequencies. Special broadband wireless communication systems aim to provide reliable wireless communication in complicated geospatial environments as well as electromagnetically complex environments.

BWTL has taken the lead in research on a new generation of mobile communication in China, proposing discrete Fourier transform-spread generalized multi-carrier (DFT-S-GMC) transmission solutions for a new generation of broadband wireless mobile communication using patented technology. It has developed two broadband wireless access systems based on millimeter-wave and time-division long-term evolution (TD-LTE), and made breakthroughs in some key technologies for a new generation of broadband wireless emergency communication. BWTL has also drafted national broadband wireless multimedia standards, participating in the international 4G standardization process and completing the integration and testing of the first Chinese 4G system. In addition, BWTL has focused on special broadband wireless communication for the future, and continues R&D on smart grid–specific wireless communication, railway location information systems, and wireless communication utilizing orbital angular momentum. BWTL will actively participate in and cooperate with standardization efforts worldwide, further develop future broadband wireless mobile communication systems, and strive to become an important base for R&D, technology innovation, and the industrialization of special broadband wireless communication technology and equipment.

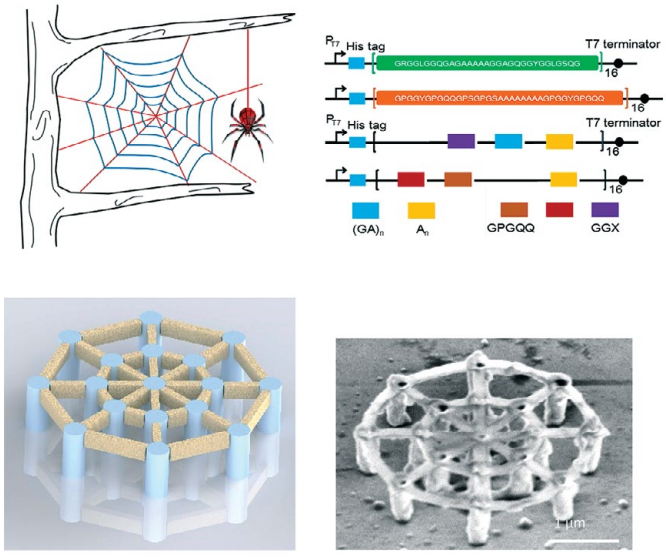
By leveraging advanced 4G/5G wireless transmission technologies combined with in-house intellectual property covering DFT-S-GMC transmission, software-defined ad hoc network architectures, and artificial intelligence spectrum management methods, BWTL can assure optimal, secure wireless communication. It has designed and developed every component of the technology while also adjusting to the ever-changing priorities and challenges of mission-critical communication. Every aspect and component of BWTL's products are designed to allow maximum flexibility and scalability. Versatile network topologies and product portfolios cater to every deployment and project scope, from small to large scale, with the greatest flexibility and reliability. BWTL develops and delivers special network solutions for many application areas, including military communication, railway transportation, disaster relief and public safety, and security and surveillance.

In the area of disaster relief and public safety, BWTL/SIMIT has been involved in many activities since 2008, and has established several emergency broadband wireless communication networks and 24-h long-distance wireless monitoring systems. For example, BWTL/SIMIT played a key role in providing technology for the 2008 Wenchuan earthquake, the 2010 Yushu earthquake, and the 2010 Shanghai World Expo. On November 18, 2016, wireless broadband communication technology was successfully used in the rescue command communication support system for the return of the Shenzhou XI spacecraft, the vehicle used for China's longest manned space mission. With eight helicopters and multiple ground systems, SIMIT built a search communication security system for the main landing area, which captured the first onsite picture for the Beijing Aerospace Control Center.

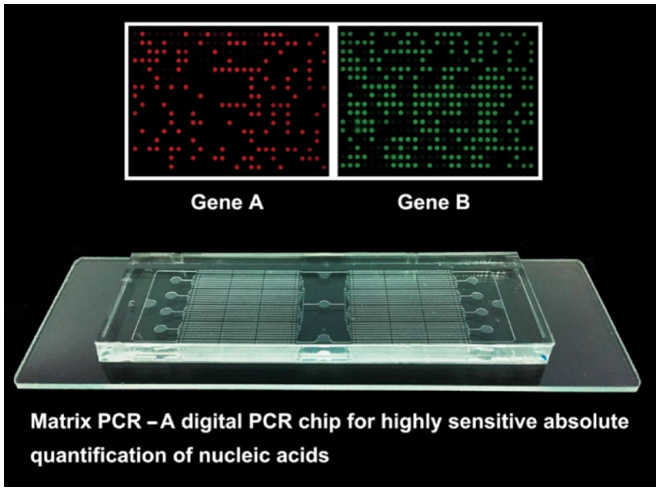
For China's railway transportation industry, SIMIT has formed a railway location information system that covers the full range of dynamic and static location information on targets such as railroad tracks, trains/vehicles, and people and facilities; developed train/cargo/human location services (positioning accuracy down to the centimeter level) and antitheft functionality for facilities; improved transport production efficiency; and ensured transport and personnel safety. The system is currently being verified and applied at the Guangzhou Railway Bureau. SIMIT has also completed several electricity projects with the State Grid Corporation and China Southern Power Grid Co., Ltd., providing solutions for automatic



Single-wafer, single-side fabrication of a multifunctional pressure sensor into an accelerometer hybrid sensor using a micro-openings inter-etch and sealing technique (*8*).



Examples of 3D biomimetic bionanostructuring on genetically engineered spider silk to create specific shapes or functions on demand (*9*).



Micro-/nanobiosensors using matrix polymerase chain reaction (PCR) technology to monitor cancer biomarkers

power distribution and video monitoring of power transmission lines over thousands of kilometers.

In the near future, SIMIT will further focus on special broadband wireless communication and continue their R&D of smart grid–specific wireless communication, railway location information systems, and wireless communication utilizing orbital angular momentum.

Micro- and nanotechnologies for sensors and transducers

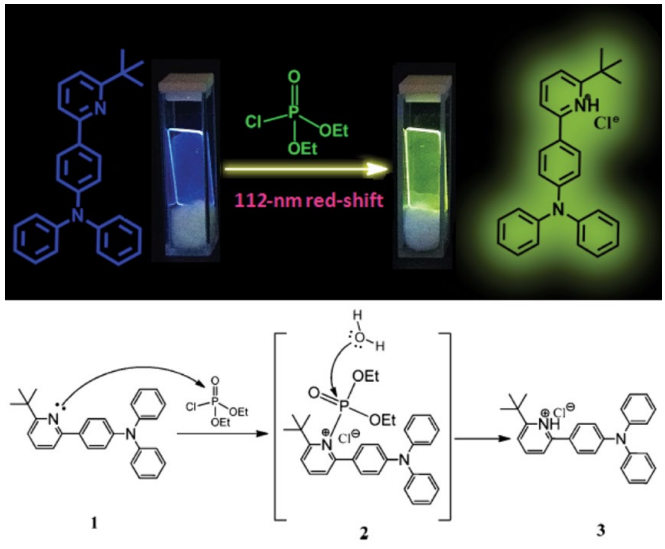
The State Key Laboratory of Transducer Technology (SKLT^2), with its south branch located at SIMIT, was approved for construction by the National Planning Commission in 1987 and officially opened in 1989. It was the first State Key Laboratory founded in China that focuses on transducer technologies based on microelectronics and micro-/nanofabrication. The major research theme of SKLT^2 is emerging micro- and nanotechnologies for sensors and transducers (MNTST), which covers major topics within this multidisciplinary field, and impacts many real-world applications ranging from consumer electronics and medical instrumentation to security systems and environmental monitoring. The main research activities at SKLT^2 include the development of micro-/nanomanufacturing techniques, new sensing materials and mechanisms, advanced sensing devices, and integrated microsystems. The goal of SKLT^2 is to meet emerging demands for transducer technologies for national security, aerospace, energy, the automotive sector, industrial applications, health care, and consumer electronics, among other areas. In 2011, SKLT^2 received the second prize of the Shanghai Science and Technology Progress Award for their work entitled "Trace Detection of Explosives Using Fluorescence Chemical Sensors." In 2012, it received the second prize of the National Science and Technology Progress Awards for "Three-Dimensional High-Density Packaging Technology of RF Electronic Systems and Its Applications," and the second prize of the State Technological Invention Award for "Key Technologies and Design Methods of MEMS 3D Manufacturing Based on Process Selectivity." Furthermore, SKLT^2 received the 2015 second prize of the Shanghai Technological Invention Award for "Key Technologies of High Sensitivity Photoelectric Sensor Chips and Portable Embedded Spectrometers."

SKLT^2 was involved in early efforts to develop key technologies in the field of MNTST, and has achieved breakthroughs in important scientific and technical areas within the following two major research directions:

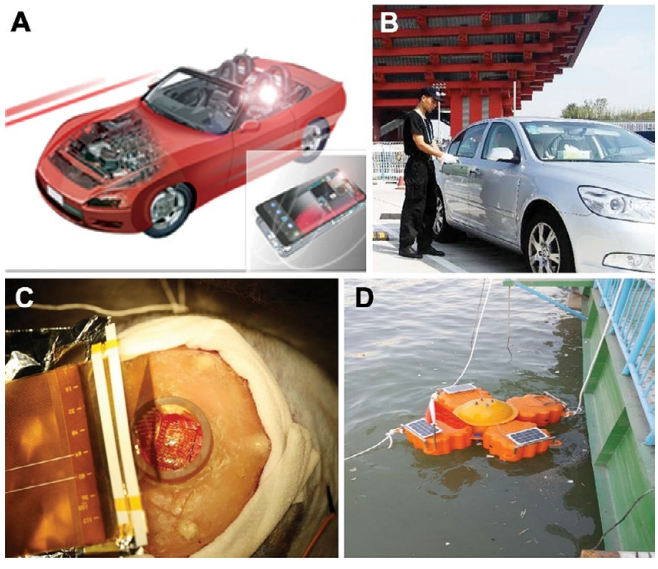
1. The development of new techniques and strategies for micro-/nanofabrication. SKLT^2 has developed a silicon-based, single-sided fabrication technique, called "micro-openings inter-etch and sealing" (MIS), which is compatible with current IC foundry protocols. In addition, three innovative microfabrication processes have been developed based on MIS, including packaging-stress-suppressed suspension (PS^3), thin film under bulk (TUB), and pressure sensor into accelerometer (PinG). They can be used individually or in combination to create 3D, multilayered, and functional MEMS structures. These processes are designed for high-yield, low-cost volume

production of MEMS devices and are compatible with standard fabrication processes in original equipment manufacturer IC foundries. A facile through-silicon-via (TSV) process has also been developed using wafer-level liquid-metal injection via a filling method. The combined effects of capillary action and liquid bridge rupture allow rapid MEMS packaging, which effectively shortens the TSV filling time from several hours to several minutes. Additionally, a novel method for precise 3D bionanostructuring—termed "protein bricks"—has been developed to create biologically functional, hierarchical, and heterogeneous micro-/nanostructures with specific shapes and functions on genetically engineered proteins. The unique combination of protein-based biomaterials and ecofriendly micro-/nanofabrication techniques offers extensive versatility for a variety of applications that require devices to be produced in a "green" way with customizable biofunctions and accurate nanoscopic geometries. This new biomanufacturing paradigm serves as a promising route to biomimetic 3D micro-/nanomanufacturing, to complement and potentially disrupt current bionanofabrication techniques including both top-down (based on lithographic methods) and bottom-up (via self-assembly, such as DNA origami) approaches, owing to its capacity to fabricate multiscale structures with facile functionalization.

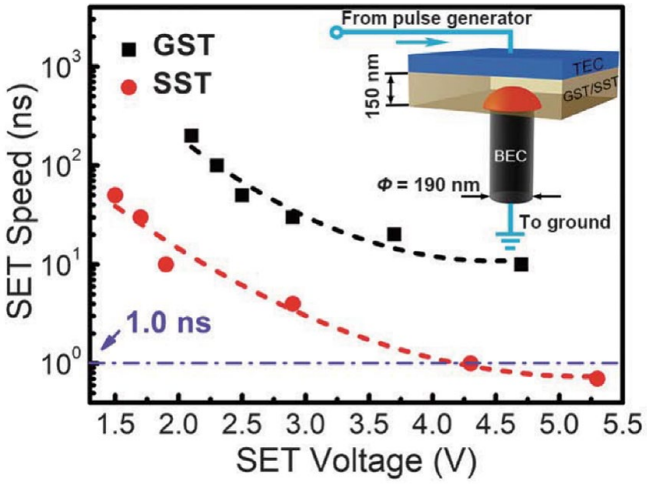
2. The discovery of new materials and mechanisms for sensing applications. SKLT[2] has been working on ecofriendly micro-/nanofabrication techniques using biocompatible and biodegradable biopolymers (e.g., naturally extracted and/or genetically engineered proteins) as the resist materials for the precise patterning of functional sensing materials and structures using standard lithography techniques such as electron-beam and ion-beam lithography. The electron- or ion-regulated nanoscale polymorphic transitions in protein-based bioresists are revealed by near-field infrared imaging and nanospectroscopy at resolutions approaching the molecular level. The ability to locally probe nanoscale protein-structure transitions combined with nanometer-precision lithography offers fine control of the structure of protein resists in both positive and negative tones as well as in two and three dimensions. SKLT[2] has further developed a method for creating a new class of optical devices that can dissolve/degrade into the surrounding environment or the human body at controlled rates, referred to as "physically transient micro-/nano-optics," for information concealment and medical implantation. Moreover, SKLT[2] has been actively engaged in the early detection of cancer and has developed a series of micro-/nanobiosensors that are capable of multiparameter monitoring of cancer biomarkers (such as nucleic acids, proteins, cells, and exosomes) at single-cell and single-molecule levels using a matrix polymerase chain reaction. These microfluidic-based sensors provide a solution for diagnosing cancer patients at a low cost, providing rapid analysis of fluids and nonturbulent flows in a device with a small footprint. SKLT[2] has also established a comprehensive set of methods and strategies for developing fluorescence-based optical chemical sensors using organic semiconductors as the key sensing element for the portable, multichannel, and ultrasensitive trace detection of explosives and nerve agents. These sensors were successfully employed at several important public events, including the Shanghai World Expo 2010, the 2016 G20 Hangzhou Summit, and the 2017 BRICS Summit. Furthermore, a "cantilever-based lab-on-a-chip" strategy has been developed to quantitatively investigate the molecular interactions between a functional material and biochemical molecules (from the perspective of the "material genome") using a comprehensive set of thermodynamic/kinetic parameters such as enthalpy, Gibbs free energy, and activation energy, which are extracted from the sensing data measured by resonant microcantilevers. A series of functional materials has also been identified and optimized for toxic gas capture and/or detoxification applications using cantilever-based microgravimetric analyzers.



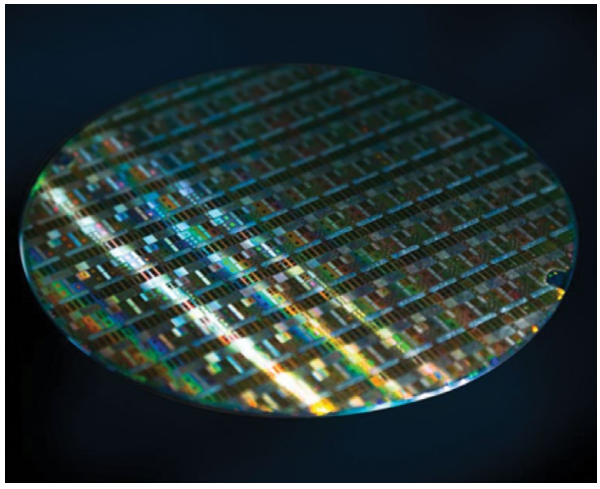
Efficient fluorescent sensor for detection of nerve-agent mimic diethyl chlorophosphate using intramolecular charge transfer (*10*).



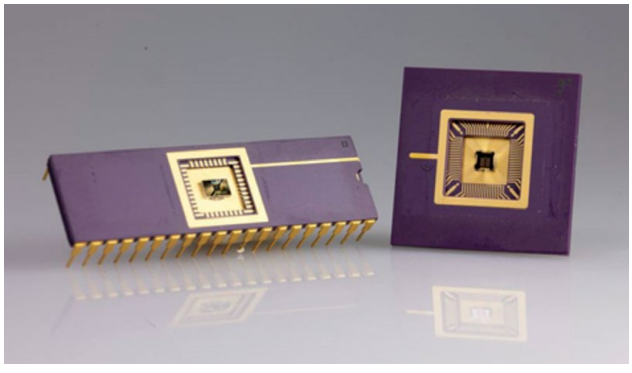
Examples of typical applications based on the four major research areas within SKLT[2]: (A) Consumer electronics, (B) security, (C) medical instrumentation, and (D) environmental monitoring.



Data from a research paper published in *Science* showing 700-ps SET speed in Sc-Sb-Te phase-change materials (*16*). GST, GeSbTe (germanium-antimony-tellurium); SST, Sc–Sb–Te (scandium-antimony-tellurium); TEC, top electrode contact; BEC, bottom electrode contact.



Phase-change random-access memory (PCRAM) units on a wafer



Packaged PCRAM chips

Mass-produced PCRAM chips used in printer ink cartridges

SKLT[2] has been a pioneering institution in research on MNTST and has contributed significantly to the advancement of transducer technologies. Currently, SKLT[2] is working on important scientific research related to emerging micro-/nanotransducer technologies, including the design and manufacture of micro-/nanosensors and transducers, MEMS and NEMS, the sensing mechanisms of nanomaterials, and integrated transducing microsystems and chips. In the future, SKLT[2] will continue its research with an expanding interdisciplinary team that embraces challenges with passion and creativity, to pursue fundamental science and meet the practical demands in the growing MNTST field.

Phase-change random-access memory (PCRAM) and applications

IC chips form the foundation of all information technology systems and therefore play a key role in the domestic economy and national safety. One third of all IC chips are used to store information, either temporarily or permanently. At present, dynamic RAM (DRAM) and Flash dominate the global market, accounting for 95% of memory storage chips sold. However, neither of these technologies are perfect: DRAM memory can be affected by external charge because of the smaller number of electrons in its capacitor relative to other forms of storage, while Flash faces serious crosstalk problems during operation, shortening its lifespan. These problems become more serious as the feature sizes on the chips shrink and approach a critical limit below 28 nm. Furthermore, storage technologies such as DRAM and Flash are incompatible with advanced CMOS technologies. Therefore, efforts are underway worldwide to develop new, non-volatile storage technologies that are compatible with advanced CMOS technology and have good scalability, 3D integration ability, fast operation speed, low power consumption, and long life. Phase-change RAM (PCRAM) shows promise, because it offers good scaling performance, fast write and erase speeds, 3D integration, and good compatibility with advanced CMOS technology. Thus, it is expected to become the most promising next-generation storage technology for mass production.

SIMIT began conducting PCRAM research in 2003 for three main reasons: First, PCRAM has the best overall properties for

a storage technology. Second, it is expected to be the most promising storage technology for mass production in international semiconductor development. And third, PCRAM is a nanoelectronics device, which aligned well with China's promotion of nanotechnology as a national development strategy at that time. SIMIT, together with Semiconductor Manufacturing International Corporation (SMIC), established a technology platform for 8- and 12-in. industrial-level PCRAM chip R&D, and achieved the integration of standard 0.18-μm/0.13-μm/40-nm PCRAM chips with standard CMOS technology. In 2008, SIMIT created China's first fully integrated, functional 8-Mbit PCRAM (the world's first embedded PCRAM), and by 2016, more than 16 million chips for printer consumables had been shipped. By 2017, the bit yield of 40-nm PCRAM chips was improved to over 99.99%, and raw 64-Mb chips were being tested in advanced information systems.

SIMIT has a research group of over 100 staff, conducting studies from basic research to engineering verification. The group has over 15 years of experience in PCRAM materials, structure, technology, testing, and circuit research. Furthermore, they have published two books in the PCRAM field and 410 papers in Science Citation Index journals, including *Science* and *Nature Communications*, and have applied for 344 patents, including 13 in the United States. Of these patents, 242 have been granted, including eight in the United States.

For a storage technology to be competitive in the market, it must offer high-density storage with high-speed read and write materials and circuits. By designing metal-centered octahedra, SIMIT developed scandium-antimony-tellurium (Sc–Sb–Te) and titanium-antimony-tellurium (Ti–Sb–Te) materials. The discovery of Sc–Sb–Te enabled reversible write and erase speeds of less than 1 ns (700 ps), and the power dissipation was 90% less than that of a mass-produced germanium-antimony-tellurium (Ge-Sb-Te) alloy, reported recently in *Science* (*11*). In the Ti–Sb–Te system, the Ti atoms partially substitute for Sb atoms in the hexagonal Sb_2Te_3 lattice and form $TiTe_2$ nanolamellae at grain boundaries, acting as nucleation centers and nanothermal insulators allowing for high speed and low power consumption properties. The results were published in *Nature Communications* (*12*, *13*). These breakthrough discoveries, accompanying the successful design of the high-speed readout circuit (US8947924), have greatly helped in the development of advanced storage technology in China.

For memory cells organized in an array, a method to select individual memory cells is necessary for reading and writing. In PCRAM, the density of memory cells is primarily determined by the size of the memory-cell selector. An Ovonic threshold switching (OTS) selector is one promising candidate for PCRAM owing to its large drive current. Researchers at SIMIT have successfully prepared an OTS device (*14*) and designed a 3D PCRAM circuit (*15*). In the future, SIMIT plans to develop one selector–one resistor (1S1R) single and integrated nanotechnology compatible with CMOS technology using a high-density crossbar structure, and further integrate them with advanced logic processes to achieve high-density storage chips.

Further PCRAM research at SIMIT will be extended to include the following aspects:

1. Basic research on PCRAM: first-principle calculations, the development and engineering of new phase-change materials, high-density 3D switching and memory devices, and brain-inspired computing;
2. PCRAM chip design, testing, and application: 130-/110-/40-/28-nm circuit design, chip technology design, an independent test platform and 8-/12-in. test systems, and chip package testing and application;
3. 8-in. and 12-in. PCRAM process development: development of nanofilling, polishing, and etching; development of single processes such as 1S (one selector), 1R (one resistor), nanoelectrodes, and dielectric coatings; and integration and optimization of 1S1R processes.

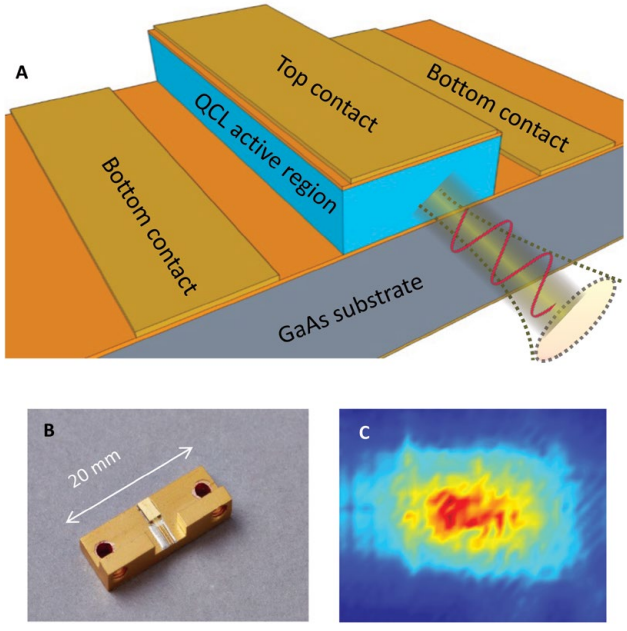
Going forward, SIMIT will focus on IoT, consumer electronics products, and national information security. These fields require a new generation of high-density memory chips that will include new materials, new structures, chip integration, and novel applications. Innovative research will comprise the screening of new low-power, high-speed phase-change memory materials; investigation of the physical nature and intrinsic properties of these materials; development of confined structure technology; development of OTS materials and devices; and clarification of the physical nature of nonlinear carrier transport. Finally, PCRAM chip products based on 130-/110-/40-/28-nm process technologies will be developed to be competitive in the international market, particularly for applications in IoT or smartphones.

Terahertz (THz) solid-state technologies

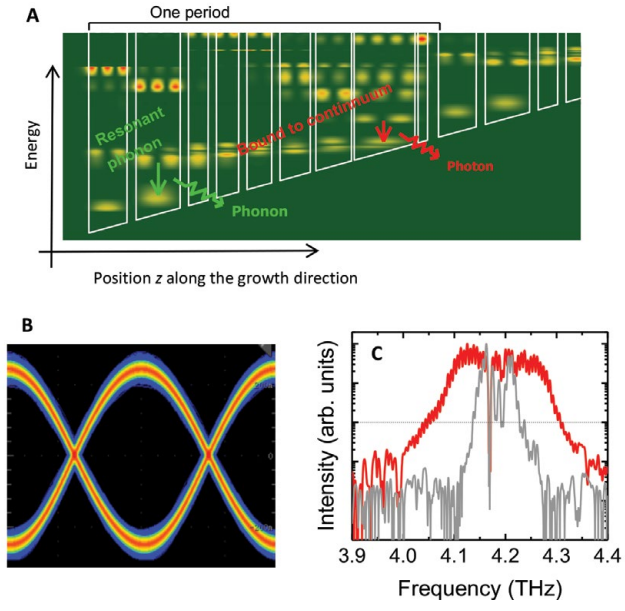
Terahertz S&T has attracted much attention in the past two decades. The THz region is defined as the frequency range from 0.1 THz to 10 THz (corresponding to wavelengths ranging from 30 μm to 3000 μm or photon energies ranging from 0.41 meV to 41 meV), located between the mid-infrared and submillimeter-wave bands. Because of the unique properties of THz waves, they have been applied in important research areas such as the investigation of the properties of semiconductor and high-temperature superconducting materials, tomography, gene discovery, cellular imaging, chemical and biological analysis, broadband communication, security imaging, and nondestructive detection of biological tissues. The study of radiation sources, detectors, and their applications in the THz band will promote the development of THz research and potentially solve major challenges in THz solid-state electronics and circuit technology.

SIMIT is one of the earliest institutions to conduct THz research in China. Although many proof-of-concept experiments have been carried out demonstrating potential applications of THz technology, the lack of compact and convenient devices for the generation and detection of THz waves needs to be addressed before widespread applications become a reality. SIMIT carries out research on THz radiation sources, detectors, and communication as well as imaging applications based on THz photonic and electronic devices. In 2010, with the approval of CAS, the CAS Key Laboratory of Terahertz Solid-State Technology was established with the support of SIMIT. This laboratory has become a leader in the THz research field in China, establishing a complete process and characterization platform for THz solid-state devices and making significant progress in building THz radiation sources and detectors. SIMIT's achievements in the area of THz quantum devices and semiconductor physics won the second prize of the Shanghai Natural Science Award in 2015.

SIMIT carries out both theoretical and experimental research in the field of THz photonics. An impact ionization model of THz radiation has been developed, and a hypothesis



Terahertz (Thz)-based technologies developed at SIMIT. (**A**) Schematic of a single-plasmon waveguide THz quantum cascade laser (QCL). (**B**) Photograph of THz QCL mounted on a copper base. (**C**) Typical far-field pattern of THz QCL.



A THz quantum cascade laser (QCL) frequency comb (*17*). (**A**) Calculated band structure of laser active region. (**B**) Eye diagram used to measure coherence of the modes. (**C**) Homogeneous THz spectral spanning of the laser comb under radio-frequency modulation.

has been proposed to explain the observation that the THz absorption coefficient varies with radiation intensity. The first domestic high-performance THz quantum cascade lasers (QCLs) and THz quantum-well detectors were built and applied in the fields of THz communication and imaging. The successful development of THz quantum-well-detector light-emitting-diode (LED) devices has overcome the challenges of THz detector arrays, and enables THz focal-plane pixelless imaging technology. Utilizing THz QCLs, SIMIT developed an engineering prototype linkage with real-time imaging and spectrum analysis, which can acquire THz images and spectra of dangerous goods. SIMIT also developed a THz modulation circuit and signal processing circuit to allow for real-time audio- and video-signal transmission.

In the fields of THz frequency combs and high-speed THz mixers, SIMIT has recently undertaken innovative research and obtained important results. Using high-performance THz QCLs with a hybrid active region design, SIMIT successfully demonstrated the application of THz frequency combs in QCLs. Furthermore, by employing RF modulation, the THz QCL frequency comb could continuously span over 330 GHz (8% of the central frequency), a record for QCLs with a bound-to-continuum active region design. SIMIT researchers, using the broadband frequency comb, also demonstrated THz spectroscopic applications such as the transmission measurement of a gallium-arsenic (GaAs) etalon as well as spectral identification of ammonia gas. In addition, by exploiting a microwave transmission line, we showed that a THz quantum-well photodetector (QWP) can work as a high-speed frequency mixer. The THz QWP mixer shows an intermediate frequency bandwidth of 6.2 GHz. We successfully measured the optical intermode beat note of a long-cavity THz QCL by employing this mixer. The high-speed QWP mixer was also used for imaging applications. The main advantage of this technique is that the frequency can be downconverted from the THz to the GHz range; therefore, the signal can be easily amplified, filtered, and read using mature microwave technology.

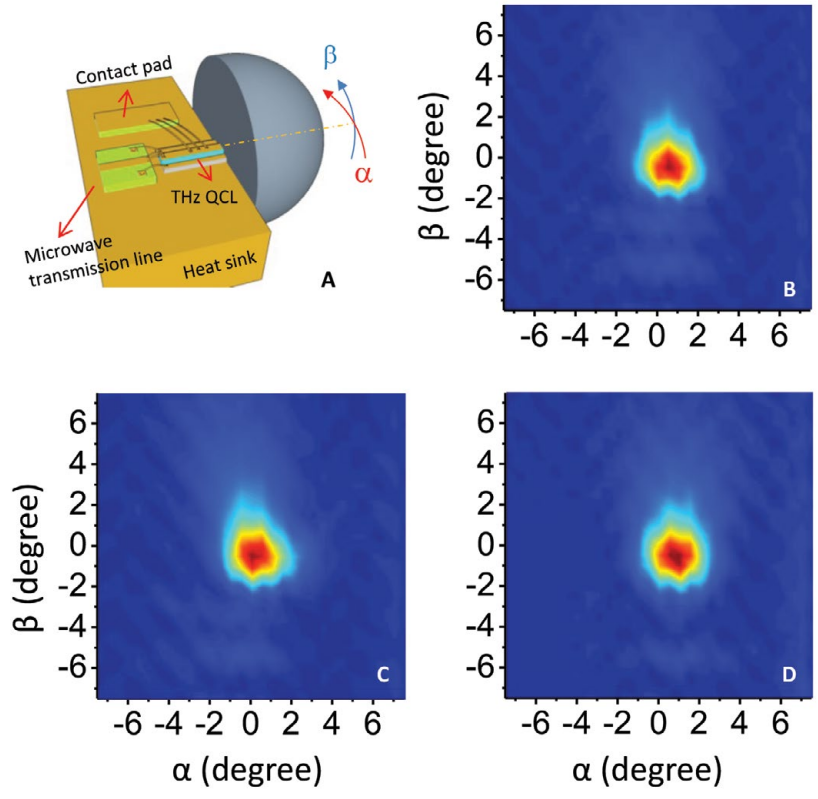
SIMIT has been conducting research on micro- and millimeter-wave solid-state electronics technology since the 1990s. At the beginning of the 21st century, SIMIT took the initiative to develop a K_a-band miniaturized anticollision radar in China, and has carried out R&D of GaAs monolithic millimeter-wave ICs since 2003. At present, K-, K_a-, V-, E-, and W-band series of monolithic millimeter integrated circuit (MMIC) chipsets, for which SIMIT holds intellectual property rights, have been manufactured and applied to millimeter-wave communication and detection systems. Additionally, SIMIT has produced China's first 60-GHz communication chipsets based on vector modulation technology, and an MMIC chip applied to 5G communication has been successfully supplied to Huawei Technologies Co., Ltd. SIMIT also successfully installed a K_a-, V-, and W-band miniature radar detector. Furthermore, SIMIT has manufactured patented human body security inspection imaging equipment, called SimImage, which won the silver award at the 2016 China International Industry Fair. SIMIT successfully completed the transfer of this technology from research to commercialization. SIMIT also performed research on THz solid-state electronic devices and circuits, Schottky diode devices, and related circuits based on GaAs and indium phosphate materials. The knowledge

derived from this work allowed for the fabrication of an integrated mixer, a frequency multiplier circuit, and a module for THz imaging using THz solid-state frequency multiplication chain technology, based on Schottky diode technology. The relevant indexes—such as cutoff frequency and junction capacitance—have reached an advanced level by international standards, and achievements include the construction of a 0.36-THz imaging system.

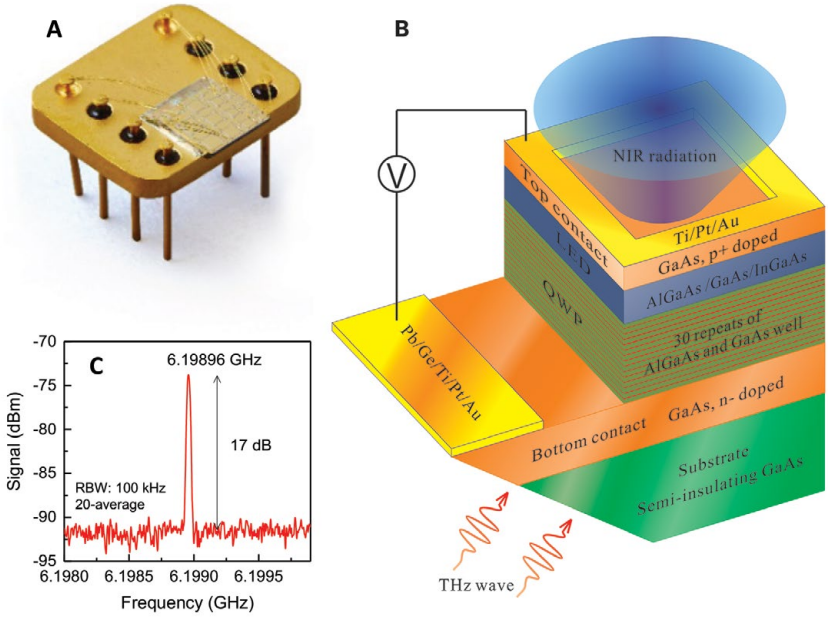
In the future, through cutting-edge research based on high-performance THz QCL frequency combs and fast THz QWPs, SIMIT plans to systematically investigate the saturation absorption, two-photon effect, and fast relaxation process in new materials illuminated by strong THz radiation. This research will provide a foundation for developing new functional devices operating in the THz range. Regarding industrialization, SIMIT will continue development of compact, high-performance THz QCLs, THz QWP modules, and systems for security imaging and space applications, and will promote market-based applications of SimImage millimeter-wave human body security inspection imaging equipment.

Brain-inspired chips and bionic vision

The field of brain science is currently attracting intense international competition. As a major application in this field, brain-inspired intelligence research aims to design formats that can simulate or even transcend human intelligence by emulating neural pathways, endowing a machine with the ability to actively acquire information, formulate concepts, perform analyses, and make decisions in a complex environment. As opposed to artificial intelligence, brain-inspired intelligence is based on human traits and characteristics. Its development is expected to lead to a new era of advanced machine intelligence. As part of SIMIT's cutting-edge research in information science, it has been investigating brain-inspired intelligence since 2012 and has gathered an international research team. The team began its study by imitating the human visual system, and now forms the backbone of the CAS Center for Excellence in Brain Science and Intelligence Technology

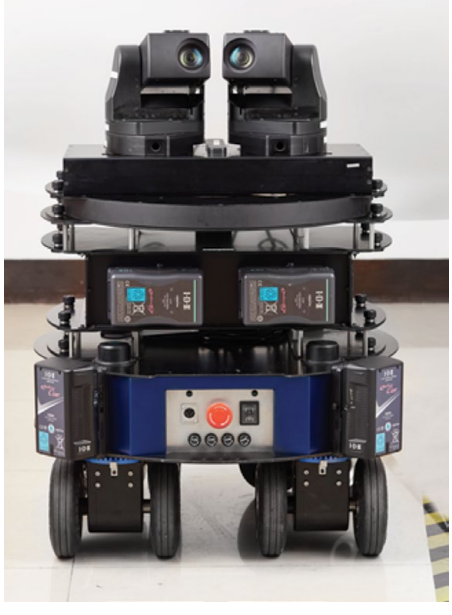


Narrow beam divergence of double-metal terahertz (Thz) quantum cascade lasers (QCLs) (*18*). (**A**) Schematic illustration of a QCL mounted with a silicon lens. **B**, **C**, and **D** show the measured far-field beam patterns of the laser at different drive currents.



A THz quantum-well photodetector (QWP) (*19*, *20*). (**A**) Photograph of THz QWP showing a 45° edge facet. (**B**) Schematic of a THz QWP-light-emitting diode (LED) device. (**C**) Intermode beat-note spectrum of THz QCL measured using a fast THz QWP. NIR, near infrared; Ga, gallium; As, arsenic; Ti, titanium; Pt, platinum; Ac, actinium; Al, aluminum; In, indium; dB, decibel; dBm, decibel-milliwatts; RBW, resolution bandwidth.

SimImage millimeter-wave human-body security inspection imaging equipment



Bionic vision demonstrated on a mobile robot

An example of bionic human eyes

(CEBSIT). Moreover, SIMIT also plays a critical role in the "Shanghai Brain-Intelligence Engineering" group, which undertakes research on human bionic vision systems and vision intelligence. To solidify SIMIT's contribution to the overall national strategy for brain research and provide reliable technical support in bionic vision and brain-inspired intelligence, it created the Bionic Vision System Laboratory.

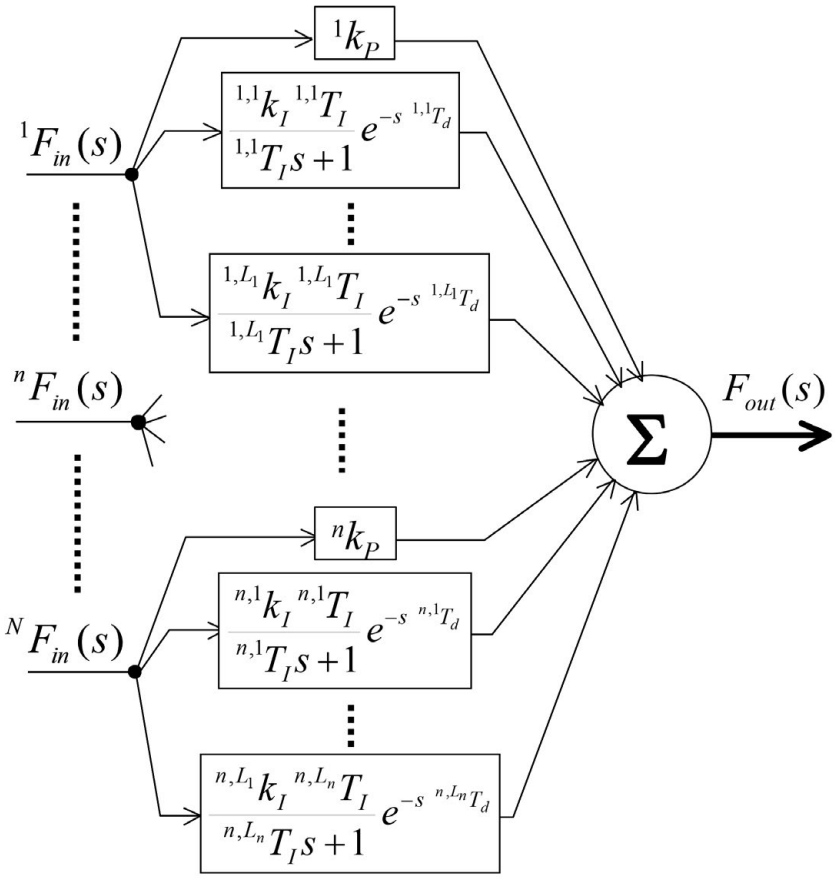
The human visual system is an integrated, intelligent system that includes information acquisition (retina and vestibular apparatus), motion control (cerebellum and brainstem), and signal processing (occipital lobe, superior colliculus, and other brain regions). It is also a representative example of what brain-inspired intelligence research is attempting to replicate. The Bionic Vision System Laboratory has succeeded in the step-by-step mathematical modeling and function simulation of the human visual system. Based on its experience of the physiological and physical characteristics of cerebellum cells, SIMIT constructed a mathematical model of a single cell and an equivalent circuit. This work was published in *Nature Precedings* in 2008 (*21*). In the paper, a neuron was considered as a signal processing unit. By analyzing the relationship between the input of a synapse and the currents across postsynaptic membranes, a dynamic pulse frequency model of a neuron could be obtained. Subsequently, the team has been researching brain-like chips. A brain-like chip architecture is a new type of microchip architecture that mimics the human brain. The processor is similar to a neuron, and the memory unit and communication system resemble a synapse and nerve fiber, respectively. The chip integrates neuronal information processing, synaptic information recording, and axon information transfer in almost the same manner as the human brain. Thus far, SIMIT has established the calculation functions of an equivalent circuit (a theoretical circuit that retains all electrical characteristics of a given circuit), involving the plus, minus, proportional, and integral functions for the pulse frequency, and has employed them to construct a practical motor neuron control system. The team analyzed how signals from the eye travel to the brain stem and impact control of the nervous system. After over 30 years of effort, researchers at SIMIT have come to understand the core theory of binocular vision control and have established mathematical models for control of human visual neurons and binocular vision. These models have been continuously improved and refined through practical engineering applications.

SIMIT has developed several bionic visual products and technologies thus far, including video capture, image processing, and video display. Focusing on the brain, which is the most biologically advanced center for signal processing, SIMIT has proposed a "Machine Consciousness Space" concept. The SIMIT team is intent on mapping the real world to the "consciousness space" of a machine, thus it has developed a series of core algorithms based on binocular vision, including algorithms for high-precision 3D reconstruction, visual odometry, visual saliency, road detection, target detection, and tracking, among others.

The bionic vision team at SIMIT has independently developed a bionic machine-brain system tailored to

performing the functions of perception, interaction, analysis, and determination. This system is modeled on brain structure, complete with binocular vision hardware, intelligence processing, and control software corresponding to the functions of the cerebellum, midbrain, and brainstem, respectively. The proposed system demonstrates outstanding abilities including human-like visual perception and interaction, 3D reconstruction of the consciousness space, and cognition and determination in dynamic environments.

The bionic vision team at SIMIT owns extensive intellectual property in the field of bionic visual control systems. It has applied for 46 patents internationally, of which 25 have been granted, in countries including China, the United States, and Japan. The core vision technology system integrates diverse technologies including ophthalmology, robotics, artificial intelligence, and computer vision, demonstrating its complexity and pointing to the high level of basic research and expertise needed for its construction. This technology has been broadly commercialized, with applications in the fields of 3D photography, robotic vision, intelligent security, virtual reality, intelligent transportation, IoT, and smart health care, among others.



Mathematical model of a neuron (*21*)

Looking to the future, SIMIT will continue to explore the mysteries of brain-inspired intelligence and begin to more deeply analyze the structure of the cerebellum, including the principles of learning and adaption, with the aim of designing a new mathematical model of neural networks that can be implemented in future hardware devices. In addition, SIMIT plans to design an equivalent circuit and a chip-based binocular motion control model. Moreover, SIMIT will focus on developing a new and efficient multimodal information fusion model based on the information integration mechanism found in the brain, in order to build an advanced brain-inspired intelligence system that integrates vision, auditognosis, and tactus based on a visual-processing mechanism. By exploiting the resources of cloud computing platforms, the machine brain will be extended to a cloud brain with decision-making capacity, to promote development toward a cloud-based intelligence platform.

References

1. W. J. Zhang *et al., Sci. China Phys. Mech.* **60**, 120314 (2017).
2. H. Zhou *et al., Opt. Express* **23**, 14603–14611 (2015).
3. X. Zhang *et al., Physica C Supercond.* **548**, 1–4 (2018).
4. Q. Liu *et al., Appl. Phys. Lett.* **110**, 222604 (2017).
5. L. Chen, H. Wang, X. Liu, L. Wu, Z. Wang, *Nano Lett.* **16**, 7726–7730 (2016).
6. D. Jiang *et al., Nat. Commun.* **5**, 5708 (2014).
7. F. H. Ji *et al., Phys. Rev. Lett.* **116**, 177601 (2016).
8. J. Wang, X. Li, *J. Microelectromech. Syst.* **24**, 531–533 (2015).
9. J. Jiang *et al., Adv. Mater.* **2018**, e1705919 (2018).
10. J. Yao *et al., Anal. Chem.* **88**, 2497–2501 (2016).
11. J. Akola, R. O. Jones, *Science* **358**, 1386 (2017).
12. M. Zhu *et al.*, *Nat. Commun.* **5**, 4086 (2014).
13. F. Rao *et al.*, *Nat. Commun.* **6**, 10040 (2015).
14. G. Liu *et al., Appl. Phys. Lett.* **111**, 252102 (2017).
15. Y. Lei, *et al., IEEE Trans. Circuits Syst. II, Exp. Briefs* **65**, 486–490 (2018).
16. F. Rao *et al.*, *Science* **358**, 1423–1427 (2017).
17. W. Wan, T. Zhou, J. C. Cao, *Sci. Rep.* **7**, 44109 (2017).
18. W. Wan, H. Li, J. C. Cao, *Opt. Express* **26**, 980–989 (2018).
19. Z. Fu *et al.*, *Sci. Rep.* **6**, 25383 (2016).
20. H. Li *et al., Sci. Rep.* **7**, 3452 (2017).
21. X. Zhang, *Nature Precedings*, hdl: 10101/npre.2008.1703.1 (2008).

Research platforms at SIMIT

In order to support existing research activities, SIMIT has established five major research platforms to meet the need for rapid development of materials investigation and characterization, device fabrication and integration, and system testing and verification. The platforms are summarized in the table and described in more detail below.

Platforms at SIMIT	Overview
8-in. "More-than-Moore" R&D Pilot Line	Over 5,000 m^2 of cleanroom space, more than 100 installed tools for wafer processing and packaging, and an established suite of micro/nano process flows. Forms an integral part of the microfabrication consortium and provides leading-edge integrated circuit (IC) capabilities in microelectromechanical systems (MEMS), silicon photonics, radio frequency (RF) communication, complementary metal-oxide semiconductor (CMOS) logic, silicon-based III–V power supplies, three-dimensional (3D) integration, tunneling magnetoresistance (TMR) magnetic sensing, and insulated-gate bipolar transistor (IGBT) back-end technologies.
Silicon-Based Photonics Technology Platform	Including standard wafer cleaning, ion implantation, diffusion and oxidation, silicon-germanium (SiGe) epitaxy, rapid thermal annealing of signal-to-distortion spikes, and metal/silicide for good contact, metallization, and dielectric [low-kappa (low-κ)] deposition; plasma etching for poly-Si and dielectrics; chemical-mechanical polishing; and inline metrology/analysis.
Shanghai Integrated Platform for Materials of Energy and Environment	The first synchrotron-based integrated research platform that combines molecular beam epitaxy (MBE)/oxide MBE (oMBE), angle-resolved photoemission spectroscopy (ARPES), scanning tunneling microscopy and spectroscopy (STM/STS), ambient-pressure photoemission spectroscopy (AP-PES), and ambient-pressure photon-in/photon-out spectroscopy (AP-PIPOS) techniques.
Superconducting ELectronics Facility (SELF)	One of the world's most advanced R&D platforms for superconducting electronic devices and circuits. Includes a 650-m^2 cleanroom facility (with a 150-m^2 class-100 section) and advanced micro-/nanofabrication equipment such as a multichamber magnetron sputtering system, high-vacuum evaporator, electron-beam lithography system, i-line stepper, mask aligner, coater/developer, reactive-ion etching system, X-ray diffractometer, atomic force microscope, and scanning electron microscope (SEM).
Microstructural Characterization Platform of Functional Materials for Informatics	Provides a series of characterization techniques ranging from microstructure and composition analysis and electrical performance analysis of materials to atomic and electronic structure analysis of materials. Includes double spherical aberration coefficient (Cs)-corrected transmission electron microscope (TEM), 3D atom probe, low-temperature/high magnetic field scanning microwave impedance microscope, focused ion beam SEM system, field emission TEM, and field emission SEM.

The 8-in. "More-than-Moore" R&D pilot production line

8-in. "More-than-Moore" R&D Pilot Line

Shanghai is a leading base for the semiconductor industry in China and home to many global enterprises in industries such as mobile phone and automobile manufacturers. It is therefore an ideal development environment for the global microelectromechanical systems (MEMS) industry, and a primary reason why the Shanghai Industrial μTechnology Research Institute (SITRI) was created by SIMIT. SITRI has taken the next step in driving the development of the MEMS industry by investing in the construction of an industry leading R&D pilot line for MEMS and other "More-than-Moore" (MtM) technologies.

The 8-in. MtM R&D pilot production line in Shanghai has over 5,000 m^2 of cleanroom space, more than 100 machines for wafer processing and packaging, and an established suite of micro/nano process flows. The line is compatible with the production capabilities of SIMIT's foundry partners, to ensure seamless transfer from R&D to high-volume production, particularly for MtM and 3D integration technologies. It is an integral part of the pilot manufacturing consortium and provides leading-edge IC capabilities in MEMS, silicon photonics, RF, CMOS logic, silicon-based III–V, 3D integration, TMR magnetic sensing, and IGBT back-end technologies.

This R&D pilot line enables SIMIT to support state-of-the-art research in advanced materials and sensor prototyping by working with its research and industry partners through an open R&D model to deliver commercially ready products to the market—products that are readily available to scale up to mass production. It is not only an important bridge between R&D and manufacturing, but also a crucial platform to enrich and improve the MtM ecosystem in China, and a key strategy to help domestic innovations gain a foothold in global markets.

Demonstrated SIMIT capabilities

Products

- Inertial sensors
- Environment sensors
- Acoustic sensors
- Microstructures
- Magnetic sensors
- TMR sensors
- Infrared radiation CMOS image sensor (IR-CIS)
- Actuators
- Power devices (RF, III–V)

Universal platforms

- Compatible with CMOS back-end-of-line
- Redistribution layer technology
- Through-silicon vias (vertical interconnect access) technology

Advantages

- Fast feedback and development cycle
- Customized and flexible process development
- Experienced technology team and consultants
- Intellectual property protection in both development and transfer
- Full analysis, testing, and characterization capabilities

Silicon-Based Photonics Technology Platform

In 2004, Intel released the first high-speed GHz silicon modulator developed using the CMOS platform, leading silicon-based photonics into a new era. Semiconductor giants including Intel and IBM, as well as top universities and institutes worldwide, played important roles in this technological breakthrough. High-performance devices including silicon electro-optical modulators, germanium-based high-speed optical detectors, low-loss silicon nanowire waveguides, tapered and grating couplers, wavelength division multiplexer components, and polarization-related devices, among others, have been made possible through advances in silicon usage, which has also enabled the production of large-scale integrated photonics circuits and optoelectronics integration. The potential of a silicon laser is still being researched because silicon is not a direct bandgap material (the bandgap is the range of energy levels within a given solid that are not possible for an electron to possess). Great strides have been made in this field through the use of hybrid bonding, quantum dot epitaxy, and bandgap engineering. Presently, driven by industrial applications such as data communications, telecommunications, autopiloting, and physical data collection (such as temperature), CMOS platforms are urgently needed for R&D into integrated silicon photonics chips. However, commercial CMOS fabrications cannot provide adequate support for silicon photonics due to process capability overload. Moreover, the Interuniversity Microelectronics Centre (IMEC) and A*STAR's Institute of Microelectronics (IME) are the only silicon photonics fabrication systems open to international users, and their availability is limited. Therefore, SIMIT initiated a 200-mm silicon photonics pilot production line in collaboration with the local government in Shanghai. This production system shares the front-end-of-line with the 200-mm MtM line. The back-end-of-line is dedicated to the silicon photonics process, in order to guarantee the quality of the fabrication by avoiding contamination with other materials and metals.

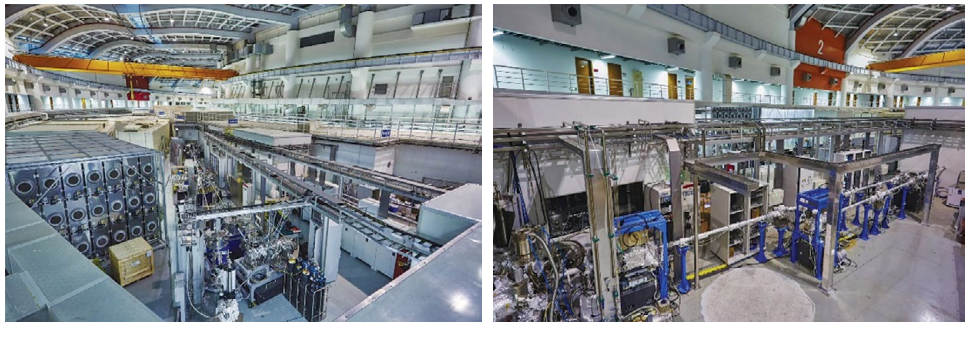
The silicon photonics pilot platform is targeted at R&D and small-volume production (~500 pieces/month) for the industry. The core process blocks include standard wafer cleaning, ion implantation, diffusion and oxidation, SiGe epitaxy, rapid thermal annealing for signal-to-distortion spikes, and metal/silicide for good contact, metallization, and dielectric (low-κ) deposition; plasma etching for poly-Si and dielectrics; chemical-mechanical polishing; and inline metrology/analysis. Equipped with 248-nm and 193-nm deep-ultraviolet photolithography systems, it can fabricate chips with a critical dimension of 90 nm, and has a passive and active device library, supporting silicon photonics products over 25 gigabits per second (Gbps) for the high-speed optical module at 100 Gbps and beyond. The platform provides a process design kit for standard fabrication and also supports customized process development. Typical products that the team can support include integrated chips for high-speed transceivers, light detection and ranging (LiDAR) devices, biosensing technologies, and integrated microwave photonics, along with fundamental research into photonic crystals, metamaterials, and plasmonics.

Multiproject wafer and customized process services are open to international customers, and these processes are available for transfer to mass production fabrications through further collaboration. 3D/2.5D optoelectronic interconnection is another important technology that is being developed on the platform, which offers China's most economical, high-performance solution for the interconnection of individually fabricated optical and electronic state-of-the-art parts. SIMIT/SITRI aims to build the first dedicated silicon photonics technology platform in China by combining advanced research, multiproject wafer services, training, and client product support.

The operations team has several years of experience in process development and technical collaboration with global silicon photonics companies, aiding them in moving their product from prototype to mass production. Furthermore, Shanghai Simgui Co., Ltd., another innovation partner and a spinoff of SIMIT, is equipped to supply customized silicon-on-insulator (SOI) wafers such as multilayered SOI and cavity SOI, thereby offering more flexibility in the design and fabrication of integrated photonics chips. An advanced silicon photonics project will be launched next year with the support of the Shanghai government, which aims to promote the city's optical communication industry by speeding up domestic production of integrated optoelectronic chips and building the silicon photonics industrial chain and ecosystem. The Shanghai municipal government is committed to building a world-class silicon photonics R&D base there, and the pilot production line is one of the most important parts of this project, together with the development of optoelectronics hybrid packaging, high-speed IC drivers, and high-performance integrated photonic chips.

Shanghai Integrated Platform for Materials of Energy and Environment (SiP.ME^2)

Traditional methods for studying the electronic structure of materials are restricted by the lack of progress in advanced light sources such as synchrotron radiation. The relatively homogenous information is inadequate to satisfy the needs of the rapidly developing field of modern energy and environmental materials science. SiP.ME^2 is the first synchrotron-based integrated research platform that combines molecular beam epitaxy (MBE)/oxide MBE (oMBE), angle-resolved photoemission spectroscopy (ARPES), scanning tunneling microscopy and spectroscopy (STM/STS), ambient-pressure photoemission spectroscopy (AP-PES), and ambient-pressure photon-in/photon-out spectroscopy (AP-PIPOS) techniques. Samples in this platform can be switched between these different measuring techniques with the aid



SiP.ME^2 beam lines and stations for in situ characterization of the electronic structure of energy and environmental materials

of the in situ or movable vacuum sample transfer device. Each part of the platform has an independent function and can handle the needs of independent research tasks. When integrated as a whole, the system's capability is significantly greater than that of each subsystem alone.

Through the latest MBE/laser MBE integrations, the growth of a material can be precisely controlled at the cell or even atomic level. The platform provides comprehensive control of all aspects of production, including the end face, surface, interface, composition, stoichiometry, structure distortion, and strain on the material, and can also enable the creation of artificial microstructures through a thermal nonequilibrium process. Therefore, not only does it circumvent restrictions encountered with traditional electronic structure research, but also shifts traditional ideas about materials research. We believe its operation will advance innovative, multidisciplinary materials design, fabrication, and characterization, significantly broadening the scope and capabilities of research into the structure of electronic materials.

The Superconducting Electronics Facility

The platform allows for in situ characterization at near-atmospheric pressure, enabling the study of the electronic structure of materials under different states, different atmospheres, and different loading conditions. For example, near-atmospheric pressure technology can measure the electronic structure of the sample surface in the presence of liquid water, which provides a novel method for application in water science, environmental science, corrosion protection, and other fields important to China. Near-atmospheric pressure is useful for analyzing in situ loading of materials and devices, and can simulate the running conditions of materials and devices. Therefore, the material surface can be measured under nonequilibrium conditions, and physical and chemical properties of materials can be accurately and reliably obtained. The platform can also provide scientific evidence for the control and improvement of new energy materials.

Incorporation of advanced light sources, including X-ray synchrotron radiation light sources, deep ultraviolet lasers, and subfemtosecond pulse light sources can significantly improve the platform's performance, overcoming the limitations of traditional ultrahigh vacuum ARPES measurement technology. It can enable the difference in electronic effect between surface and bulk materials to be distinguished, and promote the development of measurement technologies for interface electronic structure and ultrafast electron spectra.

The platform offers excellent complementarities for electronic structure measurements and analyses. Photoemission spectroscopy (PES) provides electronic structure information at the valence and core level, X-ray absorption spectroscopy (XAS) provides unoccupied state information, X-ray emission spectroscopy (XES) yields occupied state information, and the combination of ARPES and STS gives information about electron states with near-Fermi level accuracy. Thus, complete information on the electronic structure of materials can be obtained from this platform.

Utilizing the latest advances in controllable material growth, precise electronic state measurement, and advanced light sources, this platform is capable of in situ and/or in operando measurement of the electronic structure of materials with ultrahigh resolution, and has great potential for application in the fundamental and applied research of new energy and environmental materials.

Superconducting ELectronics Facility (SELF)

Since 2005, SIMIT began research on low-critical temperature (low-T_c) niobium (Nb)-based superconducting electronic devices and their applications in superconducting quantum interference devices (SQUIDs) and superconducting nanowire single-photon detectors (SNSPDs). It is well-known that device fabrication technology is one of the key issues facing superconducting electronics. It is critical to establish advanced comprehensive device fabrication facilities and develop basic technology for creating superconducting devices and circuits in China. Achieving this will significantly promote R&D of domestic superconducting electronic devices and subsequent cutting-edge applications in geological surveying, biomagnetic imaging, and quantum communication. For these reasons, the Chinese Academy of Sciences (CAS) began a support project in 2010 to establish a micro-/nanofabrication facility for superconducting devices and circuits, called the Superconducting ELectronics Facility (SELF). SELF aims to establish fabrication and characterization capabilities and develop process technologies with high reliability and stability for SQUID, SNPSD, transition-edge sensors, superconducting quantum bit (qubit), and single flux quantum (SFQ) superconducting devices and circuits, among others.

SELF is among the world's most advanced R&D platforms for superconducting electronic devices and circuits. A RMB

A selection of the equipment available within the Microstructural Characterization Platform of Functional Materials for Informatics. SEM, scanning electron microscope; 3DAP, three-dimensional atom probe; TEM, transmission electron microscope; Cs-TEM, spherical aberration coefficient TEM; FIB-SEM, focused ion beam SEM; sMIM, scanning microwave impedance microscope.

100 million (USD 15.8 million) investment has been committed for phase one of SELF, which includes a 650-m^2 cleanroom facility (with a 150-m^2 class-100 section) and advanced micro-/nanofabrication equipment such as a multichamber magnetron sputtering system (to fabricate Nb, Nb/AlOx/Nb, NbN, NbN/AlN/NbN, and NbTiN films), a high-vacuum evaporator, an electron beam lithography system (8-nm resolution and 9-nm overlay accuracy), an i-line stepper (0.35-μm resolution and 40-nm overlay accuracy), a mask aligner (2-μm resolution and 0.5-μm overlay accuracy), a coater/developer, a reactive-ion etching system, an X-ray diffractometer, an atomic force microscope, and a SEM. SELF can perform micro-, submicro-, and nanoscale patterning with high overlay accuracy as well as fabricate superconducting devices and circuits on 2-, 4-, and 6-in. substrates. Approximately RMB 70 million (USD 11 million) has been invested in phase two of SELF, which will add another 600-m^2 cleanroom as well as multiple pieces of advanced process equipment including a chemical-mechanical polisher, an ultrahigh vacuum (UHV) sputtering cluster, a UHV evaporation cluster, a plasma-enhanced chemical vapor deposition system, an ion-beam etching system, an inductively coupled plasma etching system, and a deep silicon etching system. After the construction of phase two, SELF will be a state-of-the-art platform for the fabrication of high-performance superconducting devices and large-scale integrated digital circuits. With support from SELF, SIMIT has successfully developed Nb- and NbN-based Josephson junctions, SQUIDs/nanoSQUIDs, and high-efficiency SNSPDs for a variety of applications.

SELF is an open, shared platform, available on a fee basis or for use in collaborative research projects by government, industry, and universities. We expect and encourage scientists worldwide to utilize SELF.

Microstructural Characterization Platform of Functional Materials for Informatics

With the increase in research into functional micro-/nanostructure materials for informatics, the new micro-/nanoscale characterization technology has grown in importance in materials research. The development of nanoscale detection and characterization technologies—used to observe the atomic and electronic structures of nanostructural materials and measure the force, electrical, light, and magnetic properties of various nanostructures—is an essential function of the Microstructural Characterization Platform of Functional Materials for Informatics. SIMIT established the platform more than a decade ago, and it now provides a series of characterization techniques ranging from microstructure and composition analysis to electrical performance analysis to atomic and electronic structure analysis. The platform can provide deep technical support for scientific research in many fields. It can simultaneously provide high-level testing and consulting services as well as training for research institutes, universities, and enterprises.

The platform focuses on solving important, basic scientific problems in the development and application of functional materials for informatics. It is equipped with a range of high-end equipment, including an 80-picometer ultrahigh-resolution double spherical aberration coefficient (Cs)-corrected TEM that can be used to observe atomic images at nanometer scale and perform a range of in situ experiments. The platform also boasts a 3D atom probe that can analyze the elemental makeup of complex structures, intuitively reconstruct the 3D distribution pattern of the different elements in nanometer space, and provide precise element space content. Other technology includes a low-temperature/high magnetic field scanning microwave impedance microscope, a focused ion beam–scanning electron microscope, a field emission TEM, and a field emission SEM.

Since the completion of the platform, it has been providing extensive technical support for scientific research in many areas, such as superconducting materials and electronics applications, advanced silicon-based materials and applications, novel nanoelectronic materials and devices, and compound semiconductor materials, devices, and applications. It has promoted the rapid development of world-class, original scientific research into functional materials for informatics.

"Three-in-One" Collaborative Innovation System

With the rapid development of China's economy, the Chinese Academy of Sciences (CAS) has adopted a guiding principle for a new area of pioneering research focused on the frontiers of science, national priorities, and economic development, and is committed to achieving breakthrough developments in science and technology (S&T). To promote economic progress and make significant contributions to the development of emerging industries, CAS initiated the knowledge innovation program Innovation 2020, which aims to further promote innovation and attempts to turn scientific discoveries into technologies that power economic growth and sustainable development.

During its 90 years of development, SIMIT has assumed a key role as a driver of national technological innovation and supporter of nationwide S&T development. SIMIT has defined its development strategy by emphasizing greater reliance on innovation to meet national strategic priorities. International collaboration helps SIMIT access global knowledge and innovation environments. Furthermore, SIMIT develops research from an international perspective. In July 2017, SIMIT was approved by the General Office of the State Council of the People's Republic of China as a "Demonstration Base of Mass Innovation and Entrepreneurship," to undertake the mission of continuous innovation and promote the development of entrepreneurship.

Establishment of the "Three-in-One" Collaborative Innovation System to improve the R&D ecosystem

Building upon SIMIT's years of experience in scientific research and industrialization, and to demonstrate its support for the Innovation 2020 initiative, the "Three-in-One" Collaborative Innovation System was established, with SIMIT as the center of advanced research, the Shanghai Industrial μTechnology Research Institute (SITRI) as the technology transfer and pilot platform, and SIMIC Holdings as the industrialization platform focused on capital operation.

SIMIT: Advanced research

SIMIT contains several state key laboratories and CAS key laboratories that provide strong R&D capabilities for basic S&T, while the Center for Excellence in Superconducting Electronics conducts research in strategic advanced technologies. In recent years, SIMIT has focused its research on materials and information S&T, to promote the technology transfer of new innovations from CAS. Additionally, the laboratories, in concert with the SIMIT platforms, are undertaking advanced research that improves the analysis of how emerging industries impact public welfare.

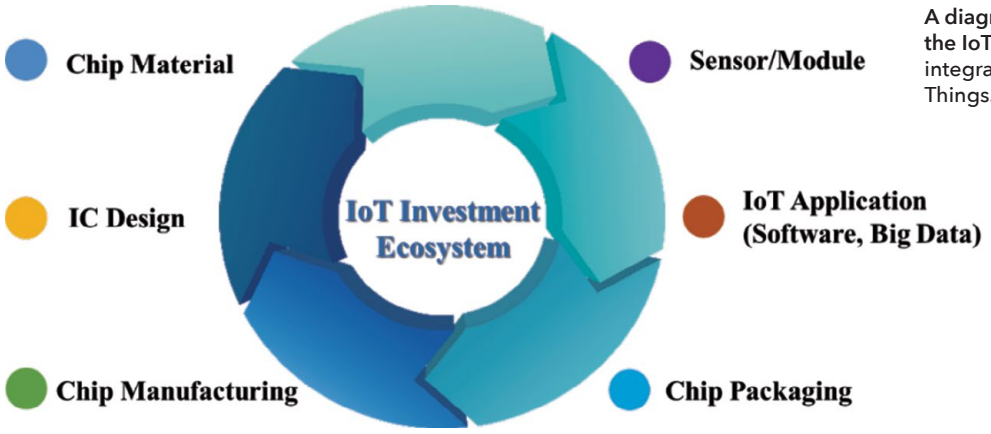
"Three-in-One" Collaborative Innovation System at SIMIT

SITRI: Accelerating technology development and commercialization

In 2013, SIMIT established SITRI, a new type of international innovation center committed to accelerating innovation and commercialization of "More-than-Moore" technologies to power the Internet of things (IoT). SITRI serves as the technology transfer and pilot platform for commercial partners and startups, and was approved by the State Council as the lead institute for the construction of the Shanghai Technology Innovation Center.

SIMIC Holdings: Leading industrialization

SIMIC Holdings Co., Ltd. is the industrialization and asset management platform of SIMIT. Thus far, SIMIC Holdings has invested over RMB 330 million (USD 52 million) in S&T programs, and has raised over RMB 20 billion (USD 3.1 billion) in outside investment to promote the industrialization of S&T innovations. SIMIC Holdings has established a successful business model consisting of technology R&D (SIMIC R&D), incubation (SIMIC incubator), venture capital (SIMIC capital), and an industrial cluster (SIMIC industry) to support the rapid development of innovative high-tech enterprises.



A diagrammatic representation of the IoT investment ecosystem. IC, integrated circuit; IoT, Internet of Things.

Three platforms supporting technological innovation

The carrier platforms

SIMIT has three campuses, in Changning, Jiading, and Xuhui districts, which have all assumed the role of incubator platforms to support the "Three-in-One" Collaborative Innovation System. SIMIC Holdings established the Shanghai IoT Center in the Jiading district, where the planned Sino-Finnish Innovation Science and Technology Industrial Park will be located, with the aim of creating an international innovation and entrepreneurship site. To meet the needs of the international R&D institutions and companies incubated in SIMIC Innospring (located in the Shanghai IoT Center), SIMIT created an advanced platform for intelligent sensor and IoT development as well as a professional incubator and makerspace for intelligent sensors and IoT at the Changning and Xuhui campuses, which are intended to provide professional services. Similar resources related to sensor and IoT technology will also be integrated in the "Three-in-One" Collaborative Innovation System.

The technology platforms

The establishment of open technology platforms for sensor, silicon photonic, and superconducting quantum communication has been supported through the "Three-in-One" Collaborative Innovation System, together with platforms for technology R&D and technology transfer, and the incubation of advanced R&D platforms. The development of the 8-in. "More-than-Moore" R&D pilot production line, superconductor production line, and 5G system platform have all contributed significantly to entrepreneurship and the emergence of new industries. Cooperation between the technology platforms and domestic research institutes, universities, and foreign research institutions has been strengthened, leading to the growth and increased capabilities of these platforms.

The funding platforms

The IoT Fund Ecosystem was established to support entrepreneurship among high-tech IoT enterprises at all stages of development. The system allows for complication-free investment in quality projects. It employs a collaborative process across the different stages of the funding process (see Figure 1), encouraging the support of quality projects while controlling investment costs. It also engenders improved social benefits and has seen superior economic returns.

Through the "Three-in-One" Collaborative Innovation System, the IoT Fund taps into the resources of the Yangtze River Delta, including technology, talent, and capital, to strengthen the fund ecosystem. By promoting rapid innovation in sensor technology, IoT, and information and communication technology, the IoT industry has made significant strides in recent years. Investing in enterprises at all stages of product development diversifies investment risk. The IoT Fund will act as a bridge to connect those enterprises in which SIMIT has invested with those supported by the fund as well as the fund's shareholders and partners. These connections will create synergies across the IoT ecosystem, covering upstream design, manufacture, packing, and testing of materials; midstream production of core components, sensors, modules, and network transmission devices; and downstream software application, data processing, problem solving, system integration, technology services, and software operation.

Replication of the "Three-in-One" innovation

The Fuzhou IoT Open Laboratory (FIOT-LAB) was established jointly by SIMIT and the Fuzhou Municipal Government in April 2017, based on the "Three-in-One" Collaborative Innovation System. FIOT-LAB focuses on narrowband IoT (NB-IoT) technology R&D, which enables multiple IoT devices to be connected through cellular networks. Incubation of new industry startups is an important means to accelerate the growth of the IoT industry. FIOT-LAB emphasizes the integration of the IoT and semiconductor industries, and aims to become a national base for NB-IoT innovation.

With minimal effort, the "Three-in-One" Collaborative Innovation System could be replicated across China in cities like Chongqing and Nantong. SIMIT will continue to consider innovation as a formula for success as well as an effective means to advance science and drive R&D that can power economic expansion and sustainable development. Against the background of Shanghai as an innovation center with global influence, the system will play an important role in supporting innovation that contributes to China's economic and technological growth.

International cooperation

SIMIT is committed to establishing international cooperation with top universities and research centers worldwide. Their collaborations with Forschungszentrum Jülich GmbH (FZJ) in electronics and with VTT Technical Research Centre of Finland and the Finland National Bureau of Technical Innovation in information technology began over 10 years ago.

Map showing location of SIMIT's international partners

Forschungszentrum Jülich GmbH

SIMIT is proud of its long-standing collaboration with FZJ to advance the field of electronics. FZJ, an affiliate of the Helmholtz Association, was established in 1956. It has more than 5,700 researchers and a strong reputation for cutting-edge scientific research, particularly in information technology, neuroscience, and energy and the environment.

SIMIT's collaboration with FZJ began in 2006 and has included personnel exchanges and cooperation on various projects, including the establishment of joint laboratories on superconducting and bioelectronics in 2010, and on quantum materials in 2014. SIMIT and FJZ signed an agreement to engage in the joint construction of the "Sino-German Virtual Joint Institute on Functional Materials and Electronics" in 2015. SIMIT and FZJ have undertaken four international programs, published 47 papers, generated 20 meeting reports, applied for two patents (one of which was granted), and have had seven joint Ph.D. students (four of whom were hired as assistant professors by SIMIT). Furthermore, they alternatively held seven workshops that provided the Sino-German scientific researchers with an opportunity to present their findings. In 2013, ZHANG Yi from FZJ won the Shanghai International Science and Technology Cooperation Award, while in 2014, YU Yuehui from SIMIT won the Helmholtz International Fellow Award.

On October 20, 2016, a high-level delegation comprising the chairman of the Helmholtz Association, the director of FZJ, and heads of several research centers affiliated with the Helmholtz Association visited SIMIT. The two institutes signed an agreement for a third round of strategic cooperation, under which both parties determined that it was to their mutual benefit to continue their strategic cooperation in areas such as superconducting technology, sensor technology, bioelectronics, energy and the environment, and quantum information. This cooperation will include constructing joint research laboratories and the sharing of research facilities, organizing academic seminars, and the exchange of personnel.

Signing of the strategic cooperation agreement between the Helmholtz Association Forschungszentrum Jülich GmbH and SIMIT; (left) WANG Xi , director of SIMIT, and **(right)** Sebastian M. Schmidt, vice-chairman of Forschungszentrum Jülich GmbH.

Finland

Focusing on the development of the new generation of broadband wireless mobile communication, SIMIT has maintained an excellent cooperative relationship with several countries including Finland, the United Kingdom, Sweden, Canada, and Australia, as well as with enterprises, universities, and research institutions known for their expertise in communication technology. SIMIT has signed cooperation agreements and undertaken joint international science and technology projects sponsored by national ministries and local governments, which have paved the way for the development of sustainable international collaboration.

SIMIT has actively participated in international scientific and technological cooperation and exchange projects sponsored by many countries and regions. Under the guidance of the Chinese Ministry of Science and Technology and the Finnish Ministry of Economic Affairs and Employment, SIMIT (on behalf of China) signed a memorandum of understanding for the Finland-China Strategic ICT Alliance Cooperation. It also officially launched the China-Finland Information and Communications Technology (ICT) Alliance for Production, Teaching, and Research, which marked the formal establishment of intergovernmental cooperation between China and Finland in science and technology innovation. SIMIT has hosted and participated eight times in Sino-Finnish international conferences on ICT, and more than 1,000 experts have been invited to visit SIMIT. This cooperation promotes scientific and technological innovation and knowledge sharing between the two governments, and advances the formulation of domestic and international research and production standards used by industry leaders. It also achieves the goal of using technology as a driver of innovation, and allows for the development of key technologies for wireless mobile communication in China. Recently, SIMIT has been working with the VTT Technical Research Centre of Finland—a state-owned nonprofit research entity—and the Finland National Bureau of Technical Innovation, to promote the construction of the Sino-Finnish Innovation Science and Technology Industrial Park.

SIMIT has actively participated in international scientific and technological cooperation and exchange projects sponsored by many countries and regions.

YU Yuehui (left) from SIMIT receiving the Helmholtz International Fellow Award from Jürgen Mlynek, president of the Helmholtz Association

Some of the attendees at the China-Finland International Cooperation Workshop and kickoff meeting for the collaborative research projects on 5G networks

Research talent at SIMIT

As a national institute, SIMIT is committed to furthering the frontiers of science, advancing national needs, and supporting the domestic economy. To do this, it has built a highly qualified team by providing a clear career structure and ensuring that it constantly improves the work environment it offers.

In September 2009, the Chinese Academy of Sciences (CAS) launched the Talent Cultivation and Recruitment Program, a system-wide project aimed at increasing recruitment and developing talent in various fields, with the overall goal of creating greater adaptability to the constantly changing requirements for expertise in China. Through this program, which includes the national Thousand Talents Plan, the Shanghai Thousand Talents Plan, and the CAS Hundred Talents Program, SIMIT selects top-level scientists who demonstrate ambition, ability, and passion. It also focuses on attracting and sponsoring outstanding overseas scholars and international scientists who are active at the forefront of science and technology to visit or work at SIMIT.

At the end of 2017, SIMIT had a staff of 610, of whom 80 were from outside China. There were 37 high-level researchers, including seven winners of the National Thousand Talents Plan award, two winners of the National Thousand Young Talents Plan award, five winners of the Shanghai Thousand Talents Plan award, and 22 winners of the CAS Hundred Talents Program award.

WANG Xi
Academician of CAS, Materials Scientist

WANG Xi's research interests include electronic materials, novel integrated circuit (IC) processes, and wireless sensor networks. Wang has long engaged in research on the physical phenomenon of the interaction between ion beams and solids under load, and has applied his research findings to the development of silicon-on-insulator (SOI) products. Based on research by Wang, particularly on the physical and chemical processes of ion-implantation SOI synthesis, SIMIT has developed a series of key technologies for industrial production of SOI materials technology, and has established an R&D and production base of SOI materials.

Wang has published over 300 scientific papers and has presented at international scientific conferences. He has received several awards and prizes for his academic achievements including the first prize of the National Science and Technology Progress Award.

ZOU Schichang
Academician of CAS, Materials Scientist

ZOU Shichang engages in the systematic study of ion beam–solid interaction and the use of ion beams for the doping, synthesis, fabrication, and surface layer analysis of semiconductor materials and devices. In the 1990s, Zou was involved in the development of the semiconductor industry in mainland China, and participated in the establishment of several IC companies, including Shanghai Huahong Grace, Shanghai Simconix Electronics Co., Ltd., Toppan Photomasks, and Shanghai Ericsson Simtek Electronics Co., Ltd.

Zou has won 14 awards and prizes, including the first prize for the National Invention Award. He has published over 300 scientific papers and mentored over 30 Ph.D. students. Zou is also an honorary member of the International Conference on Ion Beam Modification of Materials Committee.

LI Aizhen
Foreign Associate of the U.S. National Academy of Sciences, Materials Scientist

LI Aizhen has been engaged in studying semiconductor materials for over 50 years. She founded the Laboratory of Molecular Beam Epitaxy Semiconductor Microstructure Materials and Devices of the Shanghai Institute of Metallurgy, and participated in the establishment of the State Key Laboratory of Functional Materials for Informatics in 1993. From 1958 to 1980, she studied materials science, engineering, and applications of the single-crystal, micron-scale group III–V heterostructure.

Since the 1980s, she has been studying atomic layer molecular beam epitaxy growth and interface control, and group III–V compound semiconductor quantum structure materials, particularly their characterization and application in high-speed electronic devices and detectors. More recently, she has focused her research on mid-infrared band introns and transition quantum-cascade lasers, interband transition multiquantum-well laser materials and devices, and has conducted related physics research.

Li has published 235 papers and three treatises, and has been granted 17 national patents for her inventions. She has won one National Invention Award, four National Scientific and Technological Progress Awards, eight Natural Science Awards at the provincial and ministerial levels, and first and second prizes in the CAS Scientific and Technological Progress Awards, as well as the Golden Bull Award from the Ministry of Science and Technology. She was honored as a Woman Pacesetter of the Nation, a National Urban Female Contribution Model of the Nation, and a Shanghai Model Worker. Other honors received include "National Outstanding Returning Overseas Chinese Academics, and Their Families," and "Advanced Returning Individual Overseas Chinese, and Their Families," as well as honors from the Overseas Chinese Affairs Office of the State Council and the All-China Federation of Returning Overseas Chinese. In addition, she was granted the honorable title of "Outstanding Contribution Teacher" by the CAS Graduate School in 2008, was elected as a foreign associate of the U.S. National Academy of Sciences in May 2007, and won the Third World Academy of Sciences (TWAS) Engineering Science Award in 2004.

High-level staff at SIMIT

WANG Zhen

WANG Zhen received his B.S. from Nanjing University, China, in 1978, and his M.S. and Ph.D. from Nagaoka University of Technology, Japan, in 1988 and 1991, respectively. From 1991 to 2013, he worked at the National Institute of Information and Communications Technology (NICT) in Japan, where he served as a senior researcher, group leader, and distinguished researcher. He led the Superconducting Electronics Group in NICT from 1993 to 2012, and was honored as a NICT life fellow in 2011. After Wang was appointed as a Thousand Talents Plan professor in 2010, he joined SIMIT in 2013, where he is currently leading the Division of Superconducting Devices and Circuits. He is also a professor at the University of CAS and ShanghaiTech University.

Wang has conducted research primarily on the fundamental technology and applications of superconducting electronics, particularly niobium nitride (NbN)-based superconducting films and Josephson junctions, superconductor–insulator–superconductor (SIS) mixers, superconducting nanowire single-photon detectors (SNSPDs), superconducting single flux quantum (SFQ) circuits, and superconducting quantum bit (qubit) devices. He developed epitaxial growth technology for the fabrication of single-crystal NbN films and NbN/aluminum nitride (AlN)/NbN Josephson tunnel junctions. He was the first to demonstrate a low-noise SIS mixer using NbN/AlN/NbN tunnel junctions, and applied the NbN SIS mixers to the submillimeter-wave telescope for astronomical observation in collaboration with SHI Sheng-Cai of Purple Mountain Observatory, CAS, in China. He also pioneered the development of a qubit using NbN/AlN/NbN junctions in collaboration with HAN Siyuan of the University of Kansas and Yoshihiko Nakamura of Tokyo University. Since 2005, Wang and the NICT research group have developed high-performance SNSPD devices using NbN thin films, and applied the SNSPD system in the Tokyo quantum key distribution (QKD) Network, a field demonstration of long-distance QKD.

Wang has published over 300 papers in peer-reviewed academic journals and written chapters in the *Handbook of Thin Films* and the *Handbook of Terahertz Technology*. He has served as a committee member for numerous conferences in the field and chaired the 2nd International Workshop on Superconducting Sensors and Detectors in 2014. He received an award from Japan's Ministry of Posts and Telecommunications in 2000, and from Japan's Ministry of Education, Culture, Sports, Science and Technology in 2011; he also received Japan's Superconducting Science and Technology Award in 2011.

WANG Shumin

WANG Shumin received his B.S. and M.S. in physics from Fudan University, China, in 1985 and 1988, respectively. He earned his Ph.D. from the Department of Physics at Gothenburg University in Sweden, in 1994. He joined Chalmers University of Technology, Sweden, in 1994, and was promoted to associate professor in 1999 and full professor in 2008. In 2013, he was appointed as a Thousand Talents Plan professor at SIMIT. He has also been a professor at ShanghaiTech University since 2013.

Wang is an expert on the epitaxial growth and physics of low-dimensional semiconductor heterostructures and nanostructures. His current research interests include dilute bismide materials and devices, integration of light-emitting materials with silicon, group III–V nanowires and quantum dots, and epitaxial growth of 2D materials. He is the editor of *Lattice Engineering: Technologies and Applications,* has written seven book chapters, published over 420 international journal and conference papers, and holds seven Chinese patents.

Wang has worked as a leader or a principal investigator for over 10 years in Sweden under the European Union's Seventh Framework Programme, as well as on several Chinese national projects. He has been officially invited to nominate outstanding candidates for the Japan Prize and the Nobel Prize in Physics. He was appointed as a management committee member for the European Cooperation in Science and Technology Actions MP0805 and MP1204, and as a foreign reviewer for the Engineering and Physical Sciences Research Council in the United Kingdom, the Agency for Science, Technology, and Research (A*STAR) in Singapore, and the Polish Science Foundation. He has served as a committee member for several international conferences including the 14th (2006), 18th (2014), and 19th (2016) International Conference on Molecular Beam Epitaxy, and the European Molecular Beam Epitaxy Workshop each year since 2011. He was the chair for the 7th International Workshop on Bismuth-Containing Semiconductors in 2016 and has been a senior member of the Institute of Electrical and Electronics Engineers (IEEE) since 2009.

LIU Zhi

LIU Zhi received his B.Sc. in geophysics/physics from Beijing University, and his M.S. in electrical engineering and doctorate in physics from Stanford University. Before joining the Lawrence Berkeley National Laboratory as a career staff scientist, he was a research associate at Stanford University and the Stanford Synchrotron Radiation Laboratory. He is currently a professor at SIMIT and vice dean of the School of Physical Science and Technology at ShanghaiTech University. Liu is a recipient of the National Thousand Talents Plan award. His research interest is surface and interface science, particularly phenomena at gas–solid and liquid–solid interfaces, synchrotron/free-electron laser–based in situ characterization techniques, and advanced instrumentation development. He has published over 200 papers in peer-reviewed academic journals.

ZHANG Xiaolin

ZHANG Xiaolin received his B.S. from the Department of Power Systems of Northeast China Institute of Electric Power Engineering in 1985, and his M.S. in engineering (1989) and D. Eng. (1995) from Yokohama National University in Japan, where he worked as a research associate from 1989 to 1992. After completing his D. Eng., he worked as an assistant professor at the Laboratory of Biomedical Engineering of Tokyo Medical and Dental University. In 2003 he was made an associate professor at the Tokyo Institute of Technology (TIT), and in 2012, he became a full professor at the TIT Precision and Intelligence Laboratory. He has been a Thousand Talents Plan professor at SIMIT since 2013, and director of the Bionic Vision System Laboratory there since 2015, where he is currently researching brain-inspired intelligent vision systems. He has authored or coauthored over 80 technical journal papers and conference proceedings, and has over 30 granted and pending patents in countries including China, Japan, and the United States.

LIU Zhengxin

LIU Zhengxin obtained his B.S. in 1989 from the Department of Optical Physics, Changchun University of Science and Technology (Changchun College of Optics and Fine Mechanics), China, and his M.S. and Ph.D. from the Department of Electronic and Information Engineering at Toyohashi University of Technology in Japan, in 1997 and 2000, respectively. He worked in Japan at System Engineering Consultants Co., Ltd., Tateyama Kagaku Industry Co., Ltd., and the National Institute of Advanced Industrial Science and Technology before being appointed as a Thousand Talents Plan professor at SIMIT in 2010. He has also been a professor at ShanghaiTech University since 2012.

Liu is an expert on solar cell materials and devices, renewable energy, and measurement technology for solar cells. His current research interests include high-efficiency crystalline silicon solar cells, the physics of heterojunction structures, and solar cell standardization. He has translated a book entitled *Solar Cells* from Japanese to Chinese, authored over 100 journal papers, and holds four Japanese patents and five Chinese patents.

Liu has worked as a lead investigator for several Japanese and Chinese national projects and has served as a committee member for several international conferences. He is vice-chairman of Technical Committee 82 of the International Electrotechnical Commission (IEC/TC82).

TAO Hu

TAO Hu received his Ph.D. in mechanical engineering from Boston University in 2010, winning the Best Dissertation Award. His research interests have primarily focused on terahertz metamaterials using micro-electromechanical systems technology. After graduating, Tao joined the Department of Biomedical Engineering at Tufts University as a postdoctoral associate, and later advanced to research assistant professor.

In 2014, he was appointed as a Thousand Young Talents Plan professor at SIMIT where his research focuses on green nanotechnology and micro/nanotechnology-enhanced novel electronic and photonic devices for biomedical applications. Tao has published over 50 papers in peer-reviewed scientific journals including *Science* (cover), *Nature, Nature Photonics* (cover), *Nature Nanotechnology* (cover), *Nature Communications, Proceedings of the National Academy of Sciences of the United States of America, Advanced Materials* (cover), *Small* (cover), and *Physical Review Letters*.

LU Yumin

LU Yumin received his B.S. in materials science and engineering from Fudan University, Shanghai, China in 1996, his M.S. in materials science and engineering from The Ohio State University, Columbus, Ohio in 2001, and his Ph.D. in electrical engineering from the University of Michigan at Ann Arbor in 2005.

Lu has been a Thousand Young Talents Plan professor at SIMIT since 2014. He also serves as vice president of radio frequency (RF) technology for Shanghai Industrial μTechnology Research Institute (SITRI), focusing on the development of high-performance RF front-end components and millimeter-wave integrated circuits. Before joining SIMIT/SITRI, he worked in the United States for NXP Semiconductors, Autoliv, and M/A-COM Technology Solutions.

International Scientists at SIMIT

John Clarke

Superconducting quantum interference devices (SQUIDs) provide extremely sensitive measurements of magnetic flux or of any other physical quantity that can be converted to magnetic flux. One of the world's renowned SQUID pioneers, John Clarke from the University of California, Berkeley, agreed to ongoing cooperation with SIMIT on SQUIDs and their applications, with a current focus on ultralow field magnetic resonance imaging (ULFMRI). A major advantage of ULFMRI is that the longitudinal relaxation time generally depends on the type of tissue being imaged, allowing, for example, imaging of tumors without contrast agents. Another evolving application is the use of a multichannel SQUID system that combines ULFMRI with magnetoencephalography to enable simultaneous structural and functional imaging of the human brain. Clarke shares his knowledge and experience of SQUID design and the ULFMRI system with SIMIT researchers and students by visiting Shanghai regularly. In addition, two young scientists from SIMIT were trained in Clarke's group at Berkeley to perform ex vivo and in vivo brain imaging, with much success.

Makato Sato

Makoto Sato, Distinguished Professor of SIMIT, is professor emeritus of the Tokyo Institute of Technology (TIT), former director of the Precision and Intelligence Laboratory of TIT, and former chairman of the Virtual Reality Society of Japan. He has dedicated his career to researching pattern recognition, image processing, human–machine interfaces, virtual reality, and haptic interaction. In 1989, Sato was the first to develop the haptic display system known as SPIDAR (SPace Interface Device for Artificial Reality), which is based on force feedback. The system connects hand-operated devices to a haptic interaction system that gives a machine realistic tactile sensation. The system can provide the operator with an authentic touch response for true-to-life interactions in different virtual environments. Through cooperation with Sato, SIMIT has realized the fusion of bionic vision and touch in the fields of robot control and mixed reality, and has expanded the multimodal biomimetic senses of brain-inspired intelligence, also providing technical support for brain-inspired intelligence research at SIMIT.

ZHANG Yi

ZHANG Yi, a German professor who obtained his Ph.D. from Justus Liebig University in Germany, is engaged in researching SQUIDs and their applications at the Forschungszentrum Jülich GmbH in Germany, and is a leading international scientist in the SQUID field.

Since 2006, Zhang has helped SIMIT to develop its superconductivity program through cooperative research and personnel training, thus aiding it in establishing a leading, international position in the SQUID field. A fruitful cooperative relationship has led to a breakthrough in SQUID technology that has advanced the field internationally. Zhang directed the SIMIT team in its development of the first multichannel, low-temperature superconducting magnetometer in China, and its successful installation in the Shanghai Sixth People's Hospital and Xuhui District Central Hospital. Zhang has given seminars on scientific research cooperation, setting up joint laboratories, and establishing an international cooperation institute. In 2014, he won the Shanghai International Science and Technology Cooperation Award.

PHOTO: (JOHN CLARKE) PEG SKORPINSKI

SIMIT Talent Cultivation and Recruitment Program

Eligibility criteria

(candidates should fulfill any one of the following)

- Meet the requirements for Innovative Talents within the Thousand Talents Plan, or of Young Professionals within the Thousand Talents Plan;
- Meet the requirements of the Shanghai Thousand Talents Plan;
- Meet the requirements of the CAS Hundred Talents Program;
- Should be outstanding young and middle-aged scholars, doctors, and postdoctoral students from well-known foreign universities and research institutes.

Areas of research interest

- Intelligent sensing microsystems
- Superconducting quantum devices and circuits
- Advanced silicon-based materials and applications
- Special broadband wireless communication technologies and equipment
- Micro- and nanotechnologies for sensors and transducers
- Phase-change random access memory and applications
- Terahertz solid-state technologies
- Brain-inspired chips and bionic vision

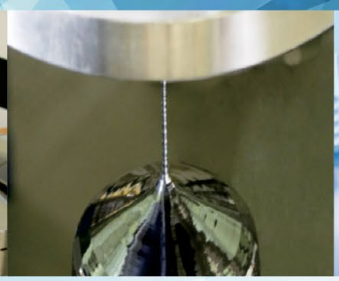
Support and benefits

- The institute will fully support and assist candidates who meet the requirements of national talent projects such as the Thousand Talents Plan, the Shanghai Thousand Talents Plan, and the Hundred Talents Program of CAS.
- In accordance with the official provisions stipulated by SIMIT, Shanghai Municipality, and the Chinese government, the project candidates will receive a research and living allowance.
- The institute will offer a competitive salary and benefits, including social insurance, a housing fund, housing subsidies, and assistance to apply for permanent residence in Shanghai for project candidates, their spouses, and their children, in accordance with official policy.
- Temporary residences, necessary offices, laboratory rooms, and financial support will be provided for selected candidates.

Contact Information:

Please email your resume to lihua@mail.sim.ac.cn, or call +86-21-62511070-5701 if you are interested in applying.
For more information about SIMIT, please go to www.sim.cas.cn/rczpw




National Silicon Industry Group (NSIG)—the only officially sanctioned materials group in China—was formed in December 2015 and is headquartered in Shanghai. NSIG is the sole shareholder of Okmetic (producing wafers 200 mm in diameter and smaller), and the largest shareholder of Simgui, Soitec (manufacturer of silicon-on-insulator (SOI) substrates), and ZingSemi (300 mm diameter wafer producer). The shareholders of NSIG are Sino IC Capital, Shanghai Guosheng Group, SummitView Capital, SIMIC (the investment platform of SIMIT), and Shanghia Jiading Industrial Zone. The goal of NSIG is to enhance its leading position in the silicon materials industry and to be an internationally competitive platform.

NSIG

National Silicon Industry Group

300 mm wafers

- **ZingSemi**

The first 300 mm silicon wafer producer in China, focused on 300 mm wafer production technology bottlenecks.

SOI wafers

- **Soitec**

The world's largest SOI wafer supplier.

- **Simgui**

Focused on the SOI market; expanding through construction of state-of-the-art SOI production line.

Customized wafers

- **Okmetic**

The world's leading MEMS silicon/SOI supplier, focused on customized, value-added wafer products, including 200 mm wafer production.

ZINGSEMI

www.zingsemi.com

ZingSemi is a 300 mm semiconductor wafer manufacturer located in Shanghai and a part of the NSIG family. It produces 300 mm epitaxial and polished wafers, aiming to generate over 600,000 wafers per month.

SIMGUI

www.simgui.com.cn

Simgui is the largest epitaxial wafer provider in China, producing 150 mm-200 mm epitaxial and SOI wafers. It is the only company in the world with a full range of SOI technology including SIMOX, Bongding, SIMBOND™, and Soitec Smart Cut™.

OKMETIC

www.okmetic.com

Okmetic is the world's leading supplier of sensor wafers for microelectromechanical systems (MEMS) manufacturing, producing 150 mm-200 mm bonded SOI, cavity SOI, and double-sided polished wafers as well as specialized silicon wafers.

中国科学院上海微系统与信息技术研究所

SHANGHAI INSTITUTE OF MICROSYSTEM AND INFORMATION TECHNOLOGY, CHINESE ACADEMY OF SCIENCES