Tomorrow's Earth & "The Commons" *pp. 1217, 1236, & 1242*

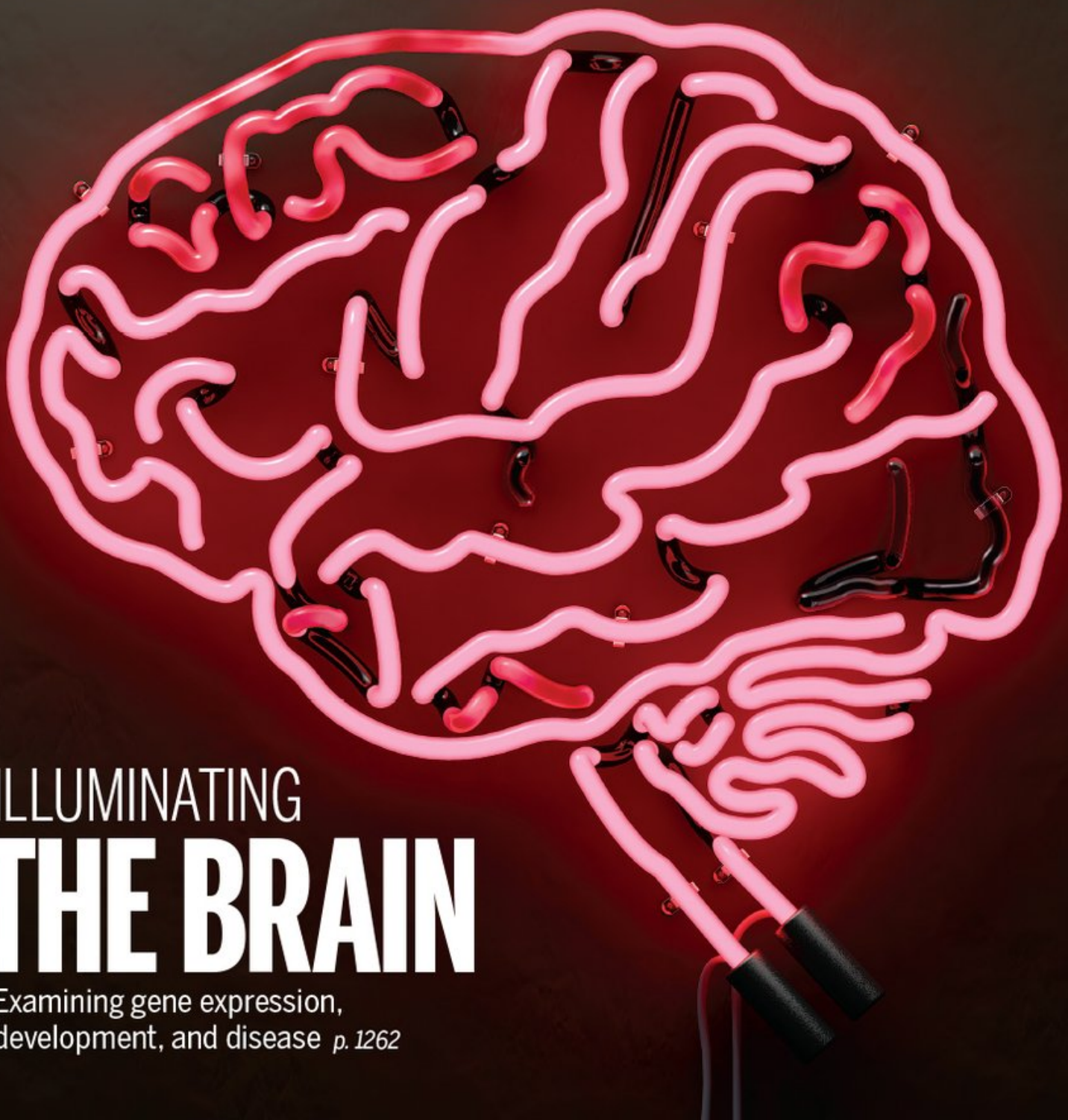
Need for human genome editing consensus *p. 1215*

Dreaming up the next great space telescope *p. 1230*

Science

$15
14 DECEMBER 2018
sciencemag.org

AAAS

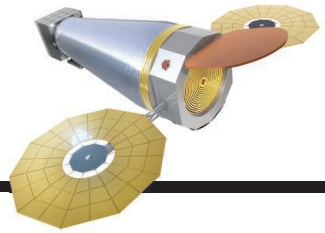


ILLUMINATING THE BRAIN

Examining gene expression, development, and disease *p. 1262*

CONTENTS

14 DECEMBER 2018 • VOLUME 362 • ISSUE 6420



ILLUMINATING THE BRAIN

ON THE COVER

The human brain is the product of myriad molecular and genetic interactions. Here, a neon brain illustration represents individual genetic variability, some of which may lead to disease (denoted by dim or dark segments), as investigated by the PsychENCODE Consortium. This issue sheds light on neurogenetic and epigenetic variation in developing and adult neurotypical brains, as well as in schizophrenia, autism spectrum disorder, and bipolar disorder. See page 1262.
*Illustration: V. Altounian/*Science

SEE ALSO ▸ *SCI. TRANSL. MED.* 10.1126/scitranslmed.aat8178 ▸ *SCI. TRANSL. MED.* 10.1126/scitranslmed.aat6912 ▸ *SCI. ADV.* 10.1126/sciadv.aav8550 ▸ NEWS STORY P. 1227

NEWS

IN BRIEF

IN DEPTH

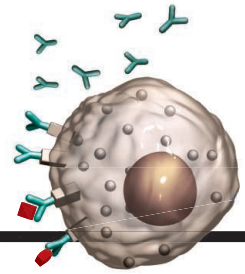
FEATURE

INSIGHTS

POLICY FORUM

PERSPECTIVES

CONTENTS



1247 & 1306
Allergen-specific antibodies cloned

14 DECEMBER 2018 • VOLUME 362 • ISSUE 6420

RESEARCH

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.

EDITORIAL

Wake-up call from Hong Kong

Victor J. Dzau
is the president of the U.S. National Academy of Medicine.
vdzau@nas.edu

Marcia McNutt
is president of the U.S. National Academy of Sciences.
mmcnutt@nas.edu

Chunli Bai
is president of the Chinese Academy of Sciences.
clbai@cas.cn

The Second International Summit on Human Genome Editing, held in Hong Kong last month, was rocked by the revelation from a researcher from Shenzhen that twins were born whose healthy embryonic genomes had been edited to confer resistance to HIV. Despite widespread condemnation by the summit organizing committee, world scientific academies, and prominent scientific leaders that such research was "deeply disturbing" and "irresponsible," and the launch of an investigation in China into the researcher's actions, it is apparent that the ability to use CRISPR-Cas9 to edit the human genome has outpaced nascent efforts by the scientific and medical communities to confront the complex ethical and governance issues that they raise. The current guidelines and principles on human germline genome editing are based on sound scientific and ethical principles. However, this case highlights the urgent need to accelerate efforts to reach international agreement upon more specific criteria and standards that have to be met before human germline editing would be deemed permissible.

"We need...broad agreement on...criteria for human germline genome editing research..."

Together, we call upon international academies to quickly convene international experts and stakeholders to produce an expedited report that will inform the development of these criteria and standards to which all genome editing in human embryos for reproductive purposes must conform, and to engage scientific bodies around the world in this effort. The United States National Academies are willing to lead in this endeavor. Academies are well-positioned to convene needed international expertise and to help foster broad scientific consensus on the responsible pursuit of human genome editing research and clinical applications. We strongly believe that international consensus on such standards is important to avoid the potential for researchers to rationalize the justification or seek out convenient locales for conducting dangerous and unethical experimentation. The establishment of international scientific standards is not intended to substitute for national regulation but could inform such regulation.

To maintain the public's trust that someday genome editing will be able to treat or prevent disease, the research community needs to take steps now to demonstrate that this new tool can be applied with competence, integrity, and benevolence. Unfortunately, it appears that the case presented in Hong Kong might have failed on all counts, risking human lives as well as rash or hasty political reaction.

Establishing standards alone will not suffice. We also need an international mechanism that would enable scientists to raise concerns about cases of research that are not conforming to the accepted principles or standards. The Second International Summit organizers have called for establishing an ongoing international forum on human genome editing that could provide such a mechanism, along with other important functions such as helping to speed the development of regulatory science, providing a clearinghouse for information about governance options, contributing to the long-term development of common regulatory standards, and enhancing coordination of research and clinical applications through an international registry of planned and ongoing experiments.

More than 40 years ago, scientists organized the renowned Asilomar Conference on Recombinant DNA amid concerns about safety and efficacy of what was then a revolutionary new biomedical technology. They publicly discussed and debated the issues, and ultimately, they were able to reach consensus on a set of research guidelines that eventually formed the basis for official government policy. The model of Asilomar still offers important lessons. We need to build upon the work done at recent international summits and the guidance provided by numerous organizations to achieve broad agreement on specific standards and criteria for human germline genome editing research and clinical applications—agreement that should include not only the scientific and clinical communities, but also society as a whole.

–Victor J. Dzau, Marcia McNutt, Chunli Bai

10.1126/science.aaw3127

ILLUSTRATION: DAVIDE BONAZZI/SALZMANART

EDITORIAL

Choices in the climate commons

Climate change is a tragedy of the commons of existential importance. At the annual United Nations climate summit that concludes this week, parties will affirm the necessity to avoid dangerous climate change. But between now and next year's summit, these same countries will in many ways act so as to hasten the outcome that they say must be avoided. This disjunction between what countries say and what they do has been repeated every year since the first summit in 1995. It is a pattern of behavior that seems irrational, but that can be explained. American ecologist Garrett Hardin's classic article, "The Tragedy of the Commons," published in *Science* 50 years ago this week, vividly describes the dilemma that causes this behavior (see page 1236).

Herders, wrote Hardin, are motivated by private gain, so have incentives to add animals to their shared pasture "without limit." As in the climate change game, all herders also want their pasture to be saved, but none is willing to bear the personal sacrifice needed to prevent its destruction. Saving the pasture requires collective action. Hardin's proposed corrective is "mutual coercion." Writing in 1651, British philosopher Thomas Hobbes similarly concluded that a sovereign is needed to tie people "by fear of punishment to the performance of their covenants."

However, a critical difference between climate change and Hardin's parable is that the players in the climate game are nation states. Although individuals can be subjected to coercion by a higher authority, human organization has not evolved to give any institution sovereignty over the nation state. Solutions to global collective action problems must involve covenants (treaties) among states that are self-enforcing.

To stabilize the climate, a treaty must get all states to (i) participate in and (ii) comply with an agreement that (iii) drives emissions to zero. The Paris Agreement, adopted at the 2015 summit, secures the first requirement, and possibly the second, but only because it is a voluntary agreement and will fall short of meeting the third requirement. The Montreal Protocol, negotiated in 1987 to protect the stratospheric ozone layer, meets all three requirements, thanks partially to a ban on trade in chlorofluorocarbons between parties to the protocol and nonparties. Because of the ban, once the vast majority of countries joined the agreement, all others wanted to join. William Nordhaus, a recipient of this year's Nobel Memorial Prize in Economic Sciences, has recently analyzed a similar cure for climate change in which members of a "climate club" who agree to curb emissions impose a tariff on imports from nonmembers to encourage their participation. Unfortunately, his analysis shows that as the carbon tax rises to the level needed to stabilize the climate, participation in the club collapses.

"...we will have to choose between risks to address the scale of this problem..."

Breaking up the problem may provide more leverage for enforcement. The Kigali Amendment to the Montreal Protocol, adopted in December 2016, phases down hydrofluorocarbons, a group of greenhouse gases, and this will be effective in addressing this particular cause of climate change for the same reasons that the Montreal Protocol has been effective in protecting the ozone layer. Other climate agreements, adopted in parallel with the Paris Agreement, should be negotiated for individual sectors, such as aluminum and steel and international aviation and shipping, all linked to trade.

However, the time has come to contemplate other, more radical solutions. The October 2018 Intergovernmental Panel on Climate Change special report concluded that limiting temperature change to 1.5°C cannot be achieved by simply curbing emissions, but requires removing CO_2 from the atmosphere. The only true "backstop" for limiting climate change is removal of CO_2 by industrial processes, which converts the problem from one of changing behavior into one of joint financing of a large-scale project. Another option, solar geoengineering, acts directly on global mean temperature, but is considered risky. Of course, not using it could also be risky. In the end, regardless of pathways forward, we will have to choose between risks to address the scale of this problem and achieve, rather than merely aspire to, global collective action on climate change.

–Scott Barrett

Scott Barrett
is the Lenfest-Earth Institute Professor of Natural Resource Economics at Columbia University, New York, NY, USA.
sb3116@columbia.edu

TOMORROW'S EARTH
Read more articles online at scim.ag/TomorrowsEarth

10.1126/science.aaw2116

NEWS

> “We are not going to punish them for the actions of their leaders.”
>
> **Associate Provost Richard Lester** of the Massachusetts Institute of Technology, in *The Boston Globe*, on accepting research money and partners from Saudi Arabia.

IN BRIEF

Edited by **Jeffrey Brainard**

PUBLISHING

China supports open-access plan

Plan S, the push by European science funders for immediate open access (OA) to research publications, got a boost last week when China's largest government research funder and two national science libraries said they back its goals of making journal articles freely available after 1 January 2020. Chinese agencies now allow papers developed with their funding to reside behind a journal paywall for up to 12 months, after which they must be made OA. But in position papers released last week, the three agencies state they intend to require immediate OA as soon as possible. "The exact timing of implementing the new policy is now being discussed, but surely it will not be long," says Zhang Xiaolin of the National Science Library of the Chinese Academy of Sciences in Beijing, which issued one of the statements; the others came from the National Natural Science Foundation of China and the National Science and Technology Library. Chinese funders won't necessarily endorse Plan S formally, but the statements do call for comparable measures, including capping OA article-processing charges. In 2016, China produced more scientific papers than any other country, the U.S. National Science Foundation reported.

Explosion kills Indian researcher

LAB SAFETY | A researcher was killed and three others seriously injured last week when a gas cylinder exploded at one of India's premier research facilities for unknown reasons. Manoj Kumar, 32, an employee of a startup named Super-Wave Technology, died at the Laboratory for Hypersonic and Shock Wave Research of the Indian Institute of Science in Bengaluru. The laboratory houses four tubes that use liquid hydrogen, oxygen, nitrogen, and helium to generate shock waves. The institute's students and researchers are not mandated to take safety training, says its director, Anurag Kumar. "It is left to individual professors to instruct the staff on safety as they are the most knowledgeable about the equipment they handle."

Panel confirms bullying claims

WORKPLACE | Empathy expert Tania Singer has resigned as director of the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany, effective 1 January 2019, after a commission confirmed allegations of bullying, the Max Planck Society (MPG) announced last week. In August, *Science* reported that researchers at the institute said that Singer had created an "atmosphere of fear" and mistreated female employees who became pregnant (*Science*, 17 August, p. 630). In a letter to her former lab members dated 2 December, Singer apologized "for the mistakes I made as a young director of a big Max Planck Department." She will work as a neuroscientist in Berlin with a small group under the guidance of the vice president of MPG.

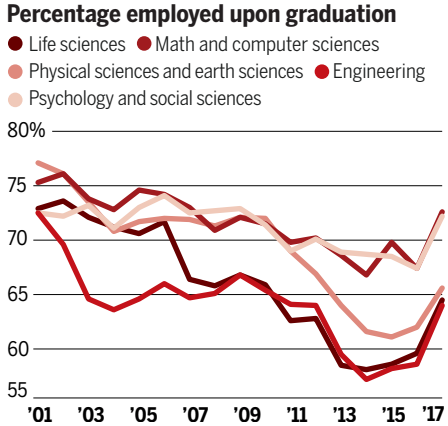
Canada denies researchers' visas

INTERNATIONAL AFFAIRS | Dozens of African researchers were denied visas for an artificial intelligence (AI) meeting in Montreal, Canada, last week, even as the government takes steps to advance the country's standing in AI and the field aims for greater inclusivity. Black in AI, a daylong workshop for scientists of African descent, held during a leading AI conference called Neural Information Processing Systems, invited more than 200 African scientists to participate. About half of the requested visas were denied or delayed until too late, in many cases because researchers were suspected of not planning to return to their home countries. Timnit Gebru, a researcher at Google in Mountain View, California, and co-founder of Black in AI, said African researchers' difficulty obtaining visas for Canada is "a long-standing problem" that demands attention. An immigration official said in an interview that people from all countries were evaluated using the same criteria.

More jobs for Ph.D. recipients

CAREERS | Employment prospects for graduating U.S. doctoral students may be looking up. In the 2016–17 academic year, the percentage who reported landing jobs, including postdoctoral appointments, rose after declining for more than a decade, according to the National Science Foundation's Survey of Earned Doctorates. The report didn't describe how many of these jobs were science related.

Percentage employed upon graduation



CREDITS: (GRAPHIC) K. LANGIN/*SCIENCE*; (DATA) SURVEY OF EARNED DOCTORATES/U.S. NATIONAL SCIENCE FOUNDATION

Published by AAAS

NCI cuts operations budget 5%

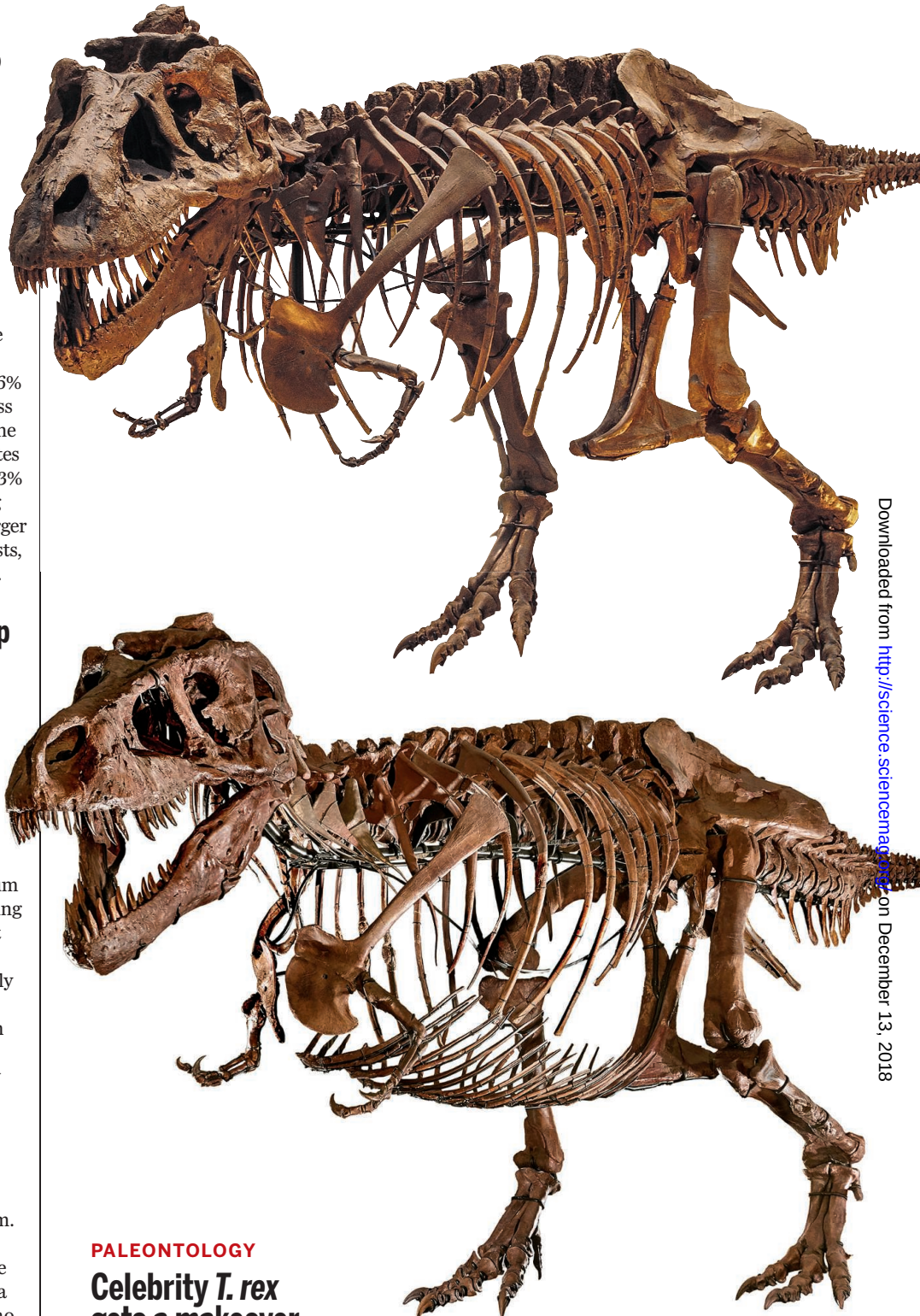
FUNDING | Despite a growing budget, the U.S. National Cancer Institute (NCI) in Bethesda, Maryland, is trimming internal operating budgets by 5% to free up funds for a growing number of grants generated by a rise in applications. It will also shave continuing grants to extramural researchers by 3% during the 2019 fiscal year, except for cancer centers, "moonshot" grants for cancer cures, and training centers. The cuts reflect the institute's struggle to maintain success rates—the odds that a proposal will be funded—in the face of a 46% rise in applications since 2013. NCI's success rate of 12% in 2017 was much lower than the 19% rate across the entire National Institutes of Health. Although NCI's budget will rise 3% in 2019, to $5.74 billion, its funds are being stretched thin by rising federal salaries, larger grants and training stipends, and other costs, NCI Director Ned Sharpless said last week.

Carbon dioxide emissions tick up

CLIMATE SCIENCE | As world leaders gathered last week for the annual United Nations talks on climate change in Katowice, Poland, they were greeted with grim news: For the second year in a row, carbon dioxide emissions from fossil fuels will hit a new high, growing 2.7% this year to a record 37.1 gigatons, according to an estimate by the Global Carbon Project, an international consortium of scientists. The increases follow a flatlining from 2014 to 2016 that had bred hope that the world had begun to limit emissions of greenhouse gases. Emissions rose especially in India, up 6.3% thanks to expanded use of coal-fired power; China, up 4.7%, driven by burning of natural gas; and the United States, up 2.5%, attributed to an unusually cold winter and hot summer.

U.K. science minister named

LEADERSHIP | Brexit is causing more tumult for science in the United Kingdom. Chris Skidmore was appointed minister for science and universities last week, the third person in the position in less than a year. Skidmore replaces Sam Gyimah, who resigned in late November to protest the plan for leaving the European Union that Prime Minister Theresa May negotiated last month. Like other science ministers, Skidmore brings a political background as a member of Parliament rather than research experience to the post. His tenure may be even shorter, as the future of May's government hangs in the balance over Brexit politics.



PHOTOS: JOHN WEINSTEIN/FIELD MUSEUM

PALEONTOLOGY

Celebrity *T. rex* gets a makeover

Sue the *Tyrannosaurus rex*, the most complete and largest *T. rex* fossil known—and the only one with a Twitter account (@SUEtheTrex)—is going back on display this month with a new look. In February, staff at the Field Museum in Chicago, Illinois, disassembled Sue to make room for another exhibit. She returns to display on 21 December, and paleontologists have given the fossil a scientific update, adding riblike bones called gastralia that scientists think helped the dinosaur breathe. They also repositioned the wishbone and the arms, based on research done since Sue first went on display in 2000 (top). Her more rotund look (bottom) fits with new estimates of her weight—now 9 or 10 tons, up from 5 to 7 tons—based on 3D scans of her bones.

TRUMP TRACKER

A Trump administration plan would reduce protected habitat for the sage grouse.

White House targets environment, climate measures

U.S. President Donald Trump's administration continues to draw controversy for its environmental proposals and policies involving science, including these this month.

Sage grouse

The Department of the Interior on 6 December released a new plan for protecting the sage grouse, a bird threatened by development in six western states. It weakens protections for nearly 85%, or 3.6 million hectares, of grouse habitat.

Ocean science

The White House deleted a chapter on climate change from a new 10-year plan for federal investments in ocean science and technology. Plans issued in 2007 and 2010 had climate chapters.

Coal power plants

The Environmental Protection Agency on 6 December said new coal-fired power plants can emit more carbon dioxide than allowed under a plan proposed by former President Barack Obama. But analysts say the move won't reverse the decline of coal as a U.S. electricity source.

Wetlands

The administration on 11 December proposed to slash the number of wetlands protected by federal law. Ephemeral streams that run only during wet periods would lose protection.

Climate change

U.S. diplomats at global climate talks in Poland this week joined Russia and Saudi Arabia in opposing a statement acknowledging the growing risks of global warming. Earlier this month, the United States refused to sign a similar statement at the G20 economic summit in Argentina.

Collider idles for upgrades

PARTICLE PHYSICS | The world's largest atom smasher turned off last week for a 2-year pit stop. The Large Hadron Collider (LHC) at the European particle physics laboratory, CERN, near Geneva, Switzerland, will remain offline until 2021 while researchers upgrade accelerators and detectors to handle more collisions at slightly higher energies. The work is the first step in a plan to boost the LHC's collision rate by as much as five times by 2026. In 2012, the LHC blasted out the long-predicted Higgs boson, central to explaining how all other fundamental particles get their mass. However, the LHC has yet to discover anything unexpected. The increased data will help physicists search for rarer decays and subtler signs of something new.

'Chaperoned' authors prosper

PUBLISHING | Researchers who first publish in prestigious, interdisciplinary journals as junior team members increasingly have an advantage later, when they seek to publish in these journals as the senior authors, a study of "chaperoning" has found. In *Nature*, for example, the share of papers by senior authors (defined as last authors) who earlier published there in a different author position grew from 16% to 22% since 1990, whereas senior authors without this experience dropped from 39% to 31%. The benefit of being chaperoned was strongest in interdisciplinary journals, and then, in decreasing order, in journals for biology, medicine, chemistry, physics, and math, the authors report in the 10 December issue of the *Proceedings of the National Academy of Sciences*; they examined papers published in 386 scientific journals from 1960 to 2012. Chaperoned authors may be increasing because of better mentoring and opportunities to build reputation and connections, the study suggests.

Voyager 2 heads for the stars

SPACE EXPLORATION | NASA's Voyager 2 probe has become only the second humanmade object to enter interstellar space. The craft, launched in 1977, left the heliosphere—the protective bubble of particles and magnetic fields created by the sun—on 5 November, mission scientists announced this week after examining the probe's instruments. The craft's sister, Voyager 1, crossed this boundary in 2012; both continue to send back useful data.

Little Foot's big debut

PALEOANTHROPOLOGY | The world's most complete skeleton of an early hominin known as Little Foot got her long-awaited close-up last week, when researchers published the first detailed analyses of her fossil remains. Little Foot was discovered in a South African cave in the late 1990s and excavated gradually for more than a decade. She lived some 3.67 million years ago and was primarily bipedal, researchers reported in four un–peer-reviewed papers on the bioRxiv preprint server. What's more, they claim, her features don't closely match those of *Australopithecus africanus*, a species from around that time found in the same cave. Instead, the researchers argue, Little Foot (so named for the small size of her foot bones) was a member of *A. prometheus*, a species proposed in 1948 but never fully accepted. Other researchers remain skeptical and say more research is needed to work out how much variation is expected within these ancient species.

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

PHOTO: DANITA DELIMONT/GETTY IMAGES

IN DEPTH

ENERGY AND ECOLOGY

Ireland slashes peat power to lower emissions

Harvested from drained and denuded bogs, peat is more polluting than coal

By **Emily Toner**

On a cold, gray morning in November, the Corneveagh Bog in central Ireland is a scene of industrial harvest. Like other Irish bogs, it has been drained and stripped of its moss and heather to reveal the rich, black soil beneath: peat. The peat is scored with tread marks left by the machines that shaved off a crumbly layer and turned it over to dry. A long mound of peat, stripped and dried earlier in the season, is covered in plastic, waiting to be piled into rail cars and taken to a nearby power plant. There, the carbon-rich soil will be burned to generate electricity.

But not for much longer, says Barry O'Loughlin, an ecologist employed by Bord na Móna, a state-owned peat harvesting and energy company based in Newbridge that owns Corneveagh Bog. Bord na Móna, which means "Peat Board," will soon retire dozens of bogs like Corneveagh from energy production. Its team of four ecologists will rehabilitate many of them by blocking drains, soaking the ground, and reestablishing plant life, O'Loughlin says as his boots crunch through the frosty soil. "We bring life back into the bog again."

In Ireland, peat has been used for centuries to warm homes and fire whiskey distilleries. For a country with little coal, oil, and gas, peat—deep layers of partially decayed moss and other plant matter—is also a ready fuel for power plants. Peat power peaked in the 1960s, providing 40% of Ireland's electricity. But peat is particularly polluting. Burning it for electricity emits more carbon dioxide than coal, and nearly twice as much as natural gas. In 2016, peat generated nearly 8% of Ireland's electricity, but was responsible for 20% of that sector's carbon emissions. "The ceasing of burning peat is a no-brainer," says Tony Lowes, a founder of Friends of the Irish Environment in Eyeries.

> ***"There's a lot of bare peat around. There's a lot of hemorrhaging carbon."***
>
> **Catherine O'Connell**,
> Irish Peatland Conservation Council

That is now beginning to happen. By the end of 2019, the Irish government will eliminate all of the roughly €100 million in annual industry subsidies it now pays for peat-generated electricity. Bord na Móna, which supplies peat to the three remaining power stations burning it for electricity, announced in October that it would cut its peat supply for electricity by a third by 2020 and end it completely by 2027. Ireland will need to find alternative, lower carbon sources of electricity. And approximately 60 bogs no longer needed for fuel will be converted back to wetlands or put to commercial uses such as land for wind farms.

Behind the phaseout is Ireland's promise to the European Union to reduce greenhouse gas emissions by 20% in 2020, compared with 2005 levels. "The country's decarbonization agenda is driving Bord na Móna's step down from peat," says Joe Lane, the company's chief operating officer. Even so, Ireland will miss its goal. Despite rapid growth in wind power and increasingly energy efficient homes and vehicles, it will struggle to reduce emissions by even 1%, says Phillip O'Brien, scientific officer for the Irish Environmental Protection Agency in Dublin.

Like any energy transition, this one comes with a human cost. Up to 430 jobs will be lost, Lane says. "Most of the people who will lose their jobs are people who have worked for Bord na Móna for a long time—people whose fathers, grandfathers, and villages are all tied to the company."

And replacing peat with biomass, as the power companies plan to do, is not a panacea. A decade ago, Bord na Móna began to cofuel a peat-burning station with mixtures of biomass including a grass called miscanthus, olive pits, almond shells, palm kernel shells, and beet pulp, much of it imported from all over the world. Because biomass

PHOTO: KLAUS-WERNER FRIEDRICH/ALAMY STOCK PHOTO

Corrected 13 December 2018. See full text.

Published by AAAS

Industrial peat extraction has stripped dozens of Irish bogs of their heather and moss.

takes up carbon from the atmosphere as it grows, the European Union counts it as a carbon-neutral, renewable resource—even though transportation, processing, and land-use costs make it less so. "The unregulated or unfettered use of biomass would lead to serious problems," says Robert Matthews, a scientist at Forest Research in Surrey, U.K. In 2021, European legislation will tighten biomass standards, reducing the advantages of burning it from a carbon accounting standpoint.

Rehabilitating the harvested peatlands, however, is a clear plus for climate. When bogs are drained to harvest peat, or for any other use, such as agriculture, grazing, or forestry, exposure to oxygen jump-starts the decomposition of the stored organic matter, releasing carbon into the atmosphere. A 2013 study of Irish peatland carbon emissions, published in *Irish Geography*, found that each hectare of industrially drained and stripped peatland emits 2.1 tons of carbon per year—the equivalent of driving a car 30,000 kilometers. And that's before the harvested peat is burned.

Those emissions cease as soon as drains are blocked and the water table rises to resaturate the peat, cutting off oxygen. As a result, say ecologists, conserving peatlands has a triple benefit: reducing emissions from both power plants and exposed fields and, with restored plant life, sequestering more carbon in future peat deposits. "Peatlands are our rainforest, our carbon sink," Lowes says.

Moreover, healthy peatlands improve water quality and provide needed habitat for threatened species such as curlews and marsh fritillary butterflies. "Our goal is to make things as wet as we can, where we can," says Catherine Farrell, an ecologist at Bord na Móna. She says that of the 80,000 hectares of peatland under company management, 18,000 hectares have been rehabilitated.

But in a country where peat smoke rises from chimneys every day, that's just a start. People cut peat to burn in their houses from another 600,000 hectares of peatlands, and there are few plans for rehabilitating these degraded bogs. Catherine O'Connell, director of the Irish Peatland Conservation Council in Lullymore, would like to see more action to heal the bogs. "There's a lot of bare peat around," she says. "There's a lot of hemorrhaging carbon." ■

Emily Toner is a geographer and journalist on a Fulbright–National Geographic fellowship in Tullamore, Ireland.

U.S. RESEARCH POLICY

Trump officials move to limit human fetal tissue research

NIH ordered staff scientists not to procure new tissue

By **Meredith Wadman** *and* **Jocelyn Kaiser**

President Donald Trump's administration is taking steps to limit the use of human fetal tissue from elective abortions in biomedical research. Last week, administration officials told researchers at one California university that their contract work involving fetal tissue would not receive the usual 1-year extension. And *Science* has learned that in September, officials quietly ordered scientists employed by the National Institutes of Health (NIH) in Bethesda, Maryland, to stop acquiring new fetal tissue for experiments.

Both moves come as the administration is reviewing all federally funded research with fetal tissue, which is used to study several diseases. The actions have prompted fears that NIH-funded university scientists who work with fetal tissue could face a broader clampdown.

The administration's actions are already affecting research, scientists say. The NIH order, which was not made public, disrupted one study of the virus that causes AIDS. "We were all poised to go and then the bombshell was dropped," says HIV researcher Warner Greene, director of the Gladstone Center for HIV Cure Research in San Francisco, California. "The decision completely knocked our collaboration off the rails. We were devastated."

Research using human fetal tissue from elective abortions is legal in the United States, but antiabortion groups and some lawmakers in Congress fiercely oppose federal funding for such work. In September, the Trump administration canceled a contract under which the Food and Drug Administration (FDA) acquired fetal tissue for testing drugs. Last week, the Department of Health and Human Services (HHS), which oversees NIH, told researchers at the University of California (UC), San Francisco, that it would be extending a contract for work involving fetal tissue for just 90 days instead of 1 year. HHS denied reports it was planning to cancel the contract.

Last week, an NIH spokesperson also confirmed that earlier this year the agency told staff scientists in NIH's intramural program "to pause procurements of fetal tissue" pending the outcome of the HHS review. NIH officials say the pause affects two laboratories, one operated by the National Eye Institute in Bethesda and one at the Rocky Mountain Laboratories (RML) in Hamilton, Montana, a part of the National Institute of Allergy and Infectious Diseases (NIAID). RML researchers use fetal tissue to create humanized mice, which have immune systems that behave like a human's. They obtain the tissue from Advanced Bioscience Resources (ABR), a nonprofit based in Alameda, California.

Researchers at the Rocky Mountain Laboratories in Hamilton, Montana, halted an ongoing HIV research project.

PHOTO: NATIONAL INSTITUTE OF ALLERGY AND INFECTIOUS DISEASES/NATIONAL INSTITUTES OF HEALTH/CC BY-NC

Corrected 13 December 2018. See full text.

According to emails provided by Greene, RML researcher Kim Hasenkrug had prepared humanized mice for testing an antibody that might prevent HIV from quickly establishing reservoirs in the human body. (Hasenkrug could not be reached for comment.) On 11 September, Greene's group got word from Hasenkrug that the mice were ready, and it sent him the antibody. But on 28 September, Hasenkrug informed Greene by email that HHS had "directed me to discontinue procuring fetal tissue from ABR, the only source for us. ... This effectively stops all of our research to discover a cure for HIV." The ban made it impossible to produce enough mice to complete a statistically convincing study. (NIH says it asked Hasenkrug to inform it if his group required new tissue. But an NIAID email to Hasenkrug that the agency provided to *Science* simply instructs him to hold off on "additional" purchases from ABR "for the next 2-3 months," a pause that "maybe[sic] extended depending on events at levels above all of us.")

The order roughly coincided with the launch of the HHS review and the decision to kill the FDA contract, which was also with ABR. (HHS said it was "not sufficiently assured" that ABR's contract "included the appropriate protections applicable to fetal tissue research.")

Academic scientists with federal grants now worry that they, too, could face restrictions. "Everything I am doing involves humanized mice. It would shut my lab down if we were not able to use fetal tissues," says Jerome Zack, a virologist who studies HIV at UC Los Angeles and has been using humanized mice for 25 years. The mice, he notes, are also used by cancer scientists developing immunotherapy drugs.

One concern is that HHS will cut off supplies from ABR, the largest commercial source of fetal tissue in the United States, which could hurt a swath of U.S. scientists who rely on the firm. "ABR is the most reliable," Zack says.

A House of Representatives panel, meanwhile, planned to hold a 13 December hearing on alternatives to using fetal tissue, and HHS has scheduled an 18 December workshop on the same topic. This week, NIH also said it will spend up to $20 million over 2 years on research into alternatives.

"Why are we having this discussion?" about alternatives, asks biologist Irving Weissman of Stanford University in Palo Alto, California, who is invited to the HHS event. The impetus for seeking alternatives, he says, is "not from scientists working in the field and trying to understand and treat diseases. It's a political force apparently coming from above the NIH level." ■

PUBLIC HEALTH

U.N. HIV/AIDS agency assailed for culture of harassment

Independent panel calls for replacing executive director

By **Jon Cohen**

The Joint United Nations Programme on HIV/AIDS (UNAIDS) in Geneva, Switzerland, is the global command center for the fight against HIV/AIDS. With a $220 million budget and a staff of 700, it stages campaigns to spur treatment and prevention and battles discrimination, especially against "marginalized" groups heavily affected by HIV/AIDS. But the agency itself is rife with harassment, bullying, and abuse of power, a report said last week, citing a "broken organisational culture" and calling for a new leader.

Michel Sidibé hopes to stay at the helm of the Joint United Nations Programme on HIV/AIDS.

PHOTO: T. J. MAPOSHERE

At press time, the agency's Programme Coordinating Board was meeting to decide the fate of its executive director, Michel Sidibé, who initiated the review. Sweden, UNAIDS's second largest donor, said it would freeze support until he leaves. In the long run, the future of UNAIDS itself may be in jeopardy, says Sten Vermund, who heads the Yale School of Public Health. "It was really hard to read that report," he says.

Based on surveys of more than 60% of the staff and interviews or written submissions from 100 of them, the report praises Sidibé for his "outstanding contribution" over the past decade. But it faults him for creating "a patriarchal culture tolerating harassment and abuse of authority." It further criticizes him for "setting a tone of favouritism, preferment, opaqueness, license for wrongdoing, and retaliation against those who speak up." Chris Beyrer, an epidemiologist at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland, and former head of the International AIDS Society, says the findings are "particularly troubling" because gender inequalities are a central driver of the HIV/AIDS epidemic.

Sidibé called for the inquiry in the wake of two allegations of sexual harassment. One involved his deputy executive director; an independent investigation by the World Health Organization (WHO) found no wrongdoing. The second allegation, against a UNAIDS country director, is now being evaluated by WHO. The four-member panel of UNAIDS outsiders who authored the new report was sharply critical of how agency leadership responded to these incidents and described "widespread harassment within the organization." Although only 3.8% of survey respondents reported sexual harassment within the past year, 43.2% said they had experienced abuse of authority.

The panel asserts that in interviews, Sidibé "accepted no responsibility for actions and effects of decisions and practices creating the conditions that led to this review." In a lengthy rebuttal to the panel's report, he challenges some of its details, spells out an "agenda for change," and asks to remain. "I want to lead this change. I want to leave a UNAIDS that is fit for purpose for the next generation."

The panel recommends a different course. "A trustworthy, energetic leader should be appointed who can earn the confidence of the staff and return UNAIDS to its fundamental commitment to non-discrimination, due process, and good governance," the report says.

Some researchers praised the report. "The panel did a really great job in undertaking their task, and that is reflected in the comprehensiveness and specificity of their findings and recommendations," says Quarraisha Abdool Karim, an epidemiologist at the Centre for the AIDS Programme of Research in South Africa in Durban and a special UNAIDS ambassador for adolescents.

The panel contends that UNAIDS's problems stem from its unique position within the U.N. system, which has led to it being "governed in a way that has produced a vacuum of accountability." Vermund suggests this could lead to soul searching about whether UNAIDS should continue to exist as a special agency. "Is UNAIDS serving the purpose for which it was formed and could those functions be better subsumed in the WHO?" he wonders. "Ultimately, you have to ask that question." ■

A baby who may have Ebola is carried to a treatment center in the Democratic Republic of the Congo.

PHOTO: JOHN WESSELS/AFP/GETTY IMAGES

INFECTIOUS DISEASES

Worries about Ebola outbreak grow, despite use of vaccine

DRC conflict hampers effort to track and contain virus

By **Jon Cohen**

As an Ebola outbreak in a conflict-plagued region of the Democratic Republic of the Congo (DRC) continues to spread after 4 months, there's a glimmer of hope: An experimental Ebola vaccine appears to be helping the communities it reaches. More than 40,000 people have received the vaccine, by far the largest use of it since a trial in 2015 showed it worked well. The vaccine's effectiveness in this outbreak has not been formally assessed. But Peter Salama, who heads the Ebola response for the World Health Organization (WHO) in Geneva, Switzerland, says, "I think it's having a major impact."

WHO, which works in concert with the DRC's Ministry of Public Health, can't distribute the vaccine as widely as it would like, however, because of limited supplies, Salama notes. And the obvious targets for vaccination—people who have had contact with cases—have been difficult to identify and reach because of the ongoing conflict; a small number of front-line health care workers have even been caught in the crossfire.

So far the outbreak has tallied some 500 cases, about half of whom have died, according to the DRC. It spans a region of the DRC's northeast that abuts four other countries, and Salama and many others worry about the deadly virus jumping a border, which would require separate response teams and raise the risk of wider spread. Without more financial and personnel support from wealthy countries, the situation could expand quickly and become a long-running calamity similar to the Ebola epidemic that devastated three West African countries from 2014 to 2016, warns an editorial published last month in *The New England Journal of Medicine* (*NEJM*). A consensus statement from 25 public health and policy experts, also published the month, in *The Journal of the American Medical Association*, calls the outbreak "exceptionally" dangerous.

The editorials urge the U.S. government to change a policy that prevents its Centers for Disease Control and Prevention (CDC) from sending staff to the region because of security concerns. And many are calling for a WHO-established review panel to designate the outbreak a Public Health Emergency of International Concern, which could drive more countries to contribute to the response.

Although the toll so far is much smaller than the West African epidemic's 28,000 cases and 11,000 deaths, it's now the second largest Ebola outbreak ever documented, and one of the longest running. The outbreak has hit mothers and their young children especially hard, because many sought care for malaria at health centers that unknowingly have Ebola cases. Women have made up 62% of all cases, and 24% were children under age 15. Only about 50% of new infections are in people identified as having been in contact with a case, which underscores how the violence has interfered with contact tracing.

Salama says funding for the response in the DRC has been ample so far, but will run out in January 2019. Programs to support readiness and preparedness efforts in nine neighboring countries need another $45 million, he says. He's particularly worried about Uganda, Burundi, Rwanda, and South Sudan, all bordering the DRC. "If there is a case, as is very likely in surrounding countries, we want to pick up that first one so we can have a very robust response," Salama says.

He adds that the front-line responders from the DRC, WHO's 300 staffers on the ground, and personnel from nongovernmental organizations such as Doctors Without Borders and ALIMA are "frankly exhausted" from working long days in a conflict zone. "Where can we keep finding these brave people who are expert in viral hemorrhagic fever and know how to operate in a conflict affected area?" he asks.

CDC has perhaps more people with this twin skill set than any other institution, and it's working closely with WHO and others—but not in the hot zone. Salama says the complexities of this outbreak, such as the surprising role of malaria cases, require a nimble response, which would make CDC's seasoned staff invaluable. "A senior leadership cadre has long-term experience and can help drive teams in the right direction," he says.

Epidemiologist Jennifer Nuzzo, an author on the *NEJM* editorial and signer of the consensus statement, hopes the pressure will lead the U.S. government to rethink the policy of keeping CDC out of the region. "The situation is serious enough that we're pulling all the levers we can," says Nuzzo, who works at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. Salama, however, is skeptical the editorials will change the U.S. government's position.

Salama says if malaria cases drop, fewer people will visit health clinics, which could slow Ebola's spread. To that end, Ebola responders are distributing insecticide-treated bed nets. They're also beginning to offer the Ebola vaccine at malaria clinics. But fewer than 260,000 doses remain, and there are competing demands, including a push to vaccinate health care workers in the bordering countries. (Uganda has started to do so, and South Sudan plans to begin 19 December.)

Salama estimates that even in a best-case scenario the outbreak will run another 6 months. And it could be far worse. "This is the kind of massive, massive priority that the whole world should be very much focused on and willing to contribute to solving." ■

To map the understory, the GEDI laser will penetrate treetops in tropical forests such as the Amazon.

PHOTO: HEMIS/ALAMY STOCK PHOTO

REMOTE SENSING

Space laser to map trees in 3D

GEDI data will yield maps of forest carbon and biodiversity

By **Gabriel Popkin**

Tallying up the biomass in a forest—and monitoring changes to it—is no easy task. You can cordon off a patch of forest and use tape measures to assess tree growth, hoping your patch is representative of the wider forest. Or you can turn to aerial or satellite photography—if the pictures are available and sharp enough. But even the best cameras can't see past the forest canopy to the understory.

On 5 December, scientists gained a new tool for this tricky business when NASA's Global Ecosystem Dynamics Investigation (GEDI) was launched on a SpaceX rocket. The instrument, the size of a large refrigerator, is now mounted on the International Space Station, where it will soon gather data on the height and 3D structure of tropical and temperate forests. The campaign will help scientists understand whether forests are slowing or amplifying climate change, and identify prime habitat for valued species. "We've wanted this data set desperately," says Ralph Dubayah, a geographer at the University of Maryland in College Park and the project's principal investigator.

GEDI will harness a technology called light detection and ranging (lidar). Like its cousin radar, lidar sends out electromagnetic pulses and measures the reflections. But whereas radar uses radio waves, GEDI's lidar uses laser light, firing 242 times per second in the near-infrared. The focused, high-frequency radiation offers sharp resolution and can penetrate dense forests, bouncing not only off the treetops, but also off midstory leaves, branches, and the ground. Dubayah and his colleagues will combine GEDI data with ground measurements and statistical models to produce maps of tropical forest carbon that, at 1-kilometer resolution, should vastly shrink the errors of previous maps.

Countries that want to use the carbon stored in their forests to help meet Paris agreement climate targets may use those maps to gauge progress, says Naikoa Aguilar-Amuchastegui, director of forest carbon science at the World Wildlife Fund in Washington, D.C. Researchers tracking forest degradation, due to the selective logging of individual trees and fuelwood harvesting from the understory, are eager for the data, too. Those activities are invisible to imaging satellites such as Landsat, says Laura Duncanson, a research scientist at NASA's Goddard Space Flight Center in Greenbelt, Maryland. "GEDI gets you that third dimension," she says.

The 3D maps could also identify the forests with the rich structure and diverse vegetation favored by at-risk species such as the orangutan, says Scott Goetz, an ecologist at Northern Arizona University in Flagstaff and a mission deputy principal investigator. The maps could find priority areas for conservation, and even help plan habitat corridors for wildlife migrating because of climate change.

The finely tuned laser will resolve the heights of treetops and the ground more precisely than previous instruments—crucial for monitoring the health of the carbon-dense mangrove forests that shroud tropical coastlines, says Goddard research scientist Lola Fatoyinbo Agueh. Knowing how high the mangroves sit above the water will determine whether they will keep pace with sea level rise or die back, releasing stored carbon—a key input for climate models, she says.

GEDI's perch on the space station—chosen to keep its cost below a $94 million cap—comes with a drawback, however. Its view will be confined to latitudes between 51.6° north and south. That means it will miss the boreal forests of North America and Asia. And it will likely be displaced by a Japanese instrument after 2 years. The short mission will make it harder to answer an urgent question: Are tropical forests overall a carbon sink, capturing some of the emissions from vehicles and industry, or a source? That depends on whether forest growth is sequestering more carbon than deforestation and degradation are releasing. But seeing such a trend requires years of continuous data, says Wayne Walker of the Woods Hole Research Center in Falmouth, Massachusetts. "Nothing's better than a long-term record."

GEDI also can't distinguish tree species, which vary in carbon density. Dubayah is using species-specific measurements from about 5000 field plots to calibrate the GEDI data. But with more than 40,000 tree species in the tropics, even more field plots would help, says Oliver Phillips, an ecologist at the University of Leeds in the United Kingdom who runs a large tropical forest plot network. "A large ground effort is needed to get maximum value from this," Phillips says.

Researchers may be able to work around some of these limitations. Alessandro Baccini, a remote sensing scientist also at Woods Hole, hopes to train machine learning algorithms to extend carbon estimates into the past and future by using GEDI's carbon maps to calibrate long-term forest-cover data from imaging satellites. He adds that by combining data from GEDI and ICESat-2, a NASA lidar satellite launched in September that primarily measures ice sheets but is flying over the whole planet, investigators could construct a global carbon map—one that includes the boreal forest. Still, Baccini wants more. "Why can't we have a proper mission designed for vegetation that is global?" he asks. ■

Gabriel Popkin is a journalist in Mount Rainier, Maryland.

NEUROGENETICS

Human brain samples yield a genomic trove

Regulatory DNA takes center stage in search for mechanisms of disease

By **Kelly Servick**

More than 2000 human brains stored in tissue banks are giving up their genetic secrets. Genome scans have already revealed hundreds of locations where DNA tends to differ between people with and without a particular psychiatric disease. But those studies don't pin down specific culprit genes or what they do in the brain. "There was kind of a missing link," says Daniel Geschwind, a neurogeneticist at the University of California (UC), Los Angeles. He and others in the 3-year-old PsychENCODE Consortium, fueled by roughly $50 million from the U.S. National Institutes of Health (NIH) in Bethesda, Maryland, have tried to bridge that gap by tracking which genes are expressed, and where.

The consortium focuses on regulatory regions, which control the expression of protein-coding genes, and which previous studies implicated as drivers of psychiatric disease risk. PsychENCODE collaborators have cataloged differences in the activity of these regulatory regions in different parts of the brain, at different stages of brain development, and in brains affected by different disorders—chiefly schizophrenia, autism, and bipolar.

The result, outlined this week in a series of papers in *Science* and its sister journals *Science Advances* and *Science Translational Medicine*, is the most complete picture yet of how regulatory regions influence the brain. In one of the new papers, for example, researchers describe DNA sites where a variation in a sequence changes the expression of a protein-coding gene elsewhere. Before PsychENCODE, that list consisted of fewer than 5000 locations, Geschwind says, but the consortium's work has brought the total to roughly 16,000.

"These data allow us to do things we've been wanting to do for a while," says Gerome Breen, a psychiatric geneticist at King's College London who was not in the consortium but plans to use its publicly available data set. Not all researchers are optimistic that the new data set will directly lead to new drugs for illnesses. But many expect it to reveal clues to how complex diseases develop.

The collaborators analyzed their brain samples with RNA sequencing to find out which genes were transcribed. They also did various epigenetic analyses, such as measuring how DNA's folded structure brings regulatory regions into contact with distant protein-coding regions.

The immense data set allows researchers to identify genome "modules"—groups of genes that tend to be expressed together and have common functions. Unique patterns of gene expression in a module might reveal a nuanced genetic feature of a disease. For example, previous studies have shown the expression of genes involved in neural signaling tends to be unusually low in autism, and to a lesser extent, in bipolar disorder and schizophrenia. But PsychENCODE data enabled a finer-grained analysis. They revealed modules including one containing genes that control how cells package and release their chemical messengers into synapses. That set of genes, it turns out, is especially active in schizophrenia and bipolar disorder, but not in autism. Such details might point to brain processes that could be targets for therapies.

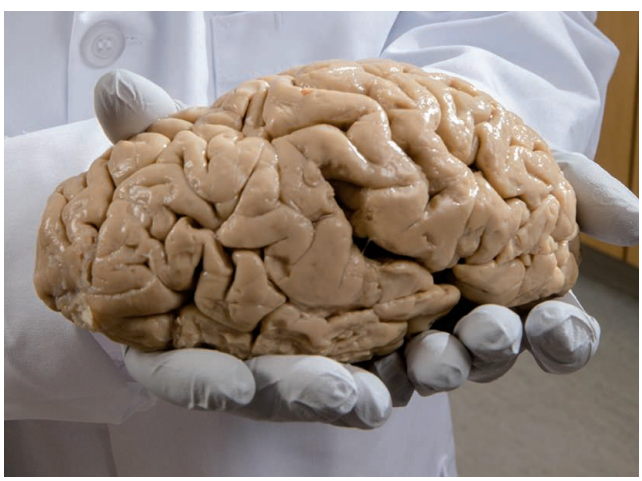
The new data set can also reveal windows of brain development when disease-associated genes seem to have the most influence, says Geetha Senthil, the NIH program director who has coordinated and overseen PsychENCODE. Those windows, in turn, might be the times when intervention would be most valuable. Doctors can already observe, based on a patient's symptoms, when a disease seems to take hold, but, she says, "having a biological clue would be thrilling."

The project's namesake, ENCODE (Encyclopedia of DNA Elements), was a broader quest to map noncoding regions of the human genome. Its initial results, unveiled in 2012, stirred controversy. Scientists disputed the team's claim that most of the genome was functional and questioned whether the project's insights would be worth NIH's $185 million investment (*Science*, 21 March 2014, p. 1306).

Dan Graur, an evolutionary geneticist at the University of Houston in Texas and one of the most outspoken critics of ENCODE, also finds fault with some of the initial PsychENCODE results. The project targets psychiatric disorders that are themselves poorly defined, he says. "If you take something vague and correlate it with millions of genetic and epigenetic variations, you are bound to get statistical significance that will have little biological significance."

Neurogeneticist Kevin Mitchell of Trinity College Dublin echoes some of Graur's concerns. "I'm not fully convinced that we know more today than we did yesterday," he says. He doubts that a profile of gene expression can define disorders as heterogeneous as schizophrenia or autism—or give new insights into how to treat them. "It's a huge amount of work, very well intended and very well done," he says, "but there are some limits to what you can do with genomics."

But many researchers defend the project's value. "I'm sure there are researchers out there who will look at these first papers and say, … 'Where is our paradigm-shifting finding?'" says Alexander Nord, a neurogeneticist at UC Davis who was not in the consortium. "That's a bit of a straw man, expecting us to find that in one set of analyses." The data set will grow richer as researchers work to interpret it, he says. "It's not going to go out of style." ■



Tissue from brain banks fed a genomic data set that may hold clues to the origins of schizophrenia, autism, and other conditions.

PHOTO: MCLEAN HOSPITAL

MATERIALS RESEARCH

Bioelectronics that vanish in the body

Wire-free devices that dissolve could expand the use of electric pulses in medicine

By **Robert F. Service**, *in Boston*

Implanted electronics can steady hearts, calm tremors, and heal wounds—but at a cost. These machines are often large, obtrusive contraptions with batteries and wires, which require surgery to implant and sometimes need replacement. That's changing. At a meeting of the Materials Research Society here last month, biomedical engineers unveiled bioelectronics that can do more in less space, require no batteries, and can even dissolve when no longer needed.

"Huge leaps in technology [are] being made in this field," says Shervanthi Homer-Vanniasinkam, a biomedical engineer at University College London. By making bioelectronics easier to live with, these advances could expand their use. "If you can tap into this, you can bring a new approach to medicine beyond pharmaceuticals," says Bernhard Wolfrum, a neuroelectronics expert at the Technical University of Munich in Germany. "There are a lot of people moving in this direction."

One is John Rogers, a materials scientist at Northwestern University in Evanston, Illinois, who is trying to improve on an existing device that surgeons use to stimulate healing of damaged peripheral nerves in trauma patients. During surgery, doctors suture severed nerves back together and then provide gentle electrical stimulation by placing electrodes on either side of the repair. But because surgeons close wounds as soon as possible to prevent infection, they typically provide this stimulation for an hour or less.

Rogers and his collaborators wondered whether they could extend the treatment by harnessing the soft, flexible, dissolvable electronic materials they developed a few years ago (*Science*, 28 September 2012, p. 1640). They used a mix of metals, semiconductors, and polymers to fashion a simple coil with two electrodes. The coil was designed to act as an antenna, picking up radiofrequency pulses transmitted wirelessly from outside the body, and converting them into mild electrical pulses. Rogers and his team implanted the devices in 25 rats in which they had cut the sciatic nerve to one of the hind legs, and stimulated the nerve ends for 1 hour a day for up to 6 days.

The stimulation sped nerve healing by about 50% compared with animals that received no stimulation or just one or a few days of it, they reported in the 8 October issue of *Nature Medicine*. And there was no need to reopen the wounds to remove the gadgets. The materials broke down and were excreted. "After 21 days the device is completely gone, and there appeared to be no adverse effect" from degradation, Rogers says.

"There is no doubt there is a potential clinical application here," Homer-Vanniasinkam says. However, she notes that before dissolvable electronics make their way into people, researchers will need to confirm that all the materials from the devices degrade safely.

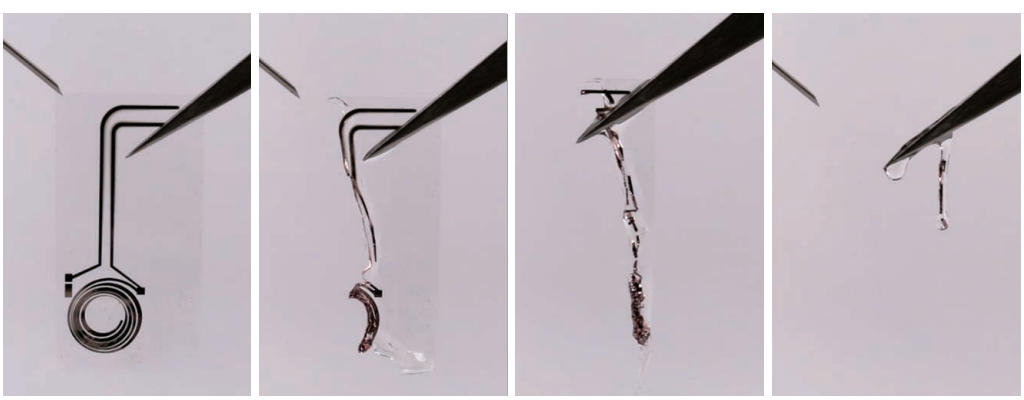
Xudong Wang, a bioelectronics expert at the University of Wisconsin in Madison, is developing miniature, wireless devices that take advantage of a technology pioneered by others to convert the body's motion into electrical current. In one study reported on 29 November in *ACS Nano*, a fingertip-size generator that delivered a stream of tiny electrical pulses to wounds on rats' skin sped healing. And at the meeting, Wang described similar generators that mimic commercially available implanted electrodes meant to help patients with obesity lose weight.

These devices stimulate a branch of the vagus nerve, which runs from the colon and stomach to the brain stem, helping relay signals of fullness after eating. Available devices are pacemaker-size and contain batteries that often need replacement, requiring repeated surgeries. Wang and his colleagues wanted to see whether their much smaller device, which requires no batteries, could do the same job.

They implanted their device on the outer wall of a rat's stomach, so the organ's motions during eating would power the generator. At the meeting, Wang reported that animals with the generator ate at normal times, but less than control animals. The rats lost 38% of their weight over 18 days, at which point their weight stabilized.

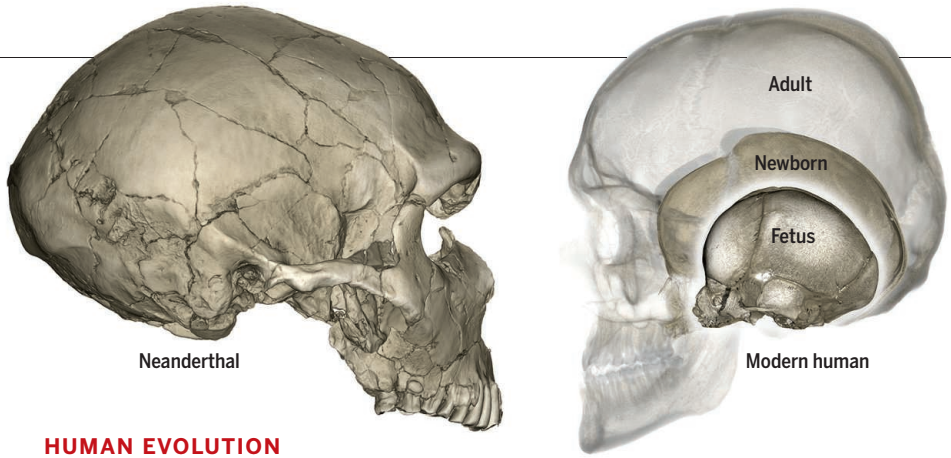
Jacob Robinson, an applied physicist at Rice University in Houston, Texas, shrank his implantable stimulator even further, to the size of a grain of rice. It is powered not by movement, but by magnetic field pulses delivered from outside the body, and is intended to replace the large, battery-powered brain stimulators used to control tremors in some patients with Parkinson's disease. In rats with a version of the disease, Robinson implanted his minuscule device in the subthalamic nucleus, the same brain region targeted by larger devices. The animals' tremors disappeared, and their movements became normal, he said at the meeting.

"It's very encouraging," Rogers says. Robinson and others are aiming their stimulators at well-established clinical areas with an urgent need for better devices, he notes. "Having immediate use is going to be very powerful," Rogers says, because it could help speed the approval of such devices by regulators—and smooth their way into patients. ■



This implantable electronic device can speed nerve healing and dissolves when its work is done.

PHOTOS: J. ROGERS/NORTHWESTERN UNIVERSITY



HUMAN EVOLUTION

Why modern humans have round heads

Neanderthal DNA points to genes that influence brain

By **Ann Gibbons**

Ever since researchers first got a good look at a Neanderthal skull in the 1860s, they were struck by its strange shape: stretched from front to back like a football rather than round like a basketball, as in living people. But why our heads and those of our ice age cousins looked different remained a mystery.

Now, researchers have found an ingenious way to identify genes that help explain the contrast. By analyzing traces of Neanderthal DNA that linger in Europeans from their ancestors' trysts, researchers have identified two Neanderthal gene variants linked to slightly less globular head shape in living people, the team reports this week in *Current Biology*. The genes also influence brain organization, offering a clue to how evolution acting on the brain might have reshaped the skull. This "very important study" pinpoints genes that have a "direct effect on brain shape and, presumably, brain function in humans today," says paleoanthropologist Chris Stringer of the Natural History Museum in London, who was not a part of the work.

Cradle a newborn and you'll see that infants start life with elongated skulls, somewhat like Neanderthals. It's only when the modern human brain nearly doubles in size in the first year of life that the skull becomes globular, says paleoanthropologist Philipp Gunz of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany. He and his colleagues analyzed computerized tomography scans of modern human and Neanderthal skulls to develop a "globularity index" of human brains.

To explore the underlying differences in brain tissue, they applied that index to MRI scans from 4468 people of European ancestry whose DNA had been genotyped. The team identified two Neanderthal DNA fragments that were correlated with slightly less globular heads. These DNA fragments affect the expression of two genes: *UBR4*, which regulates the development of neurons, and *PHLPP1*, which affects the development of myelin sheaths that insulate axons, or projections of neurons.

The Neanderthal variants may lower *URB4* expression in the basal ganglia and also lead to less myelination of axons in the cerebellum, a structure at the back of the brain. This could contribute to subtle differences in neuronal connectivity and how the cerebellum regulates motor skills and speech, says senior author Simon Fisher of the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands. But any effects of the Neanderthal genes in living people would be slight because so many genes shape the brain.

Tying Neanderthal DNA to brain scans in living people is an "innovative and exciting approach" because "soft tissue in the brain is impossible to access from the fossil record," says anthropologist Katerina Harvati of the University of Tübingen in Germany. She'd like to see the findings confirmed in more people.

Indeed, Gunz and Fisher plan to delve into the UK Biobank, a giant database of British people's health records and DNA. They hope to use Biobank brain scans to find more genes and to explore how Neanderthal brains would have functioned. "The Neanderthal DNA that remains in us can help us think about what their brains were like," says geneticist Tony Capra of Vanderbilt University in Nashville. ■

Scans of skulls show modern human infants start out with elongated heads—somewhat like Neanderthals—but they round out in adulthood.

IMAGE: PHILIPP GUNZ/CC BY-NC-ND

SCIENTIFIC COMMUNITY

Conferences score well on child care

Male-dominated disciplines lead the pack

By **Katie Langin**

This year, 68% of major scientific conferences held in North America provided child care support for parent attendees, *Science* found after examining resources available at 34 meetings, each attended by more than 1000 people. An even larger share—94%—made a lactation room available for nursing mothers.

"That's good," says Rebecca Calisi, an animal physiologist at the University of California, Davis, and author of an opinion piece published in March arguing that conferences need to do a better job supporting parent attendees. But, she adds, they still aren't good enough—those statistics should be 100%.

Of the conferences that offered support, 83% arranged for licensed providers to operate at conference facilities, where parents were charged between $40 and $110 a day. Two societies offered free child care at their annual meetings: the American Chemical Society and the American Association of Physical Anthropologists. Five conferences awarded child care grants that parents could use for a variety of child care-related expenses, for example, to pay for their child's travel, for travel expenses incurred by a caregiver, or to hire a nanny.

The disciplines with the most room for improvement are the ones that tend to have a greater share of women. Only about half of the 18 conferences in the life sciences and social sciences offered child care accommodations for parents—a much lower percentage than in the physical sciences, math, and computer sciences (85% of 13). Of three multidisciplinary conferences, two provided child care accommodations.

"There's still so much to do, but it's great to see" so many conferences helping parents, Calisi says. "Whether it's one small baby step, or a huge leap, as long as we're going in the right direction that's what's important." ■

Read more about the results and the personal stories behind the data at https://scim.ag/ConfChildcare.

FEATURES

STARRY EYES

NASA is planning four of the largest space telescopes ever. But which one will fly?

By **Daniel Clery**; *Illustration by* **Eiko Ojala**; *Graphics by* **Chris Bickel**

For NASA astronomers, this was not a good year. In June, a review board found that the agency's prized observatory—the already overdue and vastly overbudget $8.8 billion James Webb Space Telescope (JWST)—was still years away from taking flight and capturing the faint light of the universe's first stars. The holdup: torn sunshields and loose bolts. Also in trouble was the next big astrophysics mission in line, the Wide Field Infrared Survey Telescope (WFIRST), intended to pin down the nature of mysterious dark energy by surveying wide swaths of the sky. Not even off the drawing board, WFIRST was predicted to burst its $3.2 billion budget by $400 million, another review panel found—not a plus for a mission that the administration of President Donald Trump was already thinking of canceling.

Yet astronomers are about to look skyward and dream even bigger dreams. The decadal survey in astrophysics, which sets priorities for future missions by NASA, the Department of Energy, and the National Science Foundation, began last month. Dozens of astronomers, broken into committees, will identify science goals and develop a wish list of telescopes, both on the ground and in space, that could best address them. One of the toughest tasks will be to decide which—if any—of four proposed successors to the JWST and WFIRST most deserves to fly as a NASA flagship observatory. It would be launched in the 2030s to L2, a gravitationally balanced spot between the sun and Earth.

On the following pages, *Science* examines those dream telescopes. The Large UV Optical Infrared Surveyor (LUVOIR), a 15-meter-wide giant with 40 times the light-collecting power of the Hubble Space Telescope, is a bid to look back at the universe's first galaxies, and to answer the question: Is there life elsewhere in the universe? The Habitable Exoplanet Observatory (HabEx) would also focus on that question, but with a smaller mirror. HabEx would fly in tandem with a separate spacecraft carrying a starshade the size of a soccer field. By blocking the glare of a star, the starshade would reveal Earth-like exoplanets, enabling HabEx to scrutinize their faint light for signatures of life. The Lynx X-ray Observatory would gather x-rays from the universe's first black holes to learn how they help galaxies form and evolve. And the Origins Space Telescope, with machinery to chill its telescope to just 4° above absolute zero, would study a little-explored kind of infrared radiation emanating from the cold gases and dust that fuel star and planet formation.

Whichever concept rises to the top, researchers hope it has a smoother path to space than the missions chosen in previous surveys. The 2001 survey picked the JWST as its top priority, but that telescope will be lucky to meet its scheduled launch in 2021, 2 decades later. WFIRST was the top pick of the 2010 survey, but it won't fly before 2025. There's a general sense that the initial proposals were immature and unrealistic, says Roger Blandford of Stanford University in Palo Alto, California, who chaired the 2010 survey. "There's frustration all around."

This time, NASA wants the concepts on a firmer footing. Not only did the agency identify the four flagship concepts early, back in 2015, but it has since funded teams to work up rough designs for each one. In June 2019, the teams will deliver to NASA a report that includes two concepts—one expensive and

big, the other constrained and relatively affordable at less than $5 billion in most cases. (Here, *Science* examines the larger concepts.)

"This prepreparation will put the survey in a better situation to evaluate the possibilities," says Fiona Harrison, a high-energy astrophysicist at the California Institute of Technology in Pasadena who was named last month as co-chair of the survey along with Robert Kennicutt of Texas A&M University in College Station. The product of the decadal survey—a prioritized list of missions delivered in 2020—is supposed to be consensual, in part so that agencies and scientists can lobby Congress for funding with a unified voice. But competition among the four flagships will be fierce.

LUVOIR's backers tout its wide appeal as a general-purpose observatory in the mold of Hubble. LUVOIR's instruments cover the parts of the spectrum where the universe is brightest, and the huge size of its mirror means it can peer the farthest, at the faintest objects, with the sharpest vision. "It transcends astrophysics," says Jason Kalirai of the Space Telescope Science Institute (STScI) in Baltimore, Maryland. Critics argue that LUVOIR's huge mirror will lead to a huge price tag and inevitable delays, as the JWST's 6.5-meter mirror already has.

Proponents of the cheaper HabEx hope it will ride high on surging enthusiasm for exoplanets—and a concern for simplicity and thrift. But flying in formation with a distant starshade is an untested technique. And though HabEx can study a few nearby planets in detail, its smaller mirror—4 meters compared with LUVOIR's 15 meters—means more distant worlds will be out of reach. LUVOIR and HabEx will compete head-to-head for the committee's attention, and HabEx and LUVOIR team member Chris Stark of STScI says there won't be a need to launch both. "There are only so many nearby stars."

Origins would look back in time to see how dust and molecules coalesced to create the first galaxies and black holes and how the disks around young stars clump into exoplanets. But the JWST and the Atacama Large Millimeter/submillimeter Array in Chile can capture some of the same wavelengths, squeezing Origins's discovery space.

Lynx would take up the mantle of NASA's aging Chandra X-ray Observatory, zooming in on hot gas swirling into a black hole or jetting from the center of a galaxy. That would placate x-ray astronomers still smarting from the low rating their International X-ray Observatory proposal received in the 2010 decadal survey. "We got robbed at the last decadal," says STScI x-ray astronomer Rachel Osten. "Is it time for x-rays?"

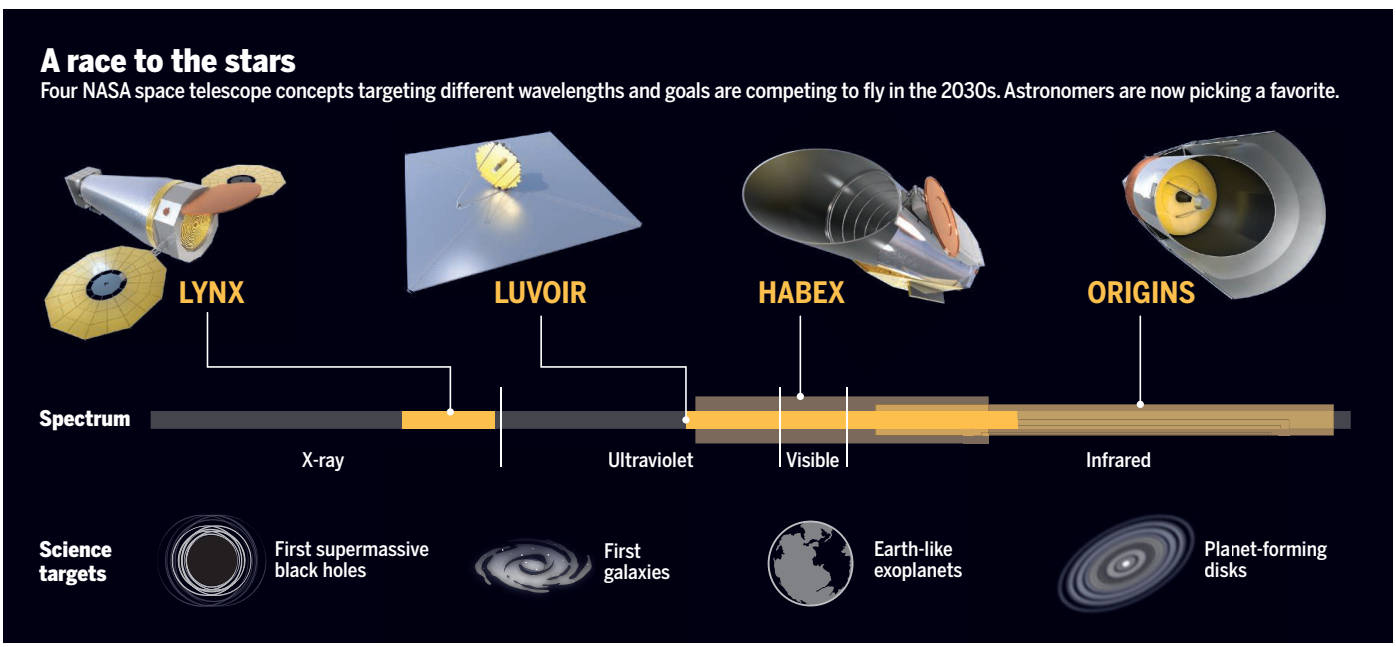
Whichever mission wins the decadal's favor, funders will ask: How do we know it won't be another JWST, swallowing up budgets and delaying other projects? Study director Dwayne Day of the National Academies of Sciences, Engineering, and Medicine (NASEM) in Washington, D.C., which organizes the decadals, says the survey is taking a sophisticated approach to estimating costs, hoping "to avoid sticker shock, committing to something that is too expensive to afford."

Day says project teams usually estimate costs by tallying labor, materials, and testing. "It's good, but it leaves out unforeseen circumstances, threats." So, for the past decade NASEM has been paying The Aerospace Corporation of El Segundo, California, to apply a cost model called CATE (for Cost And Technical Evaluation) to any proposals a decadal wishes to consider.

CATE draws on a database that goes back decades and contains details of cost and performance for more than 150 NASA missions and 700 instruments. When presented with a new mission, CATE can say how similar missions have fared in the past. The model is particularly powerful in assessing the things that can go wrong. "The best forecasters can't have hands on all the unknown unknowns," says Debra Emmons, a senior manager with Aerospace in Chantilly, Virginia. For example, if a sensor takes longer than expected to develop, or if an international partner delivers an instrument late, the project can be delayed and costs can rise. "[CATE] assesses technical threats, monetizes them, and makes a forward projection," she says. Paul Hertz, NASA's astrophysics chief in Washington, D.C., calls it "a great addition to the tool set."

The project teams are wary of the exercise, fearing that if they produce a scientifically bold and technically challenging proposal, CATE might judge it to be risky and expensive, Emmons says. And NASA wants the four project teams to be ambitious. "The missions had better be hard to do because the questions are hard," Hertz says.

But with the still-grounded JWST on everybody's mind, astronomers are eager to ensure that the winner of the great space telescope bake-off is at once dreamy and real. Blandford says: "It gives a rationale for making these terrible decisions." ■



GRAPHIC: C. BICKEL/*SCIENCE*

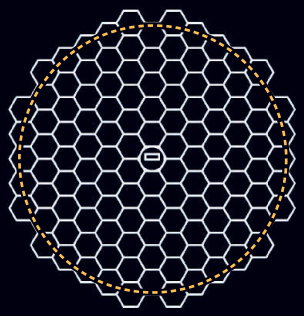
A giant eye to see to the beginning of time

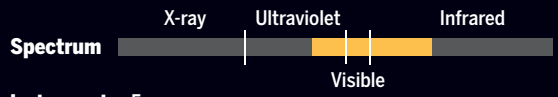
The Large UV Optical Infrared Surveyor (LUVOIR) "is a Swiss army knife," says LUVOIR study scientist Aki Roberge of NASA's Goddard Space Flight Center in Greenbelt, Maryland. Much like its multipurpose predecessor, the Hubble Space Telescope, LUVOIR would gather light over a broad spectrum. But Hubble has a 2-meter mirror, whereas LUVOIR's would be up to 15 meters across in one version, larger than that of any of today's ground-based telescopes.

Like its chief competitor, the Habitable Exoplanet Observatory, LUVOIR will scrutinize Earth-like exoplanets for signs of life. But the telescope's extraordinary light-gathering power would allow it to see more of those worlds. Another big question will be within its reach: How do galaxies form and evolve? By capturing ultraviolet wavelengths invisible from the ground, LUVOIR will see gas cycling in and out of galaxies to fuel star formation. The observatory will even be able to pick out individual stars in distant galaxies, giving a picture of what sort of stars are born where.

LUVOIR comes with risks. Fitting the mirror inside a rocket fairing will require origami even more complex than that for the 6.5-meter James Webb Space Telescope (JWST), which LUVOIR would supersede. And the planned heavy-lift rocket—a future version of NASA's troubled Space Launch System—may never materialize. At more than twice the JWST's size, LUVOIR will more than double its $8 billion cost, critics say.

Not so, supporters say: The mirror is only a fraction of the mission's cost and LUVOIR won't need the elaborate sunshield or cryocoolers that were essential for the JWST's infrared instruments. And LUVOIR's mirror will be made of glass, not the JWST's trickier beryllium. "There's no magic involved. All the technology is feasible," says LUVOIR team member John O'Meara, chief scientist of the Keck Observatory in Hawaii.



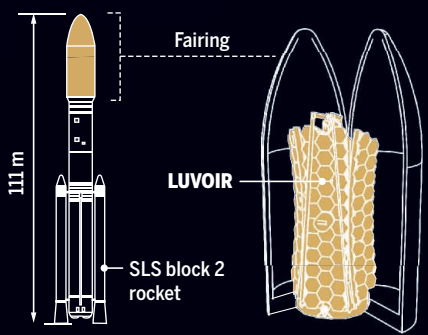


Instruments: Four
Orbit location: Sun-Earth L2
Launcher: Space Launch System block 2
Launch mass: 25 metric tons
Primary science targets: Earth-like exoplanets and first galaxies

LUVOIR

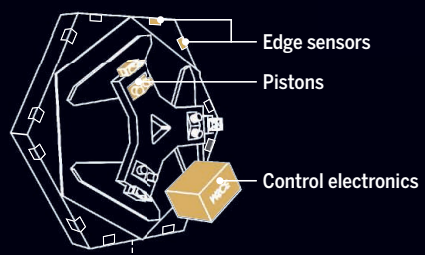
Folded for liftoff

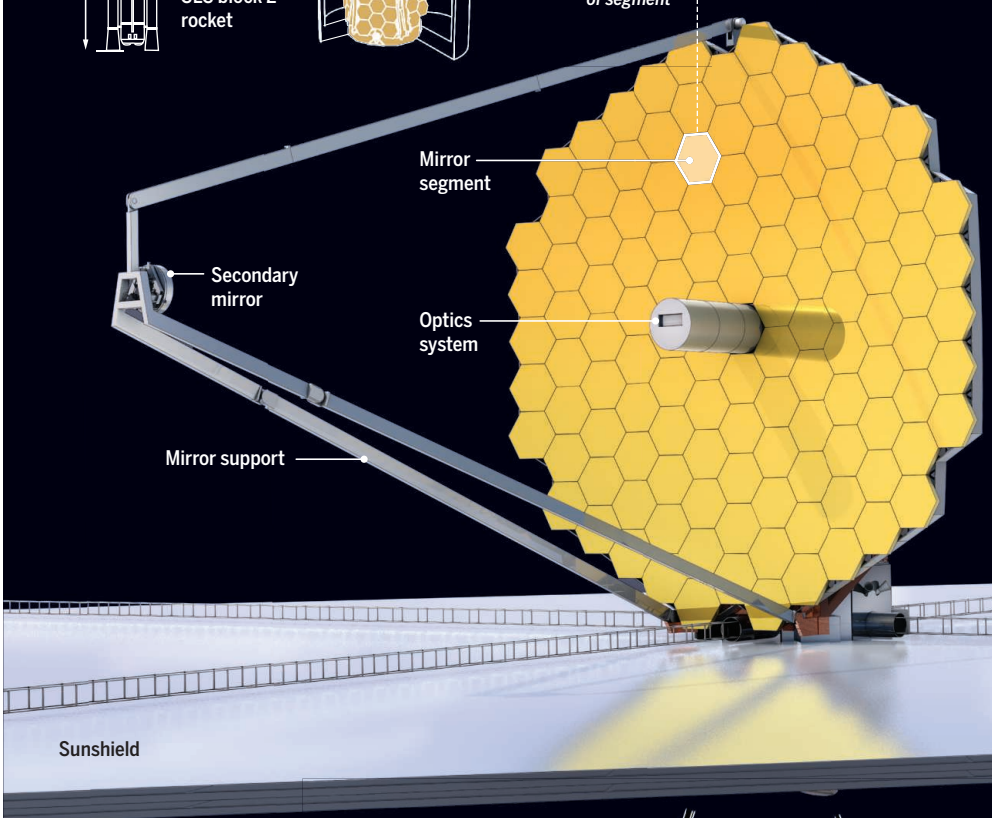
LUVOIR's mirror will fold to fit inside the 8.4-meter-wide fairing of NASA's Space Launch System (SLS) block 2. The troubled heavy-lift rocket isn't expected until the 2030s, however, and it may never fly.



Movable mirrors

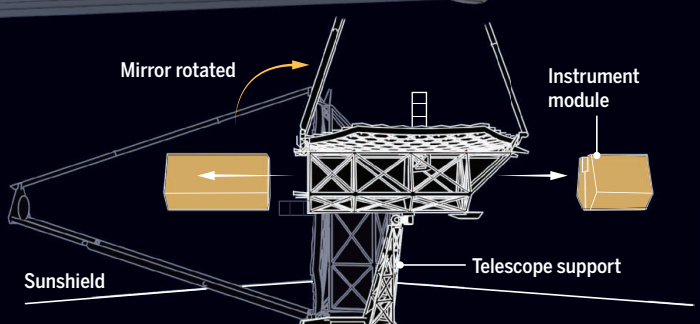
Tiny pistons will tip and tilt LUVOIR's 120 mirror segments into a perfect shape with the help of 622 edge sensors.



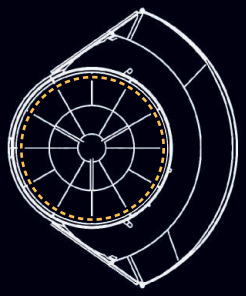


Built to last

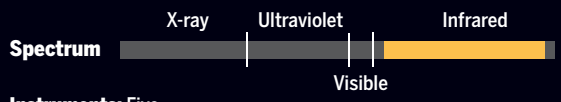
Robotic servicing missions could extend LUVOIR's life to several decades. Standardized valves, latches, and rails ease the replacement of batteries, solar panels, computers, reaction wheels, and propellant. Rotating the mirror away from the sunshield eases instrument replacement.



CREDITS: (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) DREW JONES, NASA/GSFC

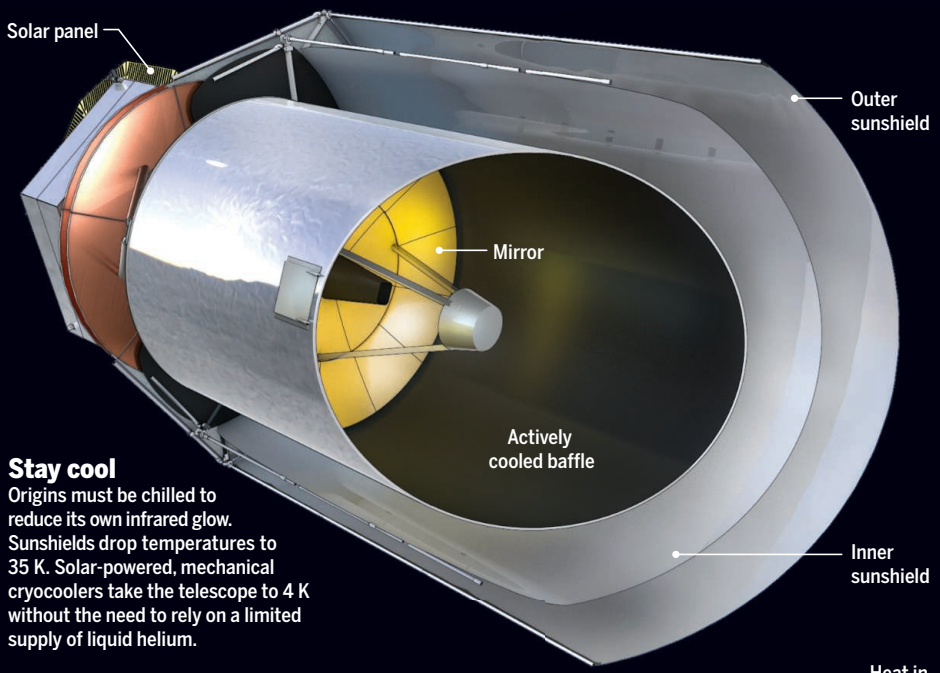


9.1-meter primary mirror



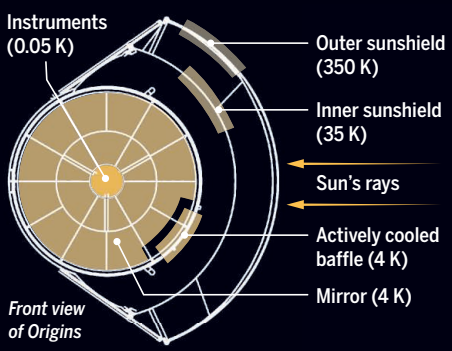
Instruments: Five
Orbit location: Sun-Earth L2
Launcher: Space Launch System block 2
Launch mass: 30 metric tons
Primary science targets: Gas clouds and planet-forming disks

ORIGINS



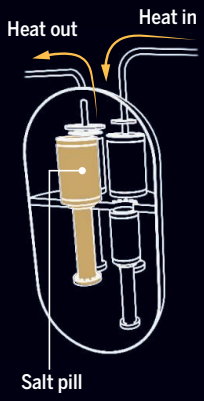
Stay cool

Origins must be chilled to reduce its own infrared glow. Sunshields drop temperatures to 35 K. Solar-powered, mechanical cryocoolers take the telescope to 4 K without the need to rely on a limited supply of liquid helium.



Approaching absolute zero

Detectors must be cooled even further, to 0.05 K. A magnetic field aligns salt molecules in a "salt pill." As they drift out of alignment they absorb heat. Realignment pumps heat out of the capsule.

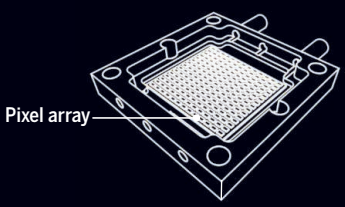


Sensing the far infrared

Far-infrared photons are feeble. Two rival detector types, never flown in space, each rely on superconducting circuits with zero resistance. Detector arrays must be scaled up from 1000 to as many as 16,000 pixels.

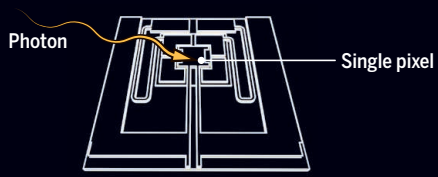
Microwave kinetic inductance detector

Incoming photons break up the pairs of electrons that confer superconductivity in a resonant circuit, resulting in a detectable change in electrical properties.



Transition edge sensor

The detector is kept right at its superconducting transition temperature. The slight heating from a photon creates a detectable rise in resistance.



CREDITS: (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) ORIGINS SPACE TELESCOPE STUDY TEAM

A cold stare at the faint glow of gas and dust

The Origins Space Telescope will stare at the cold universe: galactic gas clouds, planet-forming disks, exoplanet atmospheres, and other objects that don't burn bright but glow feebly in the far infrared. That means the telescope itself must be frigid, chilled to 4° above absolute zero to stanch its own infrared light. Earth's atmosphere largely blocks the far infrared, and few instruments have studied the range of wavelengths targeted by Origins. One pioneer was Europe's Herschel Space Observatory, which from 2009 to 2013 cooled its instruments by boiling off a limited supply of liquid helium. Origins will be much more sensitive as well as long-lived: Solar-powered mechanical cryocoolers will chill the entire 9.1-meter telescope and its five instruments while a sunshield fends off the sun's heat.

The three biggest challenges in developing Origins are "detectors, detectors, and detectors," says Origins study scientist Dave Leisawitz of NASA's Goddard Space Flight Center in Greenbelt, Maryland. Neither industry nor the military has much interest in far-infrared detectors, so astronomers are doing the R&D themselves, weighing three rival technologies. "There is a clear path to choosing one or the other," Leisawitz says. Such detectors have not flown in space before, and Origins co-leader Margaret Meixner of the Space Telescope Science Institute (STScI) in Baltimore, Maryland, says, "We want to make them bigger, more sensitive, and more efficient."

By tracking infrared emissions from simple molecules, dust, and aromatic hydrocarbons, Origins could follow gas clouds collapsing into stars and dust disks spawning planets. Water also falls into Origins's spectral sweet spot. By monitoring water's spectral lines, Origins could track it from interstellar clouds to protoplanetary disks and onto habitable worlds. "The greatest discoveries," says Origins team member Klaus Pontoppidan at STScI, "will be things we haven't even thought about yet."

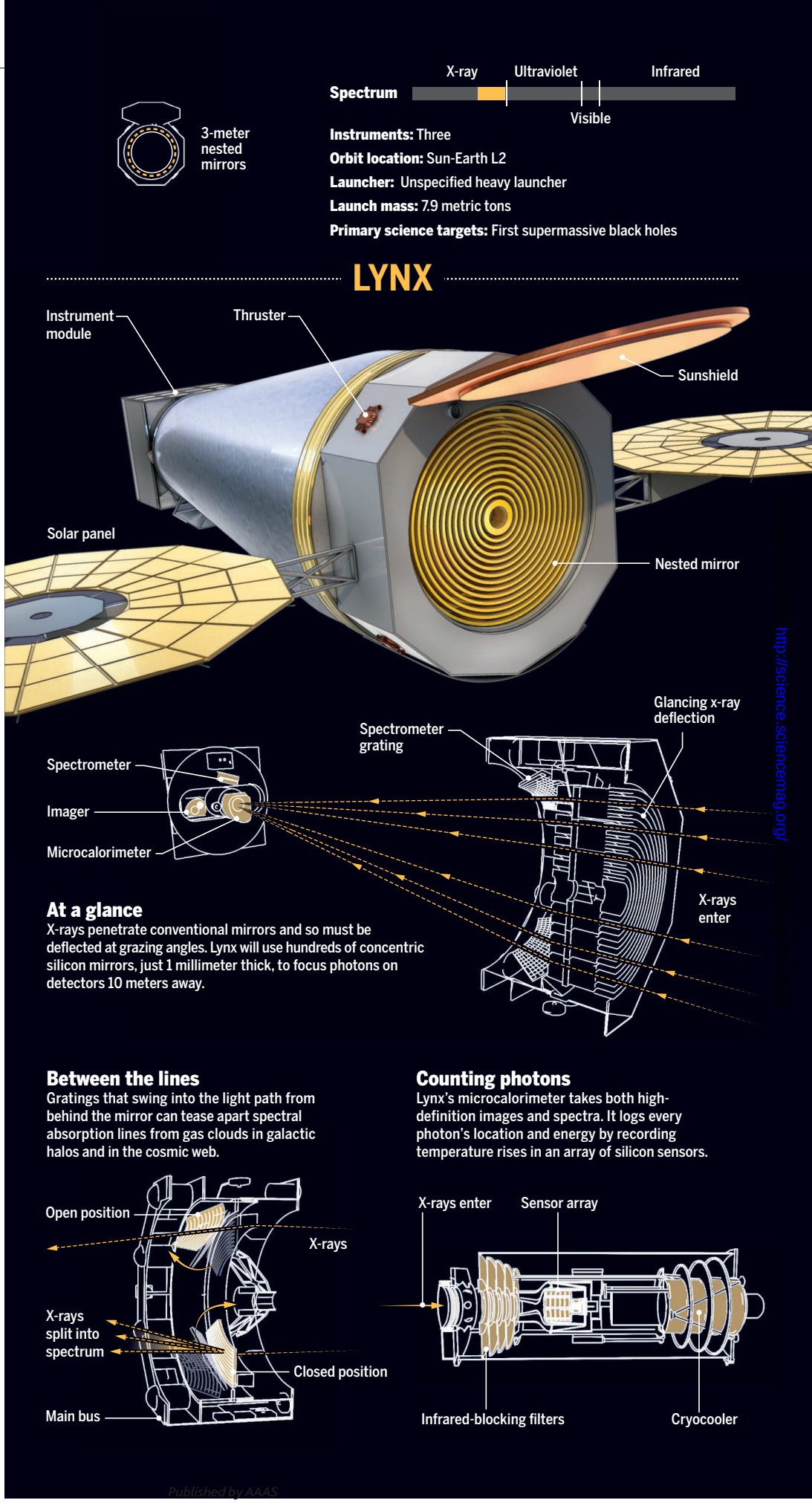
An x-ray journey to the dawn of black holes

X-rays, so useful in penetrating the body, are a pain for astronomers to gather. Earth's atmosphere blocks them, so astronomers must get to space to see the million-degree gases that shine in x-rays. But even in space the energetic photons are elusive, passing straight through conventional mirrors instead of reflecting. Only a few thousand x-ray sources are known, despite the work of pioneering missions such as NASA's Chandra X-ray Observatory and Europe's X-ray Multi-Mirror Mission–Newton.

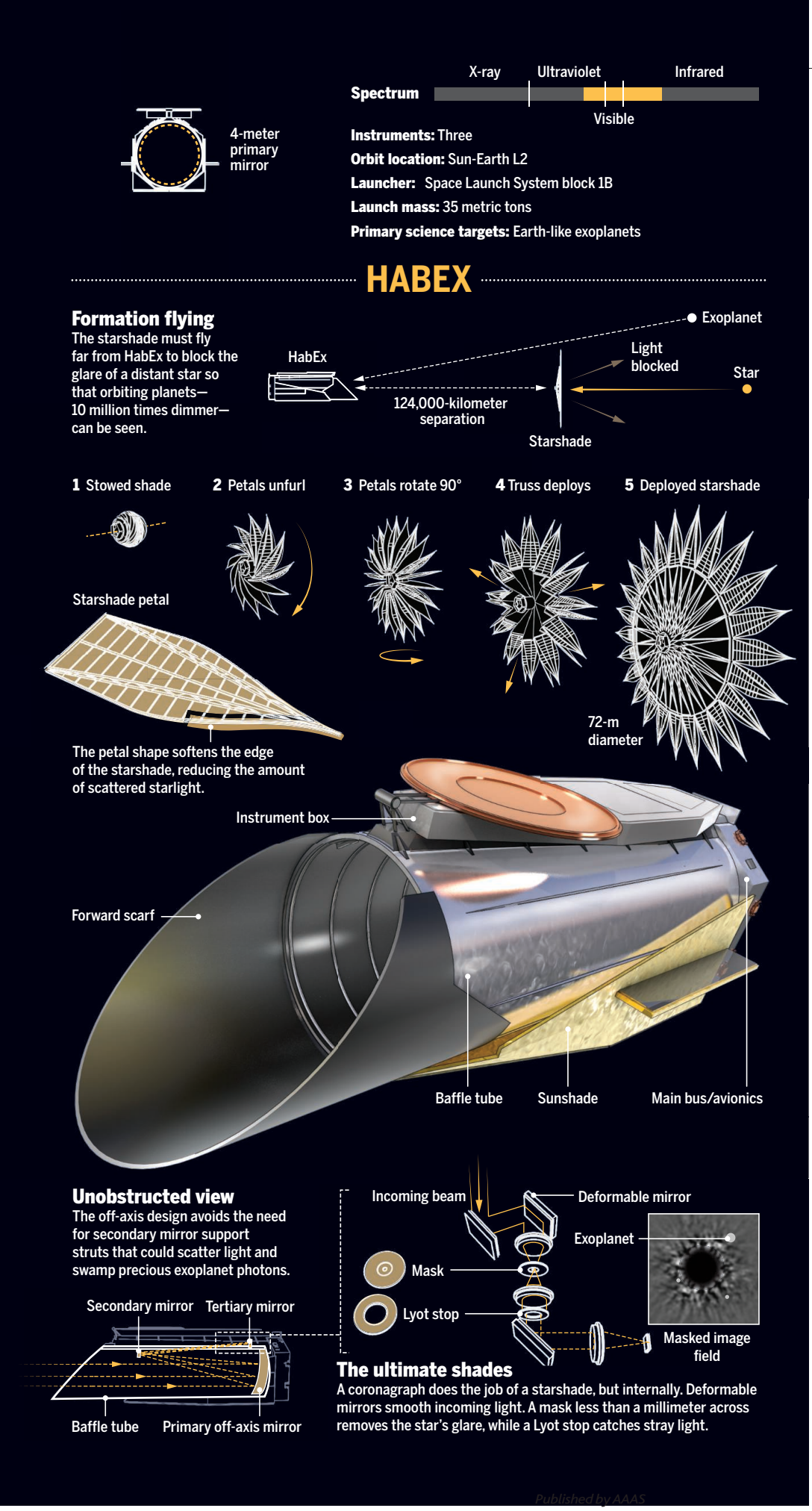
The Lynx X-ray Observatory is designed to find thousands more sources by going deeper and fainter. It would gain its unprecedented sensitivity from hundreds of silicon mirrors, each just a millimeter thick, arranged in nested shells to focus the x-rays in glancing reflections.

One target will be supermassive black holes in the early universe. They are a puzzle because they could not have grown so big, so fast simply by gobbling the star-size black holes they are thought to dine on. Seeing the gas being sucked into them may yield clues to the puzzle. Lynx would also capture stellar winds, supernovae, and the energetic jets that expel hot gases from galaxies, quenching their star formation. "We will unlock the secrets of galaxy evolution," says project co-chair Alexey Vikhlinin of the Smithsonian Astrophysical Observatory in Cambridge, Massachusetts.

U.S. x-ray astronomers have been unlucky in recent years. They built novel instruments for three Japanese x-ray satellites that failed. And in 2012, NASA pulled out of the International X-ray Observatory, a joint effort with Europe and Japan that became Europe's Athena mission, planned for launch in 2028. But the Lynx team thinks it has a compelling case. "Black holes are very easy for people to understand, and we have a unique way to see them," Vikhlinin says.



CREDITS: (GRAPHIC) C. BICKEL/*SCIENCE*; (DATA) M. BAYSINGER/NASA



GRAPHIC: C. BICKEL/*SCIENCE*

Seeking the light of Earth-like worlds

The Habitable Exoplanet Observatory (HabEx) would look for signs of life light-years away. Although thousands of exoplanets have been discovered indirectly, only a few large ones have emerged shyly from the glare of their star for a snapshot. No current telescope can capture the faint light of small rocky worlds like our own, let alone tease it apart for signs of oxygen, methane, and other biosignatures. "We want to design [HabEx] from the ground up to image Earth-sized planets," says team co-leader Scott Gaudi of Ohio State University in Columbus.

HabEx's monolithic 4-meter mirror is designed to work in concert with a starshade, a flower-shaped mask 72 meters across, which would float 124,000 kilometers away from the telescope. With the starshade blocking light from a star, HabEx could see planets around it that are one ten-billionth as bright. "These are potentially the faintest objects ever studied with telescopes," says team member Chris Stark of the Space Telescope Science Institute in Baltimore, Maryland. HabEx will also have a coronagraph, a complex internal device that blocks starlight, but less effectively than the starshade and over a narrower range of wavelengths.

Using just the coronagraph, HabEx would survey about 50 nearby planetary systems, identifying promising Earth-like planets. Then the fuel-hungry starshade would maneuver into place for observations of about 10 systems that host exo-Earths. No starshade has ever flown. But Gaudi says HabEx is still a cheaper, safer choice than its primary competitor, the giant Large UV Optical Infrared Surveyor. "HabEx is the least risky telescope to do this," he says.

With report after report arguing for the importance of finding life on an Earth-like planet—as well as public and congressional support for the quest—the team believes it has momentum. "It's a goal for many astronomers: the ultimate answer to the question, are we alone?" Stark says.

INSIGHTS

POLICY FORUM

GOVERNANCE

Tragedy revisited

"Freedom in a commons brings ruin to all." So argued ecologist Garrett Hardin in "The Tragedy of the Commons" in the 13 December 1968 issue of *Science* (*1*). Hardin questioned society's ability to manage shared resources and avoid an environmentally and socially calamitous free-for-all. In the 50 years since, the essay has influenced discussions ranging from climate change (see page 1217) to evolution, from infectious disease to the internet, and has reached far beyond academic literature—but not without criticism. Considerable work, notably by Nobelist Elinor Ostrom (*2*), has challenged Hardin, particularly his emphasis on property rights and government regulatory leviathans as solutions. Instead, research has documented contexts, cases, and principles that reflect the ability of groups to collectively govern common resources. To mark this anniversary and celebrate the richness of research and practice around commons and cooperation, *Science* invited experts to share some contemporary views on such tragedies and how to avert them. —**Brad Wible**

Collective actions, cultural norms

By **Robert Boyd**[1] *and* **Peter J. Richerson**[2]

The enduring influence of Hardin's essay testifies to the power of a clear argument. Should a selfish herdsman add animals to his flock? The benefit of additional animals flows to the herdsman, while the costs are spread among all who share the commons. Each herdsman decides to add animals, and the commons is over-grazed. Genes or ideas that encourage selflessness will be out-reproduced by those that encourage selfishness, so collective action problems can only be solved with coercive institutions such as police and courts.

This argument is clear and powerful, but wrong. Many village-scale human societies have organized hundreds of people to produce irrigation works and military action and solve commons problems, regulated not by formal coercive institutions but by informal, culturally evolved moral norms. Much evidence suggests that the propensity to be guided by culturally transmitted beliefs is a powerful adaptive tool that has been favored by natural selection (*3*). People in every human society acquire moral beliefs about what sorts of behaviors are right and wrong, and these beliefs can support solutions to collective action problems. For example, in the Turkana, an East African pastoral group, hundreds of warriors cooperate in cattle raids against other ethnic groups. The Turkana have no police, courts, or other formal coercive institutions, but cowards and deserters, tempted by selfish motives to free-ride, are punished by members of the community (*4*). Because norm violators suffer costs, those who adhere to the local norms do better than those who don't. Adherence to norms is self-interested, so genes and ideas that undermine successful norms do not spread.

This means that once they are established, very different norms can persist, even in similar environments. To understand why norms sometimes support collective action and sometimes don't, we need to understand the processes that shape norm content.

PHOTO: ALECTRUSLER2015/SHUTTERSTOCK

As R. Boyd and P. J. Richerson point out, "The enduring influence of Hardin's essay testifies to the power of a clear argument. Should a selfish herdsman add animals to his flock?"

Competition among culturally different groups is one such mechanism: Groups with norms that lead to economic success attract imitators, and norms that lead to military success spread through conquest (*5*). As societies become larger and more complex, political institutions play a major role in determining norm content and creating supporting formal institutions. However, there are many examples of norm shifts that cannot be explained as a consequence of group competition or deliberate political choices, such as the disappearance of norms supporting dueling in 19th-century Britain and shifts in norms regarding tobacco smoking, premarital sex, and same-sex marriage during the 20th century.

Although historians provide plausible narratives for particular norm shifts (*6*), plausible quantitative theory is scarce. Models based on drift-like random fluctuations make clear predictions but seem too slow to account for change in larger societies (*7*), whereas those based on self-reinforcing cascades (*8*) are fast but depend on an improbable balancing of processes. We think that developing such a theory is crucial for understanding human cooperation. Darwin argued in *The Descent of Man* that selection for cooperation in ancient tribes, acting over the long run, favored prosocial emotions such as sympathy and patriotism. These emotions, coupled with "approbation of our fellow men," contributed to changes in norms, which in turn supported legal initiatives such as the end of slavery in the British Empire in 1833. We have argued for a modern version of his idea (*3*, *5*).

When societies are small, and collective action problems are local, group beneficial norms often spread. The most difficult problems are those such as climate change that spill over into many different societies and require people from societies that share few norms or political institutions to create new norms. On the time scale of a century, progress in solving global commons problems has been impressive. It is not clear that for some problems we have another century to spare.

TOMORROW'S EARTH

Read more articles online at scim.ag/TomorrowsEarth

Playing games in a common pool

By **Ruth Meinzen-Dick**[3]

Water is a classic common pool resource: What one person consumes is not available for others, and water's mobility makes it costly to exclude other users. But classic studies of irrigation institutions (*9*) showing that people can and do cooperate to sustainably manage water have been instrumental in refuting the notion of an inevitable tragedy of the commons (*2*). Yet cooperation does not always emerge or survive, particularly in large irrigation systems built and managed by government agencies. Community organizers have been able to strengthen irrigation institutions, but this is generally time- and labor-intensive and difficult to scale up. Millions of dollars have been invested in large-scale programs to introduce, formalize, or strengthen water users' associations, but success in such programs has been limited (*10*). Groundwater is particularly problematic because it is a mostly invisible resource and it is difficult to understand the boundaries of the aquifers and how one person's use affects others.

What then can increase collective action over water? A strong tradition of interdisciplinary and transdisciplinary research brings together social sciences with irrigation engineering and hydrology, using case studies and comparative studies (*2*, *10*). Elinor Ostrom identified design principles underlying effective governance of common resources: clearly defined boundaries, rules adapted to local needs, with users' participation and respected by outsiders, monitoring, graduated sanctions, dispute resolution, and nested layers of governance that fit the resource system (*2*). In addition to these, water scarcity, type of infrastructure, market integration, and social ties among users can all affect cooperation over water. For example, when many farmers in India get wells and no longer depend on surface irrigation for all their water, they stop contributing to the irrigation organizations. Or those at the head end

ILLUSTRATION: ADAM SIMPSON/HEART AGENCY

of canals, who get water first, may take too much unless they also depend on the tail enders for other things, such as contributions to maintain the whole system.

Behavioral experiments, originally designed as games simulating commons dilemmas in the laboratories, have been adapted to be played with real commoners in the field. These games have shown the importance of communication, repeated interactions, information, and perceived fairness of the distribution of costs and benefits in influencing collective action. We are testing whether these games could be adapted from a research instrument to a tool that can also help water users understand the trade-offs and potential value of cooperation. In our groundwater game, players choose between crops with different water consumption and profitability and see the simulated effects on aquifer sustainability, showing that short-term profits by some come at long-term costs borne by all. In India, sites where this game was played were significantly more likely to adopt rules governing groundwater use, compared with control communities (*11*).

At a larger scale, multistakeholder participatory processes can sometimes create common understanding and consensus about opportunities for improving the complex governance of multiple water uses and users in river basins, including water quality improvement and reservoir reoperation for restoring more natural flow regimes in rivers (*12*). Ostrom's concepts of polycentric governance (*4*) and the rich literature on multistakeholder platforms and comanagement arrangements between the state and communities (*10*) provide insights—though not blueprints—for ways to better manage water commons in the future. Payment for environmental services financed by downstream users such as municipal water systems can encourage upstream conservation, such as seen in the Delaware County watershed that feeds New York City, but building trust between government agencies and different types of water users is key.

Revealing historical resilience

By **Tine De Moor**[4]

The practice of managing and using land and other natural resources in common—what the term "commons" originally referred to—has a long history. "Commoners" exercised rights to use resources over large expanses of permanently uncultivated, or only temporarily cultivated, open country such as heathland, rough pasture, or woodland. Commons were an essential component of early modern agriculture in many parts of Europe until the 19th century; their disappearance (through enclosures) was a key political issue at the time and has been the subject of considerable historiographical debate since. Historians, whose work on commons was for a long time mainly descriptive, have provided evidence that—contrary to Hardin's assumption—historical commons were dynamic institutions, with continuous rule-making, changing, intensive communication between the commoners and with effective monitoring mechanisms (*13*). Contrary to arguments in favor of their dissolution, common resources were used in an efficient manner, and improvements associated with enclosing common land and limiting access to commoners were probably not as large as originally thought by reformers (*14*).

A more analytic approach to commons' history, using archival records for many commons dating back to medieval times (in Europe), can provide insights about what makes a self-governing institution resilient for major crises and external shocks. After all, true resilience can take multiple generations and even centuries to surface. Historical sources are often still available, in the form of extensive written rulebooks, in many cases for commons with a lifetime of several hundreds of years during which rules changed frequently (*15*, *16*). The reconstruction of these rules demonstrates that regulation often adapted to changing circumstances, and that survival over many centuries was not an exception, but the norm. Those rule books provide essentially the same type of data as collected through fieldwork by Ostrom and colleagues (*2*), but whereas Ostrom's list of design principles is the common denominator of a large set of commons studied at a specific moment in time, the historical data allow for a longitudinal study of the temporal dynamics of a common, of governance that needed to adapt or else collapse. An ongoing study of large datasets of 30 historical commons across the Netherlands, Spain, and the United Kingdom (15) is suggesting some ways in which Ostrom's list, and work building on it, may need to be updated. For example, sanctioning—in particular, graduated sanction, incrementally based on the repetition of violations—has been seen as an essential component to make self-governing commons work, yet graduated sanctioning is hardly ever found in commons surviving more than 200 years (the minimum years of survival as set in the study) (*17*). This suggests that in order to achieve long-term survival, this particular type of sanctioning may have been less essential than suggested in Ostrom's principles, and that those commons with graduated sanctioning in Ostrom's database may have been through a severe period, with many trials and errors of sanctioning, with the graduated version as the very last resort. Futhermore, analyzing rules and sanctions over the lifetime of several commons, there appears to be an inverse correlation between the effort put into developing sanctions (expressed as the number of rules accompanied by a sanction) and the longevity of a common (expressed as the number of years between emergence and dissolution), suggesting that commons that managed to survive longest invested least in designing and applying sanctions (*18*). This counterintuitive result may be explained by the longer-lasting commons investing more time and effort in (compulsory) commoners' meetings, leading to a more thorough understanding by commoners of why rules—and changes thereof—were necessary, and possibly, as a consequence, leading to less free-riding. Historical analysis can add unexpected insights to our understanding of which methods can be used to keep commons functioning in the long run, steering them away from a tragedy.

Couple issues to address conflict

By **Matthew O. Jackson**[5,6,7]

Over the past five decades, we have come to a deep understanding of commons problems and how to solve them: They are not zero-sum games, but instead offer substantial gains from cooperation. Game theory and market design have helped us understand how to provide appropriate incentives (*19*–*21*). For instance, taxes as well as cap-and-trade systems can be designed to make the price of emitting carbon include its ultimate social/climate cost, and subsidies can make the prices of alternative technologies reflect their ultimate social benefit. However, a challenge with global commons problems is that solving the incentive problems often leads the collective gains to be distributed very unevenly (*22*); the costs can even outweigh the benefits for some parties. There are many players with enormous differences in wealth and interests around the planet—both within and across countries—facing different consequences from commons problems and abilities to pay for them. Yet, universal cooperation is needed, including coordinated limits and the willingness and the ability to enforce those limits. Thus, the main challenges that we face are political. Crafting a policy that addresses everyone's needs becomes an even bigger challenge when combined with constantly changing political leadership with short-term perspectives and impatient citizens who make it difficult to incur large costs today for

Uncoordinated management of fishing, shipping, and seabed mining challenges the health, productivity, and resilience of the global ocean commons.

PHOTO: BRUSINI AURÉLIEN/HEMIS.FR/HEMIS/ALAMY STOCK PHOTO

benefits that may not accrue for decades and involve considerable uncertainty and may affect others more than themselves. A natural reaction to this is to try to simplify things by concentrating on one issue at a time. Although this may seem sensible at first blush, the key to crafting policies that address a multitude of conflicting interests is actually to couple issues together (*23*). If there is an issue on which a group has little to gain and much to lose, then one gets their consent by including some other issue on which they have much to gain and little to lose. This is a principle underlying omnibus legislation: the packaging of unrelated issues into one large bill (*24*). Global organizations such as the United Nations have wide scope and can envision such compromise, but they are funded at a handful of billions of dollars when tens of trillions are at stake, and they lack full international buy-in and trust. The exception is the World Trade Organization (WTO); more than half of world gross domestic product crosses country borders. However, the WTO's scope is limited to trade agreements. In the absence of a world organization with sufficient jurisdiction and large enough carrots and sticks, there is a need for the leadership of key countries to step up and craft an omnibus agreement that couples commons problems with other issues, with something for everyone. Packaging issues produces an attractive agreement that entices participation, rather than coercing it by threatening nonparticipants with trade sanctions that may run afoul of existing treaties, fuel a trade war, or be costly to follow through with. Coupling global commons problems with other large issues will complicate our lives, but it is the only way to forge and enforce agreements at an appropriate scale, which everyone will sign onto. Without powerful international leadership, large global commons problems will continue to be ceded to humanitarian organizations and the voluntary behaviors of groups here and there.

An ocean of opportunity

By **Kristina M. Gjerde**[8] *and* **Harriet Harden-Davies**[9]

In many ways, the global ocean beyond national boundaries—two-thirds of the ocean's surface—epitomises the tragedy of the commons. Access remains difficult to control, resources are declining, and pollution pervades the deepest abyss (*25*). Combined with ocean warming, deoxygenation, and acidification, these impacts undermine ocean health, productivity, and resilience, exacerbating the challenge of achieving equitable and sustainable management of our shared ocean (*26*).

Since Hardin in 1968, the concept of the global ocean commons has evolved. The 1982 United Nations Convention on the Law of the Sea (UNCLOS) tempered the right of States to access resources of the high seas and international seabed ("the Area") with obligations to build capacity, advance scientific knowledge, and protect the environment. UNCLOS further designated the Area and its mineral resources as the "common heritage of mankind" to be managed by the International Seabed Authority for "the benefit of mankind as a whole." In the 1990s, States acknowledged that biodiversity loss and climate change were "common concerns" (*27*). More recently, concepts such as precaution, ecosystem-based approaches, and marine protected areas (MPAs) have been incorporated into international commitments (*27*), including United Nations (UN) Sustainable Development Goal 14.

However, global ocean health remains under threat because mechanisms to enable and enforce existing UNCLOS obligations remain weak (*25*). Despite new technologies to monitor activities and impacts (*28*), the current system of managing fishing, shipping, and seabed mining separately begets inconsistent, conflicting, and frequently unsustainable results (*25*). For example, illegal fishing is worse in some places than others; mineral exploration rights are being granted atop important fishing, scientific research, and cable sites; and biodiversity values are frequently ignored (*25*). Meanwhile, the lack of centralized reporting hinders efforts to hold accountable the few that block conservation measures despite treaty requirements (*27*, *29*) and compelling evidence of need (*26*). In the Southern Ocean, for instance, compromises made to secure consensus for the Ross Sea MPA (*29*) highlighted the power of a very few states to weaken protections.

Conversely, on the rare occasions that the UN has called on sectoral bodies to implement specific requirements to tackle threats to biodiversity, substantial progress has been made. A 2006 UN resolution requiring states sponsoring bottom fishing to conduct prior assessments, adopt measures to avoid substantial impacts, and crucially, report to the UN has protected vast areas of the deep seabed. However, as ocean stressors multiply, the UN has recognized the need for a more comprehensive approach to biodiversity conservation and use (*25*).

In September, the UN convened the first intergovernmental conference to negotiate a legally binding agreement under UNCLOS for conservation and sustainable use of marine biodiversity beyond national jurisdiction. The negotiations present an opportunity to elaborate and modernize existing requirements to conduct environmental impact assessments; proactively adopt conservation measures, including MPAs; avoid substantial harm to biodiversity; and improve accountability through regular reporting. The agreement can thus create rules, monitoring systems, and sanctioning powers to enhance compliance while ensuring more sustainable outcomes at the global, regional, and sectoral levels.

Science also has a major role to play as a catalyst for unifying stakeholders behind common concerns (*30*). The agreement can boost capacity and understanding by fostering collaboration in marine science, knowledge exchange, and technology transfer, including on marine genetic resources (*30*). The UN Decade of Ocean Science 2021–2030 could further facilitate knowledge advancement and collective capacity to enable informed, equitable, and sustainable management of our global ocean commons. The question is, will states adopt the mutual restraints and allocate the required resources to evade tragedy and renew ocean health? There is hope, but little time. An ambitious agreement is needed by 2020 to protect our common interest in a healthy, productive, and resilient ocean in the challenging decades to come.

Common knowledge

By **Brett M. Frischmann[10], Michael J. Madison[11], Katherine J. Strandburg[12]**

Intellectual resources have their own tragedy-of-the-commons allegory. Replace Hardin's pasture with an idea, and consider what happens when the resource, the idea, is openly accessible to all. Everyone who can profitably make use of the idea will do so, as much and as often and in whatever manner suits them. But ideas are public goods, not common pool resources; ideas are not congested or depleted by overuse. Unlike the pasture, unconstrained consumption of ideas seems good, and often it is.

But there's a catch. Ideas are products of human intellect, often requiring investment of time, effort, and capital. Unconstrained consumption by free riders, who invest little or nothing in creating the ideas, presents a risk for those who might make such investments in creating knowledge because they may struggle to recover a sufficient return on their investment. Anticipating this, they may underinvest, contributing to tragic underproduction of intellectual resources.

Avoiding cultural, technological, and scientific stagnation thus seems to require collective action to ensure adequate investment in knowledge creation. To facilitate this, many analysts assume two options: government subsidies or intellectual property-enabled markets. Though both are indeed important drivers of knowledge production, so are "knowledge commons," which we should not take for granted.

Knowledge commons refers to institutionalized community governance of the sharing and, in many cases, creation and curation of intellectual and cultural resources (*31*). Examples range from scientific research commons, including data, literature, and research materials (*32*), to intellectual property pools, entrepreneurial/user innovation commons, rare-disease clinical research consortia, open-source software projects, and Wikipedia (*31*). Understanding how such communities share and develop knowledge is crucial in today's "information society."

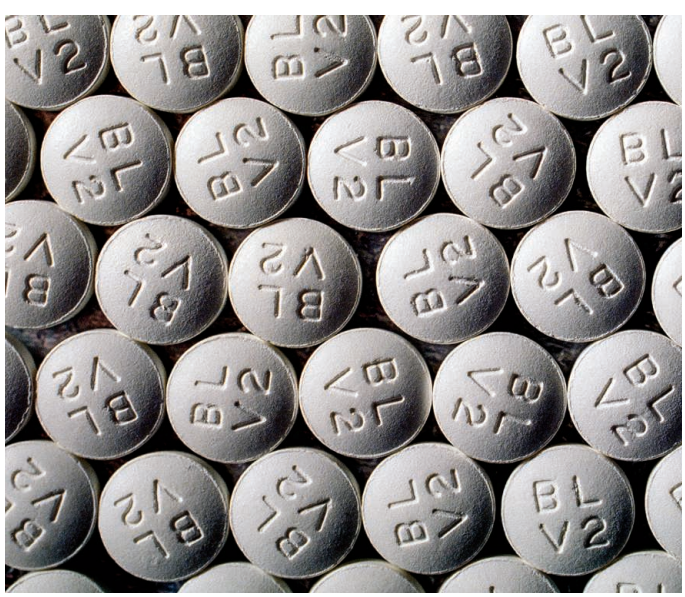
Following Ostrom (*2*, *33*) and Hess and Ostrom (*34*), we have worked to systematize the study of knowledge commons and build a new field of interdisciplinary research in which law, economics, sociology, political science, network science, and other fields converge. Dozens of case studies have begun to reveal an empirical picture of knowledge commons. A representative theme is that knowledge commons confront diverse social dilemmas not reducible to the simple free rider or tragic commons. Rare-disease research consortia, for example, address numerous governance challenges, including allocating research funding, authorship credit, and other rivalrous resources; overcoming potential anticommons dilemmas arising from researchers' incentives to hoard access to patients and their data; maintaining privacy, security, and the trust of patients and their families; reducing transaction costs of cooperation between widely dispersed researchers; and managing interactions with outsiders, such as pharmaceutical companies. The diversity of dilemmas is matched by the surprising diversity of participants critical to successful collaboration. Hardin's sheep-herder must be replaced with researchers, clinicians, patients, site coordinators, funders, third-party data custodians, and even government officials.

Despite growing evidence, we're still far from design principles, much less strong prescriptions. Yet social demand for trusted governance of shared knowledge resources, ranging from medical data (*35*) to algorithmically generated intelligence, is growing, even as public trust in governments and markets as sources of governance seems tenuous. Many researchers and policymakers understood the scope of Ostrom's commons-based framework as limited, for example, to small communities managing local resources. Now, more than ever, we need to explore if, when, and how commons governance can scale.

The antimicrobial commons

By Angela R McLean[13,14] and Christopher Dye[13]

It has become commonplace (*36*–*38*) to refer to the rise of antimicrobial resistance (AMR) as a tragedy of the commons. Each individual wishes to use the common-pool resource of functioning antimicrobials whenever they might have a beneficial effect



Antimicrobial use could be decreased if overuse led to loss of good reputation, and rules for prescribing established boundaries of "reputable" behaviors.

PHOTO: RICHARD PASLEY - DOCTOR STOCK/GETTY IMAGES

(whether in treating human illness or in raising livestock), but overuse accelerates the spread of drug-resistant pathogens, so the drugs become useless to all—and therein lies the tragedy. One way or another, some individual freedoms must be sacrificed in order to maintain a valuable resource for the common good. Whereas Hardin emphasized private or state ownership to achieve this, Ostrom argued that those who share in exploiting a common-pool resource can develop their own rules to prevent its overuse. She identified factors that are conducive to the establishment of effective institutions to regulate the exploitation of a resource: Users have common interests; they place a high value on the resource far into the future; users support effective monitoring; accurate information is valued and easily communicated; and it is feasible to establish binding and enforceable regulations. Ostrom warned that large groups often struggle to govern common pool resources and that boundary rules are needed to determine rights and responsibilities.

Many of Ostrom's observations are starting to be fulfilled in the search for solutions to the problems of AMR, even if few people in this area explicitly set out to apply her work. The growing threat of AMR is increasingly understood by medical professionals, policy professionals, and the public alike. The associated discourse reflects the common, long-term interests of these diverse users (*39*). The widely accepted need for better surveillance of AMR signals rising support for effective monitoring and accurate, shared information. In a growing search for effective rules, physicians are adhering more strictly to evidence-based guidance for diagnosing infections; for infection control in hospitals; for procuring, prescribing and dispensing antimicrobials; and for ensuring that patients complete treatments. Beyond codes of practice, governments have in some settings introduced methods of enforcement, such as restricting the use of essential drugs to certified treatment centres. And public health specialists have called for AMR to be included among the International Health Regulations, a legally binding agreement to prevent the international spread of disease. Last, the global nature of the challenge is acknowledged in the World Health Organization's leadership in developing new norms for using existing antimicrobials and investing in new ones (*40*).

Some other useful ideas arise when AMR is viewed as a tragedy of the commons. For example, a desire not to be seen as selfish offers a potential solution: antimicrobial use could be decreased if overuse led to loss of good reputation, and rules for appropriate prescribing helped establish boundaries of "reputable" behaviors (*41*). Further, the "large groups" problem may be less acute if local effects are strong enough that a region or nation can benefit from reducing their own usage, even if their neighbors do not (*42*).

In 1968, Hardin remarked that the tragedy of the commons was understood mostly as a set of special cases rather than as a general problem of resource management. The AMR tragedy will benefit from the application of the broad principles of governing a wide range of common pool resources. That will bring focus, for example, to the question of "boundary rules". Can one country ever manage AMR alone, and can AMR for human infections be controlled without also controlling agricultural use? Also uncertain is the best mechanism of control: When are binding and enforceable regulations preferred over guidelines and codes of practice? How can the principles laid out by Hardin and Ostrom guide the creation of new resources (discovery of antimicrobials), besides conserving the ones we already have? In the face of these pressing questions, taking a broader view of the AMR tragedy, and of its resolution, will show how best to govern the antimicrobial commons.

REFERENCES AND NOTES

1. G. Hardin, *Science* **162**, 1243 (1968).
2. E. Ostrom, *Governing the Commons. The Evolution of Institutions for Collective Action* (Cambridge Univ. Press, 1990).
3. P. J. Richerson, R. Boyd, *Not by Genes Alone* (Univ. of Chicago Press, 2005).
4. S. Mathew, R. Boyd, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 11375 (2011).
5. P. Richerson *et al.*, *Behav. Brain Sci.* **39**, e30 (2016).
6. K. Appiah, *The Honor Code, How Moral Revolutions Happen* (Norton, 2010).
7. H. P. Young, *Annu. Rev. Econ.* **7**, 359 (2015).
8. T. Kuran, 1995. Private Truths, Public Lies, Harvard University Press, Cambridge
9. E. W. Coward Jr., Ed., *Irrigation and Agricultural Development in Asia: Perspectives from the Social Sciences* (Cornell Univ. Press, Ithaca, NY, 1980).
10. R. Meinzen-Dick, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15200 (2007).
11. R. Meinzen-Dick *et al.*, *World Dev.* **107**, 40 (2018).
12. P. A. Sabatier *et al.*, *Swimming Upstream: Collaborative Approaches to Watershed Management* (MIT Press, Cambridge, MA, 2005).
13. P. Warde, M. De Moor, L. Shaw-Taylor, in *The Management of Common Land in North West Europe, c. 1500-1850* (CORN publication series no. 8), M. De Moor *et al.*, Eds. (Brepols, 2002), pp. 15–32.
14. G. Clark, *J. Econ. Hist.* **58**, 73 (1998).
15. T. De Moor, M. Laborda-Pemán, J. M. Lana-Berasain, R. Van Weeren, A. Winchester, *Int. J. Commons* **10**, 529 (2016).
16. M. Casari, C. R. Plott, *J. Econ. Behav. Organ.* **51**, 217 (2003).
17. R. Van Weeren, T. De Moor, *Agric. Hist. Rev.* **II**, 256 (2014).
18. T. De Moor, A. Tukker, in *Jahrbuch für Geschichte des ländlichen Raumes* (Rural History Yearbook), N. Grüne, J. Hübner, G. Siegl, Eds. (Studien Verlag, 2015), pp. 175–206.
19. B. Harstad, *Rev. Econ. Stud.* **79**, 1527 (2012).
20. M. Golosov, J. Hassler, P. Krusell, A. Tsyvinski, *Econometrica* **82**, 41 (2014).
21. W. D. Nordhaus, *Am. Econ. Rev.* **105**, 1339 (2015).
22. P. K. Dutta, R. Radner, *J. Econ. Behav. Organ.* **71**, 187 (2009).
23. M. O. Jackson, H. F. Sonnenschein, *Econometrica* **75**, 241 (2007).
24. G. S. Krutz, *Hitching a Ride: Omnibus Legislating in the US Congress* (Ohio State Univ. Press, 2001).
25. G. Wright, J. Rochette, K. M. Gjerde, L. A. Levin, *Nature* **554**, 163 (2018).
26. D. J. McCauley *et al.*, *Science* **347**, 1255641 (2015).
27. K. M. Gjerde *et al.*, *Aquat. Conserv.* **26** (suppl. 2), 45 (2016).
28. D. C. Dunn *et al.*, *Fish Fish.* **19**, 729 (2018).
29. C. M. Brooks *et al.*, *Science* **354**, 185 (2016).
30. H. Harden-Davies, *ICES J. Mar. Sci.* **75**, 426 (2017).
31. B. M. Frischmann, M. J. Madison, K. J. Strandburg, Eds., *Governing Knowledge Commons* (Oxford Univ. Press, 2014).
32. J. H. Reichman, P. F. Uhlir, T. Dedeurwaerdere, *Governing Digitally Integrated Genetic Resources, Data, and Literature: Globally Intellectual Property Strategies for a Redesigned Microbial Research Commons* (Cambridge Univ. Press, 2016).
33. E. Ostrom, *Understanding Institutional Diversity* (Princeton Univ. Press, 2005).
34. C. Hess, E. Ostrom, Eds., *Understanding Knowledge as a Commons: From Theory to Practice* (MIT Press, 2007).
35. K. J. Strandburg, B. M. Frischmann, M. J. Madison, Eds., *Governing Medical Knowledge Commons* (Cambridge Univ. Press 2017).
36. B. R. Levin, *Clin. Infect. Dis.* **33**, S161 (2001).
37. K. S. O'Brien *et al.*, *Comput. Math. Methods Med.* **2014**, 837929 (2014).
38. A. Hollis, P. Maybarduk, *J. Law Med. Ethics* **43**, 33 (2015).
39. R. Laxminarayan *et al.*, *Lancet Infect. Dis.* **13**, 1057 (2013).
40. WHO, Global Action Plan on Antimicrobial Resistance (2015); www.wpro.who.int/entity/drug_resistance/resources/global_action_plan_eng.pdf.
41. F. Baquero, J. Campos, *Rev. Esp. Quimioter.* **16**, 11 (2003).
42. K. R. Foster, H. Grundmann, *PLOS Med.* **3**, e29 (2006).

ACKNOWLEDGMENTS

Funding for R.M.-D. was provided by the CGIAR Research Program on Policies, Institutions, and Markets.

10.1126/science.aaw0911

[1]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA; robert.t.boyd@gmail.com. [2]Department of Environmental Science and Policy, University of California, Davis, CA, USA; pjricherson@ucdavis.edu. [3]International Food Policy Research Institute, Washington, DC, USA; r.meinzen-dick@cgiar.org. [4]Utrecht University, Utrecht, Netherlands; t.demoor@uu.nl. [5]Stanford University, Stanford, CA, USA; jacksonm@stanford.edu. [6]Sante Fe Insitute, Santa Fe, NM, USA. [7]Canadian Institute For Advanced Research, Toronto, ON, Canada. [8]IUCN Global Marine and Polar Programme and World Commission on Protected Areas, Cambridge, MA, USA; kgjerde@eip.com.pl. [9]Australian National Centre for Ocean Resources and Security (ANCORS), University of Wollongong, NSW, Australia. [10]Villanova University, Charles Widger School of Law, Villanova, PA, USA; brett.frischmann@law.villanova.edu. [11]University of Pittsburgh, School of Law, Pittsburgh, PA, USA; madison@pitt.edu. [12]New York University School of Law. New York, NY, USA; katherine.strandburg@nyu.edu. [13]All Souls College, Oxford University, Oxford OX1 4AL, UK; angela.mclean@zoo.ox.ac.uk, chrisdye56@gmail.com. [14]Oxford Martin Programme on Collective Responsibility for Infectious Disease, Oxford University, Oxford OX1 3BD, UK.

PERSPECTIVES

ESSAY

Reimagining the human

A human-centric worldview is blinding humanity to the consequences of our actions

By **Eileen Crist**

CREDITS: (LEFT TO RIGHT) ADAM SIMPSON/HEART AGENCY; IVAN OYARZUN/GETTY IMAGES

Earth is in the throes of a mass extinction event and climate change upheaval, risking a planetary shift into conditions that will be extremely challenging, if not catastrophic, for complex life (*1*). Although responsibility for the present trajectory is unevenly distributed, the overarching drivers are rapid increases in (i) human population, (ii) consumption of food, water, energy, and materials, and (iii) infrastructural incursions into the natural world. As the "trends of more" on all these fronts continue to swell, the ecological crisis is intensifying (*2*–*4*). Given that human expansionism is causing mass extinction of nonhuman life and threatening both ecological and societal stability, why is humanity not steering toward limiting and reversing its expansionism?

TOMORROW'S EARTH
Read more articles online at scim.ag/TomorrowsEarth

The rational response to the present-day ecological emergency would be to pursue actions that will downscale the human factor and contract our presence in the realm of nature. Yet in mainstream institutional arenas, economic, demographic, and infrastructural growth are framed as inevitable, while technological and management solutions to adverse impacts are pursued single-mindedly. Although pursuing such solutions is important, it is also clear that reducing humanity's scale and scope in the ecosphere is the surest approach to arresting the extinction crisis, moderating climate change, decreasing pollution, and providing sorely needed leeway to tackle problems of poverty, food insecurity, and forced migration (*5*). The question that arises is why the approach of contracting the human enterprise tends to be ignored.

The answer lies in the deeper cause of the ecological crisis: a pervasive worldview that imbues the trends of more with a cachet of inevitability and legitimacy. This worldview esteems the human as a distinguished entity that is superior to all other life forms and is entitled to use them and the places they live. The belief system of superiority and entitlement—or human supremacy—manifests in a range of anthropocentric commonplace assumptions, linguistic constructs, institutional regimes, and everyday actions of individual, group, nation-state, and corporate actors (*6*). For example, the human is invested with powers of life and death over all other beings and with the prerogative to control and manage all geographical space. The all-encompassing manifestation of the belief system of human supremacy is precisely what constitutes it as a worldview.

This worldview is not necessarily an explicitly articulated narrative. Rather, it forms the tacit postulate from which people source meaning and justification to disregard virtually any limitation of action or way of life in the ecosphere and toward nonhumans. Human supremacy is the underlying big story that normalizes the trends of more, and the consequent displacements and exterminations of nonhumans—as well as of humans who oppose that worldview (*7*, *8*). In this context, it is crucial to recognize that human supremacy is neither culturally nor individually universal, nor is it derived in any straightforward way from human nature. However, western civilization has elaborated its most forceful, long-standing expression, and through the West's ascendancy the influence of this worldview has spread across the globe (*9*).

BLIND TO THE WISDOM OF LIMITATIONS

The planetwide sense of entitlement bequeathed by a supremacist worldview blinds the human collective to the wisdom of limitations in several ways, thereby hindering efforts to address the ecological crisis by downscaling the human enterprise and withdrawing it from large portions of land and sea.

First, because the worldview demotes the nonhuman in favor of the human, it blocks the human mind from recognizing the intrinsic existence and value of nonhumans and their habitats. Nonhumans are rendered as resources and considered dispensable or killable; it is assumed that natural areas can be taken over and converted at will.

Second, a worldview founded on the elevation of the human impairs the experience of awe for this living planet, inducing instead the perception that viewing the ecosphere as a container of natural resources, raw materials, and goods and services makes sense. If humanity inhabited Earth with a profound sense of awe, news of an impending mass extinction would galvanize the world into action. Instead, what we find is that the response to anthropogenic mass extinction is muted in mainstream media and other social arenas.

Third, based on the conviction of the special distinction of the human, the worldview fosters the belief that humans are resourceful, intelligent, and resilient enough to face any challenges that may come. This tacit missive bolsters societal torpor and

Department of Science, Technology and Society, Virginia Tech, Blacksburg, VA 24061, USA. Email: ecrist@vt.edu

Rising human consumption is driving widespread destruction of natural systems, such as this forest in British Columbia.

political inaction, because it is widely assumed that technological innovations and interventions will overcome problems.

Fourth, the worldview impedes humans from recoiling from, or even seeing, the violence of an expansionism that fuels extinctions, population plunges, mass mortality events, and starvations of nonhumans. Because these experiences are happening to "the merely living," they are nonissues for mainstream media and the political sphere, which are focused almost exclusively on human affairs. For example, humanity's impact has become so pervasive that migratory animal species are in decline and the very phenomenon of migration is disappearing around the world. Yet neither the loss of animal migrations nor the suffering of the animals involved seem to be matters of concern in public arenas.

Lastly, the supremacist worldview insinuates that embracing limitations is unbefitting of human distinction. Whether openly or implicitly, limitations are resisted as oppressive and unworthy of humanity's stature.

By operating on all these levels, the worldview of human distinction-and-prerogative obstructs the capacity to question human hegemony for the sake of Earth's inherent splendor and in the service of a high-quality human life within a downsized, equitable global civilization nested in an all-species commonwealth. Instead, the trends of more—on the population, consumption, and infrastructure fronts—are left to persist their course seemingly unassailable.

TOWARD SCALING DOWN AND PULLING BACK

The reigning human-nature hierarchical worldview thus hinders the recognition that scaling down and pulling back is the most farsighted path forward. Scaling down involves reducing the overall amount of food, water, energy, and materials that humanity consumes and making certain shifts in what food, energy, and materials are used. This quantitative and qualitative change can be achieved by actions that can lower the global population within a human-rights framework, shrink animal agriculture, phase out fossil fuels, and transform an extractionist, overproducing, throwaway, and polluting economy into a recycling, less busy, thrifty, more ecologically benign economy (*10*–*12*). These shifts must align with a new ethos in civil society toward shared norms of mindfulness around dietary choices, avoidance of waste, conservation of energy, and reuse and recycling of materials.

Scaling down can be complemented with substantially pulling back our presence from the natural world. Achieving continental-scale protection of terrestrial and marine habitats will enable sharing Earth generously with all its life forms (*13*). Recent research reveals that large-scale nature conservation is also a powerful counter to climate change by absorbing a sizable portion of the carbon dioxide of the industrial age and preventing additional carbon (stored in the ecosphere) from being released (*14*, *15*). Vastly expanding marine protected areas will support the resurgence of marine life. Ambitious forest, grasslands, freshwater ecologies, and wetlands protection and restoration will prevent extinctions and preempt an anthropogenic mass extinction event. A robust global network of green and blue protected areas will save wildlife populations and animal migrations from their current downward spirals. Preserving the night sky in extensive swathes of wild nature will keep an open portal into the cosmos we inhabit.

Many of the global approaches called for in this pivotal moment may lack the glamor of technological and engineering breakthroughs, but they promise far-reaching strides in resolving the ecological crisis and preventing human and nonhuman suffering. Paramount examples include state-of-the-art family planning services for all (including modern contraceptive technologies), universal education from the age of 4 to 17 or 18, substantial reduction of animal-product consumption, adoption of the reduce-reuse-recycle paradigm as an everyday norm, massive protection of wild nature, and adoption of sustainable and ethical food production practices on land and sea.

BEYOND HUMAN DOMINANCE

The dominant framework of technofixes, technological schemes, and fine-tuning efficiencies is by itself no match for the tidal wave of human expansionism expected in this century. Looming before us is the imminent escalation of food, energy, materials, and commodities production, and resulting increases in wildlands destruction, species

extinctions, wildlife extirpations, freshwater appropriation, ocean degradation, extractionist operations, and the production of industrial, pesticide, nitrogen, manure, plastic, and other waste—all unfolding amid climate-change ordeals.

In the face of this juggernaut, a singular focus on a techno-managerial portfolio seems fueled by a source other than pragmatism alone. That portfolio—which would include such initiatives as climate geoengineering, desalination, de-extinction, and off-planet colonization—is in keeping with the social rubric of human distinction. The prevalent corpus resonates with a Promethean impulse to sustain human hegemony while avoiding the most expeditious approach to the ecological predicament—contracting humanity's scale and scope by means that will simultaneously strengthen human rights, facilitate the abolition of poverty, elevate our quality of life, counter the dangers of climate change, and preserve Earth's magnificent biodiversity.

To pursue scaling down and pulling back the human factor requires us to reimagine the human in a register that no longer identifies human greatness with dominance within the ecosphere and domination over nonhumans. The present historical time invites opening our imagination toward a new vision of humanity no longer obstructed by the worldview of human supremacy. Learning to inhabit Earth with care, grace, and proper measure promises material and spiritual abundance for all. ■

REFERENCES

1. W. Steffen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 8252 (2018).
2. G. Ceballos *et al.*, *The Annihilation of Nature: Human Extinction of Birds and Mammals* (Johns Hopkins Univ. Press, 2015).
3. T. Butler, Ed., *Overdevelopment, Overpopulation, Overshoot* (Grof, 2015).
4. W. Ripple, *BioScience* **67**, 1026 (2017).
5. P. Ehrlich, J. Harte, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14743 (2015).
6. G. Steiner, *Anthropocentrism and Its Discontents: The Moral Status of Animals in the History of Western Philosophy* (Univ. of Pittsburgh Press, 2010).
7. J. Mander, V. Tauli-Corpuz, Eds., *Paradigm Wars: Indigenous Peoples' Resistance to Globalization* (Sierra Club Books, 2006).
8. J. Watts, J. Vidal, *The Guardian*, 13 July 2017; www.theguardian.com/environment/2017/jul/13/environmental-defenders-being-killed-in-record-numbers-globally-new-research-reveals.
9. E. Crist, *Abundant Earth: Toward an Ecological Civilization* (Univ. of Chicago Press, 2018).
10. R. Dietz, D. O'Neill, *Enough Is Enough: Building a Sustainable Economy in a World of Finite Resources* (Berrett-Koehler, 2013).
11. B. Machovina *et al.*, *Sci. Total Environ.* **536**, 419 (2015).
12. R. Engelman, in *A Future Beyond Growth: Toward a Steady State Economy*, H. Washington, P. Twomey, Eds. (Routledge, 2016), pp. 32–42.
13. E. O. Wilson, *Half Earth: Our Planet's Fight for Life* (Norton, 2016).
14. B. Griscom *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11645 (2017).
15. C. M. Roberts *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6167 (2017).

10.1126/science.aau6026

3D PRINTING

Printing nanomaterials in shrinking gels

Photopatterning of reactive sites in gels enables arbitrary patterning of nanoparticles

By **Timothy E. Long** *and* **Christopher B. Williams**

The creation of nanoscale electronics, photonics, plasmonics, and mechanically robust metamaterials will benefit from nanofabrication processes that allow a designer full control in manipulating nanomaterial precursors in a programmable and volumetric manner. Despite decades of research, it remains challenging to design nanofabrication processes that can produce complex free-form three-dimensional (3D) objects at the scale of tens of nanometers. On page 1281 of this issue, Oran *et al.* (*1*) report on the photopatterning of reactive sites into water-swollen, chemically cross-linked acrylic gels for the subsequent site-specific deposition of nanomaterials and nanoparticles. After chemical and thermal dehydration, the gel scaffold holds the nanomaterials in a distinct 3D arrangement. This process, termed implosion fabrication (ImpFab) because the scaffold of the gel effectively "implodes" upon solvent removal, provides an opportunity to fabricate centimeter-scale assemblies of nanomaterials that possess multiple functionalities.

"...Oran et al. avoid any detrimental interactions of nanoparticles during exposure..."

The macroscopic dimension of a solvent-swollen gel provides sufficient molecular mobility to host efficient chemical reactions. However, the utility of a covalently cross-linked gel as a "nanomanufacturing reactor" for the creation of programmable nanomaterials has remained unrealized until now. Top-down processes such as photolithography can create structures with spatial resolutions approaching tens of nanometers (*2*), but the fundamental process methodologies limit the creation of arbitrary geometries in three dimensions.

Researchers are now implementing bottom-up nanofabrication processes that are similar to more recent efforts in additive manufacturing (often termed 3D printing), in that they can pattern materials in 3D space without a photomask (*3*). One such process, direct laser writing, is an exceptional process for the preparation of arbitrary 3D geometries (*4, 5*). Rastering femtosecond laser pulses through microscope optics into a photopolymer precursor enables selective photocuring anywhere in the material through the interaction of multiple photons to create discrete, polymerized voxels (3D pixels).

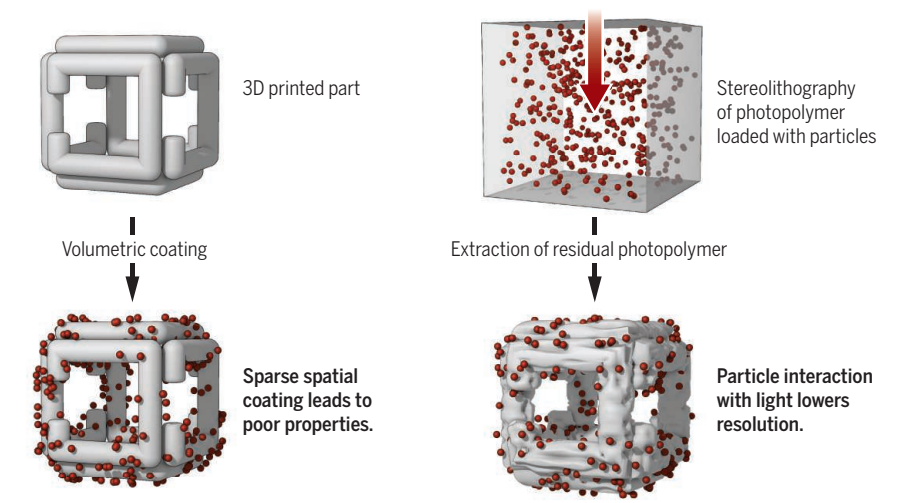
Although this technique creates 3D structures of any arbitrary geometry, its fabrication resolution is often limited by the wavelength of ultraviolet light to hundreds of nanometers (*6, 7*). Expanding the material selection for the process beyond electrical insulators has also proven challenging. Creating functional metallic materials with this process is only permitted through patterning polymer-particle nanocomposites (*8*), metal-coating the entirety of the printed surface, or multiphoton-induced reduction of metal ions. Postprocess coating does not allow for selective deposition and limits the geometries that are achievable (see the figure, right). Irradiating polymer composites and multiphoton-induced reduction of metal ions constrain resolution through refraction effects and the limited control of growth and aggregation during photoreduction, respectively (*9*).

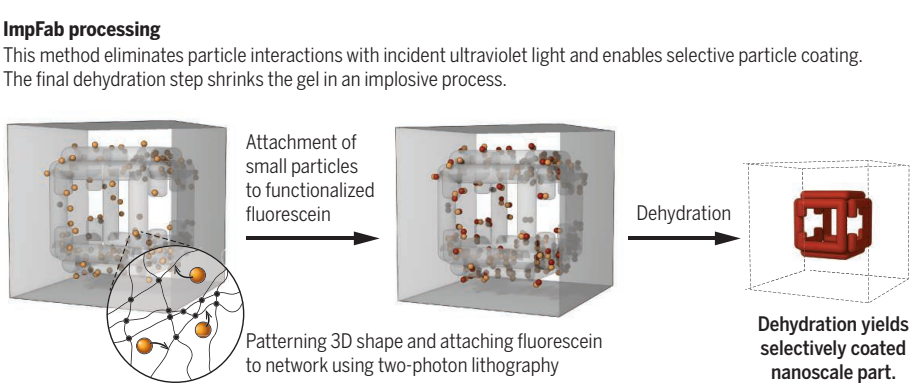
As such, fabricating truly arbitrary 3D metallic shapes at the scale of tens of nanometers has yet to be demonstrated. Researchers remain challenged to circumvent the resolution and material selection constraints imposed by direct laser writing. Oran *et al.* combined the unusual volumetric reduction properties of water-swollen gels (hydrogels) and a templating approach to fabricate complex 3D metallic nanostructures at an unprecedented scale (see the figure, left). They leveraged the stable deswelling performance of a hydrogel in the context of metallic nanofabrication. In particular, they photopatterned water-swollen gels with two-photon laser direct writing to

Department of Chemistry, Virginia Tech, Blacksburg, VA 24061, USA. Email: telong@vt.edu

Precisely placing nanomaterials

The ImpFab method of Oran *et al*. enables selective nanomaterial compositions with 3D geometries rather than simply coating 3D printed parts.





create reactive sites that enable site-specific postprocess functionalization of nanomaterials and nanoparticles. Dehydration then rapidly shrinks the fabricated structure to 1/10 its original size.

Oran *et al*. build on earlier efforts that reported the efficient reaction of fluorescein with carboxylate-containing hydrogels during two-photon excitation (*10*). Their key realization was that fluorescein derivatives also potentially serve as chaperones for the concurrent introduction of functionality and create sites for subsequent colocation of nanomaterials. This multistep segregation of defining geometry and defining material ensures that the nanomaterials are not present during the patterning step. Thus, Oran *et al*. avoid any detrimental interactions of nanoparticles during exposure that can occur in mask-projection stereolithographic printing processes (see the figure, middle). Moreover, the addition of compounds after the initial conjugation of nanomaterials can intensify the concentration of materials as well as form a spatially arranged multinanomaterial structure. Repetition of the process chain also allows the introduction of multiple nanomaterials as well as multiple patterns of nanomaterial structure.

The modularity of the methodology of Oran *et al*. for creating 3D patterns is an important aspect of their contribution. Writing into a 3D swollen gel and delivering a patterned array of functionality represent an important departure from traditional 2D and 3D lithographic printing where the patterned energy defines a printed photopolymer structure. This approach addresses a key challenge in 3D direct laser writing in terms of precisely depositing nanomaterials onto printed objects, versus the more prevalent stochastic introduction of nanoparticles that degrades both performance and printing resolution. Furthermore, two-photon laser writing allows for patterning energy with voxel-level control in 3D space, so the process can create discontinuous shapes along the hydrogel surface.

Although direct laser writing is suitable for patterning materials onto substrates and in free-form shapes, the process cannot create discontinuous multimaterial structures at the resolution Oran *et al*. demonstrated. The precise delivery of nanomaterials in multiple, complex patterns that they report enables unprecedented formation of nanomaterials of controlled geometry and high performance. The process chain effectively separates geometry definition through direct laser writing, material definition through chemical templating and sintering, and pattern resolving through gel deswelling. Separation of these steps circumvents the traditional materials, resolution, and geometric complexity constraints imposed by existing nanofabrication processes.

The work by Oran *et al*. also creates opportunities for studying the influence of the molecular architecture of the gel. Gels are complex structures that can vary in chemical composition, molecular weight between cross-link points, and dangling chain ends, and in whether they are physical versus chemical networks. These parameters in structure will influence the precise location in the gel of reactive sites for two-photon excitation and also must be considered in efforts to expand the available photoinduced chemistries in the aqueous state. Extension to other materials only depends on developing deposition chemistry that can proceed at room temperature in aqueous media. Thus, the method developed by Oran *et al*. should allow researchers to consider a myriad of new materials and reaction pathways, including other semiconductors or metals. The ability to process free-form, multimaterial nanostructures with discontinuous nanowires will enable next-generation designs of photonic, electrical, and mechanical metamaterials, as well as microelectronics, actuators, and sensors. ■

REFERENCES

1. D. Oran *et al*., *Science* **362**, 1281 (2018).
2. K. Fukuda, T. Someya, *Adv. Mater.* **29**, 1602736 (2017).
3. J. R. Tumbleston *et al*., *Science* **347**, 1349 (2015).
4. Y.-L. Zhang, Q.-D. Chen, H. Xia, H.-B. Sun, *Nano Today* **5**, 435 (2010).
5. A. Selimis, V. Mironov, M. Farsari, *Microelectron. Eng.* **132**, 83 (2015).
6. A. Marino, C. Filippeschi, V. Mattoli, B. Mazzolai, G. Ciofani, *Nanoscale* **7**, 2841 (2015).
7. S. Ali, M. L. Cuchiara, J. L. West, *Methods Cell Biol.* **121**, 105 (2014).
8. Q. Hu *et al*., *Sci. Rep.* **7**, 17150 (2017).
9. Z.-C. Ma, Y.-L. Zhang, B. Han, Q.-D. Chen, H.-B. Sun, *Small Methods* **2**, 1 (2018).
10. M. A. Skylar-Scott, M.-C. Liu, Y. Wu, A. Dixit, M. F. Yanik, *Adv. Healthcare Mater.* **5**, 1233 (2016).

ACKNOWLEDGMENTS

We thank V. Meenakshisundaram for help in preparing the figure.

10.1126/science.aav5712

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

ANTHROPOLOGY

Did maize dispersal precede domestication?

Unraveling the history of maize domesticates reveals a complex journey into South America

By **Melinda A. Zeder**

The domestication of plants and animals and their dispersal across the globe triggered a millennia-long process by which human activity has become the dominant influence on climate and the environment (*1*). Domestication was a watershed development that ushered in the Anthropocene (*2*). How, when, where, and why humans embarked on this path is central to understanding how we might chart our way in an uncertain future. On page 1309 of this issue, Kistler *et al.* (*3*) report on the dispersal of maize into and across northern South America. The study contributes to the growing appreciation of domestication as a complex, coevolutionary journey taken by humans and receptive plant and animal species over hundreds, if not thousands, of years. The study also joins others in showing how human populations incorporated dispersing domesticates into indigenous systems of exploitation and manipulation of local resources. More broadly, this research speaks to the promise of domestication research in assessing fundamental questions about evolution and the interface of natural and cultural systems that shape it.

Maize was probably "semi-domesticated" in Mexico before it dispersed into South America.

Domestication was once viewed as a binary process by which a free-living wild organism crossed a threshold to become a domesticated one under human control. It is now clear that domestication is a nonbinary process that involves complex interactions between humans and target species over long stretches of time and space (*4*). Several recent studies have combined genetic and archaeological techniques to document the progressive fixation of different domestication genes in maize over a 2000-year period as this crop plant dispersed from central Mexico into the southwestern United States (*5*, *6*). Kistler *et al.* track the dispersal of maize into South America, taking this research several steps further. Conventional wisdom had been that maize dispersed into South America well after it was fully domesticated. In a major departure, Kistler *et al.* demonstrate that the maize lineage that made its way into South America began its journey out of central Mexico in a state of partial domestication shortly after initial domestication. At the same time, the authors show that other semi-domesticated lineages followed independent trajectories through Mexico and beyond, as they diversified into various extant landraces of maize. Despite separate histories, some lineages—those that experienced subsequent gene flow with maize's wild progenitor teosinte and those, like the South American lineage, that did not—came to possess the full suite of fixed domestication traits of modern maize. This further implies, Kistler *et al.* argue, that each partially domesticated pioneer possessed the "building blocks" of fully domesticated maize. The subsequent fixation and linkage of modern maize alleles in these different lineages were the outcome of continued parallel, but independent, interactions between this evolving crop plant and humans in the different regions into which the plant dispersed. This conclusion raises questions about the nature of the human-plant interactions that, although proceeding at different rates and with different sequences of allelic fixation, nevertheless produced the same suite of domestication traits characteristic of what is known as the "domestication syndrome"—not only in maize but also in wheat, barley, rice, legumes, and other plants and animals domesticated at different times in different parts of the globe (*4*).

Traditional dispersal scenarios envisioned a process in which domesticates moved out of a restricted number of domestication centers in a wave-like fashion through demic movement or though trade. As this wave advanced, indigenous hunter-gatherers were either displaced by colonizing farming populations or induced to adopt farming and herding as more productive alternatives to traditional foraging strategies (*7*). We now know the process was much more complex. Colonizing populations did indeed move out of the Near East into Europe with their domesticates following two paths—by sea around the Mediterranean Basin and over land through central and into western Europe (*8*). In each case, however, they established pioneering farming communities in areas that were largely devoid of indigenous foragers. Complex hunter-gatherers in Europe, reliant on broad-spectrum strategies that included wild and managed local resources, often resisted adopting these domesticates, sometimes for hundreds of years, before selectively incorporating some of them into their existing subsistence regimes in highly individualistic ways. In parts of Africa unsuited to Near Eastern crop plants, low numbers of domesticated caprines and cattle introduced from the Near East were incorporated into the subsistence economies of mobile foraging populations that followed the seasonal round of migrating herd animals.

Kistler *et al.* provide another example of this process in which low-level food producing societies in the Amazon and Andes folded maize into a mix of locally domesticated, loosely managed, and wild resources. This pattern echoes that observed in eastern North America, where maize was incorporated, as a minor component, into existing food-producing economies based on a mix of local domesticates and wild resources (*9*). It was only after hundreds of years of subsequent evolution of maize and, in the case of Amazonia, human-mitigated landscape transformation that intensive maize production replaced

Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. Email: zederm@si.edu

PHOTO: FABIO DE OLIVEIRA FREITAS

broad-based food producing economies. In each case, recipient populations made strategic decisions about the utility of incorporating introduced domesticates into existing subsistence practices that were encoded in systems of ecological knowledge about local environments and biotic resources and produced stable subsistence economies.

The study of Kistler *et al.* also reveals the value of domestication as a model for exploring evolution—both biological and cultural. A debate is currently roiling evolutionary biology over the need to revise and extend traditional evolutionary theory through the development of an extended evolutionary synthesis (EES) (*10*). As Kistler *et al.* demonstrate, the domestication of plants and animals touches on all the areas of contention in this debate (*4*). The expanded time frame for the manifestation and fixation of key domestication traits documented by Kistler *et al.* provides an opportunity to evaluate the role of constructive developmental processes (especially phenotypic plasticity and niche construction) that advocates of the EES believe lend directional bias to the variation on which evolution operates (*11*). It also provides a window into how traits that arise through these processes become fixed parts of the domesticate's genome. The coevolutionary relationships between humans and target species responsible for the initial domestication of maize and its later evolution allow for an assessment of the evolutionary consequences of ecological inheritance and social learning that EES proponents see as additional inheritance systems guiding evolution (*12*). As such, the study of Kistler *et al.* joins other studies of initial domestication in providing a robust body of genetic, archaeological, and archaeobiological data within a well-constrained temporal framework to serve as models for evaluating core EES assumptions about evolution and the interface of human and natural systems that shape it (*4*). ■

REFERENCES

1. N. L. Boivin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6388 (2016).
2. B. D. Smith, M. A. Zeder, *Anthropocene* **4**, 8 (2013).
3. L. Kistler *et al.*, *Science* **362**, 1309 (2018).
4. M. A. Zeder, *Evol. Anthropol.* 10.1002/evan.21747 (2018).
5. R. R. da Fonseca *et al.*, *Nat. Plants* **1**, 14003 (2015).
6. M. Vallebueno-Estrada *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14151 (2016).
7. A. J. Ammerman, L. L. Cavalli-Sforza, *The Neolithic Transition and the Genetics of Populations in Europe* (Princeton Univ. Press, 1984).
8. M. A. Zeder, in *From Colonization to Globalization: Species Movement in Human History*, N. Boivin, R. Crassard, M. Petraglia, Eds. (Cambridge Univ. Press, 2017), chap. 11.
9. B. D. Smith, in *From Colonization to Globalization: Species Movement in Human History*, N. Boivin, R. Crassard, M. Petraglia, Eds. (Cambridge Univ. Press, 2017), chap. 13.
10. K. Laland *et al.*, *Nature* **514**, 161 (2014).
11. G. Müller, *Interface Focus* **7**, 20170015 (2017).
12. K. Laland *et al.*, *Proc. R. Soc. B* **282**, 20151019 (2015).

10.1126/science.aav7358

GRAPHIC: N. DESAI/*SCIENCE*

IMMUNOLOGY

Peanut allergen–specific antibodies go public

Characterizing peanut-specific antibodies may identify targets to treat food allergy

By **Hannah J. Gould** *and* **Faruk Ramadani**

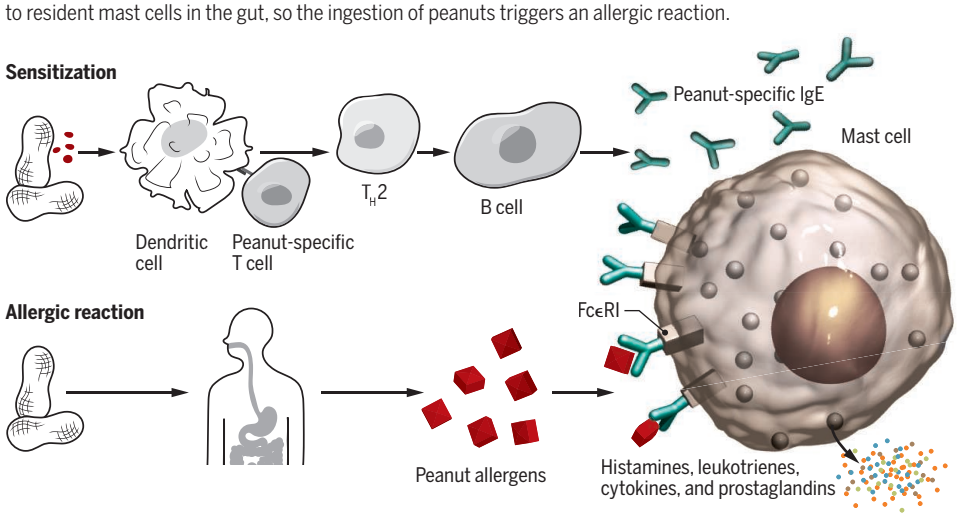
Changes in the human environment and activities over the past few decades have caused an epidemic of food allergies (*1*). People suffering from allergies often feel that they live on a cliff edge, as the allergens to which they react are potentially fatal (*2*). For example, tiny amounts of peanut picked up on skin or contaminating other foods can be dangerous to peanut-sensitized individuals (*2*–*4*). Immunoglobulin E (IgE) antibodies mediate the allergic response. They bind to specific receptors on inflammatory immune cells: mast cells in mucosal tissues lining body surfaces and cavities, and basophils in the circulation. These cells mediate allergic responses triggered by specific antigens (allergens) that are recognized by IgE. B cells expressing IgG antibodies have long served as the paradigm for the development of B cells into antibody-secreting plasma cells in the immune response. Until recently, the far less abundant IgE-expressing B cells have proved to be elusive. On page 1306 of this issue, Croote *et al.* (*5*) have analyzed single B cells from six individuals with peanut allergy, which enabled the identification of the natural Ig heavy- and light-chain pairs from IgE-expressing B cells that are responsible for peanut allergy. With this information they produced recombinant antibodies, identified the peanut allergen–specific antibodies, and used site-directed mutagenesis to suppress their activity. The mutated antibodies could be used to treat peanut allergy.

Whole-exome sequencing of single B cells from peanut-allergic individuals yielded two principal components of gene expression, representing naïve or memory B cells and plasmablasts (the circulating precursors of plasma cells). The majority of IgE-expressing cells were plasmablasts, whereas the majority of cells expressing IgG or IgA (the more abundant antibody classes) were naïve or memory B cells. It has previously been observed that IgE-expressing B cells tend to develop into the plasma cell lineage as opposed to the memory cell lineage. The

School of Biomedical Sciences and Randall Centre in Cell and Molecular Biophysics, King's College London, London SE1 1UL, UK. Email: hannah.gould@kcl.ac.uk; faruk.ramadani@kcl.ac.uk

From sensitization to peanut allergy

Dendritic cells in the skin pick up peanut allergens and present them to peanut allergen–specific T helper 2 (T_H2) cells, which in turn present them to B cells. Interaction between peanut allergen–specific T_H2 cells and B cells solicits help from T_H2 cells for B cell proliferation, somatic hypermutation and affinity maturation, class switching to IgE, and plasma cell differentiation. Allergen-specific IgE secreted by plasma cells binds to resident mast cells in the gut, so the ingestion of peanuts triggers an allergic reaction.



IgE plasma cells inherit their antigen specificity from B cells of other antibody classes, which have undergone affinity maturation. This is advantageous for their biological function in immediate hypersensitivity to antigens as it cuts out the time that would be required for affinity maturation of IgE memory B cells (*6*, *7*).

In immune responses, antigens bind to specific B cells expressing a membrane-bound form of the antibody [the B cell receptor (BCR)], which stimulates B cell maturation through the processes of somatic hypermutation (mutations affecting the antibody affinity for antigen) and affinity maturation (the selection of cells expressing BCRs with the highest affinity for antigen). The cells may also undergo class switching (from IgM to IgG, IgA, or IgE) to the most effective antibody class for a particular location in the body. IgE expression is needed for protection from parasites at barriers to the environment (airways, gut, skin). The cost of this elaborate immune mechanism is frequently the lack of normal tolerance to harmless allergens, causing allergy.

There is compelling clinical and experimental evidence that both IgE class switching and somatic hypermutation in humans occur transiently in the respiratory tract upon allergen stimulation (*8*–*10*). Whether primary contact with peanuts through the skin (*3*, *4*) is followed by local class switching to IgE in the aerodigestive tract in food allergy remains to be investigated. Immediate hypersensitivity that is characteristic of allergic reactions mediated by IgE occurs in the gut as it does in the airways (see the figure). The IgE-expressing B cells isolated from blood by Croote *et al.* may represent peanut-specific cells that have migrated out of the tissue to other sites in the body where they continue to function (*10*, *11*).

The authors focused on B cells that were of interest because the variable region sequences in six B cells from two of the six individuals studied were similar. Such similarity between individuals is highly improbable (one in 10^{14} potential sequences in the far fewer number of B cells that occur in each individual). The similarities suggest that the antigen-binding sequences are convergent or "public" sequences (inherited sequences that are conserved in evolution). Convergent sequences have been observed in infectious disease and in vaccination studies. A rationale is to hand: The relatively small germline gene repertoire encoding the Ig variable region sequences, compared to the repertoire resulting from somatic hypermutation and affinity maturation of the B cells, may have evolved in our ancestors to protect them against commonly encountered pathogens. Whether the conserved sequences serve the same purpose now or allergens are mistaken for the pathogens that affected our ancestors is unclear (*12*).

The six convergent clones were expressed as recombinant antibodies. This revealed high levels of somatic hypermutation, reflecting affinity maturation in the B cells specific for the three most common and clinically relevant peanut allergens, Ara h 1, Ara h 2, and Ara h 3. The coincidence of convergence and peanut specificity here is remarkable. Genetic mutagenesis gave insight into the crucial residues for activity, and this could be further understood through high-resolution crystal structure determination of the allergen-antibody complexes (*13*). One other B cell was shown to express an Ara h 3–specific IgE antibody. This cell was especially interesting because the IgE was related to an IgG4 (an IgG subclass) in the same cell. This confirms previous reports of related IgG4 and IgEs in allergy (*10*). IgG4 is an antibody class that confers tolerance to allergens by competing with IgE for specific antigens (*14*, *15*) and is dramatically increased in specific allergen immunotherapy. It is reassuring that the immune system itself can operate a mechanism to prevent or ameliorate allergy, which can be exploited in the clinic.

Further research on these antibodies could lead to modified antibodies or antibody fragments that compete with IgE for allergen binding and prevent the allergic response. Future use of whole-exome sequencing, perhaps comparing the development of IgE-expressing plasma cells with those expressing other antibody classes, may identify genes that regulate IgE plasma cell development and survival that could be counteracted. The work of Croote *et al.* exemplifies a concerted approach to understanding and potentially intervening in allergic disease. ■

REFERENCES

1. T. A. E. Platts-Mills, *J. Allergy Clin. Immunol.* **136**, 3 (2015).
2. J. Couzin-Frankel, *Science* **362**, 278 (2018).
3. G. Iack *et al.*, *N. Engl. J. Med.* **348**, 977 (2003).
4. S. M. Chan *et al.*, *Allergy* **67**, 336 (2012).
5. D. Coote *et al.*, *Science* **362**, 1306 (2018).
6. A. Erazo *et al.*, *Immunity* **26**, 191 (2007).
7. F. Ramadani *et al. Allergy* **72**, 66 (2017).
8. P. Campo *et al.*, *Clin. Exp. Allergy* **45**, 872 (2015).
9. H. J Gould *et al.*, *Trends Immunol.* **27**, 446 (2006).
10. H. J. Gould, C. Y. Wu, *Int. Immunol.* **30**, 403 (2018).
11. L. Ohm-Laursen *et al.*, *Front. Imunol.* **9**, 1978 (2018).
12. N. Tyagi *et al.*, *PLOS Comput. Biol.* **11**, e1004546 (2015).
13. A. N. Mitropoulou *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E8707 (2018).
14. A. F. Santos *et al.*, *J. Allergy Clin. Immunol.* **135**, 1249 (2015).
15. L. K. James *et al.*, *J. Allergy Clin. Immunol.* **130**, 605 (2012).

10.1126/science.aav3709

AGRICULTURE

Can witchweed be wiped out?

A potent stimulant induces parasitic plant germination that causes it to die

By **Harro Bouwmeester**

Root parasitic weeds of the *Orobanchaceae* such as broomrapes and witchweeds form a serious threat to agriculture in many countries around the world (*1*). They cause large yield losses in crops such as sorghum, millet, maize, rapeseed, tomato, sunflower, and legumes (*1*). These obligate parasitic plants are dependent on a host for survival, using them to grow and reproduce on. Therefore, they only germinate in the presence of a germination stimulant exuded by the host root (*2*). On page 1301 of this issue, Uraguchi *et al.* (*3*) reveal the discovery of a potent synthetic germination stimulant. Their discovery provides the basis for the development of an agrochemical that may be used to germinate parasitic weeds in the absence of a host (so that they will die, called suicide germination) and gives insight into what may be determining host specificity of these parasites.

"...SPL7 can induce suicidal germination of Striga in soil and thus reduces Striga infection of maize..."

The tight control of germination of these root parasitic plants is caused by their ability to respond to germination stimulants (*4*). These are secreted by the roots of host plants and induce seed germination. Although several compounds, from different chemical classes, in the root exudate have been identified as germination stimulants, the most important class is the strigolactones (*5*) (see the figure). The first discovered strigolactone, strigol, was isolated from the root exudate of cotton and induced germination of the root parasitic plant *Striga*

Swammerdam Institute for Life Sciences (SILS), University of Amsterdam, Science Park 904, 1098 XH Amsterdam, Netherlands. Email: h.j.bouwmeester@uva.nl

lutea (*6*). At least 25 other strigolactones have been identified in root exudates of different plant species and shown to be germination stimulants of root parasitic *Striga*, *Orobanche*, *Alectra*, and *Phelipanche* spp. (*5*, *7*).

It took more than 50 years to answer why plants are producing and secreting strigolactones (obviously not to induce germination of parasitic plant seed). In 2005, it was reported that strigolactones induce hyphal branching in arbuscular mycorrhizal (AM) fungi (*8*). AM fungi engage in a symbiotic interaction in the roots of most land plants: They supply water and nutrients in return for assimilates produced from photosynthesis. Later, it was discovered that the strigolactones are also a plant hormone that regulate plant branching (*9*, *10*). Further studies discovered that strigolactones also regulate other aspects of plant development, including root architecture and leaf senescence (*4*).

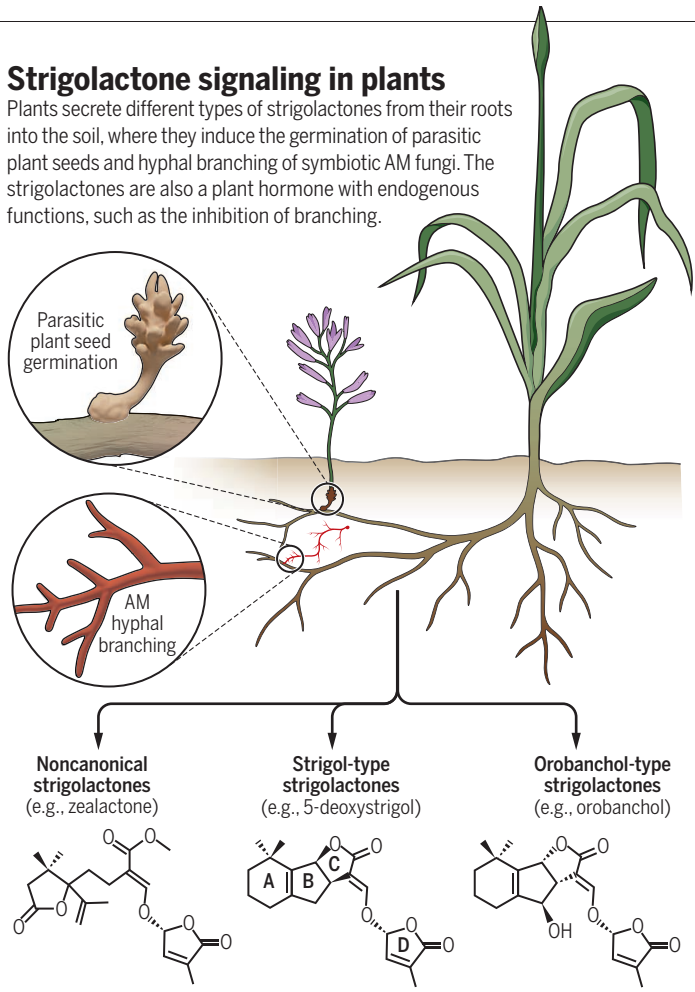
Since these discoveries, strigolactone biosynthesis was partially elucidated (*11*)—although our knowledge is far from complete (*5*). Strigolactone perception was also investigated, including the discovery of the strigolactone receptor, D14 (*12*). In the root parasitic broomrapes and witchweeds, however, a receptor homologous to D14, HYPOSENSITIVE TO LIGHT (HTL), was shown to have duplicated and evolved new ligand binding specificity, allowing these parasites to germinate upon perception of strigolactones secreted by their host (*13*, *14*). Intriguingly, the exact role and ligand of HTL in other, nonparasitic plants remains elusive (*12*).

Uraguchi *et al*. used *Striga hermonthica* (witchweed) HTL, ShHTL7, as a sensitive biosensor for germination stimulants. In a chemical screen using *Striga* germination as a readout, they identified a molecule that had considerable potency. Serendipitously, most of the activity was due to the presence of a synthetic impurity, which had the classical D-ring that is also present in all strigolactones (see the figure). Upon further optimization of this molecule, the authors generated sphynolactone-7 (SPL7), a molecule with an affinity for ShHTL7 that is comparable with the affinity of the most potent natural strigolactone known, 5-deoxystrigol. However, intriguingly, experiments in which amino acids outside the ligand binding pocket of ShHTL7 were mutated suggest that the interaction of SPL7 with ShHTL7 involves different amino acids than for 5-deoxystrigol. Although the authors do not show what the mechanism underlying this difference is, it is now clear that amino acids outside the ligand binding pocket are important in ligand specificity. This will help direct investigations into the causes of strigolactone specificity in these parasites.

This result was further underpinned with experiments in which the effect of SPL7 was compared with that of GR24 (a synthetic strigolactone with a similar D-ring as that of SPL7). SPL7 did not have the hormonal effect that GR24 has—for example, in inhibiting shoot branching or inducing root hair elongation in *Arabidopsis thaliana*. SPL7 also hardly affected AM fungi hyphal branching, in contrast to GR24. This suggests that through the structure of the rest of the molecule SPL7 has a high affinity for ShHTL7, whereas its affinity for other strigolactone receptors, such as D14 in *A. thaliana* and the as yet unknown receptor in AM fungi, is very low. Last, the authors showed that SPL7 can induce suicidal germination of *Striga* in soil and thus reduces *Striga* infection of maize that is sown afterward.

Strigolactone signaling in plants

Plants secrete different types of strigolactones from their roots into the soil, where they induce the germination of parasitic plant seeds and hyphal branching of symbiotic AM fungi. The strigolactones are also a plant hormone with endogenous functions, such as the inhibition of branching.



GRAPHIC: A. KITTERMAN/*SCIENCE*

The work of Uraguchi *et al*. confirms the crucial importance of the D-ring for the biological activity of the strigolactones. Importantly, the authors touched on a phenomenon so far hardly addressed in the field: Does specificity in germination contribute to target host specificity (*5*)? A number of *S. hermonthica* hosts produce quite different strigolactones (*5*, *7*). Sorghum produces mainly strigol-type strigolactones, such as the 5-deoxystrigol that was also used by Uraguchi *et al*. (*3*, *5*, *7*). Millet produces mainly orobanchol-type strigolactones, whereas maize produces noncanonical strigolactones (*5*, *7*). Yet, all three are severely infected by *S. hermonthica*, albeit by different strains. Whether selectivity to the strigolactones produced by these hosts plays a role in this strain preference, and whether ligand specificity of the different ShHTLs is important, is a conundrum.

SPL7 is an interesting lead for the development of suicide germination stimulants that could be used to clear fields from *Striga*, before a crop is planted. There are, however, several challenges that need to be overcome. For application in the African continent, the molecules must be extremely cheap, if not free. In addition, the application on a field and sufficient penetration into the soil will probably need large amounts of water (*15*). Clearly, a lot of research is still needed to bring this finding to the field. However, the study of Uraguchi *et al*. may lead to new approaches, such as engineering of the strigolactone profile of the crops, which could also result in solutions for this tremendous agricultural problem that causes hardship for millions of African farmers. ■

REFERENCES

1. C. Parker, *Weed Sci.* **60**, 269 (2012).
2. H. J. Bouwmeester, R. Matusova, S. Zhongkui, M. H. Beale, *Curr. Opin. Plant Biol.* **6**, 358 (2003).
3. D. Uraguchi *et al*., *Science* **362**, 1301 (2018).
4. S. Al-Babili, H. J. Bouwmeester, *Annu. Rev. Plant Biol*. **66**, 161 (2015).
5. Y. Wang, H. J. Bouwmeester, *J. Exp. Bot.* **69**, 2219 (2018).
6. C. E. Cook *et al*., *Science* **154**, 1189 (1966).
7. K. Yoneyama *et al*., *J. Exp. Bot.* **69**, 2231 (2018).
8. K. Akiyama *et al*., *Nature* **435**, 824 (2005).
9. V. Gomez-Roldan *et al*., *Nature* **455**, 189 (2008).
10. M. Umehara *et al*., **455**, 195 (2008).
11. A. Alder *et al*., *Science* **335**, 1348 (2012).
12. M. T. Waters *et al*., *Annu. Rev. Plant Biol.* **68**, 291 (2017).
13. C. E. Conn *et al*., *Science* **349**, 540 (2015).
14. S. Toh *et al*., *Science* **350**, 203 (2015).
15. B. Zwanenburg *et al*., *Pest Manag. Sci.* **65**, 478 (2009).

10.1126/science.aav8482

NEURODEGENERATION

Alzheimer's disease: The right drug, the right time

Lessons from failed clinical trials can improve the development of Alzheimer's disease–modifying therapies

By **Todd E. Golde[1], Steven T. DeKosky[1], Douglas Galasko[2]**

Alzheimer's disease (AD) is an age-associated neurodegenerative disease that is reaching epidemic proportions as a result of the aging of the world's population. Impressive gains in our understanding of AD pathogenesis have not yet translated into disease-modifying therapies that benefit patients. Is this because the knowledge that guides target identification and, hence, therapeutics, is insufficient? Are current clinical trial designs not optimal? Or are other factors contributing? Here, we highlight the challenges of developing effective AD therapies and discuss how lessons learned from failed trials must be implemented to increase the likelihood of success.

Compelling data support a contemporary version of the amyloid cascade hypothesis (ACH) in the pathogenesis of AD (*1*) (see the figure). The ACH posits that slow, progressive accumulation of aggregates of the amyloid-β protein (Aβ) in the brain triggers AD by initiating a complex pathological cascade that accelerates tau pathological pathways and leads to neurodegeneration and clinical dementia. Factors such as genetics [for example, apolipoprotein E (*APOE*) e4 variant and others], head trauma, lifestyle (for example, exercise, sleep), systemic inflammation, and vascular disease may interact to influence risk or pathologic processes. The ACH provides the rationale for therapeutics designed to (i) alter Aβ aggregate accumulation and the "toxic" actions of these aggregates; (ii) prevent tau accumulation; and (iii) target subsequent cellular dysfunction contributing to the complex downstream neurodegenerative processes that result in symptomatic AD. These diagnostic pathological features of AD can now be assessed by a research classification scheme using imaging- and fluid-based biomarkers in humans, the A/T/N (Aβ/tau/neurodegeneration) diagnostic staging system (*2*). Further, the ACH provides a framework for aligning different therapeutic interventions with disease stage (*3*) (see the figure). This framework has not been applied consistently in clinical trials of drugs that target AD. Instead, many drugs were tested at disease stages where there was concern that limited efficacy would be predicted by the ACH, primarily because testing in symptomatic patients was the most feasible route forward. Further, several trials did not define optimal doses or show evidence of sufficient target engagement. To optimize the chances of success, therapies must be tested at a disease stage where they are most likely to show efficacy (i.e., the right time) and do so only when target engagement and an effective dose have been established in early-phase clinical trials (i.e., the right drug). It is also necessary to ensure that preclinical studies supporting advancement of a therapy to human studies are rigorous and reproducible, and to evaluate, to the extent possible, the stage of disease where the therapy is most likely to show efficacy.

Completed disease-modifying AD clinical trials, primarily of drugs that target Aβ, have tested limited aspects of the ACH; many failed in phase 3, the final stage with the potential for U.S. Food and Drug Administration (FDA) approval (see supplementary materials). Only trials with proven target engagement, such as those of solanezumab and verubecestat, truly tested some aspect of the ACH (*4, 5*). Solanezumab and verubecestat both targeted soluble Aβ, which might slow accumulation in presymptomatic stages, but should have limited effects on preexisting Aβ pathology, as predicted from preclinical studies in mouse models (*6, 7*). In retrospect, such negative results are not surprising—by the time clinical symptoms appear, Aβ aggregates have accumulated over many years and the brain has undergone extensive degeneration.

Can better clinical trials be designed based on the ACH? Assessing disease modification in AD requires multiyear cycles of innovation and optimization. Practical, safety, financial, and regulatory considerations have contributed to suboptimal clinical studies. In some studies, a potentially effective drug may have been tested at the wrong disease stage, but in many studies, it has simply been the wrong drug. Although methods are available to assess target engagement and assess efficacy with biomarkers, they have not been applied consistently in early-phase trials. Moreover, evidence for sufficient target engagement was often underemphasized in go-no-go decisions to move therapies into pivotal clinical trials.

In several concluded phase 3 studies of AD therapies, ~20% of individuals enrolled with a clinical AD diagnosis did not have AD when biomarker studies were assessed postenrollment (*8*). Most trials now use Aβ imaging or cerebrospinal fluid–based biomarkers to document AD pathology in participants. This is a critical and ethical step if the therapy is targeting mechanisms underlying AD. Ongoing advances in blood-based AD biomarkers will likely increase efficiency and reduce the costs of cohort selection. Additional progress with biomarkers and more sensitive cognitive assessments that accurately track degeneration and functional decline from the earliest signs of pathology will also improve the chances of success.

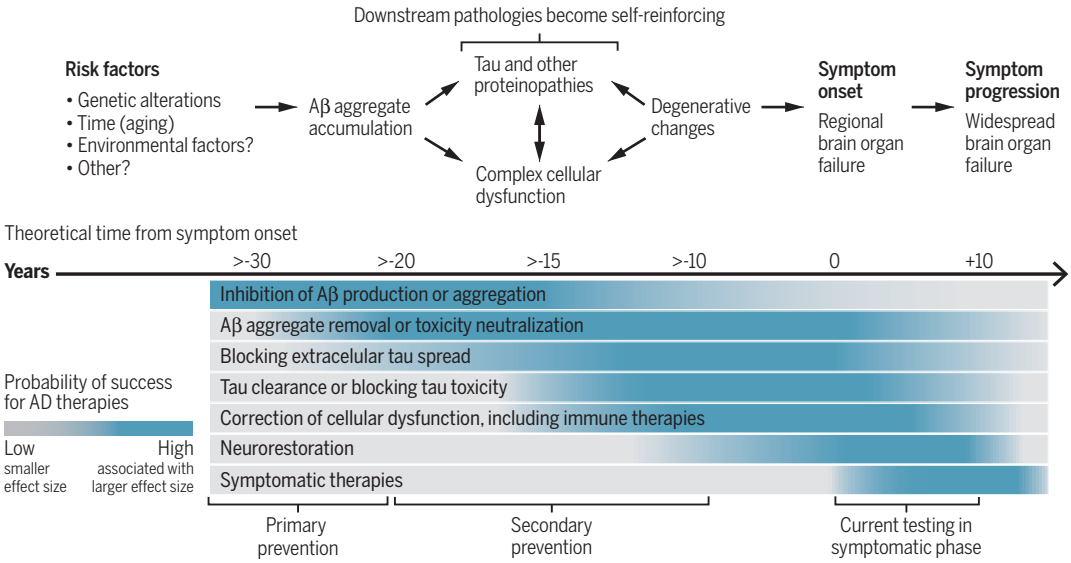
AD clinical trials have been powered to detect relatively small changes in rates of cognitive or functional decline (typically, 25 to 30% slowing of decline over 18 months) when AD is symptomatic. These trials require large cohorts, increasing costs and recruitment time. If a statistically significant slowing of decline was achieved, such an effect might be sufficient for FDA approval but may not be clinically meaningful to patients and families. Testing drugs appropriate for disease stage with biomarker-defined participants and using enough patients for larger clinical effect sizes (for example, 40 to 50% slowing of decline over 18 months) could reduce costs and increase predictive power, especially of early-phase trials.

Efforts now focus on testing agents at earlier disease stages where efficacy may be more likely. Secondary prevention trials in asymptomatic individuals who are positive for AD biomarkers and, in some instances, with high genetic risk for AD are testing interventions that target Aβ [for example, the Alzheimer's Prevention Initiative (*9*), the A4 study (*10*), and the Dominantly Inherited Alzheimer Network trials unit (*11*)]. In contrast to intervention in symptomatic AD, a therapy with modest impact on Aβ could show clinical benefit over time, because presymptomatic patients are less affected by tau deposition and structural damage occurring in symptomatic patients. However, not all individuals with positive Aβ biomarkers will develop AD, and they are healthy; these secondary prevention trial drugs require a benign safety profile. Recent guidance from

[1]Evelyn F. and William L. McKnight Brain Institute, Center for Translational Research in Neurodegenerative Disease, Departments of Neuroscience and Neurology, College of Medicine, University of Florida, Gainesville, FL, USA. [2]Department of Neurosciences and Shiley-Marcos Alzheimer's Disease Research Center, University of California San Diego, La Jolla, CA, USA. Email: tgolde@ufl.edu; steven.dekosky@neurology.ufl.edu; dgalasko@ucsd.edu

A framework for selecting the right time for the right drug

The amyloid cascade hypothesis provides a framework for timing interventions, depending on the target and likelihood that a therapy will be successful at a given stage of AD, inferred from cross-sectional autopsy studies and in vivo human biomarker studies (*1*–*3*).



GRAPHIC: N. CARY/*SCIENCE*

the FDA (*12*) suggests that acceptance of biomarker endpoints in clinical trials might be sufficient for drug approval—a considerable change to the requirement for clinical endpoints, which would take far longer. The nature of the biomarker result that might enable FDA approval remains uncertain and, similar to the history of approval of statins for cardiovascular disease, subsequent postmarket (phase 4) studies evaluating clinical efficacy would be required.

Aβ antibodies (aducanumab, BAN2401, gantenerumab, LY3002813) being tested in symptomatic AD patients appear capable of reducing Aβ aggregates, as assessed by Aβ positron emission tomography (PET), in some cases eliminating the Aβ signal (*13, 14*). Although hints of clinical benefit have emerged from these studies, the effects reported to date are small and potentially influenced by unbalanced cohorts or small group sizes, and will need to be reproduced in phase 3 trials. Autopsy studies will be needed to determine the impact of diminished Aβ PET ligand signal on brain levels of Aβ, tau, and downstream pathology.

The ultimate test of the ACH, and the test most likely to have the greatest health impact, will be in primary prevention studies—where an Aβ-targeting therapy is initiated prior to detectable Aβ accumulation in the brain. No such study has yet been launched, although planning is under way. Such studies will likely require many years to obtain a biomarker readout and even longer to test definitively that an intervention prevents or slows development of AD symptoms. Thus, the therapy needs to be extremely safe and well tolerated.

If Aβ aggregate clearance does not have clinical benefit in phase 3 studies in symptomatic AD, there are concerns that financial considerations may limit enthusiasm for further trials—even though primary and secondary prevention studies are the logical path forward. The expense and cost of trials to show benefit in a slowly progressive disease, coupled with multiple failures, have already resulted in a decline in private sector investments. Loss of investment may accelerate if failures continue.

Patients with mild AD still progress after their PET-Aβ burden is reduced (albeit, possibly, at a slower rate), reinforcing the possibility that downstream changes become independent of Aβ pathology. The point at which this independence emerges is almost certain to be defined by ongoing anti-Aβ trials. Moreover, identification of therapeutic targets beyond Aβ is essential. A limited number of current trials target tau, despite considerable interest and long-standing knowledge of its pathophysiological roles (*15*). Indeed, the extent of tau aggregation has long been known to have a direct relationship to symptoms; biomarkers, including tau PET imaging, allow it to be assessed in patients. Nevertheless, tau remains a challenging therapeutic target. First-generation tau immunotherapy trials are under way, as are efforts to lower tau levels using modified antisense oligonucleotide and a few small-molecule studies.

Given the unmet medical need and the impact of lifestyle and vascular mechanisms on dementia risk, evaluations of nonpharmacologic interventions such as exercise, behavioral therapies, and diet are important. Such interventions may have benefit in trials, although the effect size is typically small. On a larger scale, and initiated early enough (midlife), these strategies could lower population risk and have public health benefit.

As interventions are tested to prevent symptom onset, lack of therapeutic success in symptomatic studies may lead to diminishing efforts to develop therapies that benefit those who already suffer from AD. Despite the less certain biology, imperfect animal models, and challenges of treating complex neurodegenerative dysfunction, efforts must continue to identify new therapeutic approaches for the millions of individuals who have AD and the millions who will become symptomatic before an effective prophylactic treatment is identified. Selecting the right drug or drug combination to combat the pathological changes in symptomatic patients is a huge challenge, but one we must take on. We must continue to build a more predictive, translational road map and adhere to the principles of good drug development to ensure that efforts from basic science translated to clinical trial design meet the challenges of treatment and prevention. ■

REFERENCES AND NOTES

1. J. Hardy, D. J. Selkoe, *Science* **297**, 353 (2002).
2. C. R. Jack Jr. *et al.*, *Neurology* **87**, 539 (2016).
3. T. E. Golde, *J. Neurochem* **139** (suppl 2), 224 (2016).
4. M. F. Egan *et al.*, *N. Engl. J. Med.* **378**, 1691 (2018).
5. R. S. Doody *et al.*, *N. Engl. J. Med.* **370**, 311 (2014).
6. P. Das *et al.*, *Mol. Neurodegener.* **7**, 39 (2012).
7. P. Seubert *et al.*, *Neurodegener. Dis.* **5**, 65 (2008).
8. S. Salloway *et al.*, *N. Engl. J. Med.* **370**, 322 (2014).
9. E. M. Reiman *et al.*, *J. Alzheimers Dis.* **26** (suppl. 3), 321 (2011).
10. R. I. Sperling *et al.*, *Arthritis Rheum.* **33**, 1149 (1990).
11. R. J. Bateman *et al.*, *Alzheimers Dement.* **13**, 8 (2017).
12. www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM596728.pdf
13. S. Ostrowitzki *et al.*, *Alzheimers Res. Ther.* **9**, 95 (2017).
14. J. Sevigny *et al.*, *Nature* **537**, 50 (2016).
15. J. L. Cummings *et al.*, *Alzheimers Res. Ther.* **6**, 37 (2014).

ACKNOWLEDGMENTS

The authors are supported by NIH grants (AG046139, AG047266, AG005131). T.E.G. is cofounder of Lacerta Inc. and inventor on patents relating to AD therapeutics. D.G. is a consultant for vTv Therapeutics and on the data and safety monitoring board for Cognition Therapeutics. S.T.DeK. serves on a science advisory board for Amgen, a data monitoring committee for Biogen, and medical advisory board for Cognition Therapeutics.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/1250/suppl/DC1

10.1126/science.aau0437

Fog envelops Mexico's Large Millimeter Telescope, one of the Event Horizon Telescope instruments.

BOOKS *et al.*

ASTROPHYSICS

Searching for the singularity

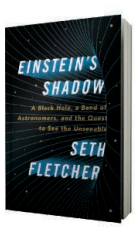
An embedded journalist tells the tale of an Earth-sized telescope that could provide the first image of a black hole

By **Matthew Kleban**

Size matters, especially when it comes to telescopes. This is partly because larger instruments collect more light and see better in the dark. But just as two separated eyes allow for stereo perception, the larger the distance between points on a telescope, or the farther apart several coordinated telescopes are, the more precisely distant objects can be resolved.

Seth Fletcher's *Einstein's Shadow* is the story of the Event Horizon Telescope (EHT)—an astrophysical endeavor on an extraordinary scale that knits radio telescopes at far-flung locations across the globe into what is, in effect, a single telescope the size of Earth. The goal of the EHT is to capture a direct image of the supermassive black hole believed to lurk at the center of our Galaxy and another even more massive hole at the center of the M87 galaxy.

Although nearly any scientist in the field (including this writer) would bet at long odds that there is, in fact, a black hole there, as Fletcher writes (paraphrasing astrophysicist Avery Broderick): "[T]he first picture of a black hole could be just as important as Pale Blue Dot." However, such a picture would say something different, "it would say, there are monsters out there."



Einstein's Shadow
Seth Fletcher
Ecco, 2018. 283 pp.

Fletcher, a writer and a senior editor at *Scientific American*, spent 6 years embedded with teams of astronomers as they traveled to distant telescopes, set up finicky equipment, and wrestled over control of, and individual credit for, the forthcoming science. The result is an ambitious and richly detailed account told mainly from the viewpoint of Shep Doeleman of the Massachusetts Institute of Technology as he conceives the idea for the EHT, struggles with technical obstacles, and absorbs a rival group. Far from the romantic image of the lone astronomer glued to his eyepiece, Doeleman (now at Harvard University and head of the EHT) is portrayed rushing around the world, simultaneously filling the roles of astronomer, technician, administrator, politician, and occasionally, weatherman.

If one's brain received a signal from the left ear with a delay relative to the right, the listener would struggle to localize where the sound was coming from. In the same way, each telescope in the EHT's network must observe the same part of the sky at the same time. "At the same time," in this case, does not just mean on the same night or even during the same few minutes. It means each telescope must collect a data stream digitally stamped with the time of the observation to an accuracy of a minuscule fraction of a second so that the data can later be precisely aligned, combined, and correlated.

The state-of-the-art clocks capable of such accuracy are called hydrogen masers. These massive, finicky beasts have to be transported to each site and carefully installed and calibrated. In one of the most entertaining parts of his book, Fletcher describes a high-altitude maser installation at a telescope in the mountains of Mexico that was nearly thwarted by muddy roads, sudden snowfall, and bandits. In the end, the delicate machine was swung, "Tarzan-style," into place.

Then there is the weather. If it does not cooperate at even one site on the night the network is supposed to be observing, the remaining telescopes might not be able to collect enough data to resolve anything of interest. But when the weather is clear, the seeing can be glorious. As Fletcher puts it, if the black hole at the center of our Galaxy "were to develop sentience and look back, it would see a conveyor belt of silver dishes mounted on mountains, a sparsely mirrored disco ball spinning at the speed of night and day."

Negotiating the politics of the EHT collaboration may be the largest challenge facing the endeavor. Questions of who's involved, who's in charge, and who gets credit for what are a recurring theme in the book. "'You know what they're fighting about, don't you?... They're fighting over who gets their name on the Nobel Prize,'" an anonymous astronomer confides to Fletcher.

It was only in the past 2 years that the EHT matured to the point that it had the capability to image these distant black holes. Some unexpected technical glitch might have prevented it from producing any image at all. On the other hand, the data may constitute a beautiful confirmation of Einstein's theory or possibly even something completely unexpected and revolutionary. Unfortunately for Fletcher, the 6 years he was embedded with the team did not suffice to reveal the answer.

Here lies the book's one notable shortcoming—it is a story without a climax. With the possible exception of a few researchers bound by secrecy, no one knows what the EHT observed, and so Fletcher's narrative abruptly fades to black. This is not a fatal flaw, but it detracts from what is otherwise a refreshingly fast-paced account of this extraordinary scientific enterprise. ■

The reviewer is at the Department of Physics, New York University, New York, NY 10003, USA. Email: kleban@nyu.edu

10.1126/science.aaw0776

PHOTO: HUGO ORTUÑO/GETTY IMAGES

SOCIAL SCIENCE

Unlocking the science of success

A complexity expert reveals how social networks create recognition and acclaim

By **Raissa M. D'Souza**

Want to master your professional and social networks to maximize recognition? Want to learn how to build productive teams that create lasting impact? In his new book, *The Formula: The Universal Laws of Success*, Albert-László Barabási translates almost a decade of scholarly research on the science of success into a lively and compelling narrative woven together with captivating stories and his own deeply personal experiences.

The book reveals the scientific underpinnings behind many informal "rules of thumb" used by successful people and provides scientific explanations for why our efforts to succeed often yield counterintuitive results. For instance, why are some ideas ignored in their own time but then catch like wildfire later? Why do two individuals with seemingly similar levels of skill and performance achieve widely different levels of notoriety?

In order to understand "success," we must first define it. At its most basic, success is about achieving a specified goal. Typically, we also associate success with recognition from our peers, fame, and profit. In *The Formula,* Barabási shows us that achieving this sort of success relies inherently on the workings of the invisible professional and social networks that shape our world. He defines "success" as the intangible things, separate from performance, that bring about recognition.

Taking us on a wildly entertaining journey from the precision-measurement world of individual performance sports, such as running and tennis, to the intangible world of art and music, to team-based efforts, Barabási reveals how to extract five "laws" that govern the recognition we will receive. He begins by showing us that when performance cannot be quantified directly, it is the perceptions of others that matter most. And even when performance can be measured, for the highest achievers, just a small increase in performance can lead to an exponential increase in how we perceive their value and in the amount of recognition they receive. Performance is ultimately limited by our personal abilities, but recognition, which comes from the networks, is unbounded.

Even when performance can be precisely quantified, measurement biases can creep in. For instance, a judge of a gymnastics competition is unlikely to give perfect marks to the first competitor, placing athletes in the first performance slot at a disadvantage. That opening competitor is further penalized if the second-round performances are ordered from worst to best, as is often the case in Olympic sports. In a world where minuscule score differences between the ultra-elite can lead to extremely varying levels of success, it is important to be aware of such effects.

High-quality work is essential, but visibility can also help breed success.

The Formula also shows us how to quantify the old adage that "success breeds success." An initial kickstart in visibility, coupled with high performance and ability (which Barabási calls "fitness"), compounds. Although a kickstart to a performer of low intrinsic quality may initially lead to high visibility, given enough time that performer should ultimately fade into obscurity.

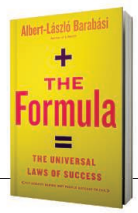
No success story is that of a single individual. Moreover, we collectively form the networks that create success, so the scientific study of success also reveals a lot about human nature. Although teamwork is typically at the core of any major success, we are quick to reward the credit to one lone individual. We like to create superstars to idolize and aspire to.

Society must also be ready to entertain a new idea for it to be adopted. We are excited about new things that balance comfort and discomfort, asimilarity and innovation. Too little innovation is boring, and too much is incomprehensible. How to strike the right balance? One strategy presented is to build a team that includes "forbidden triads" like Miles Davis did in creating his timeless masterpiece, *Kind of Blue*. This means, for instance, bringing in the strong collaborators of your strong collaborators. Of course, as Barabási shows, dumb luck, grit, and perseverance all play a role in success too.

It is worth noting that "success" measured in terms of recognition is not synonymous with happiness. Arguably, success is about achieving goals that matter to us personally, and most of us do like to receive recognition from our peers, all of which can enhance our happiness.

The Formula is an important book for us all to read. It weaves together meticulously researched historical context with more than a decade of Barabási's and other scholars' "eureka moments" and research findings to extract scientific principles and actionable insights for achieving success. And it shows us how the numerous social and professional networks that are embedded in society shape the success stories of individuals and provides an intimate portrait of a great scientist and his own path to resounding success. ■

The Formula
The Universal Laws of Success
Albert-László Barabási
Little, Brown, 2018. 316 pp.



The reviewer is at the Department of Computer Science and the Department of Mechanical and Aerospace Engineering, University of California, Davis, Davis, CA 95616, USA. Email: raissa@cse.ucdavis.edu

10.1126/science.aav7140

PHOTO: QUAVONDO/ISTOCK PHOTO

HEALTH AND MEDICINE

A new hope for beating back cancer

Vivid portraits of patients, scientists, and physicians reveal the promise of immunotherapy

By **Carolyn Wong Simpkins**

Every cancer is a living, ever-evolving, mutated derivation of a body's own cells. This makes fighting a cancer like fighting the mythological many-headed hydra. Cut off one head, and two may grow in its place.

In *The Breakthrough*, journalist Charles Graeber tells the story of how we may finally slay the beast. The "breakthrough" referenced in the title is not a single drug or treatment but a series of revelations regarding how the body's immune system regulates itself and how cancer can hijack it to avoid our defenses. This, Graeber argues, is cancer's "penicillin moment," opening the door to a radically new therapeutic approach.

Graeber is remarkably skilled at explaining complex immunological phenomena and captures the convoluted dynamics of scientific discovery. He centers each part of the narrative on a character or two, whom he brings to vivid and sympathetic life, highlighting not just their work or their disease but also their humanity, their personality, and the emotional challenges they face.

Nowhere are these strands woven together as powerfully as in the first chapter. To illustrate the game-changing nature of cancer immunotherapy, Graeber introduces "Patient 101006 JDS," a finance guy turned music industry executive named Jeff Schwartz. Diagnosed with stage 4 kidney cancer in February 2014 and fading fast, on 20 December, Schwartz secured the last spot in a clinical trial of a new immunotherapy candidate.

The physician overseeing the trial, Dan Chen, recalls wrestling with the decision of whether to admit Schwartz—whose advanced disease and poor performance status made him a less-than-ideal candidate—into a trial that could launch or doom the nascent hopes of cancer immunotherapy. "My initial reaction upon seeing him … was, 'Are you kidding me?'" However, Schwartz was ultimately admitted to the study and responded beautifully to the treatment. "Right away, I just came back to life," Schwartz would tell Graeber.

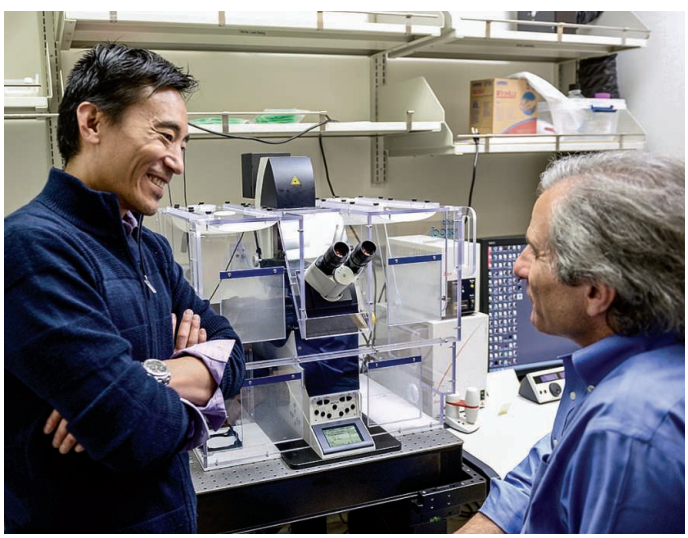
Later, Graeber steps back to examine the historical origins of the concept that the immune system could be unleashed to combat cancer. This story is centered on a series of finely drawn characters, including Elizabeth "Bessie" Dashiell, a childhood companion of John D. Rockefeller Jr. whose untimely death would inspire her surgeon, William Coley, portrayed by Graeber as a cross between Indiana Jones and Sherlock Holmes, to chase down every lead—from scientific clues in the laboratory to sociological data in the tenements of Manhattan's lower east side—ultimately earning him the moniker the "father of immunotherapy." Far from a dry accounting of historical events, look for the themes of chance observation, persistence, and fantastical luck that find resonance throughout the rest of the story.

Graeber intersperses portraits of the scientists seeking to uncover the immune system's inner workings throughout the book, alongside accessible explanations of their discoveries. Along the way, we meet many luminaries of the field, including a certain recent Nobel recipient who is colorfully introduced as "a hard living harmonica-playing Texan who … looks like something between Jerry Garcia and Ben Franklin."

Graeber also crafts beautifully evocative phrases that illuminate the workings of the body and the harsh reality of disease, describing, for example, "tumors leapfrogging each other like kids grabbing a bat handle for dibs." My favorite analogy was one in which he likened the kidney's filtering glomeruli to "a demolition worker clearing out asbestos from a condemned building" to explain the particular vulnerability of the kidney to aggressive malignancy.

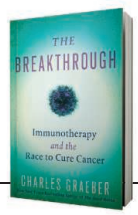
The book's final chapter takes readers back to the characters we met at the start of the book but adds little to the scientific or clinical story. I recommend saving chapter 6 to read as your final chapter. It also revisits some of the story's early characters while offering a frank discussion of the limitations of immunotherapy, the latter of which is much needed in the current era of scientific and medical hype.

Readers who come to this book with some knowledge of the immune system's workings will find a very satisfying read, with entertaining and largely accurate overviews of the workings of the immune system, an exciting flyover of the scientific journey that's brought us to our current understanding, and important reminders of the humanity of every player in this saga. Scientists and clinicians who work in cancer or immune-related disorders may wish to gift this to partners, children, parents, and friends who've never quite grasped what it is they work on. But this book really shines as a resource for laypeople who seek a better understanding of the immune system, of cancer, and of the research process. ■



In *The Breakthrough*, Dan Chen (left) candidly reflects on a memorable clinical trial.

The Breakthrough
Immunotherapy and the Race to Cure Cancer
Charles Graeber
Twelve, 2018. 320 pp.



The reviewer is a physician and molecular biologist who consults on health systems and data for technology and health care organizations. Chevy Chase, MD 20815, USA.
Email: carolyndca@gmail.com

10.1126/science.aav6272

LETTERS

Edited by **Jennifer Sills**

Sanctioning to extinction in Iran

The lifting of economic sanctions on Iran in early 2016 raised hopes among conservationists that much-needed support would finally be made available (*1*) to protect the country's unique and threatened biodiversity (*2*). Unfortunately, on 4 November, economic sanctions were reimposed, likely leading to serious repercussions on biodiversity conservation (*3*).

Conservation of threatened biodiversity often relies heavily on international cooperation, which can become impossible under economic sanctions. Sanctions reduce opportunities to transfer international expertise and skills (*2*) and erect barriers to international financial support (*4*), which together limit the capacity of conservationists within sanctioned countries to enact effective conservation interventions. These factors have hampered conservation efforts to save the critically endangered Asiatic cheetah (*Acinonyx jubatus venaticus*) (*5*), the population of which is confined entirely to Iran and now numbers fewer than 50 individuals (*6*).

Rightly, international law enshrines peoples' right to humanitarian relief during conflicts and embargos (*7*). Recently, the United Nations has taken steps to protect globally important cultural heritage sites during conflict (*8*). Biodiversity, which has global value and is critical for human well-being (*9*), requires similar protections. The UN Convention on Biological Diversity (CBD) (*10*) enshrines international responsibilities to safeguard ecosystems and biodiversity. Additional measures are needed to ensure that countries meet their CBD obligations during conflicts. Exemptions should allow the international cooperation and resources needed to save threatened species. Countries must also be required to adhere to their responsibilities (*11*) to safeguard conservation personnel ("In letter, researchers call for 'fair and just' treatment of Iranian researchers accused of espionage," R. Stone, 21 November; https://scim.ag/IranLetter). Without such measures, we may see the first continent-wide extinction of a big cat, the Asiatic cheetah, in modern times (*12*).

L. Khalatbari[1,2], J. C. Brito[1,2], A. Ghoddousi[3], H. Abolghasemi[4], U. Breitenmoser[5,6], Ch. Breitenmoser-Würsten[5], G. H. Yusefi[1,2], S. Ostrowski[7], S. M. Durant[8]*

[1]CIBIO/InBIO, Centro de Investigação em Biodiversidade e Recursos Genéticos da Universidade do Porto, 4485-661 Vairão, Portugal. [2]Departamento de Biologia da Faculdade de Ciências da Universidade do Porto, 4169-007 Porto, Portugal. [3]Geography Department, Humboldt University of Berlin, Berlin, Germany. [4]Tehran, Tehran, Iran. [5]IUCN/SSC Cat Specialist Group, c/o KORA, 3074 Muri, Switzerland. [6]Center of Fish and Wildlife Health, University of Bern, Bern, Switzerland. [7]IUCN/SSC Cat Specialist Group, Paris, France. [8]Institute of Zoology, Zoological Society of London, London, UK.
*Corresponding author.
Email: sarah.durant@ioz.ac.uk

REFERENCES

1. A. Ghoddousi *et al.*, *Front. Ecol. Environ.* **15**, 9 (2017).
2. H. Jowkar *et al.*, *Iran. Stud.* **49**, 1065 (2016).
3. M. Rao *et al.*, *AMBIO* **43**, 789 (2013).
4. D. O'Driscoll, "Impact of economic sanctions on poverty and economic growth," K4D Helpdesk Report (Institute of Development Studies, Brighton, UK, 2017).
5. H. Jowkar *et al.*, *The IUCN Red List of Threatened Species* (2008); www.iucnredlist.org/details/220/0.
6. L. Khalatbari *et al.*, *Cat News* **66**, 10 (2017).
7. Unite4Heritage, UNESCO, "UNESCO's response to protect culture in crises" (Paris, France, 2016).
8. K. King *et al.*, "Understanding humanitarian exemptions: U.N. Security Council sanctions and principled humanitarian action" (2016).
9. B. J. Cardinale *et al.*, *Nature* **486**, 59 (2015).
10. United Nations, Convention on Biological Diversity (1992).
11. IUCN, Resolution 2.37, "Support for environmental defenders," adopted at World Conservation Congress 2000 in Amman, Jordan (2000); https://portals.iucn.org/library/sites/library/files/resrecfiles/WCC_2000_RES_37_EN.pdf.
12. S. Ostrowski, *Cat News* **66**, 5 (2017).

10.1126/science.aav8221

PHOTO: BY KATOOSHA/SHUTTERSTOCK

Sanctions could drive Iran's Asiatic cheetah to extinction.

Defending the return of results and data

The National Academies of Science, Engineering, and Medicine recently published a committee report on return of individuals' research results and data, proclaiming commitment to increasing research participants' access (*1*). Our Policy Forum, "Return of results and data to study participants" (12 October, p. 159), showed that the report's recommendations would actually constrict participants' access, eroding crucial federal privacy protections and rejecting two decades of consensus recommendations on how to return results safely and ethically. In their Letter, "Standardizing return of participant results" (J. R. Botkin *et al.*, 16 November, p. 759), committee members defend their report. Their letter again shows misunderstanding of the law and reluctance to trust research participants with access to their own data and results.

The committee's report is based on a disputed position by the Centers for Medicare and Medicaid Services (CMS), which maintains that a laboratory must be certified under the Clinical Laboratory Improvement Amendments of 1988 (CLIA) in order to return individual-specific results (*1*). Unfortunately, the report's Statement of Task directed the committee to evaluate current regulations and recommend alternatives but prohibited them from analyzing "the scope or applicability of CLIA" and whether this CMS position is correct (*1*). Our Policy Forum showed that the CMS position is incorrect. Under the CLIA statute and regulations, CMS can require CLIA compliance only if a research laboratory provides information for clinical use; other purposes fall outside CLIA, including providing results to trigger clinical confirmation or allow participants to contribute data to further research. Basic administrative law analysis shows the defect in the CMS position.

Botkin *et al.* claim that "there is no

consensus" about the defect in CMS's position and that it "has not been overruled by the courts." However, the federal Secretary's Advisory Committee on Human Research Protections (SACHRP) found the CMS position "at odds with the plain language" of the CLIA regulation (*2*), which follows the statute's language. Neither source that Botkin *et al.* cite actually defends the CMS position under the administrative law principles on which our Policy Forum relied. Such legal analysis is based on established administrative law doctrines and does not depend on consensus, but on the plain language of enacted statutes and regulations.

One also cannot assume that a federal agency's position is legally correct simply because it has not yet been "overruled." Various legal doctrines limit courts' ability to hear challenges to agency position statements (*3*). It is naïve to assume courts promptly "overrule" errant agencies.

Botkin *et al.* claim that "many research institutions" are following the CMS position but cite no support (and the report indicates that others return non-CLIA results). Whatever some institutions may be doing to minimize risk in a confusing legal landscape says nothing about what an Academies committee should recommend normatively as a solution. To devise sound recommendations for law and policy, the committee needed to fully analyze the relevant statutes and legal options. We did not urge "ignoring" the CMS position; we urged the opposite—thorough analysis. The committee did not provide this, as the Statement of Task forbade it.

Botkin *et al.* also defend their recommendation to amend the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule to exclude much research data and results from the individually accessible Designated Record Set. This similarly suffers from inadequate legal analysis. The HIPAA access right clearly applies to research information, including from non-CLIA laboratories. As SACHRP notes, the Designated Record Set may include test results "from non–CLIA-certified research laboratories" (*2*). And when CLIA-confirmation is unavailable, "the results should still be provided upon the individual's request," as this is "required by law" (*2*). Congress extended HIPAA access rights to genetic information, including from research (*4*, *5*). People need access, regardless of data quality, to assess their privacy risks.

As our Policy Forum and others recognize, individuals have strong interests in access to their research results and data, especially as research transitions to more participatory models (*6*–*8*). The barriers advocated by the committee are based on inadequate legal analysis, inaccurate synthesis of current guidelines, and refusal to trust research participants. We urge regulatory agencies, research institutions, and investigators to perform a full analysis of the law and literature before acting on the recommendations of this Academies report.

Susan M. Wolf[1]* and Barbara J. Evans[2]
[1]Law School, Medical School, Consortium on Law and Values in Health, Environment & the Life Sciences, University of Minnesota, Minneapolis, MN 55455, USA. [2]Law Center, Department of Electrical and Computer Engineering, Center on Biotechnology and Law, University of Houston, Houston, TX 77204, USA.
*Corresponding author. Email: swolf@umn.edu

REFERENCES

1. The National Academies of Science, Engineering, and Medicine, "Returning individual research results to participants: Guidance for a new research paradigm" (Consensus Study Report, 2018).
2. U.S. Department of Health and Human Services, Office for Human Research Protections, *Secretary's Advisory Committee for Human Research Protections (SACHRP)*, "Attachment C: Return of individual results and special consideration of issues arising from amendments of HIPAA and CLIA" (2015).
3. N. A. Mendelson, *Cornell Law Rev.* **92**, 397 (2007).

4. Genetic Information Nondiscrimination Act of 2008, Pub. L. 110–233, 122 Stat. 881 (May 21, 2008), §§ 102, 105; www.gpo.gov/fdsys/pkg/PLAW-110publ233/pdf/PLAW-110publ233.pdf.
5. Public Health Service Act, 42 U.S.C. §§ 300gg-91(d)(16)–(18); www.gpo.gov/fdsys/pkg/USCODE-2010-title42/pdf/USCODE-2010-title42-chap6A-subchapXXV-partC-sec300gg-91.pdf.
6. J. Bollinger *et al.*, *Genet. Med.* **14**, 451 (2012).
7. J. L. Ohayon *et al.*, *Environ. Res.* **153**, 140 (2017).
8. A. Thorogood *et al.*, *Hum. Genom.* **12**, 7 (2018).

10.1126/science.aaw1851

Working governance for working land

In their Review "Landscapes that work for biodiversity and people" (19 October, p. eaau6020), C. Kremen and A. M. Merenlender discuss techniques that can preserve both ecosystem services and biodiversity in landscapes that have been modified by humans. They suggest that working lands can form useful peripheries to core protected areas. However, if appropriately managed, working lands can do more than just provide appropriate land use around strictly protected areas. Some working lands and less-strict forms of protection afford comparable conservation outcomes to state-controlled protected areas (*1*, *2*).

Whether as core or periphery, the critical challenge is to understand what governance works best to conserve the biodiversity of private, communal, and state-managed resources (*3*, *4*). On working lands, the potential for biodiversity-rich management depends on who owns and controls land or water use, on what terms, and with what objectives. Rights to resources, the rules controlling their use, and the arrangements by which these are forged, enforced, and revised are critical to conservation success (*5*, *6*). Even as there are calls for improved governance, knowledge about the relative effectiveness of different governance arrangements, and the political and social coalitions necessary to support them, remains in its infancy.

Rural people play a vital role in the protection of biodiversity in most landscapes, both within and outside protected areas (*7*, *8*). The conservation challenge lies in identifying what specific forms of governance arrangements will work in particular locations and with which rural peoples. Models must vary; we should design governance arrangements for different contexts. Only solutions tailored to the particularities of each region can win the enduring social and political support needed for maintaining biodiversity in the long term.

Dan Brockington[1]*, William M. Adams[2], Bina Agarwal[3], Arun Agrawal[4], Bram Büscher[5], Ashwini Chhatre[6], Rosaleen Duffy[7], Robert Fletcher[5], Johan A. Oldekop[3]

[1]Sheffield Institute for International Development, University of Sheffield, Sheffield S10 2TN, UK. [2]Department of Geography, University of Cambridge, Cambridge CB2 3EN, UK. [3]Global Development Institute, University of Manchester, Manchester M13 9PL, UK. [4]School for Environment and Sustainability, University of Michigan, Ann Arbor, MI 48109, USA. [5]Sociology of Development and Change, Department of Social Sciences, Wageningen University, 6700AC Wageningen, Netherlands. [6]Department of Economics and Public Policy, Indian School of Business, Hyderabad, Telangana 500 111, India. [7]Department of Politics, University of Sheffield, Sheffield S10 2TN, UK.
*Corresponding author.
Email: d.brockington@sheffield.ac.uk

REFERENCES

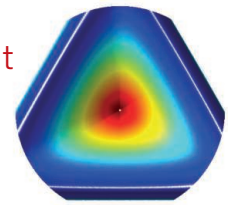
1. D. Nepstad *et al.*, *Conserv. Biol.* **20**, 65 (2006).
2. E. Ostrom, H. Nagendra, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19224 (2006).
3. J. Schleicher *et al.*, *Sci. Rep.* **7**, 11318 (2017).
4. J. A. Oldekop *et al.*, *Conserv. Biol.* **30**, 133 (2016).
5. A. Agrawal *et al.*, *Science* **320**, 1460 (2008).
6. L. Persha *et al.*, *Science* **331**, 1606 (2011).
7. S. T. Garnett *et al.*, *Nat. Sustain.* **1**, 369 (2018).
8. B. Agarwal, *Gender and Green Governance: The Political Economy of Women's Presence Within and Beyond Community Forestry* (OUP, Oxford, 2010).

10.1126/science.aav8452

RESEARCH

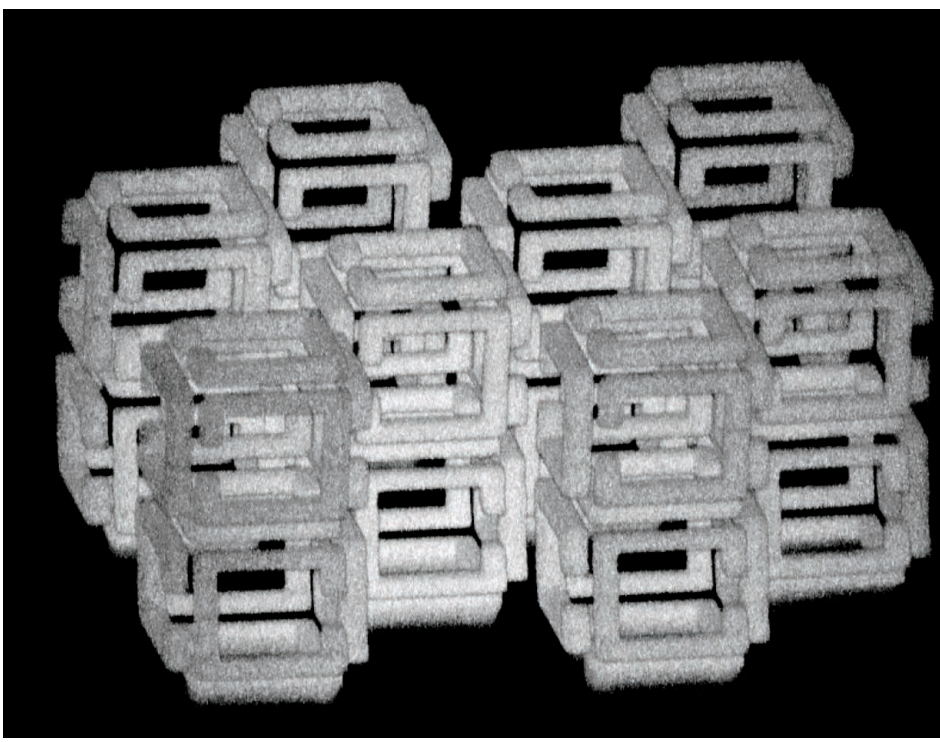
Simulation and experiment in chemistry align

Yuan et al., *p. 1289*



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**



3D PRINTING

Shrinking problems in 3D printing

Although a range of materials can now be fabricated using additive manufacturing techniques, these usually involve assembly of a series of stacked layers, which restricts three-dimensional (3D) geometry. Oran *et al.* developed a method to print a range of materials, including metals and semiconductors, inside a gel scaffold (see the Perspective by Long and Williams). When the hydrogels were dehydrated, they shrunk 10-fold, which pushed the feature sizes down to the nanoscale. —MSL

Science, this issue p. 1281; see also p. 1244

Use of a gel scaffold allows for more-complex 3D printing

ENZYMOLOGY

Evolution trains a from-scratch catalyst

Metal-bound peptides can catalyze simple reactions such as ester hydrolysis and may have been the starting point for the evolution of modern enzymes. Studer *et al*. selected progressively more-proficient variants of a small protein derived from a computationally designed zinc-binding peptide. The resulting enzyme could perform the trained reaction at rates typical for naturally evolved enzymes and serendipitously developed a strong preference for a single enantiomer of the substrate. A structure of the final catalyst highlights how small, progressive changes can remodel both catalytic residues and protein architecture in unpredictable ways. —MAF

Science, this issue p. 1285

ELECTROCATALYSIS

Combine and conquer

Platinum (Pt)–group metals, which are scarce and expensive, are used for the demanding oxygen reduction reaction (ORR) in hydrogen fuel cells. One competing approach for reducing their use is to create nanoparticles with earth-abundant metals to increase their activity and surface area; another is to replace them with metals such as cobalt (Co) in carbide or nitride sites. Chong *et al*. thermally activated a Co metal-organic framework compound to create ORR-active Co sites and then grew PtCo alloy nanoparticles on this substrate. The resulting catalyst had high activity and durability, despite its relatively low Pt content. —PDS

Science, this issue p. 1276

ARCHAEOLOGY

Early humans in northern Africa

Evidence for the earliest stone tools produced by human ancestors (from ~2.6 million years ago) has hitherto come from East Africa. Sahnouni *et al*. report the discovery of Oldowan stone artifacts and associated cutmarks on fossil bones excavated in Algeria, with the earliest dated to 2.4 million years ago. Thus, hominins inhabited the Mediterranean fringe in North Africa earlier than commonly believed. Furthermore, either stone tool manufacture and use dispersed early from East Africa or stone tool manufacture and use originated in both North and East Africa. —AMS

Science, this issue p. 1297

PLANT SCIENCE

A step toward control of a noxious weed

The parasitic plant *Striga hermonthica* causes extensive crop losses, particularly in Africa. Strigolactone hormones can be used to initiate germination of *Striga* seeds when no host crop is present, which causes the nascent *Striga* plants to die. Unfortunately, strigolactones are also used by crop plants to establish beneficial mutualisms.

IMAGES: (TOP TO BOTTOM) YUAN *ET AL.*; ORAN *ET AL.*

Published by AAAS

Uraguchi *et al*. developed a hybrid molecule that can initiate *Striga* germination without interfering with strigolactone-dependent events in the host (see the Perspective by Bouwmeester). The compound has the potential to diversify routes toward protecting fields from *Striga* infestation. —PJH

Science, this issue p. 1301; see also p. 1248

IMMUNOGENOMICS

IgE B cells unmasked

Immunoglobulin E (IgE) antibodies play a central role in immune responses against helminth and protozoan parasites; however, they also contribute to allergies. IgE antibodies (and the B cells generating them) are rare and thus poorly characterized. Croote *et al*. performed single-cell RNA sequencing of peripheral blood B cells from patients with peanut allergies and delineated each cell's gene expression, splice variants, and antibody sequences (see the Perspective by Gould and Ramadani). Unlike other isotypes, circulating IgE B cells were mostly immature plasmablasts. Surprisingly, certain IgE antibodies manifested identical gene rearrangements in unrelated individuals. These IgE antibodies showed high affinity and unexpected cross-reactivity to peanut allergens. —STS

Science, this issue p. 1306; see also p. 1247

MAIZE DOMESTICATION

The complexity of maize domestication

Maize originated in what is now central Mexico about 9000 years ago and spread throughout the Americas before European contact. Kistler *et al*. applied genomic analysis to ancient and extant South American maize lineages to investigate the genetic changes that accompanied domestication (see the Perspective by Zeder). The origin of modern maize cultivars likely involved a "semidomesticated" lineage that moved out of Mexico. Later improvements then occurred among multiple South American populations, including those in southwestern Amazonia. —LMZ

Science, this issue p. 1309; see also p. 1246

DIAGNOSTICS

Differentiating febrile disease in the field

Many infectious diseases present with common clinical symptoms, such as fever, which complicates diagnosis at the point of need. Sebba *et al*. used surface-enhanced Raman scattering (SERS) nanotags to distinguish Ebola virus infections from Lassa fever and malaria. The no-wash triplex assay workflow adds a small volume of blood and buffer to dried SERS reagents and delivers a readout within 30 minutes. The assay detected parasite- and virus-specific antigens spiked into blood, Ebola infections in nonhuman primates, and Ebola and malaria infections in human blood samples collected from endemic regions during field testing. —CC

Sci. Transl. Med. **10**, eaat0944 (2018).

EDUCATION

Later school start helps sleep and grades

Chronic sleep deprivation during adolescence is a growing problem. In 2017, the Seattle school district became the largest U.S. school district to delay secondary-school start times by nearly an hour. During this transition, Dunster *et al*. used activity wristwatches to collect quantitative evidence about the effects of a later school start time. The change increased daily sleep by more than a half hour, improved the median of students' grades by 4.5%, and reduced absenteeism and tardiness. —PJB

Sci. Adv. 10.1126/sciadv.aau6200 (2018).

IN OTHER JOURNALS

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

GALAXY EVOLUTION

Galaxy pairs follow filaments

Galaxies are nonuniformly distributed in the Universe, forming a cosmic web of filaments and clusters. Filaments occupy about 5% of the volume of the Universe but contain about a third of the galaxies, which grow by merging. Mesa *et al*. identified pairs of neighboring galaxies embedded within filaments. They found that the orientation of the pairs preferentially align with the axes of the surrounding filaments, with the effect being more pronounced for elliptical galaxies than for spirals. Because galaxy spins are known to follow the filament direction, this implies that major merger events have a preferred orientation in this environment. —KTS

Astron. Astrophys. **619**, A24 (2018).

ORGANIC CHEMISTRY

Searching for the best conditions

The vastness of the archival chemistry literature is both a blessing and a curse. The reaction that you're looking for is probably in there, provided you take enough time to search for it. Gao *et al*. trained a neural network model on 10 million known reactions to speed up this process. Specifically, the model was charged with predicting a catalyst, reagents, solvents, and temperature to achieve a given transformation. When tested, the model's top-10 list of suggestions produced a close match to actual conditions nearly 70% of the time, with a ±20°C error margin in temperature. —JSY

ACS Cent. Sci. **4**, 1465 (2018).

BIOCHEMISTRY

A cage for catalysts

The biosynthetic reactions that power cells often require unstable or toxic intermediates that must be contained and kept at low concentrations. One strategy to manage transient species is physical encapsulation, which can occur at many different size scales. Bernhardsgrütter *et al*. characterized a protein cage formed by conjoined catalytic domains, creating an incredibly small "nanoreactor" for three sequential reactions in a carbon-fixation pathway. A crystal structure revealed that each domain houses an independent active site facing the interior compartment. Enzyme kinetics suggest that the cage can close upon substrate and cofactor binding, preventing release of reaction intermediates, which have reactive moieties. —MAF

Nat. Chem. Biol. **14**, 1127 (2018).

PROTEIN FOLDING

Folding to self-destruct

The bacterial enzyme glucosamine-6-phosphate synthase (GlmS) is essential for synthesis of the cell wall. Its expression is regulated by a structured messenger RNA (mRNA) element, the *glmS* riboswitch. Most riboswitches are stabilized in an "on" conformational state by binding a ligand. In GlmS, however, ligand binding leads to self-cleavage, and this, in turn, targets the mRNA for degradation. Savinov and Block used optical tweezers to measure folding dynamics

PHOTO: ESA/HUBBLE & NASA

Spiral galaxy pairs tend to align their spin axes with the direction of the spine of their host cosmic filament.

and cleavage rates for the core *glmS* ribozyme with and without ligand. A specific duplex called P2.2 folds last and transiently. Ligand binding does not stabilize the P2.2 duplex; it is only when ligand binds this structure that cleavage occurs. A compound that stabilizes the duplex could make an antibiotic candidate. —VV

Proc. Natl. Acad. Sci. U.S.A. **115**, 11976 (2018).

TISSUE REGENERATION

Repairing injured muscle

As we become older, it takes longer to heal. Aging skeletal muscle loses its capacity to regenerate after injury. Sahu *et al*. report that α-Klotho, a protein that suppresses aging phenotypes in other tissues, may rescue muscle vitality. Muscle progenitor cells from aged mice showed decreased α-Klotho expression. Moreover, young muscle progenitor cells deficient in α-Klotho were senescent, with damaged mitochondrial DNA, compromised structural integrity, and impaired bioenergetics. The result is defective myofiber structure and an impaired repair response to injury. However, when treated with α-Klotho, older animals with muscle injury could regenerate muscle fiber and function. —LC

Nat. Commun. **9**, 4859 (2018).

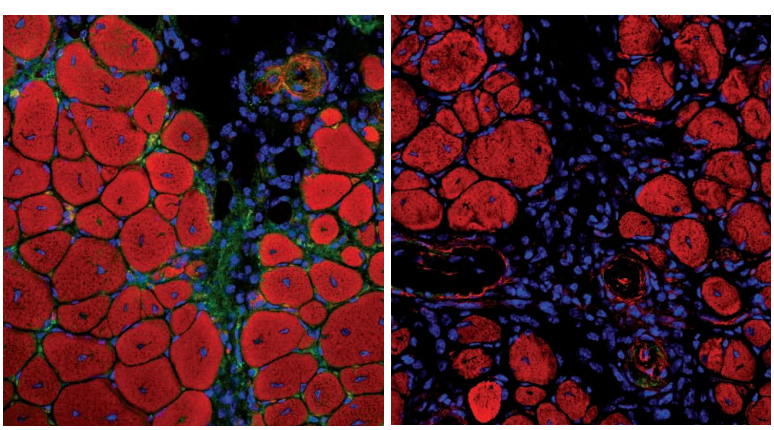


IMAGE: SAHU *ET AL.*

Immunofluorescence shows that the increase in α-Klotho (green) in damaged tissue is reduced with aging (right).

RNA MEDICINE

RNA treats preeclampsia

Small interfering RNAs (siRNAs) bound to cholesterol can be nonselectively taken up by a range of tissues with high blood flow and porous (fenestrated) endothelium. Turanov *et al*. showed that such hydrophobic siRNA accumulates in the placenta, which offers possibilities for a range of therapies for pregnancy-related diseases. Preeclampsia is a pregnancy disorder caused by a circulating tyrosine kinase called sFLT1, which inhibits blood vessel formation in the placenta, thus risking damage to the pregnancy. Placenta-originated sFLT1 has a different sequence than FTL1 in other tissues, which means an siRNA can be designed to selectively silence it. This approach was tested in both mouse and baboon preeclampsia models. —SYM

Nat. Biotech. **36**, 1164 (2018).

COMPARATIVE COGNITION

Robotic rat friends

Robots are becoming increasingly prevalent throughout society. Surprisingly perhaps, humans can feel a sense of altruism and empathy with robots that have human or animal traits. Such responses raise questions about how robots might affect social interactions. Quinn *et al*. show that rats, a highly social species that displays several types of reciprocity and empathy, will help small robots "escape" from a cage. Help is even more prompt for those robots that show rat-like social and helping behaviors. These results raise important questions about the impact of robot deployment, not just for humans but for other social species too. Importantly, these findings also dispel some of the questions that have been raised about the validity of empathy findings in species other than our own. —SNV

Anim. Behav. Cogn. **5**, 368 (2018).

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

SUPERCONDUCTIVITY

Revealing spin-orbit coupling in a cuprate

Strong coupling between the spin and orbital degrees of freedom is crucial in generating the exotic band structure of topological insulators. The combination of spin-orbit coupling with electronic correlations could lead to exotic effects; however, these two types of interactions are rarely found to be strong in the same material. Gotlieb *et al*. used spin- and angle-resolved photoemission spectroscopy to map out the spin texture in the cuprate Bi2212. Surprisingly, they found signatures of spin-momentum locking, not unlike that seen in topological insulators. Thus, in addition to strong electronic correlations, this cuprate also has considerable spin-orbit coupling. —JS

Science, this issue p. 1271

CHEMICAL PHYSICS

Pinpointing the role of geometric phase

During chemical reactions, electrons usually rearrange more quickly than nuclei. Thus, theorists often adopt an adiabatic framework that considers vibrational and rotational dynamics within single electronic states. Near the regime where two electronic states intersect, the dynamics get more complicated, and a geometric phase factor is introduced to maintain the simplifying power of the adiabatic treatment. Yuan *et al*. conducted precise experimental measurements that validate this approach. They studied the elementary H + HD reaction at energies just above the intersection of electronic states and observed angular oscillations in the product-state cross sections that are well reproduced by simulations that include the geometric phase. —JSY

Science, this issue p. 1289

RADIOCARBON

The whole story

An accurate, precise record of the carbon-14 (^{14}C) content of the atmosphere is important for developing chronologies in climate change, archaeology, and many other disciplines. Cheng *et al*. provide a record that covers the full range of the ^{14}C dating method (~54,000 years), using paired measurements of $^{14}C/^{12}C$ and thorium-230 (^{230}Th) ages from two stalagmites from Hulu Cave, China. The advantage of matching absolute ^{230}Th ages and $^{14}C/^{12}C$ allowed the authors to fashion a seamless record from a single source with low uncertainties, particularly in the older sections. —HJS

Science, this issue p. 1293

ECOLOGY

A new path for humanity

Scientific evidence of an ecological and climatic crisis caused by human actions is compelling, yet humanity is largely continuing on its current, heavily resource-dependent path. In a Perspective, Crist argues that the main reasons why humanity is not changing course lie in a human-centric worldview that discounts the value and needs of nonhuman life. As a result, placing limits on consumption appears oppressive, and technological solutions gain supremacy over efforts to reduce human impacts. Resolving the ecological and climatic crisis will instead require humanity to scale back its impacts. This will only be possible if we humans reimagine ourselves as part of the ecosphere. —JFU and SNV

Science, this issue p. 1242

NEURODEGENERATION

Improving Alzheimer's disease drug development

There has been considerable investment and effort in developing drugs to slow the progress of Alzheimer's disease, but clinical trials have been disappointing. In a Perspective, Golde *et al*. discuss the problems that have thwarted Alzheimer's disease drug development, in particular, treating patients too late during disease progression. Efforts to improve treatment and prevention strategies require a mechanism-based approach that also ensures disease progression is followed accurately during clinical trials. —GKA

Science, this issue p. 1250

NEUROSCIENCE

Treating stroke with a microRNA mimic

The loss and subsequent return of blood flow in the brain that occurs with a stroke damages brain tissue and can be lethal or severely impair cognitive and motor functions. Kim *et al*. treated rodents with an oligonucleotide mimicking the microRNA miR-7 either before or within 30 minutes of an experimentally induced stroke. The approach successfully reduced the amount of brain damage and improved motor recovery in the animals. The mimic appeared to work by repressing the expression of the protein α-synuclein, which is associated with neuronal death in various diseases. —LKF

Sci. Signal. **11**, eaat4285 (2018).

REVEALING THE BRAIN'S MOLECULAR ARCHITECTURE

By **The PsychENCODE Consortium***

The brain, our most complex organ, is at the root of both the cognitive and behavioral repertoires that make us unique as a species and underlies susceptibility to neuropsychiatric disorders. Healthy brain development and neurological function rely on precise spatiotemporal regulation of the transcriptome, which varies substantially by brain region and cell type. Recent advances in the genetics of neuropsychiatric disorders reveal a highly polygenic risk architecture involving contributions of multiple common variants with small effects and rare variants with a range of effects. Because most of this genetic variation resides in noncoding regions of the genome, establishment of mechanistic links between variants and disease phenotypes is impeded by a lack of a comprehensive understanding of the regulatory and epigenomic landscape of the human brain.

To address this matter, the PsychENCODE Consortium was established in 2015 by the National Institute of Mental Health (NIMH) to characterize the full spectrum of genomic elements active within the human brain and to elucidate their roles in development, evolution, and neuropsychiatric disorders. To reach this objective, a multidisciplinary team of investigators across 15 research institutes has generated an integrative atlas of the human brain by analyzing transcriptomic, epigenomic, and genomic data of postmortem adult and developing human brains at both the tissue and single-cell levels. Samples from more than 2000 individuals were phenotypically characterized as neurotypical or diagnosed with schizophrenia, autism spectrum disorder (ASD), or bipolar disorder.

In *Science*, *Science Translational Medicine*, and *Science Advances*, we present manuscripts that provide insights into the biology of the developing, adult, and diseased human brain. These papers are organized around three flagship articles, the first analyzing human development, the second examining disease transcriptomes, and the third describing integration of tissue and single-cell data with deep-learning approaches.

The consortium's integrative genomic analyses elucidate the mechanisms by which cellular diversity and patterns of gene expression change throughout development and reveal how neuropsychiatric risk genes are concentrated into distinct coexpression modules and cell types. Developmental analysis of macaque and human brains reveals shared and divergent spatiotemporal features and expression of neuropsychiatric risk genes. Another study shows how the transcriptomes of affected and neurotypical brains exhibit differences in gene regulatory networks and mRNA splicing, thus highlighting the importance of isoform-level regulation and cell type specificity in neuropsychiatric disorders. Because we examined a large number of individuals, quantitative trait loci (QTL) identification is improved, and QTLs are found to be associated with variation in cell type proportions in the brain, as well as those affecting chromatin, DNA hydroxymethylation, and gene expression.

Additional investigations highlight the role of noncoding regions, particularly promotors, in ASD, as well as the three-dimensional structure of the genome and specific noncoding RNAs and transcription factors in schizophrenia. For these papers, the consortium developed analytical and biological tools. These include model systems for delineating regulatory networks: human induced pluripotent stem cell–derived cerebral organoids and primary cultured olfactory neuroepithelial cells. Finally, all data and associated analysis products are available from the consortium website (psychencode.org).

Overall, efforts such as the PsychENCODE project address how to link molecules, genes, and their regulatory elements to higher levels of biological complexity, from a single cell to human behavior. However, continued investigations are necessary, and the NIMH and the PsychENCODE Consortium envision future work that will provide additional insights into human brain origin, development, and function in health and disease.

We dedicate this series of papers to Pamela Sklar, one of the chief architects and leaders of the PsychENCODE Consortium. Pamela's vision and ideas resonate throughout our studies.

**Corresponding author: Nenad Sestan (nenad.sestan@yale.edu)*

INSIDE

RESEARCH ARTICLES

RELATED ITEMS

- *SCI. TRANSL. MED.* 10.1126/scitranslmed.aat8178
- *SCI. TRANSL. MED.* 10.1126/scitranslmed.aat6912
- *SCI. ADV.* 10.1126/sciadv.aav8550
- NEWS STORY P. 1227

Genetic variants may lead to disease, denoted here by a dimmed letter representing a nucleotide. The PsychENCODE Consortium presents research to link the effects of genetic variation to gene expression in the brain.

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

RESEARCH ARTICLE SUMMARY

PSYCHIATRIC GENOMICS

Integrative functional genomic analysis of human brain development and neuropsychiatric risks

Mingfeng Li*, Gabriel Santpere*, Yuka Imamura Kawasawa*, Oleg V. Evgrafov*, Forrest O. Gulden*, Sirisha Pochareddy*, Susan M. Sunkin*, Zhen Li*, Yurae Shin*, Ying Zhu, André M. M. Sousa, Donna M. Werling, Robert R. Kitchen, Hyo Jung Kang, Mihovil Pletikos, Jinmyung Choi, Sydney Muchnik, Xuming Xu, Daifeng Wang, Belen Lorente-Galdos, Shuang Liu, Paola Giusti-Rodríguez, Hyejung Won, Christiaan A. de Leeuw, Antonio F. Pardiñas, BrainSpan Consortium, PsychENCODE Consortium, PsychENCODE Developmental Subgroup, Ming Hu, Fulai Jin, Yun Li, Michael J. Owen, Michael C. O'Donovan, James T. R. Walters, Danielle Posthuma, Pat Levitt, Daniel R. Weinberger, Thomas M. Hyde, Joel E. Kleinman, Daniel H. Geschwind, Michael J. Hawrylycz, Matthew W. State, Stephan J. Sanders, Patrick F. Sullivan, Mark B. Gerstein†, Ed S. Lein†, James A. Knowles†, Nenad Sestan†

INTRODUCTION: The brain is responsible for cognition, behavior, and much of what makes us uniquely human. The development of the brain is a highly complex process, and this process is reliant on precise regulation of molecular and cellular events grounded in the spatiotemporal regulation of the transcriptome. Disruption of this regulation can lead to neuropsychiatric disorders.

RATIONALE: The regulatory, epigenomic, and transcriptomic features of the human brain have not been comprehensively compiled across time, regions, or cell types. Understanding the etiology of neuropsychiatric disorders requires knowledge not just of endpoint differences between healthy and diseased brains but also of the developmental and cellular contexts in which these differences arise. Moreover, an emerging body of research indicates that many aspects of the development and physiology of the human brain are not well recapitulated in model organisms, and therefore it is necessary that neuropsychiatric disorders be understood in the broader context of the developing and adult human brain.

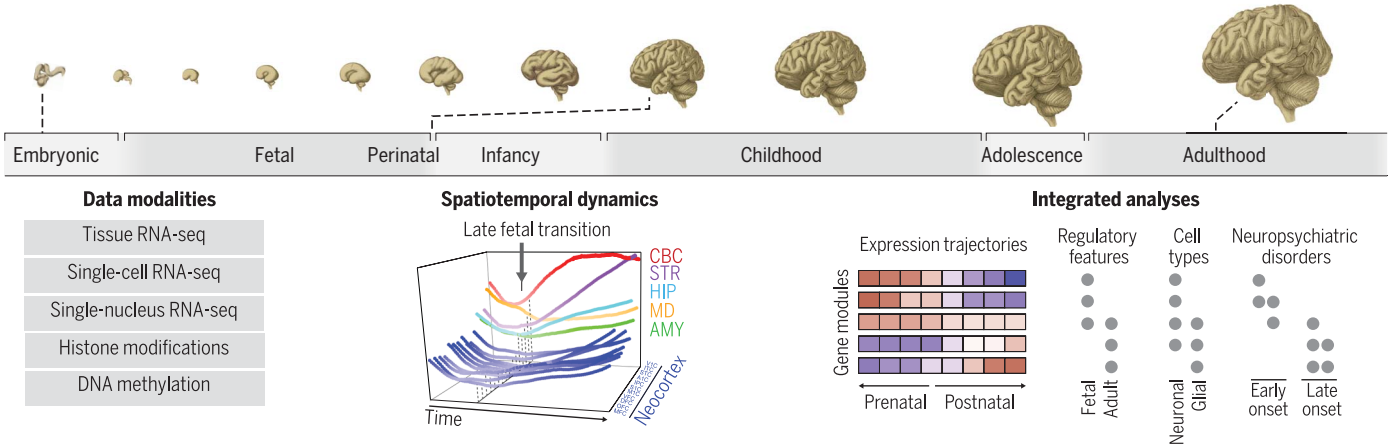
RESULTS: Here we describe the generation and analysis of a variety of genomic data modalities at the tissue and single-cell levels, including transcriptome, DNA methylation, and histone modifications across multiple brain regions ranging in age from embryonic development through adulthood. We observed a widespread transcriptomic transition beginning during late fetal development and consisting of sharply decreased regional differences. This reduction coincided with increases in the transcriptional signatures of mature neurons and the expression of genes associated with dendrite development, synapse development, and neuronal activity, all of which were temporally synchronous across neocortical areas, as well as myelination and oligodendrocytes, which were asynchronous. Moreover, genes including *MEF2C*, *SATB2*, and *TCF4*, with genetic associations to multiple brain-related traits and disorders, converged in a small number of modules exhibiting spatial or spatiotemporal specificity.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat7615

CONCLUSION: We generated and applied our dataset to document transcriptomic and epigenetic changes across human development and then related those changes to major neuropsychiatric disorders. These data allowed us to identify genes, cell types, gene coexpression modules, and spatiotemporal loci where disease risk might converge, demonstrating the utility of the dataset and providing new insights into human development and disease. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: mark.gerstein@yale.edu (M.B.G.); edl@alleninstitute.org (E.S.L.); james.knowles@downstate.edu (J.A.K.); nenad.sestan@yale.edu (N.S.)
Cite this article as M. Li *et al.*, *Science* **362**, eaat7615 (2018). DOI: 10.1126/science.aat7615



Spatiotemporal dynamics of human brain development and neuropsychiatric risks. Human brain development begins during embryonic development and continues through adulthood (top). Integrating data modalities (bottom left) revealed age- and cell type–specific properties and global patterns of transcriptional dynamics, including a late fetal transition (bottom middle). We related the variation in gene expression (brown, high; purple, low) to regulatory elements in the fetal and adult brains, cell type–specific signatures, and genetic loci associated with neuropsychiatric disorders (bottom right; gray circles indicate enrichment for corresponding features among module genes). Relationships depicted in this panel do not correspond to specific observations. CBC, cerebellar cortex; STR, striatum; HIP, hippocampus; MD, mediodorsal nucleus of thalamus; AMY, amygdala.

RESEARCH ARTICLE

PSYCHIATRIC GENOMICS

Integrative functional genomic analysis of human brain development and neuropsychiatric risks

Mingfeng Li[1]*, Gabriel Santpere[1]*, Yuka Imamura Kawasawa[1,2]*, Oleg V. Evgrafov[3]*, Forrest O. Gulden[1]*, Sirisha Pochareddy[1]*, Susan M. Sunkin[4]*, Zhen Li[1]*, Yurae Shin[1,5]*, Ying Zhu[1], André M. M. Sousa[1], Donna M. Werling[6], Robert R. Kitchen[7,8], Hyo Jung Kang[1,9], Mihovil Pletikos[1,10], Jinmyung Choi[1], Sydney Muchnik[1], Xuming Xu[1], Daifeng Wang[11], Belen Lorente-Galdos[1], Shuang Liu[1,7], Paola Giusti-Rodríguez[12], Hyejung Won[12,13], Christiaan A. de Leeuw[14], Antonio F. Pardiñas[15], BrainSpan Consortium†, PsychENCODE Consortium†, PsychENCODE Developmental Subgroup†, Ming Hu[16], Fulai Jin[17], Yun Li[18], Michael J. Owen[15], Michael C. O'Donovan[15], James T. R. Walters[15], Danielle Posthuma[14], Pat Levitt[19,20], Daniel R. Weinberger[21], Thomas M. Hyde[21], Joel E. Kleinman[21], Daniel H. Geschwind[22,23,24], Michael J. Hawrylycz[4], Matthew W. State[6], Stephan J. Sanders[6], Patrick F. Sullivan[11], Mark B. Gerstein[7,25,26,27]‡, Ed S. Lein[4]‡, James A. Knowles[3]‡, Nenad Sestan[1,8,28,29,30]‡

To broaden our understanding of human neurodevelopment, we profiled transcriptomic and epigenomic landscapes across brain regions and/or cell types for the entire span of prenatal and postnatal development. Integrative analysis revealed temporal, regional, sex, and cell type–specific dynamics. We observed a global transcriptomic cup-shaped pattern, characterized by a late fetal transition associated with sharply decreased regional differences and changes in cellular composition and maturation, followed by a reversal in childhood-adolescence, and accompanied by epigenomic reorganizations. Analysis of gene coexpression modules revealed relationships with epigenomic regulation and neurodevelopmental processes. Genes with genetic associations to brain-based traits and neuropsychiatric disorders (including *MEF2C*, *SATB2*, *SOX5*, *TCF4*, and *TSHZ3*) converged in a small number of modules and distinct cell types, revealing insights into neurodevelopment and the genomic basis of neuropsychiatric risks.

The development of the human central nervous system is an intricate process that unfolds over several decades, during which time numerous distinct cell types are generated and assembled into functionally distinct circuits and regions (*1*–*4*). These basic components of the brain are neither born mature nor static throughout their lifetimes; over the course of development, they undergo a variety of molecular and morphological changes. As a consequence, the characteristics of a given cell, circuit, or brain region described at a given time offer only a snapshot of that unit.

The processes guiding the development of the nervous system are reliant on the diversity and precise spatiotemporal regulation of the transcriptome (*1*–*4*). There is increasingly persuasive evidence that dysregulation of the transcriptional, regulatory, and epigenetic processes underlying the spatial architecture and temporal progression of human neurodevelopment can have dire consequences for brain function or strongly affect the risk of neuropsychiatric disorders (*5*–*7*). Indeed, many of the regulatory and epigenomic features governing the transcriptome of the developing human nervous system may be specific to particular developmental contexts in humans or closely related primate species. As such, it is difficult to identify or fully study human functional genomic elements using most common model organisms or cell culture systems (*8*). Assaying human cells and postmortem tissues solves some of these problems, but challenges, including the availability and quality of developmental tissue, limit the scale of such analyses. Consequently, despite ongoing efforts, our understanding of different facets of the transcriptional, regulatory, and epigenetic architecture of the human nervous system, particularly during early developmental periods, remains highly incomplete (*8*–*21*).

To begin rectifying this deficiency, the National Institutes of Health–funded PsychENCODE (http://psychencode.org) and BrainSpan Consortia (www.brainspan.org) sought to generate and analyze multidimensional genomics data from the developing and adult human brain in healthy and disease states.

Study design and data generation

Here we describe the generation and integrated analysis of multiple genomic data modalities, including transcriptomic profile, DNA methylation status, histone modifications, CTCF binding sites, and genotype generated from bulk tissue (1230 samples from 48 brains) or at the single-cell or single-nucleus level (18,288 cells or nuclei from 12 brains) from 60 de-identified postmortem brains obtained from clinically and histopathologically unremarkable donors of both sexes and multiple ancestries. Subject ages ranged from 5 postconceptional weeks (PCW) to 64 postnatal years (PY) (Fig. 1 and tables S1 to S6). Genotyping of DNA extracted from brain with a HumanOmni2.5-8 BeadChip confirmed subject ancestry and revealed no obvious genomic abnormalities (*22*).

[1]Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT, USA. [2]Departments of Pharmacology and Biochemistry and Molecular Biology, Institute for Personalized Medicine, Pennsylvania State University College of Medicine, Hershey, PA, USA. [3]Department of Cell Biology, SUNY Downstate Medical Center, Brooklyn NY, USA. [4]Allen Institute for Brain Science, Seattle, WA, USA. [5]National Research Foundation of Korea, Daejeon, South Korea. [6]Department of Psychiatry, University of California, San Francisco, San Francisco, CA, USA. [7]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA. [8]Department of Psychiatry, Yale School of Medicine, New Haven, CT, USA. [9]Department of Life Science, Chung-Ang University, Seoul, Korea. [10]Department of Anatomy & Neurobiology, Boston University School of Medicine, MA, USA. [11]Department of Biomedical Informatics Stony Brook University, NY, USA. [12]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [13]UNC Neuroscience Center, University of North Carolina, Chapel Hill, NC 27599, USA. [14]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, VU University, Amsterdam, Netherlands. [15]MRC Centre for Neuropsychiatric Genetics and Genomics, Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, UK. [16]Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH, USA. [17]Department of Genetics and Genome Science, Case Western Reserve University, Cleveland, OH, USA. [18]Department of Genetics and Department of Biostatistics, University of North Carolina, Chapel Hill, NC, USA. [19]Department of Pediatrics, Institute for the Developing Mind Keck School of Medicine of USC, Los Angeles, CA, USA. [20]Children's Hospital Los Angeles, Los Angeles, CA, USA. [21]Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA. [22]Department of Neurology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [23]Center for Autism Research and Treatment, Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [24]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA. [25]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT, USA. [26]Department of Computer Science, Yale University, New Haven, CT, USA. [27]Department of Statistics & Data Science, Yale University, New Haven, CT, USA. [28]Department of Genetics, Yale School of Medicine, New Haven, CT, USA. [29]Department of Comparative Medicine, Program in Integrative Cell Signaling and Neurobiology of Metabolism, Yale School of Medicine, New Haven, CT, USA. [30]Program in Cellular Neuroscience, Neurodegeneration, and Repair and Yale Child Study Center, Yale School of Medicine, New Haven, CT, USA.

*These authors contributed equally to this work. †For each consortium, authors and affiliations are listed in the supplementary materials. ‡Corresponding author. Email: mark.gerstein@yale.edu (M.B.G.); edl@alleninstitute.org (E.S.L.); james.knowles@downstate.edu (J.A.K.); nenad.sestan@yale.edu (N.S.)

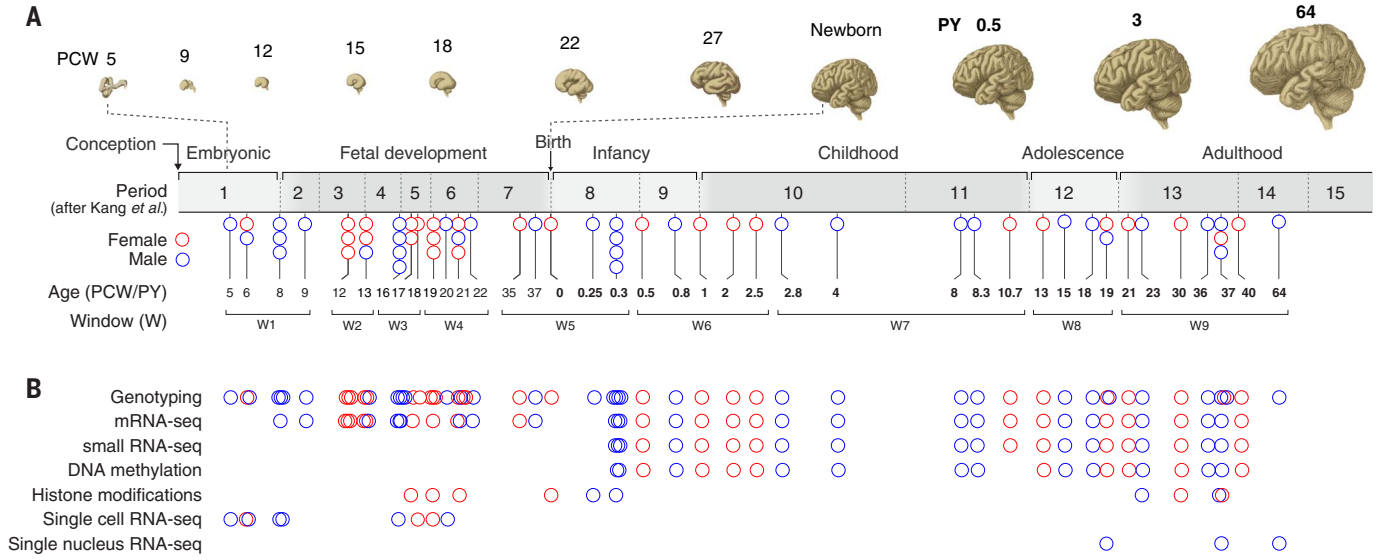


Fig. 1. Overview of the data generated in this study. (**A**) The developmental time span of the human brain, from embryonic ages (≤8 PCW) through fetal development, infancy, childhood, adolescence, and adulthood, with PCW and PY indicated. Below is the distribution of samples in this study across broad developmental phases (embryonic to adulthood), age [5 PCW to 64 PY (*19*)], and developmental windows (W1 to W9). Each circle represents a brain, and color indicates the sex [red circles (female) and blue circles (male)]. (**B**) Postmortem human brains sampled for different data modalities in this study are indicated.

For transcriptome analysis, tissue-level mRNA sequencing (mRNA-seq) was performed on a total of 607 histologically verified, high-quality tissue samples from 16 anatomical brain regions [11 areas of the neocortex (NCX), hippocampus (HIP), amygdala (AMY), striatum (STR), mediodorsal nucleus of thalamus (MD), and cerebellar cortex (CBC)] involved in higher-order cognition and behavior [Fig. 2A, (*22*)]. These regions were systematically dissected from 41 brains ranging in age from 8 PCW to 40 PY [18 females and 23 males; postmortem interval (PMI) = 12.9 ± 10.4 hours; tissue pH = 6.5 ± 0.3; RNA integrity number = 8.8 ± 1] (Fig. 1 and table S1). Because of the limited amounts of prenatal samples, small-RNA sequencing (smRNA-seq) was performed on 16 regions of 22 postnatal brains, with 278 samples passing quality control measures (Fig. 1 and table S2). These tissue-level RNA-seq analyses were complemented by single-cell RNA sequencing (scRNA-seq) data generated from 1195 cells collected from embryonic fronto-parietal neocortical wall and mid-fetal fronto-parietal neocortical plate and adjacent subplate zone of an independent set of nine brains ranging in age from 5 to 20 PCW (Fig. 1 and table S3) and single-nuclei RNA sequencing data (snRNA-seq) generated from 17,093 nuclei from the dorsolateral prefrontal cortex (DFC, also termed DLPFC) of three adult brains (Fig. 1 and table S4). For epigenome analyses, DNA cytosine methylation was profiled with the Infinium HumanMethylation450 BeadChip in 269 postnatal samples covering the same 16 brain regions analyzed by RNA-seq (Fig. 1 and table S5). Additional epigenomic data was generated with chromatin immunoprecipitation sequencing (ChIP-seq) for histone marks H3K4me3 (trimethylated histone H3 lysine 4), H3K27me3 (trimethylated histone H3 lysine 27), and H3K27ac (acetylated histone H3 lysine 27) and the epigenetic regulatory protein CTCF, which together identify a large fraction of promoters, repressors, active enhancers, and insulators. These data were generated from DFC and CBC of a subset of samples from mid-fetal, infant, and adult brains (Fig. 1 and table S6). Stringent quality control measures (figs. S1 to S8) were applied to all datasets before in-depth analyses. We also validated some results by applying independent approaches (figs. S9, S10, and S18). Finally, to enable more powerful comparisons, we grouped specimens into nine time windows (W1 to W9) on the basis of major neurodevelopmental milestones and unsupervised transcriptome-based temporal arrangement of constituent specimens (Fig. 1A and tables S1 to S6).

Global spatiotemporal dynamics

We found that most protein-coding genes were temporally (67.8%) or spatially (54.5%) differentially expressed (*22*) between at least two time windows or regions, respectively, with the majority of spatially differentially expressed genes (95.8%) also temporally differentially expressed. To gain a broad understanding of this transcriptomic variation, we analyzed the level of similarity between individual samples in the mRNA-seq dataset using multidimensional scaling applied to both gene and isoform transcript-level analyses (Fig. 2B and figs. S11 and S12). In both analyses, we found a clear divide between samples from embryonic through late mid-fetal development (W1 to W4) and samples from late infancy through adulthood (W6 to W9), with samples from the late fetal period through early infancy (W5) generally spanning this divide. To determine the relationship between these three groups, we performed unsupervised hierarchical clustering analysis and found that all samples from W5, including the late fetal samples, were more similar to early postnatal samples than to late mid-fetal samples (fig. S13). Analysis of large-scale, intraregional changes in the transcriptome across time also suggest a major transition that begins before birth. The transcriptomes of major brain regions and neocortical areas correlated well across both embryonic and early to mid-fetal (W1 to W4) and later postnatal (W6 to W9) development but displayed a sharp decrease in correlation across late fetal development and early infancy (W5) (Fig. 2C and fig. S14). This transition was also apparent at the inter-regional level. Pairwise comparisons of gene expression across all 16 brain regions found a reduction in the number of genes showing differential regional expression during W5 relative to all other windows (fig. S15). Taken together, our observation of high variation during embryonic and early to mid-fetal ages followed by a decrease across late fetal ages and the subsequent resumption of higher levels of inter- and intraregional variation during late childhood and adolescence revealed a cup-shaped, or hourglass-like, pattern of transcriptomic development (Fig. 2D).

To further explore how regional transcriptomic profiles change with age, we applied the adjustment for confounding principal components analysis algorithm (AC-PCA) (*23*), which adjusts for interindividual variations. Within any given developmental window, AC-PCA exhibited a clear separation of brain regions, but the average dissimilarity between transcription profiles of brain regions declined from W1 to W5 and then increased with age after W5 (Fig. 2, E and F, and fig. S16). Implying a relationship between transcriptional signatures and developmental origin, we

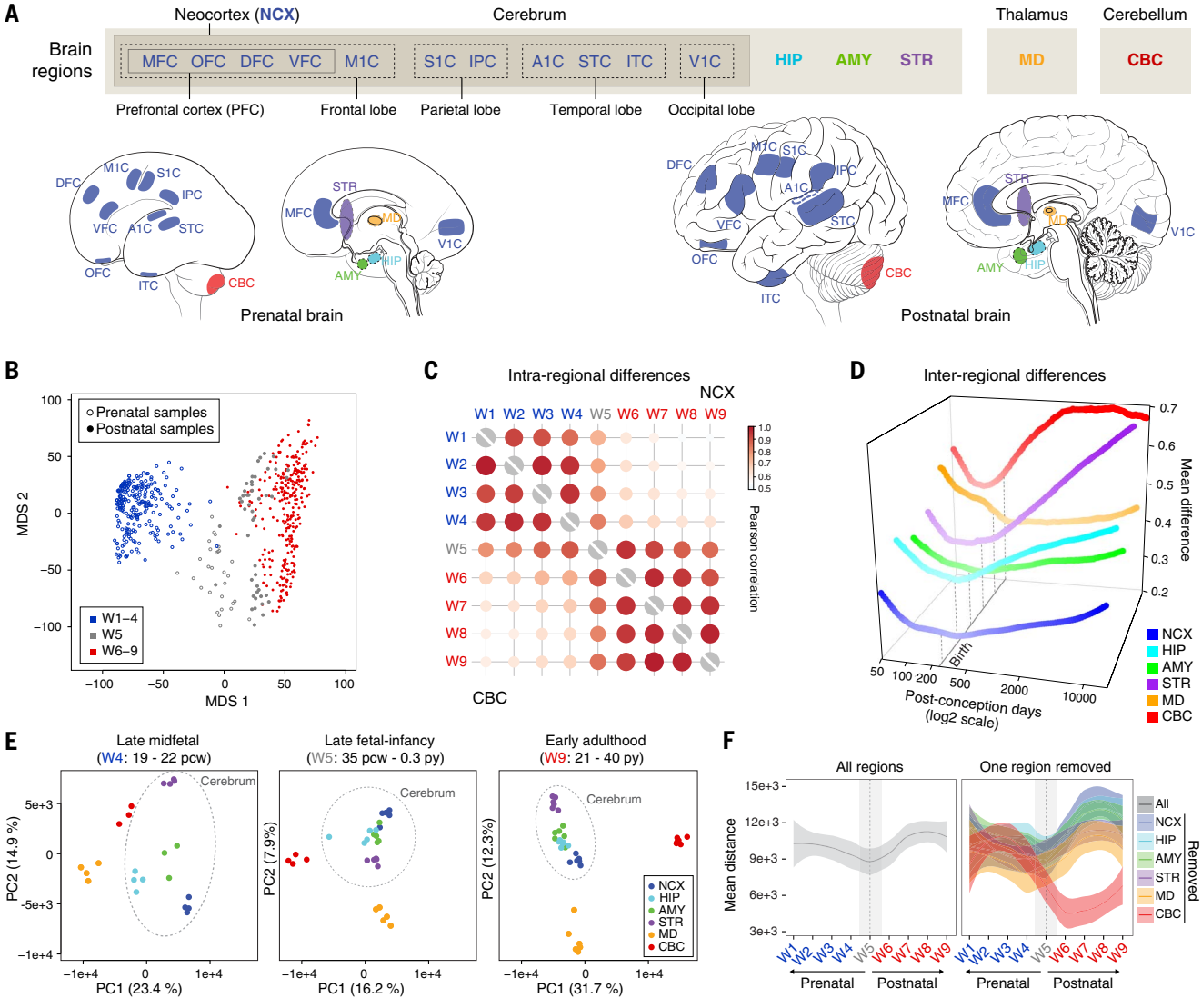


Fig. 2. Global transcriptomic architecture of the developing human brain. (**A**) mRNA-seq dataset includes 11 neocortical areas (NCX) and five additional regions of the brain. IPC, posterior inferior parietal cortex; A1C, primary auditory (A1) cortex; STC, superior temporal cortex; ITC, inferior temporal cortex; V1C, primary visual (V1) cortex. (**B**) The first two multidimensional scaling components from gene expression showed samples from late fetal ages and early infancy (W5, gray) clustered between samples from exclusively prenatal windows (W1 to W4, blue) and exclusively postnatal windows (W6 to W9, red). (**C**) Intraregional Pearson's correlation analysis found that samples within exclusively prenatal (W1 to W4) or postnatal (W6 to W9) windows correlated within, but not across, those ages. (**D**) Interregional transcriptomic differences revealed a developmental cup-shaped pattern in brain development. The interregional difference was measured as the upper quartile of the average absolute difference in gene expression of each area compared to all other areas. (**E**) AC-PCA for samples from all brain regions at late mid-fetal ages (W4), late fetal ages and early infancy (W5), and early adulthood (W9) showed that interregional differences were generally greater during W4 and W9 but reduced across W5. (**F**) Pairwise distance across samples using the first two principal components for all regions (left) or excluding one region at a time (right) demonstrated that the reduction of variation we observed is common across multiple brain regions, once the most differentiated transcriptomic profile (the cerebellum) is excluded. The shaded bands are 95% confidence intervals of the fitted lines.

found that dorsal pallium–derived structures of the cerebrum (i.e., NCX, HIP, and AMY) as well as STR became increasingly similar across prenatal development, whereas CBC and MD remained most distinct across all time windows. To confirm these observations and to evaluate the contribution of each brain region to the regional variation described by AC-PCA, we quantified the mean distance in the first two principal components across brain regions, excluding from the AC-PCA one region at a time. Because of the relative transcriptomic uniqueness of the CBC, its exclusion unmasked a qualitatively distinct and pronounced cup-shaped pattern with a transition beginning before birth and spanning the late fetal period and early infancy (Fig. 2F). CBC was again the most distinct region of the brain after multidimensional scaling analysis for expressed mature microRNAs (miRNAs), a small RNA species enriched within our smRNA-seq dataset, and the dominant contributor to miRNA expression variance (fig. S17).

The global late fetal transition and overall cup-shaped developmental dynamics we observed were also apparent when this analysis was repeated for the 11 neocortical areas included in this study (Fig. 3A and fig. S16). We observed greater dissimilarity across areas at early fetal ages (Fig. 3A), with prefrontal areas [medial prefrontal cortex (MFC), orbital prefrontal cortex (OFC), DFC, and ventrolateral prefrontal cortex (VFC)] being the most distinct. In addition, reflecting the spatial and functional topography of

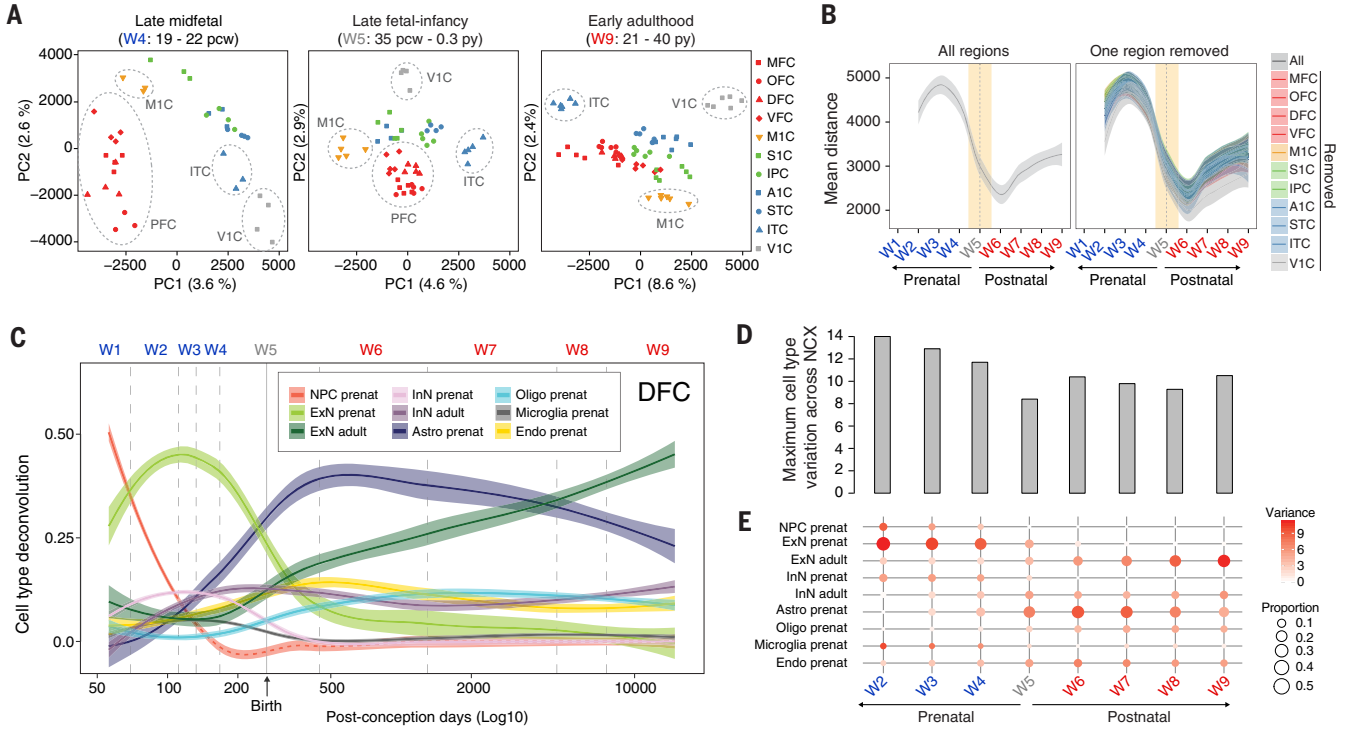


Fig. 3. Dynamics of cellular heterogeneity in the human neocortex. (**A**) AC-PCA conducted on 11 neocortical areas showed decreased interareal variation across W5, similar to our observations of interregional variation in major brain regions. (**B**) Pairwise distance across samples using the first two principal components identified a late fetal transition in all of the neocortical areas we assessed, similar to what we observed across other brain regions. (**C**) Deconvolution of tissue-level data using cell type–enriched markers identified through single-cell sequencing of primary cells from 5 to 20 PCW postmortem human brains as well as from single-nuclei sequencing of adult human brains (*27*). (**D**) Maximum interareal variance across cell types for each window. (**E**) Neocortical areal variation in the transcriptomic signatures of each major cell type assayed in each developmental window. Because of dissection protocols and rapid brain growth across early fetal development, progenitor cell proportions are nonreliable estimates after W2 [red dashed line in (C)]. The shaded bands are 95% (B) and 50% (C) confidence intervals of the fitted lines. NPC, neural progenitor cells; ExN, excitatory neurons; InN, interneurons; Astro, astroglial lineage; Oligo, oligodendrocytes; Endo, endothelial cells.

the NCX, both rostro-caudal and dorsal-ventral axes were evident in the transcriptome during fetal development. Areal differences were also seen at later ages, with functional considerations likely taking precedence over topographical arrangements. For example, VFC clustered closely with primary motor (M1C) and somatosensory (S1C) cortex, likely reflecting functional relationships with orofacial regions of the motor and somatosensory perisylvian cortex (fig. S16). Across the entirety of human brain development, transcriptomic variation between cortical regions also showed a pronounced decrease centered on the late fetal and early infancy samples of W5 (i.e., perinatal window), again reminiscent of a cup-shaped pattern (Fig. 3, A and B, and fig. S16).

Similar to gene expression, global measures of alternative splicing, such as the ratio between reads including or excluding exons [i.e., the percent spliced in index (PSI)], were higher during prenatal than postnatal ages (fig. S18 and table S7). So too was the gene expression of 68 RNA-binding proteins selected because of their involvement in RNA splicing and their analysis in adulthood by the Genotype-Tissue Expression (GTEx) project (*24*). Hierarchical clustering of expression data for these proteins also revealed a late fetal transition (fig. S19). Coincident with these observations, we found that genes exhibiting the highest interregional variation in expression in any given window [see (*22*)] exhibited a higher PSI during that window than iteratively chosen control groups of genes (fig. S18). Taken together, these analyses suggest that broad phenomena in the developing human brain, including a late fetal transition in intra- and interregional transcriptomic variation, may be amplified by alternative splicing.

Cellular heterogeneity and developmental dynamics

The high interareal variation observed during embryonic and early to mid-fetal development (Fig. 3B) coincides with a crucial period in neural development and the suspected etiology of psychiatric diseases (*4*). To help understand the temporal dynamics underlying this variation in gene expression, we analyzed our scRNA-seq data from embryonic fronto-parietal neocortical wall and mid-fetal fronto-parietal neocortical plate and adjacent subplate zone alongside our snRNA-seq data from adult human NCX and other independent datasets from overlapping developmental time points (*12*, *25*, *26*). To do so, we first applied a clustering and classification algorithm (*27*, *28*) to the prenatal scRNA-seq data after an initial division of the dataset on the basis of the age of the donor brain (i.e., embryonic or fetal), obtaining 24 transcriptomically distinct cell clusters (fig. S20). Reflecting the rapid developmental change occurring across embryonic and fetal development and the relative homogeneity of cell-type composition as compared to adult ages, as well as the specific distribution of samples in our dataset, a number of these clusters were comprised of cells from only a single donor brain, and vice versa. Suggesting that this resulted from spatiotemporal changes across brain development rather than artifactual changes related to data processing, we confirmed broad classifications of individual cells and general relationships between cell clusters and donor brains using an alternative clustering algorithm (fig. S21). Differential expression analysis and measurements of expression specificity recovered well-known gene markers of distinct types of neuronal and non-neuronal progenitor and postmitotic cell types (figs. S20 and S22 and table S8), as well as closely related groups of cell types (i.e., markers enriched in all prenatal excitatory neuron clusters) (fig. S22).

We complemented these data with snRNA-seq from adult human DFC (fig. S20), from which we identified 29 transcriptomically distinct cell

clusters representing various populations of glutamatergic excitatory projection neurons, GABAergic interneurons, oligodendrocyte progenitor cells, oligodendrocytes, astrocytes, microglia, endothelial cells, and mural cells (i.e., pericytes and vascular smooth muscle cells) (fig. S21). Alignment of our prenatal data with adult snRNA-seq data revealed hierarchical relationships and similarities between major cell classes, reflecting their developmental origins and functional properties (fig. S23). Notably, putative embryonic and fetal excitatory neurons clustered near, but did not wholly overlap with, their adult counterparts. We also observed transient transcriptomic entities, such as fetal cells in the oligodendrocyte lineage that clustered separately from their adult counterparts. Similarly, nascent excitatory neurons generally did not cluster with progenitor cells nor with fetal or adult excitatory neurons, indicating their maturationally distinct status. Confirming the validity of our prenatal scRNA-seq and adult snRNA-seq data, alignment of our prenatal data with cells from a previously published dataset (*9*) consisting of mid-fetal and adult human neocortical cells yielded similar relationships between prenatal and adult cell types (fig. S23). Comparison of neuronal transcriptomes from our prenatal single cells with both our adult single-nucleus data and independently generated adult single-nucleus data (*27*) also confirmed key differences between embryonic, mid-fetal, and adult populations. We observed limited transcriptional diversity in embryonic and mid-fetal excitatory and inhibitory neuron populations in the NCX as compared to the adult counterparts. The clusters identified in our prenatal dataset did not express specific combinations of marker genes described for the adult excitatory (fig. S24) and inhibitory (fig. S25) neurons. For example, the embryonic and mid-fetal neocortical excitatory neurons expressed combinations of genes known to be selectively enriched in different layers in adult human or mouse NCX (*29*–*31*), as previously shown in the prenatal human and mouse NCX (*12*, *31*). Notably, genes enriched in adult excitatory projection neuron subtypes located in layer (L) 5 and L6, such as *BCL11B* (*CTIP2*) and *FEZF2* (*FEZL*, *ZFP312*, or *ZNF312*), were coexpressed with L2 to L4 intracerebral excitatory projection neuron markers, such as *CUX2*, in certain embryonic and mid-fetal excitatory cell types (figs. S24 and S26). We also observed temporal changes in the coexpression patterns of cell type–specific marker genes in other cell types. For example, single-cell data from mid-fetal NCX revealed frequent coexpression of *RELN*, a marker for L1 Cajal-Retzius neurons (*32*), and *PCP4* [75.9% of 133 *PCP4*-expressing cells; reads per kilobase of exon model per million mapped reads (RPKM) ≥ 1], a marker previously shown to be expressed by deep-layer excitatory neurons (*33*). By contrast, analysis of snRNA-seq data suggested only sporadic coexpression of these genes [10.8% of 6084 *PCP4*-expressing cells; unique molecular identifier (UMI) ≥ 1] in the adult human DFC. Subsequent immunohistochemistry on independent specimens confirmed the robust coexpression of these genes in L1 of the prenatal cortex, but not in L1 or in other cortical layers of the adult cortex (fig. S26). These data imply that the molecular identities of many neuronal cell types are not fully resolved before the end of mid-fetal development and are likely malleable during early postmitotic differentiation.

Next, we utilized our single-cell and single-nucleus datasets to deconvolve bulk tissue mRNA-seq samples and estimate temporal changes in the relative proportions of major cell types in the NCX. The combined analysis revealed the cellular architecture of distinct neocortical areas and their variations across development. We observed temporal changes in cellular composition and maturational states, including the most dramatic changes during a late fetal transition (Fig. 3, C to E). For example, transcriptomic signatures for fetal excitatory neurons and fetal interneurons were generally inversely correlated with progenitor cell signatures during embryonic and early fetal development, but fetal neuron signatures nonetheless decreased across mid-fetal to late fetal development despite a concomitant reduction in the progenitor cell signature, an observation that was likely affected by our dissection strategy [Fig. 3C, (*22*)]. Similarly, signatures for adult excitatory neurons increased rapidly across the late fetal period and early infancy, coincident with the decrease in signatures of fetal excitatory neurons and interneurons (Fig. 3C). As expected, the molecular signatures for early born, deep-layer excitatory neurons preceded those for late born, upper-layer excitatory neurons (fig. S27). Transcriptomic signatures for prenatal oligodendrocytes and prenatal astrocytes also began to emerge during mid-fetal periods and increased rapidly across the late fetal transition and early infancy (Fig. 3C). Demonstrating the robustness of these observations, independent deconvolution using two alternate fetal single-cell datasets (*12*, *26*) yielded similar results (figs. S27 and S30).

Given the increase in adult cell-type signatures during W5, we next reasoned that the observed decrease in interregional transcriptomic divergence during late fetal periods and infancy may reflect a synchronized transition from fetal to more mature features of neural cells. Consequently, we analyzed the variance in cell type–specific signatures across neocortical areas, which varies in accordance with their relative proportion, and found that the maximum cell type interareal variation through time recapitulated the developmental cup-shaped pattern (Fig. 3D), with large variation in the proportion of neural progenitor cells and fetal excitatory neurons (figs. S28 and S29). Beginning during early postnatal periods, we observed increased proportions and variance in the signatures of astrocytes and, by adulthood, mature excitatory neurons (Fig. 3E). These observed temporal differences in the magnitudes and variances of the relative proportions of certain cell types and the global heterogeneity of the cell type composition at each window at least partially explain the observed pattern of interareal differences across development. Gene Ontology (GO) enrichment analysis using the top variant genes in each window, with all genes expressed in each window as background, provided further support for these changes in cell composition across areas and time. Commensurate with the changes we observed in discrete cell populations, biological processes—including neurogenesis in early developmental windows (W1 to W4), myelination in the perinatal window (W5), and sensory and ion activity calcium-related biological processes in later postnatal windows (W7 to W9), among others—exhibited regional variation in the global brain transcriptome (fig. S31 and table S9). Similar patterns of interregional variation involving discrete cell types were also observed in the macaque neocortical transcriptome (*34*), indicating that these are conserved and consistent features of prenatal primate NCX.

Other lines of evidence also suggested pronounced and qualitatively distinct regional differences in myelination, synaptic function, and neuronal activity. For example, although we observed differences in the expression of genes associated with these processes (*10*) across the NCX (fig. S31 and table S9), TempShift, a Gaussian-based model that allows the quantification of temporal shifts in the trajectories of groups of genes represented by their first principal components (*34*), indicated that of these processes, only genes associated with myelination displayed such a shift (Fig. 4A). Conversely, perhaps reflecting functional or areal diversity in cell subtypes, we observed no similar temporal shift in the expression of genes associated with synaptogenesis or neuronal activity, confirming these results through reference to published posttranslational analyses of myelinated fiber density (*35*) and synaptic density (*36*) conducted across multiple neocortical areas (Fig. 4B). Crucially, although genes associated with these processes were expressed across the late fetal transition (Fig. 4C), of the processes analyzed, only myelination contributed to the increased interareal differences we observed during this period (Fig. 4D). Suggesting that these differences are a conserved feature of primate development, we also observed similar areal differences in the transcriptional signatures of oligodendrocytes in the macaque NCX.

Overall, these observations indicate that higher levels of divergence during early prenatal and later postnatal development reflect regional variations in cell type composition, likely arising from topographical variation in progenitor populations and neuron development during prenatal ages and cell type and functional diversification during later postnatal ages.

Spatiotemporal and multimodal integration

We next sought to assess temporal variation in epigenetic signatures and their relationships to gene expression, development, and biological processes. Global DNA methylation profiling revealed that most CpG loci were either hypermethylated [37.5%; beta value (β) ≥ 0.8] or

hypomethylated (31.8%; β ≤ 0.2) in at least one sample (fig. S32), but only about 10% of the tested methylation sites were progressively hyper- or hypomethylated through prenatal windows, postnatal windows, or both. Similarly, most methylation sites also exhibited regional variation, with 64% of tested sites differentially methylated between at least two brain regions at postnatal ages. Additionally, 16% of tested sites were differentially methylated between at least two neocortical areas. Conversely, most putative promoters (66%) and a substantial proportion of putative enhancers (43%) were not differentially enriched between DFC and CBC at either fetal or adult ages. However, a greater proportion of putative enhancers [H3K27ac-enriched regions not overlapping H3K4me3-enriched regions or proximal to a transcription start site (TSS)]

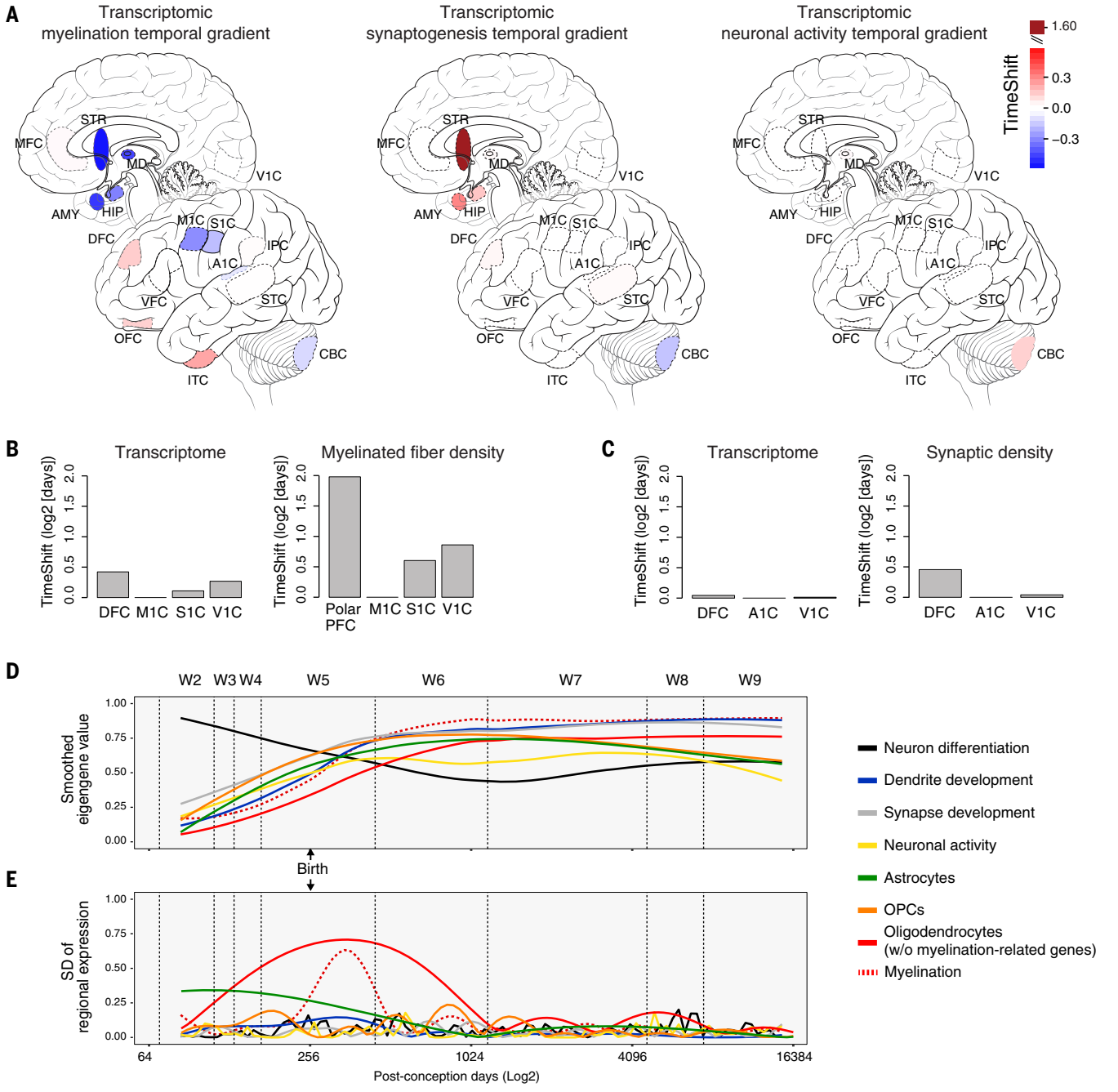


Fig. 4. Timing and temporal variation of gene expression associated with key neurodevelopmental processes. (**A**) Temporal variation, as determined by the TempShift algorithm (*34*), in the expression of genes associated with myelination showed a broad gradient across the NCX and other brain regions, whereas synaptogenesis showed only a shift between brain regions (but not neocortical areas) and neuronal activity indicated the distinct nature of the cerebellum. (**B** and **C**) Application of the TempShift algorithm to previously published posttranslational analyses of myelinated fiber density (*35*) (B) and synaptic density (*36*) (C) in multiple neocortical areas yielded relationships between areas similar to those observed in the transcriptome. (**D**) Expression of genes associated with assorted biological processes highlights pronounced change during the late fetal period and W5. (**E**) Variation in myelination-associated genes peaks during W5, as evidenced by the standard deviation of the fitted regional mean, driving interregional variation during this and neighboring (W4 and W6) windows.

were regionally (15%), temporally (17%), or spatiotemporally (24%) enriched than putative promoters (8, 14, and 12%, respectively). These differences, which suggest a greater role for enhancers relative to promoters in contributing to differential spatiotemporal gene expression, were selectively validated using quantitative droplet digital polymerase chain reaction (ddPCR) (fig. S10). We next explored correlations between methylation, histone modifications, and gene expression (figs. S32 to S34). In the adult, we found that TSSs that were more highly methylated were associated with genes that were expressed at low levels at the corresponding age, and vice versa. These relationships were not strongly indicated for methylation at other locations in the gene body (fig. S32). The presence of CBC-enriched H3K4me3 and H3K27ac marks in

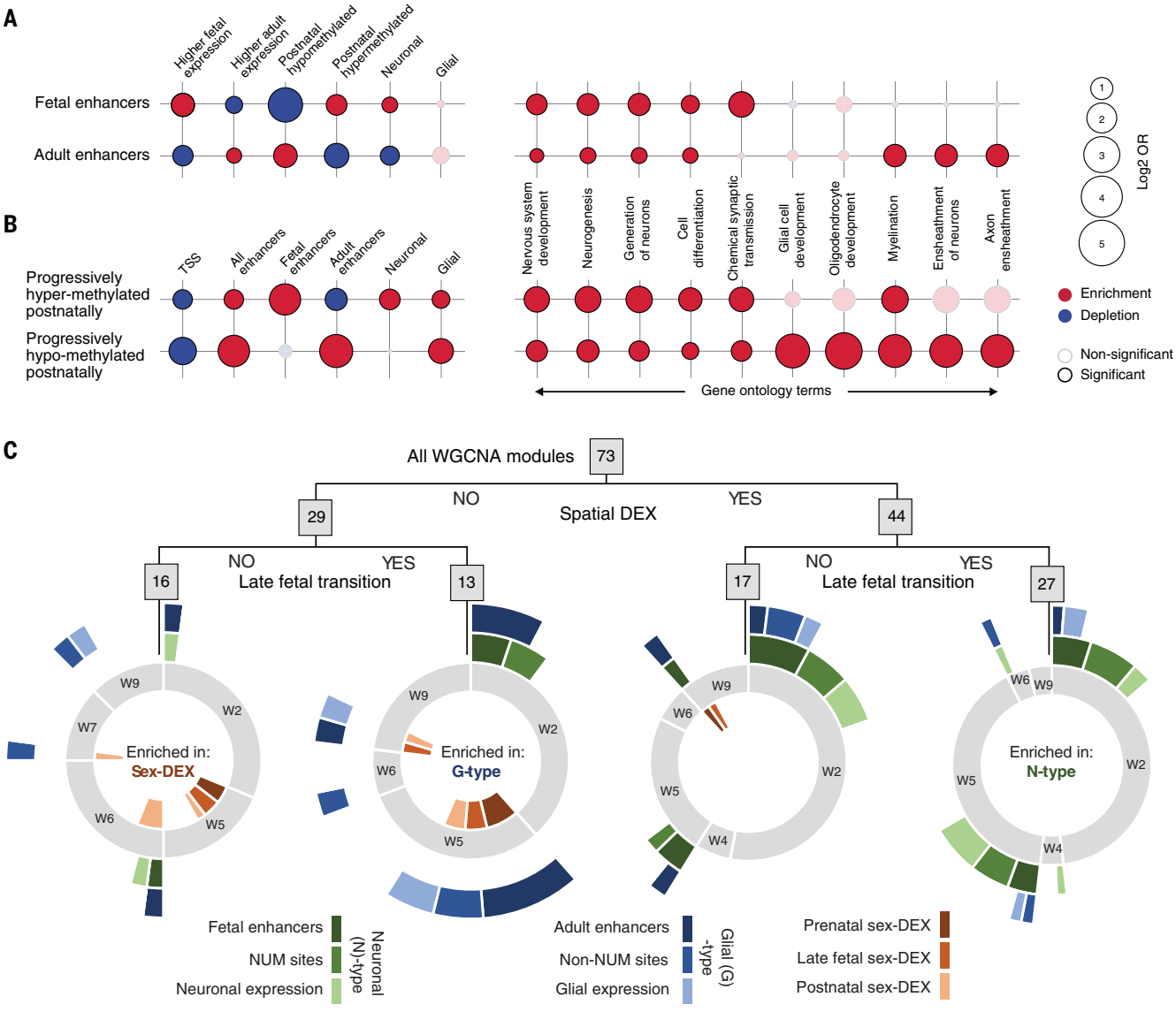


Fig. 5. Integration of gene expression and epigenetic regulation with cell types and biological processes. (**A**) Fetal-active enhancers (top left) were generally enriched for sites where methylation progressively increased across postnatal ages and associated with genes whose expression was higher during fetal development than adulthood and whose expression was enriched in neurons as compared to glia. Conversely, adult-active enhancers were enriched for sites exhibiting progressively lower methylation across postnatal ages and depleted for associations with higher fetal gene expression and expression in neurons. These enhancers were also enriched for gene ontology terms generally involving neurons and glia, respectively. OR, odds ratio. (**B**) Sites where methylation progressively increased across postnatal ages and where methylation progressively decreased across postnatal ages were generally enriched for fetal enhancers and genes whose expression was enriched in neurons, or adult enhancers and genes whose expression was enriched in glia, respectively, as well as related gene ontology terms. (**C**) Modules identified through WGCNA were segregated by regulation across brain regions, prenatal and postnatal gene expression in the NCX, both, or neither. Spatiotemporal modules (right) were enriched for modules that are themselves enriched for genes associated with enhancers active in the fetal DFC, associated with sites undermethylated in NeuN-positive (neuronal) cells, and/or enriched in neurons (N-type associations). Temporal, nonspatial modules (second from left) were enriched for modules that are themselves enriched for genes associated with enhancers active in the adult DFC, associated with sites undermethylated in non-NeuN-positive (non-neuronal) cells, and/or genes enriched in glia (G-type associations). Modules exhibiting no spatial or temporal specificity (left) were enriched for genes exhibiting sex-biased gene expression across neocortical development. Full circles (gray) indicate the proportion of modules in each category of modules exhibiting their greatest rate of change in W1 through W9.

the adult human brain also correlated strongly with increased gene expression in CBC relative to DFC (fig. S33), and vice versa. Similarly, putative fetal-active and adult-active enhancers were associated with higher fetal or adult gene expression, respectively.

In addition to epigenetic effects on gene expression, we observed discrete relationships between specific enhancers, methylation sites, and cell type–specific signatures. For example, enhancers identified during the fetal period were enriched for methylation sites that were progressively more methylated across postnatal ages (post-up), whereas adult-active enhancers were enriched for methylation sites that were progressively less methylated across postnatal ages (post-down) ($P < 0.05$, Fisher's exact test) [Fig. 5A and fig. S35, (*22*)]. Both post-up and post-down sites were themselves depleted at TSSs and enriched for sites undermethylated in neurons [neuron undermethylated (NUM) sites] and undermethylated in non-neurons (non-NUM sites) (fig. S35). They were also enriched for fetal and adult enhancers, respectively (Fig. 5B). Post-up sites were also enriched in both neuron- and glia-enriched-genes, whereas post-down sites were enriched only in glial genes (Fig. 5B) ($P < 0.05$, Fisher's exact test). Further suggesting a relationship between enhancer activity, methylation, and cell type, genes associated with fetal-active enhancers, as well as those associated with differentially methylated regions (DMRs) composed of post-up sites (*22*), were enriched for GO terms related to early events in neural development—such as neurogenesis, cell differentiation, and synaptic transmission—but generally not for processes occurring later in development (Fig. 5B and fig. S35). By contrast, genes near adult-active enhancers and post-down DMRs exhibited enrichment for postnatal or adult processes including myelination and axon ensheathment ($P < 0.01$, Fisher's exact test) (Fig. 5B and fig. S35). Taken together, these data demonstrate relationships between gene expression and epigenetic modifications, including methylation status and putative regulatory elements, as well as signatures of specific cell types and developmental programs.

We next sought further evidence that cellular dynamics contributed to the late fetal transition through the analysis of cell type– and spatiotemporal-specific patterns of gene expression and epigenetic regulation. We curated 73 gene coexpression modules resulting from weighted gene correlation network analysis (WGCNA) according to spatial relationships between brain regions and the temporal relationships of gene expression in the NCX across the late fetal transition (fig. S36 and tables S10 and S11). We found 44 modules that showed expression differences among regions in the brain (spatial), 40 modules that showed expression differences between prenatal and postnatal neocortical areas (temporal), 16 modules that were neither spatially nor temporally dynamic, and 27 modules that exhibited both spatial and temporal differences (Fig. 5C). A significantly greater than expected number of these spatiotemporally dynamic modules (including modules 2, 10, 32, and 37) exhibited their greatest change in neocortical expression from W2 through W5 ($P < 0.0118$, hypergeometric test) (Fig. 5C, fig. S37, and table S12). Genes whose expression was enriched in excitatory neurons, genes associated with putative fetal-active enhancers, and/or genes associated with NUM sites—a selection of characteristics we refer to collectively as neuronal (N)–type associations—were also enriched in spatiotemporal dynamic modules ($P < 0.0029$, hypergeometric test) (Fig. 5C, fig. S37, and table S12). Conversely, genes associated with adult-active enhancers, methylation sites hypomethylated in non-NUM sites, and glial genes [glial (G)–type modules or associations in Fig. 5C, fig. S37 and table S12] were enriched among the 13 modules where temporal ($P < 0.0002$, hypergeometric test), but not spatial, specificity was observed. These observations indicate increased spatial diversity of neuronal cell types relative to glial cell populations.

Analyses by sex revealed that modules enriched for the 783 genes exhibiting sex-differential expression (sex-DEX) in at least two consecutive windows in at least one brain region were enriched among modules with no spatial or temporal differential expression in the NCX ($P < 0.0029$, hypergeometric test) and depleted among spatiotemporal modules ($P < 0.0021$, hypergeometric test) (Fig. 5C and fig. S37). There were four modules exhibiting temporal expression differences in the NCX that were also enriched for sex-biased genes, as well as glial and other cell type–enriched markers, but these did not represent a significant enrichment in sex-DEX enriched modules among strictly temporal modules ($P < 0.132$, hypergeometric test). In addition, no module comprised of autosomal genes exhibited persistent male or female dimorphism across both prenatal development and later postnatal ages such as adolescence or adulthood (figs. S38 and S39); in cases in which an autosomal module was sex-DEX throughout development, the sex exhibiting higher expression reversed between early and late postnatal development (fig. S39). This observation was upheld when multiple thresholds were used for the identification of sexual dimorphism (fig. S40). Similarly, we identified no autosomal genes that exhibited sexual dimorphism throughout development in all brain regions or neocortical areas (figs. S38 and S39).

Cellular and temporal convergence of neuropsychiatric disease risks

Loci implicated in several neuropsychiatric disorders have been identified through genome-wide

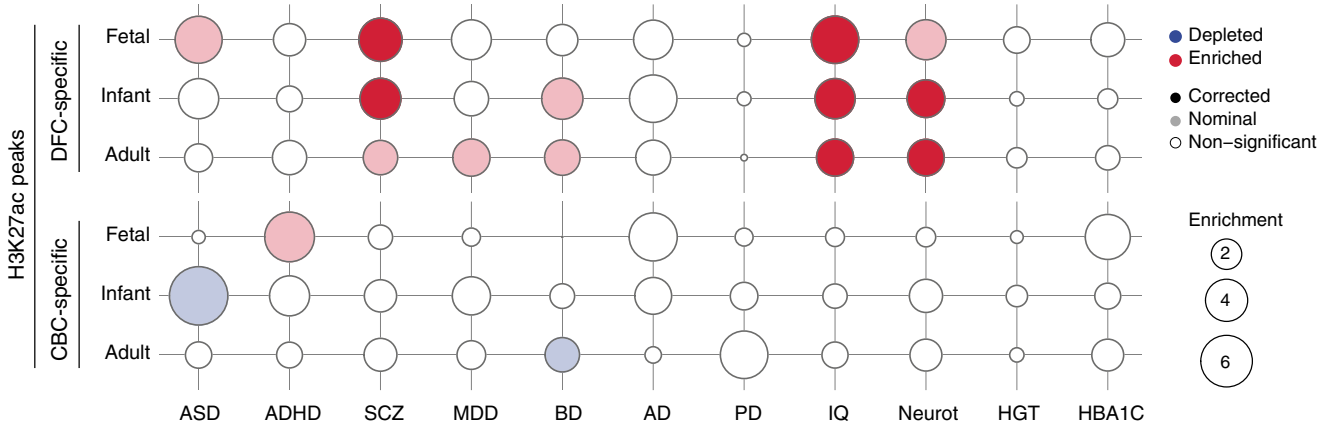


Fig. 6. Enrichment analysis for GWAS loci among putative regulatory elements. Putative promoters and enhancers (H3K27ac peaks) specific for DFC or CBC in the fetal, infant, or adult were enriched for SNP heritability identified through partitioned LD score regression analysis from GWASs for autism spectrum disorder [ASD, (*40*)], attention-deficit hyperactive disorder [ADHD, (*41*)], schizophrenia [SCZ, (*37*)], major depressive disorder [MDD, (*42*)], bipolar disorder [BD, (*43*)], Alzheimer's disease [AD, (*38*)], Parkinson's disease [PD, (*39*)], IQ, (*44*), or neuroticism [Neurot, (*45*)] but not for non-neural disorders or traits such as height [HGT, (*46*)] or diabetes [HBA1C, (*49*)]. Solid color indicates significance for Bonferroni adjusted *P* value, and faint color indicates nominal significance at LD score regression $P < 0.05$.

association studies (GWAS) and are enriched in putative noncoding regulatory elements (29–31). We sought to determine whether the proportion of phenotypic variance explained by common single-nucleotide polymorphisms (SNPs) in large neuropsychiatric GWAS (i.e., SNP heritability) was enriched in the cis-regulatory elements we identified at W1, W4, W5, and W9 in DFC and CBC. Toward this end, we collected GWAS data concerning neuropsychiatric disorders or personality traits including schizophrenia (SCZ) from CLOZUK (37), Alzheimer's disease (AD) from IGAP (38), Parkinson's disease (PD) (39), autism spectrum disorder (ASD) (40), attention deficit hyperactivity disorder (ADHD) from iPSYCH (41), major depressive disorder (MDD) (42), bipolar disorder (BD) (43), intelligence quotient (IQ) (44), and neuroticism (45), as well as non-neural traits such as height from GIANT (46), inflammatory bowel disease (IBD) (47), total cholesterol levels (48), and an endophenotype associated with diabetes (HBA1C) (49). Using partitioned linkage disequilibrium (LD) score regression analysis, we found that SNP heritability in SCZ, IQ, and neuroticism were exclusively enriched in DFC-specific, but not CBC-specific, regulatory elements as identified by peak regions of H3K27ac activity. By contrast, SNP heritability in AD or PD rendered no significant associations, and the analysis on ASD, ADHD, BD, and MDD was only nominally enriched or not enriched in putative region-specific fetal enhancers [Fig. 6 and fig. S41, (22)]. Non-neural traits (such as height and HBA1C) were also not enriched in either DFC- or CBC-specific regulatory elements but were instead enriched in regulatory elements active in the two brain regions (fig. S41), indicating a general enrichment of many of our tested GWASs in H3K27ac regions when considering a set of more ubiquitous regulatory regions.

After aggregating GWAS SNPs and identifying candidate associated regions on the basis of their *P* values and LD patterns in individuals of northwest European ancestry (50), we next leveraged partially overlapping Hi-C datasets, derived from mid-fetal and adult NCX and processed by two independent research groups (51–53), as well as H3K27ac activity in the brain, to develop two lists of genes putatively associated with those GWAS-associated regions. To do so, we initially populated both lists of disease-associated genes by identifying TSSs overlapping H3K27ac peaks that themselves overlapped a GWAS significant region, as well as genes directly affected by GWAS significant variants within the LD region, as predicted by EnsemblV78. We next expanded these lists of disease-associated genes by identifying TSSs that interact with H3K27ac peaks overlapping GWAS significant regions, excluding interactions that did not overlap with at least one H3K27ac peak at each end or where peak-to-peak interactions were not concordant in time and brain region. In the first, less stringent list (list 1), a single interaction from either of the two Hi-C datasets was sufficient to associate a gene to a GWAS locus (table S13). For the second, more stringent list (list 2), we excluded those genes whose only association to a GWAS locus was via Hi-C interactions identified in only one of the two Hi-C datasets (table S14).

We next sought to determine the cell types enriched for the expression of the high-stringency genes implicated in neuropsychiatric disorders or brain-based traits, using our prenatal scRNA-seq and adult snRNA-seq datasets and matching prenatal and adult datasets generated from the macaque (34). We found numerous cell types enriched for disease-associated loci in both human and macaque (fig. S42). For example, neocortical excitatory neurons were enriched for the expression of genes we associated with IQ in both the fetal and adult human as well as the fetal and adult macaque. However, we found no other excitatory neuron populations in the macaque AMY, STR, HIP, thalamus, or cerebellum enriched for genes associated with IQ. Similarly, neural progenitors in the prenatal macaque AMY, but not progenitors in the prenatal macaque HIP, thalamus, NCX, or STR, were enriched for the expression of genes associated with MDD, a finding especially intriguing given the variable or potentially increased size of some amygdalar nuclei in MDD patients (54, 55). Similarly confirmatory was the enrichment of SCZ risk genes in cortical excitatory neurons (56), with enrichment also observed in embryonic and/or fetal progenitor cells and adult cortical interneurons.

Analysis of gene coexpression modules found that genes in the more-stringent early-onset disease (ADHD, SCZ, and MDD) risk lists converged on 7 of 73 coexpression modules, whereas adult-onset disease (AD and PD) risk-gene lists converged on five partially overlapping modules (fig. S37 and table S12). Eight of these 10 total disease-associated modules (Fig. 7A) exhibited spatiotemporal or temporal specificity, and all modules exhibited their greatest spatiotemporal change during either W2 or W5 (fig. S37). A significant number of modules associated with adult-onset disorders were enriched for signatures of glial gene expression ($P < 0.0266$, hypergeometric test, table S12), and of particular interest were modules ME3 and ME7, which, in addition to glial signatures, were enriched for non-NUM sites, adult-active enhancers, sex-DEX genes, and AD-associated risk genes (Fig. 7A).

Another module of interest was ME37, a module of 145 genes enriched for NUM sites and fetal enhancers and whose expression was enriched specifically in neurons as opposed to neural progenitors or glia. ME37 was also exceptional for its disease association, as it was enriched for genes associated with SCZ, IQ, and neuroticism but not for non-neurological characteristics such as height or a HBA1C-related trait (Fig. 7A). Complementary module-based association analysis with Multi-marker Analysis of GenoMic Annotation (MAGMA), which tested for an enrichment in association to disease specifically around genes in any given module, confirmed enrichment for SCZ, IQ, and neuroticism in ME37 [MAGMA *P* values < 0.01; the false discovery rate (FDR) for all traits and modules was <0.3] (table S11). At the gene level, multiple genes in ME37 identified using our less stringent criteria for interaction were associated with up to four or more different traits and disorders, including *MEF2C*, *ZNF184*, *TCF4*, and *SATB2*, all genes critical for neurodevelopment and/or implicated in neurodevelopmental disorders (57–65) (Fig. 7, B and C). We also found that ME37 was specifically enriched in clusters of excitatory neurons in the fetal and adult NCX (Fig. 7D), and further analysis of adult excitatory neuron populations identified in this study and an independent database of adult single nucleus data (27) suggested that this enrichment was selective for deep-layer neocortical neurons (fig. S43).

As the ASD GWAS resulted in only 13 significant genes, eight of which were non-protein coding, and because de novo germline mutations are known to contribute to ASD risk (66), we next developed two nonoverlapping lists of neurodevelopmental disorders (NDDs) [ASD, intellectual disability (ID), and developmental delay (DD)]. The first list was comprised of 65 high-confidence ASD risk genes (hcASD) associated with de novo mutations (66). The second list included all ASD genes documented in the SFARI database (http://gene.sfari.org) under categories "syndromic" or with scores from 1 to 4, as well as an independent list of genes associated with DD (67), with genes overlapping the hcASD list removed. We found that these genes were also significantly enriched in ME37 (FDR < 0.0001, Fisher's exact test), and, commensurate with the cell-type enrichment found in ME37, the expression of genes in both of these lists was also enriched in several clusters of fetal and adult excitatory neurons identified in our single-cell dataset (Fig. 7D). Medium spiny neurons in the STR, a population that has also been previously linked to ASD (68), were also enriched for the expression of ASD risk genes in the prenatal macaque (Fig. 7D).

We finally studied the overlap between WGCNA modules and modules significantly enriched in differentially expressed genes in postmortem brains from patients of SCZ, BD, and ASD (69). Interestingly, we found little overlap between modules enriched in genes exhibiting postmortem differences in expression between SCZ, BD, or ASD, as compared with neurotypical controls, and modules enriched in GWAS risk genes for these same disorders ($P > 0.05$, hypergeometric test) (fig. S37). Emphasizing the necessity of studying neurotypical brain development, these observations may suggest a decoupling between the primary genetic causes of some neurological or psychiatric disorders and second-order effects manifesting as changes in gene expression months or years after disease onset.

Discussion

In this study, we have presented a comprehensive dataset and a multiplatform functional genomic analysis of the developing and adult human brain. The presence of these multiple data modalities in a unified resource, and largely from the same

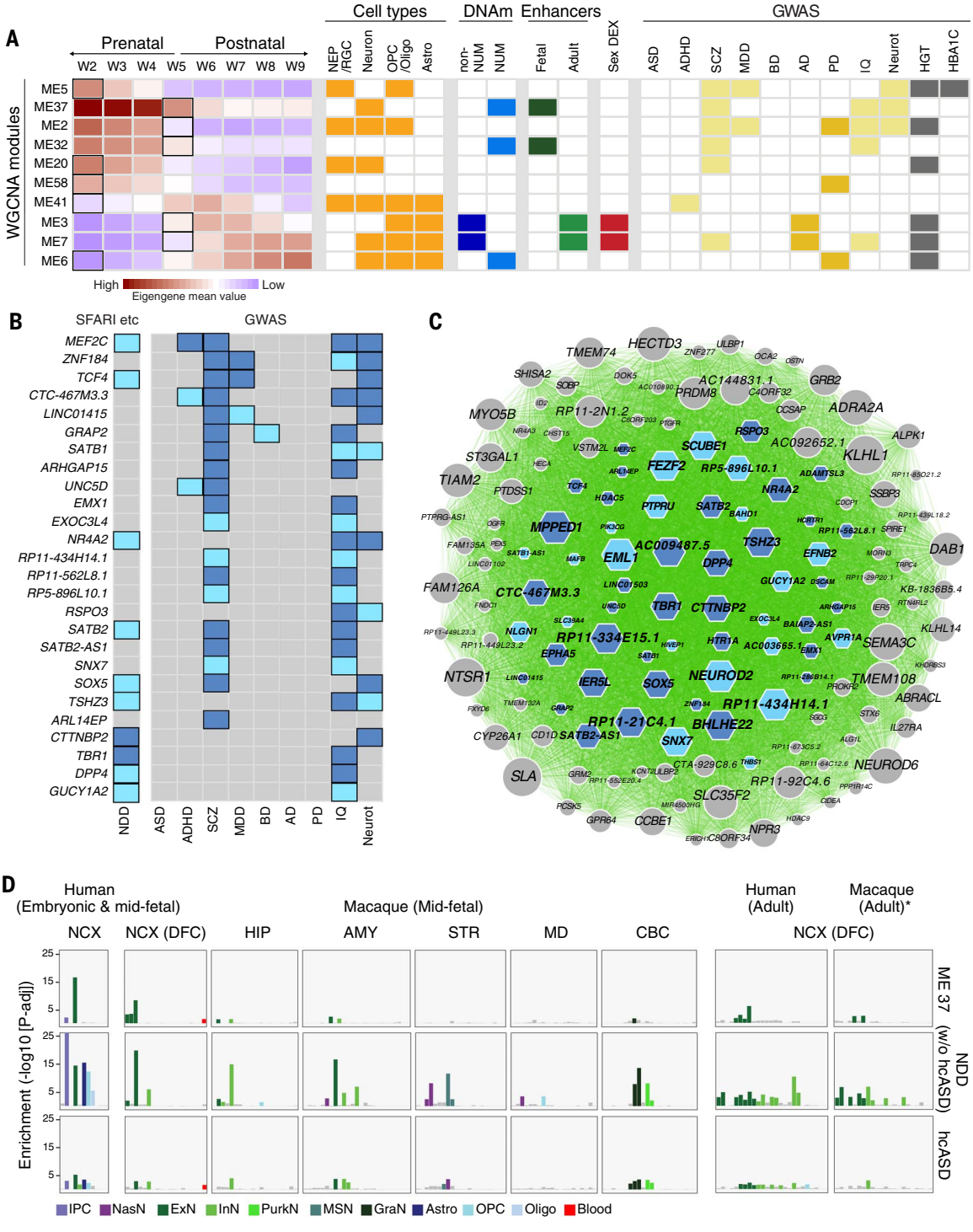


Fig. 7. Convergence of risk for brain-based traits and disorders on discrete coexpression modules and cell types. (**A**) Genes associated with disease risk (right; light yellow indicates neuropsychiatric disorder or brain-based trait, and dark yellow indicates adult-onset disorder) were identified by integrating GWAS, Hi-C, and H3K27ac data and converged on 10 WGCNA modules. Many of these modules exhibited dynamic expression across time; the bold rectangles in the left panel indicate the windows with greatest rate of change. Many were also enriched for gene expression associated with distinct cell types (orange), putative active enhancers (green), and/or sites undermethylated in NeuN-positive (NUM) or NeuN-negative cells (blue, non-NUM). (**B**) Schematic highlighting genes in ME37 that were implicated by our study in multiple neuropsychiatric disorders (ADHD, SCZ, MDD, or BD) and neurological traits (IQ or Neurot) (list 1, light blue; list 2, dark blue), as well as neurodevelopmental disorder (NDD) risk genes, including two independent lists of high-confidence risk genes associated with ASD through de novo mutations or copy number variants [dark blue, (66)] as well as ASD risk genes identified from the SFARI dataset (light blue, http://gene.sfari.org) or for developmental delay (67). Genes implicated in only a single disorder or trait are not shown in this panel. (**C**) Network representation of ME37 showing connectivity between genes based on Pearson correlation. Genes linked to NDDs or neurological characteristics in our study are indicated using either dark blue–shaded or light blue–shaded hexagons, as in (B). The size of a given hexagon (or circle, indicating no association in this study) is proportional to the degree of each gene under a minimum correlation value of 0.7. (**D**) Enrichment for genes in ME37 or two lists of ASD risk genes among the fetal and adult cell types we identified from human NCX and multiple regions of the macaque (34) brain. For graphical representation, $\log_{10} P$ values are capped at 25. *Adult macaque cells were classified into human adult clusters using Random Forest. NEP/RGC, neural epithelial progenitor/radial glial lineage; MSN, medium spiny neurons; NasN, nascent neurons; GraN, granule neurons; PurkN, Purkinje neurons; IPC, intermediate progenitor cells; OPC, oligodendrocyte progenitor cells.

tissue samples, allows the integration of information spanning prenatal and postnatal human brain development. Resource description and access are available at development.psychencode.org and www.brainspan.org.

Although transcriptomic differences between distinct brain regions remain across time, they are developmentally specified and exhibit an overall cup-shaped pattern centered on a late fetal transition after a period of high intra- and interregional variation during embryonic and early or mid-fetal development. Multiple analyses of distinct transcriptomic features all confirm this transition begins well before birth. Our complementary transcriptomic study of the developing rhesus macaque brain (*34*) also revealed a similar global developmental pattern, with a first transition beginning before birth, indicating that this is a conserved feature of catarrhine primate neurodevelopment and not due to an artifact resulting from difficulties acquiring samples from late fetal and early postnatal development. Such a phenomenon is consistent with previously observed differences in transcriptomic and methylomic profiles of mid-fetal and postnatal human NCX (*17*–*20*) and coincident with processes involved in region-specific cell type generation, differentiation, and maturation (*2*). Crucially, this transition is notably distinct from previously reported phylogenetic hourglass-like patterns that occur during the embryonic organogenetic period in several invertebrate and vertebrate species (*70*, *71*). Moreover, the developmental (ontogenetic) cup-shaped pattern we observe coincides with an "evolutionary" (phylogenetic) cup-shaped pattern, in which developmental periods exhibiting high levels of interregional differences (for example, early to mid-fetal periods) also exhibit less conservation in gene expression patterns between human and macaque (*34*).

Among the processes that become prominent during the late fetal period are astrogliogenesis, synaptogenesis, dendritogenesis, and neuronal activity. In contrast to a previous report of robust areal differences in the progression of synaptogenesis during the same time period in humans (*36*), this and an accompanying study (*34*) found that genes associated with these processes exhibit largely synchronous expression trajectories across the developing NCX in both humans and macaque. However, myelination—which sharply increases during late fetal development, peaks after birth, and extends through childhood and adolescence (*72*)—is temporally asynchronous. This asynchronicity in oligodendrocyte development and myelination is not apparent at the level of oligodendrocyte progenitor cells (OPCs), which suggests that the maturation of OPCs into myelinating oligodendrocytes is a process with a variable onset and pace across areas. Similar observations were made in macaque (*34*), indicating that this may be another conserved catarrhine feature.

Transcriptomic variation may reflect several distinct cellular and maturational reorganizational events. For example, as first described by Brodmann (*73*), an ontogenetic six-layered Grundtypus foreshadows the adult NCX and transiently transforms the entirety of the neocortical plate beginning in the late fetal period, or in our W5. Furthermore, consistent with the extensive changes we observed in the cerebellar transcriptome during late fetal development and early postnatal ages, cerebellar granule cells, a cell type that represents about two-thirds of all neurons in the brain, are also generated predominately during this period (*74*). The late fetal transition may therefore follow an inflection point after which developmental and spatiotemporal transcriptomic variations are transiently consolidated in advance of the emergence of cellular and functional differences between adult brain regions.

The mid-fetal period of high intra- and interregional divergence that immediately precedes the late fetal transition also coincides with a key developmental period previously associated with the etiology of ASD and SCZ (*63*, *65*, *75*). Consequently, understanding the developmental and evolutionary history of this period may be essential for understanding neuropsychiatric disease. Integrating our multiple data modalities with gene coexpression modules allowed us to organize and characterize the whole-brain developmental transcriptome and identify modules with dynamic spatiotemporal trajectories, many of them showing a sharp late fetal transition, and enrichment in specific cell types, epigenetic activity, and disease-associated genes. Of particular interest is ME37, a module displaying the greatest rate of change in the NCX within the late fetal transition and in which putative risk genes for ASD, NDD, SCZ, IQ, and neuroticism converged. Several of the genes in ME37 were implicated by our study in multiple disorders and traits and have been linked previously to neurodevelopment and human disease. For example, *MEF2C* controls activity-dependent expression of neuronal genes, including those linked to synapse function and ASD (*61*, *63*), and *Mef2c*-mutant mice display numerous behaviors reminiscent of ASD, ID, and SCZ (*58*). Similarly, *TCF4* regulates key neurodevelopmental processes, such as neurogenesis and synaptic plasticity, DNA methylation, and memory function processes (*62*, *64*). Moreover, mutations in both *MEF2C* and *TCF4* result in intellectual disability in humans (*57*, *59*, *60*). Numerous other genes in this module are similarly involved in neurodevelopment, have been implicated in human brain disease, and are highly plausible disease-risk genes and potentially therapeutic candidates. For example, *NR4A2*, a gene encoding another transcription factor in ME37 that we linked to neuroticism and IQ, has been linked to ASD and SCZ, among other disorders. Our study also links the gene for the transcription factor *TSHZ3* to neuroticism and IQ, and previous efforts have linked murine *Tshz3* to ASD and the fetal development of cortical excitatory projection neurons (*76*), a cell type and developmental period also implicated in ASD (*63*, *65*). Other genes in ME37, such as *SATB2*, *FEZF2*, *SOX5*, and *TBR1*, play critical roles in the development of cortical excitatory projection neurons and are mutated in NDDs (*29*–*31*, *65*, *77*, *78*). Similarly, the population of genes included in ME37, as well as genes linked to ASD and NDD, also exhibit regional and cell type–specific convergence in neocortical excitatory neurons. Moreover, the identification of ME37 and the overlap of genes in this module with those implicated in ASD and NDD illustrates how disease-association signals from common variants unveiled by GWAS for any given neuropsychiatric disorder can identify genes that have also been associated with the etiology of a different disease through the study of de novo mutations in patient populations (*76*). Although not every gene in ME37 is likely to contribute to neuropsychiatric disease etiology, the coincident enrichment within this module of genes associated with multiple disorders or neurological traits, along with the multitude of genes in this module that are associated directly, suggests that neuropsychiatric disease might be considered through a broader lens encompassing additional aspects of brain dysfunction.

Interestingly, there is little overlap between the risk gene–associated modules we identified and modules enriched in genes that are differentially expressed in postmortem brains of SCZ, ASD, and BD, as compared to controls (*69*). This comparison may help discriminate gene networks that are primary causes from those that are secondary or reactive in these neuropsychiatric disorders while emphasizing the importance of studying disease in the context of neurotypical development.

Taken together, these observations demonstrate the utility of this resource to perform integrated analysis for the understanding of brain development and function and for the rapid interpretation of findings from neuropsychiatric genomics.

Materials and methods summary

A full description of the materials and methods is available in the supplementary materials. Briefly, we precisely dissected multiple brain regions (HIP, STR, AMY, cerebellum, thalamus, and 11 neocortical areas) in more than 60 postmortem human brains ranging in age from 5 PCW to 64 PY. We then applied bulk tissue RNA-seq, scRNA-seq and snRNA-seq, smRNA-seq, DNA methylation assay, or ChIP-seq to generate multimodal datasets, often from the same brain. After applying stringent quality control checks and independent analysis of each dataset, we performed integrated analyses to gain insights into human brain development, function, and disease.

REFERENCES AND NOTES

1. E. S. Lein, T. G. Belgard, M. Hawrylycz, Z. Molnár, Transcriptomic perspectives on neocortical structure, development, evolution, and disease. *Annu. Rev. Neurosci.* **40**, 629–652 (2017). doi: 10.1146/annurev-neuro-070815-013858; pmid: 28661727
2. J. C. Silbereis, S. Pochareddy, Y. Zhu, M. Li, N. Sestan, The cellular and molecular landscapes of the developing human central nervous system. *Neuron* **89**, 248–268 (2016). doi: 10.1016/j.neuron.2015.12.008; pmid: 26796689

3. D. H. Geschwind, P. Rakic, Cortical evolution: Judge the brain by its cover. *Neuron* **80**, 633–647 (2013). doi: 10.1016/j.neuron.2013.10.045; pmid: 24183016
4. J. H. Lui, D. V. Hansen, A. R. Kriegstein, Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011). doi: 10.1016/j.cell.2011.06.030; pmid: 21729779
5. T. Paus, M. Keshavan, J. N. Giedd, Why do many psychiatric disorders emerge during adolescence? *Nat. Rev. Neurosci.* **9**, 947–957 (2008). doi: 10.1038/nrn2513; pmid: 19002191
6. S. A. McCarroll, S. E. Hyman, Progress in the genetics of polygenic brain disorders: Significant new challenges for neurobiology. *Neuron* **80**, 578–587 (2013). doi: 10.1016/j.neuron.2013.10.046; pmid: 24183011
7. B. I. Bae, D. Jayaraman, C. A. Walsh, Genetic changes shaping the human brain. *Dev. Cell* **32**, 423–434 (2015). doi: 10.1016/j.devcel.2015.01.035; pmid: 25710529
8. J. M. Keil, A. Qalieh, K. Y. Kwan, Brain transcriptome databases: A user's guide. *J. Neurosci.* **38**, 2399–2412 (2018). doi: 10.1523/JNEUROSCI.1930-17.2018; pmid: 29437890
9. S. Darmanis *et al*., A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7285–7290 (2015). doi: 10.1073/pnas.1507125112; pmid: 26060301
10. M. B. Johnson *et al*., Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009). doi: 10.1016/j.neuron.2009.03.027; pmid: 19477152
11. J. A. Miller *et al*., Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014). doi: 10.1038/nature13185; pmid: 24695229
12. 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
13. L. de la Torre-Ubieta *et al*., The dynamic landscape of open chromatin during human cortical neurogenesis. *Cell* **172**, 289–304.e18 (2018). doi: 10.1016/j.cell.2017.12.014; pmid: 29307494
14. R. N. Doan *et al*., Mutations in human accelerated regions disrupt cognition and social behavior. *Cell* **167**, 341–354.e12 (2016). doi: 10.1016/j.cell.2016.08.071; pmid: 27667684
15. 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
16. M. B. Johnson *et al*., Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nat. Neurosci.* **18**, 637–646 (2015). doi: 10.1038/nn.3980; pmid: 25734491
17. R. Lister *et al*., Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013). doi: 10.1126/science.1237905; pmid: 23828890
18. C. Colantuoni *et al*., Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519–523 (2011). doi: 10.1038/nature10524; pmid: 22031444
19. H. J. Kang *et al*., Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011). doi: 10.1038/nature10523; pmid: 22031440
20. M. Pletikos *et al*., Temporal specification and bilaterality of human neocortical topographic gene expression. *Neuron* **81**, 321–332 (2014). doi: 10.1016/j.neuron.2013.11.018; pmid: 24373884
21. S. K. Reilly *et al*., Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015). doi: 10.1126/science.1260943; pmid: 25745175
22. Materials and methods are available as supplementary materials.
23. Z. Lin *et al*., Simultaneous dimension reduction and adjustment for confounding variation. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14662–14667 (2016). doi: 10.1073/pnas.1617317113; pmid: 27930330
24. K. G. Ardlie *et al*., The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015). doi: 10.1126/science.1262110; pmid: 25954001
25. B. B. Lake *et al*., Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018). doi: 10.1038/nbt.4038; pmid: 29227469
26. S. Zhong *et al*., A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018). doi: 10.1038/nature25980; pmid: 29539641
27. B. B. Lake *et al*., Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016). doi: 10.1126/science.aaf1204; pmid: 27339989
28. D. Mi *et al*., Early emergence of cortical interneuron diversity in the mouse embryo. *Science* **360**, 81–85 (2018). doi: 10.1126/science.aar6821; pmid: 29472441
29. S. Lodato, P. Arlotta, Generating neuronal diversity in the mammalian cerebral cortex. *Annu. Rev. Cell Dev. Biol.* **31**, 699–720 (2015). doi: 10.1146/annurev-cellbio-100814-125353; pmid: 26359774
30. A. S. Nord, K. Pattabiraman, A. Visel, J. L. R. Rubenstein, Genomic perspectives of transcriptional regulation in forebrain development. *Neuron* **85**, 27–47 (2015). doi: 10.1016/j.neuron.2014.11.011; pmid: 25569346
31. M. Shibata, F. O. Gulden, N. Sestan, From trans to cis: Transcriptional regulatory networks in neocortical development. *Trends Genet.* **31**, 77–87 (2015). doi: 10.1016/j.tig.2014.12.004; pmid: 25624274
32. G. D'Arcangelo *et al*., A protein related to extracellular matrix proteins deleted in the mouse mutant *reeler*. *Nature* **374**, 719–723 (1995). doi: 10.1038/374719a0; pmid: 7715726
33. J. J. Rowell, A. K. Mallik, J. Dugas-Ford, C. W. Ragsdale, Molecular analysis of neocortical layer structure in the ferret. *J. Comp. Neurol.* **518**, 3272–3289 (2010). doi: 10.1002/cne.22399; pmid: 20575059
34. Y. Zhu *et al*., Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, eaat8077 (2018). doi: 10.1126/science.aat8077
35. D. J. Miller *et al*., Prolonged myelination in human neocortical evolution. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16480–16485 (2012). doi: 10.1073/pnas.1117943109; pmid: 23012402
36. P. R. Huttenlocher, A. S. Dabholkar, Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* **387**, 167–178 (1997). doi: 10.1002/(SICI)1096-9861(19971020)387:2<167:AID-CNE1>3.0.CO;2-Z; pmid: 9336221
37. A. F. Pardiñas *et al*., Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018). doi: 10.1038/s41588-018-0059-2; pmid: 29483656
38. J. C. Lambert *et al*., Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013). doi: 10.1038/ng.2802; pmid: 24162737
39. M. A. Nalls *et al*., Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014). doi: 10.1038/ng.3043; pmid: 25064009
40. J. Grove *et al*., ASD Working Group of the PGC; BUPGEN; Major Depressive Disorder Working Group of the PGC; 23andMe Research Team, Common risk variants identified in autism spectrum disorder. bioRxiv 224774 [Preprint]. 27 November 2017. doi: 10.1101/224774
41. D. Demontis *et al*., ADHD Working Group of the Psychiatric Genomics Con; Early Lifecourse & Genetic Epidemiology (EAGLE); 23andMe Research Team, Discovery of the first genome-wide significant risk loci for ADHD. bioRxiv 145581 [Preprint]. 3 June 2017. doi: 10.1101/145581
42. N. R. Wray *et al*., Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018). doi: 10.1038/s41588-018-0090-3; pmid: 29700475
43. D. M. Ruderfer *et al*., Genomic dissection of bipolar disorder and schizophrenia, including 28 subphenotypes. *Cell* **173**, 1705–1715.e16 (2018). doi: 10.1016/j.cell.2018.05.046; pmid: 29906448
44. J. E. Savage *et al*., Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat. Genet.* **50**, 912–919 (2018). doi: 10.1038/s41588-018-0152-6; pmid: 29942086
45. M. Luciano *et al*., Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. *Nat. Genet.* **50**, 6–11 (2018). doi: 10.1038/s41588-017-0013-8; pmid: 29255261
46. A. R. Wood *et al*., Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173–1186 (2014). doi: 10.1038/ng.3097; pmid: 25282103
47. J. Z. Liu *et al*., Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015). doi: 10.1038/ng.3359; pmid: 26192919
48. C. J. Willer *et al*., Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013). doi: 10.1038/ng.2797; pmid: 24097068
49. E. Wheeler *et al*., Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLOS Med.* **14**, e1002383 (2017). doi: 10.1371/journal.pmed.1002383; pmid: 28898252
50. 1000 Genomes Project Consortium, A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010). doi: 10.1038/nature09534; pmid: 20981092
51. P. M. Giusti-Rodriguez, P. F. Sullivan, Schizophrenia and a high-resolution map of the three-dimensional chromatin interactome of adult and fetal cortex. *bioRxiv* (2018).
52. D. Wang *et al*., Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018). doi: 10.1126/science.aat8464
53. H. Won *et al*., Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016). doi: 10.1038/nature19847; pmid: 27760116
54. J. P. Hamilton, M. Siemer, I. H. Gotlib, Amygdala volume in major depressive disorder: A meta-analysis of magnetic resonance imaging studies. *Mol. Psychiatry* **13**, 993–1000 (2008). doi: 10.1038/mp.2008.57; pmid: 18504424
55. M. J. Rubinow *et al*., Basolateral amygdala volume and cell numbers in major depressive disorder: A postmortem stereological study. *Brain Struct. Funct.* **221**, 171–184 (2016). doi: 10.1007/s00429-014-0900-z; pmid: 25287512
56. N. G. Skene *et al*., Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018). doi: 10.1038/s41588-018-0129-5; pmid: 29785013
57. J. Amiel *et al*., Mutations in *TCF4*, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt-Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am. J. Hum. Genet.* **80**, 988–993 (2007). doi: 10.1086/515582; pmid: 17436254
58. A. J. Harrington *et al*., MEF2C regulates cortical inhibitory and excitatory synapses and behaviors relevant to neurodevelopmental disorders. *eLife* **5**, e20059 (2016). doi: 10.7554/eLife.20059; pmid: 27779093
59. C. Zweier *et al*., Haploinsufficiency of *TCF4* causes syndromal mental retardation with intermittent hyperventilation (Pitt-Hopkins syndrome). *Am. J. Hum. Genet.* **80**, 994–1001 (2007). doi: 10.1086/515583; pmid: 17436255
60. M. Zweier *et al*., Mutations in *MEF2C* from the 5q14.3q15 microdeletion syndrome region are a frequent cause of severe mental retardation and diminish *MECP2* and *CDKL5* expression. *Hum. Mutat.* **31**, 722–733 (2010). doi: 10.1002/humu.21253; pmid: 20513142
61. D. H. Ebert, M. E. Greenberg, Activity-dependent neuronal signalling and autism spectrum disorder. *Nature* **493**, 327–337 (2013). doi: 10.1038/nature11860; pmid: 23325215
62. A. J. Kennedy *et al*., *Tcf4* regulates synaptic plasticity, DNA methylation, and memory function. *Cell Reports* **16**, 2666–2685 (2016). doi: 10.1016/j.celrep.2016.08.004; pmid: 27568567
63. N. N. Parikshak *et al*., Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013). doi: 10.1016/j.cell.2013.10.031; pmid: 24267887
64. U. Schmidt-Edelkraut, G. Daniel, A. Hoffmann, D. Spengler, Zac1 regulates cell cycle arrest in neuronal progenitors via Tcf4. *Mol. Cell. Biol.* **34**, 1020–1030 (2014). doi: 10.1128/MCB.01195-13; pmid: 24396065
65. A. J. Willsey *et al*., Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013). doi: 10.1016/j.cell.2013.10.020; pmid: 24267886
66. S. J. Sanders *et al*., Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. *Neuron* **87**, 1215–1233 (2015). doi: 10.1016/j.neuron.2015.09.016; pmid: 26402605
67. Deciphering Developmental Disorders Study, Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017). doi: 10.1038/nature21062
68. X. Xu, A. B. Wells, D. R. O'Brien, A. Nehorai, J. D. Dougherty, Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J. Neurosci.* **34**, 1420–1431 (2014). doi: 10.1523/JNEUROSCI.4488-13.2014; pmid: 24453331
69. M. J. Gandal *et al*., Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018). doi: 10.1126/science.aad6469; pmid: 29439242

70. A. T. Kalinka *et al*., Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814 (2010). doi: 10.1038/nature09634; pmid: 21150996
71. T. Domazet-Lošo, D. Tautz, A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815–818 (2010). doi: 10.1038/nature09632; pmid: 21150997
72. P. R. Huttenlocher, Synaptic density in human frontal cortex—Developmental changes and effects of aging. *Brain Res.* **163**, 195–205 (1979). doi: 10.1016/0006-8993(79)90349-4; pmid: 427544
73. K. Brodmann, *Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues* (Barth, 1909).
74. M. C. Kiessling *et al*., Cerebellar granule cells are generated postnatally in humans. *Brain Struct. Funct.* **219**, 1271–1286 (2014). doi: 10.1007/s00429-013-0565-z; pmid: 23716277
75. S. Gulsuner *et al*., Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013). doi: 10.1016/j.cell.2013.06.049; pmid: 23911319
76. X. Caubit *et al*., *TSHZ3* deletion causes an autism syndrome and defects in cortical projection neurons. *Nat. Genet.* **48**, 1359–1369 (2016). doi: 10.1038/ng.3681; pmid: 27668656
77. K. Y. Kwan, Transcriptional dysregulation of neocortical circuit assembly in ASD. *Int. Rev. Neurobiol.* **113**, 167–205 (2013). doi: 10.1016/B978-0-12-418700-9.00006-X; pmid: 24290386
78. A. N. Lamb *et al*., Haploinsufficiency of *SOX5* at 12p12.1 is associated with developmental delays with prominent language delay, behavior problems, and mild dysmorphic features. *Hum. Mutat.* **33**, 728–740 (2012). doi: 10.1002/humu.22037; pmid: 22290657

ACKNOWLEDGMENTS

We are grateful to the many individuals in our laboratories and at our institutions who have provided support and contributed to these projects. In addition, we would like to thank the National Institute of Mental Health (NIMH), and, in particular, A. Beckel-Mitchener, M. Freund, T. Lerner, and G. Senthil, for providing invaluable institutional support and guidance for these projects. **Funding:** Data was generated as part of the PsychENCODE Consortium, supported by NIMH grants U01MH103392, U01MH103365, U01MH103346, U01MH103340, U01MH103339, R21MH109956, R21MH105881, R21MH105853, R21MH103877, R21MH102791, R01MH111721, R01MH110928, R01MH110927, R01MH110926, R01MH110921, R01MH110920, R01MH110905, R01MH109715, R01MH109677, R01MH105898, R01MH105898, R01MH094714, R01MH109901, P50MH106934, and 5R24HD000836 awarded to S. Akbarian (Icahn School of Medicine at Mount Sinai), G. Crawford (Duke University), S. Dracheva (Icahn School of Medicine at Mount Sinai), P. Farnham (University of Southern California), M.B.G. (Yale University), D.H.G. (University of California, Los Angeles), F. Goes (Johns Hopkins University), T. M. Hyde (Lieber Institute for Brain Development), A. Jaffe (Lieber Institute for Brain Development), J.A.K. (University of Southern California), C. Liu (SUNY Upstate Medical University), D. Pinto (Icahn School of Medicine at Mount Sinai), P. Roussos (Icahn School of Medicine at Mount Sinai), S.J.S. (University of California, San Francisco), N.S. (Yale University), P. Sklar (Icahn School of Medicine at Mount Sinai), M.W.S. (University of California, San Francisco), P.F.S. (University of North Carolina), F. Vaccarino (Yale University), D.R.W. (Lieber Institute for Brain Development), S. Weissman (Yale University), K. White (University of Chicago), J. Willsey (University of California, San Francisco), and P. Zandi (Johns Hopkins University). The BrainSpan Project Consortium was supported by grants MH089929, MH090047, and MH089921 from NIMH. Additional support was provided by the Kavli Foundation, the James S. McDonnell Foundation, the Beatriu de Pinós program (BP-DGR 2014 to B.L.-G.), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (5R24HD000836). S.J.S. and D.M.W. were supported by a grant from the Simons Foundation (SFARI #307705) and the Autism Science Foundation (Postdoctoral Fellowship #16-009), respectively. The human embryonic and some of the fetal material was provided by the Joint MRC–Wellcome Trust (MR/R006237/1) Human Developmental Biology Resource (www.hdbr.org). This work also received funding from The Netherlands Organization for Scientific Research (NWO VICI 435-14-005), Sophia Foundation for Scientific Research. Part of the analyses were carried out on the Genetic Cluster Computer, which is financed by the Netherlands Scientific Organization (NWO, 480-05-003), by the VU University, Amsterdam, Netherlands, and by the Dutch Brain Foundation, and is hosted by the Dutch National Computing and Networking Services SurfSARA. **Author contributions:** Specific contributions to the work are provided in the annotated author list for the consortia in the supplementary materials. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Scripts used in this manuscript can be found at https://github.com/GSantpere/PEC_DEV/tree/master/Code. scRNA-seq and snRNA-seq data have been deposited at http://psychencode.org. Tissue RNA-seq, ChIP-seq, DNA methylation, and SNP array data have been deposited in dbGAP under accession number phs000755. All processed files and scripts pertaining to this manuscript are available on http://development.psychencode.org.

SUPPLEMENTARY MATERIALS

2 April 2018; accepted 15 November 2018
10.1126/science.aat7615

RESEARCH ARTICLE SUMMARY

PSYCHIATRIC GENOMICS

Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder

Michael J. Gandal*, Pan Zhang, Evi Hadjimichael, Rebecca L. Walker, Chao Chen, Shuang Liu, Hyejung Won, Harm van Bakel, Merina Varghese, Yongjun Wang, Annie W. Shieh, Jillian Haney, Sepideh Parhami, Judson Belmont, Minsoo Kim, Patricia Moran Losada, Zenab Khan, Justyna Mleczko, Yan Xia, Rujia Dai, Daifeng Wang, Yucheng T. Yang, Min Xu, Kenneth Fish, Patrick R. Hof, Jonathan Warrell, Dominic Fitzgerald, Kevin White, Andrew E. Jaffe, PsychENCODE Consortium†, Mette A. Peters, Mark Gerstein, Chunyu Liu*, Lilia M. Iakoucheva*, Dalila Pinto*, Daniel H. Geschwind*

INTRODUCTION: Our understanding of the pathophysiology of psychiatric disorders, including autism spectrum disorder (ASD), schizophrenia (SCZ), and bipolar disorder (BD), lags behind other fields of medicine. The diagnosis and study of these disorders currently depend on behavioral, symptomatic characterization. Defining genetic contributions to disease risk allows for biological, mechanistic understanding but is challenged by genetic complexity, polygenicity, and the lack of a cohesive neurobiological model to interpret findings.

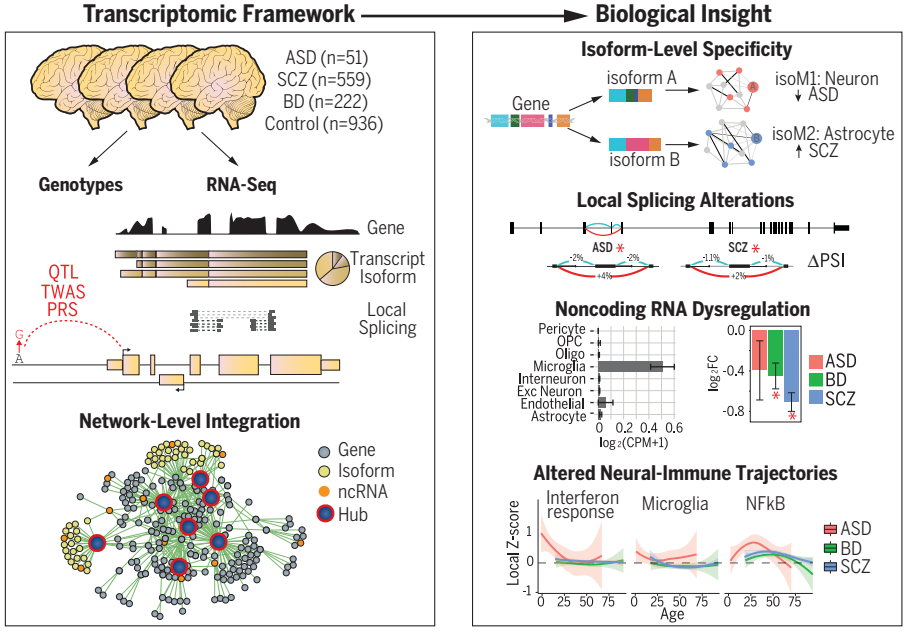
RATIONALE: The transcriptome represents a quantitative phenotype that provides biological context for understanding the molecular pathways disrupted in major psychiatric disorders. RNA sequencing (RNA-seq) in a large cohort of cases and controls can advance our knowledge of the biology disrupted in each disorder and provide a foundational resource for integration with genomic and genetic data.

RESULTS: Analysis across multiple levels of transcriptomic organization—gene expression, local splicing, transcript isoform expression, and coexpression networks for both protein-coding and noncoding genes—provides an in-depth view of ASD, SCZ, and BD molecular pathology. More than 25% of the transcriptome exhibits differential splicing or expression in at least one disorder, including hundreds of noncoding RNAs (ncRNAs), most of which have unexplored functions but collectively exhibit patterns of selective constraint. Changes at the isoform level, as opposed to the gene level, show the largest effect sizes and genetic enrichment and the greatest disease specificity. We identified coexpression modules associated with each disorder, many with enrichment for cell type–specific markers, and several modules significantly dysregulated across all three disorders. These enabled parsing of down-regulated neuronal and synaptic components into a variety of cell type– and disease-specific signals, including multiple excitatory neuron and distinct interneuron modules with differential patterns of disease association, as well as common and rare genetic risk variant enrichment. The glial-immune signal demonstrates shared disruption of the blood-brain barrier and upregulation of NFkB-associated genes, as well as disease-specific alterations in microglial-, astrocyte-, and interferon-response modules. A coexpression module associated with psychiatric medication exposure in SCZ and BD was enriched for activity-dependent immediate early gene pathways. To identify causal drivers, we integrated polygenic risk scores and performed a transcriptome-wide association study and summary-data–based Mendelian randomization. Candidate risk genes—5 in ASD, 11 in BD, and 64 in SCZ, including shared genes between SCZ and BD—are supported by multiple methods. These analyses begin to define a mechanistic basis for the composite activity of genetic risk variants.

CONCLUSION: Integration of RNA-seq and genetic data from ASD, SCZ, and BD provides a quantitative, genome-wide resource for mechanistic insight and therapeutic development at Resource.PsychENCODE.org. These data inform the molecular pathways and cell types involved, emphasizing the importance of splicing and isoform-level gene regulatory mechanisms in defining cell type and disease specificity, and, when integrated with genome-wide association studies, permit the discovery of candidate risk genes. ■

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat8127



The PsychENCODE cross-disorder transcriptomic resource. Human brain RNA-seq was integrated with genotypes across individuals with ASD, SCZ, BD, and controls, identifying pervasive dysregulation, including protein-coding, noncoding, splicing, and isoform-level changes. Systems-level and integrative genomic analyses prioritize previously unknown neurogenetic mechanisms and provide insight into the molecular neuropathology of these disorders.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: mgandal@mednet.ucla.edu (M.J.G.); liuch@upstate.edu (C.L.); lilyak@ucsd.edu (L.M.I.); dalila.pinto@mssm.edu (D.P.); dhg@mednet.ucla.edu (D.H.G.)
†PsychENCODE Consortium authors and affiliations are listed in the supplementary materials.
Cite this article as M. J. Gandal *et al.*, *Science* **362**, eaat8127 (2018). DOI: 10.1126/science.aat8127

RESEARCH ARTICLE

PSYCHIATRIC GENOMICS

Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder

Michael J. Gandal[1,2,3,4]*, Pan Zhang[5], Evi Hadjimichael[6,7,8,9], Rebecca L. Walker[2,3,4], Chao Chen[10,11], Shuang Liu[12], Hyejung Won[2,3,4,13,14], Harm van Bakel[7], Merina Varghese[9,15], Yongjun Wang[16], Annie W. Shieh[17], Jillian Haney[1,2,3], Sepideh Parhami[1,2,3], Judson Belmont[6,7,8,9], Minsoo Kim[1,4], Patricia Moran Losada[5], Zenab Khan[7], Justyna Mleczko[18], Yan Xia[10,17], Rujia Dai[10,17], Daifeng Wang[19], Yucheng T. Yang[12], Min Xu[12], Kenneth Fish[18], Patrick R. Hof[9,15,20], Jonathan Warrell[12], Dominic Fitzgerald[21], Kevin White[21,22,23], Andrew E. Jaffe[24,25], PsychENCODE Consortium†, Mette A. Peters[26], Mark Gerstein[12], Chunyu Liu[10,17,27]*, Lilia M. Iakoucheva[5]*, Dalila Pinto[6,7,8,9]*, Daniel H. Geschwind[1,2,3,4]*

Most genetic risk for psychiatric disease lies in regulatory regions, implicating pathogenic dysregulation of gene expression and splicing. However, comprehensive assessments of transcriptomic organization in diseased brains are limited. In this work, we integrated genotypes and RNA sequencing in brain samples from 1695 individuals with autism spectrum disorder (ASD), schizophrenia, and bipolar disorder, as well as controls. More than 25% of the transcriptome exhibits differential splicing or expression, with isoform-level changes capturing the largest disease effects and genetic enrichments. Coexpression networks isolate disease-specific neuronal alterations, as well as microglial, astrocyte, and interferon-response modules defining previously unidentified neural-immune mechanisms. We integrated genetic and genomic data to perform a transcriptome-wide association study, prioritizing disease loci likely mediated by cis effects on brain expression. This transcriptome-wide characterization of the molecular pathology across three major psychiatric disorders provides a comprehensive resource for mechanistic insight and therapeutic development.

Developing more-effective treatments for autism spectrum disorder (ASD), schizophrenia (SCZ), and bipolar disorder (BD), three common psychiatric disorders that confer lifelong disability, is a major international public health priority (*1*). Studies have identified hundreds of causal genetic variants robustly associated with these disorders and thousands more that likely contribute to their pathogenesis (*2*). However, the neurobiological mechanisms through which genetic variation imparts risk, both individually and in aggregate, are still largely unknown (*2*–*4*).

The majority of disease-associated genetic variation lies in noncoding regions (*5*) enriched for noncoding RNAs (ncRNAs) and cis-regulatory elements that regulate gene expression and splicing of their cognate coding gene targets (*6*, *7*). Such regulatory relationships show substantial heterogeneity across human cell types, tissues, and developmental stages (*8*) and are often highly species specific (*9*). Recognizing the importance of understanding transcriptional regulation and noncoding genome function, several consortia (*8*, *10*–*12*) have undertaken large-scale efforts to provide maps of the transcriptome and its genetic and epigenetic regulation across human tissues. Although some have included central nervous system (CNS) tissues, a more comprehensive analysis focusing on the brain in both healthy and disease states is necessary to accelerate our understanding of the molecular mechanisms of these disorders (*13*–*16*).

We present results of the analysis of RNA sequencing (RNA-seq) data from the PsychENCODE Consortium (*16*), integrating genetic and genomic data from more than 2000 well-curated, high-quality postmortem brain samples from individuals with SCZ, BD, and ASD, as well as controls (*17*). We provide a comprehensive resource of disease-relevant gene expression changes and transcriptional networks in the postnatal human brain (see Resource.PsychENCODE.org for data and annotations). Data were generated across eight studies (*18*, *19*, *20*), uniformly processed, and combined through a consolidated genomic data processing pipeline (*21*) (fig. S1), yielding a total of 2188 samples passing quality control (QC) for this analysis, representing frontal and temporal cerebral cortices from 1695 individuals across the human life span, including 279 technical replicates (fig. S2). Extensive QC steps were taken within and across individual studies, resulting in the detection of 16,541 protein-coding and 9233 noncoding genes based on Gencode v19 annotations (*21*) (fig. S3). There was substantial heterogeneity in RNA-seq methodologies across cohorts, which was accounted for by including 28 surrogate variables and aggregate sequencing metrics as covariates in downstream analyses of differential expression (DE) at gene, isoform, and local splicing levels (*21*). DE did not overlap with experimentally defined brain RNA degradation metrics indicating that results were not driven by RNA-quality confounds (fig. S4) (*22*).

To provide a comprehensive view of the genomic architecture of these disorders, we characterized several levels of transcriptomic organization—gene-level, transcript isoform, local splicing, and coexpression networks—for protein-coding and noncoding gene biotypes.

[1]Department of Psychiatry, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, 695 Charles E. Young Drive South, Los Angeles, CA 90095, USA. [2]Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. [3]Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, 695 Charles E. Young Drive South, Los Angeles, CA 90095, USA. [4]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. [5]Department of Psychiatry, University of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093, USA. [6]Department of Psychiatry, and Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [7]Department of Genetics and Genomic Sciences, and Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [8]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [9]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [10]The School of Life Sciences, Central South University, Changsha, Hunan 410078, China. [11]National Clinical Research Center for Geriatric Disorders, Central South University, Changsha, Hunan 410078, China. [12]Program in Computational Biology and Bioinformatics, Departments of Molecular Biophysics and Biochemistry, Computer Science, and Statistics & Data Science, Yale University, New Haven, CT 06520, USA. [13]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA. [14]UNC Neuroscience Center, University of North Carolina, Chapel Hill, NC 27599, USA. [15]Fishberg Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [16]The Second Xiangya Hospital, Central South University, Changsha, Hunan 410011, China. [17]Department of Psychiatry, SUNY Upstate Medical University, Syracuse, NY 13210, USA. [18]Departments of Medicine and Cardiology, Cardiovascular Research Center, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [19]Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY 11794, USA. [20]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [21]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. [22]Institute for Genomics and Systems Biology, University of Chicago, Chicago, IL 60637, USA. [23]Tempus Labs, Chicago, IL 60654, USA. [24]Lieber Institute for Brain Development, Baltimore, MD 21205, USA. [25]Departments of Psychiatry and Behavioral Sciences, Johns Hopkins School of Medicine, Baltimore, MD 21287, USA. [26]CNS Data Coordination Group, Sage Bionetworks, Seattle, WA 98109, USA. [27]School of Psychology, Shaanxi Normal University, Xian, Shaanxi 710000, China.
***Corresponding author. Email: mgandal@mednet.ucla.edu (M.J.G.); liuch@upstate.edu (C.L.); lilyak@ucsd.edu (L.M.I.); dalila.pinto@mssm.edu (D.P.); dhg@mednet.ucla.edu (D.H.G.)**
†PsychENCODE Consortium authors and affiliations are listed in the supplementary materials.

We integrated results with common genetic variation and disease genome-wide association study (GWAS) results to identify putative regulatory targets of genetic risk variants. Although each level provides important disease-specific and shared molecular pathology, we find that isoform-level changes show the largest effects in diseased brains, are most reflective of genetic risk, and provide the greatest disease specificity when assembled into coexpression networks.

We recognize that these analyses involve a variety of steps and data types and are necessarily multifaceted and complex. We therefore organize results into two major sections. The first is at the level of individual genes and gene products, starting with gene-level transcriptomic analyses, as well as isoform and splicing analyses, followed by identification of potential genetic drivers. The second section is anchored in gene network analysis, where we identify coexpression modules at both gene and isoform levels and assess their relationship to genetic risk. As these networks reveal many layers of biology, we provide an interactive website to permit their in-depth exploration (Resource.PsychENCODE.org).

Gene and isoform expression alterations

RNA-seq–based quantifications enabled assessment of coding and noncoding genes and transcript isoforms, imputed using the RSEM software package guided by Gencode v19 annotations (*21*, *23*). In accordance with previous results (*13*), we observed pervasive differential gene expression (DGE) in ASD, SCZ, and BD [n = 1611, 4821, and 1119 genes at false discovery rate (FDR) < 0.05, respectively; Fig. 1A and table S1]. There was substantial cross-disorder sharing of this DE signal and a gradient of transcriptomic severity with the largest changes in ASD compared with SCZ or BD (ASD versus SCZ, mean $|\log_2FC|$ 0.26 versus 0.10, $P < 2 \times 10^{-16}$, Kolmogorov-Smirnov (K-S) test; ASD versus BD, mean $|\log_2FC|$ 0.26 versus 0.15, $P < 2 \times 10^{-16}$, K-S test), as observed previously (*13*). Altogether, more than one-quarter of the brain transcriptome was affected in at least one disorder (Fig. 1, A to C; complete gene list, table S1).

DGE results were concordant with previously published datasets for all three disorders (fig. S4), although some had overlapping samples. We observed significant concordance of DGE effect sizes with those from a microarray meta-analysis of each disorder [ASD: ρ = 0.8, SCZ: ρ = 0.78, BD: ρ = 0.64, Spearman ρ of $\log_2FC$, all P values < 10^{-16} (*13*)] and with previous RNA-seq studies of individual disorders [ASD: ρ = 0.96 (*19*); SCZ ρ = 0.78 (*18*); SCZ ρ = 0.80 (*24*); BD ρ = 0.85 (*13*); Spearman ρ of $\log_2FC$, all P values < 10^{-16}]. These DE genes exhibited substantial enrichment for known pathways and cell type–specific markers derived from single-nucleus RNA-seq in the human brain (Fig. 1, D and E) (*21*), consistent with previously observed patterns (*13*, *19*).

Expanding these analyses to the transcript isoform level, we observed widespread differential transcript expression (DTE) across ASD, SCZ, and BD (n = 767, 3803, and 248 isoforms at FDR < 0.05, respectively; table S1). Notably, at the DTE level, the cross-disorder overlap was significantly attenuated (Fig. 1C), suggesting that alternative transcript usage and/or splicing confers a substantial portion of disease specificity. In addition, isoform-level alterations in disease exhibited substantially larger effect sizes compared with gene-level changes (mean $|\log_2FC|$ 0.25 versus 0.14, $P < 2 \times 10^{-16}$, K-S test), particularly for protein-coding biotypes (Fig. 1A), consistent with recent work demonstrating the importance of splicing dysregulation in disease pathogenesis (*25*). Furthermore, although isoform and gene-level changes exhibited similar pathway and cell type enrichments (e.g., Fig. 1, D and E), isoform-level analysis identified DE transcripts that did not show DGE (isoform-only DE), including 811 in SCZ, 294 in ASD, and 60 in BD. These isoform-only DE genes were more likely to be down-regulated than up-regulated in disease (one-sample t test, $P < 10^{-16}$), exhibited greatest overlap with excitatory neuron clusters [odds ratios (ORs) > 4, Fisher's exact test, FDRs < 10^{-10}], and showed significant enrichment for neuron projection development, mRNA metabolism, and synaptic pathways (FDR < 3×10^{-3}; table S1). To validate DTE results, we performed polymerase chain reaction (PCR) on several selected transcripts in a subset of ASD, SCZ, and control samples (*21*) and found significant concordance in fold-changes compared with those from RNA-seq data (fig. S5, A and B). Together, these results suggest that isoform-level changes are most reflective of neuronal and synaptic dysfunction characteristic of each disorder.

Differential expression of the noncoding transcriptome

ncRNAs represent the largest class of transcripts in the human genome and have increasingly been associated with complex phenotypes (*26*). However, most have limited functional annotation, particularly in the human brain, and have been only minimally characterized in the context of psychiatric disease. On the basis of Gencode annotations, we identified 944 ncRNAs exhibiting gene- or isoform-level DE in at least one disorder [hereafter referred to as neuropsychiatric (NP) ncRNAs (*21*)], 693 of which were differentially expressed in SCZ, 178 in ASD, and 174 in BD. Of these, 208, 60, and 52 are annotated as intergenic long ncRNAs (lincRNAs) in each disorder, respectively. To place these NPncRNAs within a functional context, we examined expression patterns across human tissues, cell types, and developmental time periods, as well as sequence characteristics including evolutionary conservation, selection, and constraint. We highlight several noncoding genes exhibiting DE across multiple disorders (fig. S6) and provide comprehensive annotations for each NPncRNA (table S2), including cell type specificity, developmental trajectory, and constraint, to begin to elucidate a functional context in the human brain.

As a class, NPncRNAs were under greater selective constraint compared with all Gencode annotated ncRNAs (Fig. 1F), consistent with the observed increased purifying selection in brain-expressed genes (*27*). We identified 74 NPncRNAs (~8%) under purifying selection in humans, with average exon-level context-dependent tolerance scores (CDTS) below the 10th percentile (*21*). More than 200 NPncRNAs exhibited broad and nonspecific expression patterns across cell types, whereas 66 were expressed within a specific cell type class (table S2). Notable examples are: *LINC00996*, which is down-regulated in SCZ ($\log_2FC$ −0.71, FDR < 5×10^{-11}) and BD ($\log_2FC$ −0.45, FDR = 0.02) and restricted to microglia in the brain (fig. S6); *LINC00343*, which is expressed in excitatory neurons and down-regulated in BD ($\log_2FC$ −0.33, FDR = 0.012) with a trend in SCZ ($\log_2FC$ −0.15, FDR 0.065); and *LINC00634*, an unstudied brain-enriched lincRNA down-regulated in SCZ ($\log_2FC$ −0.06, FDR 0.027) with a genome-wide significant SCZ TWAS association as described below.

Local splicing dysregulation in disease

Isoform-level diversity is achieved by combinatorial use of alternative transcription start sites, polyadenylation, and splicing (*28*). We used LeafCutter (*29*) to assess local differential splicing (DS) in ASD, SCZ, and BD compared with controls using de novo aligned RNA-seq reads, controlling for the same covariates as DGE and DTE (fig. S7). This approach complements DTE by considering aggregate changes in intron usage affecting exons that may be shared by multiple transcripts and is consequently not restricted to the specified genome annotation (*21*). Previous studies have identified alterations in local splicing events in ASD (*19*, *30*) and in smaller cohorts in SCZ (*18*, *24*) and BD (*31*).

We identified 515 DS intron clusters in 472 genes across all disorders (FDR < 0.1), 117 of which (25%) contained one or more previously unidentified exons (table S3 and Fig. 2A). Validation of DS changes for 9 genes in a subset of cases and controls (n = 5 to 10 in each group) by semiquantitative reverse transcription (RT)–PCR showed percent spliced-in (PSI) changes consistent with those reported by LeafCutter (fig. S5, C to E). The most commonly observed local splicing change was exon skipping (41 to 60%), followed by alternative 5′ exon inclusion (e.g., due to alternative promoter usage; 11 to 21%) and alternative 3′ splice site usage (5 to 18%) (table S3 and fig. S8A). DS genes overlapped significantly with DTE results for ASD and SCZ (fig. S8B), but not BD, which likely still remains underpowered. There was significant cross-disorder correlation in PSI changes (Spearman's ρ = 0.59 SCZ-BD, ρ = 0.52 SCZ-ASD, all $P < 10^{-4}$) and, subsequently, overlap among DS genes (Fig. 2, A and B), although the majority of splicing changes still are disorder specific. Only two genes, *DTNA* and *AHCYL1*, were significantly differentially spliced in all three disorders (fig. S9). Differentially spliced genes showed significant (FDR < 0.05) enrichment for signaling, cell communication, actin cytoskeleton, synapse, and neuronal development

pathways across disorders (Fig. 2C and fig. S8C) and were relatively broadly expressed across cell types (Fig. 2D). Disorder-specific pathways implicated by splicing dysfunction include plasma membrane receptor complex, endocytic vesicle, regulation of cell growth and cytoskeletal protein binding in ASD; angiotensin receptor signaling in BD; and guanosine triphosphatase receptor activity, neuron development, and actin

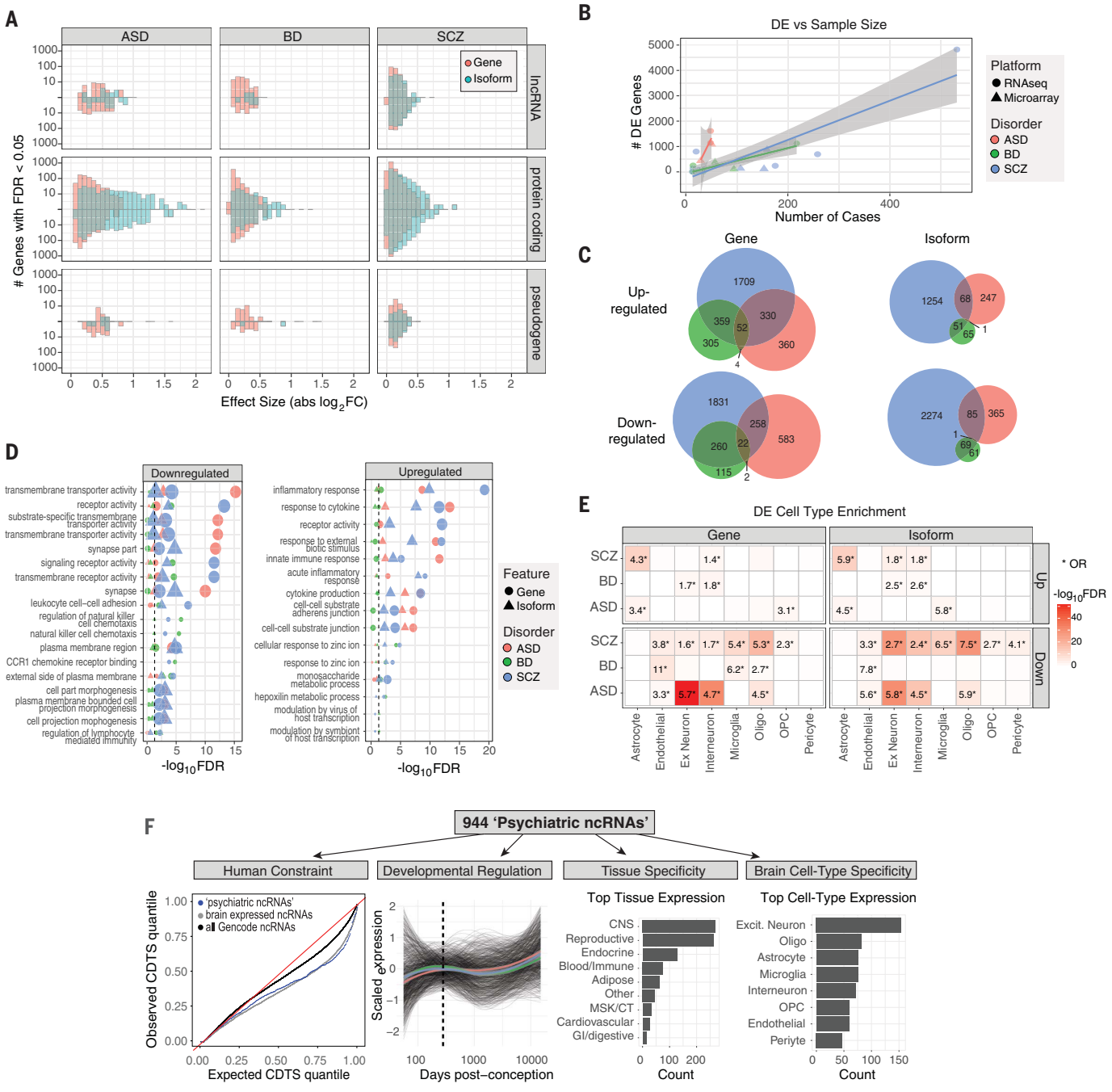


Fig. 1. Gene and isoform expression dysregulation in brain samples from individuals with psychiatric disorders. (**A**) DE effect size (|log$_2$FC|) histograms are shown for protein-coding, lncRNA, and pseudogene biotypes up- or down-regulated (FDR < 0.05) in disease. Isoform-level changes (DTE; blue) show larger effect sizes than at the gene level (DGE; red), particularly for protein-coding biotypes in ASD and SCZ. (**B**) A literature-based comparison shows that the number of DE genes detected is dependent on study sample size for each disorder. (**C**) Venn diagrams depict overlap among up- or down-regulated genes and isoforms across disorders. (**D**) Gene ontology enrichments are shown for differentially expressed genes or isoforms. The top five pathways are shown for each disorder. (**E**) Heatmap depicting cell type specificity of enrichment signals. Differentially expressed features show substantial enrichment for known CNS cell type markers, defined at the gene level from single-cell RNA-seq. (**F**) Annotation of 944 ncRNAs DE in at least one disorder. From left to right: Sequence-based characterization of ncRNAs for measures of human selective constraint; brain developmental expression trajectories are similar across each disorder (colored lines represent mean trajectory across disorders); tissue specificity; and CNS cell type expression patterns.

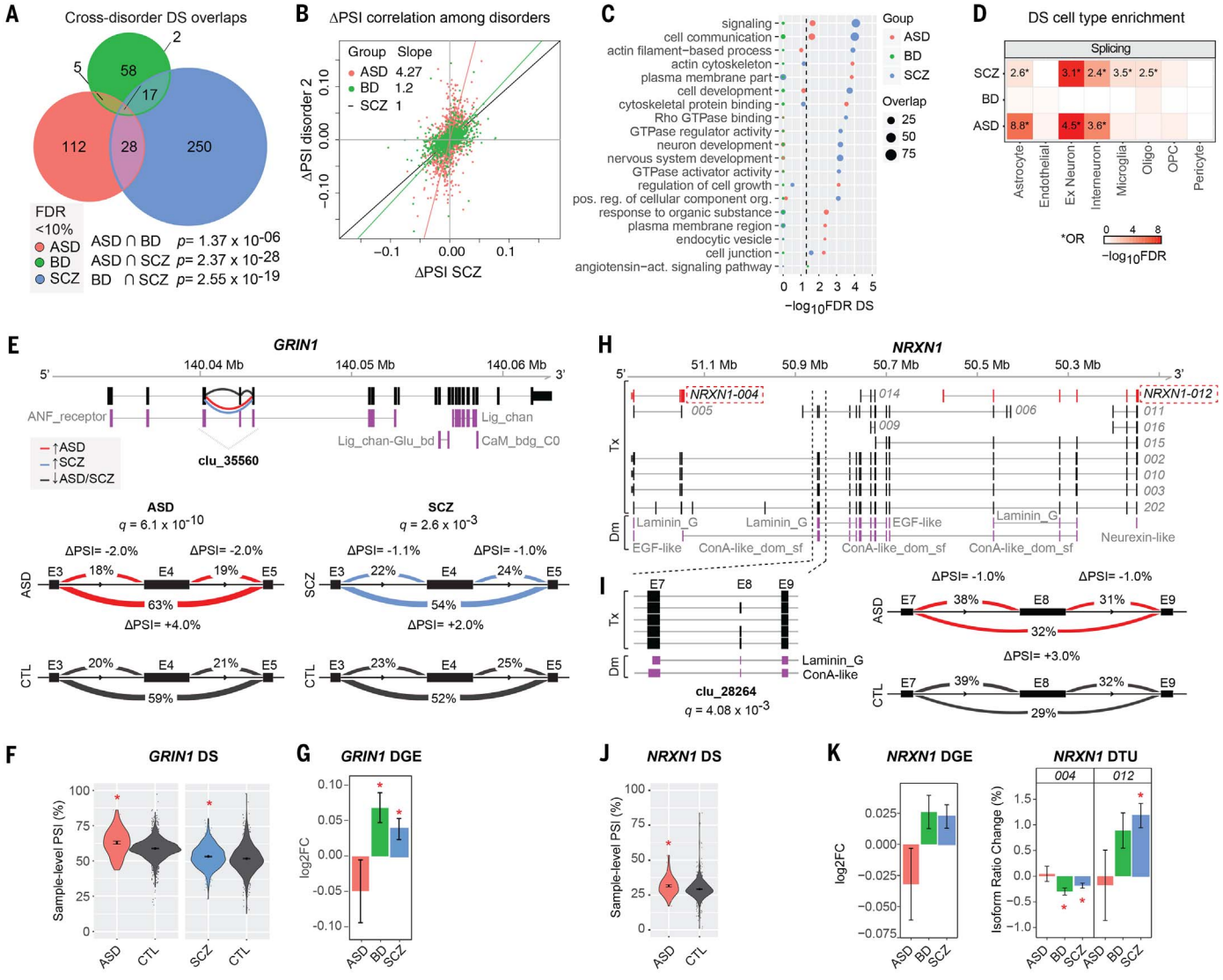


Fig. 2. Aberrant local splicing and isoform usage in ASD, SCZ, and BD. (**A**) Venn diagram showing cross-disorder overlap for 472 genes with significant differentially spliced (DS) intron clusters (FDR < 10%) identified by LeafCutter. *P* values for hypergeometric tests of pairwise overlaps between each disorder are shown at the bottom. (**B**) Scatter plots comparing PSI changes for all 1287 introns in 515 significant DS clusters in at least one disorder, for significant disease pairs SCZ versus ASD and SCZ versus BD (Spearman's ρ = 0.52 and 0.59, respectively). Principal component regression lines are shown in red, with regression slopes for ASD and BD ΔPSI compared to SCZ in the top-left corner. (**C**) Top 10 gene ontology (GO) enrichments for DS genes in each disorder (see also fig. S8C). (**D**) Significant enrichment for neuronal and astrocyte markers (ASD and SCZ), as well as oligodendrocyte and microglia (SCZ) cell type markers in DS genes. The odds ratio (*OR) is given only for FDR < 5% and OR > 1. Oligo, oligodendrocytes; OPC, oligodendrocyte progenitor cells. (**E**) A significant DS intron cluster in *GRIN1* (clu_35560; chr9:140,040,354-140,043,461) showing increased exon 4 (E4) skipping in both ASD and SCZ. Increased or decreased intron usage in ASD and SCZ cases compared to controls is highlighted in red and blue, respectively. Protein domains are annotated as ANF_receptor, extracellular receptor family ligand binding domain; Lig_chan, ionotropic glutamate receptor; Lig_chan-Glu_bd, ligated ion channel L-glutamate- and glycine-binding site; CaM_bdg_C0, calmodulin-binding domain C0 of NMDA receptor NR1 subunit. Visualization of splicing events in cluster clu_35560 with the change in PSI (ΔPSI) for ASD (left) and SCZ (right) group comparisons. FDR-corrected *P* values (q) are indicated for each comparison. Covariate-adjusted average PSI levels in ASD or SCZ (red) versus CTL (blue) are indicated at each intron. (**F**) Violin plots with the distribution of covariate-adjusted PSI per sample for the intron skipping E4 are shown for each disease group comparison. (**G**) DGE for *GRIN1* in each disorder (*FDR < 5%). (**H**) Whole-gene view of *NRXN1* highlighting (dashed lines) the intron cluster with significant DS in ASD (clu_28264; chr2:50,847,321-50,850,452), as well as transcripts *NRXN1-004* and *NRXN1-012* that show significant DTU in SCZ and/or BD. Protein domain mappings are shown in purple. DM, protein domains; Tx, transcripts; ConA-like_dom_sf, concanavalin A–like lectin/glucanase domain; EGF-like, epidermal growth factor-like domain; laminin_G, laminin G domain; neurexin-like, neurexin/syndecan/glycophorin C domain. (**I**) (Left) Close-up of exons and protein domains mapped onto the DS cluster and FDR-corrected P value (*q*). (Right) Visualization of introns in cluster clu_28264 with their change in percent spliced in (ΔPSI). Covariate-adjusted average PSI levels in ASD (red) versus CTL (blue) are indicated for each intron. (**J**) Violin plots with the distribution of covariate-adjusted PSI per sample for the largest intron skipping exon 8 (E8). (**K**) Bar plots for changes in gene expression and transcript usage for *NRXN1-004* and *NRXN1-012* (*FDR < 5%).

cytoskeleton in SCZ. We also found significant enrichment of splicing changes in targets of two RNA binding proteins that regulate synaptic transmission and whose targets are implicated in both ASD and SCZ, the neuronal splicing regulator *RBFOX1* (FDR = 5.16×10^{-11}) (*32*) and the fragile X mental retardation protein (FMRP) (FDR = 3.10×10^{-21}) (*33*). Notably, 48 DS genes (10%; FDR = 8.8×10^{-4}) encode RNA binding proteins or splicing factors (*34*), with at least six splicing factors also showing DTE in ASD (*MATR3*), SCZ (*QKI*, *RBM3*, *SRRM2*, *U2AF1*), or both (*SRSF11*).

Many differential splicing events show predictable functional consequences on protein isoforms. Notable examples include *GRIN1* and *NRXN1*, which are known risk loci for neurodevelopmental disorders (*35*, *36*). *GRIN1* encodes the obligatory subunit of the *N*-methyl-D-aspartate (NMDA)–type glutamate ionotropic receptors, is up-regulated in SCZ and BD, and shows increased skipping of exon 4 in both ASD and SCZ that affects its extracellular ligand-binding domain (Fig. 2, E to G). *NRXN1* is a heterotypic, presynaptic cell adhesion molecule that undergoes extensive alternative splicing and plays a key role in the maturation and function of synapses (*35*, *37*). We observed various DS and/or differential transcript usage (DTU) changes in *NRXN1* in ASD, SCZ, and/or BD (Fig. 2, H to K). An exon skipping event in ASD disrupts a laminin domain in *NRXN1* (Fig. 2, I and J), changes that are predicted to have major effects on its function (Fig. 2H). Another example is *CADPS*, which is located within an ASD GWAS risk locus and supported by high-resolution chromosome conformation capture (Hi-C)–defined chromatin interactions as a putative target gene (*38*) and manifests multiple isoform and splice alterations in ASD (fig. S9 and tables S1 and S3).

We found significant overlap (42%, $P = 3.42 \times 10^{-27}$; Fisher's exact test) of the ASD DS intron clusters and splicing changes identified in a previous study (*19*) that used a different method and only a subset of the samples in our ASD and control cohorts (table S3). Overall, this examination of local splicing across three major neuropsychiatric disorders, coupled with the analysis of isoform-level regulation, emphasizes the need to understand the regulation and function of transcript isoforms at a cell type–specific level in the human nervous system.

Identifying drivers of transcriptome dysregulation

We next sought to determine whether changes observed across levels of transcriptomic organization are reflective of the same, or distinct, underlying biological processes. Further, transcriptomic changes may represent a causal pathophysiology or may be a consequence of disease. To begin to address this, we assessed the relationships among transcriptomic features and with polygenic risk scores (PRS) for disease, which provide a directional, genetic anchor (Fig. 3A). Across all three disorders, there was strong concordance among differential gene, isoform, and ncRNA signals, as summarized by their first principal component (Fig. 3A). Notably, DS exhibited greatest overlap with the ncRNA signal, suggesting a role for noncoding genes in regulating local splicing events.

Significant associations with PRS were observed for DGE and DTE signals in SCZ, with greater polygenic association at the isoform level in accordance with the larger transcript isoform effect sizes observed. Transcript-level DE also showed the greatest enrichment for SCZ single-nucleotide polymorphism (SNP) heritability, as measured by stratified LD (linkage disequilibrium) score regression (*21*, *39*) (Fig. 3B). The overall magnitude of genetic enrichment was modest, however, suggesting that most observed transcriptomic alterations are less a proximal effect of genetic variation and more likely the consequence of a downstream cascade of biological events following earlier-acting genetic risk factors.

We were also interested in determining the degree to which genes showed increases in the magnitude of DE over the duration of illness, as a positive relationship would be expected if age-related cumulative exposures (e.g., drugs, smoking) were driving these changes. To assess this, we fit local regression models to case and control sample-level expression measurements as a function of age and computed age-specific DE effect sizes (fig. S10). Of 4821 differentially expressed genes in SCZ, only 143 showed even nominal association between effect size magnitude and age. Similar associations were seen in 29 of 1119 differentially expressed genes in BD and 85 of 1611 differentially expressed genes in ASD. Consequently, this would not support substantial age-related environmental exposures as the mechanism for the vast majority of differentially expressed genes.

Using gene expression data from animal models, we investigated whether exposure to commonly used psychiatric medications could recapitulate observed gene expression changes in disease (fig. S11). Overall, with the exception of lithium, chronic exposure to medications—including antipsychotics (clozapine, haloperidol), mood stabilizers (lamotrigine), and SSRI antidepressants (fluoxetine)—had a small effect on the transcriptome, in many cases with no differentially expressed genes at traditional FDR thresholds (*21*). Even at more liberal thresholds, the overlap between medication-driven and disease signal remains sparse. One notable exception was a module that reflects major components of a well-described (*40*) neural activity–dependent gene expression program, whose disease relationships are refined in the network analysis section below. Finally, we note that other unmeasured factors could potentially contribute

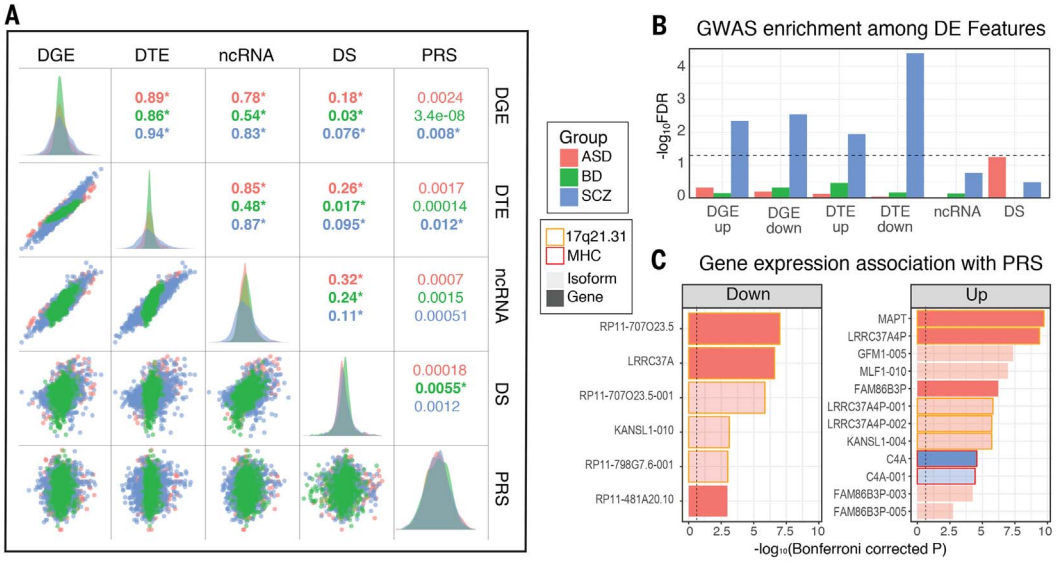


Fig. 3. Overlap and genetic enrichment among dysregulated transcriptomic features. (**A**) Scatterplots demonstrate overlap among dysregulated transcriptomic features, summarized by their first principal component across subjects (R^2 values; *$P < 0.05$). PRS show greatest association with differential transcript signal in SCZ. (**B**) SNP heritability in SCZ is enriched among multiple differentially expressed transcriptomic features, with down-regulated isoforms showing the most substantial association via stratified LD-score regression. (**C**) Several individual genes and isoforms exhibit genome-wide significant associations with disease PRS. Plots are split by direction of association with increasing PRS. In ASD, most associations localize to the 17q21.31 locus, harboring a common inversion polymorphism.

to gene expression variation in postmortem tissue, including agonal events or smoking (22, 41, 42) in addition to those measured and used as covariates, such as RNA integrity and postmortem interval. We used surrogate variable correction in our analyses to account for such unmeasured confounders (43), which is a standard approach (44).

Transcriptome-wide association

We next sought to leverage this transcriptomic dataset to prioritize candidate disease risk genes with predicted genetically driven effects on expression in brain. We identified 18 genes or isoforms whose expression was significantly associated with PRS [(21); Bonferroni-corrected $P < 0.05$]: 16 in ASD and 2 in SCZ, with none in BD (Fig. 3C and table S4). In ASD, the majority of associations map to 17q21.31, which harbors a common inversion polymorphism and rare deleterious structural variants associated with intellectual disability (45). Additional associations for ASD included two poorly annotated pseudogenes, *FAM86B3P* and *RP11-481A20.10*. In SCZ, PRS was associated with up-regulation of the established risk gene *C4A* (3). Concordantly, we found a strong positive correlation between *C4A* expression and genetically imputed *C4A* copy number ($R = 0.36$, $P = 6 \times 10^{-21}$) and imputed number of *C4-HERV* elements ($R = 0.35$, $P = 4 \times 10^{-20}$) but a slight negative association with *C4B* copy number [$R = -0.087$, $P = 0.03$ (21)]. At less stringent thresholds (FDR-corrected $P < 0.05$), we identified BD PRS associations with isoforms of the neuronal calcium sensor *NCALD* and *SNF8*, an endosomal sorting protein, as well as several additional associations in the major histocompatibility complex (MHC) region in SCZ, which harbors the largest GWAS peak composed of multiple independent signals (3) but is difficult to parse due to complex patterns of LD. These included two lncRNAs, *HCG17* and *HCG23*, as well as the MHC class I heavy-chain receptor *HLA-C*. However, expression of all three was also significantly ($P < 0.05$) correlated with imputed *C4A* copy number, suggesting pleiotropic effects.

Taking an orthogonal approach, we performed a formal transcriptome-wide association study (TWAS) (46) to directly identify genes whose cis-regulated expression is associated with disease (21). TWAS and related methods have the advantage of aggregating the effects of multiple SNPs onto specific genes, reducing multiple comparisons and increasing power for association testing, although results can still be influenced by LD and pleiotropy (46, 47). Further, by imputing the cis-regulated heritable component of brain gene expression into the association cohort, TWAS enables direct prediction of the transcriptomic effects of disease-associated genetic variation, identifying potential mechanisms through which variants may impart risk. However, the limited size of brain eQTL (expression quantitative trait loci) datasets to date has necessitated the use of non-CNS tissues to define TWAS weights (46). Given the enrichment of psychiatric GWAS signal within CNS-expressed regulatory elements (39), we reasoned that our dataset would provide substantial power and specificity. Indeed, we identified 14,750 genes with heritable cis-regulated brain expression in the PsychENCODE cohort, enabling increased transcriptomic coverage for detection of association signal (Fig. 4). In BD, TWAS prioritizes 17 genes across 14 distinct loci (Bonferroni-corrected $P < 0.05$; Fig. 4 and table S4), none of which exhibited DE. At loci with multiple hits, we applied conditional analyses to further fine-map these regions (21). For orthogonal validation, we conducted summary-data–based Mendelian randomization (SMR), a complementary method that tests for pleiotropic associations in the cis window with an accompanying HEIDI test to distinguish linkage from pleiotropy (48). Eleven genes—*BMPR1B*, *DCLK3*, *HAPLN4*, *HLF*, *LMAN2L*, *MCHR1*, *UBE2Q2L*, *SNAP91*, *TTC39A*, *TMEM258*, and *VPS45*—showed consistent association (21) across multiple analyses (table S4). The two isoforms with PRS associations in BD (*NCALD*, *SNF8*) were nonsignificant in TWAS, perhaps owing to lack of a nearby genome-wide significant locus or isoform-specific regulation, which suggests that those expression changes may be driven by trans-acting factors.

In ASD, TWAS prioritizes 12 genes across three genomic loci (Bonferroni-corrected $P < 0.05$; Fig. 4). This includes the 17q21.31 region, which showed multiple PRS associations as described above but did not reach genome-wide significance in the largest GWAS to date (38). Of the seven TWAS-significant genes at this locus, conditional analysis prioritizes one—*LRRC37A*, which is further supported by SMR and Hi-C interaction in fetal brain (38). *LRRC37A* is intriguing due to its primate-specific evolutionary expansion, loss-of-function intolerance, and expression patterns in the brain and testis (45). However, common variants in GWAS are also likely tagging the common inversion and other recurrent structural variants present at this locus (45). TWAS additionally prioritizes genes on chromosomes 8 and 20 (Fig. 4). Altogether, five genes showed consistent associations with ASD across multiple methods: *LRRC37A*, *FAM86B3P*, *PINX1*, *XKR6*, and *RP11-481A20.10* (table S4) (21).

In SCZ, TWAS identifies 193 genes, of which 107 remain significant after conditional analysis at each gene within multi-hit loci. Excluding the MHC region, there remained 164 significant genes representing 78 genome-wide significant GWAS loci (Fig. 4 and table S4). A previous TWAS study in SCZ primarily based on non-neural tissue prioritized 157 genes, 37 of which are identified here, a significant overlap (OR = 61, $P < 10^{-42}$, Fisher's exact test). Moreover, 60 TWAS-prioritized genes overlapped with the list of 321 high-confidence SCZ risk genes in a companion manuscript (17), identified using gene regulatory networks and a deep learning approach (OR = 34.7, $P < 10^{-60}$, Fisher's exact test). Of the 107 conditionally significant genes prioritized by TWAS, 62 were further supported by SMR ($P_{SMR} < 0.05$, $P_{HEIDI} > 0.05$), and 11 were also concordantly differentially expressed in SCZ brains in the same direction as predicted by TWAS. Altogether, 64 genes were consistently prioritized across multiple methods, including 10 ncRNAs (table S4) (21). These included a number of previously unknown candidates for SCZ: two down-regulated lysine methyltransferases (*SETD6*, *SETD8*); *RERE*, a down-regulated, mutationally intolerant nuclear receptor co-regulator of retinoic acid signaling associated with a rare neurodevelopmental genetic syndrome; *LINC00634*, a down-regulated poorly annotated brain-enriched lincRNA; and *SLC12A5*, which encodes a mitochondrial Ca^{2+} binding aspartate/glutamate carrier protein, associated with a recessive epileptic encephalopathy. Most genes identified in this analysis show disease-specific effects, as only four genes (*MCHR1*, *VPS45*, *SNAP91*, and *DCLK3*) showed overlap between SCZ and BD TWAS, and none overlapped with ASD. Overall, this analysis provides a core set of strong candidate genes implicated by risk loci and provides a mechanistic basis for the composite activity of disease risk variants.

Networks refine shared cross-disorder signals

To place transcriptomic changes within a systems-level context and more fully investigate the specific molecular neuropathology of these disorders, we performed weighted gene correlation network analysis (WGCNA) to create independent gene- and isoform-level networks (14, 49, 50), which we then assessed for disease association and GWAS enrichment by using stratified LD score regression [(21); see Resource.PsychENCODE.org for interactive visualization]. Although calculated separately, gene- and isoform-level networks generally reflected equivalent biological processes, as demonstrated by hierarchical clustering (Fig. 5A). However, the isoform-level networks captured greater detail, and a larger proportion were associated with disease GWAS than gene-level networks (61% versus 41% with nominal GWAS enrichment, $P = 0.07$, χ^2; Fig. 5A). Consistent with expectations, modules showed enrichment for gene ontology pathways, and we identified modules strongly and selectively enriched for markers of all major CNS cell types (Fig. 5, A and B, and fig. S12), facilitating computational deconvolution of cell type–specific signatures (14, 49, 51). For ease of subsequent presentation, we grouped gene-isoform module pairs that co-cluster, have overlapping parent genes, and represent equivalent biological processes.

The large sample sizes, coupled with the specificity of isoform-level quantifications, enabled refinement of previously identified gene networks related to ASD, BD, and SCZ (13–15, 18, 19, 52). Of a combined 90 modules, including 34 gene-level (geneM) and 56 isoform-level (isoM) modules, 61 (68%) showed significant association with at least one disorder, demonstrating the pervasive nature of transcriptome dysregulation in psychiatric disease. Five modules are shared across all three disorders, 3 up-regulated and 2 down-regulated; 22 modules are shared by two of the

three disorders, and 36 demonstrate more specific patterns of dysregulation in either ASD, SCZ, or BD (Fig. 5 and table S5). It is notable that of these 61 coexpression modules with a disease-association, 41 demonstrate cell type enrichments, consistent with the strong cell type disease-related signal that was observed via both supervised and unsupervised methods in a companion study (*17*). This demonstrates the importance of cell type–specific changes in the molecular pathology of these major psychiatric disorders; the cell type relationships defined by the disease modules substantially enhance our knowledge of these processes, as we outline below.

The five modules shared between ASD, BD, and SCZ can be summarized to represent three distinct biological processes. Two of these processes are up-regulated, including an inflammatory NFkB (nuclear factor κB) signaling module pair (geneM5/isoM5; further discussed in the "Distinct neural-immune trajectories" section) and a module (geneM31) enriched primarily for genes with roles in the postsynaptic density, dendritic compartments, and receptor-mediated presynaptic signaling that are expressed in excitatory neurons and, to a lesser extent, inhibitory neurons (Fig. 5C). Notably, *DCLK3*, one of the hubs of geneM31, is a genome-wide significant TWAS hit in both SCZ and BD. The third biological process, geneM26/isoM22 (Fig. 5C), is down-regulated and enriched for endothelial and pericyte genes, with hubs that represent markers of the blood-brain barrier, including *ITIH5*, *SLC38A5*, *ABCB1*, and *GPR124*, a critical regulator of brain-specific angiogenesis (*53*, *54*). This highlights specific, shared alterations in neuronal-glial-endothelial interactions across these neuropsychiatric disorders.

In contrast to individual genes or isoforms, no modules were significantly associated with PRS after multiple-testing correction. However, 19 modules were significantly (FDR < 0.05) enriched for SNP heritability on the basis of published GWASs (*21*) (Fig. 5A and fig. S13). A notable example is geneM2/isoM13, which is enriched for oligodendrocyte markers and neuron projection developmental pathways and is down-regulated in ASD and SCZ, with a trend in BD (Fig. 5C). isoM13 showed the greatest overall significance of enrichment for SCZ and educational attainment GWAS and was also enriched in BD GWAS to a lesser degree. Further, this module is enriched for genes harboring ultrarare variants identified in SCZ (*55*) (fig. S13). Finally, we also observe pervasive and distinct enrichments for syndromic genes and rare variants identified through whole-exome sequencing in individuals

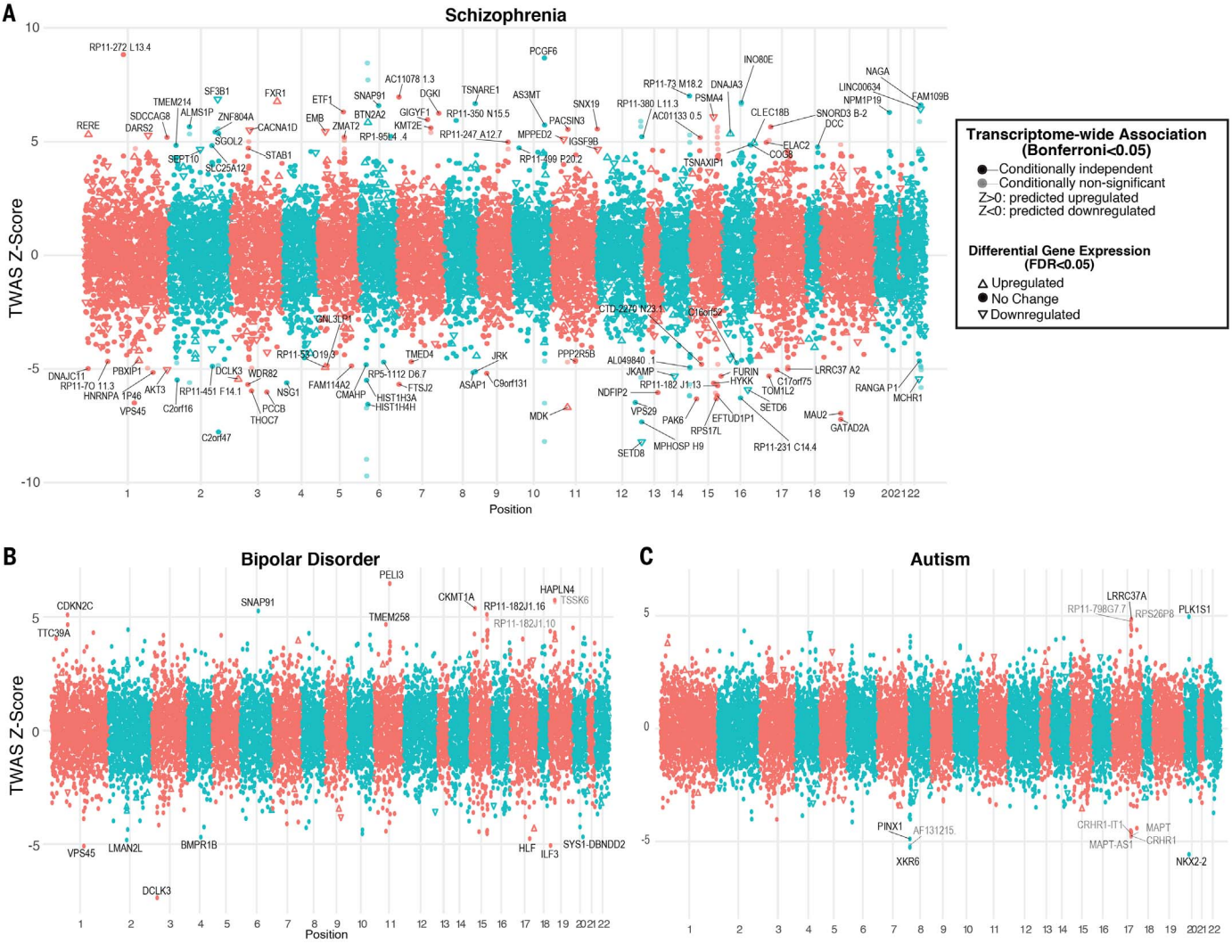


Fig. 4. Transcriptome-wide association. Results from a TWAS prioritize genes whose cis-regulated expression in brain is associated with disease. Plots show conditionally-independent TWAS prioritized genes, with lighter shades depicting marginal associations. The sign of TWAS *z*-scores indicates predicted direction of effect. Genes significantly up- or down-regulated in diseased brain are shown with arrows, indicating directionality. (**A**) In SCZ, 193 genes (164 outside of MHC) are prioritized at Bonferroni-corrected $P < 0.05$, including 107 genes with conditionally independent signals. Of these, 23 are also differentially expressed in SCZ brains with 11 in the same direction as predicted. (**B**) Seventeen genes are prioritized in BD, of which 15 are conditionally independent. (**C**) In ASD, a TWAS prioritizes 12 genes, of which 5 are conditionally independent.

with neurodevelopmental disorders (table S5 and fig. S13).

Neuronal isoform networks capture disease specificity

Multiple neuronal and synaptic signaling pathways have been previously shown to be down-regulated in a diminishing gradient across ASD, SCZ, and BD brains without identification of clear disease-specific signals for these neuronal-synaptic gene sets (*13*, *15*, *18*, *19*, *56*, *57*). We do observe neuronal modules broadly dysregulated across multiple disorders, including a neuronal/synaptic module (isoM18) with multiple isoforms of the known ASD risk gene, *ANK2*, as hubs. However, the large sample size, coupled with the specificity of isoform-level qualifications, enabled us to identify synaptic modules containing isoforms with distinct disease associations and to separate signals from excitatory and inhibitory neurons (Fig. 5B).

A salient example of differential module membership and disease association of transcript isoforms is *RBFOX1*, a major neuronal splicing regulator implicated across multiple neurodevelopmental and psychiatric disorders (*15*, *32*, *58*, *59*). Previous work has identified down-regulated neuronal modules in ASD and SCZ containing *RBFOX1* as a hub (*13*, *15*). In this study, we identified two neuronal modules with distinct *RBFOX1* isoforms as hub genes (Fig. 6A). The module pair geneM1/isoM2, down-regulated only in ASD (Fig. 6B), contains the predominant brain-expressed *RBFOX1* isoform and includes several cation channels (e.g., *HCN1*, *SCN8A*). The second most abundant *RBFOX1* isoform is in another module, isoM17, which is down-regulated in both ASD and SCZ (Fig. 6B). Experiments in mouse indicate that *RBFOX1* has distinct nuclear and cytoplasmic isoforms with differing functions, the nuclear isoform primarily regulating pre-mRNA alternative splicing, and the cytoplasmic isoform binding to the 3′ untranslated region to stabilize target transcripts involved in regulation of neuronal excitability (*28*, *32*, *58*, *60*). isoM17 shows greater enrichment for nuclear *RBFOX1* targets (Fig. 6C), whereas isoM2 shows stronger overlap with cytoplasmic targets (*32*). Consistent with a

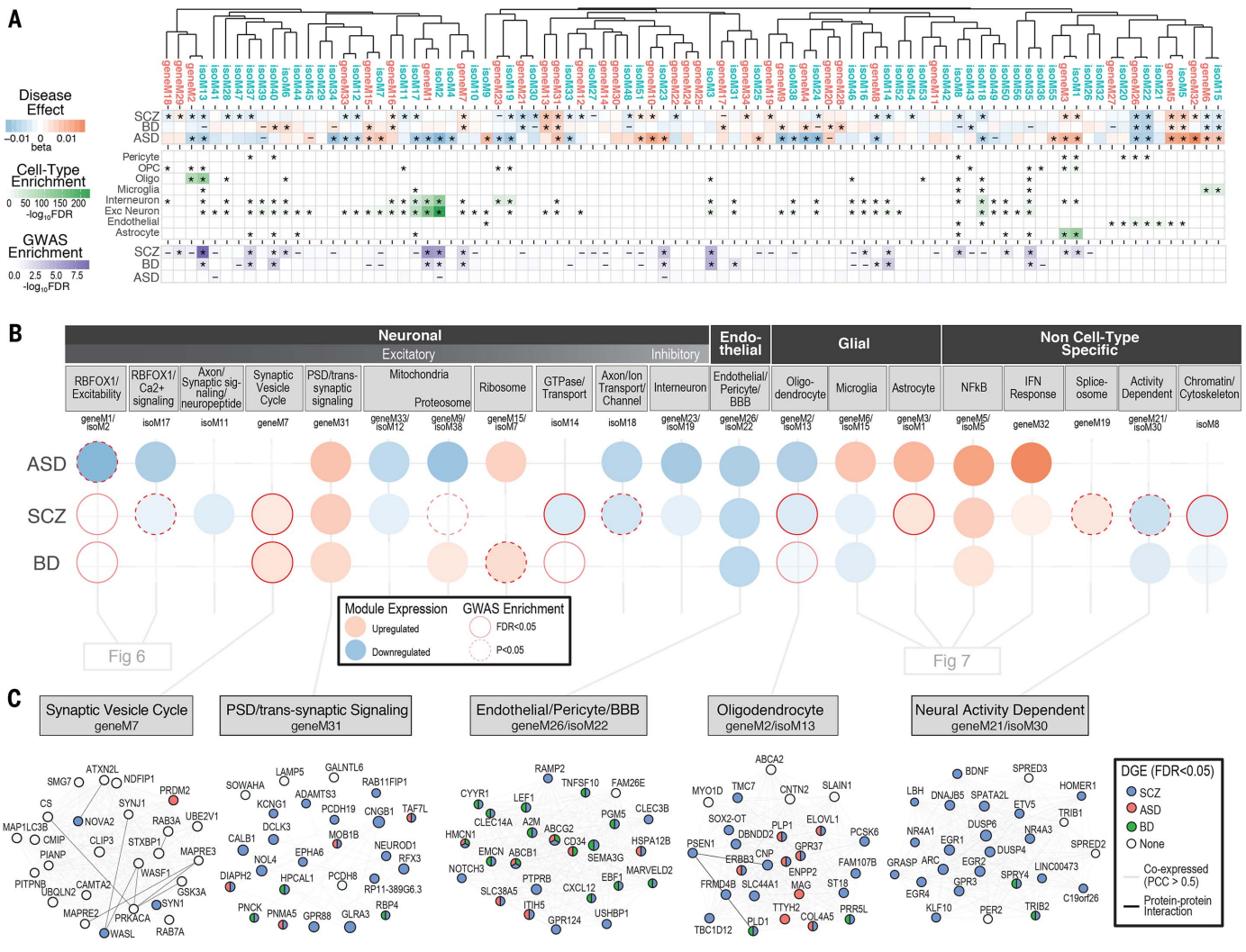


Fig. 5. Gene and isoform coexpression networks capture shared and disease-specific cellular processes and interactions. (**A**) Coexpression networks demonstrate pervasive dysregulation across psychiatric disorders. Hierarchical clustering shows that separate gene- and isoform-based networks are highly overlapping, with greater specificity conferred at the isoform level. Disease associations are shown for each module (linear regression β value, *FDR < 0.05, –*P* < 0.05). Module enrichments (*FDR < 0.05) are shown for major CNS cell types. Enrichments are shown for GWAS results from SCZ (*59*), BD (*97*), and ASD (*38*), using stratified LD score regression (*FDR < 0.05, –*P* < 0.05). (**B**) Coexpression modules capture specific cellular identities and biological pathways. Colored circles represent module DE effect size in disease, with red outlines representing GWAS enrichment in that disorder. Modules are organized and labeled based on CNS cell type and top gene ontology enrichments. (**C**) Examples of specific modules dysregulated across disorders, with the top 25 hub genes shown. Edges represent coexpression (Pearson correlation > 0.5) and known protein-protein interactions. Nodes are colored to represent disorders in which that gene is differentially expressed (*FDR < 0.05).

predicted splicing-regulatory effect, isoM17 shows greater enrichment for genes exhibiting DS in ASD and SCZ (Fig. 6D). In accordance with a predicted role in regulating excitability, isoM2 shows strong enrichment for epilepsy risk genes (Fig. 6E). Moreover, the two modules show differential association with common genetic risk (Fig. 6E), with isoM2 exhibiting GWAS enrichment across SCZ, BD, and major depressive disorder (MDD). This widespread enrichment of neurodevelopmental and psychiatric disease risk factors—from rare variants in epilepsy to common variants in BD, SCZ, and MDD—is consistent with a model in which broad neuropsychiatric liability emanates from myriad forms of dysregulation in neuronal excitability, all linked via *RBFOX1*. These results highlight the importance of further studies focused on understanding the relationship between human *RBFOX1* transcript diversity and functional divergence, as most of what is known is based on mouse, and the human shows far greater transcript diversity (*32*, *58*, *61*).

Previous transcriptional networks related to ASD, BD, and SCZ did not separate inhibitory and excitatory neuron signals (*13*). The increased resolution here allowed us to identify several modules enriched in inhibitory interneuron markers (Fig. 5B), including geneM23/isoM19, which is down-regulated in ASD and SCZ, with a trend toward down-regulation observed in BD; downsampling in the SCZ dataset suggests that the lack of significance in BD may be due to a smaller sample size (fig. S14). This module pair contained as hubs the two major γ-aminobutyric acid (GABA) synthesizing enzymes (*GAD1*, *GAD2*), multiple GABA transporters (*SLC6A1*, *SLC24A3*), many other known interneuron markers (*RELN*, *VIP*), as well as *DLX1* and the lncRNA *DLX6-AS1*, both critical known regulators of inhibitory neuron development (*62*). This inhibitory neuron–related module is not enriched for common or rare genetic disease–associated variation, although other studies have found enrichment for SCZ GWAS signal among interneuron markers defined in other ways (*63*).

Several neuronal modules that distinguish between the disorders differentiate BD and SCZ from ASD, including the module pair geneM21/isoM30 (Fig. 5C), which captures known elements of activity-dependent neuronal gene regulation, whose hubs include classic early-response (*ARC*, *EGR1*, *NPAS4*, *NR4A1*) and late-response genes (*BDNF*, *HOMER1*) (*40*). Although these modules were not significantly down-regulated in ASD, subsampling indicates that the differences between disorders could be driven by sample size (fig. S14). These genes play critical roles in regulating synaptic plasticity and the balance of excitatory and inhibitory synapses (*40*). Of note, a nearly identical module was recently identified as a sex-specific transcriptional signature of major depression and stress susceptibility (*64*). We further observed that these modules may be affected by medication exposure. Indeed, geneM21/isoM30 was associated with genes down-regulated by chronic high doses of the antipsychotic haloperidol, as well as genes up-regulated by the antidepressant fluoxetine (fig. S11A). Furthermore, geneM21/isoM30 expression was negatively correlated with the degree of lifetime antipsychotic exposure in the subset of patients for whom these data were available (P = 0.001, Pearson correlation; fig. S11B). As such, it will be worthwhile to determine whether this module is a core driver of the therapeutic response, as has been suggested (*65*). Finally, other neuronal modules distinguished SCZ and BD from ASD (Fig. 5B), including geneM7, enriched for synaptic and metabolic processes with the splicing regulator *NOVA2* (Fig. 5C). This neuronal module was significantly enriched for both BD and SCZ GWAS signals, supporting a causal role for this module.

Distinct neural-immune trajectories

Previous work has identified differential activation of glial and neural-immune processes in brains from patients with psychiatric disorders (*15*, *52*, *57*, *66*–*69*), including up-regulation of astrocytes in SCZ and BD (*13*, *57*) and both microglia and astrocytes in ASD (*19*, *70*). Evidence supports hyperactive complement-mediated synaptic pruning in SCZ pathophysiology, presumably through microglia (*3*), although postmortem microglial up-regulation was observed only in ASD (*13*, *19*, *70*). We examined whether our large cohort of ~1000 control brains, capturing an age range from birth to 90 years, would enable refinement of the nature and timing of this neuroinflammatory signal and potential relationship to disease pathogenesis (Fig. 7A). Four modules were directly related to neural-immune processes

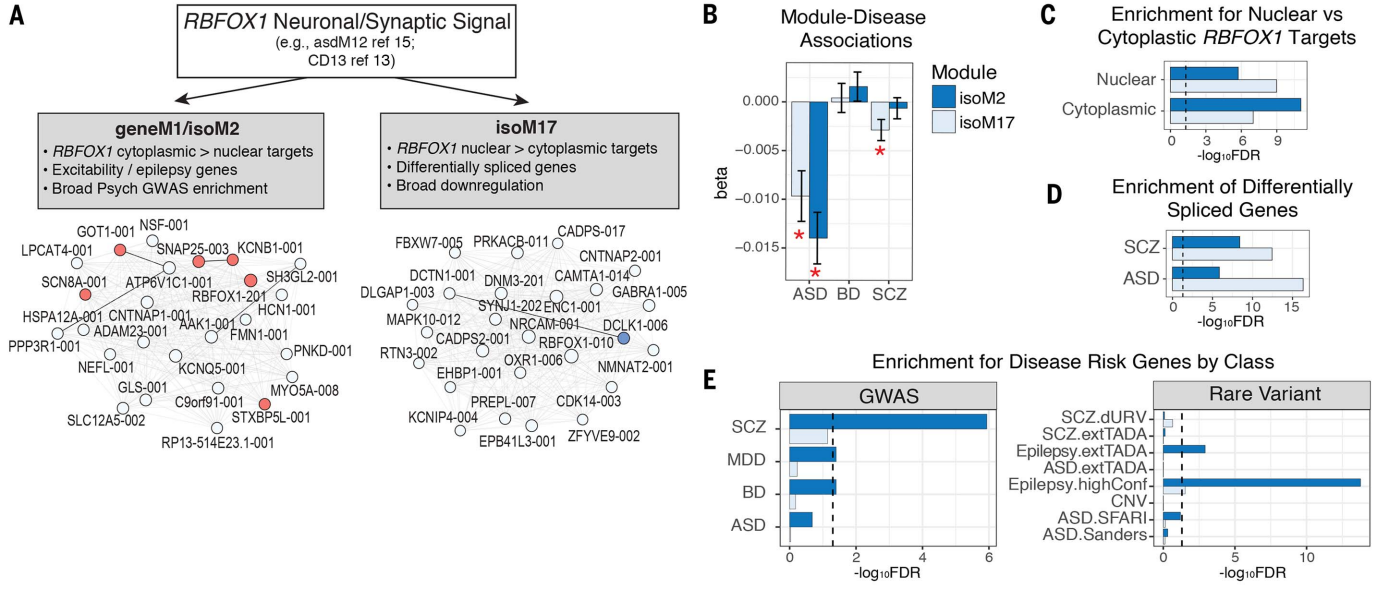


Fig. 6. Two *RBFOX1* isoform modules capture distinct biological and disease associations. (**A**) Previous studies have identified *RBFOX1* as a critical hub of neuronal and synaptic modules down-regulated across multiple psychiatric disorders (*13*, *15*). We identified two pairs of modules with distinct *RBFOX1* isoforms as hub genes. Plots show the top 25 hub genes of modules isoM2 and isoM17, following the same coloring scheme as in Fig. 5C. (**B**) Distinct module-eigengene trait associations are observed for isoM2 (down-regulated in ASD only) compared with isoM17, which is down-regulated in ASD and SCZ. (**C**) Modules show distinct enrichments for nuclear and cytoplasmic *RBFOX1* targets, defined experimentally in mouse (*32*). (**D**) Genes harboring DS events observed in ASD and SCZ show greater overlap with isoM17, consistent with its association with nuclear *RBFOX1* targets. (**E**) Modules show distinct patterns of genetic association. isoM2 exhibits broad enrichment for GWAS signal in SCZ, BD, and MDD, as well as for epilepsy risk genes, whereas isoM17 shows no apparent genetic enrichment (*21*).

(Fig. 7, A to C), two of which are gene/isoform module pairs that correspond clearly to cell type–specific gene expression: one representing microglia (geneM6/isoM15) and the other astrocytes (geneM3/isoM1), as they are strongly and selectively enriched for canonical cell type–specific marker genes (Fig. 7, C to E). Two additional immune-related modules appear to represent more broadly expressed signaling pathways: interferon (IFN) response (geneM32) and NFkB (geneM5/isoM5). The IFN-response module (geneM32) contains critical components of the IFN-stimulated gene factor 3 (ISGF3) complex that activates the transcription of downstream IFN-stimulated genes, which comprise 59 of the 61 genes in this module (*71*). The NFkB module pair (geneM5/isoM5) includes four out of five NFkB family members (*NFkB1*, *NFkB2*, *REL*, *RELA*), as well as many downstream transcription factor targets and upstream activators of this pathway.

The dynamic trajectories of these processes in cases with respect to controls reveal distinct patterns across disorders (Fig. 7F). The IFN-response and microglial modules are most strongly up-regulated in ASD, peaking during early development, coincident with clinical onset. In contrast, in SCZ and BD, the microglial module is actually down-regulated, driven by a later dynamic decrease, dropping below controls after age ~30. The NFkB module, which is up-regulated across all three disorders, maximally diverges from controls during early adulthood, coincident with typical disease onset in SCZ and BD. Accordingly, this NFkB module contained *C4A*, the top GWAS-supported, and strongly up-regulated, risk gene for SCZ (*3*). This pattern is distinct from that of ASD, which shows a dynamic trajectory but remains up-regulated throughout (Fig. 7F).

Noncoding modules and lncRNA regulatory relationships

As many lncRNAs are predicted to have transcriptional regulatory roles, we next assessed whether mRNA-based coexpression networks could provide additional functional annotation for ncRNAs. As a subset of lncRNAs are thought to function by repressing mRNA targets (*72*), we applied csuWGCNA (*73*) to identify potential regulatory relationships (*21*). We identified 39 modules (csuM) using csuWGCNA, all preserved in the signed networks with strong cell type and GWAS enrichments, which captured 7186 negatively correlated lncRNA-mRNA pairs within the same module (fig. S15). We provide a table of putative mRNA targets for these brain-expressed lncRNAs, including 209 exhibiting DE in ASD, 122 in BD, and 241 in SCZ (table S6).

A salient example of the power of this approach for functional annotation is *LINC00473*, a hub of the neuronal activity–dependent gene regulation module (geneM21/isoM30; Fig. 5C). Expressed in excitatory neurons and down-regulated in SCZ (log_2FC −0.16, FDR < 0.002), *LINC00473* is regulated by synaptic activity and down-regulates immediate early gene expression (*74*), consistent with its hub status in this module. Similarly, we identify the lncRNA *DLX6-AS1*, a known developmental regulator of interneuron specification (*62*), as the most central hub gene in the interneuron module (geneM23/isoM19), which is down-regulated in ASD and SCZ. This interneuron module also contains *LINC00643* and *LINC01166*, two poorly annotated, brain-enriched lncRNAs. *LINC00643* is down-regulated in SCZ (log_2FC −0.06, FDR = 0.04), whereas

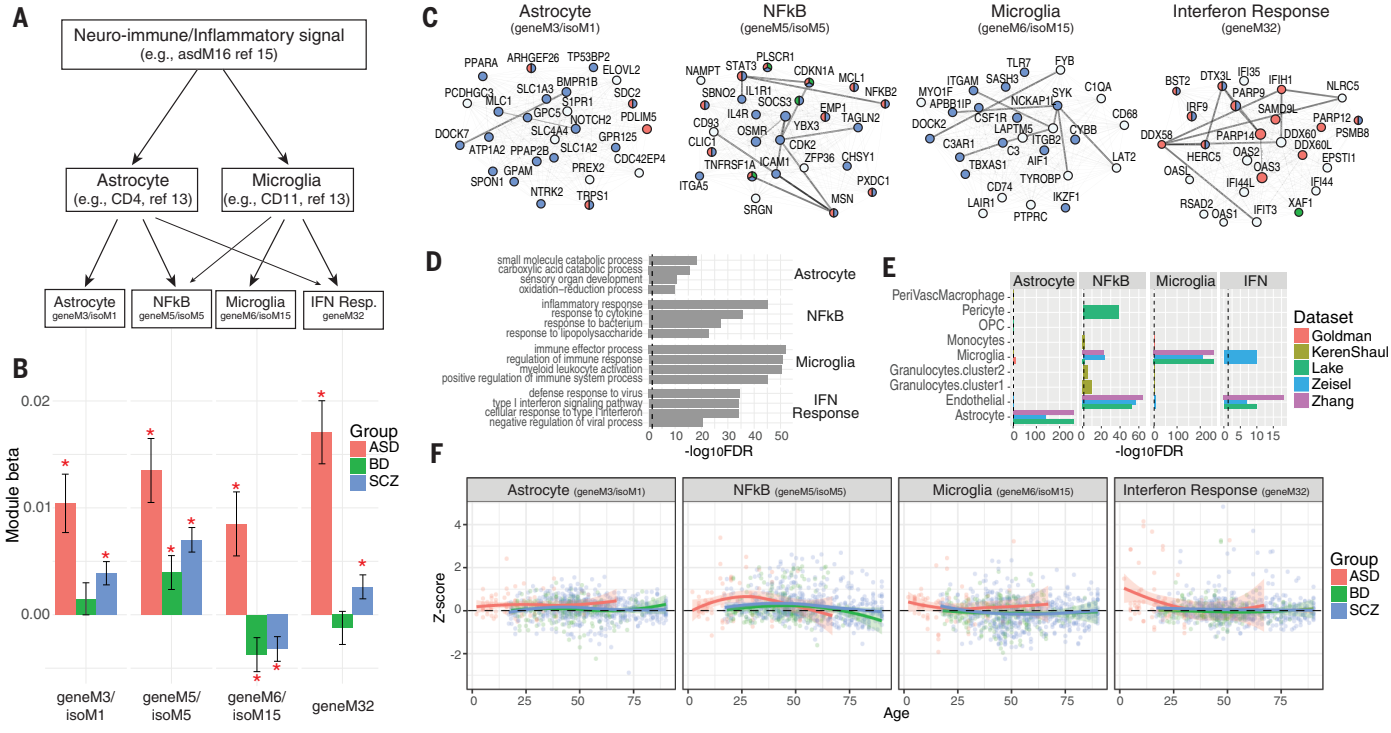


Fig. 7. Distinct neural-immune trajectories in disease. (**A**) Coexpression networks refine the neural-immune/inflammatory processes up-regulated in ASD, SCZ, and BD. Previous work has identified specific contributions to this signal from astrocyte and microglial populations (*13*, *19*). Here, we identify additional contributions from distinct IFN-response and NFkB signaling modules. (**B**) Eigengene-disease associations are shown for each of four identified neural-immune module pairs. The astrocyte and IFN-response modules are up-regulated in ASD and SCZ. NFkB signaling is elevated across all three disorders. The microglial module is up-regulated in ASD and down-regulated in SCZ and BD. (**C**) Top hub genes for each module are shown, along with edges supported by coexpression (light gray; Pearson correlation > 0.5) and known protein-protein interactions (dark lines). Nodes follow the same coloring scheme as in Fig. 5C. Hubs in the astrocyte module (geneM3/isoM1) include several canonical, specific astrocyte markers, including *SOX9*, *GJA1*, *SPON1*, and *NOTCH2*. Microglial module hub genes include canonical, specific microglial markers, including *AIF1*, *CSF1R*, *TYROBP*, and *TMEM119*. The NFkB module includes many known downstream transcription factor targets (*JAK3*, *STAT3*, *JUNB*, and *FOS*) and upstream activators (*IL1R1*, nine TNF receptor superfamily members) of this pathway. (**D**) The top four GO enrichments are shown for each module. (**E**) Module enrichment for known cell type–specific marker genes, collated from sequencing studies of neural-immune cell types (*98*–*102*). (**F**) Module eigengene expression across age demonstrates distinct and dynamic neural-immune trajectories for each disorder.

LINC01166 is significantly down-regulated in BD (log_2FC −0.17, FDR < 0.05) with trends in ASD and SCZ (FDR < 0.1). Our data suggest a role for these lncRNAs in interneuron development, making them intriguing candidates for follow-up studies. Using fluorescence in situ hybridization (FISH), we confirmed that both *LINC00643* and *LINC01166* are expressed in GAD1$^+$ GABAergic neurons in area 9 of the adult brain, present both in the cell nucleus and the cytoplasm (Fig. 8A and fig. S16), although expression was also detected in other non-GAD1$^+$ neurons as well.

Multiple ncRNAs including *SOX2-OT*, *MIAT*, and *MEG3* are enriched in oligodendrocyte modules (geneM2/isoM13/csuM1; Fig. 5C) that are down-regulated in both SCZ and ASD. *SOX2-OT* is a heavily spliced, evolutionarily conserved lncRNA exhibiting predominant brain expression and a hub of these oligodendrocyte modules, without previous mechanistic links to myelination (*75*, *76*). The lncRNAs *MIAT* and *MEG3* are negatively correlated with most of the hubs in this module, including *SOX2-OT* (fig. S15). *MIAT* is also known to interact with *QKI*, an established regulator of oligodendrocyte-gene splicing also located in this module (*77*, *78*). These analyses predict critical roles for these often overlooked noncoding genes in oligodendrocyte function (*77*, *78*) and potentially in psychiatric conditions.

Isoform network specificity and switching

To more comprehensively assess whether aspects of disease specificity are conferred by alternative transcript usage or splicing, versus DE, we surveyed genes exhibiting DTU across disorders (*21*). We identified 134 such "switch isoforms," corresponding to 64 genes displaying different DTU between ASD and SCZ (table S7). As an example, isoforms of *SMARCA2*, a member of the BAF-complex strongly implicated in several neurodevelopmental disorders including ASD (*79*), are up- and down-regulated in ASD and SCZ, respectively (fig. S17). Conversely, the isoforms of *NIPBL*, a gene associated with Cornelia de Lange syndrome (*80*), are down- and up-regulated in ASD and SCZ, respectively (fig. S17). Such opposing changes in isoform expression of various genes may represent differences in disease progression or symptom manifestation in diseases such as ASD and SCZ, mediated by genetic risk variants that create

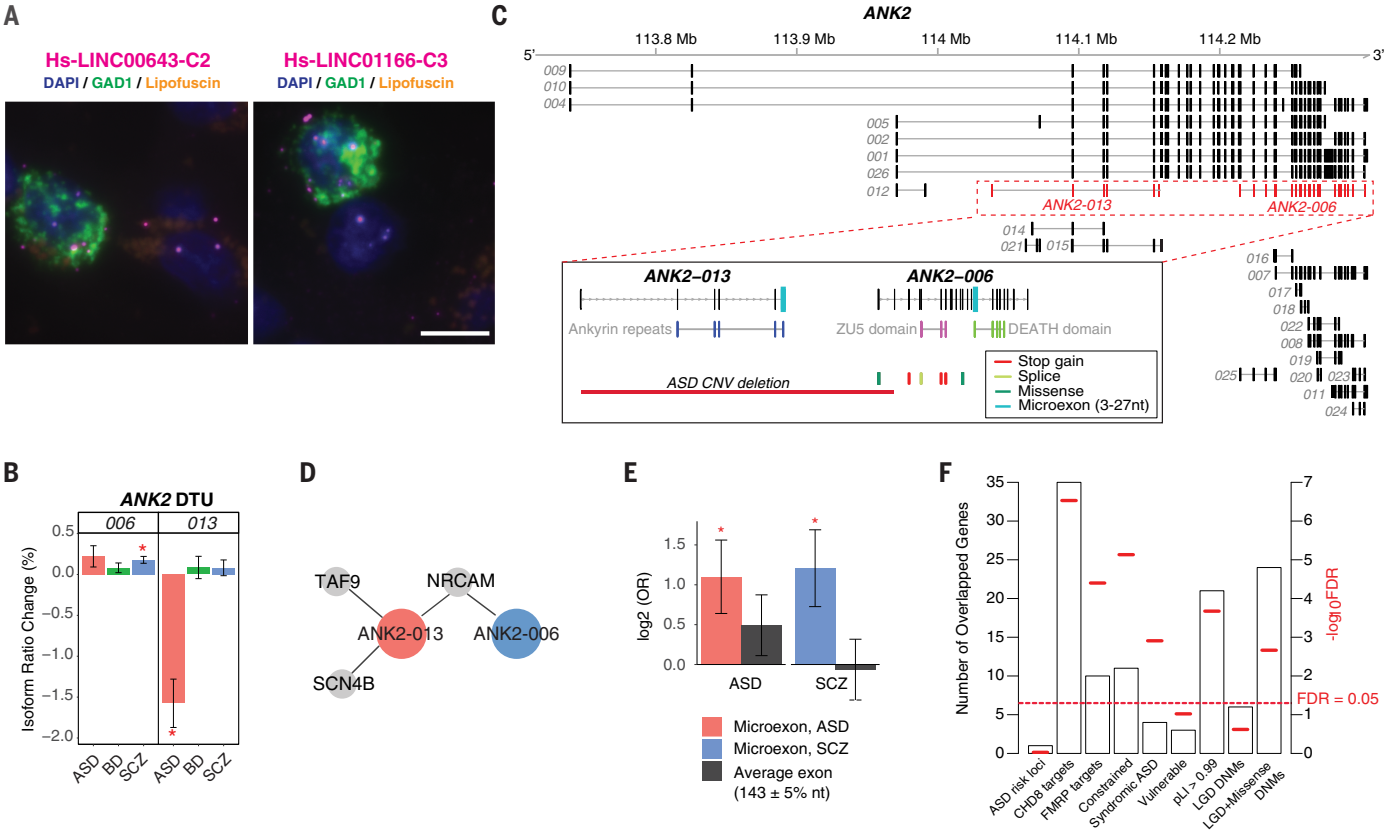


Fig. 8. LncRNA annotation, *ANK2* isoform switching, and microexon enrichment. (**A**) FISH images demonstrate interneuron expression for two poorly annotated lincRNAs—*LINC00643* and *LINC01166*—in area 9 of adult human prefrontal cortex. Sections were labeled with *GAD1* probe (green) to indicate GABAergic neurons and lncRNA (magenta) probes for *LINC00643* (left) or for *LINC01166* (right). All sections were counterstained with DAPI (blue) to reveal cell nuclei. Lipofuscin autofluorescence is visible in both the green and red channels and appears orange. Scale bar, 10 μm. FISH was repeated at least twice on independent samples (table S9) (*21*), with similar results (see also fig. S16). (**B**) *ANK2* isoforms *ANK2-006* and *ANK2-013* show significant DTU in SCZ and ASD, respectively (*FDR < 0.05). (**C**) Exon structure of *ANK2* highlighting (dashed lines) the *ANK2-006* and *ANK2-013* isoforms. (Inset) These isoforms have different protein domains and carry different microexons. *ANK2-006* is affected by multiple ASD DNMs, while *ANK2-013* could be entirely eliminated by a de novo CNV deletion in ASD. (**D**) Disease-specific coexpressed PPI network. Both *ANK2-006* and *ANK2-013* interact with *NRCAM*. The ASD-associated isoform *ANK2-013* has two additional interacting partners, *SCN4B* and *TAF9*. (**E**) As a class, switch isoforms are significantly enriched for microexon(s). In contrast, exons of average length are not enriched among switch isoforms. The *y* axis displays odds ratio on a log_2 scale. *P* values are calculated using logistic regression and corrected for multiple comparisons. (**F**) Enrichment of 64 genes with switch isoforms for: ASD risk loci (*81*); CHD8 targets (*103*); FMRP targets (*33*); mutationally constraint genes (*104*); syndromic and highly ranked (1 and 2) genes from SFARI Gene database; vulnerable ASD genes (*105*); genes with probability of loss-of-function intolerance (pLI) > 0.99 as reported by the Exome Aggregation Consortium (*106*); genes with likely-gene-disruption (LGD) or LGD plus missense de novo mutations (DNMs) found in patients with neurodevelopmental disorders (*21*).

subtle differences in isoforms within the same gene that exhibit distinct biological effects in each disorder. A noteworthy example is the ASD risk gene *ANK2* (*81*), whose two alternatively spliced isoforms, *ANK2-006* and *ANK2-013*, are differentially regulated in SCZ and ASD (Fig. 8B). These switch isoforms show markedly different expression patterns, belonging to different coexpression modules, geneM3/isoM1 (Fig. 7C) and isoM18, which are enriched in astrocyte and neuronal cell types, respectively (Fig. 5A and fig. S12). The protein domain structure of these transcripts is also nonoverlapping, with *ANK2-006* carrying exclusively ZU5 and DEATH domains and *ANK2-013* carrying exclusively ankyrin repeat domains (Fig. 8C). Both isoforms are affected by a de novo ASD CNV, and *ANK-006* also carries de novo mutations from neurodevelopmental disorders. Both isoforms bind to the neuronal cell adhesion molecule *NRCAM*, but *ANK2-013* has two additional partners: *TAF9* and *SCN4B* (Fig. 8D), likely cell type–specific interactions that suggest distinct functions of the isoforms of this gene in different neural cell types and diseases.

Finally, several studies have demonstrated that genes carrying microexons are preferentially expressed in the brain and their splicing is dysregulated in ASD (*30*, *82*, *83*). This PsychENCODE sample provided the opportunity to assess the role of microexons in a far larger cohort and across disorders. Indeed, we found that switch isoforms with microexons (3 to 27 base pairs) are significantly enriched in both ASD (FDR = 0.03) and SCZ (FDR = 0.03, logistic regression) (Fig. 8E) (*21*). Genes with switch isoforms are also enriched for the regulatory targets of two ASD risk genes, *CHD8* and *FMRP*, as well as highly mutationally constrained genes (pLI > 0.99), syndromic ASD genes, and in genes with de novo exonic mutations in ASD, SCZ, and BD (Fig. 8F and table S7) (*21*). These data confirm the importance of microexon regulation in neuropsychiatric disorders beyond ASD, and its potential role in distinguishing among biological pathways differentially affected across conditions. This role for microexons further highlights local splicing regulation as a potential mechanism conferring key aspects of disease specificity, extending the larger disease signal observed at the isoform level in coexpression and DE analyses.

Discussion

We present a large-scale RNA-seq analysis of the cerebral cortex across three major psychiatric disorders, including extensive analyses of the noncoding and alternatively spliced transcriptome, as well as gene- and isoform-level coexpression networks. The scope and complexity of these data do not immediately lend themselves to simple mechanistic reduction. Nevertheless, at each level of analysis, we present concrete examples that provide proofs-of-principle and starting points for investigations targeting shared and distinct disease mechanisms to connect causal drivers with brain-level perturbations.

Broadly, we find that isoform-level changes exhibit the largest effect sizes in diseased brain, are most enriched for genetic risk, and provide the greatest disease specificity when assembled into coexpression networks. Notably, disturbances in the expression of distinct isoforms of more than 50 genes are differentially observed in SCZ and ASD, which in the case of the ASD risk gene *ANK2* is predicted to affect different cell types in each disorder. Moreover, we observe disease-associated changes in the splicing of dozens of RNA-binding proteins and splicing factors, most of whose targets and functions are unknown. Similarly, nearly 1000 ncRNAs are dysregulated in at least one disorder, many with significant CNS enrichment but, until now, limited functional annotation.

This work highlights isoform-level dysregulation as a critical, and relatively underexplored, proximal mechanism linking genetic risk factors with psychiatric disease pathophysiology. In contrast to local splicing changes, isoform-level quantifications require imputation from short-read RNA-seq data guided by existing genomic annotations. Consequently, the accuracy of these estimates is hindered by incomplete annotations, as well as by limitations of short-read sequencing, coverage, and genomic biases like GC content (*84*, *85*). This may be particularly problematic in the brain, where alternative splicing patterns are more distinct than in other organ systems (*82*). We present experimental validations for several specific isoforms but try to focus on the class of dysregulated isoforms, and the modules and biological processes they represent, rather than individual cases, which may be more susceptible to bias. Longer-read sequencing, which provides a more precise means for isoform quantification, will be of great utility as it becomes more feasible at scale.

Several broad shared patterns of gene expression dysregulation have been observed in postmortem brain samples in previous studies—most prominently, a gradient of down-regulation of neuronal and synaptic signaling genes and up-regulation of glial-immune or neuroinflammatory signals. In this study, we refine these signals by distinguishing both up and down-regulated neuron-related processes that are differentially altered across these three disorders. Furthermore, we extend previous work that identified broad neuroinflammatory dysregulation in SCZ, ASD, and BD by identifying specific pathways involving IFN-response, NFkB, astrocytes, and microglia that manifest distinct temporal patterns across conditions. A module enriched for microglial-associated genes, for example, shows a clear distinction between disorders, with strong up-regulation observed in ASD and significant down-regulation in SCZ and BD. Overall, these results provide increased specificity to the observations that ASD, BD, and SCZ are associated with elevated neuroinflammatory processes (*69*, *86*–*88*).

By integrating transcriptomic data with genetic variation, we identify multiple disease-associated coexpression modules enriched for causal variation, as well as mechanisms potentially underlying specific disease loci in each of the diseases. In parallel, by performing a well-powered brain-relevant TWAS in SCZ, and to a lesser extent in BD and ASD, we are further able to elucidate candidate molecular mechanisms through which disease-associated variants may act. TWAS prioritizes dozens of previously unidentified candidate disease genes, including many that are dysregulated in diseased brain. Similar to the eQTLs identified in a companion study (*17*), the majority of these loci do not overlap with disease GWAS association signals. Rather, most are outside of the LD block and distal to the original association signal, highlighting the importance of orthogonal functional data types, such as transcriptome or epigenetic data (*16*, *47*, *82*, *89*), in deciphering the underlying mechanisms of disease-associated genetic effects.

As with any case-control association study, multiple potential factors, many of which may represent reactive processes, contribute to gene expression changes in postmortem human brain samples. At each step of analysis, we have attempted to mitigate the contribution of these factors through known and hidden covariate correction, assessment of age trajectories, and enrichment for causal genetic variation. Supporting the generalizability of our results, we find significant correlations of the $\log_2$FC between randomly split halves of the data (fig. S3). This likely varies by transcript class, and some of the modest correlations are likely due to low-abundance genes, such as ncRNAs, which we prefer to include, though we recognize the inherent tension between expression level and measurement accuracy. We provide access to this extensive resource, both in terms of raw and processed data and as browsable network modules (Resource.PsychENCODE.org).

A large proportion of disease-associated coexpression modules are enriched for cell type–specific markers, as is overall disease DE signal, indicating that transcriptomic alterations in disease are likely driven substantially by (even subtle) shifts in cell type proportions, or cell type–specific pathways, consistent with our previous observations (*13*) and those in a companion study (*17*). Functional genomic studies often remove such cell type–specific signals, through the use of large numbers of expression-derived principal components or surrogate variables as covariates, to mitigate unwanted sources of variation and maximize detection of cis eQTLs (*44*). We retain the cell type–specific signals as much as possible, reasoning that cell type–related alterations may directly inform the molecular pathology of disease in psychiatric disorders, in which there is no known microscopic or macroscopic pathology. This rationale is supported by the consistent observation of the dynamic and disease-specific microglial up-regulation observed in ASD and the shared astrocyte up-regulation in SCZ and ASD. This approach, however, reduces the ability to detect genetic enrichment from GWAS, as current methods predominantly capture cis-acting regulatory effects. The modesty of genetic enrichments among disease-associated transcriptomic alterations may also indicate that gene expression changes reflect an indirect cascade of

molecular events triggered by environmental as well as genetic factors or that genetic factors may act earlier, such as during development.

Finally, these data, while providing a unique, large-scale resource for the field, also suggest that profiling additional brains, especially from other implicated brain regions, will continue to be informative. Similarly, these data suggest that although isoform-level analyses, including the identification of isoform-specific protein-protein interactions (PPI) and cell type specificity, pose major challenges for high-throughput studies, they are likely to add substantial value to our understanding of brain function and neuropsychiatric disorders. Finally, as GWAS studies in ASD and BD increase in size and subsequently in power, their continued integration with these transcriptome data will likely prove critical in identifying the functional impact of disease-associated genetic variation.

Materials and methods summary

The data generated for this manuscript represent Freeze 1 and 2 of the PsychENCODE Consortium dataset. Postmortem human brain samples were collected as part of eight studies, detailed in fig. S1. RNA-seq and genotype array data were generated by each site and then processed together through a unified pipeline (fig. S1) by a central data analysis core. Raw data are available at (*90*), with processed summary-level data available at http://Resource.PsychENCODE.org.

For this study, we restricted analysis to frontal and temporal cortex brain samples from postnatal time points with at least 10 million total reads (fig. S2). RNA-seq reads were aligned to the GRCh37.p13 (hg19) reference genome via STAR 2.4.2a with comprehensive gene annotations from Gencode v19. Gene- and isoform-level quantifications were calculated using RSEM v1.2.29 (*25*). QC metrics were calculated from PicardTools v1.128, RNA-SeQC v1.1.8, featureCounts v1.5.1, cutadapt, and STAR. This generated a matrix of 187 QC metrics, which was then summarized by its top principal components, which were used as covariates in downstream analyses.

Genes were filtered to include only those on autosomes longer than 250 base pairs with transcripts per million reads (TPM) > 0.1 in at least 25% of samples, removing immunoglobulin biotypes. Outlier samples with discordant sex or low network connectivity *z*-scores were identified within each individual study and removed (*91*). Surrogate variable analysis was performed to identify hidden confounding factors (*43*). Count-level quantifications were corrected for library size by using trimmed mean of M-values (TMM) normalization and were $\log_2$ transformed. DE was assessed using a linear mixed effects model, accounting for known biological, technical, and four surrogate variables as fixed effects and subject-level technical replicates as random effects. Analogous assessments of DTE and DTU were performed using isoform-level expression quantifications and isoform ratios, respectively. *P* values were corrected for multiple testing using the Benjamini-Hochberg method, with significance set at 5%. Stratified LD-score regression (*39*) was used to investigate GWAS enrichment among DE gene sets. DE ncRNAs were further annotated for tissue-specificity using GTEx v6 data (*10*, *82*), evolutionary conservation using phyloP and phastCons scores (*92*, *93*), and exon-level selective constraint via CDTS (*27*). DS analysis was performed using LeafCutter (*29*), controlling for the same covariates as above after randomly selecting a single technical replicate for each distinct subject.

Robust WGCNA was performed to identify signed coexpression modules using gene- and isoform-level quantifications separately, after first regressing out all covariates except for the diagnostic group (*94*). Modules were summarized by their first principal component (eigengene), and disease associations were evaluated using a linear mixed-effects model as above. Significance values were FDR-corrected to account for multiple comparisons.

Genotype calls from SNP arrays were generated at each data production site separately and centralized for imputation, as detailed in a companion manuscript (*17*). Parallel haplotype prephasing and imputation were done using Eagle2, Minimac3, with the HRC reference panel for imputation. Calculation of gene-level eQTL and isoform-level expression QTLs (isoQTL) was done using QTLtools, as described in a companion manuscript (*17*). PRS were calculated for individuals of European ancestry using LDPred (*95*) with GWAS summary statistics and 1000 Genomes Phase 3 European subset as an LD reference panel.

TWAS was performed using the FUSION package [http://gusevlab.org/projects/fusion/ (*46*)] with custom SNP-expression weights generated from our adult transcriptome dataset. We used GCTA (*96*) to estimate cis SNP heritability for each gene in our dataset, and analysis was restricted to those exhibiting significant heritability (cis h^2_g $P < 0.05$). Association statistics were Bonferroni corrected ($P < 0.05$). SMR and the associated HEIDI test were performed as implemented in the SMR software package [http://cnsgenomics.com/software/smr/ (*48*)]. Experimental validations of selected splicing and isoform-level changes were performed using RT-PCR. See the supplementary materials and methods for full details.

REFERENCES AND NOTES

1. H. A. Whiteford, A. J. Ferrari, L. Degenhardt, V. Feigin, T. Vos, The global burden of mental, neurological and substance use disorders: An analysis from the Global Burden of Disease Study 2010. *PLOS ONE* **10**, e0116820 (2015). doi: 10.1371/journal.pone.0116820; pmid: 25658103
2. M. J. Gandal, V. Leppa, H. Won, N. N. Parikshak, D. H. Geschwind, The road to precision psychiatry: Translating genetics into disease mechanisms. *Nat. Neurosci.* **19**, 1397–1407 (2016). doi: 10.1038/nn.4409; pmid: 27786179
3. A. Sekar *et al.*, Schizophrenia risk from complex variation of complement component 4. *Nature* **530**, 177–183 (2016). doi: 10.1038/nature16549; pmid: 26814963
4. L. de la Torre-Ubieta, H. Won, J. L. Stein, D. H. Geschwind, Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361 (2016). doi: 10.1038/nm.4071; pmid: 27050589
5. M. T. Maurano *et al.*, Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012). doi: 10.1126/science.1222794; pmid: 22955828
6. L. D. Ward, M. Kellis, Interpreting noncoding genetic variation in complex traits and human disease. *Nat. Biotechnol.* **30**, 1095–1106 (2012). doi: 10.1038/nbt.2422; pmid: 23138309
7. A. Visel, E. M. Rubin, L. A. Pennacchio, Genomic views of distant-acting enhancers. *Nature* **461**, 199–205 (2009). doi: 10.1038/nature08451; pmid: 19741700
8. K. G. Ardlie *et al.*, Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015). doi: 10.1126/science.1262110; pmid: 25954001
9. S. K. Reilly *et al.*, Evolutionary genomics. Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015). doi: 10.1126/science.1260943; pmid: 25745175
10. The ENCODE Project Consortium, Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799–816 (2007). doi: 10.1038/nature05874; pmid: 17571346
11. A. Kundaje *et al.*, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015). doi: 10.1038/nature14248; pmid: 25693563
12. R. Andersson *et al.*, An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461 (2014). doi: 10.1038/nature12787; pmid: 24670763
13. M. J. Gandal *et al.*, Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018). doi: 10.1126/science.aad6469; pmid: 29439242
14. N. N. Parikshak, M. J. Gandal, D. H. Geschwind, Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. *Nat. Rev. Genet.* **16**, 441–458 (2015). doi: 10.1038/nrg3934; pmid: 26149713
15. I. Voineagu *et al.*, Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011). doi: 10.1038/nature10110; pmid: 21614001
16. S. Akbarian *et al.*, The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015). doi: 10.1038/nn.4156; pmid: 26605881
17. D. Wang *et al.*, Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, eaat8464 (2018).
18. M. Fromer *et al.*, Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016). doi: 10.1038/nn.4399; pmid: 27668389
19. N. N. Parikshak *et al.*, Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016). doi: 10.1038/nature20612; pmid: 27919067
20. H. J. Kang *et al.*, Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011). doi: 10.1038/nature10523; pmid: 22031440
21. See supplementary materials and methods.
22. A. E. Jaffe *et al.*, qSVA framework for RNA quality correction in differential expression analysis. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 7130–7135 (2017). doi: 10.1073/pnas.1617384114; pmid: 28634288
23. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011). doi: 10.1186/1471-2105-12-323; pmid: 21816040
24. A. E. Jaffe *et al.*, Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat. Neurosci.* **21**, 1117–1125 (2018). doi: 10.1038/s41593-018-0197-y; pmid: 30050107
25. Y. I. Li *et al.*, RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604 (2016). doi: 10.1126/science.aad9417; pmid: 27126046
26. P. J. Batista, H. Y. Chang, Long noncoding RNAs: Cellular address codes in development and disease. *Cell* **152**, 1298–1307 (2013). doi: 10.1016/j.cell.2013.02.012; pmid: 23498938
27. J. di Iulio *et al.*, The human noncoding genome defined by genetic diversity. *Nat. Genet.* **50**, 333–337 (2018). doi: 10.1038/s41588-018-0062-7; pmid: 29483654
28. C. K. Vuong, D. L. Black, S. Zheng, The neurogenetics of alternative splicing. *Nat. Rev. Neurosci.* **17**, 265–281 (2016). doi: 10.1038/nrn.2016.27; pmid: 27094079

29. Y. I. Li *et al*., Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158 (2018). doi: 10.1038/s41588-017-0004-9; pmid: 29229983
30. M. Irimia *et al*., A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159**, 1511–1523 (2014). doi: 10.1016/j.cell.2014.11.035; pmid: 25525873
31. N. Akula *et al*., RNA-sequencing of the brain transcriptome implicates dysregulation of neuroplasticity, circadian rhythms and GTPase binding in bipolar disorder. *Mol. Psychiatry* **19**, 1179–1185 (2014). doi: 10.1038/mp.2013.170; pmid: 24393808
32. J.-A. Lee *et al*., Cytoplasmic Rbfox1 Regulates the Expression of Synaptic and Autism-Related Genes. *Neuron* **89**, 113–128 (2016). doi: 10.1016/j.neuron.2015.11.025; pmid: 26687839
33. J. C. Darnell *et al*., FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011). doi: 10.1016/j.cell.2011.06.013; pmid: 21784246
34. E. Sebestyén *et al*., Large-scale analysis of genome and transcriptome alterations in multiple tumors unveils novel cancer-relevant splicing networks. *Genome Res.* **26**, 732–744 (2016). doi: 10.1101/gr.199935.115; pmid: 27197215
35. J. Gauthier *et al*., Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia. *Hum. Genet.* **130**, 563–573 (2011). doi: 10.1007/s00439-011-0975-z; pmid: 21424692
36. F. F. Hamdan *et al*., Excess of de novo deleterious mutations in genes associated with glutamatergic systems in nonsyndromic intellectual disability. *Am. J. Hum. Genet.* **88**, 306–316 (2011). doi: 10.1016/j.ajhg.2011.02.001; pmid: 21376300
37. B. Treutlein, O. Gokce, S. R. Quake, T. C. Südhof, Cartography of neurexin alternative splicing mapped by single-molecule long-read mRNA sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E1291–E1299 (2014). doi: 10.1073/pnas.1403244111; pmid: 24639501
38. J. Grove *et al*., Common risk variants identified in autism spectrum disorder. bioRxiv 224774 [Preprint]. 27 November 2017. .doi: 10.1101/224774
39. H. K. Finucane *et al*., Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015). doi: 10.1038/ng.3404; pmid: 26414678
40. A. E. West, M. E. Greenberg, Neuronal activity-regulated gene transcription in synapse development and cognitive function. *Cold Spring Harb. Perspect. Biol.* **3**, a005744 (2011). doi: 10.1101/cshperspect.a005744; pmid: 21555405
41. Y. Zhu, L. Wang, Y. Yin, E. Yang, Systematic analysis of gene expression patterns associated with postmortem interval in human tissues. *Sci. Rep.* **7**, 5435 (2017). doi: 10.1038/s41598-017-05882-0; pmid: 28710439
42. J. Z. Li *et al*., Systematic changes in gene expression in postmortem human brains associated with tissue pH and terminal medical conditions. *Hum. Mol. Genet.* **13**, 609–616 (2004). doi: 10.1093/hmg/ddh065; pmid: 14734628
43. J. T. Leek, J. D. Storey, Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLOS Genet.* **3**, e161 (2007). doi: 10.1371/journal.pgen.0030161; pmid: 17907809
44. GTEx Consortium, Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017). doi: 10.1038/nature24277; pmid: 29022597
45. 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
46. A. Gusev *et al*., Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016). doi: 10.1038/ng.3506; pmid: 26854917
47. E. R. Gamazon *et al*., A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015). doi: 10.1038/ng.3367; pmid: 26258848
48. Z. Zhu *et al*., Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016). doi: 10.1038/ng.3538; pmid: 27019110
49. M. C. Oldham *et al*., Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11**, 1271–1282 (2008). doi: 10.1038/nn.2207; pmid: 18849986
50. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, e17 (2005). doi: 10.2202/1544-6115.1128; pmid: 16646834
51. J. A. Miller, M. C. Oldham, D. H. Geschwind, A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J. Neurosci.* **28**, 1410–1420 (2008). doi: 10.1523/JNEUROSCI.4098-07.2008; pmid: 18256261
52. C. Chen *et al*., Two gene co-expression modules differentiate psychotics and controls. *Mol. Psychiatry* **18**, 1308–1314 (2013). doi: 10.1038/mp.2012.146; pmid: 23147385
53. R. Daneman *et al*., The mouse blood-brain barrier transcriptome: A new resource for understanding the development and function of brain endothelial cells. *PLOS ONE* **5**, e13741 (2010). doi: 10.1371/journal.pone.0013741; pmid: 21060791
54. B. Obermeier, R. Daneman, R. M. Ransohoff, Development, maintenance and disruption of the blood-brain barrier. *Nat. Med.* **19**, 1584–1596 (2013). doi: 10.1038/nm.3407; pmid: 24309662
55. G. Genovese *et al*., Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016). doi: 10.1038/nn.4402; pmid: 27694994
56. S. E. Ellis, R. Panitch, A. B. West, D. E. Arking, Transcriptome analysis of cortical tissue reveals shared sets of downregulated genes in autism and schizophrenia. *Transl. Psychiatry* **6**, e817 (2016). doi: 10.1038/tp.2016.87; pmid: 27219343
57. R. C. Ramaker *et al*., Post-mortem molecular profiling of three psychiatric disorders. *Genome Med.* **9**, 72 (2017). doi: 10.1186/s13073-017-0458-5; pmid: 28754123
58. C. K. Vuong *et al*., Rbfox1 Regulates Synaptic Transmission through the Inhibitory Neuron-Specific vSNARE Vamp1. *Neuron* **98**, 127–141.e7 (2018). doi: 10.1016/j.neuron.2018.03.008; pmid: 29621484
59. A. F. Pardiñas *et al*., Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018). doi: 10.1038/s41588-018-0059-2; pmid: 29483656
60. L. T. Gehman *et al*., The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* **43**, 706–711 (2011). doi: 10.1038/ng.841; pmid: 21623373
61. B. L. Fogel *et al*., RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* **21**, 4171–4186 (2012). doi: 10.1093/hmg/dds240; pmid: 22730494
62. A. M. Bond *et al*., Balanced gene regulation by an embryonic brain ncRNA is critical for adult hippocampal GABA circuitry. *Nat. Neurosci.* **12**, 1020–1027 (2009). doi: 10.1038/nn.2371; pmid: 19620975
63. N. G. Skene *et al*., Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* **50**, 825–833 (2018). doi: 10.1038/s41588-018-0129-5; pmid: 29785013
64. B. Labonté *et al*., Sex-specific transcriptional signatures in human depression. *Nat. Med.* **23**, 1102–1111 (2017). doi: 10.1038/nm.4386; pmid: 28825715
65. A. de Bartolomeis *et al*., Immediate-Early Genes Modulation by Antipsychotics: Translational Implications for a Putative Gateway to Drug-Induced Long-Term Brain Changes. *Front. Behav. Neurosci.* **11**, 240 (2017). doi: 10.3389/fnbeh.2017.00240; pmid: 29321734
66. R. Birnbaum *et al*., Investigating the neuroimmunogenic architecture of schizophrenia. *Mol. Psychiatry* **23**, 1251–1260 (2018). doi: 10.1038/mp.2017.89; pmid: 28485405
67. S. G. Fillman *et al*., Increased inflammatory markers identified in the dorsolateral prefrontal cortex of individuals with schizophrenia. *Mol. Psychiatry* **18**, 206–214 (2013). doi: 10.1038/mp.2012.110; pmid: 22869038
68. R. Pacifico, R. L. Davis, Transcriptome sequencing implicates dorsal striatum-specific gene network, immune response and energy metabolism pathways in bipolar disorder. *Mol. Psychiatry* **22**, 441–449 (2017). doi: 10.1038/mp.2016.94; pmid: 27350034
69. J. S. Rao, G. J. Harry, S. I. Rapoport, H. W. Kim, Increased excitotoxicity and neuroinflammatory markers in postmortem frontal cortex from bipolar disorder patients. *Mol. Psychiatry* **15**, 384–392 (2010). doi: 10.1038/mp.2009.47; pmid: 19488045
70. S. Gupta *et al*., Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat. Commun.* **5**, 5748 (2014). doi: 10.1038/ncomms6748; pmid: 25494366
71. I. Rusinova *et al*., Interferome v2.0: An updated database of annotated interferon-regulated genes. *Nucleic Acids Res.* **41**, D1040–D1046 (2013). doi: 10.1093/nar/gks1215; pmid: 23203888
72. A. Necsulea *et al*., The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* **505**, 635–640 (2014). doi: 10.1038/nature12943; pmid: 24463510
73. R. Dai, Y. Xia, C. Liu, C. Chen, csuWGCNA: A combination of signed and unsigned WGCNA to capture negative correlations. bioRxiv 288225 [Preprint]. 27 September 2018. . doi: 10.1101/288225
74. P. Pruunsild, C. P. Bengtson, H. Bading, Networks of Cultured iPSC-Derived Neurons Reveal the Human Synaptic Activity-Regulated Adaptive Gene Program. *Cell Reports* **18**, 122–135 (2017). doi: 10.1016/j.celrep.2016.12.018; pmid: 28052243
75. P. P. Amaral *et al*., Complex architecture and regulated expression of the Sox2ot locus during vertebrate development. *RNA* **15**, 2013–2027 (2009). doi: 10.1261/rna.1705309; pmid: 19767420
76. T. R. Mercer, M. E. Dinger, S. M. Sunkin, M. F. Mehler, J. S. Mattick, Specific expression of long noncoding RNAs in the mouse brain. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 716–721 (2008). doi: 10.1073/pnas.0706729105; pmid: 18184812
77. K. Aberg, P. Saetre, N. Jareborg, E. Jazin, Human QKI, a potential regulator of mRNA expression of human oligodendrocyte-related genes involved in schizophrenia. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7482–7487 (2006). doi: 10.1073/pnas.0601213103; pmid: 16641098
78. G. Barry *et al*., The long non-coding RNA Gomafu is acutely regulated in response to neuronal activation and involved in schizophrenia-associated alternative splicing. *Mol. Psychiatry* **19**, 486–494 (2014). doi: 10.1038/mp.2013.45; pmid: 23628989
79. M. Koga *et al*., Involvement of SMARCA2/BRM in the SWI/SNF chromatin-remodeling complex in schizophrenia. *Hum. Mol. Genet.* **18**, 2483–2494 (2009). doi: 10.1093/hmg/ddp166; pmid: 19363039
80. I. D. Krantz *et al*., Cornelia de Lange syndrome is caused by mutations in NIPBL, the human homolog of Drosophila melanogaster Nipped-B. *Nat. Genet.* **36**, 631–635 (2004). doi: 10.1038/ng1364; pmid: 15146186
81. S. J. Sanders *et al*., Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015). doi: 10.1016/j.neuron.2015.09.016; pmid: 26402605
82. M. Melé *et al*., The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015). doi: 10.1126/science.aaa0355; pmid: 25954002
83. M. Quesnel-Vallières *et al*., Misregulation of an Activity-Dependent Splicing Network as a Common Mechanism Underlying Autism Spectrum Disorders. *Mol. Cell* **64**, 1023–1034 (2016). doi: 10.1016/j.molcel.2016.11.033; pmid: 27984743
84. T. Steijger *et al*., Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods* **10**, 1177–1184 (2013). doi: 10.1038/nmeth.2714; pmid: 24185837
85. M. I. Love, J. B. Hogenesch, R. A. Irizarry, Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. *Nat. Biotechnol.* **34**, 1287–1291 (2016). doi: 10.1038/nbt.3682; pmid: 27669167
86. M. L. Estes, A. K. McAllister, Immune mediators in the brain and peripheral tissues in autism spectrum disorder. *Nat. Rev. Neurosci.* **16**, 469–486 (2015). doi: 10.1038/nrn3978; pmid: 26189694
87. U. Meyer, J. Feldon, O. Dammann, Schizophrenia and autism: Both shared and disorder-specific pathogenesis via perinatal inflammation? *Pediatr. Res.* **69**, 26R–33R (2011). doi: 10.1203/PDR.0b013e318212c196; pmid: 21289540
88. J. D. Rosenblat *et al*., Inflammation as a neurobiological substrate of cognitive impairment in bipolar disorder: Evidence, pathophysiology and treatment implications. *J. Affect. Disord.* **188**, 149–159 (2015). doi: 10.1016/j.jad.2015.08.058; pmid: 26363613
89. T. Lappalainen, Functional genomics bridges the gap between quantitative genetics and molecular biology. *Genome Res.* **25**, 1427–1431 (2015). doi: 10.1101/gr.190983.115; pmid: 26430152

90. PsychENCODE Capstone Data Collection (2018); www.doi.org/10.7303/syn12080241.
91. M. C. Oldham, P. Langfelder, S. Horvath, Network methods for describing sample relationships in genomic datasets: Application to Huntington's disease. *BMC Syst. Biol.* **6**, 63 (2012). doi: 10.1186/1752-0509-6-63; pmid: 22691535
92. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010). doi: 10.1101/gr.097857.109; pmid: 19858363
93. A. Siepel *et al*., Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005). doi: 10.1101/gr.3715005; pmid: 16024819
94. N. N. Parikshak *et al*., Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013). doi: 10.1016/j.cell.2013.10.031; pmid: 24267887
95. B. J. Vilhjálmsson *et al*., Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015). doi: 10.1016/j.ajhg.2015.09.001; pmid: 26430803
96. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011). doi: 10.1016/j.ajhg.2010.11.011; pmid: 21167468
97. D. M. Ruderfer *et al*., Genomic Dissection of Bipolar Disorder and Schizophrenia, Including 28 Subphenotypes. *Cell* **173**, 1705–1715.e16 (2018). doi: 10.1016/j.cell.2018.05.046; pmid: 29906448
98. B. B. Lake *et al*., Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018). doi: 10.1038/nbt.4038; pmid: 29227469
99. T. Goldmann *et al*., Origin, fate and dynamics of macrophages at central nervous system interfaces. *Nat. Immunol.* **17**, 797–805 (2016). doi: 10.1038/ni.3423; pmid: 27135602
100. H. Keren-Shaul *et al*., A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell* **169**, 1276–1290.e17 (2017). doi: 10.1016/j.cell.2017.05.018; pmid: 28602351
101. A. Zeisel *et al*., Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015). doi: 10.1126/science.aaa1934; pmid: 25700174
102. Y. Zhang *et al*., Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* **89**, 37–53 (2016). doi: 10.1016/j.neuron.2015.11.013; pmid: 26687838
103. B. Wilkinson *et al*., The autism-associated gene chromodomain helicase DNA-binding protein 8 (CHD8) regulates noncoding RNAs and autism-related genes. *Transl. Psychiatry* **5**, e568 (2015). doi: 10.1038/tp.2015.62; pmid: 25989142
104. K. E. Samocha *et al*., A framework for the interpretation of de novo mutation in human disease. *Nat. Genet.* **46**, 944–950 (2014). doi: 10.1038/ng.3050; pmid: 25086666
105. I. Iossifov *et al*., Low load for disruptive mutations in autism genes and their biased transmission. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5600–E5607 (2015). doi: 10.1073/pnas.1516376112; pmid: 26401017
106. K. J. Karczewski *et al*., The ExAC browser: Displaying reference data information from over 60 000 exomes. *Nucleic Acids Res.* **45** (D1), D840–D845 (2017). doi: 10.1093/nar/gkw971; pmid: 27899611

ACKNOWLEDGMENTS

We acknowledge the National Institute of Mental Health (NIMH) for funding. Also, we acknowledge program staff, in particular T. Lehner, L. Bingaman, D. Panchision, A. Arguello, and G. Senthil, for providing institutional support and guidance for this project. We also thank A. Silva, M. López-Aranda, B. Pasanuic, N. Mancuso, and C. Giambartolomei for helpful discussions. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories, the Banner Sun Health Research Institute, the Harvard Brain Bank as part of the Autism Tissue Project (ATP), the Stanley Medical Research Institute, and the NIMH Human Brain Collection Core. Tissue for the RNAscope experiments was obtained from the Mount Sinai Neuropathology Research Core and Brain Bank. **Funding:** The work is funded by the U.S. National Institute of Mental Health (NIMH) (grants P50-MH106438, D.H.G.; R01-MH094714, D.H.G.; U01-MH103339, D.H.G.; R01-MH110927, D.H.G.; R01-MH100027, D.H.G.; R01-MH110920, C.L.; U01-MH103340, C.L.; MH109885, LMI; MH104766, LMI; MH105524, LMI; MH108528, LMI; R21-MH105881, D.P.; R01-MH110555, D.P.; R01-MH109715, D.P.), the Simons Foundation for Autism Research Initiative (SFARI grant 206733, D.H.G.; SFARI grant 345469, LMI; SFARI Bridge to Independence Award, M.J.G.; Simons Foundation Grant FA 345922, P.R.H.), the National Natural Science Foundation of China (grants 81401114, C.C.; 31571312, C.C.), the National Key Plan for Scientific Research and Development of China (2016YFC1306000, C.C.), the Innovation-Driven Project of Central South University (2015CXS034, C.C.; 2018CX033, C.C.), and the Icahn School of Medicine at Mount Sinai computational resources by the Office of Research Infrastructure of the NIH (award number S10OD018522, D.P.). Data were generated as part of the PsychENCODE Consortium, supported by grants U01MH103339, U01MH103365, U01MH103392, U01MH103340, U01MH103346, R01MH105472, R01MH094714, R01MH105898, R21MH102791, R21MH105881, R21MH103877, and P50MH106934 awarded to S. Akbarian (Icahn School of Medicine at Mount Sinai), G. Crawford (Duke), S. Dracheva (Icahn School of Medicine at Mount Sinai), P. Farnham (USC), M. Gerstein (Yale), D. Geschwind (UCLA), T. M. Hyde (LIBD), A. Jaffe (LIBD), J. A. Knowles (USC), C. Liu (SUNY Upstate), D. Pinto (Icahn School of Medicine at Mount Sinai), N. Sestan (Yale), P. Sklar (Icahn School of Medicine at Mount Sinai), M. State (UCSF), P. Sullivan (UNC), F. Vaccarino (Yale), S. Weissman (Yale), K. White (UChicago), and P. Zandi (JHU). RNA-seq data from the CommonMind Consortium used in this study (Synapse accession no. syn2759792) was supported by funding from Takeda Pharmaceuticals Company, F. Hoffmann-La Roche and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, R01MH075916, P50M096891, P50MH084053S1, R37MH057881, R37MH057881S1, HHSN271201300031C, AG02219, AG05138, and MH06692. **Author contributions:** Data were generated by the PsychENCODE Consortium, M.J.G., E.H., S.L., H.W., H.v.B., J.M., Y.X., R.D., D.W., K.F., D.F., K.W., A.E.J., M.A.P., M.G., C.L., D.P., and D.H.G. Data analysis was performed by M.J.G., P.Z., E.H., R.L.W., C.C., S.L., H.W., H.v.B., M.v.V., J.H., S.P., J.B., M.K., P.M.L., Z.K., J.M., D.W., Y.T.Y., M.X., K.F., P.R.H., J.W., D.F., M.G., L.M.I., D.P., and D.H.G. The manuscript was written by M.J.G., P.Z., R.L.W., C.C., J.H., S.P., M.K., L.M.I., D.P., and D.H.G. Supervision was performed by the corresponding authors. **Competing interests:** K.W. is associated with Tempus Labs. The other authors declare no direct competing interests. **Data and materials availability:** All data are available at (*90*). Summary data are available on our companion website at Resource.PsychENCODE.org.

SUPPLEMENTARY MATERIALS

10 April 2018; accepted 13 November 2018
10.1126/science.aat8127

RESEARCH ARTICLE SUMMARY

PSYCHIATRIC GENOMICS

Comprehensive functional genomic resource and integrative model for the human brain

Daifeng Wang*, Shuang Liu*, Jonathan Warrell*, Hyejung Won*, Xu Shi*, Fabio C. P. Navarro*, Declan Clarke*, Mengting Gu*, Prashant Emani*, Yucheng T. Yang, Min Xu, Michael J. Gandal, Shaoke Lou, Jing Zhang, Jonathan J. Park, Chengfei Yan, Suhn Kyong Rhie, Kasidet Manakongtreecheep, Holly Zhou, Aparna Nathan, Mette Peters, Eugenio Mattei, Dominic Fitzgerald, Tonya Brunetti, Jill Moore, Yan Jiang, Kiran Girdhar, Gabriel E. Hoffman, Selim Kalayci, Zeynep H. Gümüş, Gregory E. Crawford, PsychENCODE Consortium†, Panos Roussos, Schahram Akbarian, Andrew E. Jaffe, Kevin P. White, Zhiping Weng, Nenad Sestan, Daniel H. Geschwind‡, James A. Knowles‡, Mark B. Gerstein‡

INTRODUCTION: Strong genetic associations have been found for a number of psychiatric disorders. However, understanding the underlying molecular mechanisms remains challenging.

RATIONALE: To address this challenge, the PsychENCODE Consortium has developed a comprehensive online resource and integrative models for the functional genomics of the human brain.

RESULTS: The base of the pyramidal resource is the datasets generated by PsychENCODE, including bulk transcriptome, chromatin, genotype, and Hi-C datasets and single-cell transcriptomic data from ~32,000 cells for major brain regions. We have merged these with data from Genotype-Tissue Expression (GTEx), ENCODE, Roadmap Epigenomics, and single-cell analyses. Via uniform processing, we created a harmonized resource, allowing us to survey functional genomics data on the brain over a sample size of 1866 individuals.

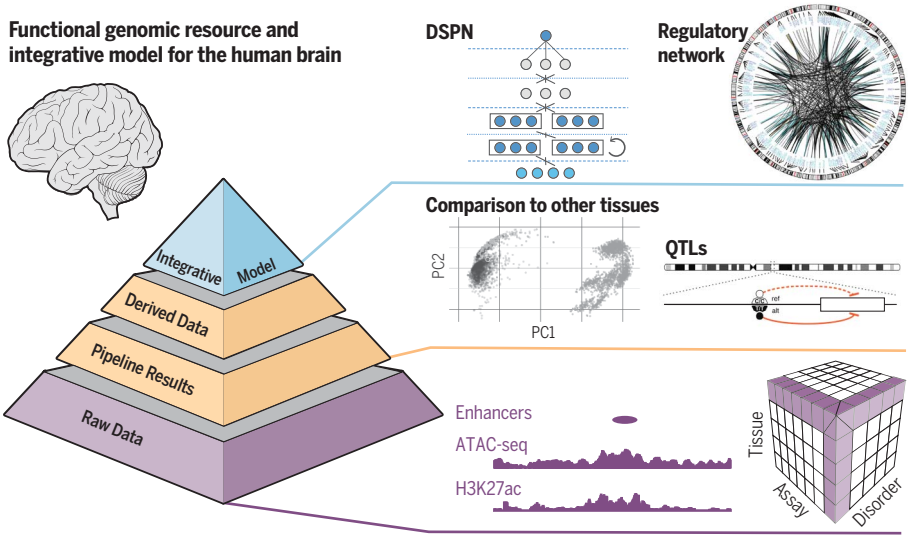
From this uniformly processed dataset, we created derived data products. These include lists of brain-expressed genes, coexpression modules, and single-cell expression profiles for many brain cell types; ~79,000 brain-active enhancers with associated Hi-C loops and topologically associating domains; and ~2.5 million expression quantitative-trait loci (QTLs) comprising ~238,000 linkage-disequilibrium–independent single-nucleotide polymorphisms and of other types of QTLs associated with splice isoforms, cell fractions, and chromatin activity. By using these, we found that >88% of the cross-population variation in brain gene expression can be accounted for by cell fraction changes. Furthermore, a number of disorders and aging are associated with changes in cell-type proportions. The derived data also enable comparison between the brain and other tissues. In particular, by using spectral analyses, we found that the brain has distinct expression and epigenetic patterns, including a greater extent of noncoding transcription than other tissues.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat8464

The top level of the resource consists of integrative networks for regulation and machine-learning models for disease prediction. The networks include a full gene regulatory network (GRN) for the brain, linking transcription factors, enhancers, and target genes from merging of the QTLs, generalized element-activity correlations, and Hi-C data. By using this network, we link disease genes to genome-wide association study (GWAS) variants for psychiatric disorders. For schizophrenia, we linked 321 genes to the 142 reported GWAS loci. We then embedded the regulatory network into a deep-learning model to predict psychiatric phenotypes from genotype and expression. Our model gives a ~6-fold improvement in prediction over additive polygenic risk scores. Moreover, it achieves a ~3-fold improvement over additive models, even when the gene expression data are imputed, highlighting the value of having just a small amount of transcriptome data for disease prediction. Lastly, it highlights key genes and pathways associated with disorder prediction, including immunological, synaptic, and metabolic pathways, recapitulating de novo results from more targeted analyses.

CONCLUSION: Our resource and integrative analyses have uncovered genomic elements and networks in the brain, which in turn have provided insight into the molecular mechanisms underlying psychiatric disorders. Our deep-learning model improves disease risk prediction over traditional approaches and can be extended with additional data types (e.g., microRNA and neuroimaging). ■



A comprehensive functional genomic resource for the adult human brain. The resource forms a three-layer pyramid. The bottom layer includes sequencing datasets for traits, such as schizophrenia. The middle layer represents derived datasets, including functional genomic elements and QTLs. The top layer contains integrated models, which link genotypes to phenotypes. DSPN, Deep Structured Phenotype Network; PC1 and PC2, principal components 1 and 2; ref, reference; alt, alternate; H3K27ac, histone H3 acetylation at lysine 27.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†PsychENCODE Consortium authors with their affiliations appear at the end of the full article online.
‡Corresponding author. Email: dhg@mednet.ucla.edu (D.H.G.); james.knowles@downstate.edu (J.A.K.); mark@gersteinlab.org (M.B.G.)
Cite this article as D. Wang *et al.*, *Science* **362**, eaat8464 (2018). DOI: 10.1126/science.aat8464

RESEARCH ARTICLE

PSYCHIATRIC GENOMICS

Comprehensive functional genomic resource and integrative model for the human brain

Daifeng Wang[1,2,3]*, Shuang Liu[1,2]*, Jonathan Warrell[1,2]*, Hyejung Won[4,5]*, Xu Shi[1,2]*, Fabio C. P. Navarro[1,2]*, Declan Clarke[1,2]*, Mengting Gu[1]*, Prashant Emani[1,2]*, Yucheng T. Yang[1,2], Min Xu[1,2], Michael J. Gandal[6], Shaoke Lou[1,2], Jing Zhang[1,2], Jonathan J. Park[1,2], Chengfei Yan[1,2], Suhn Kyong Rhie[7], Kasidet Manakongtreecheep[1,2], Holly Zhou[1,2], Aparna Nathan[1,2], Mette Peters[8], Eugenio Mattei[9], Dominic Fitzgerald[10], Tonya Brunetti[10], Jill Moore[9], Yan Jiang[11], Kiran Girdhar[12], Gabriel E. Hoffman[12], Selim Kalayci[12], Zeynep H. Gümüş[12], Gregory E. Crawford[13], PsychENCODE Consortium†, Panos Roussos[11,12], Schahram Akbarian[11,14], Andrew E. Jaffe[15], Kevin P. White[10,16], Zhiping Weng[9], Nenad Sestan[17], Daniel H. Geschwind[18,19,20]‡, James A. Knowles[21]‡, Mark B. Gerstein[1,2,22,23]‡

Despite progress in defining genetic risk for psychiatric disorders, their molecular mechanisms remain elusive. Addressing this, the PsychENCODE Consortium has generated a comprehensive online resource for the adult brain across 1866 individuals. The PsychENCODE resource contains ~79,000 brain-active enhancers, sets of Hi-C linkages, and topologically associating domains; single-cell expression profiles for many cell types; expression quantitative-trait loci (QTLs); and further QTLs associated with chromatin, splicing, and cell-type proportions. Integration shows that varying cell-type proportions largely account for the cross-population variation in expression (with >88% reconstruction accuracy). It also allows building of a gene regulatory network, linking genome-wide association study variants to genes (e.g., 321 for schizophrenia). We embed this network into an interpretable deep-learning model, which improves disease prediction by ~6-fold versus polygenic risk scores and identifies key genes and pathways in psychiatric disorders.

Disorders of the brain affect nearly one-fifth of the world's population (*1*). Decades of research have led to little progress in our understanding of the molecular causes of psychiatric disorders. This contrasts with cardiac disease, for which lifestyle and pharmacological modification of environmental risk factors has had profound effects on morbidity, or cancer, which is now understood to be a direct disorder of the genome (*2*–*5*). Although genome-wide association studies (GWAS) have identified many genomic variants strongly associated with neuropsychiatric disease risk—for instance, the Psychiatric Genomics Consortium (PGC) has identified 142 GWAS loci associated with schizophrenia (SCZ) (*6*)—for most of these variants, we have little understanding of the molecular mechanisms affecting the brain (*7*).

Many of these variants lie in noncoding regions, and large-scale studies have begun to elucidate the changes in genetic and epigenetic activity associated with these genomic alterations, suggesting potential molecular mechanisms. In particular, the Genotype-Tissue Expression (GTEx) project has associated many noncoding variants with expression quantitative-trait loci (eQTLs), and the ENCODE and Roadmap Epigenomics (Roadmap) projects have identified noncoding regions acting as enhancers and promoters (*8*–*10*). However, none of these projects have focused their efforts on the human brain. Initial work focusing on brain-specific functional genomics has provided greater insight but could be enhanced with larger sample sizes (*11*, *12*). Moreover, new methodologies, such as Hi-C and single-cell sequencing, have yet to be fully integrated at scale with brain genomics data (*13*–*16*).

Hence, the PsychENCODE Consortium has generated large-scale data to provide insight into the brain and psychiatric disorders, including data derived through genotyping, bulk and single-cell RNA sequencing (RNA-seq), chromatin immunoprecipitation with sequencing (ChIP-seq), assay for transposase-accessible chromatin using sequencing (ATAC-seq), and Hi-C (*17*). All data have been placed into a central, publicly available resource that also integrates relevant reprocessed data from related projects, including ENCODE, the CommonMind Consortium (CMC), GTEx, and Roadmap. By using this resource, we identified functional elements, quantitative-trait loci (QTLs), and regulatory-network linkages specific to the adult brain. Moreover, we combined these elements and networks to build an integrated deep-learning model that predicts high-level traits from genotype via intermediate molecular phenotypes. By "intermediate phenotypes," we mean the readouts of functional genomic information on genomic elements (e.g., gene expression and chromatin activity). In some contexts, these are also referred to as "molecular endophenotypes" (*18*). However, we include additional low-level "phenotypes," such as cell fractions, so we use the more general term "intermediate phenotype." We also refer to the high-level traits as "observed phenotypes," which include both classical clinical variables and characteristics of healthy individuals, such as gender and age.

Resource construction

The PsychENCODE resource (*19*) is the central website for this paper. It organizes data hierarchically, with a base of raw data files, a middle layer of uniformly processed and easily shareable results (such as open chromatin regions and gene expression quantifications), and a top-level "cap" of an integrative, deep-learning model, based on regulatory networks and QTLs. To build the base layer, we included all adult brain data from PsychENCODE and merged these with relevant data from ENCODE, CMC, GTEx, Roadmap, and

[1]Program in Computational Biology and Bioinformatics, Yale University, New Haven, CT 06520, USA. [2]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA. [3]Department of Biomedical Informatics, Stony Brook University, Stony Brook, NY 11794, USA. [4]Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA. [5]UNC Neuroscience Center, University of North Carolina, Chapel Hill, NC 27599, USA. [6]Department of Psychiatry, Semel Institute, David Geffen School of Medicine, University of California–Los Angeles, 695 Charles E. Young Drive South, Los Angeles, CA 90095, USA. [7]Keck School of Medicine and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90007, USA. [8]Sage Bionetworks, Seattle, WA 98109, USA. [9]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA. [10]Institute for Genomics and Systems Biology, Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA. [11]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [12]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [13]Center for Genomic and Computational Biology, Department of Pediatrics, Duke University, Durham, NC 27708, USA. [14]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [15]Lieber Institute for Brain Development, Johns Hopkins Medical Campus, and Departments of Mental Health and Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA. [16]Tempus Labs, Chicago, IL 60654, USA. [17]Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT 06520, USA. [18]Department of Human Genetics, David Geffen School of Medicine, University of California–Los Angeles, Los Angeles, CA 90095, USA. [19]Program in Neurobehavioral Genetics, Semel Institute, David Geffen School of Medicine, University of California–Los Angeles, Los Angeles, CA 90095, USA. [20]Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California–Los Angeles, Los Angeles, CA 90095, USA. [21]SUNY Downstate Medical Center College of Medicine, Brooklyn, NY 11203, USA. [22]Department of Computer Science, Yale University, New Haven, CT 06520, USA. [23]Department of Statistics and Data Science, Yale University, New Haven, CT 06520, USA.
*These authors contributed equally to this work. †PsychENCODE Consortium authors with their affiliations appear at the end of this paper.
‡Corresponding author. Email: dhg@mednet.ucla.edu (D.H.G.); james.knowles@downstate.edu (J.A.K.); mark@gersteinlab.org (M.B.G.)

recent single-cell studies (table S1 and Fig. 1). In total, the resource contains 3810 genotype, transcriptome, chromatin, and Hi-C datasets from PsychENCODE and 1662 datasets obtained by using similar bulk assays merged from outside the consortium. Overall, the datasets from the prefrontal cortex (PFC) involve sampling from 1866 individuals. The resource also has single-cell RNA-seq data for 18,025 cells from PsychENCODE and 14,012 cells from outside sources (*20*). These data represent a range of psychiatric disorders, including SCZ, bipolar disorder (BPD), and autism spectrum disorder (ASD). The individual genotyping and raw next-generation sequencing of transcriptomics and epigenomics are restricted for privacy protection, but access can be obtained upon approval. The protocols for all associated data are readily available (fig. S1). Finally, PsychENCODE has developed a reference brain project on the PFC by using matched assays on the same set of brain tissues, which we used to develop an anchoring annotation (*21*).

Transcriptome analysis: Bulk and single cell

To identify the genomic elements exhibiting transcriptional activities specific to the brain, we took a conservative approach and used the standardized and established ENCODE pipeline to uniformly process RNA-seq data from PsychENCODE, GTEx, and Roadmap (figs. S2 and S3). This consistency makes our expression data and subsequent results (including eQTLs and single-cell analyses) comparable with previous work. Using these data, we identified noncoding regions of transcription and sets of differentially expressed and coexpressed genes (*21*, *22*).

Brain tissue is composed of a variety of basic cell types. Gene expression changes observed at the tissue level may be due to changes in the proportions of basic cell types (*23*–*28*). However, it is unclear how these changes in cell proportions can contribute to the variation in tissue-level gene expression observed across a population of individuals. To address this question, we used two complementary strategies across our cohort of 1866 individuals.

First, we used standard pipelines to uniformly process single-cell RNA-seq data from PsychENCODE, in conjunction with other single-cell studies on the brain (*14*, *16*, *20*). Then we assembled profiles of brain cell types, including both excitatory and inhibitory neurons (denoted as Ex1 to Ex9 and In1 to In8, respectively, according to previous conventions), major nonneuronal types (e.g., microglia and astrocytes), and additional cell types associated with development (*21*). Depending on the underlying sequencing and quantification, our profiles were of two fundamentally different formats, transcripts per

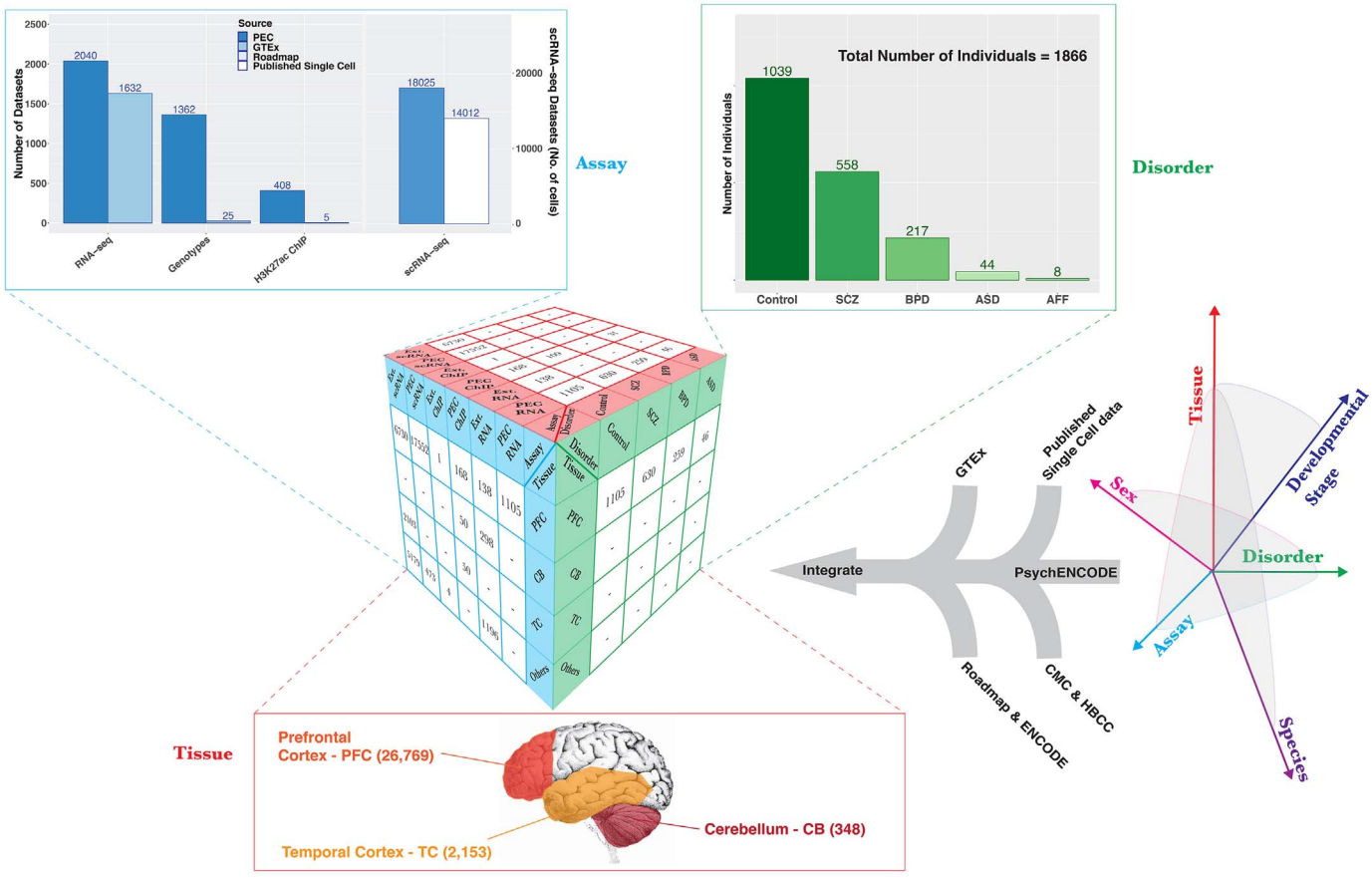


Fig. 1. Comprehensive data resource for functional genomics of the human brain. The functional genomics data generated by the PsychENCODE Consortium (PEC) constitute a multidimensional exploration across tissue, developmental stage, disorder, species, assay, and sex. The central data cube represents the results of our data integration for the three dimensions of disorder, assay, and tissue, where the numbers of datasets in the analysis are depicted. Projections of the data onto each of these three parameters are shown as graphs for assay and disorder and as a schematic for the primary brain regions of interest. Assay: Dataset numbers for a subset of assays are shown, including RNA-seq (2040 PsychENCODE samples and 1632 GTEx samples, used in multiple downstream analyses), genotypes (1362 PsychENCODE and 25 GTEx individuals for a total of 1387 individuals matched to RNA-seq samples for QTL analysis after quality control filtering), and H3K27ac ChIP-seq (408 PsychENCODE and 5 Roadmap samples). The number of cells assayed by small conditional RNA sequencing (scRNA-seq) (right-hand *y* axis) is 18,025 for PsychENCODE and 14,012 for external (ext.) datasets. Disorder: Across all assays, there are 113 GTEx and 926 PsychENCODE control individuals and 558 SCZ, 217 BPD, 44 ASD, and 8 affective disorder (AFF) individuals from PsychENCODE, resulting in 1866 individuals. Tissue: Three brain regions are considered—the PFC (n = 26,769 samples), TC (n = 2153 samples), and CB (n = 348 samples). See table S11 and (*19*) for more details. HBCC, Human Brain Collection Core.

kilobase million (TPM) and unique molecular identifier (UMI) counts. The former (TPM profiles) includes the uniformly processed PsychENCODE developmental single-cell data merged with published adult and developmental data (fig. S4 and table S2) (*14*, *16*). By contrast, the UMI profiles are built by merging PsychENCODE adult single-cell profiles with other recently published datasets (*14*). Both formats share common neuronal and major nonneuronal cell types and are used interchangeably in various analyses in this study (fig. S5 and tables S3 and S4). Moreover, the expression values of biomarker genes for the same cell type were correlated between two formats (figs. S6 and S7). However, our TPM profiles have additional development-specific cell types, such as quiescent and replicating.

From both sets of profiles, we can generate a matrix **C** of expression signatures, comprising marker genes and their expression levels across various cells (fig. S8). In this matrix, a number of genes (e.g., the gene for dopamine receptor DRD3) had expression levels that varied more across cell types than they did in bulk tissue measurements across individuals in a population (Fig. 2A). This suggests that cell-type changes across individuals could contribute substantially to variation in individual bulk expression levels.

Second, we used an unsupervised analysis to identify the primary components of bulk expression variation. We decomposed the bulk gene expression matrix by using nonnegative matrix factorization (NMF) (**B** ≈ **VH**, where **B**, **V**, and **H** represent matrices) and determined whether the top components (NMF-TCs), capturing the majority of covariance (columns of **V**) (Fig. 2B), were consistently associated with the single-cell signatures (Fig. 2C) (*21*). A number of NMF-TCs were, in fact, highly correlated with cell types from matrix **C** for both TPM and UMI data—e.g., component NMF-17 is correlated with the Ex2 cell type (correlation coefficient $r = 0.63$) (Fig. 2C and fig. S9). This demonstrates that an unsupervised analysis derived solely from bulk data can roughly recapitulate the single-cell signatures, partially corroborating them.

We then examined how variation in the proportions of basic cell types contributes to variation in bulk expression. To this end, we estimated the relative proportions of various cell types ("cell fractions") for each tissue sample. In particular, we deconvolved the bulk tissue-level expression matrix by using the single-cell signatures to estimate cell fractions across individuals (matrix **W**), solving **B** ≈ **CW** (Fig. 2B) (*21*). As a validation, our estimated fractions of $NEU^{+/-}$ cells matched the experimentally determined fractions from reference brain samples (median difference = 0.04) (fig. S10). Overall, our analyses demonstrated that variation in cell types contributed substantially to bulk variation. That is, weighted combinations of single-cell signatures could account for most of the population-level expression variation, with an accuracy of >88% (Fig. 2D) ($1 - ||\mathbf{B} - \mathbf{CW}||^2/||\mathbf{B}||^2 > 88\%$), and when calculated on a per-person basis, this quantity varies ±4% over the 1866 individuals in our cohort (figs. S11 and S12). Also, our results explained more variation than previous deconvolution approaches (fig. S13) (*21*).

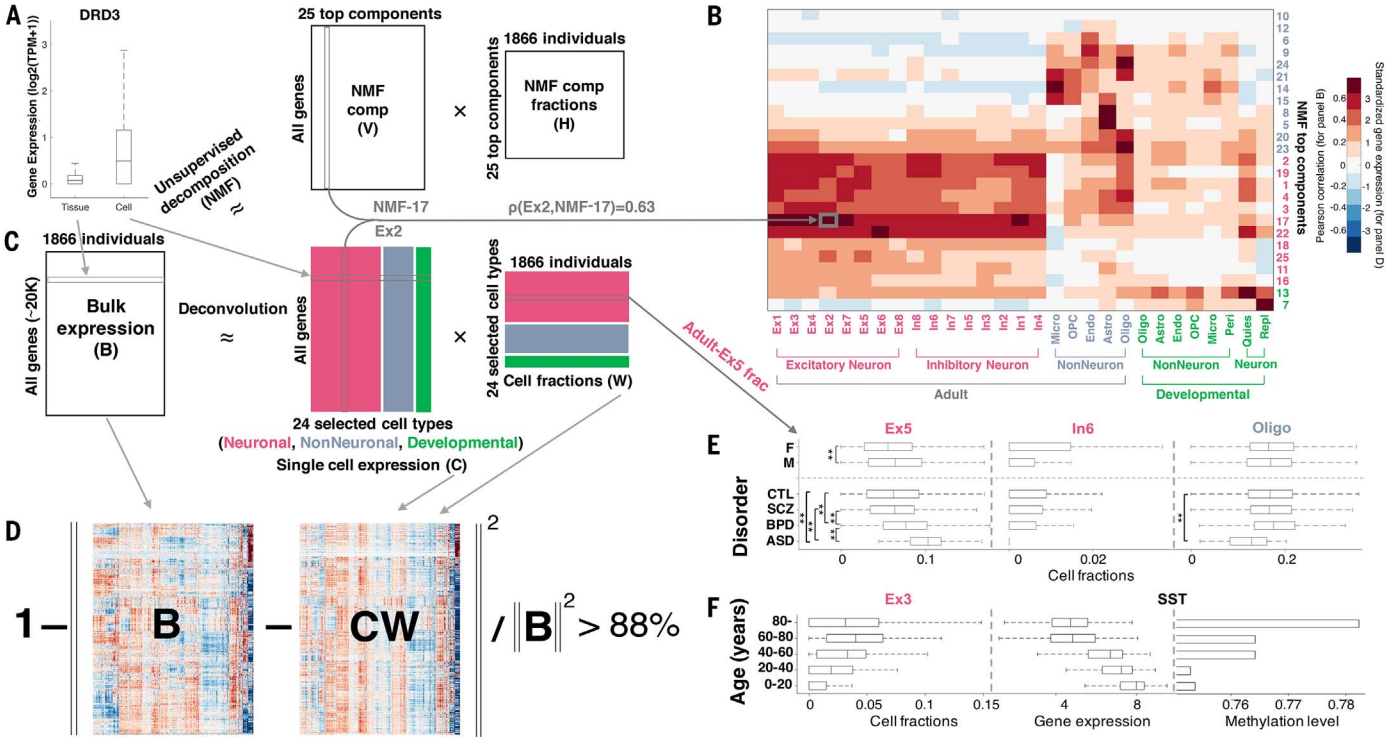


Fig. 2. Deconvolution analysis of bulk and single-cell transcriptomics reveals cell fraction changes across the population. (**A**) Genes had significantly higher expression variability across single cells sampled from different types of brain cells than across equivalent tissue samples taken from a population of individuals. (Left) Dopamine gene *DRD3*. (**B**) The heatmap shows the Pearson correlation coefficients of gene expression between the NMF-TCs and single-cell signatures (for $n = 457$ biomarker genes) (*15*). Micro, microglia; OPC, oligodendrocyte progenitor cells; endo, endothelial cells; astro, astrocytes; oligo, oligodendrocytes; peri, pericytes; quies, quiescent cells; repl, replicating cells. (**C**) (Top) The bulk tissue gene expression matrix (**B**, genes by individuals) can be decomposed by NMF (see fig. S52). (Bottom) The bulk tissue gene expression matrix **B** can be also deconvolved by the single-cell gene expression matrix (**C**, genes by cell types) to estimate the cell fractions across individuals (the matrix **W**); i.e., **B** ≈ **CW**. The three major cell types analyzed are depicted with neuronal cells in red, nonneuronal cells in blue, and developmental cells in green, as highlighted by column groups in matrix **C** (also row groups in **W**). frac, fraction. (**D**) The estimated cell fractions can account for >88% of the bulk tissue expression variation across the population. (**E**) Cell fraction changes across genders and brain disorders. **Differences from control samples are significant (via a Kolmogorov-Smirnov test) after accounting for age distributions. See table S12 for more detail. CTL, control. (**F**) Changing cell fractions (for Ex3), gene expression (for SST), and promoter methylation level (median level, for SST) across age groups are shown. With increasing age, the fractions of Ex3 and Ex4 significantly increase, and some nonneuronal types decrease (Ex3 trend analysis, $P < 6.3 \times 10^{-10}$).

We identified cell fraction changes associated with different traits (Fig. 2E and figs. S14 to S17). For example, particular types of excitatory and inhibitory neurons (such as In6) are present in different fractions in male and female samples (Fig. 2E). Also, in individuals with ASD, the fraction of Ex5 was higher and that of oligodendrocytes, lower, with some commensurate increase for microglia and astrocytes (Fig. 2E and fig. S18) (*24*, *29*).

Lastly, we observed an association with age. In particular, with increasing age, the fractions of Ex3 and Ex4 significantly increased and the fractions of some nonneuronal types decreased (Fig. 2F and fig. S19). These changes may be associated with differential expression of specific genes, e.g., the gene for somatostatin (SST), known to be associated with aging and neurotransmission (Fig. 2F) (*30*). Also, SST exhibits increasing promoter methylation with age, perhaps explaining its decreasing expression. Other genes known to be associated with brain aging, such as those for EGR1 (early growth response) and CP (ceruloplasmin), exhibit different trends (Fig. 2F and figs. S20 and S21) (*21*, *31*).

Enhancers

To annotate brain-active enhancers, we used chromatin modification data from the reference brain, supplemented by deoxyribonuclease sequencing (DNase-seq) and ChIP-seq data from Roadmap PFC samples. All data were processed by standard ENCODE ChIP-seq pipelines to ensure maximal compatibility of our results (fig. S22). Consistent with ENCODE, we define active enhancers as open chromatin regions enriched in H3K27ac (histone H3 acetylation at lysine 27) and depleted in H3K4me3 (histone H3 trimethylation at lysine 4) (Fig. 3A and fig. S23) (*21*). Overall, we annotated a reference set of 79,056 enhancers in the PFC. [We also provide a filtered subset (*21*).]

Assessing the variability across individuals and tissues is more difficult for enhancers than for gene expression (*32*). Not only is the variability in chromatin-mark level at enhancers across different individuals and tissues high, but the boundaries of enhancers can grow and shrink, sometimes disappearing altogether (e.g., for H3K27ac) (Fig. 3A). To investigate this in more detail, we uniformly processed the H3K27ac data from the PFC, temporal cortex (TC), and cerebellum (CB) on a cohort of 50 individuals, primarily of European descent and sequenced to similar depths (*21*) (fig. S24). Aggregating data across the cohort resulted in a total of 37,761 H3K27ac "peaks" (enriched regions) in the PFC, 42,683 in the TC, and 26,631 in the CB—where each peak is

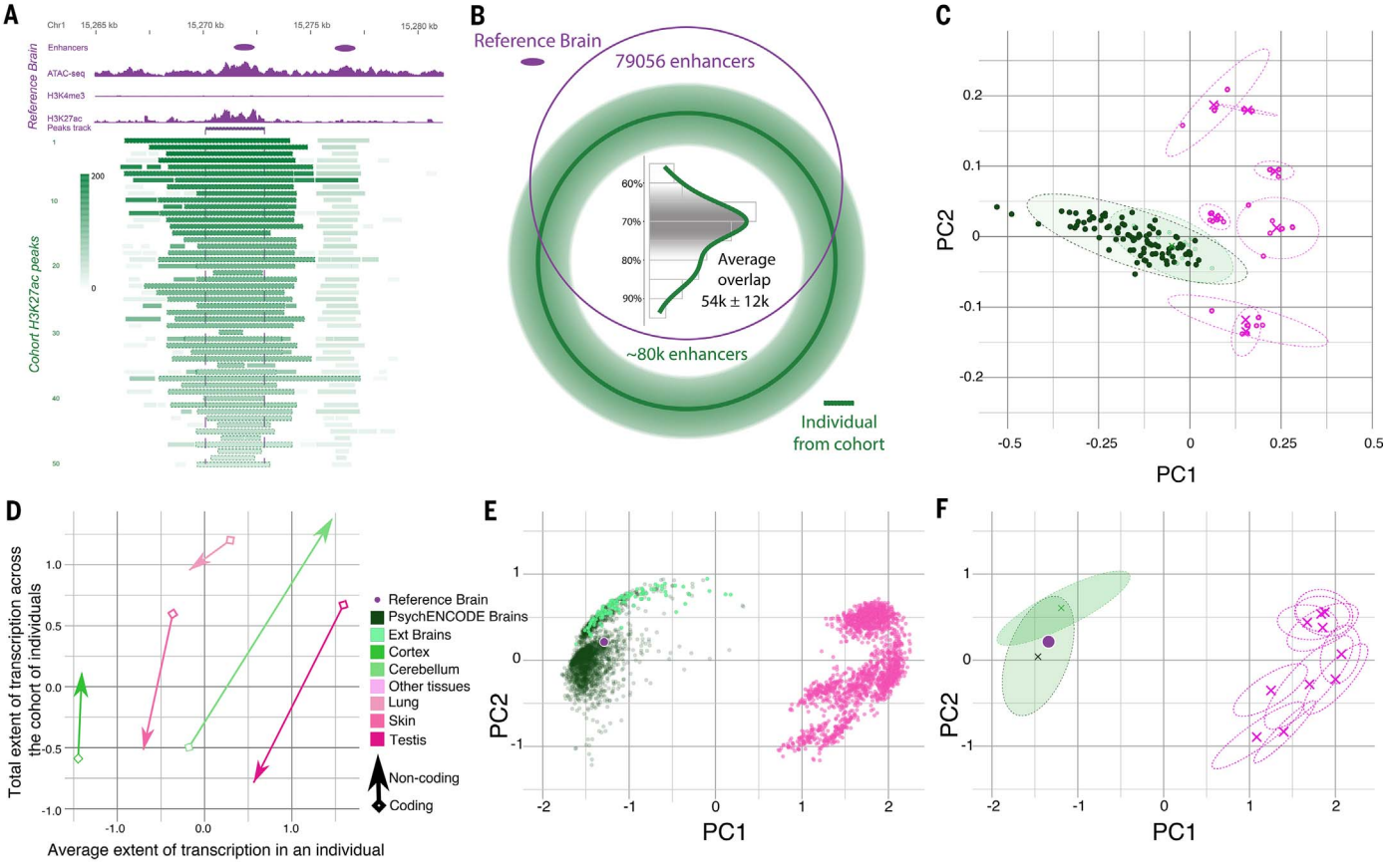


Fig. 3. Comparative analysis of transcriptomics and epigenomics between the brain and other tissues. (**A**) Epigenetics signals of the reference brain (purple) were used to identify active enhancers with the ENCODE enhancer pipeline. The H3K27ac signal tracks at the corresponding enhancer region from each individual in the cohort are shown in green, with the gradient showing the normalized signal value for each H3K27ac peak. (**B**) The overlap of the H3K27ac peaks from an individual in the population with the reference brain enhancers is shown as a Venn diagram. The histogram shows the varying percentages of overlapped H3K27ac peaks across individuals. (**C**) The tissue clusters of RCA coefficients [principal component 1 (PC1) versus PC2] for chromatin data of any potential regulatory elements are shown. Clusters of PsychENCODE samples (dark green ellipses), external brain samples (light green ellipses), and other non-brain tissues (magenta ellipses) are plotted. (**D**) The extent of transcription for coding (arrowhead) and noncoding (diamond) regions. The average transcription extent (*x* axis) is shown compared with the cumulative extent of transcription across a cohort of individuals (*y* axis) for select tissue types, including the CB, cortex, lung, skin, and testis, by using polyadenylate RNA-seq data. (**E** and **F**) Similar to (C), but now for transcription rather than epigenetics. (E) RCA coefficients for gene expression data from PsychENCODE, GTEx brains, and other tissue samples are shown in dark green, light green, and magenta, respectively. (F) The center (cross) and ranges of different tissue clusters (dashed ellipses) are shown on an RCA scatterplot of (E).

present in more than half of the individuals surveyed. In a comparison of aggregated sets for these three brain regions, the PFC was more similar to the TC than the CB (~90% versus 34% overlap in peaks). This difference is consistent with previous reports and suggests potentially different cell-type composition in the CB and the cortex (*33*, *34*).

We also examined how many of the enhancers in the reference brain are active (i.e., have enriched H3K27ac) in each of the individuals in our cohort. As expected, not every reference enhancer was active in each individual. On average, only ~70% ± 15% (~54,000) of the enhancers in the reference brain were active in an individual in the cohort, and a similar fraction of the reference enhancers was active in more than half the cohort (68%) (Fig. 3B). To estimate the total number of enhancers in the PFC, we calculated the cumulative number of active regions across the cohort (fig. S25). This increased for the first 20 individuals sampled but saturated at the 30th. Thus, we hypothesize that pooling PFC enhancers from ~30 individuals is sufficient to cover nearly all

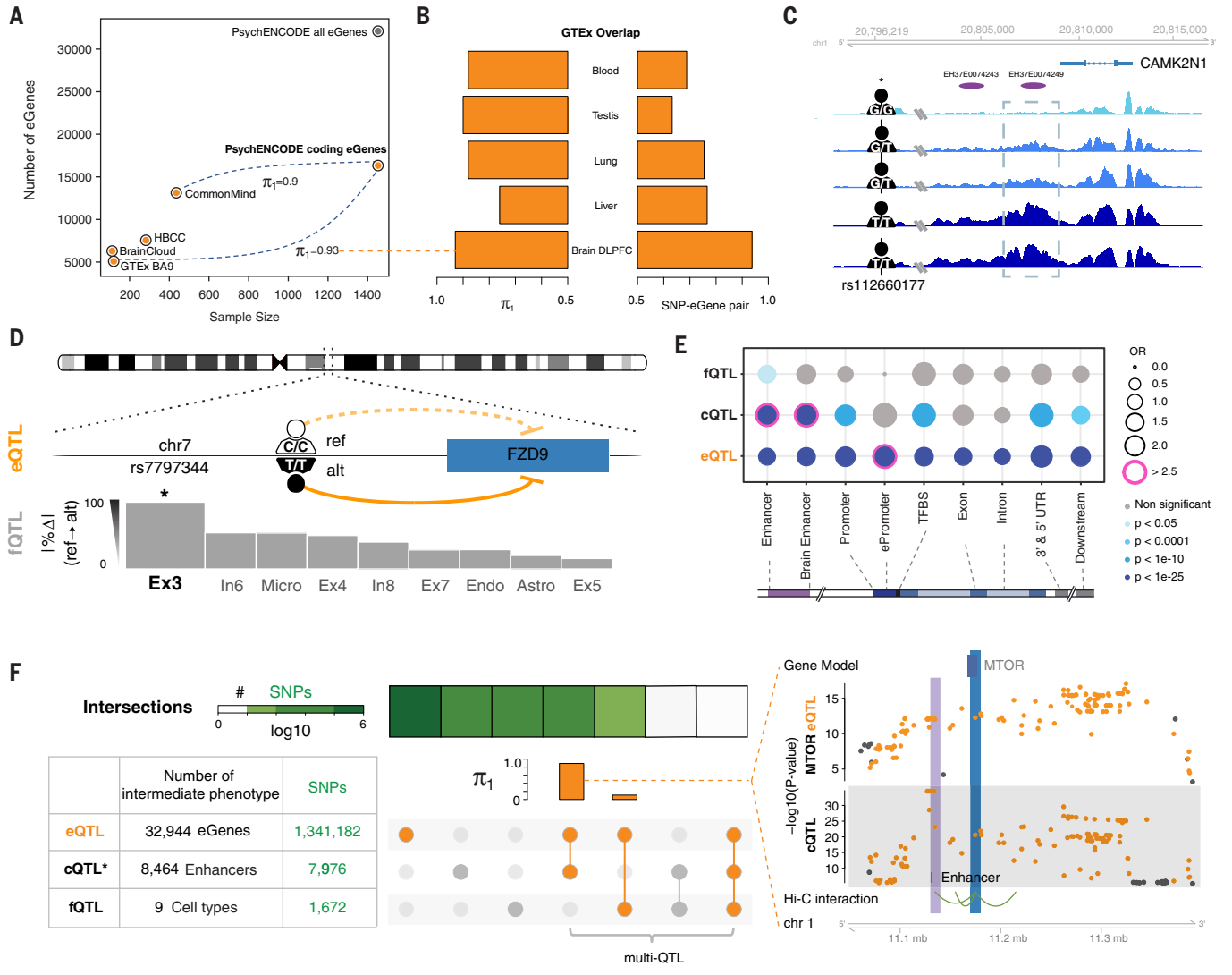


Fig. 4. QTLs in the adult brain. (**A**) The frequency of genes with at least one eQTL (eGenes) is shown across different studies. The number of eGenes increased as the sample size increased. PsychENCODE eGenes are close to saturation for protein-coding genes. The estimated replication π_1 values for GTEx and CMC eQTLs versus PsychENCODE are shown (*36*). (**B**) The similarity between PsychENCODE brain dorsolateral PFC (DLPFC) eQTLs and GTEx eQTLs of other tissues are evaluated by π_1 values and SNP-eGene overlap rates. Both π_1 values and SNP-eGene overlap rates are higher for brain DLPFC than for the other tissues. (**C**) An example of an H3K27ac signal across individuals in a representative genomic region, showing largely congruent identification of regions of open chromatin. The region within the dashed rectangle represents a cQTL; the signal magnitudes for individuals with a G/G or G/T genotype were lower than those for individuals with a T/T genotype. chr1, chromosome 1; rs, reference SNP. (**D**) An example of the mechanism by which an fQTL may affect phenotype. This fQTL overlaps with an eQTL for *FZD9*, a gene located in the 7q11.23 region that is deleted in Williams syndrome. The fQTL may affect the fraction of Ex3 by regulating FZD9 expression. Only Ex3 constitutes a statistically significant fQTL with this SNP (as designated by the asterisk). ref, reference; alt, alternate. (**E**) The enrichment of QTLs in different genomic annotations is shown. Pink circles indicate highly significant enrichment ($P < 1 \times 10^{-25}$ and OR > 2.5). OR, odds ratio; TFBS, TF binding site; UTR, untranslated region. (**F**) Numbers of identified QTL-associated elements (eGenes, enhancers, and cell types) and QTL SNPs are shown in the bottom left table. Asterisks indicate that, for cQTLs, we show only the number of top SNPs for each enhancer. Overlaps of all QTL SNPs are shown in heatmaps (square rows). The linked circles show the overlap of QTL types. The intersections of other QTLs with eQTLs are evaluated by using π_1 values in the orange bar plot. The greatest intersection is between cQTLs and eQTLs. An example is displayed on the right: the intersection of eQTL SNPs (for the MTOR gene) and cQTL SNPs (for the H3K27ac signal on an enhancer ~50 kb upstream of the gene). Hi-C interactions (bottom) indicate that the enhancer interacts with the promoter of MTOR, suggesting that the cQTL SNPs potentially mediate the expression modulation manifest by the eQTL SNPs.

possible PFC enhancer regions, estimated at ~120,000.

Consistent comparison: Transcriptome and epigenome

As we uniformly processed the transcriptomic and epigenomic data across the PsychENCODE, ENCODE, GTEx, and Roadmap datasets, we could compare the brain with other organs in a consistent fashion and also compare transcriptome variation with that of the epigenome (Fig. 3, C to F). Several approaches, including principal components anaylsis (PCA), *t*-distributed stochastic neighbor embedding (t-SNE), and reference component analysis (RCA), were tested to determine the best method for comparison. We found that, although popular and interpretable, PCA deemphasizes local structure and is overly influenced by outliers; by contrast, t-SNE preserves local relationships but "shatters" global structure. RCA is a compromise (*21*): It captures local structure while maintaining meaningful distances globally. We used RCA to project gene expression from PsychENCODE samples against a reference panel of gene expression for different tissues derived from GTEx and then reduced the dimensionality of the projections with PCA. RCA thus allowed us to represent high-dimensional expression data in a simple two-coordinate diagram.

For gene expression, RCA revealed that the brain separates from the other tissues in the first component (Fig. 3E and fig. S26). In particular, for the brain, intertissue comparisons exhibit more differences than intratissue ones (figs. S27 to S30). A different picture emerged for chromatin. The H3K27ac chromatin levels at all regulatory positions were, overall, less distinguishable between the brain and other tissues (Fig. 3C) (*21*). At first glance, this is surprising, as one expects great differences in enhancer usage between tissues. However, our analysis compares chromatin signals over all regulatory elements from ENCODE (including enhancers and promoters), which is logically consistent with our expression comparison across all protein-coding genes (Fig. 3, F versus C, and tables S5 to S7). As the total number of human regulatory elements is much larger than the number of brain-active enhancers (~1.3 million versus ~79,000), our results likely reflect the fact that there are proportionately fewer brain-active regulatory elements than protein-coding genes (6% versus 60%).

Up to this point, our analysis has focused on annotated regions (genes, promoters, and enhancers). However, in addition to the canonical expression differences in protein-coding genes, we also found differences in unannotated

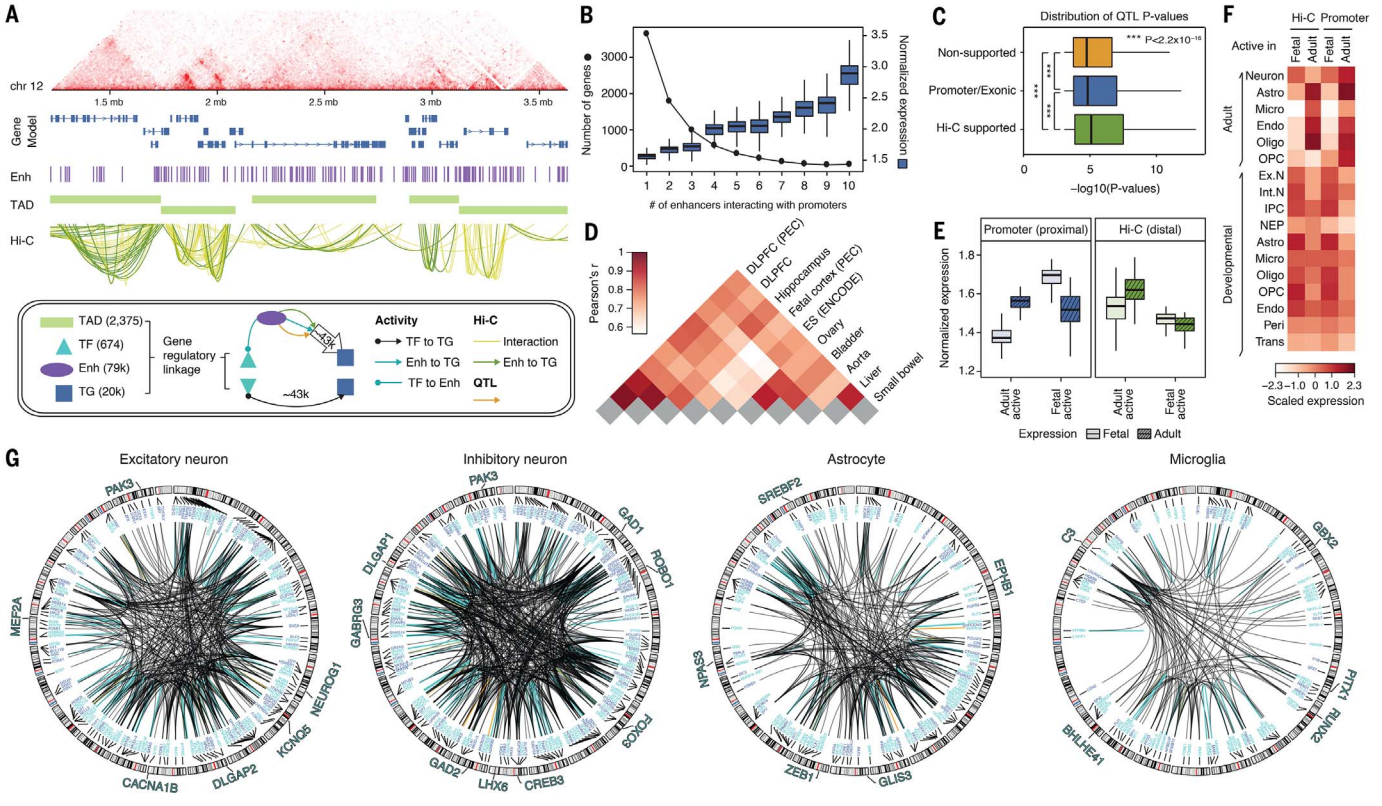


Fig. 5. Building a gene regulatory network (GRN) from Hi-C and data integration. (**A**) A full Hi-C dataset from adult brain reveals the higher-order structure of the genome, ranging from contact maps (top) to TADs and promoter-based interactions. (Bottom) A schematic of how we leveraged gene regulatory linkages involving TADs, TFs, enhancers (Enh), and target genes (TG) to build a full GRN (fig. S42) and a high-confidence subnetwork consisting of 43,181 TF–to–target gene promoter and 42,681 enhancer–to–target gene promoter linkages (*21*). (**B**) We compared the number of genes (left *y* axis, dotted line) and the normalized gene expression levels (right *y* axis, boxes) with the number of enhancers that interact with the gene promoters. Boxes show means and SDs. (**C**) QTLs that were supported by Hi-C evidence (174,719) showed more significant *P* values than those that were not (promoter or exonic QTLs, 130,155; nonsupported QTLs, 1,065,311). (**D**) Cross-tissue comparison of chromatin architecture indicates that adult brains in PsychENCODE and Roadmap (e.g., DLPFC and hippocampus tissues) share chromatin architecture more than nonrelated tissue types. Fetal brain shows chromatin architecture distinct from that in adult brain, indicating extensive rewiring of chromatin structures during brain development. ES, embryonic stem cell. (**E**) Genes assigned to fetal active elements are prenatally enriched, whereas genes assigned to adult active elements are postnatally enriched. (**F**) Genes assigned to fetal active elements are relatively more enriched in neurons in the adult brain and fetal (developmental) brain, whereas genes assigned to adult active elements are relatively more enriched in glia (adult astrocytes, endothelial cells, and oligodendrocytes). Ex. N, excitatory neuron; Int. N, inhibitory neuron; IPC, intermediate progenitor cells; NEP, neuroepithelial cells; trans, transient cell type. (**G**) The circos plots show the linkages from the full regulatory network targeting the cell-type–specific biomarker genes. The biomarker genes for excitatory or inhibitory neuronal type are the biomarker genes shared by at least five excitatory or inhibitory subtypes (*20*). Selected TFs for particular cell types are highlighted.

noncoding and intergenic regions (fig. S30). In particular, testes and lung have the largest extent of transcription overall (the most genes transcribed) for protein-coding genes (Fig. 3D). However, when we shift to unannotated regions, the ordering changes: Brain tissues, such as the cortex and CB, now have a greater extent of transcription than any other tissue.

QTL analysis

We used the data in the brain resource to identify QTLs affecting gene expression and chromatin activity. We calculated expression, splicing-isoform, chromatin, and cell fraction QTLs (eQTLs, isoQTLs, cQTLs, and fQTLs, respectively). For eQTLs, we adopted a standard approach, closely adhering to the GTEx pipeline for maximal compatibility (figs. S31 to S33) (*35*). (However, for maximal utility of the resource, we also provide alternate lists, filtered more conservatively.) In the PFC, we identified ~2.5 million cis-eQTLs involving ~33,000 eGenes (expressed genes) [~17,000 noncoding and ~16,000 coding, with a false discovery rate of <0.05] (Fig. 4A). We found 1,341,182 eQTL single-nucleotide polymorphisms (SNPs) from ~5.3 million total SNPs tested in 1-Mb windows around genes, constituting 238,194 independent SNPs after linkage-disequilibrium (LD) pruning. This estimate identified substantially more eQTLs and associated eGenes than previous studies, reflecting our large sample size (*8*, *11*, *21*). The number of eGenes, in fact, approaches the total number of genes estimated to be expressed in the brain. That said, a very large fraction of the smaller GTEx and CMC brain eQTL sets was contained within our set (as evident from overlap testing with the π_1 statistic) (Fig. 4A) (*36*). Moreover, as expected, our brain eQTL set showed higher π_1 similarity to and SNP-eGene overlap with GTEx brain eQTLs than with those from other tissues (Fig. 4B and fig. S31). Lastly, we applied the QTL pipeline to isoform levels to calculate a set of isoQTLs. We

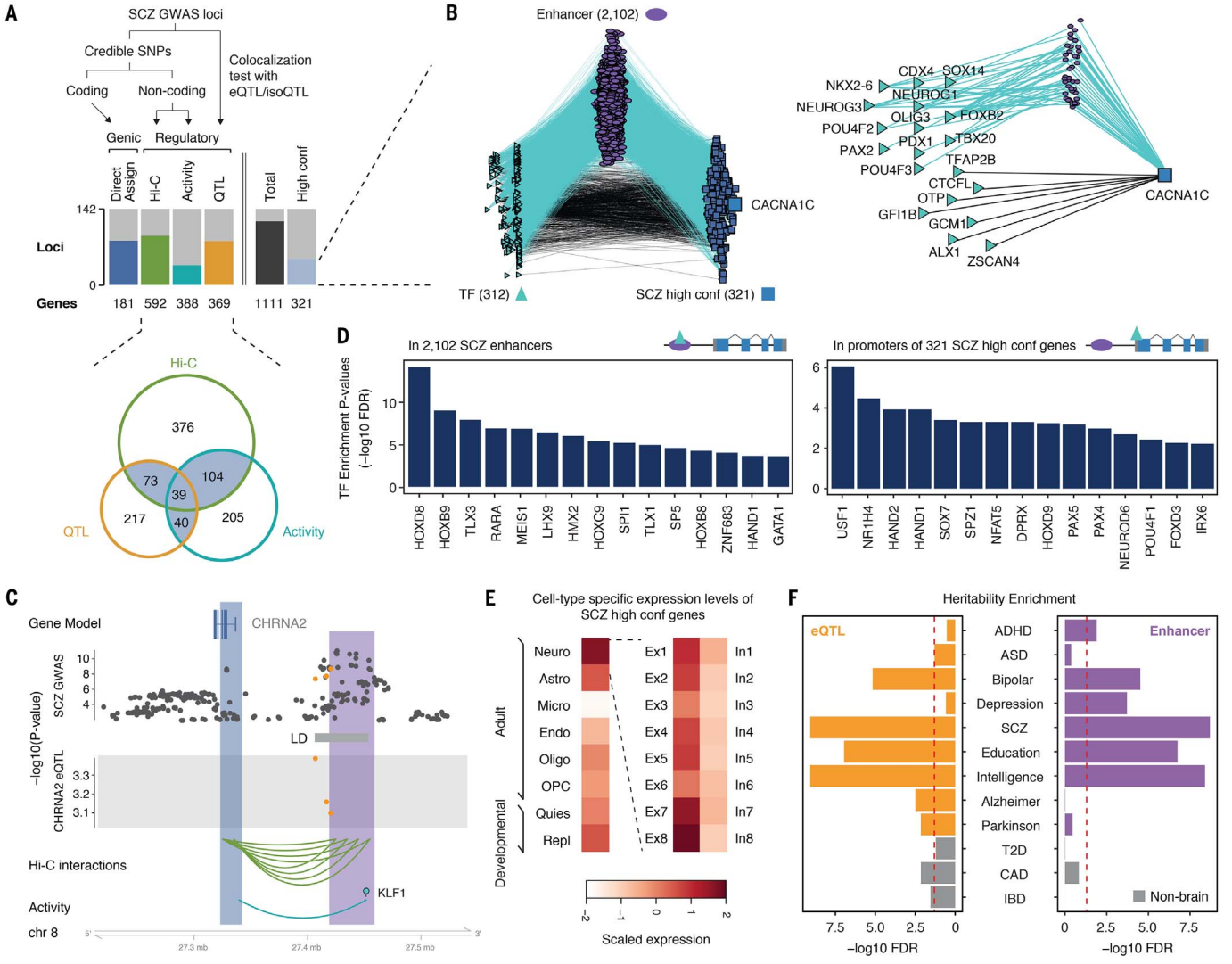


Fig. 6. GRNs assign genes to GWAS loci for psychiatric disorders. (**A**) A schematic depicting how SCZ GWAS loci were assigned to putative genes. The number of SCZ GWAS loci and their putative target genes (SCZ genes) annotated by each assignment strategy is indicated (top). The overlap between SCZ genes defined by QTL associations (QTL), chromatin interactions (Hi-C), and activity relationships (activity) is depicted in a Venn diagram (bottom). SCZ genes with more than two evidence sources were defined as high-confidence (high conf.) genes. (**B**) A GRN of TFs, enhancers, and 321 SCZ high-confidence genes, on the basis of TF activity linkages. A subnetwork for *CACNA1C* is highlighted on the right. (**C**) An example of the evidence indicating that GWAS SNPs that overlap with *CHRNA2* eQTLs also have chromatin interactions and activity correlations with the same gene. Orange dots refer to SNPs that overlap between eQTLs and GWAS plots. (**D**) TFs that are significantly enriched in enhancers (left) and promoters (right) of SCZ genes. FDR, false discovery rate. (**E**) SCZ genes show higher expression levels in neurons (particularly excitatory neurons) than in other cell types. (**F**) Brain disorder GWAS show stronger heritability enrichment in brain regulatory variants (eQTLs) and elements (enhancers) than non–brain disorder GWAS. ADHD, attention-deficit/hyperactivity disorder; T2D, type 2 diabetes; CAD, coronary artery disease; IBD, inflammatory bowel disease.

performed filtering in a variety of different ways, generating a number of different lists (*21*).

For cQTLs, no established methods exist for large-scale data, although there have been previous efforts (*37*, *38*). To identify cQTLs, we focused on our reference set of enhancers and examined how H3K27ac activity varied at these loci across 292 individuals (Fig. 4C) (*21*). Overall, we identified ~2000 cQTLs in addition to 6200 identified from individuals within the CMC cohort (*39*).

We next identified SNPs associated with changes in the relative abundances of specific cell types. We refer to such relationships with the term fQTLs. In total, we identified 1672 distinct SNPs constituting 4199 fQTLs (fig. S34). The excitatory neurons Ex4 and Ex5 were associated with the most fQTLs (1060 and 896, respectively). The biological mechanism governing an fQTL may involve other QTL types, such as eQTLs. An illustrative example is the *FZD9* gene (Fig. 4D): We found that the expression levels of this gene were associated with a neighboring noncoding SNP via an eQTL, and this same SNP was associated with the proportion of Ex3 cells via an fQTL. Perhaps connected to this, deletion variants upstream of *FZD9* had previously been associated with cell fraction changes related to Williams syndrome (*40*).

Next, we attempted to recalibrate the observed gene expression variation by considering fQTLs. In particular, our scheme described above for approximately deconvolving gene expression from heterogeneous bulk tissue (matrix **B**) into single-cell signatures (matrix **C**) and estimated cell fractions (matrix **W**) enables us to calculate the residual gene expression (Δ) remaining after accounting for cell fraction changes (Fig. 2). Specifically, it is the component of the bulk tissue expression variation that cannot be explained by the changing cell fractions alone: $\Delta = \mathbf{B} - \mathbf{CW}$. We can subsequently use this quantity to determine "residual QTLs" by directly correlating it with genotype. In total, this results in 202,940 SNPs involved in residual eQTLs. Potentially, one can elaborate on this further by allowing the correlations to be done in a cell-type–specific fashion (fig. S35).

To further dissect the associations between genomic elements and QTLs, we compared all of the different types of QTLs with one another and with genomic annotations (Fig. 4E). As expected, eQTLs tended to be enriched at promoters, and cQTLs, at enhancers and transcription factor (TF)–binding sites; fQTLs were spread over many different elements. Also, an appreciable number of eQTLs were enriched on the promoter of a different gene from the one regulated, suggesting the activity of an Epromoter, a regulatory element with dual promoter and enhancer functions (*41*). For the overlap among different QTLs, we expected that most cQTLs and fQTLs would be a subset of the much larger number of eQTLs; somewhat surprisingly, an appreciable number of these did not overlap (Fig. 4F). To evaluate this precisely, we calculated π_1 statistics and found that the cQTL overlap was larger than the fQTL overlap (0.89 versus 0.11). Moreover, eQTL-cQTL overlaps often suggested that the expression-modulating function of an eQTL derived from chromatin changes (e.g., for MTOR) (Fig. 4F). Overall, the total number of overlapping QTLs was 2477 (which we dub multi-QTLs) (Fig. 4F).

Regulatory networks

We next integrated the genomic elements described above into a regulatory network. We first processed a Hi-C dataset for adult brain in the same reference samples used for enhancer identification, providing a physical basis for interactions between enhancers and promoters (Fig. 5A and table S8) (*13*, *21*). In total, we identified 2735 topologically associating domains (TADs) and ~90,000 enhancer-promoter interactions (fig. S36). As expected, ~75% of enhancer-promoter interactions occurred within the same TAD, and genes with more enhancers tended to have higher expression (Fig. 5B and fig. S36). We integrated the Hi-C data with QTLs; surprisingly, QTLs involving SNPs distal to eGenes but linked by Hi-C interactions showed significantly stronger associations (as indicated by the QTL *P* value) than those with SNPs directly in the eGene promoter or exons (Fig. 5C and fig. S37).

To gain insights into the brain chromatin, we compared the adult PsychENCODE Hi-C dataset with those from other tissues in a similar fashion to the transcriptomic and epigenomic comparisons described above. In particular, we selected a set of tissues and cell types from ENCODE and Roadmap, consistently processed their associated Hi-C data at a low resolution, and compared them with our reference-brain Hi-C data. As expected, we found that all the samples for adult brain regions tend to separate markedly from the other tissues in terms of A-B compartment similarity and other metrics (Fig. 5D and fig. S38).

In addition to data for the adult brain, we also added PsychENCODE Hi-C data for the fetal brain into the comparison, assessing the degree to which the chromatin differences between developmental stages relate to those between tissues (Fig. 5D). We found that whereas Hi-C datasets for the adult brain clustered together, the Hi-C dataset for the fetal brain was distinct (Fig. 5D and fig. S39). Only ~31% of the interactions in our adult Hi-C data were detected in the fetal dataset (figs. S39 and S40) (*13*). Though hard to exactly quantify, this difference appears to be larger than that seen from cross-tissue transcriptome comparison, with fetal samples included (fig. S41). We did a number of other comparisons between fetal and adult brain Hi-C datasets, analyzing the regulatory elements and genes linked by each. As expected, we found fetus-linked genes to be more highly expressed prenatally and adult-linked ones postnatally (Fig. 5E). In addition, the fetus-linked genes were preferentially expressed in developmental cell types (Fig. 5F). They were also highly expressed in adult neurons, whereas the adult-linked ones were preferentially expressed in glia, reflecting known cell-type composition (Fig. 5, D and F) (*42*).

In addition to Hi-C linkages, we tried to find further regulatory connections by relating the activity of TFs to target genes (Fig. 5A). In particular, for each potential target of a TF, we created a linkage if it had a "good binding site" (matching the TF's motif) in gene-proximal open chromatin regions (either promoters or brain-active enhancers) and if it had a high coefficient in a regularized, elastic net regression, relating TF activity to target expression (fig. S42) (*21*). Elastic net regression assumes that target gene expression is determined by a linear combination of the expression levels of its regulating TFs, via regression coefficients (using sparsified L_1 and L_2 regularization). Overall, we found that a subset of regulatory connections could predict the expression of 8930 genes with a mean square error (MSE) of <0.05 (fig. S43). For example, we could predict the expression of the ASD-associated gene *CHD8* with MSE = 0.034 (equivalent to coefficient of determination R^2 = 0.77 over the population) (*21*). Lastly, the enhancer-binding TFs with high regression coefficients—implying a high chance for TF regulation of the target genes via particular bound enhancers—provide a third set of putative enhancer-to-gene links.

Collectively, we generated a full regulatory network, linking enhancers, TFs, and target genes (fig. S42). This includes 43,181 proximal and 42,681 distal linkages involving 11,573 protein-encoding target genes (TF–to–target gene via promoter for proximal versus via enhancer–target gene connection for distal) (Fig. 5A) (*15*, *21*). As functioning regulatory connections reflect cell type, we also generated potential cell-type–specific regulatory networks (Fig. 5, F and G, and fig. S44). In these, we found a number of well-known TFs associated with brain development—e.g., NEUROG1, DLGAP2, and MEF2A for excitatory neurons and GAD1, GAD2, and LHX6 for inhibitory neurons (Fig. 5G) (*43*–*46*). Lastly, for broad utility on the resource website, we also provide an expanded regulatory network with slightly different parameterization (fig. S42).

Linking GWAS variants to genes

We used our regulatory network based on Hi-C, QTLs, and activity relationships to connect noncoding GWAS loci to potential disease genes. In particular, for the 142 SCZ GWAS loci, we identified a set of 1111 putative SCZ-associated genes, covering 119 loci (the SCZ genes) (Fig. 6A) (*47*). Of these, 321 constitute a "high-confidence" set supported by more than two evidence sources (e.g., QTLs and Hi-C) (Fig. 6, A and B, and fig. S45); examples include the CHRNA2 and CACNA1C genes (Fig. 6, B and C). Overall, the SCZ genes represent an increase from the 22 genes reported in an earlier QTL study and a larger number than can be linked simply by genomic proximity (176) (Fig. 6A) (*11*, *47*). The majority of SCZ genes were not even in LD with the index SNPs (~67%, or 748 of 1111 genes with r^2 < 0.6) (fig. S45), consistent with the fact that regulatory relationships often do not follow linear genome organization (*13*).

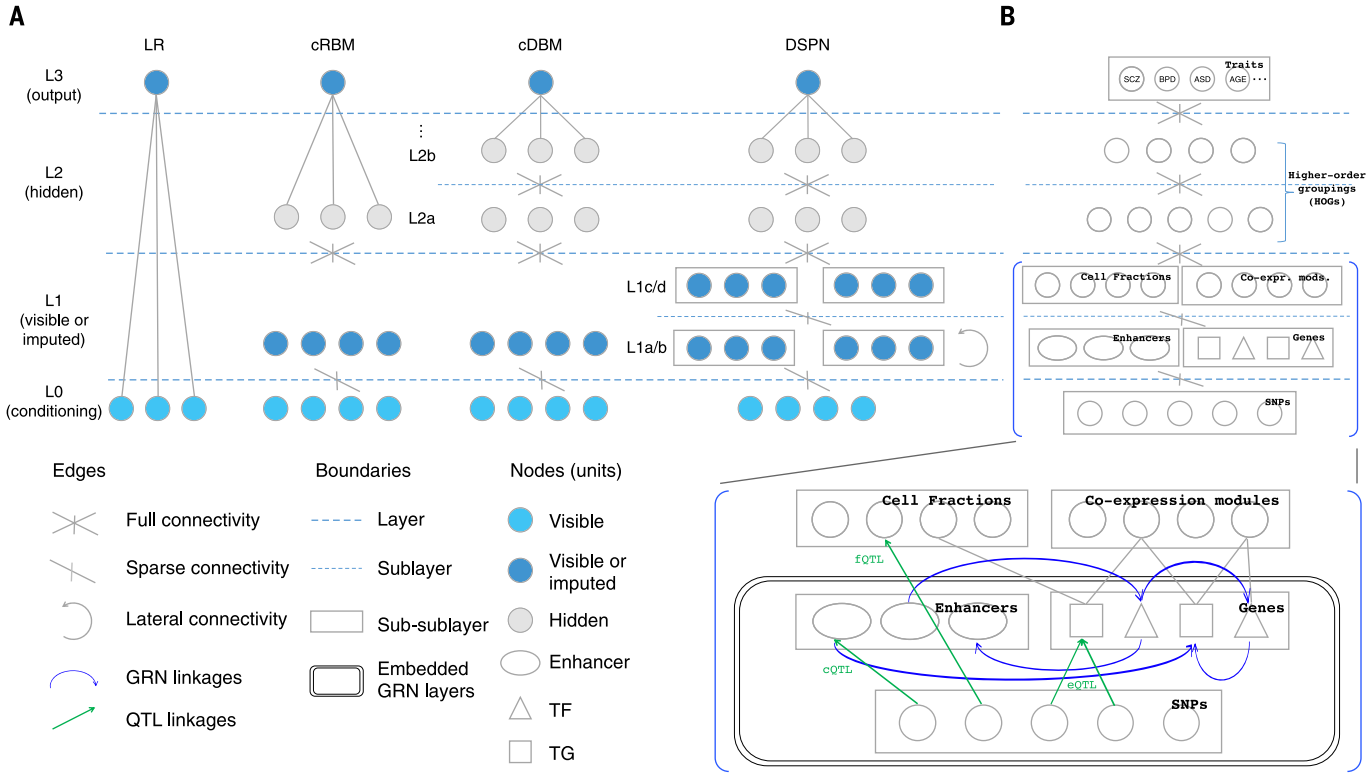
We then looked at the characteristics of the 1111 SCZ genes (and the high-confidence subset

of 321). As expected, they shared many characteristics with known SCZ-associated genes, being enriched in translational regulators, cholinergic receptors, calcium channels, synaptic genes, SCZ differentially expressed genes, and loss-of-function–intolerant genes (fig. S45) (*47*). Next, we identified the TFs regulating the SCZ genes (on the basis of our regulatory network, either directly or via an enhancer) (Fig. 6D). These include LHX9 and SOX7, TFs critical for early cortical specification and neuronal apoptosis, respectively (*48*, *49*). Lastly, we integrated the SCZ genes with single-cell profiles and found that they are highly expressed in neurons, particularly excitatory ones, consistent with the recent findings (Fig. 6E) (*47*).

In addition to SCZ, we also looked at other diseases linked by our regulatory network. In particular, we found aggregate associations between our brain eQTLs and enhancers and many brain disorder GWAS variants, much more so than for GWAS variants for non-brain diseases (Fig. 6F and table S9).

Integrative deep-learning model

The full interaction between genotype and phenotype involves many levels, beyond those encapsulated by the regulatory network. We



C

Method	SCZ	BPD	ASD	AVG (SCZ+BPD+ASD)	GEN	ETH	AGE
LR-gene	54.6% (0.5%)	56.7% (2.5%)	50.0% (0.0%)	53.8% (1.0%)	50.0%	99.0%	61.9%(AOD)
LR-trans	63.0% (4.8%)	63.3% (6.3%)	51.7% (1.8%)	59.3% (4.3%)	69.7%	86.0%	81.2%
cRBM	70.0% (31.0%)	71.1% (22.6%)	63.3% (10.8%)	68.1% (21.5%)	71.5%	89.0%	83.1%
DSPN-impute	59.0% (1.8%)	67.2% (10.7%)	58.8% (3.2%)	61.7% (5.2%)			
DSPN-full	73.6% (32.8%)	76.7% (37.4%)	68.3% (11.3%)	72.9% (27.2%)	71.5%	94.3%	86.9%

Model complexity	increasing	increasing	constant	increasing
Predictors	genotype	transcriptome	genotype->transcriptome	genotype->transcriptome

Unbracketed figures show test-set performance accuracy, with chance at 50%; bracketed figures show variance explained on liability scale

Fig. 7. DSPN deep-learning model links genetic variation to psychiatric disorders and other traits. (**A**) The schematic outlines the structure of the following models: logistic regression (LR), conditional Restricted Boltzmann Machine (cRBM), conditional Deep Boltzmann Machine (cDBM), and DSPN. Nodes are partitioned into four layers (L0 to L3) and colored according to their status as visible, visible or imputed (depending on whether nodes were observed or not at test time), or hidden. (**B**) DSPN structure is shown in further detail, with the biological interpretation of layers L0, L1, and L3 highlighted. The GRN structure learned previously (Fig. 5A) is embedded in layers L0 and L1, with different types of regulatory linkages and functional elements shown. Co-expr. mods., coexpression modules. (**C**) The performance of different models is summarized, with comparisons of performance across models of different complexity and of transcriptome versus genome predictors, corresponding to being with or without imputation for the DSPN (colors highlight relevant models for each comparison). Performance accuracy is shown first, with variance explained on the liability scale in brackets. All models were tested on identical data splits, which were balanced for predicted trait and covariates (including gender, ethnicity, age, and assay). RNA-seq, cell fraction, and H3K27ac data were binarized by thresholding at median values (per gene, cell type, and enhancer, respectively), as was age (median, 51 years) when predicted. LR-gene and LR-trans are logistic models using genetic and transcriptomic predictors, respectively; DSPN-impute and DSPN-full are models with imputed intermediate phenotypes (genotype predictors only) and fully observed intermediate phenotypes (transcriptome predictors), respectively. Differential performance is shown in terms of improvement above chance, with liability variance score increases in brackets. GEN, gender; ETH, ethnicity; AOD, age of individual at death.

addressed this by embedding our regulatory network into a larger multilevel model. In particular, we developed an interpretable deep-learning framework, the Deep Structured Phenotype Network (DSPN) (*21*). This model combines a Deep Boltzmann Machine architecture with conditional and lateral connections derived from the regulatory network (*50*). Traditional classification methods such as logistic regression predict phenotype directly from genotype, without using intermediates such as the transcriptome (Fig. 7A). In contrast, the DSPN is constructed via a series of intermediate models that add layers of structure. We included layers for intermediate molecular phenotypes associated with specific genes (i.e., their gene expression and chromatin state) and predefined gene groupings (cell-type marker genes and coexpression modules), multiple higher layers for inferred groupings (hidden nodes), and a top layer for observed traits (psychiatric disorders and other brain phenotypes). Finally, we used sparse inter- and intralevel connectivity to integrate our knowledge of QTLs, regulatory networks, and coexpression modules from the sections above (Fig. 7B). By using a generative architecture, we ensure that the model is able to impute intermediate phenotypes, as well as provide forward predictions from genotypes to traits.

Using the full model with the genome and transcriptome data provided, we demonstrated that the extra layers of structure in the DSPN allowed us to achieve substantially better trait prediction than traditional additive models (Fig. 7C). For instance, a logistic predictor was able to gain a 2.4-fold improvement when including the transcriptome versus using the genome alone (+9.3% for the transcriptome versus +3.8% for the genome, above a 50% random baseline). By contrast, the DSPN was able to gain a larger, 6-fold improvement (+22.9% versus +3.8%), which may reflect its ability to incorporate nonlinear interactions. This result clearly manifests that the transcriptome carries additional information, which the DSPN is able to extract. Moreover, the DSPN allows us to perform joint inference and imputation of intermediate phenotypes (i.e., transcriptome and epigenome) and observed traits from just the genotype alone, achieving a ~3.1-fold improvement over a logistic predictor in this context (Fig. 7C and fig. S46). Overall, these results demonstrate the usefulness of even a limited amount of functional genomic information for unraveling gene-disease relationships and show that the structure learned from such data can be used to make more accurate predictions of observed traits, even on samples for which intermediate phenotypes are imputed.

We transformed our results to the liability scale for comparison with narrow-sense heritability estimates (Fig. 7C) (*21*). Prior studies have estimated that common SNPs explain 25.6, 20.5, and 19% of the genetic variance for SCZ, BPD, and ASD, respectively (*51*). These may be taken as theoretical upper bounds for additive models, given unlimited common-variant data. By contrast, nonlinear predictors can exceed these limits. Our best liability scores (from just the genotype at QTL-associated variants) are substantially below these bounds, implying that additional data would be beneficial. By contrast, the variance explained by the full DSPN model exceeds that explained by common SNPs in SCZ and BPD, possibly reflecting the influence of rare variants and epistatic interactions (32.8 and 37.4% respectively—the variance of 11.3% for ASD is slightly lower). However, these estimates may be confounded by trait-associated variation that is environmental in origin (fig. S47).

A key aspect of the DSPN is its interpretability. In particular, we examined the specific connections learned by the DSPN between intermediate and high-level phenotypes. Here, we included coexpression modules in the model, referring to this modification as "DSPN-mod" (fig. S48). Using it, we determined which modules were prioritized, as well as the sets of genes associated with latent nodes that were found at each hidden layer (Fig. 8A and table S10) (*15*, *21*). Broadly, we take an unbiased view of all 5024 modules and higher-order groupings constructed from these and then prioritize a subset of ~180 modules and groupings for each psychiatric disorder, showing these to be enriched in specific functional categories and to intersect substantially with the modules from more disease-focused analyses (Fig. 8, B and C, and fig. S49) (*22*). [For completeness, we provide a full table showing the prioritization and functional categories for all possible modules associated with various traits (fig. S50).] In particular, we found that cross-disorder prioritized modules are associated with functional categories such as "immune processes," "synaptic activity," and "splicing," consistent with the findings from more disease-focused analyses (Fig. 8C) (*22*). Also, we showed that prioritized SCZ and BPD modules are enriched for known GWAS SNPs (fig. S51) (for ASD, the lack of GWAS SNPs precludes similar analyses). For SCZ, which is the best characterized of the three disorders, we find enrichments for pathways and genes known to be associated with the disease, including glutamatergic-synapse pathway genes, such as *GRIN1*; calcium-signaling pathway and astrocyte-marker genes; and complement cascade pathway genes such as *C4A*, *C4B*, and *CLU* (Fig. 8D) (*22*). Other prioritized modules include well-characterized genes such as *MIAT*, *RBFOX1*, and *ANK2* (SCZ); *RELA*, *NFkB2*, and *NIPBL* (ASD); and *HOMER1* (BPD), consistent with the results of (*22*). Finally, we identify modules associated with aging, finding that they are enriched in Ex4 neuronal cell-type genes, synaptic and longevity functions, and the gene *NRGN*—all consistent with differential expression analysis (Fig. 8D and fig. S20).

Conclusions

We have developed a comprehensive resource for functional genomics of the adult brain by integrating PsychENCODE data with a broad range of publicly available datasets. In closing, we review our main findings and ways that they can be improved in the future.

First, in terms of QTLs, we identified a set of eQTLs several times as large as those in previous studies, targeting a saturating proportion of protein-coding genes. Moreover, we were able to identify a substantial number of cQTLs. PsychENCODE was, in fact, among the first efforts to generate ChIP-seq data across a large cohort of brain samples, with experiments focused primarily on H3K27ac. In the future, further increasing cohort size and performing additional chromatin assays, such as STARR-seq (self-transcribing active regulatory region sequencing) and ChIP-seq for other histone modifications, will improve the identification of enhancers and cQTLs (*52*). More fundamentally, one-dimensional fluctuations in the chromatin signal reflect changes in three-dimensional chromatin architecture, and new metrics beyond cQTLs may be needed.

Second, in terms of single-cell analysis, we found that varying proportions of basic cell types (with different expression signatures) accounted for a large fraction of the expression variation across a population of individuals. However, this assumes that the expression levels characterizing a signature are fairly constant over a population of cells of a given cell type. In the future, larger-scale single-cell studies will allow us to examine this question in detail, perhaps quantifying and bounding environment-associated transcriptional variability. In addition, current single-cell techniques suffer from low sensitivity and dropouts; thus, it remains challenging to reliably quantify low-abundance transcripts (*15*, *53*). This is particularly the case for specific brain cell substructures, such as axons and dendrites (*15*).

Third, we developed a comprehensive deep-learning model, the DSPN, and used it to illustrate how functional genomics data could improve the link between genotype and phenotype. In particular, by integrating regulatory-network connectivity and latent factors, the DSPN improves trait prediction over traditional additive models. Moreover, it takes into account dependencies between gene expression levels not modeled by univariate eQTL methods. In this study, we kept our eQTL methods very standard, closely following the GTEx paradigm. This separation we make between univariate eQTL detection and multivariate integrative modeling allows us to compare our eQTLs directly with those from previous analyses, such as the CMC study. However, multivariate-based methods for QTLs have been used elsewhere and, in the future, may be combined with our approach (*54*, *55*).

Further, in the future, we can envision how our DSPN approach can be extended to modeling additional intermediate phenotypes. In particular, we can naturally embed in the middle levels of the model additional types of QTLs and phenotype-phenotype interactions—e.g., QTLs associated with microRNAs, neuroimaging, human- and primate-specific genes, and developmental brain enhancers (*56*–*59*).

We expect that the DSPN will improve accuracy mainly for complex traits with a highly

polygenic architecture, but not necessarily for traits that are strongly determined by only a few variants, such as Mendelian disorders, or are closely correlated with population structure, such as ethnicity. However, even when the DSPN performance is low, it may still provide insights about intermediate phenotypes; for instance, in our analysis, the PFC transcriptome appears substantially less predictive with respect to gender (after removing the sex chromosome genes) than age, but this very fact highlights the similarity of the transcriptome between sexes (*60*). Finally, although our focus has been on common SNPs, the DSPN may be able to capture the effects of rare variants, such as those known to be

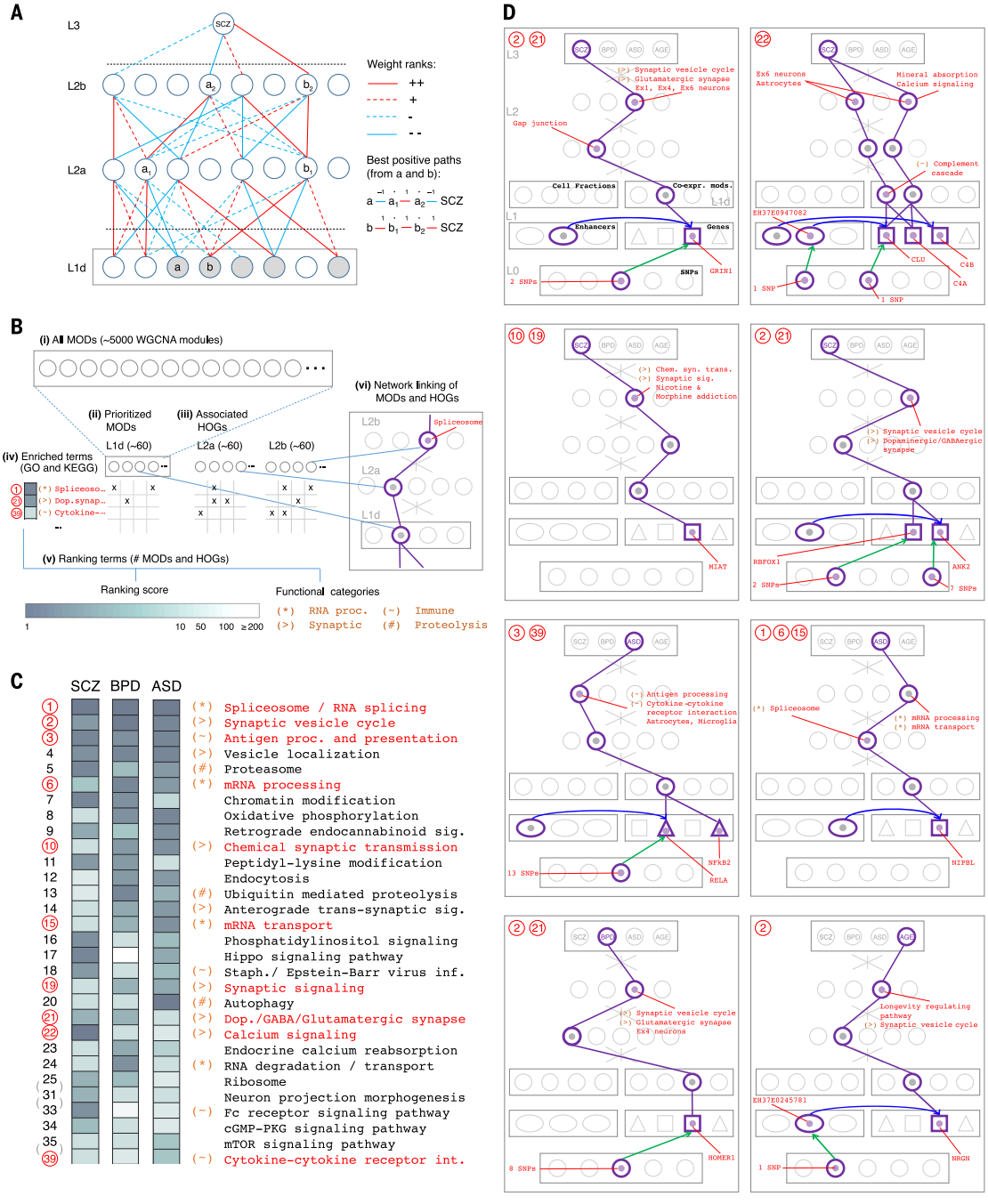


Fig. 8. Interpretation of the DSPN model highlights functional associations and shared disease mechanisms. (**A**) The schematic illustrates the module (MOD) and higher-order grouping (HOG) prioritization schemes. Red and blue lines represent positive and negative weights, respectively, and full and dotted lines represent first and second ranks by absolute value [creating a directed acyclic graph (DAG) with branching factor 4, rooted at L3]. Highlighted nodes (gray) in L1d show positive prioritized MODs, for which a positive path (containing an even number of negative links) exists connecting the module to the SCZ node. a_1/a_2 and b_1/b_2 highlight "best positive paths" from a and b, respectively, to SCZ in terms of absolute rank score. Associated HOGs are defined for a_1/a_2, containing all nodes in L1d having a path in the DAG to a_1 (respectively a_2), which is identically signed to the best path from a to a_1 (respectively a_2) (*21*). Positive prioritized HOGs are associated with nodes on best paths from all positive prioritized MODs; negative prioritized MODs and HOGs are calculated similarly. (**B**) Summary of the functional annotation scheme. (i) A total of 5024 weighted gene coexpression network analysis (WGCNA) MODs (modules and submodules) are derived from multiple data splits. (ii) MODs are prioritized as in (A) for each disorder, and (iii) associated HOGs are calculated. (iv) Gene set enrichment analysis associates functional terms with all MODs and HOGs. (v) Terms are ranked per disorder by counting the number of prioritized MODs or HOGs they associate with, and broad functional categories are defined; (vi) prioritized MODs and HOGs are linked to potentially interesting genes, enhancers, and SNPs by using GRN connectivity. proc., processing. (**C**) Upper segment of cross-disorder ranking of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) functional terms, where cross-disorder ranks are assigned by using the average per-disorder rank ordering. Ranking score levels and functional categories are as in the key in (B). Highlighted ranks and terms correspond to examples shown in (D). See fig. S49 for extended ranking. sig., signaling; staph., staphylococcus; inf., infection; dop., dopamine; cGMP-PKG, guanosine 3′,5′-monophosphate–cGMP-dependent protein kinase; int., interaction. (**D**) Examples of associations between prioritized MODs or HOGs and genes, enhancers, and SNPs for each disorder and age model. Associated functional terms and categories are as in (B). A table providing coordinates of eQTLs and cQTLs for all examples shown is provided in table S13. Chem. syn. trans., chemical synaptic transmission.

implicated in ASD (*51*), through their influence on intermediate phenotypes.

In summary, our integrative analyses demonstrate the usefulness of functional genomics for unraveling molecular mechanisms in the brain (*21*, *61*), and the results of these analyses suggest directions for further research into the etiology of brain disorders.

Materials and methods summary

The materials and methods for each section of the main text are available in the section with same heading in the supplementary materials (*21*); i.e., supplementary content for a given main text section within the supplementary materials is named in a parallel fashion. Detailed data protocols are available in the supplementary materials. Moreover, associated and derived data files are available at the PsychENCODE resource site (*19*). Often we provide multiple versions of the derived summary files with different parameterizations (e.g., for the single-cell profiles and for eQTLs).

REFERENCES AND NOTES

1. R. C. Kessler *et al*., Design and field procedures in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A). *Int. J. Methods Psychiatr. Res.* **18**, 69–83 (2009). doi: 10.1002/mpr.279; pmid: 19507169
2. P. W. Wilson *et al*., Prediction of coronary heart disease using risk factor categories. *Circulation* **97**, 1837–1847 (1998). doi: 10.1161/01.CIR.97.18.1837; pmid: 9603539
3. J. N. Weinstein *et al*., The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013). doi: 10.1038/ng.2764; pmid: 24071849
4. D. M. Lloyd-Jones *et al*., Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. *Circulation* **113**, 791–798 (2006). doi: 10.1161/CIRCULATIONAHA.105.548206; pmid: 16461820
5. M. R. Stratton, P. J. Campbell, P. A. Futreal, The cancer genome. *Nature* **458**, 719–724 (2009). doi: 10.1038/nature07943; pmid: 19360079
6. D. H. Geschwind, J. Flint, Genetics and genomics of psychiatric disease. *Science* **349**, 1489–1494 (2015). doi: 10.1126/science.aaa8954; pmid: 26404826
7. S. Cichon *et al*., Genomewide association studies: History, rationale, and prospects for psychiatric disorders. *Am. J. Psychiatry* **166**, 540–556 (2009). doi: 10.1176/appi.ajp.2008.08091354; pmid: 19339359
8. A. Battle *et al*., Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017). doi: 10.1038/nature24277; pmid: 29022597
9. A. Kundaje *et al*., Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015). doi: 10.1038/nature14248; pmid: 25693563
10. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: 10.1038/nature11247; pmid: 22955616
11. M. Fromer *et al*., Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016). doi: 10.1038/nn.4399; pmid: 27668389
12. C. Colantuoni *et al*., Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519–523 (2011). doi: 10.1038/nature10524; pmid: 22031444
13. H. Won *et al*., Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016). doi: 10.1038/nature19847; pmid: 27760116
14. B. B. Lake *et al*., Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016). doi: 10.1126/science.aaf1204; pmid: 27339989
15. A. E. Saliba, A. J. Westermann, S. A. Gorski, J. Vogel, Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Res.* **42**, 8845–8860 (2014). doi: 10.1093/nar/gku555; pmid: 25053837
16. S. Darmanis *et al*., A survey of human brain transcriptome diversity at the single cell level. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7285–7290 (2015). doi: 10.1073/pnas.1507125112; pmid: 26060301
17. S. Akbarian *et al*., The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015). doi: 10.1038/nn.4156; pmid: 26605881
18. J. T. Walters, M. J. Owen, Endophenotypes in psychiatric genetics. *Mol. Psychiatry* **12**, 886–890 (2007). doi: 10.1038/sj.mp.4002068; pmid: 17895920
19. PsychENCODE Integrative Analysis, http://resource.psychencode.org/.
20. B. B. Lake *et al*., Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat. Biotechnol.* **36**, 70–80 (2018). doi: 10.1038/nbt.4038; pmid: 29227469
21. Materials and methods are available as supplementary materials.
22. M. J. Gandal *et al*., Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science* **362**, eaat8127 (2018).
23. I. Voineagu *et al*., Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011). doi: 10.1038/nature10110; pmid: 21614001
24. M. J. Gandal *et al*., Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. *Science* **359**, 693–697 (2018). doi: 10.1126/science.aad6469; pmid: 29439242
25. M. C. Oldham *et al*., Functional organization of the transcriptome in human brain. *Nat. Neurosci.* **11**, 1271–1282 (2008). doi: 10.1038/nn.2207; pmid: 18849986
26. T. E. Bakken *et al*., A comprehensive transcriptional map of primate brain development. *Nature* **535**, 367–375 (2016). doi: 10.1038/nature18637; pmid: 27409810
27. A. E. Jaffe *et al*., Developmental regulation of human cortex transcription and its clinical relevance at single base resolution. *Nat. Neurosci.* **18**, 154–161 (2015). doi: 10.1038/nn.3898; pmid: 25501035
28. K. W. Kelley, H. Nakao-Inoue, A. V. Molofsky, M. C. Oldham, Variation among intact tissue samples reveals the core transcriptional features of human CNS cell classes. *Nat. Neurosci.* **21**, 1171–1184 (2018). doi: 10.1038/s41593-018-0216-z; pmid: 30154505
29. J. L. Rubenstein, M. M. Merzenich, Model of autism: Increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255–267 (2003). doi: 10.1034/j.1601-183X.2003.00037.x; pmid: 14606691
30. B. C. McKinney *et al*., Hypermethylation of BDNF and SST genes in the orbital frontal cortex of older individuals: A putative mechanism for declining gene expression with age. *Neuropsychopharmacology* **40**, 2604–2613 (2015). doi: 10.1038/npp.2015.107; pmid: 25881116
31. R. Tacutu *et al*., Human ageing genomic resources: New and updated databases. *Nucleic Acids Res.* **46** (D1), D1083–D1090 (2018). doi: 10.1093/nar/gkx1042; pmid: 29121237
32. M. Kasowski *et al*., Extensive variation in chromatin states across humans. *Science* **342**, 750–752 (2013). doi: 10.1126/science.1242510; pmid: 24136358
33. W. Sun *et al*., Histone acetylome-wide association study of autism spectrum disorder. *Cell* **167**, 1385–1397.e11 (2016). doi: 10.1016/j.cell.2016.10.031; pmid: 27863250
34. D. Purves, *Neuroscience* (Oxford Univ., ed. 6, 2018).
35. K. G. Ardlie *et al*., The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015). doi: 10.1126/science.1262110; pmid: 25954001
36. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 9440–9445 (2003). doi: 10.1073/pnas.1530509100; pmid: 12883005
37. R. C. del Rosario *et al*., Sensitive detection of chromatin-altering polymorphisms reveals autoimmune disease mechanisms. *Nat. Methods* **12**, 458–464 (2015). doi: 10.1038/nmeth.3326; pmid: 25799442
38. F. Grubert *et al*., Genetic control of chromatin states in humans involves local and distal chromosomal interactions. *Cell* **162**, 1051–1065 (2015). doi: 10.1016/j.cell.2015.07.048; pmid: 26300125
39. J. Bryois *et al*., Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat. Commun.* **9**, 3121 (2018). doi: 10.1038/s41467-018-05379-y; pmid: 30087329
40. T. Chailangkarn *et al*., A human neurodevelopmental model for Williams syndrome. *Nature* **536**, 338–343 (2016). doi: 10.1038/nature19067; pmid: 27509850
41. L. T. M. Dao *et al*., Genome-wide characterization of mammalian promoters with distal enhancer functions. *Nat. Genet.* **49**, 1073–1081 (2017). doi: 10.1038/ng.3884; pmid: 28581502
42. L. de la Torre-Ubieta, H. Won, J. L. Stein, D. H. Geschwind, Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361 (2016). doi: 10.1038/nm.4071; pmid: 27050589
43. C. Fode *et al*., A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes Dev.* **14**, 67–80 (2000). pmid: 10640277
44. A. H. Rasmussen, H. B. Rasmussen, A. Silahtaroglu, The DLGAP family: Neuronal expression, function and role in brain disorders. *Mol. Brain* **10**, 43 (2017). doi: 10.1186/s13041-017-0324-9; pmid: 28870203
45. M. G. Erlander, N. J. Tillakaratne, S. Feldblum, N. Patel, A. J. Tobin, Two genes encode distinct glutamate decarboxylases. *Neuron* **7**, 91–100 (1991). doi: 10.1016/0896-6273(91)90077-D; pmid: 2069816
46. P. Liodis *et al*., Lhx6 activity is required for the normal migration and specification of cortical interneuron subtypes. *J. Neurosci.* **27**, 3078–3089 (2007). doi: 10.1523/JNEUROSCI.3055-06.2007; pmid: 17376969
47. A. F. Pardiñas *et al*., Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018). doi: 10.1038/s41588-018-0059-2; pmid: 29483656
48. V. S. Mangale *et al*., Lhx2 selector activity specifies cortical identity and suppresses hippocampal organizer fate. *Science* **319**, 304–309 (2008). doi: 10.1126/science.1151695; pmid: 18202285
49. C. Wang *et al*., SOX7 interferes with β-catenin activity to promote neuronal apoptosis. *Eur. J. Neurosci.* **41**, 1430–1437 (2015). doi: 10.1111/ejn.12910; pmid: 25847511
50. R. Salakhutdinov, G. Hinton, "Deep Boltzmann machines," *Proc. 12th Int. Conf. Artif. Intell. Stat.* **5**, 448–455 (2009).
51. V. Anttila *et al*., Analysis of shared heritability in common disorders of the brain. *Science* **360**, eaap8757 (2018). doi: 10.1126/science.aap8757; pmid: 29930110
52. Y. Liu *et al*., Functional assessment of human enhancer activities using whole-genome STARR-sequencing. *Genome Biol.* **18**, 219 (2017). doi: 10.1186/s13059-017-1345-5; pmid: 29151363
53. S. Liu, C. Trapnell, Single-cell transcriptome sequencing: Recent advances and remaining challenges. *F1000Res.* **5**, 10.12688/f1000research.7223.1 (2016). pmid: 26949524
54. H. Chun, S. Keles, Expression quantitative trait loci mapping with multivariate sparse partial least squares regression. *Genetics* **182**, 79–90 (2009). doi: 10.1534/genetics.109.100362; pmid: 19270271
55. M. P. Scott-Boyer *et al*., An integrated hierarchical Bayesian model for multivariate eQTL mapping. *Stat. Appl. Genet. Mol. Biol.* **11**, /j/sagmb.2012.11.issue-4/1544-6115.1760/1544-6115.1760.xml (2012). doi: 10.1515/1544-6115.1760; pmid: 22850063
56. C. E. Bearden, P. M. Thompson, Emerging global initiatives in neurogenetics: The Enhancing Neuroimaging Genetics through Meta-analysis (ENIGMA) Consortium. *Neuron* **94**, 232–236 (2017). doi: 10.1016/j.neuron.2017.03.033; pmid: 28426957
57. T. G. M. van Erp *et al*., Cortical brain abnormalities in 4474 individuals with schizophrenia and 5098 control subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. *Biol. Psychiatry* **84**, 644–654 (2018). doi: 10.1016/j.biopsych.2018.04.023; pmid: 29960671
58. A. M. M. Sousa *et al*., Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358**, 1027–1032 (2017). doi: 10.1126/science.aan3456; pmid: 29170230
59. A. Amiri *et al*., Transcriptome and epigenome landscape of human cortical development modeled in organoids. *Science* **362**, eaat6720 (2018).
60. O. V. Evgrafov *et al*., Gene expression in patient-derived neural progenitors provide insights into neurodevelopmental aspects of schizophrenia. bioRxiv 209197 [Preprint]. 26 October 2017. doi: 10.1101/209197
61. M. Li *et al*., Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).

ACKNOWLEDGMENTS

Funding: Data were generated as part of the PsychENCODE Consortium, supported by U01MH103339, U01MH103365, U01MH103392, U01MH103340, U01MH103346, U01MH116492, R01MH105472, R01MH094714, R01MH105898, R01MH110920, R01MH110905, R01MH110926, R01MH110927, R01MH110928, R01MH109715, R01MH111721, R21MH102791, R01MH110921, R01MH109677, R21MH105881, R21MH103877, R21MH109956, R21MH105853, and P50MH106934, awarded to S.A. (Icahn School of Medicine at Mount Sinai); A. Chess (Icahn School of Medicine at Mount Sinai); G.E.C. (Duke University); S. Dracheva (Icahn School of Medicine at Mount Sinai); P. Farnham (University of Southern California); Z.W. (UMass Medical School); M.B.G. (Yale University); D.H.G. (University of California–Los Angeles); F. Goes (Johns Hopkins University); T. M. Hyde (Lieber Institute for Brain Development); A.E.J. (Lieber Institute for Brain Development); J.A.K. (SUNY Downstate Medical Center); C. Liu (SUNY Upstate Medical University); D. Pinto (Icahn School of Medicine at Mount Sinai); P.R. (Icahn School of Medicine at Mount Sinai); N.S. (Yale University); P. Sklar (Icahn School of Medicine at Mount Sinai); M. State (University of California–San Francisco); P. Sullivan (University of North Carolina); F. Vaccarino (Yale University); D. Weinberger (Lieber Institute for Brain Development); S. Weissman (Yale University); K.P.W. (University of Chicago); A. J. Willsey (University of California–San Francisco); and P. Zandi (Johns Hopkins University), L. Mangravite and M.P. (Sage Bionetworks), and A. Arguello, L. Bingaman, T. Lehner, D. Panchision, and G. Senthil (NIMH). A subset of the transcriptome (RNA-seq) data was generated as part of the CMC, supported by funding from Takeda Pharmaceuticals, F. Hoffman–La Roche, and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH-075916, P50M096891, P50MH084053S1, R37MH057881, R37MH057881S1, HHSN271201300031C, AG02219, AG05138, and MH06692. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories, and the NIMH Human Brain Collection Core. CMC leadership is as follows: P. Sklar and J. Buxbaum (Icahn School of Medicine at Mount Sinai), B. Devlin and D. Lewis (University of Pittsburgh), R. Gur and C.-G. Hahn (University of Pennsylvania), K. Hirai and H. Toyoshiba (Takeda Pharmaceuticals), E. Domenici and L. Essioux (F. Hoffman–La Roche), L. Mangravite and M.P. (Sage Bionetworks), and T. Lehner and B. Lipska (NIMH). Z.H.G. is supported by NIH/NIAID grant U19AI118610. **Author contributions:** All individually named authors contributed substantially to the paper through either data generation or analysis: data generation, H.W., M.J.G., Y.J., G.E.H., G.E.C., P.R., S.A., A.E.J., K.P.W., N.S., and D.H.G.; and data analysis, D.W., S.Li., J.W., H.W., X.S., F.C.P.N., D.C., M.G., P.E., Y.T.Y., M.X., M.J.G., S.Lo., J.Z., J.J.P., C.Y., S.K.R., K.M., H.Z., A.N., M.P., E.M., D.F., T.B., J.M., K.G., S.K., Z.H.G., G.E.C., P.R., S.A., A.E.J., K.P.W., Z.W., N.S., D.H.G., J.A.K., and M.B.G. The three corresponding authors co-led the analysis. **Competing interests:** G.E.C. is a co-founder of Element Genomics. K.P.W. is associated with Tempus Labs. The other authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Associated data files are available under individual file identifiers at the PsychEncode resource site (*19*).

The PsychENCODE Consortium

Allison E. Ashley-Koch[1], Gregory E. Crawford[1], Melanie E. Garrett[1], Lingyun Song[1], Alexias Safi[1], Graham D. Johnson[1], Gregory A. Wray[1], Timothy E Reddy[1], Fernando S. Goes[2], Peter Zandi[2], Julien Bryois[3], Andrew E. Jaffe[4], Amanda J. Price[4], Nikolay A. Ivanov[4], Leonardo Collado-Torres[4], Thomas M. Hyde[4], Emily E. Burke[4], Joel E. Kleiman[4], Ran Tao[4], Joo Heon Shin[4], Schahram Akbarian[5], Kiran Girdhar[5], Yan Jiang[5], Marija Kundakovic[5], Leanne Brown[5], Bibi S. Kassim[5], Royce B. Park[5], Jennifer R Wiseman[5], Elizabeth Zharovsky[5], Rivka Jacobov[5], Olivia Devillers[5], Elie Flatow[5], Gabriel E. Hoffman[5], Barbara K. Lipska[6], David A. Lewis[7], Vahram Haroutunian[5,8], Chang-Gyu Hahn[9], Alexander W. Charney[10], Stella Dracheva[10], Alexey Kozlenkov[10], Judson Belmont[5], Diane DelValle[5], Nancy Francoeur[5], Evi Hadjimichael[5], Dalila Pinto[5], Harm van Bakel[5], Panos Roussos[10], John F. Fullard[10], Jaroslav Bendl[10], Mads E. Hauberg[10], Lara M Mangravite[11], Mette A. Peters[11], Yooree Chae[11], Junmin Peng[12], Mingming Niu[12], Xusheng Wang[12], Maree J. Webster[13], Thomas G. Beach[14], Chao Chen[15], Yi Jiang[15], Rujia Dai[15], Annie W. Shieh[16], Chunyu Liu[16], Kay S. Grennan[16], Yan Xia[15,16], Ramu Vadukapuram[16], Yongjun Wang[15], Dominic Fitzgerald[17], Lijun Cheng[17], Miguel Brown[17], Mimi Brown[17], Tonya Brunetti[17], Thomas Goodman[17], Majd Alsayed[17], Michael J. Gandal[18], Daniel H. Geschwind[18], Hyejung Won[18], Damon Polioudakis[18], Brie Wamsley[18], Jiani Yin[18], Tarik Hadzic[18], Luis De La Torre Ubieta[18], Vivek Swarup[18], Stephan J. Sanders[19], Matthew W. State[19], Donna M. Werling[19], Joon-Yong An[19], Brooke Sheppard[19], A. Jeremy Willsey[19], Kevin P. White[17], Mohana Ray[17], Gina Giase[16], Amira Kefi[20], Eugenio Mattei[21], Michael Purcaro[21], Zhiping Weng[21], Jill Moore[21], Henry Pratt[21], Jack Huey[21], Tyler Borrman[21], Patrick F. Sullivan[22], Paola Giusti-Rodriguez[22], Yunjung Kim[22], Patrick Sullivan[22], Jin Szatkiewicz[22], Suhn Kyong Rhie[23], Christoper Armoskus[23], Adrian Camarena[23], Peggy J. Farnham[23], Valeria N. Spitsyna[23], Heather Witt[23], Shannon Schreiner[23], Oleg V. Evgrafov[24], James A. Knowles[24], Mark Gerstein[25], Shuang Liu[25], Daifeng Wang[26], Fabio C. P. Navarro[25], Jonathan Warrell[25], Declan Clarke[25], Prashant S. Emani[25], Mengting Gu[25], Xu Shi[25], Min Xu[25], Yucheng T. Yang[25], Robert R. Kitchen[25], Gamze Gürsoy[25], Jing Zhang[25], Becky C. Carlyle[25], Angus C. Nairn[25], Mingfeng Li[25], Sirisha Pochareddy[25], Nenad Sestan[25], Mario Skarica[25], Zhen Li[25], Andre M. M. Sousa[25], Gabriel Santpere[25], Jinmyung Choi[25], Ying Zhu[25], Tianliuyun Gao[25], Daniel J. Miller[25], Adriana Cherskov[25], Mo Yang[25], Anahita Amiri[25], Gianfilippo Coppola[25], Jessica Mariani[25], Soraya Scuderi[25], Anna Szekely[25], Flora M. Vaccarino[25], Feinan Wu[25], Sherman Weissman[25], Tanmoy Roychowdhury[27], Alexej Abyzov[27]

[1]Duke University, Durham, NC, USA. [2]Johns Hopkins University, Baltimore, MD, USA. [3]Karolinska Institutet, Stockholm, Sweden. [4]Lieber Institute for Brain Development, Baltimore, MD, USA. [5]Icahn School of Medicine at Mount Sinai, New York, NY, USA. [6]Human Brain Collection Core, National Institutes of Health, Bethesda, MD, USA. [7]University of Pittsburgh, Pittsburg, PA, USA. [8]James J. Peters VA Medical Center, Bronx, NY, USA. [9]University of Pennsylvania, Philadelphia, PA, USA. [10]Mount Sinai, New York, NY, USA. [11]Sage Bionetworks, Seattle, WA, USA. [12]St. Jude Children's Hospital, Memphis, TN, USA. [13]Stanley Medical Research Institute, Kensington, MD, USA. [14]Banner Sun Health Research Institute, Sun City, AZ, USA. [15]Central South University, Changsha, Hunan, China. [16]SUNY Upstate Medical University, Syracuse, NY, USA. [17]The University of Chicago, Chicago, IL, USA. [18]University of California–Los Angeles, Los Angeles, CA, USA. [19]University of California–San Francisco, San Francisco, CA, USA. [20]University of Illinois at Chicago, Chicago, IL, USA. [21]University of Massachusetts Medical School, Worcester, MA, USA. [22]University of North Carolina–Chapel Hill, Chapel Hill, NC, USA. [23]University of Southern California, Los Angeles, CA, USA. [24]SUNY Downstate Medical Center, Brooklyn, NY, USA. [25]Yale University, New Haven, CT, USA. [26]Stony Brook University, Stony Brook, NY, USA. [27]Mayo Clinic Rochester, Rochester, MN, USA.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/eaat8464/suppl/DC1
Supplementary Text
Figs. S1 to S52
Tables S1 to S13
References (*62*–*111*)

10 April 2018; accepted 15 November 2018
10.1126/science.aat8464

RESEARCH ARTICLE SUMMARY

PSYCHIATRIC GENOMICS

Spatiotemporal transcriptomic divergence across human and macaque brain development

Ying Zhu*, André M. M. Sousa*, Tianliuyun Gao*, Mario Skarica*, Mingfeng Li*, Gabriel Santpere, Paula Esteller-Cucala, David Juan, Luis Ferrández-Peral, Forrest O. Gulden, Mo Yang, Daniel J. Miller, Tomas Marques-Bonet, Yuka Imamura Kawasawa, Hongyu Zhao, Nenad Sestan†

INTRODUCTION: Improved understanding of how the developing human nervous system differs from that of closely related nonhuman primates is fundamental for teasing out human-specific aspects of behavior, cognition, and disorders.

RATIONALE: The shared and unique functional properties of the human nervous system are rooted in the complex transcriptional programs governing the development of distinct cell types, neural circuits, and regions. However, the precise molecular mechanisms underlying shared and unique features of the developing human nervous system have been only minimally characterized.

RESULTS: We generated complementary tissue-level and single-cell transcriptomic datasets from up to 16 brain regions covering prenatal and postnatal development in humans and rhesus macaques (*Macaca mulatta*), a closely related species and the most commonly studied nonhuman primate. We created and applied TranscriptomeAge and TempShift algorithms to age-match developing specimens between the species and to more rigorously identify temporal differences in gene expression within and across the species. By analyzing regional and temporal patterns of gene expression in both the developing human and macaque brain, and comparing these patterns to a complementary dataset that included transcriptomic information from the adult chimpanzee, we identified shared and divergent transcriptomic features of human brain development. Furthermore, integration with single-cell and single-nucleus transcriptomic data covering prenatal and adult periods of both species revealed that the developmental divergence between humans and macaques can be traced to distinct cell types enriched in different developmental times and brain regions, including the prefrontal cortex, a region of the brain associated with distinctly human aspects of cognition and behavior.

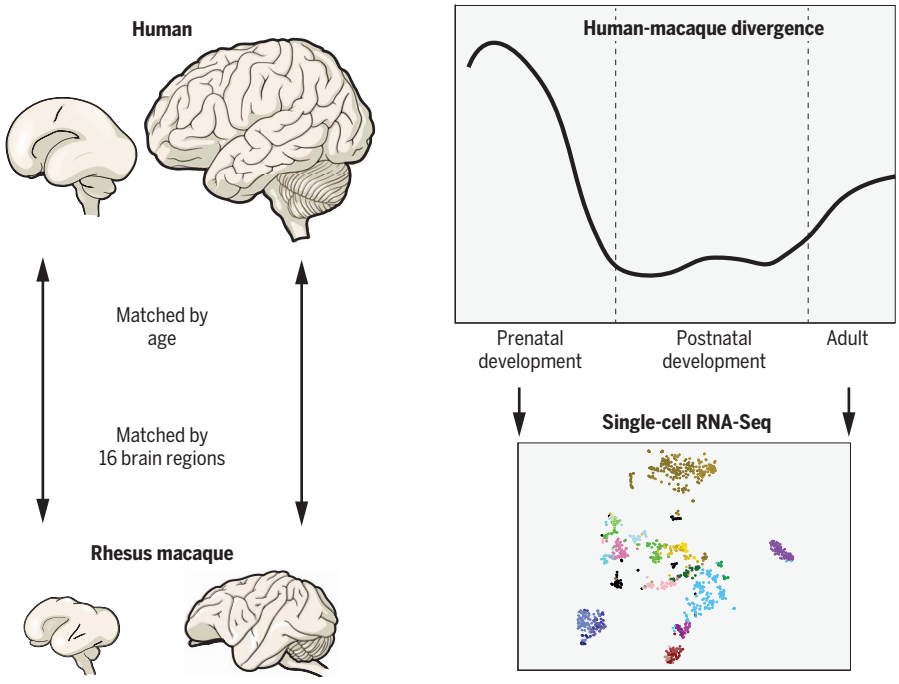
ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat8077

We found two phases of prominent species differences: embryonic to late midfetal development and adolescence/young adulthood. This evolutionary cup-shaped or hourglass-like pattern, with high divergence in prenatal development and adolescence/young adulthood and lower divergence in early postnatal development, resembles the developmental cup-shaped pattern described in the accompanying study by Li *et al.* Even though the developmental (ontogenetic) and evolutionary (phylogenetic) patterns have similar profiles, the overlap of genes driving these two patterns is not substantial, indicating the existence of different molecular mechanisms and constraints for regional specification and species divergence.

Notably, we also identified numerous genes and gene coexpression modules exhibiting human-distinct patterns in either temporal (heterochronic) or spatial (heterotopic) gene expression, as well as genes with human-distinct developmental expression, linked to autism spectrum disorder, schizophrenia, and other neurological or psychiatric diseases. This finding potentially suggests mechanistic underpinnings of these disorders.

CONCLUSION: Our study provides insights into the evolution of gene expression in the developing human brain and may shed some light on potentially human-specific underpinnings of certain neuropsychiatric disorders. ■



Concerted ontogenetic and phylogenetic transcriptomic divergence in human and macaque brain. Left: Human and macaque brain regions spanning both prenatal and postnatal development were age-matched using TranscriptomeAge. Right: Phylogenetic transcriptomic divergence between humans and macaques resembles the developmental (ontogenetic) cup-shaped pattern of each species, with high divergence in prenatal development and adolescence/young adulthood and lower divergence during the early postnatal period (from perinatal to adolescence). Single-cell transcriptomics revealed shared and divergent transcriptomic features of distinct cell types.

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: nenad.sestan@yale.edu
Cite this article as Y. Zhu *et al.*, *Science* **362**, eaat8077 (2018). DOI: 10.1126/science.aat8077

RESEARCH ARTICLE

PSYCHIATRIC GENOMICS

Spatiotemporal transcriptomic divergence across human and macaque brain development

Ying Zhu[1,2]*, André M. M. Sousa[1]*, Tianliuyun Gao[1]*, Mario Skarica[1]*, Mingfeng Li[1]*, Gabriel Santpere[1], Paula Esteller-Cucala[3], David Juan[3], Luis Ferrández-Peral[3], Forrest O. Gulden[1], Mo Yang[1], Daniel J. Miller[1], Tomas Marques-Bonet[3,4,5,6], Yuka Imamura Kawasawa[7], Hongyu Zhao[2], Nenad Sestan[1,8]†

Human nervous system development is an intricate and protracted process that requires precise spatiotemporal transcriptional regulation. We generated tissue-level and single-cell transcriptomic data from up to 16 brain regions covering prenatal and postnatal rhesus macaque development. Integrative analysis with complementary human data revealed that global intraspecies (ontogenetic) and interspecies (phylogenetic) regional transcriptomic differences exhibit concerted cup-shaped patterns, with a late fetal-to-infancy (perinatal) convergence. Prenatal neocortical transcriptomic patterns revealed transient topographic gradients, whereas postnatal patterns largely reflected functional hierarchy. Genes exhibiting heterotopic and heterochronic divergence included those transiently enriched in the prenatal prefrontal cortex or linked to autism spectrum disorder and schizophrenia. Our findings shed light on transcriptomic programs underlying the evolution of human brain development and the pathogenesis of neuropsychiatric disorders.

The development of the human nervous system is an intricate process that unfolds over a prolonged time course, ranging from years to decades, depending on the region (*1*–*6*). Precise spatial and temporal regulation of gene expression is crucial for all aspects of human nervous system development, evolution, and function (*6*–*13*). Consequently, alterations in this process have been linked to psychiatric and neurological disorders, some of which may exhibit primate- or human-specific manifestations (*11*, *14*–*18*). However, our ability to explain many aspects of human nervous system development and disorders at a mechanistic level has been limited by our evolutionary distance from genetically tractable model organisms, such as the mouse (*15*, *16*, *19*–*22*), and by a lack of contextual and functional interpretations of polymorphisms and disease-associated variations in the human and nonhuman primate (NHP) genomes (*11*, *17*, *21*, *23*). Moreover, neither the extent of molecular changes underlying human-specific differences nor the specific developmental programs affected by these changes have been thoroughly studied.

The rhesus macaque (*Macaca mulatta*) is the most widely studied NHP in neuroscience and medicine (*24*–*26*). The macaque nervous system parallels the human nervous system with its complex cellular architecture and extended development, and thereby offers a unique opportunity to study features of neurodevelopment that are shared and divergent between the two closely related primates. Furthermore, studies of post mortem NHP tissues provide a unique opportunity to validate results obtained using post mortem human tissue, especially those from critical developmental periods that can be confounded by ante mortem and post mortem factors and tissue quality. Finally, substantial advances in transgenic and genome-editing technologies now allow the possibility of creating more precise genetic models for human disorders in macaques (*24*–*26*). This will facilitate the interrogation of the effects of specific gene mutations in a model that is closer to the human brain than any other experimental animal.

Comparative transcriptomic profiling offers unbiased insight into conserved and clade- or species-specific molecular programs underlying cellular and functional development of the human nervous system (*27*–*31*). However, a systematic characterization of the spatial and temporal transcriptomic landscapes of the macaque brain at the region-specific and single-cell levels, as well as the identification of shared and divergent features between humans and macaques, are lacking. Data and analyses such as we present here should provide both retrospective and prospective benefits to the fields of neuroscience, evolutionary biology, genomics, and medicine.

Study design, data generation, and integrated analysis

RNA sequencing (RNA-seq) data were obtained from bulk tissue (366 samples from 26 prenatal and postnatal brains) or single cells/nuclei (113,274 cells or nuclei from two fetal and three adult brains) from post mortem rhesus macaque specimens. Both tissue and single cell/nucleus datasets were subjected to multiple quality control measures (figs. S1 to S6 and tables S1 and S2) (*32*). Tissue-level samples covered the entire span of both prenatal and postnatal neurodevelopment (Fig. 1, A and B, and table S1) and included 11 areas of the cerebral neocortex (NCX), hippocampus (HIP), amygdala (AMY), striatum (STR), mediodorsal nucleus of thalamus (MD), and cerebellar cortex (CBC). Subject ages ranged from 60 post-conception days (PCD) to 11 postnatal years (PY) and were matched by age and brain region to 36 human brains from an accompanying study (*33*) and five adult chimpanzee brains from a previous study (*34*) (Fig. 1A). To investigate the contribution of different factors to the global transcriptome dynamics, we applied unsupervised clustering and principal components analysis, which revealed that age, species, and regions contributed more to the global transcriptomic differences than did other tested variables (figs. S3 and S4).

To explore cell type origins of tissue-level interspecies differences, we conducted single-cell RNA-seq (scRNA-seq) on 86,341 cells from six matching regions of two 110-PCD fetal macaque brains [i.e., the dorsolateral prefrontal neocortex (DFC, also called DLPFC), HIP, AMY, STR, MD, and CBC] and single-nucleus RNA-seq (snRNA-seq) of 26,933 nuclei from three adult macaque DFCs (8, 11, and 11 PY; tables S2 and S3) (*32*). These data were complemented by 17,093 snRNA-seq samples from adult humans [see (*33*)] as well as two scRNA-seq datasets from embryonic and fetal human NCX (*33*, *35*). In the six fetal macaque brain regions, we identified 129 transcriptomically distinct clusters of cell types (i.e., 19 in DFC, 20 in HIP, 25 in AMY, 22 in STR, 20 in MD, and 23 in CBC) (figs. S7 to S12 and tables S3 and S4). In the adult human DFC (fig. S13) and adult macaque DFC (fig. S14), we identified 29 and 21 transcriptomically distinct cell types, respectively (tables S3, S5, and S6). Alignment of our macaque fetal data with the adult single-nucleus data revealed hierarchical relationships and similarities between major cell classes, reflecting their ontogenetic origins and functional properties (fig. S15). Cell clusters were categorized

[1]Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT, USA. [2]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA. [3]Institute of Evolutionary Biology (UPF-CSIC), PRBB, Barcelona, Spain. [4]Catalan Institution of Research and Advanced Studies (ICREA), Barcelona, Spain. [5]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [6]Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Barcelona, Spain. [7]Departments of Pharmacology and Biochemistry and Molecular Biology, Institute for Personalized Medicine, Penn State University College of Medicine, Hershey, PA, USA. [8]Departments of Genetics, Psychiatry, and Comparative Medicine, Program in Cellular Neuroscience, Neurodegeneration and Repair, and Yale Child Study Center, Yale School of Medicine, New Haven, CT, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: nenad.sestan@yale.edu

by their gene expression patterns and assigned identities commensurate with their predicted cell type and, in the case of human adult neocortical excitatory neurons, their putative laminar identity. Although the majority of cell clusters were composed of cells derived from all brains, we found a few clusters in subcortical regions (AMY, 2 of 25 clusters; CBC, 1 of 23 clusters; STR, 1 of 22 clusters) that included cells from a single donor brain. This might be due to variations in dissection, age (even though both fetal macaques were 110 PCD, a 3- to 4-day variation remains), individual differences, and other technical bias. We used the single-cell datasets in this and the accompanying study (*33*) to deconvolve tissue-level RNA-seq data, identify temporal changes in cell type–specific signatures, analyze differences in cell types and their

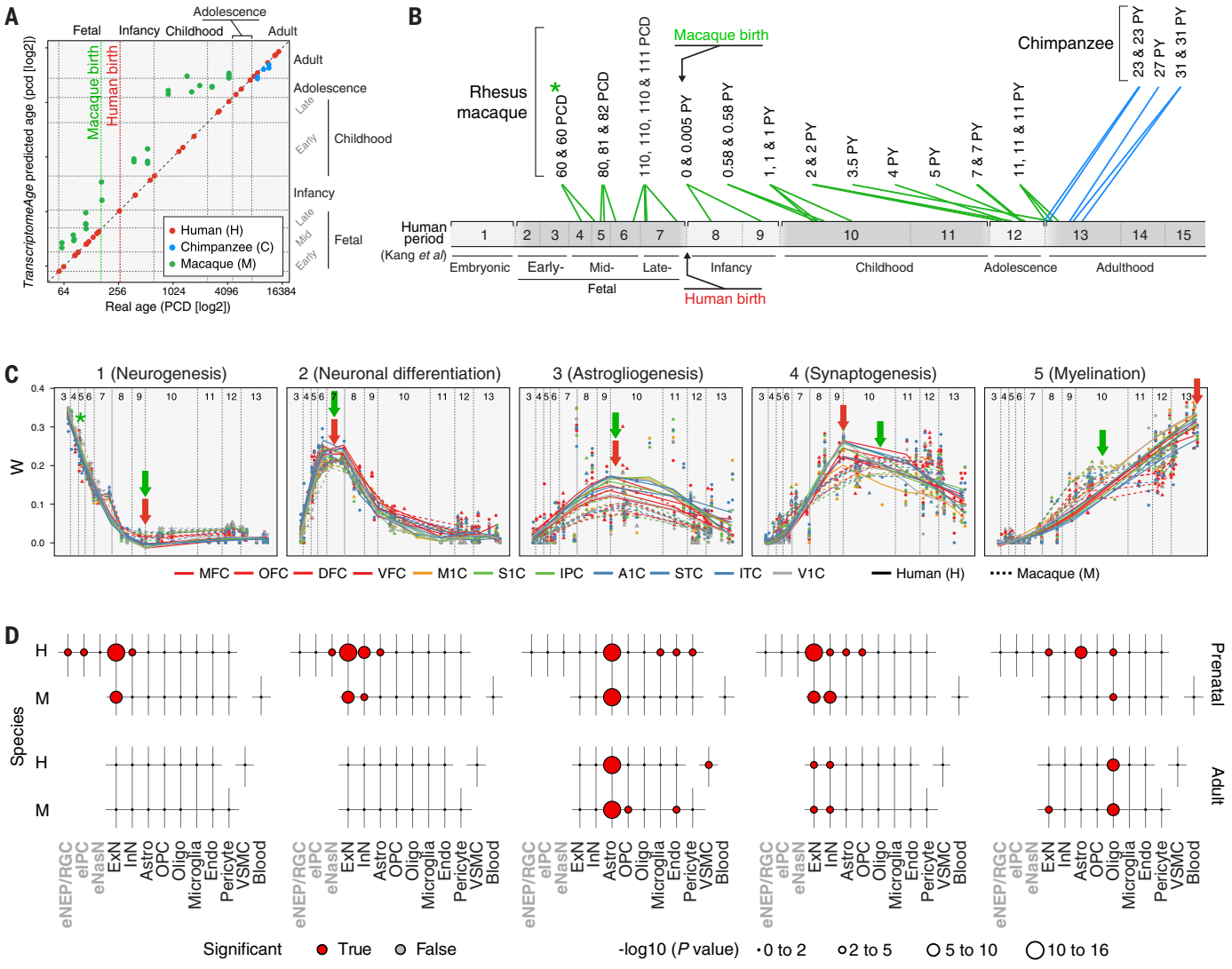


Fig. 1. Conserved and divergent transcriptomic features of human and macaque neurodevelopmental processes. (**A**) Plot depicting the real age (*x* axis) and the age predicted by TranscriptomeAge (*y* axis) of human, chimpanzee, and macaque. Macaque (164 PCD) and human (266 PCD) births are shown as green and red dashed lines, respectively. (**B**) Schematic showing human developmental periods as described in Kang *et al*. (*29*) and the matched macaque developmental and chimpanzee adult datasets. Each line corresponds to one macaque or one chimpanzee specimen and the corresponding predicted age when compared to human neurodevelopment. PCD, post-conception day; PY, postnatal year. The asterisk indicates the extension of the early fetal period, in which early fetal macaques (60 PCD) cluster with midfetal humans. (**C**) The weight (W) of five transcriptomic signatures in the developing human (solid line) and macaque (dashed line) NCX and the respective association with neurodevelopmental processes. In signature 1 (neurogenesis), the arrow indicates the point at which the signature reaches the minimum in humans (red) and macaques (green). The asterisk indicates the same as in (B). In transcriptomic signatures 2, 3, 4, and 5, arrows indicate the point at which the signatures reach the maximum in humans (red) and macaques (green). Note that for transcriptomic signatures 2 and 3 (neuronal differentiation and astrogliogenesis), there is a synchrony between humans and macaques, whereas for transcriptomic signatures 4 and 5 (synaptogenesis and myelination), there is heterochrony between the species, with acceleration in human synaptogenesis and delay in human myelination. Prefrontal cortical areas are plotted in red, primary motor cortex in orange, parietal areas in green, temporal areas in blue, and primary visual cortex in gray. MFC, medial prefrontal cortex; OFC, orbital prefrontal cortex; DFC, dorsolateral prefrontal cortex; VFC, ventrolateral prefrontal cortex; M1C, primary motor cortex; S1C, primary somatosensory cortex; IPC, inferior posterior parietal cortex; A1C, primary auditory cortex; STC, superior temporal cortex; ITC, inferior temporal cortex; V1C, primary visual cortex. (**D**) Cell type enrichment is shown for each signature. *P* values adjusted by Benjamini-Hochberg procedure are plotted (with ranges indicated by size of dots); significance is labeled by color (red, true; gray, false). H, human; M, macaque; eNEP/RGC, embryonic neuroepithelial progenitor/radial glial cell; eIPC, embryonic intermediate progenitor cell; eNasN, embryonic nascent neuron; ExN, excitatory neuron; InN, interneuron; Astro, astrocyte; OPC, oligodendrocyte progenitor cell; Oligo, oligodendrocyte; Endo, endothelial cell; VSMC, vascular smooth muscle cell.

transcriptomic profiles, and conduct cell type enrichment analyses.

Similarities and differences in the spatiotemporal dynamics of the human and macaque brain transcriptomes

Unsupervised hierarchical clustering and principal components analysis of bulk tissue revealed common principles of transcriptomic regional architecture across development in macaques and humans (figs. S3 and S4). Among macaque regions, these analyses showed distinct and developmentally regulated clustering of the NCX (combination of 11 areas), HIP, and AMY, with CBC exhibiting the most distinctive transcriptional profile—an observation shared with our complementary study in humans (*27*, *29*, *30*, *33*, *36*). A hierarchical clustering of both fetal and postnatal NCX areal samples revealed their grouping by topographical proximity and functional overlap, similar to those relationships that we observed in the human brain (fig. S3). Thus, these results show that the transcriptomic architecture of the macaque brain is regionally and temporally

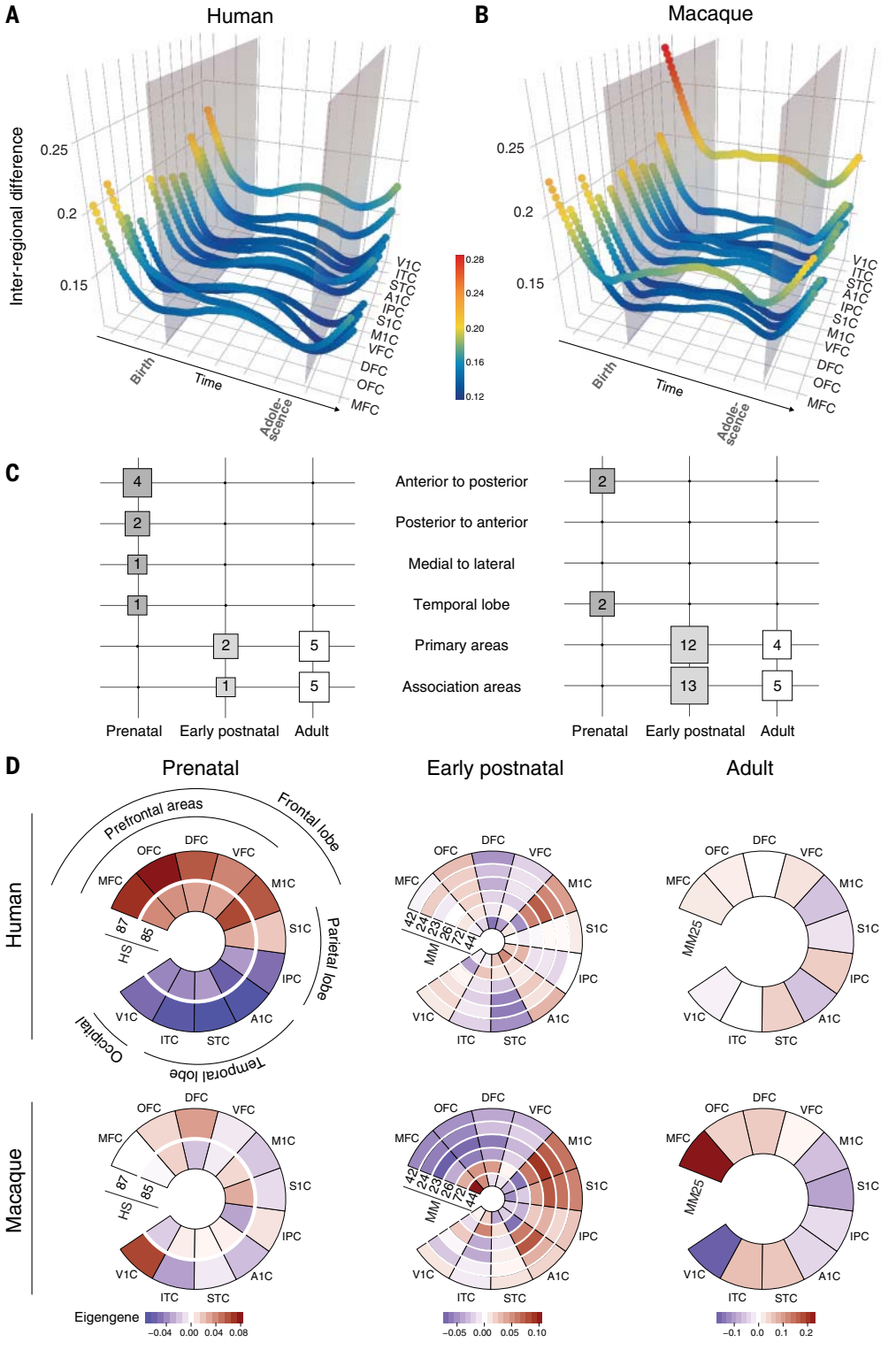


Fig. 2. Ontogenetic interregional transcriptomic differences display a cup-shaped pattern in humans and macaques. (**A** and **B**) The interregional difference was measured as the average distance of each neocortical area to all other areas in the human (A) and macaque (B) neocortices across development. The upper-quartile interregional difference among all genes is plotted; the color scale indicates magnitude. The gray planes represent the transition from prenatal to early postnatal development (late fetal transition) and from adolescence to adulthood. (**C**) The number of coexpression modules that display gradient-like expression (anterior to posterior, posterior to anterior, medial to lateral, temporal lobe–enriched) and enrichment in primary areas or enrichment in association areas in each developmental phase. Left, human modules; right, macaque modules. (**D**) Donut plots depicting the modules from (C) that exhibited species-distinct interregional differences. The expression pattern of each species-distinct module is shown for humans (top) and macaques (bottom). Color scales indicate expression level of the genes in each module. Prenatal modules show a human-distinct anterior-to-posterior expression gradient (left); macaque-distinct early postnatal modules show enrichment in primary or association areas (center); and a macaque-distinct adult module is enriched in association areas, especially in MFC (right). HS, human (*Homo sapiens*) module; MM, macaque (*Macaca mulatta*) module.

specified and reflects conserved global patterns of ontogenetic and functional differences that are also found in humans.

To explore species similarities and differences in the spatiotemporal dynamics of the brain transcriptome, we used the XSAnno computational framework (*37*) to minimize biases in comparative data analyses arising from the disparate quality of gene annotation for the two species. We created common annotation sets of 27,932 and 26,514 orthologous protein-coding and noncoding mRNA genes for human-macaque and human-chimpanzee-macaque comparisons, respectively (fig. S2) (*32*). Next, we developed TranscriptomeAge, an algorithm to unbiasedly predict the equivalent

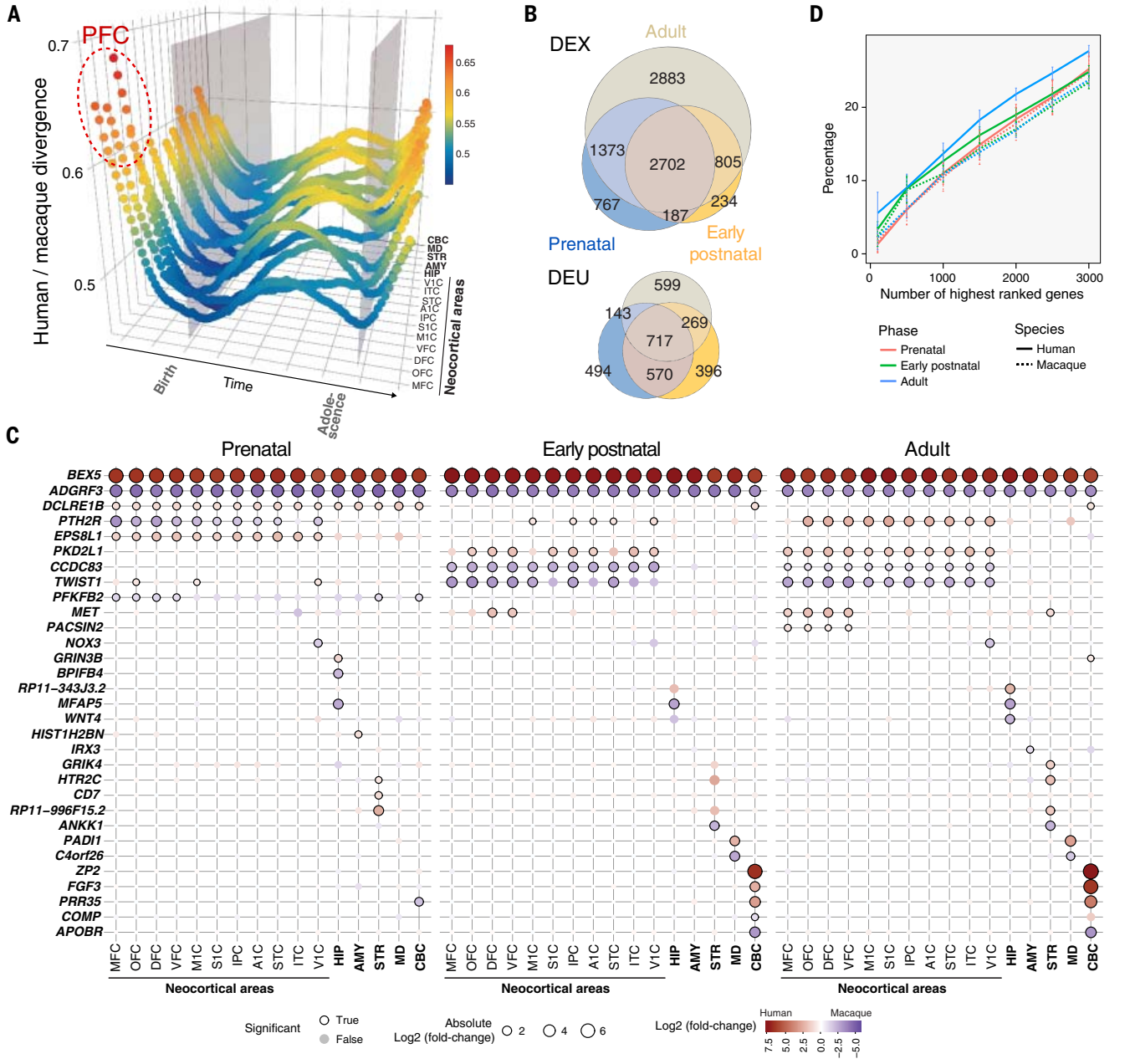


Fig. 3. Transcriptomic divergence between humans and macaques throughout neurodevelopment reveals a phylogenetic cup-shaped pattern. (**A**) Interspecies divergence, measured as absolute difference in gene expression, between humans and macaques in each brain region throughout development (coded as in Fig. 2A). The upper-quartile divergence among all genes is plotted. The gray planes represent the transition from prenatal to early postnatal development (late fetal transition, left) and from adolescence to adulthood (right). (**B**) Venn diagrams displaying the number of differentially expressed genes (DEX, top) or genes with differential exon usage (DEU, bottom) between humans and macaques in at least one brain region during prenatal development, early postnatal development, and adulthood. (**C**) Bubble matrix with examples of genes showing global or regional interspecies differential expression. Brain regions displaying significant differential expression between humans and macaques are shown with black circumference. Red circles show up-regulation in humans; blue, up-regulation in macaques. Circle size indicates absolute $\log_2$ fold change. (**D**) Percentage of overlap between genes showing the highest interspecies divergence in each region (driving the evolutionary cup-shaped pattern) and genes with the largest pairwise distance between brain regions in prenatal, early postnatal, and adult human and macaque brains (driving the developmental cup-shaped pattern). The result is plotted using a variable number of the highest-ranked genes based on interregional difference and interspecies divergence. Data are means ± SD across regions.

ages of human and macaque samples on the basis of temporal transcriptomic changes (*32*). We chose to optimize this model for age-matching the aforementioned 11 neocortical areas, which are highly similar in terms of their transcriptomes, cellular composition, and developmental trajectories when compared to other brain regions [see (*33*)]. TranscriptomeAge confirmed transcriptomic similarities in both species coinciding with major prenatal and postnatal developmental phases, including fetal development, infancy, childhood, and adulthood (Fig. 1, A and B, and figs. S16 to S18). However, we identified two human developmental periods where alignment suggested that they are transcriptomically distinct from macaques and/or are especially protracted. First, 60-PCD macaque specimens [which correspond to the human early fetal period (*29*) according to the Translating Time model (*38*)] were most closely aligned with midfetal human samples (102 to 115 PCD, i.e., 14.5 to 16.5 post-conception weeks). This suggests that, transcriptomically, human brain development is protracted even at early fetal periods. Second, we found that 2-, 3.5-, 4-, 5-, and 7-PY macaque specimens, of which at least the youngest should chronologically match to human childhood (*39*), did not align with any of our human specimens from early or late childhood [1 to 12 PY, or periods 9 and 10 according to (*29*)] but did align with adolescent and adult humans (Fig. 1, A and B). Consistent with previous morphophysiological and behavioral studies (*5*), these results indicate that macaques lack global transcriptomic signatures of late childhood and/or that humans have a

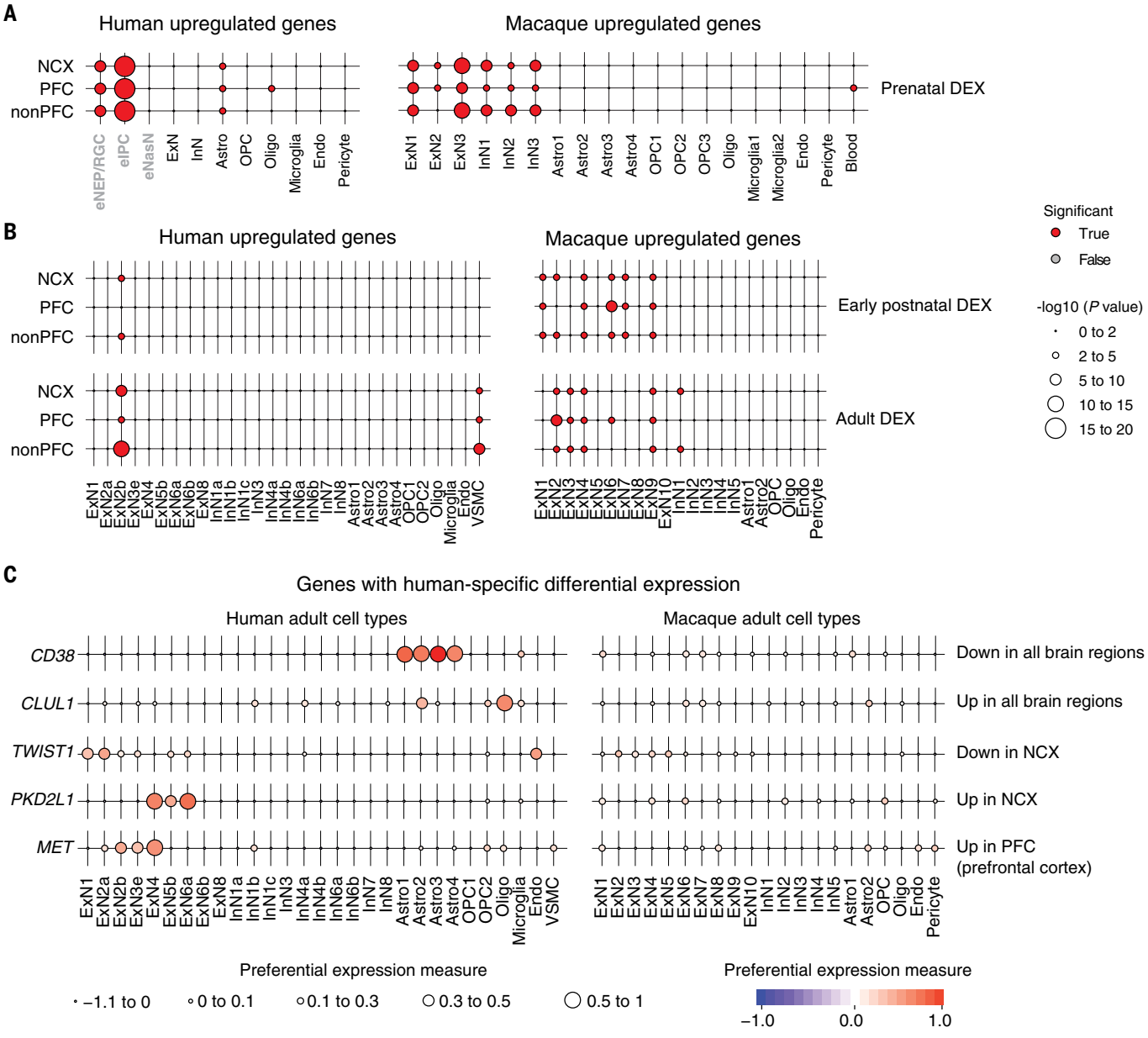


Fig. 4. Cell type specificity of species differences. (**A**) Cell type enrichment for differentially expressed genes up- or down-regulated in human neocortical areas. Enrichment of genes up-regulated in humans or macaques was tested using single cells from prenatal human NCX (*33*) or macaque DFC, respectively. The plot shows $-\log_{10} P$ values adjusted by Benjamini-Hochberg procedure averaged across all neocortical areas (NCX), prefrontal areas (PFC), and non-prefrontal areas (nonPFC). Significance (average $-\log10\ P > 2$) is labeled by color (red, true; gray, false). (**B**) Same as (A) for early postnatal and adult periods. (**C**) Cell type enrichment of selected genes showing human-distinct up- or down-regulation in adult brain regions or neocortical areas (*34*). Preferential expression measure is plotted to show the cell type specificity.

prolonged childhood relative to macaques (Fig. 1, A and B).

Species differences in the timing of concerted neurodevelopmental processes

We hypothesized that the observed developmental differences between humans and macaques might be grounded on transcriptomic changes in concerted biological processes in developmental timing (i.e., heterochrony). By decomposing the gene expression matrix of human neocortical samples, we identified five transcriptomic signatures underlying neocortical development (*32*). Using top cell type–specific genes derived from our prenatal single-cell and adult single-nucleus data, we analyzed cell type enrichment of each of the five signatures, and ascribed them to neurogenesis, neuronal differentiation, astrogliogenesis, synaptogenesis, and oligodendrocyte differentiation and myelination (Fig. 1, C and D, and fig. S19). To determine whether the transcriptomic signatures we identified were correctly assigned, we compared their developmental patterns to the timing of major human neurodevelopmental processes, expression trajectories of key genes previously implicated in those processes, and trajectories of cell type proportions identified by the deconvolution of tissue-level data (figs. S19 and S20). We found that the developmental trajectories of genes associated with neuronal differentiation, synaptogenesis, and myelination, as well as the cell type proportions of fetal human or macaque excitatory neurons, astrocytes, and oligodendrocytes, matched those of the corresponding transcriptomic signatures (fig. S20). Moreover, the identities we assigned to these transcriptomic signatures were confirmed by comparison of transcriptomic signatures to independently generated nontranscriptomic data predicting the start and end of human neocortical neurogenesis (for neurogenesis) (*40*) and to data measuring the number of doublecortin (DCX)–immunopositive nascent neurons in the human hippocampus throughout development and adulthood (for neuronal migration and initial differentiation) (*41*), developmental variation in synaptic density in the human cortex (for synaptogenesis) (*42*), and myelinated fiber length density (for myelination) (*43*) (fig. S19).

Next, we analyzed how the shape of the five transcriptomic trajectories was conserved across the 11 neocortical areas within each species and between species. Analysis of their trajectories within each species revealed that the shape of a given trajectory is similar across neocortical areas (Fig. 1C and fig. S17). However, the transcriptomic trajectories associated with oligodendrocyte differentiation and myelination exhibited a prominent temporal shift (asynchrony) across neocortical areas in both species (fig. S17). Between species, myelination and, to a lesser extent, synaptogenesis exhibited species differences in the shapes of these trajectories; the myelination transcriptomic signature progressively increased in the human NCX beginning from late fetal development through adulthood without reaching an obvious plateau until 40 PY, but in the macaque NCX the myelination signature reached a plateau around the first postnatal year (Fig. 1C). This corresponds to early childhood in human neurodevelopment [window 6 or period 10 according to (*33*) and (*29*), respectively] and is consistent with histological studies and reflective of previously reported hierarchical maturation of neocortical areas (*43*–*47*). Similarly, we corroborated synchronous or concurrent transcriptomic patterns of neocortical synaptogenesis by analyzing previously collected data on synaptic density in multiple areas of the macaque NCX (*48*) (fig. S19). However, we observed that the synaptogenesis transcriptomic trajectory peaked earlier in humans than in macaques, at the transition between late infancy and early childhood (Fig. 1C). In addition, expression trajectories of genes induced by neuronal activity—a process critical for synaptogenesis—also showed drastic increases during late fetal development and infancy, and, like the synaptogenesis trajectory, displayed a concurrent or synchronous shape across neocortical areas [see (*33*)]. Interestingly, the developmental transcriptomic profile of *DCX* (a marker of nascent, migrating neurons) showed that macaques maintain higher expression in the hippocampus throughout postnatal development and adulthood; this suggests that postnatal neurogenesis is more prominent in the macaque hippocampus than in the human hippocampus, as recently shown (fig. S19) (*49*). Thus, both species exhibited distinct transcriptomic signatures of neoteny, such as prolonged myelination in humans and prolonged postnatal hippocampal neurogenesis in macaques. Together, these data suggest that the temporal staging of major neurodevelopmental processes, in particular with myelination beginning in primary areas before association neocortical areas, is a conserved feature of primate development, although the temporal progression of certain processes is heterochronic.

Concerted ontogenetic and phylogenetic transcriptomic divergence

After matching the global transcriptome by age between the two species, we analyzed regional differences in gene expression (heterotopy) within each species. By adopting Gaussian-process models to accommodate the spatiotemporal correlations of gene expression (*32*), we found that the developmental cup-shaped or hourglass-like pattern of transcriptomic interregional differences we observed in humans (*33*) is also present in macaque neocortices and other brain regions (Fig. 2, A and B, and fig. S21), with greater differential expression between regions observed during early and midfetal ages preceding this period and subsequent young adulthood. Notably, two brain regions—CBC and STR—exhibited greater differences, relative to other brain regions, beginning immediately after birth, rather than beginning during childhood or adolescence (fig. S21). This suggests that the development of the primate forebrain may be constrained by unique developmental or evolutionary influences, which led us to investigate the gene expression patterns, developmental processes, and cell types underlying this transcriptomic phenomenon.

To do so, we considered three phases of brain development mirroring major transitions in the cup-shaped pattern: prenatal development, early postnatal development, and adulthood. Between these three phases are two transitional periods: a steep late fetal transition (*33*) and a more moderate transition between childhood/adolescence and adulthood. We performed weighted gene coexpression network analysis (WGCNA) independently for each phase and species, resulting in *Homo sapiens* (HS) and macaque (*Macaca mulatta*, MM) modules (*32*) (table S7), with analyses conducted on 11 neocortical areas; this allowed us to identify discrete spatiotemporal expression patterns that otherwise might be comingled as a result of the highly disparate nature of CBC and other non-neocortical regions. Within the prenatal phase, we found 12 modules consisting of genes exhibiting spatial expression gradients along the anterior-posterior (8 modules) and medial-lateral (1 module) axes of the NCX and broadly reflecting prospective neocortical areal topography (Fig. 2C). For example, prenatal modules HS85 and HS87 exhibited prefrontal/frontal-enriched graded expression in the human brain, tapering to lowest expression in the temporal and occipital lobes (Fig. 2D). Furthermore, prenatal modules, such as HS15 and MM57, had their highest expression restricted to the temporal lobe (table S8 and figs. S22 and S23) during prenatal development.

In contrast to the prenatal phase, modules identified from early postnatal development (i.e., infancy, childhood, and adolescence) in either species did not exhibit anterior-to-posterior or medial-to-lateral expression gradients. Rather, the greater regional synchrony characterizing gene expression in this phase yielded differences organized not around topography but between primary and association areas of the NCX (Fig. 2C, figs. S24 and S25, and table S9). This suggests that the gradient-like transcriptomic patterns arising during prenatal development are superseded by myelination and neuronal activity–related processes postnatally, which may differentiate the separation between primary and association areas. Early postnatal modules such as MM42, MM24, and MM23, among others, exhibited greater expression in primary areas such as the primary motor cortex (M1C), primary auditory cortex (A1C), and primary visual cortex (V1C) than in association areas such as DFC and ventrolateral prefrontal cortex (VFC) (Fig. 2D).

The transition to young adulthood was marked by another decrease in interregional differences, but this reduction was not as pronounced as in the late fetal transition, nor were interregional patterns of gene expression markedly different in the adult. Thus, gene expression differences between primary and association areas continued to drive regional variation in both adult humans and macaques (Fig. 2, C and D, figs. S26 and S27, and table S10). Gene Ontology (GO) enrichment

analysis using the top variant genes in each period, with all genes expressed in each period as background, indicated differential enrichment of biological processes associated with different cell populations across areas and time. As observed in the accompanying human study (*33*) and commensurate with the developmental trajectories of the observed transcriptomic signatures, the functional terms enriched prenatally were generally related to neurogenesis and neuronal differentiation, whereas early postnatal and adult functional terms were enriched for processes related to synaptogenesis and myelination (fig. S28).

We next sought to determine whether the regional-specific expression patterns of coexpression modules detected in human brains correlated with their expression patterns in macaque brains, and vice versa (*32*). We found that two human prenatal modules contained genes exhibiting a pronounced anterior-to-posterior gradient in the human NCX, HS85 and HS87, but these genes did not exhibit enriched expression in the macaque prefrontal cortex (Fig. 2D and table S8). Among genes in these modules were *RGMA* and *SLIT3*, two genes encoding axon guidance molecules (*50*), and *BRINP2* and *CXXC5*, which encode proteins involved in retinoic acid signaling (*51*), potentially implicating this signaling pathway—critical for early brain development and neuronal differentiation (*51*)—in the patterning of the human prefrontal cortex. We also observed that several modules in macaque postnatal development that did not correlate well with human modules (MM23, MM24, MM26, and MM42) were enriched for genes that are expressed in oligodendrocytes (Fig. 2D, fig. S24, and table S9) and were up-regulated in all primary areas of macaque NCX relative to association areas. Conversely, genes in these modules were up-regulated in humans only in M1C and A1C, but not in primary somatosensory cortex (S1C) or V1C (fig. S24 and table S9). Integration with our multi-regional database of the adult chimpanzee transcriptome (*34*) indicates that the macaque gene expression pattern, rather than the human gene expression pattern, may be unique among these species (fig. S29). Many of the species-specific patterns of diversification between primary and association areas that we observed during early postnatal development were preserved in adult modules of both species (fig. S26), with some notable exceptions. For example, the adult macaque module MM25 exhibited up-regulation in association areas in both species, but prominent up-regulation in the medial prefrontal cortex (MFC) and down-regulation in V1C were observed only in macaques (Fig. 2D, fig. S26, and table S10).

These findings reaffirm a conserved framework in primate neocortical development and function (*21*), including a topographic basis for transcriptomic differences during prenatal development and functional relationships postnatally. Our analyses also suggest that interregional and interspecies differences in oligodendrocyte development and myelination, particularly during early postnatal development, mediate key aspects of transcriptomic variation both within and among species.

Heterotopic changes in human and macaque brain transcriptomes

We next investigated the transcriptomic divergence between humans and macaques for each brain region across development. We found that the developmental phases exhibiting high levels of interregional differences within each species (i.e., prenatal development and young adulthood) also displayed greater transcriptomic divergence between the two species, revealing a concerted phylogenetic (evolutionary) cup-shaped pattern (Fig. 3A). This phylogenetic cup-shaped pattern divided neurodevelopment into the same three phases as the regional ontogenetic (developmental) cup shape (Fig. 3A). However, unlike the ontogenetic (developmental) cup-shaped pattern, where CBC, MD, and STR disproportionally exhibited more intraspecies differences than NCX, HIP, and AMY, all regions appeared to exhibit a relatively similar amount of interspecies differences (Fig. 3A). Interestingly, interspecies differences among neocortical areas were distinct enough to provide clear clustering of topographically and functionally related prefrontal areas [i.e., MFC, orbital prefrontal cortex (OFC), DFC, and VFC], particularly during prenatal development, or topographically distributed nonvisual primary areas (i.e., M1C, S1C, and A1C) in adulthood. Prospective areas of the prefrontal cortex, which underlie some of the most distinctly human aspects of cognition, were more phylogenetically distinct than other neocortical areas during early prenatal development (Fig. 3A and fig. S30). Together, these findings suggest that the evolutionary and developmental constraints acting on the brain transcriptome, in particular the NCX, may share some overlapping features.

To gain insight into the transcriptomic programs driving phylogenetic divergence across neocortical areas, we conducted a functional annotation of the top 100 genes driving the observed variation along the first principal component (PC1). We found that interspecies divergence in the prenatal prefrontal cortex could be explained by an enrichment of genes related to cell proliferation [false discovery rate (FDR) < 10^{-5}]. This indicated that the observed interspecies divergence in the prefrontal cortex was likely due to a different proportion of progenitor cells in the early fetal human prefrontal tissue samples (fig. S30). In contrast, during postnatal development, PC1 separated prefrontal areas and the inferior temporal cortex (ITC) from the other neocortical areas. This pattern was mainly driven by genes associated with myelination-associated categories (FDR < 0.05; fig. S30) and genes associated with synaptic transmission (FDR < 0.05; fig. S29). Although speculative, these observations potentially link the expansion of the human prefrontal cortex, the wealth of human-specific connectivity made possible by that extension, and the altered patterns of myelination we observe between humans and macaques.

Confirming the observed regional diversification in each species, postnatal development displayed the lowest number of differentially expressed genes between species; most of these (89.3%) were also differentially expressed in adulthood, the phase where we observed the greatest number of interspecies differentially expressed genes (Fig. 3B and table S11). Genes differentially expressed between humans and macaques exhibited distinct patterns of spatiotemporal divergence (Fig. 3C) and showed diverse functional enrichment (table S12). Although 229 genes (2.6%) displayed up- or down-regulation in all the sampled brain regions throughout development and adulthood, others were specifically up- or down-regulated in a subset of brain regions and/or during a particular developmental phase.

To test whether genes with differential expression between humans and macaques showed distinct conservation profiles, we compared values of dN/dS (the ratio of nonsynonymous to synonymous substitution rates) for the whole set of genes differentially expressed in any of the 16 brain regions in at least one of the three developmental phases (*32*). We found that the differentially expressed genes between humans and macaques also show significantly higher dN/dS values associated with higher evolutionary rates than the remaining protein-coding genes (Wilcoxon-Mann-Whitney $P = 2.2 \times 10^{-8}$, $n = 4429$ genes). This result was also observed when we focused on the genes differentially expressed in prenatal development ($P = 3.7 \times 10^{-11}$, $n = 2380$ genes), early postnatal development ($P = 4.5 \times 10^{-24}$, $n = 1765$ genes), or adulthood ($P = 1.0 \times 10^{-6}$, $n = 3837$ genes) separately. Moreover, these higher dN/dS values for differentially expressed genes remained highly significant in all the brain regions and developmental phases analyzed, highlighting the consistent association between interspecies transcriptional variation and gene evolution.

Integration with our complementary dataset generated on adult chimpanzee brains (*34*) revealed that 531 (10.6%), 507 (12.9%), and 1079 (13.9%) genes differentially expressed between species in prenatal development, early postnatal development, and adulthood, respectively, showed human-specific expression in the same brain region in the adult brain. Several genes among those exhibiting species- or human-specific patterns of gene expression were developmentally and regionally regulated. *PKD2L1*, a gene that encodes an ion channel (*52*), exhibited human-specific up-regulation only postnatally (Fig. 3C). Conversely, *TWIST1*, a gene encoding a transcriptional factor implicated in Saethre-Chotzen syndrome (*53*), showed human-specific down-regulation only postnatally (Fig. 3C). In contrast, *MET*, a gene linked to autism spectrum disorders (*54*), showed human-specific up-regulation in the prefrontal cortex and STR postnatally (Fig. 3C). *PTH2R*, a gene encoding the parathyroid hormone 2 receptor, exhibited macaque-distinct up-regulation in the prenatal NCX but human-distinct up-regulation in the adult NCX, and immunohistochemistry showed that *PTH2R* is

enriched in excitatory neurons (fig. S31). These results show that at least some of the tissue-level interspecies differences we observed are due to changes at the level of specific cell types. Furthermore, even though the ontogenetic and phylogenetic patterns have similar profiles, the overlap of genes driving these two patterns is not substantial (Fig. 3D), indicating the existence of different molecular mechanisms and constraints for regional specification and species divergence.

To gain a more complete understanding of the interspecies transcriptomic differences, we performed an analysis of interspecies differential exon usage as a conservative way of exploring the impact of putative differential alternative splicing. We detected largely similar numbers of genes containing differentially used exons between species in all developmental phases (*32*) (table S13), with 1924 genes showing interspecies differential exon usage in at least one brain region during the prenatal phase, 1952 during the early postnatal phase, and 1728 during adulthood (Fig. 3B and fig. S32). In our set of differentially used exonic elements, non–protein-coding regions were overrepresented ($P < 2.2 \times 10^{-16}$, χ^2 independence test), with 4705 of the 5372 differentially used exonic elements in noncoding regions. This enrichment was especially strong for non–untranslated region (UTR) exonic elements belonging to non–protein-coding transcripts from protein-coding genes and 5′ UTR regions ($P < 2.2 \times 10^{-16}$), but was also significant

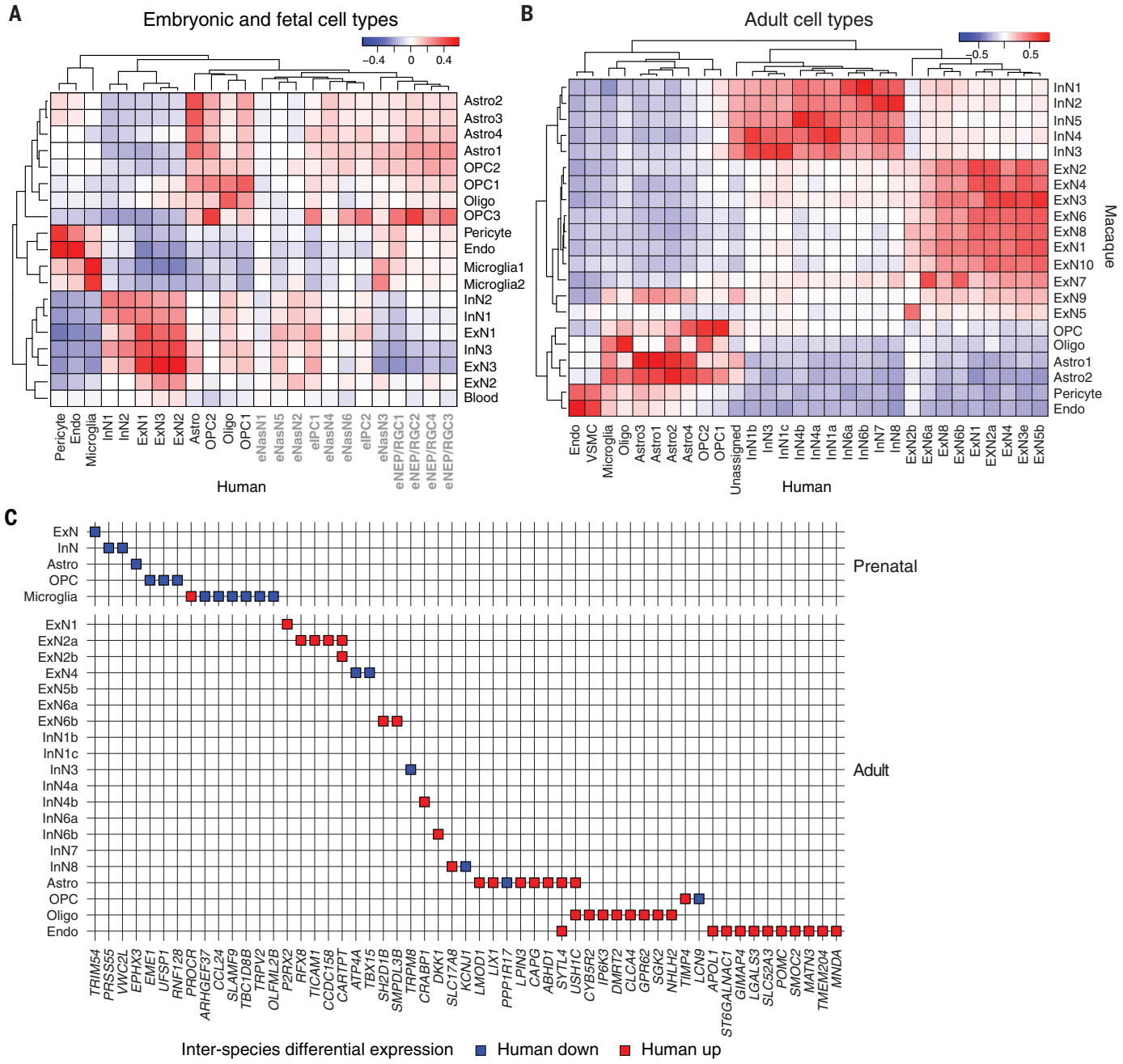


Fig. 5. Shared and divergent transcriptomic features of homologous cell types between humans and macaques. (**A**) Dendrogram and heat map showing diversity and correlation of prenatal cell types within and between the two species. The human single cells were from (*33*). (**B**) Dendrogram and heat map showing diversity and correlation of adult cell types within and between the two species. (**C**) Cell type specificity of interspecies differentially expressed genes based on the single cell/nucleus information. Blue, human down-regulated genes; red, human up-regulated genes.

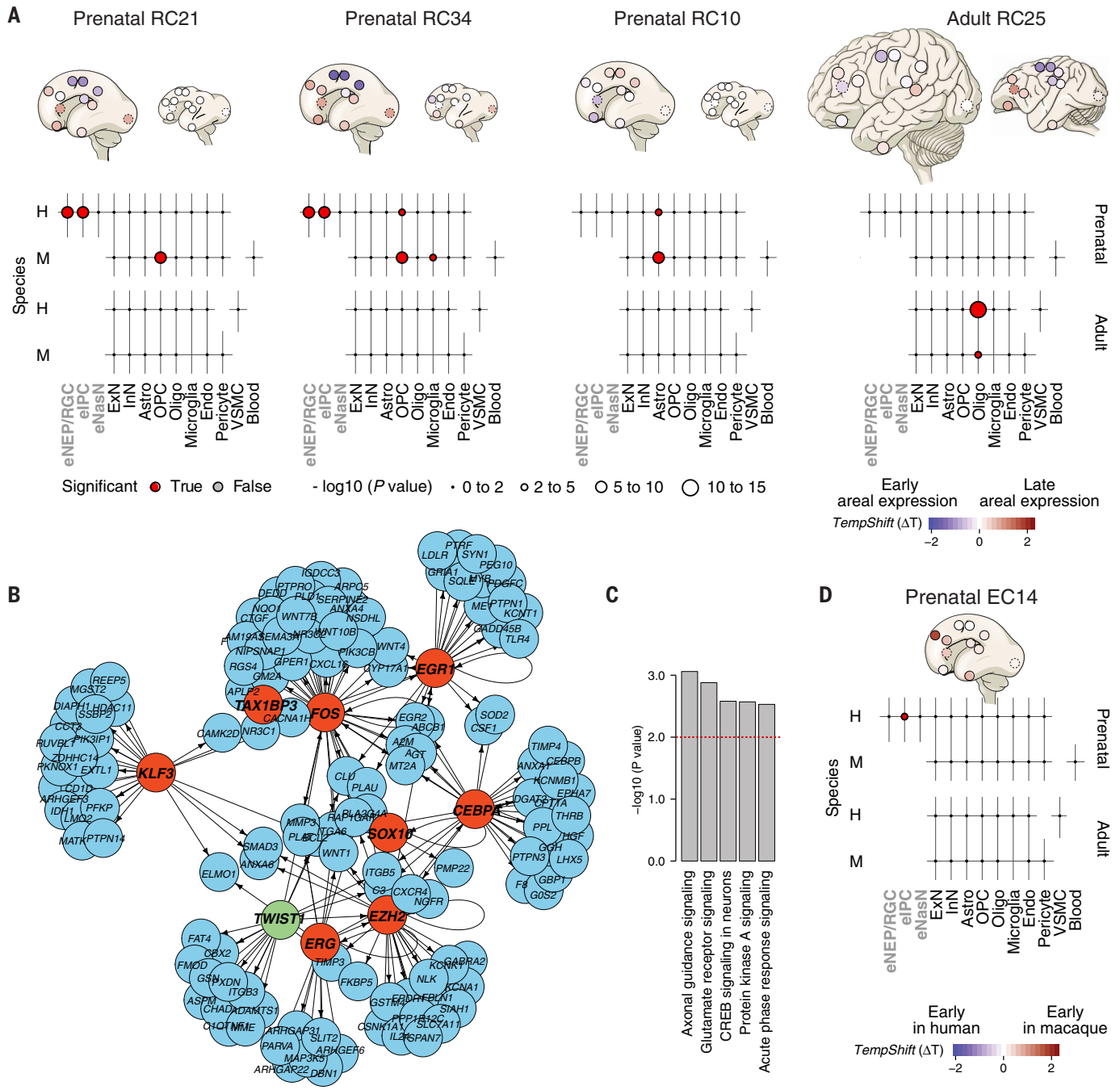


Fig. 6. Heterochronic expression of regional and interspecies gene clusters. (**A**) Clusters of genes exhibiting species-distinct regional heterochronic expression patterns in human and macaque brains at various prenatal periods and adulthood. The timing of expression of genes in the cluster is represented by a color scale (blue, earlier expression; red, later expression). Prenatal heterochronic regional clusters RC21 and RC34 show earlier expression in human prenatal frontoparietal perisylvian neocortical areas (M1C, S1C, and IPC) and enrichment in neural progenitors. RC10 is composed of genes with earlier expression in the human prenatal prefrontal cortex and enrichment in astrocytes. These observed regional expression patterns are not present in the macaque prenatal NCX. Adult heterochronic cluster RC25 shows earlier expression in primary areas of the macaque cortex and enrichment for genes associated with oligodendrocytes. (**B**) A network of 139 interspecies heterochronic genes (blue) is enriched for targets of putative upstream transcriptional regulators that include those encoded by eight genes of the same network (red) and *TWIST1* (green), a transcription factor with interspecies heterotopic expression (fig. S34). Arrows indicate direction of regulation. (**C**) Top five canonical pathways enriched among interspecies heterochronic genes in at least one neocortical area. The dashed red line corresponds to $P = 0.01$. (**D**) Cluster EC14 shows interspecies heterochronic expression, exhibits a delayed expression specifically in the human prenatal prefrontal cortex, and is enriched for genes selectively expressed by intermediate progenitor cells (IPC).

for 3′ UTR regions ($P = 1.81 \times 10^{-11}$) and non-UTR exonic elements from non–protein-coding genes ($P = 0.02364$); these results suggest that post-transcriptional regulation may contribute to species differences at the exon level.

Phylogenetic divergence in transcriptional heterotopic regulation

Because transcription factors can regulate the expression of multiple genes, the differential expression we observed between species in different brain regions might be mediated in part by differential expression of a relatively small number of transcription factors. To assess this possibility, we searched for transcription factor binding sites (TFBSs) that were enriched in the annotated promoters of interspecies differentially expressed genes for each brain region and developmental stage in our analysis (*32*). We found that the binding sites for 86 transcription factors were enriched among interspecies differentially expressed genes; 7 of these 86 transcription factors were differentially expressed between humans and macaques (table S14). *RUNX2* was differentially expressed between humans and macaques in the prenatal HIP, *PAX7* in the early postnatal AMY, *STAT6* in the prenatal NCX, *STAT4* in the early postnatal and adult NCX, *SNAI2* in the adult CBC, and *EWSR1* and *NEUROD1* in the adult NCX. Although these enriched motifs were found in only a relatively small proportion of the promoters of the interspecies differentially expressed genes (table S15), expression changes of almost 30% of the differentially expressed genes in the NCX can be explained solely by the transcription factors *STAT4*, *EWSR1*, and *NEUROD1*, which have been previously implicated in neuronal development (*55*) and brain disorders (*56*, *57*). This suggests that species differences in the expression levels of influential transcription factors could be phenotypically relevant.

To substantiate the possibility that these transcription factors might regulate interspecies differences in gene expression, we next conducted an independent analysis that integrated epigenomic data. We used previously published data on macaque-human differential regulatory elements (active promoters and enhancers) in several regions of adult brains (*58*). Using region-matched (i.e., NCX, STR, MD, and CBC) aspects of this dataset, we performed TFBS enrichments for the regions defined as up-regulated in humans as well as those down-regulated in humans relative to macaques (*32*) (tables S16 to S18). As before, we then compared TFBSs enriched among regulatory elements differentially detected in humans and macaques with the transcription factors differentially expressed in a given area or region between species. We observed a higher number of differentially expressed transcription factors associated with binding sites selective for epigenetic loci down-regulated in humans (17, 6, 6, and 1 for NCX, CBC, MD, and STR, respectively) than for loci up-regulated in humans (3, 1, and 1 for NCX, CBC, and MD, respectively). Moreover, 86% of promoters associated with interspecies differentially expressed genes in the NCX contained TFBSs for transcription factors that were differentially expressed between species in the NCX. The same was true for 33% of all differentially expressed genes retrieved from the CBC, 29% for the differentially expressed genes in the MD, and 8.5% of the differentially expressed genes in the STR.

Analysis of epigenomic data (*58*) in matched brain regions and developmental stages showed that all TFBSs enriched in differentially expressed genes were also found to be enriched in differential regulatory elements. The good agreement between the two independent datasets supports the regulatory relevance of these differentially expressed TFBSs in driving the expression changes of other differentially expressed genes.

Diversity and cell type specificity of species differences

To explore whether cell type–specific transcriptomic changes account for the interspecies divergence observed at the tissue level, we tested the enrichment of human up-regulated genes in human single cells and human down-regulated genes in macaque single cells. Furthermore, we used prenatal scRNA-seq data for prenatal differentially expressed genes and adult snRNA-seq data for the early postnatal and adulthood periods (Fig. 4, A and B, and fig. S33). In all prenatal neocortical areas, human up-regulated genes were enriched in neural progenitors, indicating that the human NCX may possess more neural progenitors at matched time points relative to macaque counterparts, although we cannot completely exclude the possibility that a lack of macaque samples matching human early fetal samples (Fig. 1, A and B) might contribute to this observation, despite the efforts we made to minimize the effects of sampling bias between species by fitting a Gaussian-process model. In contrast, macaque up-regulated genes were enriched in multiple subtypes of excitatory and inhibitory neurons in all neocortical areas (Fig. 4A). Interestingly, a specific subtype of excitatory neurons (i.e., ExN2) was enriched for the macaque up-regulated genes only in prefrontal areas. In the postnatal and adult NCX, human up-regulated genes were enriched in a single population of likely upper-layer excitatory neurons (ExN2b), which was not described in a recent snRNA-seq study of the adult human NCX (*59*). Conversely, postnatally up-regulated macaque genes were enriched in multiple subtypes of excitatory neurons (Fig. 4B). Interspecies differentially expressed genes in non-neocortical brain regions of the prenatal brain were also enriched in specific cell types (fig. S33). For example, genes displaying interspecies differential expression in HIP and CBC were enriched in a population of oligodendrocyte progenitor cells (OPCs) and external granular layer transition to granule neuron (EGL-TransGraN) cells, respectively. Furthermore, genes showing interspecies differential expression in HIP, AMY, STR, and CBC were enriched in a population of microglia (fig. S34).

By integrating our single-cell datasets with a tissue-level transcriptomic dataset of adult human, chimpanzee, and macaque brains (*34*), we identified the cell type enrichment of several genes showing human-specific up- or down-regulation in NCX or all brain regions relative to chimpanzees and macaques. For example, *CD38* was found to be down-regulated in all human brain regions and enriched in astrocytes (Fig. 4C). This gene encodes a glycoprotein that is important in the regulation of intracellular calcium, and its deletion leads to impaired development of astrocytes and oligodendrocytes in mice (*60*). *CLUL1*, a gene reported to be specifically expressed in cone photoreceptor cells (*61*), showed human-specific up-regulation in all brain regions and was enriched in oligodendrocytes and astrocytes. *TWIST1* exhibited human-specific down-regulation in all neocortical areas postnatally and was enriched in upper-layer excitatory neurons (Fig. 4C). Conversely, *PKD2L1* is up-regulated in NCX postnatally and was enriched in putative deep-layer excitatory neurons (Fig. 4C). *MET* exhibited human-specific up-regulation in the prefrontal cortex and STR postnatally and was enriched in upper-layer excitatory neurons (Fig. 4C).

Shared and divergent transcriptomic features of homologous cell types

To test whether the observed differential expression between humans and macaques was due to differences in cell type composition or due to transcriptomic differences between homologous cell types, we performed a comparative analysis between human and macaque cell types of prenatal and adult dorsolateral prefrontal cortices. The correlation between human and macaque cell types showed that all human cell types had a close homolog in macaques, and vice versa (Fig. 5, A and B). Nonetheless, we identified genes showing interspecies differential expression in homologous cell types (Fig. 5C). To avoid biases inherent to high variation in scRNA-seq or snRNA-seq, we filtered out genes that did not display differential expression between species at the tissue level and only included genes that exhibited enrichment in cell types where they showed interspecies differential expression [preferential expression measure > 0.3 (*32*)].

We identified 14 differentially expressed genes in prenatal development and 41 differentially expressed genes in adulthood (Fig. 5C). For example, *TRIM54*, which encodes a protein implicated in axonal growth (*62*), was down-regulated in human prenatal neocortical excitatory neurons (Fig. 5C). *VW2CL*, which encodes a protein associated with α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA)–type glutamate receptors (*63*), was down-regulated in prenatal human neocortical interneurons. *SLC17A8* (aka *VGLUT3*), which encodes vesicular glutamate transporter 3, is up-regulated in human postnatal somatostatin-positive interneurons (InN8). Overall, we found that human DFC cell types showed high correlation with macaque DFC cell types and that only a small set of genes displayed differential expression between these homologous cell types (Fig. 5C). Thus, the

interspecies differences identified at the tissue level are likely to result from variations in cellular diversity, abundance, and, to a lesser extent, transcriptional divergence between cell types.

Heterochronic changes in human and macaque brain transcriptomes

The observed heterotopic differences may result, in part, from changes in the timing of gene expression, or heterochrony. To identify such heterochronic differences, we created a Gaussian process–based model [TempShift (*32*)] and applied this model independently to human and macaque gene expression datasets. To maintain consistency with earlier analyses, we focused our analysis on 11 neocortical areas, which had similar transcriptomic signatures relative to other brain regions [see (*33*)]. We identified genes with interregional temporal differences within neocortical areas of each species and aggregated them into 36 regional clusters (RCs; fig. S35 and table S19). For both human and macaque brains, analysis of all heterochronic genes revealed greater interareal differences during prenatal periods than at early postnatal or adult ages (fig. S36). In addition, although we observed differences in interareal heterochrony between the early postnatal phase and the adult phase in humans, we did not observe these differences in macaques (fig. S36). This suggests that interregional synchrony in macaques precedes that in age-matched humans, possibly reflecting the protracted development of the human brain during childhood and the earlier plateauing of myelination-associated processes in macaque postnatal development (Fig. 1C and fig. S19).

Analysis of the regional clusters revealed further insights into shared and species-distinct aspects of neurodevelopment. For example, we identified five regional clusters (RC4, 21, 26, 29, and 34) enriched for genes expressed selectively by neural progenitors that exhibited temporal differences between human neocortical areas (fig. S35). Each of these clusters exhibited a gradient whereby a decrease in expression in central regions of the prenatal NCX preceded a decrease at the anterior and posterior poles, suggesting increased progenitor populations or a prolonged neurogenic period in the prefrontal cortex as well as superior temporal cortex (STC), ITC, and V1C. However, although we observed similar temporal gradients in macaques for RC4, 26, and 29, neither RC21 nor RC34—the modules exhibiting the sharpest delay in the posterior NCX—exhibited a similar central-to-polar gradient in macaques (Fig. 6A). Conversely, RC10 and RC12 exhibited an inverse gradient in humans, with decreased expression in the prefrontal NCX, STC, ITC, and V1C preceding a decrease in the central cortex. These modules, which are enriched in astrocytes, did not exhibit a similar gradient in macaques (Fig. 6A and fig. S35). This indicates that even though the transcriptomic signature associated with astrogliogenesis showed a global synchronicity between species (Fig. 1C and fig. S19), a smaller group of genes enriched in astrocytes displayed heterochrony between species.

Despite the global enrichment of heterochronic genes in prenatal development (fig. S36), we also identified clusters exhibiting higher interregional differences in postnatal development and adulthood. One example is RC25, a cluster enriched for oligodendrocyte markers that exhibited a pattern of early expression in primary motor and somatosensory areas in the macaque NCX but not the human NCX (Fig. 6A). This finding corroborates myelination-related regional asynchrony (because primary areas myelinate earlier) as well as interspecies heterochrony in oligodendrocyte maturation and myelination-associated processes. Reflective of the cup-shaped pattern of regional variation in global development, the regional clusters also suggest the asynchronous maturation of prenatal areas, a gradual synchronization during early postnatal development in both species, and additional postnatal and adult differences driven in part by myelination.

We next applied TempShift to identify genes exhibiting interspecies heterochronic divergence. Among 11 neocortical areas, we identified approximately 3.9% of coding and noncoding mRNA genes (1100 of 27,932 analyzed orthologous genes) exhibiting interspecies heterochronic expression in at least one neocortical area. We then used Ingenuity Pathway Analysis (Qiagen) to assess upstream transcriptional regulation of heterochronic genes. We found that the differential expression of 139 interspecies heterochronic genes could be explained by as few as eight coregulated heterochronic transcriptional regulators (Fig. 6B) (*32*), plus one transcription factor with heterotopic expression (down-regulated in the postnatal human NCX) between species, *TWIST1* (fig. S37). A majority (90 of 139) of these putative target genes of the nine transcriptional regulators exhibited accelerated expression in the human NCX. As mentioned above, humans exhibit an accelerated heterochronic pattern for the synaptogenesis transcriptomic signature; the presence of *FOS*, a neuronal activity–regulated gene, as one of the hubs of this transcriptional network indicates that this accelerated synaptogenesis likely drives the accelerated expression of several genes in the human NCX. Furthermore, an ontological analysis of the genes with heterochronic expression revealed an enrichment for functional categories such as "axonal guidance signaling," "glutamate receptor signaling," and "CREB signaling in neurons" (Fig. 6C), which suggests that heterochronic processes include molecular pathways related to axon guidance and synaptic activity.

We next identified 15 evolutionary clusters (ECs) on the basis of the 1100 heterochronic genes displaying interspecies neocortical heterochronic expression patterns (table S20). Among the evolutionary clusters, EC14 exhibited a delayed expression in the human dorsolateral prefrontal cortex and was enriched for intermediate progenitor cell (IPC) markers (Fig. 6D and fig. S38), in agreement with the progenitor cell population differences we observed previously in the prefrontal cortex, indicating that this neocortical prefrontal area likely has a protracted neurogenesis relative to macaques. Similarly, the species-distinct maturation gradients of neural progenitors, astrocytes, and oligodendrocytes also support observations we made concerning interspecies heterotopy. These results were supported by selective validation of the expression profiles of heterochronic genes; using droplet digital polymerase chain reaction, we selected five genes with different developmental profiles across regions and species (figs. S39 to S43), which enabled us to confirm the expression profiles of these genes as well as to ensure that our observations were not the result of biases introduced by TranscriptomeAge.

Species difference in spatiotemporal expression of disease genes

Next, we investigated whether genes associated with risk for neuropsychiatric disorders exhibited differences in their spatiotemporal expression between humans and macaques. We focused our analysis on genes linked to autism spectrum disorders (ASD) and other neurodevelopmental disorders (NDD), attention deficit hyperactivity disorder (ADHD), schizophrenia (SCZ), bipolar disorder (BD), major depressive disorder (MDD), Alzheimer's disease (AD), and Parkinson's disease (PD) in previous genetic studies or through our integrative analysis from the accompanying study (*33*) (table S21). We next sought to determine whether the expression of genes associated with these neuropsychiatric disorders were enriched in any particular developmental phase. Consistent with previous studies associating the midfetal time frame with specific high-confidence ASD (hcASD) genes (*64*), we found that a larger group of hcASD genes were more highly expressed in the prenatal brains than in the early postnatal and adult brains in both species (fig. S44). In contrast, AD-related genes were more highly expressed in the early postnatal and adult brains than in the prenatal brains in both species (fig. S44). Other groups of disease-related genes did not show any obvious global difference across development. We identified genes with heterochronic or heterotopic expression between the two species that are associated with ASD (6 and 0, respectively), non-hcASD NDD (56 and 14, respectively), and SCZ (45 and 14, respectively) (Fig. 7). This finding potentially suggests the involvement of species-specific aspects in the etiology of ASD, NDD, and SCZ. Unsupervised hierarchical clustering of SCZ-associated genes with heterotopic expression yielded five obvious spatiotemporal clusters, three of which exhibited species differences exclusively during prenatal development (fig. S44). NDD-associated genes with heterotopic expression did not yield any obvious spatiotemporal clusters. Of the prenatal clusters, cluster 1 showed enrichment in the prefrontal cortex, cluster 3 in the temporal cortex, and cluster 2 in both the frontal and temporal cortices, in humans; in macaques, cluster 4 displayed an enrichment in the postnatal and adult frontal cortex, and cluster 5 exhibited a similar enrichment in the adult prefrontal cortex (Fig. 7D).

Further analysis revealed that the ASD-associated genes *SHANK2* and *SHANK3*, which

encode synaptic scaffolding proteins at the postsynaptic density of excitatory glutamatergic synapses, exhibited earlier expression in the macaque NCX and other brain regions relative to humans (Fig. 7B). Commensurate with a role for these proteins in neural circuit development, and in agreement with analyses suggesting the involvement of neocortical projection neurons in the etiology of ASD, these two genes also became progressively more expressed across prenatal ages in both humans and macaques (fig. S45). SCZ-associated genes displaying interspecies heterochrony included *GRIA1*, a glutamate ionotropic receptor AMPA-type subunit that has different expression trajectories in MFC and OFC relative to other neocortical areas, and that is expressed earlier in human VFC, M1C, S1C, IPC, and STC (Fig. 7B and fig. S45).

These evolutionary changes in the spatiotemporal expression of certain disease-associated genes might therefore imply transcriptional underpinnings for potential human-specific aspects of neuropsychiatric disorders. For example, the presence of human-distinct heterochrony in synapse-related proteins associated with ASD, coupled with the lack of obvious heterotopic expression in hcASD genes, may suggest that conserved neurodevelopmental programs common to primate species are uniquely shifted temporally in some areas in the human brain, potentially implicating key developmental periods, places, and cell types involved in disease etiology. Similarly, the heterochronic and heterotopic changes we associated with SCZ—in particular, those affecting the prenatal prefrontal and temporal cortices—may be involved in human-specific aspect of disease etiology.

Given the importance of UTRs and other noncoding regions in the regulation of gene expression as well as disease, we next explored differences in exon usage between species in genes associated with neuropsychiatric disorders. We observed that 413 genes with differentially expressed exonic elements were linked to the studied diseases. Moreover, we detected 35 disease genes showing differentially used exonic elements with predicted binding sites (*65*) for microRNAs (miRNAs) independently associated with central nervous system diseases (*66*) (table S22). Several of these genes (e.g., *GRIN2B*, *BCL11B*, and *NKPD1*) were potentially targeted by a large number of disease-associated miRNAs (fig. S46), and gene-miRNA interactions have already been experimentally validated for 11 of the 35 genes we identified, according to miRTarBase (*67*) (table S23). For example, we detected differential exon usage of *BCL11B*, a gene involved in the development of medium spiny neurons (*68*), between humans and macaques in the adult STR (fig. S46). However, although *BCL11B* shows lower expression in the human STR than in the macaque STR, the exonic element containing the 3′UTR of *BCL11B* was itself not differentially expressed.

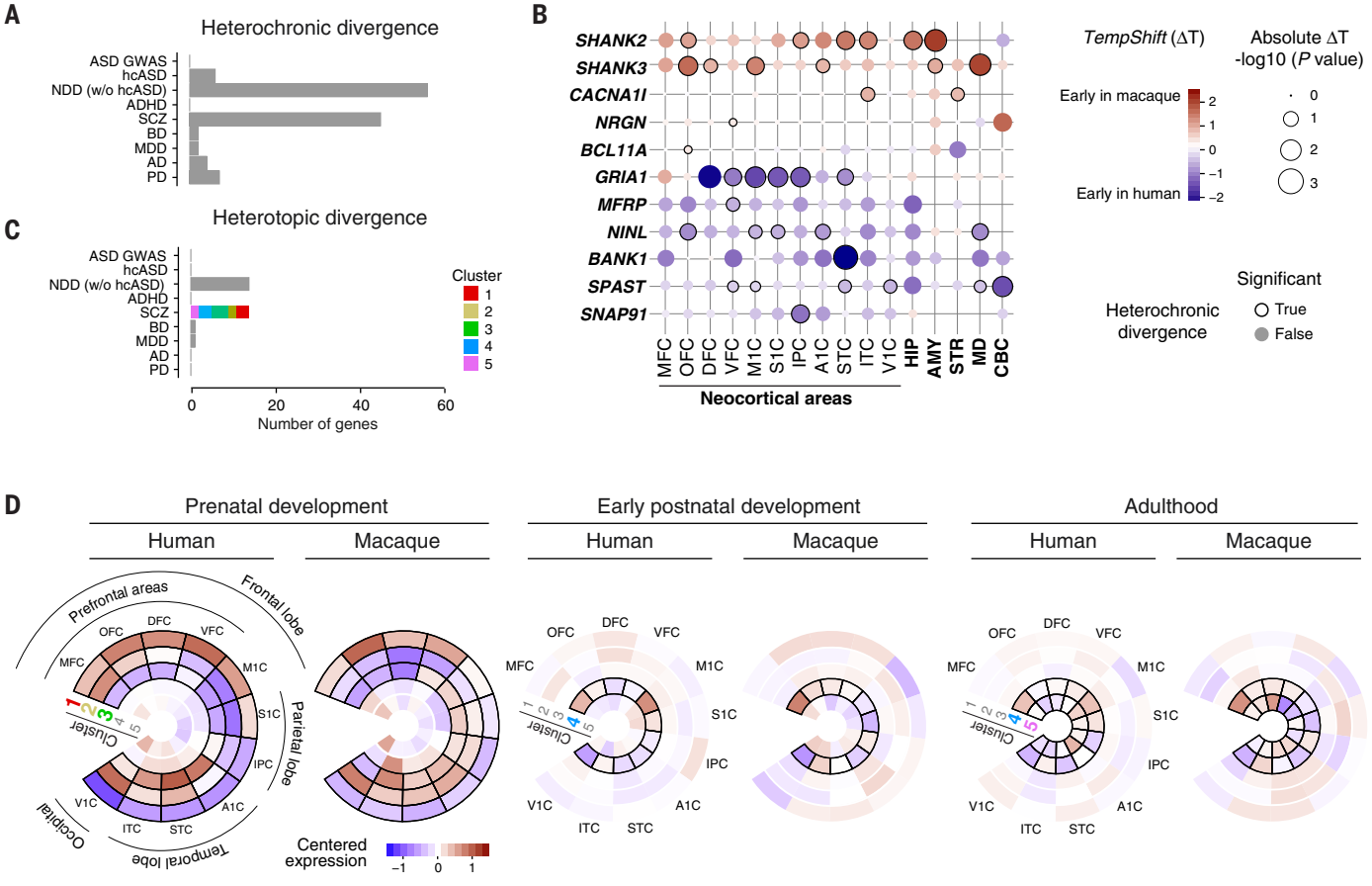


Fig. 7. Heterotopic and/or heterochronic expression of disease-associated genes between humans and macaques. (**A**) Bar plot depicting the number of genes associated with autism spectrum disorder (ASD; hc, high confidence), neurodevelopmental disorders (NDD), attention deficit hyperactivity disorder (ADHD), schizophrenia (SCZ), bipolar disorder (BD), major depressive disorder (MDD), Alzheimer's disease (AD), and Parkinson's disease (PD) that display heterochronic divergence between humans and macaques. (**B**) Bubble matrix showing the heterochronic expression of ASD- and SCZ-associated genes. Blue represents earlier expression in humans; red represents earlier expression in macaques. (**C**) Bar plot depicting the number of genes associated with neuropsychiatric disorders that exhibit heterotopic divergence between humans and macaques. The 14 SCZ-associated genes that displayed heterotopy are grouped into five clusters on the basis of their spatiotemporal expression profiles (fig. S41). (**D**) Donut plots exhibiting the centered expression of the five SCZ-associated heterotopic clusters in prenatal development, early postnatal development, and adulthood. Clusters that are not significantly divergent between species in each period are gray and do not have a black border. Red indicates high expression; blue indicates low expression.

This observation suggests that overexpression in macaques is associated with an alternative isoform containing a shorter 3′UTR region. This shorter 3′UTR lacks predicted binding sites for various miRNAs, including members of the brain-specific miR-219 family, which have been experimentally shown to interact with *BCL11B* mRNA (*69*). Together, these findings indicate that certain genes associated with neuropsychiatric disorders exhibit changes in the timing of their expression, location, and splicing pattern between human and NHP brains, and thus may lead to species differences in disease pathogenesis.

Discussion

In this study, we present a comprehensive spatiotemporal transcriptomic brain dataset of the macaque brain. Our integrative and comparative analysis involving complementary humans and adult chimpanzees (*33*, *34*) revealed similarities and differences in the spatiotemporal transcriptomic architecture of the brain and the progression of major neurodevelopmental processes between the two species. For example, we have identified shared and divergent transcriptomic features among homologous brain regions and cell types. We found transcriptomic evidence suggesting that human childhood is especially protracted relative to that of macaques. It has long been recognized that the development of the human brain is prolonged relative to that of other NHPs, and that this slower rate of maturation expands the period of neural plasticity and capacity for learning activities, memory, and complex sensory perception, all processes necessary for higher-order cognition (*1*–*4*, *14*, *28*). We also found that, relative to macaques, the early periods of human fetal neurodevelopment are transcriptomically distinct and protracted. A similar observation of early neurodevelopmental protraction was recently observed in vitro, in neural progenitors derived from pluripotent cells of human and NHPs (*70*). However, we also identified cases of neoteny in macaques, such as the protracted postnatal expression of DCX in the hippocampus, likely reflecting differences in neurogenesis between the two species, as recently shown (*49*).

We found that global patterns of spatiotemporal transcriptomic dynamics were conserved between humans and macaques, and that they display a highly convergent cup-like shape. The sharpest decrease in interregional differences occurs during late fetal ages and before birth; this is likely a consequence of reorganizational processes at this developmental period rather than extrinsic influences due to birth and subsequent events (i.e., respiratory activity or other developmentally novel stimuli). Interestingly, after this transitional period, diversification of neocortical areas appears to be driven mainly by differences between primary and association areas. In addition to these largely conserved broad developmental patterns of interregional differences, we identified numerous genes and gene modules with human-distinct heterochronic or heterotopic expression. These patterns involved brain regions such as the developing prefrontal areas, which are central to the evolution of distinctly human aspects of cognition and behavior (*19*–*21*). Surprisingly, we also found that developmental phases exhibiting high levels of interregional differences (i.e., early to midfetal periods and young adulthood) were also less conserved between the two species. The coincident convergence of the ontogenetic and phylogenetic cups during the late fetal period and infancy is strikingly distinct from the previously reported phylogenetic transcriptomic hourglass-like pattern that occurs during the embryonic organogenetic period (*71*, *72*).

Genes with divergent spatiotemporal expression patterns included those previously linked to ASD, SCZ, and NDD. These species differences in the expression of disease-associated genes linked to synapse formation, neuronal development, and function, as well as regional and species differences in synaptogenesis and myelination, might have implications for the overall development of neural circuitry and consequently human cognition and behavior. These observations are possibly relevant for recent NHP models of neuropsychiatric disease, such as the *SHANK3*-deficient macaque model (*73*), which might therefore not be capable of fully capturing human-distinct aspects of *SHANK3* regulation during neurodevelopment.

Our study reveals insights into the evolution of gene expression in the developing human brain. Future work on the development patterns and the functional validation of the genes we report to have heterotopic and/or expression patterns between humans and macaques will likely shed some light on potentially human-specific underpinnings of certain neuropsychiatric disorders.

Materials and methods

Sixteen regions of the macaque brain spanning from early prenatal to adulthood were dissected using the same standardized protocol used for human specimens and described in the accompanying study by Li *et al.* [(*33*); see also (*32*)]. The macaque brain regions and developmental time points matched human brain regions and time points analyzed in (*33*). The sampled homologous brain regions were identified using anatomical landmarks provided in the macaque brain atlas (*74*). An overview of dissected brain regions is provided in fig. S1. The Translating Time model (*38*) was used to identify equivalent time points between macaque and human prenatal development. The list of macaque brains used in this study and relevant metadata are provided in tables S1 and S2. Macaque studies were carried out in accordance with a protocol approved by Yale University's Committee on Animal Research and NIH guidelines.

We performed tissue-level RNA extraction and sequencing of all 16 regions, scRNA-seq of dorsolateral prefrontal cortex (DFC), hippocampus (HIP), amygdala (AMY), striatum (STR), mediodorsal nucleus of the thalamus (MD), and cerebellar cortex (CBC) of midfetal macaques, and snRNA-seq of DFC of adult macaques. Single cell/nucleus sample processing was done with 10X Genomics and sequencing was done with Illumina platforms.

For tissue-level analysis, we generated annotations of human-macaque orthologs using the XSAnno pipeline, and matched the developmental age of human and macaque samples based on their respective transcriptome using our algorithm TranscriptomeAge. We also developed TempShift, a method based on a Gaussian-process model, to reveal the interregional differences, interspecies divergence, and genes with heterotopic and heterochronic expression. We also queried differentially expressed genes for enrichment in transcription factor binding sites using findMotifs.pl, and analyzed interspecies differential exon usage using the R package DEXSeq.

The single cell/nucleus data were first analyzed by cellranger for decoding, alignment, quality filtering, and UMI counting. After that, data were further analyzed with Seurat according to its guidelines, and cell types were clustered for classification with SpecScore.R. To perform direct comparisons between human and macaque at the single-cell level, we focused on the homologous genes between these species and aligned monkey and human cells together to further analyze interspecies divergence of homologous cell types (fig. S47). We used MetageneBicorPlot function to examine the correlation of neuronal and glial cell subtypes, and we employed correlation analysis to detect the correspondence of excitatory neuron and interneuron subtypes. Finally, we did functional enrichment of disease-associated genes in both tissue-level and single-cell datasets.

REFERENCES AND NOTES

1. M. F. A. Montagu, Time, Morphology, and Neoteny in the Evolution of Man. *Am. Anthropol.* **57**, 13–27 (1955). doi: 10.1525/aa.1955.57.1.02a00030
2. J. Dobbing, J. Sands, Quantitative growth and development of human brain. *Arch. Dis. Child.* **48**, 757–767 (1973). doi: 10.1136/adc.48.10.757; pmid: 4796010
3. H. F. Prechtl, New perspectives in early human development. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **21**, 347–355 (1986). doi: 10.1016/0028-2243(86)90014-6; pmid: 3721046
4. B. Bogin, Evolutionary perspective on human growth. *Annu. Rev. Anthropol.* **28**, 109–153 (1999). doi: 10.1146/annurev.anthro.28.1.109; pmid: 12295621
5. A. H. Schultz, in *Human Growth*, J. M. Tanner, Ed. (Oxford, 1960), pp. 1–20.
6. J. C. Silbereis, S. Pochareddy, Y. Zhu, M. Li, N. Sestan, The Cellular and Molecular Landscapes of the Developing Human Central Nervous System. *Neuron* **89**, 248–268 (2016). doi: 10.1016/j.neuron.2015.12.008; pmid: 26796689
7. R. S. Hill, C. A. Walsh, Molecular insights into human brain evolution. *Nature* **437**, 64–67 (2005). doi: 10.1038/nature04103; pmid: 16136130
8. J. L. Fish, C. Dehay, H. Kennedy, W. B. Huttner, Making bigger brains—the evolution of neural-progenitor-cell division. *J. Cell Sci.* **121**, 2783–2793 (2008). doi: 10.1242/jcs.023465; pmid: 18716282
9. D. H. Geschwind, G. Konopka, Neuroscience in the era of functional genomics and systems biology. *Nature* **461**, 908–915 (2009). doi: 10.1038/nature08537; pmid: 19829370
10. J. H. Lui, D. V. Hansen, A. R. Kriegstein, Development and evolution of the human neocortex. *Cell* **146**, 18–36 (2011). doi: 10.1016/j.cell.2011.06.030; pmid: 21729779
11. S. A. McCarroll, S. E. Hyman, Progress in the genetics of polygenic brain disorders: Significant new challenges for neurobiology. *Neuron* **80**, 578–587 (2013). doi: 10.1016/j.neuron.2013.10.046; pmid: 24183011
12. E. S. Lein, T. G. Belgard, M. Hawrylycz, Z. Molnár, Transcriptomic Perspectives on Neocortical Structure,

Development, Evolution, and Disease. *Annu. Rev. Neurosci.* **40**, 629–652 (2017). doi: 10.1146/annurev-neuro-070815-013858; pmid: 28661727
13. G. E. Hardingham, P. Pruunsild, M. E. Greenberg, H. Bading, Lineage divergence of activity-driven transcription and evolution of cognitive ability. *Nat. Rev. Neurosci.* **19**, 9–15 (2018). doi: 10.1038/nrn.2017.138; pmid: 29167525
14. T. Paus, M. Keshavan, J. N. Giedd, Why do many psychiatric disorders emerge during adolescence? *Nat. Rev. Neurosci.* **9**, 947–957 (2008). doi: 10.1038/nrn2513; pmid: 19002191
15. B. Y. Liao, J. Zhang, Null mutations in human and mouse orthologs frequently result in different phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6987–6992 (2008). doi: 10.1073/pnas.0800387105; pmid: 18458337
16. T. Barak *et al*., Recessive LAMC3 mutations cause malformations of occipital cortical development. *Nat. Genet.* **43**, 590–594 (2011). doi: 10.1038/ng.836; pmid: 21572413
17. M. O'Bleness, V. B. Searles, A. Varki, P. Gagneux, J. M. Sikela, Evolution of genetic and genomic features unique to the human lineage. *Nat. Rev. Genet.* **13**, 853–866 (2012). doi: 10.1038/nrg3336; pmid: 23154808
18. K. Mirnics, J. Pevsner, Progress in the use of microarray technology to study the neurobiology of disease. *Nat. Neurosci.* **7**, 434–439 (2004). doi: 10.1038/nn1230; pmid: 15114354
19. T. M. Preuss, Human brain evolution: From gene discovery to phenotype discovery. *Proc. Natl. Acad. Sci. U.S.A.* **109** (suppl. 1), 10709–10716 (2012). doi: 10.1073/pnas.1201894109; pmid: 22723367
20. C. C. Sherwood, A. L. Bauernfeind, S. Bianchi, M. A. Raghanti, P. R. Hof, Human brain evolution writ large and small. *Prog. Brain Res.* **195**, 237–254 (2012). doi: 10.1016/B978-0-444-53860-4.00011-8; pmid: 22230630
21. A. M. M. Sousa, K. A. Meyer, G. Santpere, F. O. Gulden, N. Sestan, Evolution of the Human Nervous System Function, Structure, and Development. *Cell* **170**, 226–247 (2017). doi: 10.1016/j.cell.2017.06.036; pmid: 28708995
22. H. B. van der Worp *et al*., Can animal models of disease reliably inform human studies? *PLOS Med.* **7**, e1000245 (2010). doi: 10.1371/journal.pmed.1000245; pmid: 20361020
23. T. Marques-Bonet, O. A. Ryder, E. E. Eichler, Sequencing primate genomes: What have we learned? *Annu. Rev. Genomics Hum. Genet.* **10**, 355–386 (2009). doi: 10.1146/annurev.genom.9.081307.164420; pmid: 19630567
24. J. C. Izpisua Belmonte *et al*., Brains, genes, and primates. *Neuron* **86**, 617–631 (2015). doi: 10.1016/j.neuron.2015.03.021; pmid: 25950631
25. C. G. Jennings *et al*., Opportunities and challenges in modeling human brain disorders in transgenic primates. *Nat. Neurosci.* **19**, 1123–1130 (2016). doi: 10.1038/nn.4362; pmid: 27571191
26. K. Sato, E. Sasaki, Genetic engineering in nonhuman primates for human disease modeling. *J. Hum. Genet.* **63**, 125–131 (2018). doi: 10.1038/s10038-017-0351-5; pmid: 29203824
27. M. B. Johnson *et al*., Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009). doi: 10.1016/j.neuron.2009.03.027; pmid: 19477152
28. M. Somel *et al*., Transcriptional neoteny in the human brain. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5743–5748 (2009). doi: 10.1073/pnas.0900544106; pmid: 19307592
29. H. J. Kang *et al*., Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011). doi: 10.1038/nature10523; pmid: 22031440
30. M. Pletikos *et al*., Temporal specification and bilaterality of human neocortical topographic gene expression. *Neuron* **81**, 321–332 (2014). doi: 10.1016/j.neuron.2013.11.018; pmid: 24373884
31. X. Liu *et al*., Extension of cortical synaptic development distinguishes humans from chimpanzees and macaques. *Genome Res.* **22**, 611–622 (2012). doi: 10.1101/gr.127324.111; pmid: 22300767
32. See supplementary materials.
33. M. Li *et al*., Integrative functional genomic analysis of human brain development and neuropsychiatric risk. *Science* **362**, eaat7615 (2018).
34. A. M. M. Sousa *et al*., Molecular and cellular reorganization of neural circuits in the human lineage. *Science* **358**, 1027–1032 (2017). doi: 10.1126/science.aan3456; pmid: 29170230
35. S. Zhong *et al*., A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018). doi: 10.1038/nature25980; pmid: 29539641
36. J. A. Miller *et al*., Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199–206 (2014). doi: 10.1038/nature13185; pmid: 24695229
37. Y. Zhu, M. Li, A. M. Sousa, N. Sestan, XSAnno: A framework for building ortholog models in cross-species transcriptome comparisons. *BMC Genomics* **15**, 343 (2014). doi: 10.1186/1471-2164-15-343; pmid: 24884593
38. A. D. Workman, C. J. Charvet, B. Clancy, R. B. Darlington, B. L. Finlay, Modeling transformations of neurodevelopmental sequences across mammalian species. *J. Neurosci.* **33**, 7368–7383 (2013). doi: 10.1523/JNEUROSCI.5746-12.2013; pmid: 23616543
39. J. Fooden, *Systematic Review of the Rhesus Macaque, Macaca mulatta (Zimmermann, 1780)* (Field Museum of Natural History, 2000).
40. B. Clancy, R. B. Darlington, B. L. Finlay, Translating developmental time across mammalian species. *Neuroscience* **105**, 7–17 (2001). doi: 10.1016/S0306-4522(01)00171-3; pmid: 11483296
41. R. Knoth *et al*., Murine features of neurogenesis in the human hippocampus across the lifespan from 0 to 100 years. *PLOS ONE* **5**, e8809 (2010). doi: 10.1371/journal.pone.0008809; pmid: 20126454
42. P. R. Huttenlocher, A. S. Dabholkar, Regional differences in synaptogenesis in human cerebral cortex. *J. Comp. Neurol.* **387**, 167–178 (1997). doi: 10.1002/(SICI)1096-9861(19971020)387:2<167::AID-CNE1>3.0.CO;2-Z; pmid: 9336221
43. D. J. Miller *et al*., Prolonged myelination in human neocortical evolution. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16480–16485 (2012). doi: 10.1073/pnas.1117943109; pmid: 23012402
44. J. N. Giedd *et al*., Brain development during childhood and adolescence: A longitudinal MRI study. *Nat. Neurosci.* **2**, 861–863 (1999). doi: 10.1038/13158; pmid: 10491603
45. J. L. Rapoport, N. Gogtay, Brain neuroplasticity in healthy, hyperactive and psychotic children: Insights from neuroimaging. *Neuropsychopharmacology* **33**, 181–197 (2008). doi: 10.1038/sj.npp.1301553; pmid: 17851542
46. E. R. Sowell *et al*., Mapping cortical change across the human life span. *Nat. Neurosci.* **6**, 309–315 (2003). doi: 10.1038/nn1008; pmid: 12548289
47. P. Flechsig of Leipsic, Developmental (myelogenetic) localisation of the cerebral cortex in the human subject. *Lancet* **158**, 1027–1030 (1901). doi: 10.1016/S0140-6736(01)01429-5
48. P. Rakic, J. P. Bourgeois, M. F. Eckenhoff, N. Zecevic, P. S. Goldman-Rakic, Concurrent overproduction of synapses in diverse regions of the primate cerebral cortex. *Science* **232**, 232–235 (1986). doi: 10.1126/science.3952506; pmid: 3952506
49. S. F. Sorrells *et al*., Human hippocampal neurogenesis drops sharply in children to undetectable levels in adults. *Nature* **555**, 377–381 (2018). doi: 10.1038/nature25975; pmid: 29513649
50. A. Chédotal, L. J. Richards, Wiring the brain: The biology of neuronal guidance. *Cold Spring Harb. Perspect. Biol.* **2**, a001917 (2010). doi: 10.1101/cshperspect.a001917; pmid: 20463002
51. M. Rhinn, P. Dollé, Retinoic acid signalling during development. *Development* **139**, 843–858 (2012). doi: 10.1242/dev.065938; pmid: 22318625
52. A. L. Huang *et al*., The cells and logic for mammalian sour taste detection. *Nature* **442**, 934–938 (2006). doi: 10.1038/nature05084; pmid: 16929298
53. Q. Qin, Y. Xu, T. He, C. Qin, J. Xu, Normal and disease-related biological functions of Twist1 and underlying molecular mechanisms. *Cell Res.* **22**, 90–106 (2012). doi: 10.1038/cr.2011.144; pmid: 21876555
54. D. B. Campbell *et al*., A genetic variant that disrupts MET transcription is associated with autism. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 16834–16839 (2006). doi: 10.1073/pnas.0605296103; pmid: 17053076
55. A. Pataskar *et al*., NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. *EMBO J.* **35**, 24–45 (2016). doi: 10.15252/embj.201591206; pmid: 26516211
56. M. Rodriguez *et al*., STAT4- and STAT6-signaling molecules in a murine model of multiple sclerosis. *FASEB J.* **20**, 343–345 (2006). doi: 10.1096/fj.05-4650fje; pmid: 16352646
57. D. Sturm *et al*., New brain tumor entities emerge from molecular classification of CNS-PNETs. *Cell* **164**, 1060–1072 (2016). doi: 10.1016/j.cell.2016.01.015; pmid: 26919435
58. M. W. Vermunt *et al*., Epigenomic annotation of gene regulatory alterations during evolution of the primate brain. *Nat. Neurosci.* **19**, 494–503 (2016). doi: 10.1038/nn.4229; pmid: 26807951
59. B. B. Lake *et al*., Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science* **352**, 1586–1590 (2016). doi: 10.1126/science.aaf1204; pmid: 27339989
60. N. L. Rosenberg, B. L. Kotzin, M. C. Kemp, J. S. Burks, T. J. Santoro, Coronavirus SD-induced immunoregulatory disturbances in a murine model of demyelination. *Adv. Exp. Med. Biol.* **218**, 441–447 (1987). doi: 10.1007/978-1-4684-1280-2_54; pmid: 2829557
61. G. M. Sturgill *et al*., Mutation screen of the cone-specific gene, CLUL1, in 376 patients with age-related macular degeneration. *Ophthalmic Genet.* **27**, 151–155 (2006). doi: 10.1080/13816810600976871; pmid: 17148042
62. M. A. Gomez-Ferreria *et al*., Human Cep192 is required for mitotic centrosome and spindle assembly. *Curr. Biol.* **17**, 1960–1966 (2007). doi: 10.1016/j.cub.2007.10.019; pmid: 17980596
63. J. Schwenk *et al*., High-resolution proteomics unravel architecture and molecular diversity of native AMPA receptor complexes. *Neuron* **74**, 621–633 (2012). doi: 10.1016/j.neuron.2012.03.034; pmid: 22632720
64. S. J. Sanders *et al*., Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015). doi: 10.1016/j.neuron.2015.09.016; pmid: 26402605
65. V. Agarwal, G. W. Bell, J. W. Nam, D. P. Bartel, Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4**, e05005 (2015). doi: 10.7554/eLife.05005; pmid: 26267216
66. Q. Jiang *et al*., miR2Disease: A manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.* **37**, D98–D104 (2009). doi: 10.1093/nar/gkn714; pmid: 18927107
67. C. H. Chou *et al*., miRTarBase update 2018: A resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* **46**, D296–D302 (2018). doi: 10.1093/nar/gkx1067; pmid: 29126174
68. P. Arlotta, B. J. Molyneaux, D. Jabaudon, Y. Yoshida, J. D. Macklis, Ctip2 controls the differentiation of medium spiny neurons and the establishment of the cellular architecture of the striatum. *J. Neurosci.* **28**, 622–632 (2008). doi: 10.1523/JNEUROSCI.2986-07.2008; pmid: 18199763
69. C. Zhao *et al*., Gga-miR-219b targeting BCL11B suppresses proliferation, migration and invasion of Marek's disease tumor cell MSB1. *Sci. Rep.* **7**, 4247 (2017). doi: 10.1038/s41598-017-04434-w; pmid: 28652615
70. T. Otani, M. C. Marchetto, F. H. Gage, B. D. Simons, F. J. Livesey, 2D and 3D Stem Cell Models of Primate Cortical Development Identify Species-Specific Differences in Progenitor Behavior Contributing to Brain Size. *Cell Stem Cell* **18**, 467–480 (2016). doi: 10.1016/j.stem.2016.03.003; pmid: 27049876
71. T. Domazet-Lošo, D. Tautz, A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns. *Nature* **468**, 815–818 (2010). doi: 10.1038/nature09632; pmid: 21150997
72. A. T. Kalinka *et al*., Gene expression divergence recapitulates the developmental hourglass model. *Nature* **468**, 811–814 (2010). doi: 10.1038/nature09634; pmid: 21150996
73. H. Zhao *et al*., Altered neurogenesis and disrupted expression of synaptic proteins in prefrontal cortex of SHANK3-deficient non-human primate. *Cell Res.* **27**, 1293–1297 (2017). doi: 10.1038/cr.2017.95; pmid: 28741620
74. K. S. Saleem, N. Logothetis, *A Combined MRI and Histology Atlas of the Rhesus Monkey Brain in Stereotaxic Coordinates* (Academic Press, 2007).

ACKNOWLEDGMENTS

We thank M. Horn, G. Sedmak, M. Pletikos, D. Singh, G. Terwilliger, and S. Wilson for assistance with tissue acquisition and processing. We also thank A. Duque for using equipment from MacBrainResource (NIH/NIMH R01MH113257). **Funding:** Data were generated as part of the PsychENCODE Consortium, supported by NIH grants U01MH103392, U01MH103365, U01MH103346, U01MH103340, U01MH103339, R21MH109956, R21MH105881, R21MH105853, R21MH103877, R21MH102791, R01MH111721, R01MH110928, R01MH110927, R01MH110926, R01MH110921, R01MH110920, R01MH110905, R01MH109715, R01MH109677, R01MH105898, R01MH105898, R01MH094714, P50MH106934 awarded to S. Akbarian (Icahn School of Medicine at Mount Sinai), G. Crawford (Duke University), S. Dracheva (Icahn School of Medicine at Mount Sinai), P. Farnham (University of Southern California), M. Gerstein (Yale University), D. Geschwind (University of California, Los Angeles), F. Goes (Johns Hopkins University),

T. M. Hyde (Lieber Institute for Brain Development), A. Jaffe (Lieber Institute for Brain Development), J. A. Knowles (University of Southern California), C. Liu (SUNY Upstate Medical University), D. Pinto (Icahn School of Medicine at Mount Sinai), P. Roussos (Icahn School of Medicine at Mount Sinai), S. Sanders (University of California, San Francisco), P. Sklar (Icahn School of Medicine at Mount Sinai), M. State (University of California, San Francisco), P. Sullivan (University of North Carolina), F. Vaccarino (Yale University), D. Weinberger (Lieber Institute for Brain Development), S. Weissman (Yale University), K. White (University of Chicago), J. Willsey (University of California, San Francisco), P. Zandi (Johns Hopkins University), and N.S. Also supported by BFU2017-86471-P (MINECO/FEDER, UE), U01 MH106874 grant, Howard Hughes International Early Career, 3P30AG021342-16S2 (H.Z.); Obra Social "La Caixa" and Secretaria d'Universitats i Recerca and CERCA Programme del Departament d'Economia i Coneixement de la Generalitat de Catalunya (GRC 2017 SGR 880) (T.M.-B.); a Formació de Personal Investigador fellowship from Generalitat de Catalunya (FI_B00122) (P.E.-C.); La Caixa Foundation (L.F.-P.); a Juan de la Cierva fellowship (FJCI-2016-29558) from MICINN (D.J.); and NIH grants MH109904 and MH106874, the Kavli Foundation, and the James S. McDonnell Foundation. **Author contributions:** A.M.M.S. and N.S. designed the study and procured and dissected all samples; A.M.M.S. and Y.I.K. performed all tissue-level experiments and validations; T.G. and M.S. performed single-cell experiments; M.S., D.J.M., and M.Y. performed single-nucleus experiments; Y.Z. developed TranscriptomeAge and TempShift under supervision of H.Z.; Y.Z., M.L., and G.S. analyzed the data; P.E.-C., D.J., L.F.-P., and T.M.-B. performed transcription factor enrichment, differential exon usage analyses, and evolutionary conservation analyses; and Y.Z., A.M.M.S., F.O.G., and N.S. wrote the manuscript with input of the other authors. **Competing interests:** Authors have no conflict of interest. **Data and materials availability:** Data are available at NCBI BioProjects (accession number PRJNA448973) and via Synapse in psychencode.org. All algorithms, packages, and scripts are available at evolution.psychencode.org. Supplement contains additional data.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/eaat8077/suppl/DC1
Materials and Methods
Figs. S1 to S47
Tables S1 to S27
References (*75*–*98*)

6 April 2018; accepted 8 November 2018
10.1126/science.aat8077

RESEARCH ARTICLE SUMMARY

PSYCHIATRIC GENOMICS

Transcriptome and epigenome landscape of human cortical development modeled in organoids

Anahita Amiri*, Gianfilippo Coppola*, Soraya Scuderi*, Feinan Wu*, Tanmoy Roychowdhury*, Fuchen Liu, Sirisha Pochareddy, Yurae Shin, Alexias Safi, Lingyun Song, Ying Zhu, André M. M. Sousa, The PsychENCODE Consortium†, Mark Gerstein, Gregory E. Crawford, Nenad Sestan, Alexej Abyzov‡, Flora M. Vaccarino‡

INTRODUCTION: The human cerebral cortex has undergone an extraordinary increase in size and complexity during mammalian evolution. Cortical cell lineages are specified in the embryo, and genetic and epidemiological evidence implicates early cortical development in the etiology of neuropsychiatric disorders such as autism spectrum disorder (ASD), intellectual disabilities, and schizophrenia. Most of the disease-implicated genomic variants are located outside of genes, and the interpretation of noncoding mutations is lagging behind owing to limited annotation of functional elements in the noncoding genome.

RATIONALE: We set out to discover gene-regulatory elements and chart their dynamic activity during prenatal human cortical development, focusing on enhancers, which carry most of the weight upon regulation of gene expression. We longitudinally modeled human brain development using human induced pluripotent stem cell (hiPSC)–derived cortical organoids and compared organoids to isogenic fetal brain tissue.

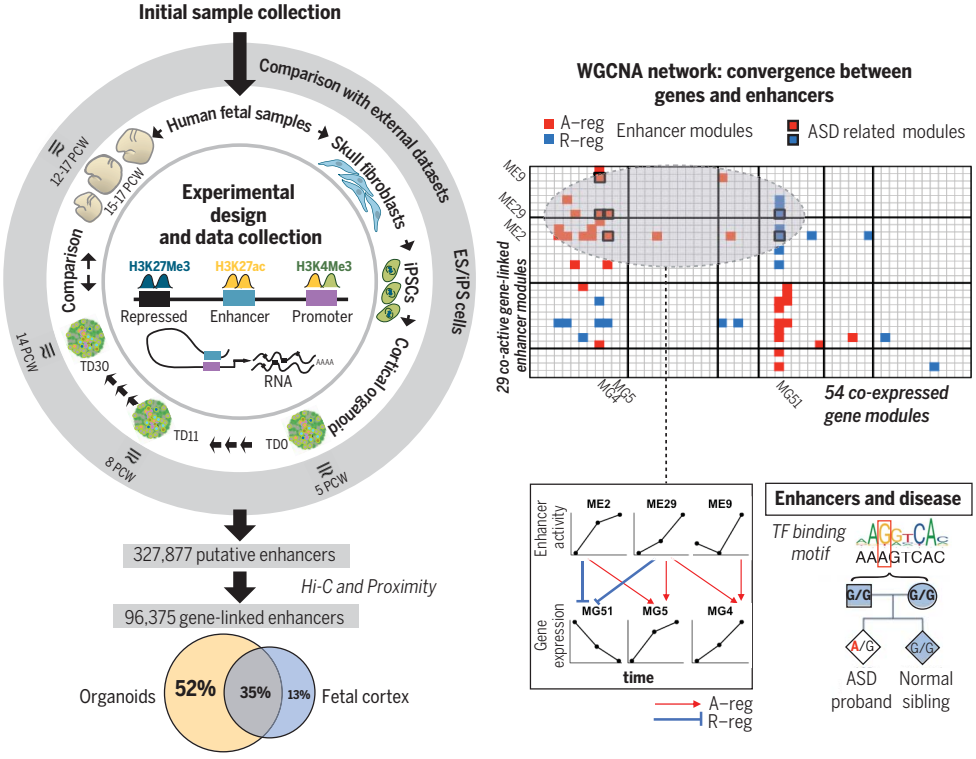
RESULTS: Fetal fibroblast–derived hiPSC lines were used to generate cortically patterned organoids and to compare oganoids' epigenome and transcriptome to that of isogenic fetal brains and external datasets. Organoids model cortical development between 5 and 16 postconception weeks, thus enabling us to study transitions from cortical stem cells to progenitors to early neurons. The greatest changes occur at the transition from stem cells to progenitors. The regulatory landscape encompasses a total set of 96,375 enhancers linked to target genes, with 49,640 enhancers being active in organoids but not in mid-fetal brain, suggesting major roles in cortical neuron specification. Enhancers that gained activity in the human lineage are active in the earliest stages of organoid development, when they target genes that regulate the growth of radial glial cells.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat6720

Parallel weighted gene coexpression network analysis (WGCNA) of transcriptome and enhancer activities defined a number of modules of coexpressed genes and coactive enhancers, following just six and four global temporal patterns that we refer to as supermodules, likely reflecting fundamental programs in embryonic and fetal brain. Correlations between gene expression and enhancer activity allowed stratifying enhancers into two categories: activating regulators (A-regs) and repressive regulators (R-regs). Several enhancer modules converged with gene modules, suggesting that coexpressed genes are regulated by enhancers with correlated patterns of activity. Furthermore, enhancers active in organoids and fetal brains were enriched for ASD de novo variants that disrupt binding sites of homeodomain, Hes1, NR4A2, Sox3, and NFIX transcription factors.

CONCLUSION: We validated hiPSC-derived cortical organoids as a suitable model system for studying gene regulation in human embryonic brain development, evolution, and disease. Our results suggest that organoids may reveal how noncoding mutations contribute to ASD etiology. ■



Summary of the study, analyses, and main results. Data were generated for iPSC-derived human telencephalic organoids and isogenic fetal cortex. Organoids modeled embryonic and early fetal cortex and show a larger repertoire of enhancers. Enhancers could be divided into activators and repressors of gene expression. We derived networks of modules and supermodules with correlated gene and enhancer activities, some of which were implicated in autism spectrum disorders (ASD).

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†The PsychENCODE Consortium authors and affiliations are listed in the supplementary materials.
‡Corresponding author: Email: abyzov.alexej@mayo.edu (A.A.); flora.vaccarino@yale.edu (F.M.V.)
Cite this article as A. Amiri *et al.*, *Science* **362**, eaat6720 (2018). DOI: 10.1126/science.aat6720

RESEARCH ARTICLE

PSYCHIATRIC GENOMICS

Transcriptome and epigenome landscape of human cortical development modeled in organoids

Anahita Amiri[1]*, Gianfilippo Coppola[1]*, Soraya Scuderi[1]*, Feinan Wu[1]*, Tanmoy Roychowdhury[2]*, Fuchen Liu[3], Sirisha Pochareddy[3], Yurae Shin[3,4], Alexias Safi[5], Lingyun Song[5], Ying Zhu[3,6], André M. M. Sousa[3], The PsychENCODE Consortium†, Mark Gerstein[7], Gregory E. Crawford[5], Nenad Sestan[3,8], Alexej Abyzov[2]‡, Flora M. Vaccarino[1,3,8]‡

Genes implicated in neuropsychiatric disorders are active in human fetal brain, yet difficult to study in a longitudinal fashion. We demonstrate that organoids from human pluripotent cells model cerebral cortical development on the molecular level before 16 weeks postconception. A multiomics analysis revealed differentially active genes and enhancers, with the greatest changes occurring at the transition from stem cells to progenitors. Networks of converging gene and enhancer modules were assembled into six and four global patterns of expression and activity across time. A pattern with progressive down-regulation was enriched with human-gained enhancers, suggesting their importance in early human brain development. A few convergent gene and enhancer modules were enriched in autism-associated genes and genomic variants in autistic children. The organoid model helps identify functional elements that may drive disease onset.

Patterning of the mammalian brain into regions of specific size and fate, demarcated by transcription factor expression and enhancer activity, is already in progress around the time the neural tube closes in the fourth postconceptional week (PCW) in humans and forestalls species-specific mechanisms of neurogenesis, connectivity, and function (*1*–*3*). A growing list of genetic and epidemiological evidence implicates early neurodevelopment in the etiology of many common neuropsychiatric disorders, such as autism spectrum disorder (ASD), intellectual disabilities, and schizophrenia (*4*–*7*). Development, including cell proliferation, interaction, and differentiation, is the result of an inherent gene regulation governed by complex interactions between enhancers, promoters, noncoding RNAs, and transcription regulatory proteins. However, the understanding of epigenetic gene regulation in the developing human brain is very limited, largely owing to the relative scarcity of available human brain tissue at early developmental time points.

The human cerebral cortex has undergone an extraordinary increase in size and complexity during mammalian evolution, in part through the symmetrical division and the exponential increase in number of radial glial (RG) cells, which are the cortical stem cells (*1*). The genetic and molecular underpinnings of this process are still unclear, perhaps because these events occur embryonically, before the cortical anlage is formed during the fetal period. Human induced pluripotent stem cells (hiPSCs) and hiPSC-derived organoids allow investigators to gain specific and direct insights into the genetic and molecular events that drive these very early aspects of human cortical development.

Brain organoids match embryonic to early fetal stages of human cortical development

We produced hiPSC lines from fibroblasts isolated from human postmortem fetuses at midgestation, and we differentiated these lines into telencephalic organoids patterned to the dorsal forebrain; samples of cerebral cortex were collected from the same specimens for comparative analyses (fig. S1). To assess the validity of hiPSC-derived telencephalic organoids as a model of human brain development, we compared overall gene expression and regulation of organoids with isogenic cortical brain tissue. Several iPSC lines were derived from skin fibroblasts of postmortem fetal specimens 310, 313, and 320, aged between 15 and 17 PCWs, for which cortical tissue was available (fig. S2 and table S1). The hiPSC lines derived from fetal fibroblasts were comparable to those derived from adult fibroblasts with regard to pluripotency, growth rate, and differentiation potential (figs. S3 and S4 and table S2) (*8*). From two hiPSC lines per each of the fetal specimens, we generated telencephalic organoids patterned to the dorsal forebrain (*6*), grew them under proliferative conditions for 11 days, and then moved them into a terminal differentiation (TD) medium. Organoids were randomly collected for RNA sequencing (RNA-seq) from whole cells as well as nuclear fractions and histone mark chromatin immunoprecipitation sequencing (ChIP-seq) from nuclear fractions at around day 0, day 11, and day 30 of TD in vitro (TD0, TD11, and TD30, respectively). The transcriptomes of whole cells and nuclear RNA were highly correlated (fig. S5) (*8*); hence, we used the cellular transcriptome for all subsequent analyses. Peaks of three histone marks [trimethylation of histone H3 on lysine 4 (H3K4me3), acetylation of histone H3 on lysine 27 (H3K27ac), and trimethylation of histone H3 on lysine 27 (H3K27me3)] were called to mark functional elements including enhancers, promoters, or polycomb-repressed regions (table S3) (*8*). To place organoids in a human developmental context, we then compared transcriptomes and chromatin marks from organoids with those from the corresponding isogenic cortical tissue, human embryonic stem cell (hESC) lines, and brain tissue of various ages obtained from the PsychENCODE developmental dataset (*9*), other PsychENCODE projects (*10*), and the Roadmap Epigenomics project (*11*) (Fig. 1A).

Hierarchical clustering of transcriptomes and histone marks revealed that fetal, perinatal, and adult brain samples formed separate clusters (Fig. 1, B to D), confirming fundamental differences in gene expression in prenatal versus postnatal stages of brain development (*12*, *13*). Furthermore, hiPSC and hESC lines from different sources (including ours) and brain organoids clustered together with fetal brain tissue and separately from adult brain tissue. However, hiPSC and hESC lines formed a distinct subcluster, highlighting differences between organoids and pluripotent cells. Within each cluster, datasets for the same cell type but from different sources were highly concordant with each other (i.e., our data, those of Roadmap Epigenomics, and the PsychENCODE developmental dataset), suggesting that batch effects were not responsible for the observed clustering.

Within our datasets, organoid transcriptomes clustered by in vitro age (i.e., TD0, TD11, and TD30) irrespective of the hiPSC lines from which they were generated, suggesting that the transcriptome reveals well-defined, stage-specific cellular differentiation processes (Fig. 1E and fig. S6). Invariably, organoids clustered separately from the corresponding isogenic fetal cortex. To understand the relationships between organoids and the developing human brain, we classified the organoids against the PsychENCODE developmental dataset (*9*), which spans a wide range of

[1]Child Study Center, Yale University, New Haven, CT 06520, USA. [2]Department of Health Sciences Research, Center for Individualized Medicine, Mayo Clinic, Rochester, MN 55905, USA. [3]Department of Neuroscience, Yale University, New Haven, CT 06520, USA. [4]National Research Foundation of Korea, Daejeon, South Korea. [5]Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC 27708, USA. [6]Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA. [7]Department of Molecular Biophysics and Biochemistry, Department of Computer Science, and Department of Statistics and Data Science, Yale University, New haven, CT 06520, USA. [8]Kavli Institute for Neuroscience, Yale University, New Haven, CT 06520, USA. *These authors contributed equally to this work. †The PsychENCODE Consortium authors and affiliations are listed in the supplementary materials. ‡Corresponding author: Email: abyzov.alexej@mayo.edu (A.A.); flora.vaccarino@yale.edu (F.M.V.)

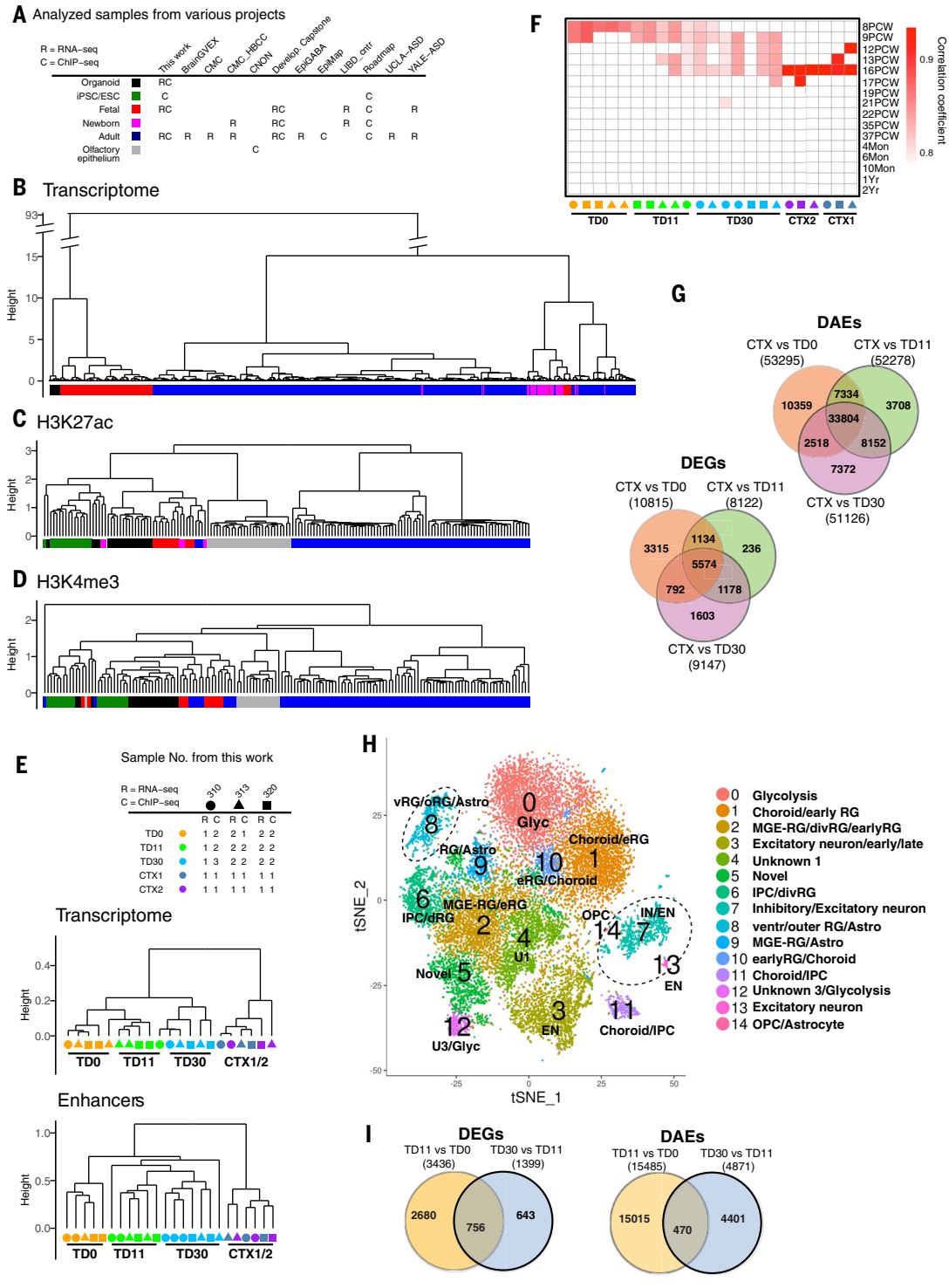


Fig. 1. Comparison of transcriptome and epigenome of organoid and isogenic fetal brain. (**A**) Dataset and sample annotation. Samples are from both our project (hiPSC lines, organoids, fetal brain samples), other PsychENCODE projects, and the Roadmap Epigenomics project. Colors correspond to datasets represented in (B) to (D). (**B** to **D**) Hierarchical clustering dendrograms of samples by transcriptomes (B) and ChIP-seq peaks of H3K27ac (C) and H3K4me3 (D). (**E**) Hierarchical clustering of organoids and isogenic postmortem cortexes by transcriptomes and gene-associated enhancer elements. Organoid and brain samples used for clustering are shown on top. Colors and shapes correspond to the datasets represented in the panels below. (**F**) Transcriptome-based classification of organoids and isogenic cortexes by age (*8*) against the tissues from the PsychENCODE developmental dataset (PCW, postconceptional week) from Li *et al*. (*9*). For each sample, red shading indicates the average of correlation coefficients above the cut-off as defined in (*8*) between our sample and those in Li *et al*. (*9*). White boxes indicate correlations below the cut-off. Correlations to brains older than 2 years of age were all below the cut-off and thus were not displayed. (**G**) Overlap of differentially expressed genes (DEGs) and differentially active enhancers (DAEs) between organoids at each differentiation time point and isogenic fetal cortex (CTX). (**H**) tSNE scatterplot of 17,837 nuclei, colored by cluster. Clusters arising predominantly from fetal cortex are circled. RG, radial glia; MGE, medial ganglionic eminence; IPC, intermediate progenitor cells; OPC, oligodendrocyte precursor cells. "Novel" means no correspondence to previous annotations. (**I**) Counts of DEGs and DAEs between organoids at different stages of development.

human ages and brain regions. Organoids' transcriptomes mapped most closely to the human neocortex between 8 and 16 PCWs of development, with the isogenic fetal brain samples mapping most consistently around 16 PCW, in good agreement with their annotated age (Fig. 1F). This analysis places the organoids substantially earlier than their corresponding mid-fetal brains, suggesting that organoids model late embryonic to early fetal stages of telencephalic development.

We next compared transcriptomes between each stage of organoid development and the postmortem fetal cortical tissue from the same individual. Overall, there was a large number of differentially expressed genes (DEGs) between each organoid stage and isogenic brain tissue, of which roughly half was up-regulated and half down-regulated (Fig. 1G and table S4). Although some stage-specific DEGs were present, particularly at TD0 (24%), most of the differences (63%) were shared across two or more organoid stages. Top Gene Ontology (GO) terms for this common set of organoid-brain DEGs were neurogenesis and regulation of nervous system development,

whereas the TD0-specific set of organoid-brain DEGs were related to DNA replication, consistent with age and cell-type differences between fetal brain tissue and organoids (table S4). We tested this hypothesis in silico, by assessing for overlap between the organoid-brain DEGs and cell type–specific transcripts identified in fetal human brain (*14*). Genes up-regulated in the fetal cortex were consistently enriched in markers for maturing excitatory neurons, interneurons, and newborn neurons compared to all organoid stages, whereas genes up-regulated in organoids at TD0 and TD11 were enriched in markers for dividing RG (fig. S6B and table S5).

To validate bulk analyses, we performed single-nucleus RNA sequencing (snRNA-seq) (*8*) and analyzed the cellular composition of organoids and the fetal brain (one sample per differentiation time point and one sample for brain). We shallow-sequenced about 10,000 cells per sample and considered the top 6000 most informative cells in each sample. We retained only cells expressing at least 500 genes, resulting in a final set of 17,837 cells that were used for analysis. Batch-corrected clustering of single cell's transcriptomes by tSNE analysis from all samples identified 15 clusters (Fig. 1H), with 11 containing cells mostly from organoids and 4 containing cells mostly from the fetal cortex (fig. S6, C and D). Differential expression analysis between any individual cluster and all the others highlighted sets of marker genes for each cluster (table S6), and we used a combination of published datasets of cell markers from single-cell RNA-seq studies of fetal human brain samples (*14*, *15*) to annotate them. The clusters largely contributed by organoid cells overlapped with those identified in human developing brains (*15*) (Fig. 1H and fig. S6E), and only one cluster, cluster 5, did not find any correspondence to the postmortem human dataset and was labeled "novel." These organoid-specific clusters comprised various types of RG cells including early RG (eRG), outer RG (oRG), ventricular RG (vRG), dividing RG (divRG), and truncated RG (tRG). In addition, cluster 3 expresses early- and late-born excitatory neuron (EN) markers, consistent with an organoid specification to dorsal cortex. Cell clusters specific to the fetal cortex contained inhibitory and excitatory neurons (IN/EN) (clusters 7, 13), RG cells (cluster 8), and a small oligodendrocyte precursor cell (OPC) cluster (cluster 14) (table S6). The presence of IN in the fetal cortex is expected, given that the cortex at PCW 17 is already receiving migrating interneurons from the developing basal ganglia. Timewise, our TD0 organoids (clusters 1, 2, 5, 6, and 10) containing RG and choroid cells matched with cells ranging from 6 to 9 PCWs in fetal brain samples (*15*). Correspondingly, our CTX1 (clusters 7, 8, 13 and 14) matched with markers (MGE-RG, RG, IN, and EN) seen in 15- to 16.5-PCW fetal brain (fig. S6, K and L). Together, the data confirmed the conclusion of bulk transcriptome analyses that organoids are younger than the fetal brain.

The fraction of cells in a cluster originating from a sample at each time point reveals some clear trends: clusters 1 (Choroid/eRG), 2 (MGE-RG/dorsal RG/eRG), 6 (IPC/divRG), and 10 (eRG/Choroid) decreased over time, consistent with their being composed of mostly immature cells originating from organoids at TD0 (fig. S6, C and D, and table S6). By contrast, clusters 0 (Glyc) and 12 (U3/Glyc), mostly from samples at TD30, increased with time, perhaps suggesting changing metabolic requirements among neural precursors (*15*). The remaining clusters, in particular clusters 3 (EN), 4, and 5 (unknown), reached a maximum at TD11, consistent with findings that some newborn neurons peak at an intermediate pseudoage (*15*). Finally, we ordered the cells along a pseudotime (fig. S6, F to I), which revealed cell trajectories along several dimensions (*8*). Cells originating from TD0 samples populated the top branch and were nearly absent after the first branch point, which is consistent with the pseudotime progression (fig. S6H) from the top branch (time 0) to the left and right bottom branches (time 15). Similarly, scoring individual cells using cell cycle markers (fig. S6I) revealed a higher frequency of actively cycling cells (G_2-M or S phase) at the early pseudotimes and larger fractions of non-cycling cells (G_1 phase) when moving along each path (*8*). In summary, from this integrated analysis emerges a highly coherent picture of organoids' temporal evolution (i.e., differentiation and maturation), representing earlier stages with respect to the corresponding 17-PCW fetal brain counterpart, and mimicking early human brain development, consistent with the classification of the bulk transcriptome with the PsychENCODE developmental Capstone dataset.

We next defined putative promoter and enhancer elements as well as repressed chromatin from histone mark data by chromatin segmentation analyses (figs. S1 and S7 and tables S7 and S8) (*8*). As a result, we identified 327,877 putative enhancers (H3K27ac peaks, which lack H3K4me3 and H3K27me3 signals) across organoids and fetal brains (table S9). Among these enhancers, H3K27ac signals are highly correlated with ATAC-seq (assay for transposase-accessible chromatin using sequencing) signals, confirming the open chromatin signatures and supporting the robustness of our approach (fig. S7). We further connected these enhancers to genes either by promoter-enhancer distance (within 20 kb) or by the strength of their physical interaction to gene promoters on the basis of Hi-C data for fetal brains (*16*). From the initial dataset of >300,000 putative enhancers, 96,375 enhancers (29.4%) were found to be associated with 22,835 protein-coding or long intergenic noncoding RNA (lincRNA) genes (out of 27,585 such genes from Gencode V25 annotation) (*17*) and were used for further analyses (table S10). The gene-associated enhancer dataset was corroborated by the observation of the trend that an increase in activity of enhancers or associated number of enhancers leads to higher expression of interacting genes (figs. S8 to S10).

Of the 96,375 gene-linked enhancers, 90% are concordant with those previously discovered by the ENCODE/Roadmap Consortia in various cell lines and tissues (*18*), and 10,243 (10%) were completely novel. Overall, 83,608 and 46,735 were active in organoids and the isogenic mid-fetal cortex, respectively. Of the former, 49,640 (59%) were active only in organoids (fig. S11E) and down-regulated in the mid-fetal brain, suggesting that organoids, and by extension, the embryonic and early fetal cortex, use roughly 1.8-fold as many enhancers as later developing cerebral cortex. Comparing enhancer numbers active in organoids across stages, an increasingly larger number became active with the progression of organoid development, with roughly 11,700 enhancers becoming active only at TD30 (fig. S11F). Furthermore, hierarchical clustering analyses based upon the degree of enhancer activity (magnitude of the H3K27ac signal) (Fig. 1E) revealed two major clusters—organoids and the fetal cortex—where organoids' enhancers clustered by in vitro age (i.e., TD0, TD11, and TD30) irrespective of genomic background of hiPSC lines, a pattern almost identical to that of transcriptome data (Fig. 1E and fig. S6). Finally, comparing enhancer activity between each stage of organoid development and fetal cortical tissue from the same individual showed that the three organoid stages shared a large number of differentially active enhancers (DAEs) with respect to the fetal cortex (Fig. 1G), as observed with transcriptome data. Together, these analyses reveal a close parallelism between gene expression and enhancer activities across early development and suggest that gene regulation in embryonic and early fetal development is driven by sets of early enhancers, most of which are not active in the mid-fetal cerebral cortex.

Expression and regulatory changes defining early developmental transitions in organoids

To better understand the gene-regulatory changes driving embryonic and early fetal development, we analyzed DEGs and DAEs in organoids between transitions TD0 to TD11 and TD11 to TD30. We found that the largest differences in gene expression and enhancer activity were at the first transition and that from ⅔ to ¾ of changes were specific for this transition (Fig. 1I and tables S10 and S11), confirming that a substantial change in gene regulation must occur at the beginning of cortical stem cell differentiation. Down-regulated genes specific for the first transition were related to mitosis and regulation of the cell cycle, including cyclin-dependent kinases (CDK2, CDK4, and CDK6) and DNA repair enzymes (TP53, BRCA1/2, and PCNA), all showing a downward trend in expression, likely reflecting top proliferative activity of precursor cells at the earliest time point that decreases during differentiation (fig. S12 and table S11). Consistent with this observation, markers for cell proliferation were progressively down-regulated at the cellular level between TD0 and TD30 (fig. S3). Top functional annotations for genes down-regulated at the second transition (from TD11 to TD30) were instead related to transcriptional regulation of pluripotent and cortical precursor cells (i.e., SOX1/2, EOMES, LHX2, FOXG1, POU3F2/3, SIX3, FEZF2, EMX2, GLI1/3, NEUROD4, HES5/6, REST, and DLL3). By contrast, genes involved in the development of the

neuronal system and synaptic transmission were up-regulated at both transitions and included cell adhesion–, guidance– and synaptic molecule–related genes, including a large number of receptors, calcium and potassium channels, and synaptic membrane recycling components, as well as intellectual disability–related genes such as several CNTN family members.

Performing ChIP-seq and RNA-seq in the same samples provided an opportunity to assess the impact of enhancers on the transcription of their gene targets. We correlated enhancer activity and expression of their associated genes across the whole dataset (organoids and brain samples) to reveal that, globally, 10.6% of gene-enhancer pairs had significant positive or negative correlations, corresponding to 15,026 enhancers and 7858 genes (table S12). Observation of both positive and negative correlations is reminiscent of the finding that H3K27ac-enriched regulatory regions, commonly referred to as enhancers, can be bound by both activators and repressors of gene transcription (*19*). We referred to 10,192 (67.8%) enhancers with positive correlations as activating regulators (A-regs) of 5605 genes, and to 4993 (33.2%) enhancers with negative correlations as repressing regulators (R-regs) of 3251 genes. Moreover, 98.9% of enhancers are either A-regs or R-regs but not both, consistent with the notion that binding sites of activators and repressors are mutually exclusive (*20*). Indeed, across both transitions, we observed more pronounced correlations between expression changes of genes and activity change of linked A-regs versus linked non-A-regs; similar observations were made for R-regs (fig. S13A). Consistently, differentially active A-regs and R-regs are associated with DEGs in the expected direction, i.e., A-regs with increased activity are enriched in up-regulated DEGs, whereas R-reg with increased activity are enriched in down-regulated DEGs (Fisher's exact test, $p < 2.2 \times 10^{-16}$ for both transitions) (fig. S13B), suggesting that differential activity of the identified enhancers is indeed driving differential gene expression across organoid development.

Gene and enhancer network analyses

To study the temporal dynamics of gene expression and enhancer activities across the three developmental time points, we used weighted gene coexpression network analysis (WGCNA) (*21*). The resulting networks grouped gene transcripts in 54 coexpressed modules (MG1 to MG54) and gene-associated enhancers into 29 coactive modules (ME1 to ME29), each showing a specific trajectory along organoid differentiation (Fig. 2, A and B, and tables S12 and S14). Unsupervised hierarchical clustering of module eigengenes, which are representative of the gene expression and enhancer activity of each module, grouped samples by differentiation time point. Using k-means clustering of each module's eigengenes, we grouped the gene and enhancer modules into six and four "supermodules," respectively, which represent higher-order clustering of the modules (Fig. 2, C and D).

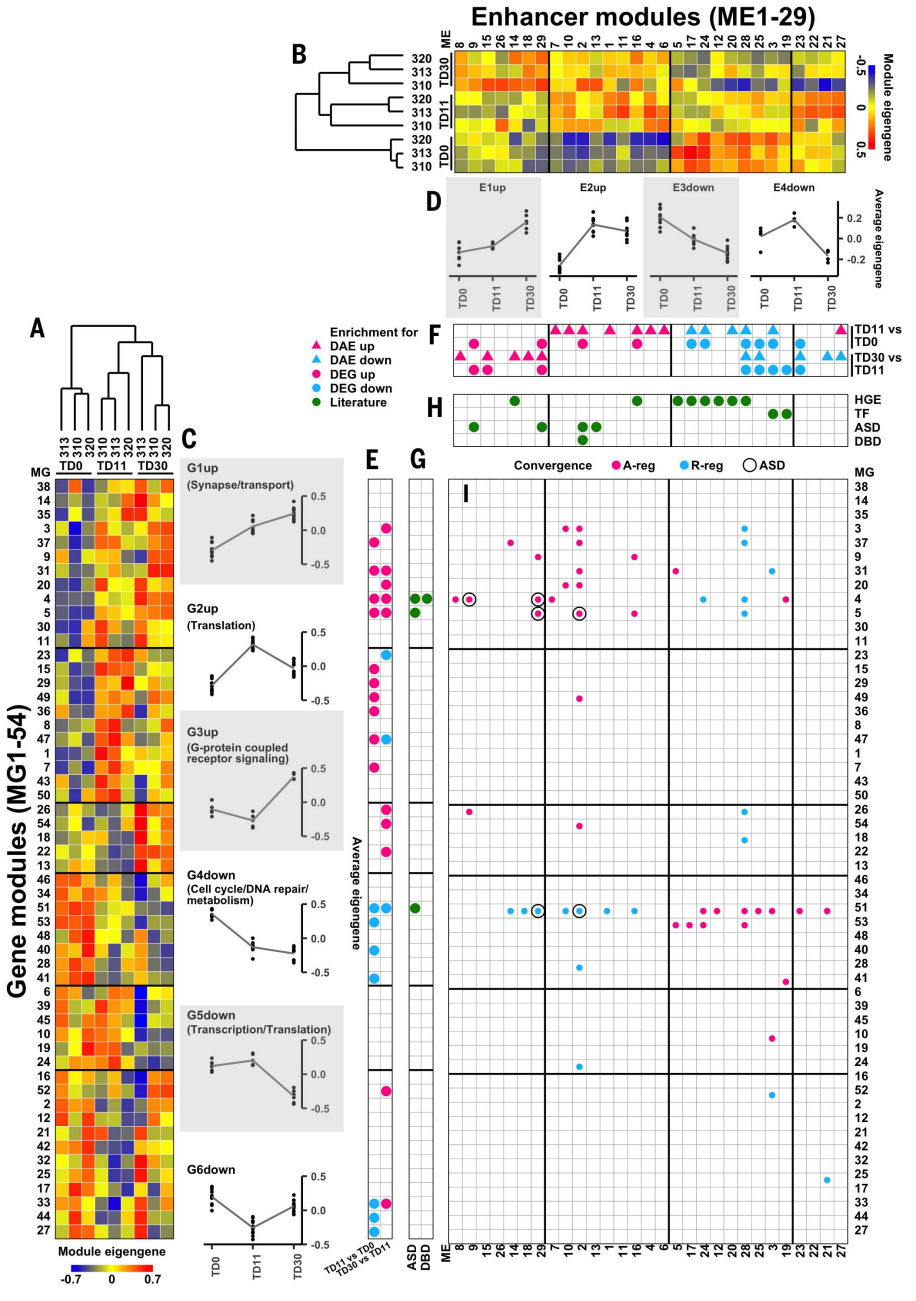


Fig. 2. Modules of coexpressed genes and coactive enhancers during organoid differentiation. (**A**) Unsupervised hierarchical clustering of gene modules (1 through 54) by expression eigengenes. Rows and columns represent gene modules and samples, respectively. (**B**) Unsupervised hierarchical clustering of enhancer modules (1 through 29) by activity eigengenes. Rows and columns represent samples and enhancer modules, respectively. (**C** and **D**) Mean module eigengenes (lines) across differentiation times grouped by gene (C) and enhancer (D) supermodules, respectively. Dots represent values of eigengenes for individual modules. (**E** to **H**) Enrichment of gene (E and G) and enhancer (F and H) modules for DEGs and DAEs and for various enhancers and genes of interest from the literature, including HGE (human-gained enhancers) (*26*), TF (genes encoding transcription factors during human fetal brain development) (*24*), ASD (genes pertinent to autism spectrum disorder) (*22*), and DBD (genes pertinent to developing brain disorder) (*23*). (**I**) Correspondence between the gene and enhancer networks. The strongest A-reg (pink dots) and R-reg (cyan dots) for a subset of gene modules are overrepresented in a number of enhancer modules. Black circles emphasize converging genes and enhancer modules, both of which are ASD-associated [as shown in (G) and (H)]. Panels (E) to (I) are aligned by the gene and enhancer modules shown in (A) and (B).

Supermodules exhibit specific profiles of activities during the two transitions (*8*) and functional annotations (table S14). The monotonically up-regulated gene supermodule G1up comprised modules related to neurons, synapses, cell adhesion, and axon guidance and was hence dubbed as governing synapse/transport. Conversely, the down-regulated supermodule G4down comprised modules enriched in DNA repair and cell cycle–related genes and was thus dubbed as governing cell cycle/DNA repair (Fig. 2C), reflecting the cell cycle annotation of TD0-to-TD11–down-regulated DEGs (fig. S12). Other supermodules exhibited transition-specific changes. G2up, which exhibited peak up-regulated gene expression at TD11, was enriched in genes related to ribosome, translation, protein folding, and degradation. The transcription supermodule G5down, down-regulated at the second transition, included major transcription factors (TFs) expressed by cortical progenitor cells, which show down-regulation at TD11 to TD30 (fig. S12). By contrast, the G3up supermodule, up-regulated at the second transition, was enriched in G protein receptor signaling, implying a previously unknown role of these molecules in the earliest stages of cortical neuron differentiation. Patterns of gene expression and enhancer activity in the modules and supermodules were further confirmed by enrichment analysis of DEG and DAEs (Fig. 2, E and F). Specifically, gene modules and linked genes of enhancer modules were enriched with DEGs for which gene expression changes were generally in the same direction as their respective module eigengenes.

Further evidence for functional relevance of the modules and supermodules arises from intersection with genes relevant to neuropsychiatric diseases. Genes within the SFARI dataset, a curated list of genes associated with ASD, including both rare mutations and common variants (*22*), were significantly overrepresented in the MG4 and MG5 neuronal and synaptic modules and the MG51 cell cycle module (Fig. 2G and table S14). SFARI genes were also enriched within gene targets of four enhancer modules (ME9 and ME29 in supermodule E1up, and ME2 and ME13 in supermodule E2up) with up-regulated patterns of activity across development, one of which, the ME2 module, was also enriched in developmental brain disorder genes (*23*) (Fig. 2H). Enrichment analysis also showed that a set of TFs pertinent to human cortical neurogenesis (*24*) was preferentially associated with gene targets of two enhancer modules (ME3 and ME19, both in supermodule E3down) that have down-regulated enhancer activity across organoid development (Fig. 2H). This evidence supports the notion that organoid culture can capture dynamic gene-regulatory events present in early human brain development and that such early events are potentially involved in disease pathogenesis.

To assess the correspondence between the gene network and the enhancer network, we examined whether enhancers linked to a gene module are overrepresented in one or a small number of enhancer modules. Such convergence between a gene module and an enhancer module would suggest that coexpressed genes are likely regulated by enhancers with correlated patterns of activity. To mitigate the ambiguity caused by multiple enhancers per gene, we focused on the strongest A-regs or R-regs of a gene, defined by the most positive or negative correlation between enhancer activity and gene expression. Indeed, we find that A-regs and R-regs of 14 and 12 gene modules, respectively, are overrepresented in a small number of enhancer modules [false discovery rate (FDR) < 0.05, Fig. 2I]. Not surprisingly, A-regs and R-regs linked to the same gene module are overrepresented in different enhancer modules with opposite trajectories over time, e.g., A-regs of MG3 in G1up converges with ME10 and ME2 in E2up, but its R-regs converges with ME28 in E3down. Such convergence between the gene network and the enhancer network suggests that coexpressed genes likely share a set of co-regulated enhancers. Moreover, enhancers discovered in organoids hint at upstream elements that regulate the expression of disease-associated genes. For example, ASD-associated MG4, MG5, and MG51 gene modules converge with ME9, ME29, and ME2, enhancer modules that are associated with ASD genes as well (Fig. 2, G to I, black circles). ME29 is particularly interesting as it contains both A-regs and R-regs for all three ASD-associated gene modules, suggesting that it may be responsible for the coordinated up- and down-regulation of genes modules involved in autism pathogenesis.

The ASD-associated gene modules—MG4, MG5, and MG51—overlapped to a significant extent with previously published ASD modules identified by in vivo analyses of differential gene expression between ASD patients and normal individuals (Fig. 3A and table S14). Our MG4 and MG5 modules were annotated by neuronal and synaptic terms (Fig. 3B) and overlapped with neuronal and synaptic modules down-regulated in the ASD postmortem cerebral cortex (*25*) as well as with a synapse module up-regulated in brain organoids from ASD individuals with macrocephaly (*6*). By contrast, our down-regulated MG51 module was annotated by cell cycle and DNA repair terms (Fig. 3B) and overlapped with M3, a module harboring protein-disrupting, rare de novo variants in ASD (*4*). No overlap was observed with modules related to immune dysfunction and microglia in ASD (*25*) (Fig. 3A). Within each ASD-associated gene module, the distribution of genes that are implicated in ASD and are targets of a member of the ME9, ME29, ME2, and ME13 ASD-associated enhancer modules appears, overall, to be skewed toward the central part of each module (i.e., the "strongest" hubs) (Fig. 3, C and D, and fig. S14). Given that hub genes are the drivers of a module, one may speculate that mutations disrupting these genes are more likely to be penetrant and/or syndromic. Looking at the first 100 hub genes (table S14), we find that the MG4 module shows two confident and two syndromic ASD-associated genes (respectively DSCAM, MYO5A, CAMK2B, and SMARCA2); the MG5 module shows three confident and three syndromic ASD-associated genes (respectively ANK3, STXBP1, ACHE, WDR26, and ATP1A3); and the MG51 module only shows DIAPH3, a lower-confidence gene (Fig. 3C and fig. S14). Orthogonal analyses by quantitative polymerase chain reaction (qPCR) confirmed the expression level of these and other ASD genes in the organoid dataset (fig. S15). Overall, the results suggest that our organoid model may be used to unravel the roles of early prenatal neurodevelopment and genetic factors in ASD.

Relevance of the organoid model to understanding human brain evolution

To determine whether the organoid model is useful to understanding the genetic mechanisms driving human brain evolution, we assessed the overlap of our enhancers with a list of 8996 human-gained enhancers (HGEs). These HGEs showed increased activity at very early stages of brain development (7 to 12 PCWs) in the human lineage, compared with their homologs in rhesus macaque and mouse brains at similar developmental time points (*26*). The majority (70%, 6295 out of 8996) of published HGEs overlapped with 9915 enhancers in our dataset, and among the latter, 3310 are associated with genes (table S15). Out of 3310 gene-associated HGEs, 2670 (85.3%) have differential activity between organoids and fetal brains, suggesting a dynamic role during brain development (fig. S16). The largest fraction of gene-associated HGEs are progressively declining in activity along organoid differentiation and from organoids to fetal brain. Among eight enhancer modules enriched with HGEs, six (all in the supermodule E3down) had decreasing activity along organoid differentiation (Fig. 2H). Genes targeted by HGEs in these six down-regulated modules were enriched in signaling pathways related to cell proliferation and cell differentiation and communication and included extracellular growth factors such as FGF7 and FGF6, FGFRL1, ERBB4, IGF2, EGFL7, VEGFA, and PDGFA (table S15). Overall, among all 2908 HGE-linked genes, 824 are differentially expressed between human and macaque brain in at least one of three brain ages—438 in fetal brains, 346 in postnatal brains, and 724 in adult brains (*27*). Together, these findings suggest that HGEs are likely to be important regulators of genes controlling cell proliferation and cell-to-cell interactions in the human cerebral cortical primordium during the very early stages of cortical morphogenesis. These data are consistent with ATAC-seq from in vivo human brain (*24*), which demonstrates that HGEs are active in germinal zones and especially enriched in outer radial glia (oRG), which are expanded in humans (*28*).

Gene regulation and relevance to disorders

More than 24% of the ASD genes in the SFARI dataset are differentially expressed in the organoid system across time, and over 80% are linked to enhancers active in organoids or fetal brain (table S16). To understand whether enhancers active in organoids or fetal brain can inform about common and rare genetic variants that

underlie ASD, we selected three subsets from the 96,375 gene-associated enhancers: 11,448 early enhancers, only active in all organoid stages; 8999 late enhancers, only active in fetal brain; and 7865 constant enhancers, active at all stages of organoid differentiation and in fetal brain (Fig. 4A). These enhancers were analyzed for enrichment with personal variants inherited from either parent in 540 families of the Simons Simplex Collection (SSC). Each family consisted of phenotypically normal parents, an ASD male proband, and a normal male sibling (Fig. 4A). Out of an average 3.6 million inherited single-nucleotide polymorphisms (SNPs) per person, 3327 with <5% minor allele frequency (MAF) were located within early, late, or constant enhancers (fig. S17, A to C). Among these, low-allele frequency SNPs (MAF 0.1% to 5%) were significantly enriched in probands relative to siblings in early but not in late or constant enhancers (p = 0.02 by one-sample t test, Fig. 4B).

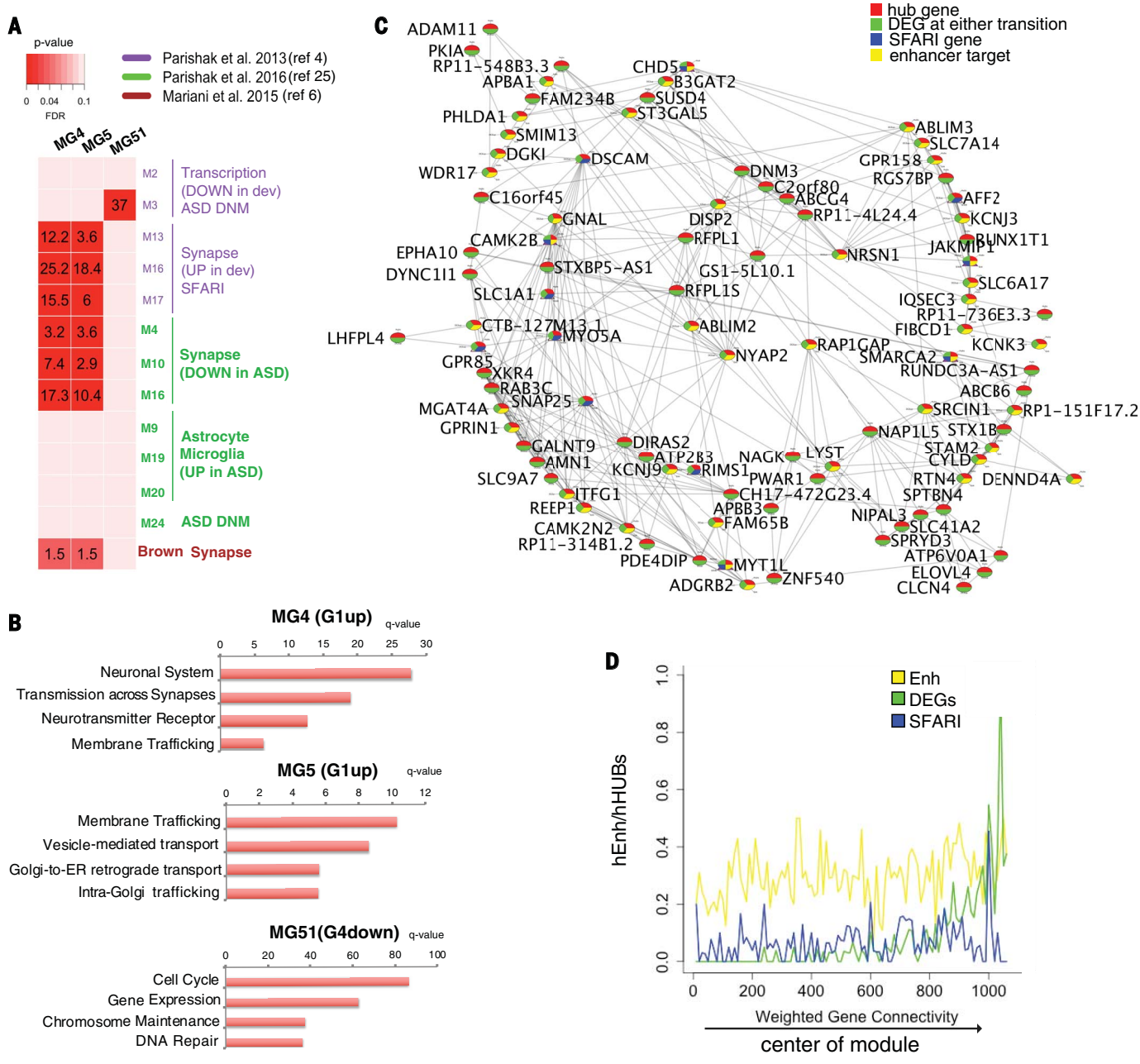


Fig. 3. ASD-associated genes modules. (**A**) Overlap of ASD gene modules MG4, MG5, and MG51 from this study with transcript modules associated with ASD from postmortem brain studies or enriched in ASD de novo mutations (DNM) (green, violet) (*4*, *25*) and from an ASD patient-derived organoid study (brown) (*6*). Rows are modules from this study and columns are modules from other studies. Red shading represents the degree of enrichment between pairs of modules. Corrected p values of significant overlaps (hypergeometric test) are numerically indicated as −log10(p value). (**B**) Bar plots of the top-scoring biological process terms for the ASD-associated modules shown in (A). (**C**) Graphical representation of the strongest interacting hub genes in the MG4 module network. Circles: genes; lines: topological overlap above 0.95. Colors in circles annotate each gene as hub (red), DEG (green), SFARI gene (blue), and enhancer target (yellow). Enhancer target: genes targeted by enhancers in the ME9, ME29, ME13, and ME2 ASD-associated enhancer modules (Fig. 2I). (**D**) Frequency plots within the MG4 module showing that enhancer targets, DEGs, and SFARI genes have higher intramodular connectivity. x axis shows the weighted gene connectivity, from low (peripheral genes) to high (central hub genes).

These SNPs were also enriched in the ME2 and ME29 enhancer modules (p = 0.05 and 0.03, respectively, by one-sample t test) (Fig. 4B), which converge with ASD-associated gene modules (Fig. 2I). These variants are relatively common, and thus our results support the hypothesis of etiology of ASD via superposition of multiple inherited variants of low effect size (*29*–*32*).

Contrary to numerous inherited SNPs, there are only a few dozen de novo mutations (DNMs) in probands, which must have deleterious effects in order to contribute to ASD phenotypes. We compared DNMs of probands and siblings of the same family cohort (*33*). Out of 66,306 total DNMs, 2422 were located in our dataset of gene-associated enhancers. There was a trend of having a larger fraction of probands' DNMs in constant enhancers, which are active during a prolonged period of development (Fig. 4C and fig. S17D). We next elucidated the effect of individual DNMs in the gene-associated enhancers on TF binding. Around 24% of DNMs (out of 1240 and 1184 from proband and sibling, respectively) overlapped with at least one TF motif (figs. S17, E and F, and S18). Overall, there was a larger number of TFs with greater count of motif-breaking DNMs in probands than in siblings (more circles below the diagonal than above in Fig. 4D). A significant difference (p < 0.05 by binomial test) was observed for TFs such as homeodomain, Hes1, NR4A2, Sox3, and NFIX (table S17), which are implicated in development, ASD, or mental disorders (*34*, *35*). De novo copy-number variants at the NR4A2 gene locus at 2.q24.1, in particular, have been associated with ASD with language and cognitive impairment across multiple datasets (*35*). These observations provide genetic support for the relevance of enhancer elements identified in organoids in the complex etiology of ASD and link noncoding variants to ASD etiology, as previously proposed (*36*). Enhancers discovered in this study also inform about the possible regulatory role of SNPs that underlie the etiology of schizophrenia (*37*) (fig. S17G).

Discussion

Using forebrain organoids, we provide an initial map of enhancer elements and corresponding transcripts that are dynamically active in the transitions between human cortical stem cells, progenitors, and early cortical neurons. Although the cataloged functional elements may require further validation of their in vivo activity, our findings suggest that human brain organoids provide an avenue to approach the study of the molecular and cellular events underlying brain development. Indeed, our brain organoids patterned to forebrain, on both transcriptome and regulatory levels, mimic the longitudinal development of the embryonic and early fetal cortical primordium. Because all organoid preparations (from other studies and with different protocols) patterned to the dorsal forebrain are derived from neural stem cells, it is likely that they share similar gene dynamics specific to the embryonic brain described here. Thus, our gene and enhancer analyses have wide implications, and the described map can aid the identification of sets of genes, enhancers, and genomic variants underlying neurodevelopmental disorders and ASD in particular, because brain development is nearly complete at the time of diagnosis (*38*).

The majority of enhancer elements active in our organoid system are not shared with isogenic mid-fetal brain tissue, which suggests that they play a role in earlier events, i.e., progenitor proliferation and the specification of neuronal lineages. However, it remains unclear whether organoids fully recapitulate developmental processes, particularly those at later stages. Organoid preparations grown for longer periods in vitro may show greater overlap with mid-fetal human brains (*39*, *40*), although a notable aspect of the organoid system is its ability to span very early developmental transitions, which map to stages earlier than those commonly available in postmortem human tissue. This finding is confirmed by single-cell transcriptome analyses, which revealed a wide diversity of RG and progenitor clusters throughout organoid development. All but one organoid-specific cell clusters find correspondence to cell clusters in embryonic-fetal human brain. The one that did not could be the result of in vitro culturing. Through longitudinal analyses, we show that many genes and their enhancer elements are differentially active in a stage-specific fashion from RG stem cells to

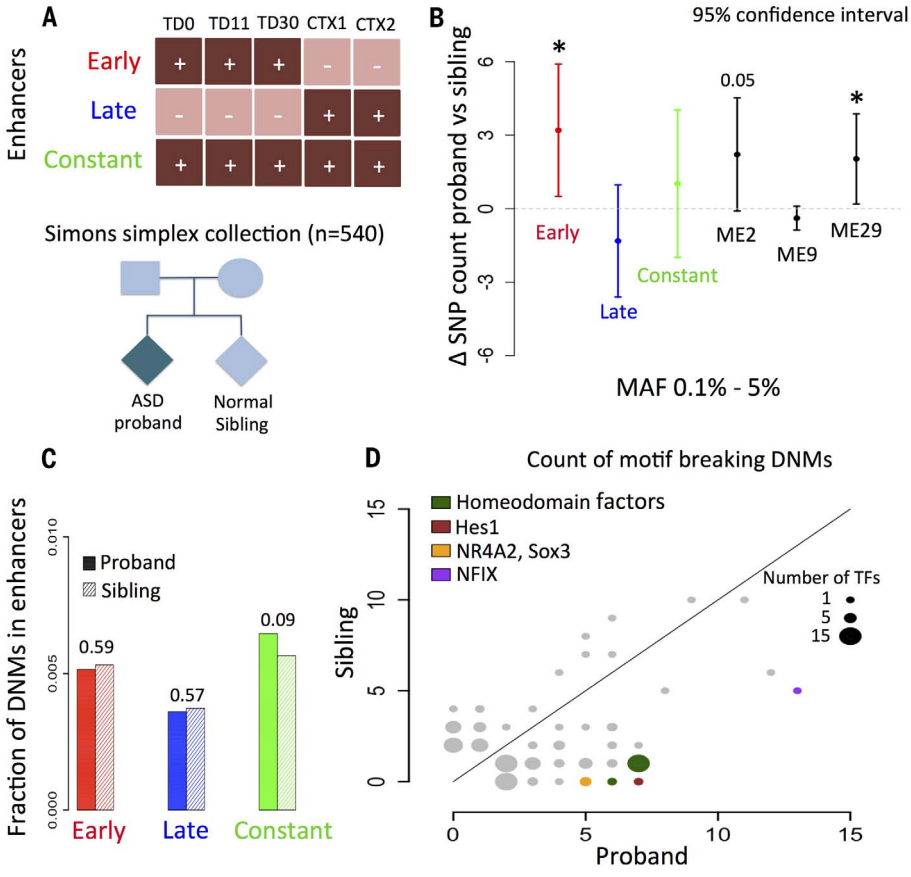


Fig. 4. Enrichment of variants in gene-associated enhancers. (**A**) Three subsets of enhancers were selected from all gene-associated enhancers. Early: enhancers active (denoted by +) in all organoid stages but inactive (denoted by -) in fetal brain (red); late: enhancers active in fetal brain but inactive in all organoid stages (blue); constant: enhancers active in all organoid stages and fetal brain (green). Variants in 540 families from the Simons Simplex Collection were analyzed for enrichment in these enhancer sets. (**B**) Comparison of inherited personal SNPs between ASD probands and normal siblings from the SSC revealed significant enrichment in probands versus siblings ($p \leq 0.05$ by one-sample t test) of low–allele frequency SNPs (MAF 0.1 to 5%) in early enhancers (red) and enhancer modules ME2 and ME29 (black). Dashed line at value of 0 represents no difference between probands and siblings. *p < 0.05. (**C**) Fractions of DNMs in enhancers were compared in probands and siblings across the whole genome. P values (shown above the bars) were calculated by using the chi-square test. (**D**) Count of motif-breaking DNMs in all gene-associated enhancers were compared between probands and siblings. Circles represent TFs with counts of broken motifs in probands and siblings plotted on x- and y axis. The size of the circles is proportional to the number of TFs. Circles away from the diagonal represent TFs enriched with motif-breaking DNMs in probands or siblings. A few TFs in the probands (colored circles) but not in the siblings were significantly enriched (p < 0.05 by binomial test) with motif-breaking DNMs.

neuronal progenitors and to young neurons. The first transition, from neural stem cells to early cortical progenitors, has the largest number of DEGs (71%) and DAEs (76%), the majority of which are specific to that step, which implies that in vivo transition from the embryonic to the fetal brain is a vulnerable step for normal brain development. Such changes reflect dynamic transitions in proliferation-related genes and transcription factors, together with the upregulation of neuronal lineage and synaptic genes as cortical stem cells (i.e., RG cells) progressively stop dividing and acquire different neuronal identities. We found that HGEs exhibit their highest activity in RG cells, after which their activity progressively declines with differentiation. Consistent with previous findings (*24*), this observation implicates HGEs as regulators of the earliest phases of human brain development. Although the exact function of HGEs remains to be determined, based on enrichment for growth factors signaling pathways, their time course and the comparison with other studies, we hypothesize that they are involved in the regulation of RG cell proliferation in the cerebral cortex.

Global integrative analyses of transcriptome and enhancer elements allowed us to classify the gene-associated enhancers into elements that activate or repress gene transcription, in which activity changes in A-regs and R-regs are correlated with changes in the expression of their gene targets at each developmental transition. Because a third of those regulators likely acted as gene-repressing elements, our results point out an underappreciated layer of trans-repression during early brain development. This level of integration allows the construction of a complex regulatory network with convergent and concordant patterns of activity between gene and enhancer modules, where enhancers of coexpressed genes also exhibit correlated activity. We propose that this network portrays fundamental developmental programs in embryonic and fetal brain.

Three gene modules were enriched in genes implicated in ASD, two of which, MG4 and MG5, regulate neuron and synapses and progressively increased in expression during development; whereas the other, MG51, regulates the cell cycle, and whose expression progressively declines. Those modules overlap gene modules previously implicated in ASD based on in vivo postmortem data (*25*). Additionally, we found that ASD-associated gene modules converged with three ASD-associated enhancer modules, implying that other genes and enhancers in those modules may also be related to ASD by shared expression and perhaps function. This supports the validity of the organoid model for the discovery and analysis of regulatory elements whose variation may underlie the risk for neuropsychiatric disorders. Indeed, enhancers active in organoids, and, by extension, embryonic and early fetal cerebral cortices, were enriched for low population-frequency personal variants carried by ASD probands relative to unaffected siblings. Furthermore, DNMs in ASD probands more frequently disrupted binding motifs of specific transcription factors within regulatory elements active at those stages. Those TFs, their disrupted binding motifs, and the gene targets of the enhancers with the motifs can be the subject of future functional studies on the etiology of ASD. Altogether, the evidence corroborates previous suggestions that single-nucleotide variants in noncoding regions contribute to ASD (*36*) and points to genes and regulatory elements underlying its onset. Thus, organoids can offer mechanistic insights into early human telencephalic development, brain evolution, and disease.

Methods summary

Detailed materials and methods can be found in the supplementary materials. hiPSC lines were derived from skull fibroblasts of three male fetal specimens aged between 15 and 17 PCWs, from which two cerebral cortical samples each were also collected for comparative analyses. iPSCs were differentiated into telencephalic organoids patterned to the dorsal forebrain as previously described (*6*). Organoids were collected at three TDs for downstream analyses. Immunohistochemistry using proliferation, glutamatergic, and GABAergic neuronal markers were used for organoids' differentiation quality control. Samples from iPSCs, iPSC-derived organoids, and fetal cerebral cortical regions were used for total stranded RNA-seq (cells and nuclei), snRNA-seq (nuclei), and ChIP-seq for three histone marks (H3K4me3, H3K27ac, and H3K27me3) (nuclei). We used edgeR (*41*) and trended dispersion estimates to infer differentially expressed genes and differentially active enhancers. We used the Seurat pipeline (*42*) for single-cell RNA-seq clustering and the Monocle pipeline (*43*) for single-cell trajectories. ConsensusPathDB (*44*) and ToppGene (*45*) were used for functional annotation. Quantitative real-time PCR was used to cross-validate RNA-seq and DEG analyses using a random subset of the DEGs as well as DEGs implicated in ASD. ChIP-seq peaks were called by MACS2 (*46*), and chromatin segmentation was done by chromHMM (*47*). Peaks were merged into consensus peaks and annotated by the corresponding chromatin states at each TD or in the fetal cortex. We used physical proximity and published chromatin conformation (Hi-C) data (*16*) from the fetal brain to link enhancers to genes. Gene and enhancer modules were identified by WGCNA (*21*), and supermodules were defined by K-means clustering of module eigengenes. To assess the relevance of the organoid model to studying noncoding pathological mutations, personal genomic variants across the whole genome were obtained from the SFARI (Simons Simplex Collection) dataset in 540 families with ASD probands and normal siblings. We also used de novo SNPs identified in Werling *et al.* from the same cohort (*33*). Transcription factor binding site motifs were obtained from the JASPAR database (*48*).

REFERENCES AND NOTES

1. 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
2. J. L. R. Rubenstein, S. Martinez, K. Shimamura, L. Puelles, The embryonic vertebrate forebrain: The prosomeric model. *Science* **266**, 578–580 (1994). doi: 10.1126/science.7939711; pmid: 7939711
3. A. Visel *et al.*, A high-resolution enhancer atlas of the developing telencephalon. *Cell* **152**, 895–908 (2013). doi: 10.1016/j.cell.2012.12.041; pmid: 23375746
4. N. N. Parikshak *et al.*, Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013). doi: 10.1016/j.cell.2013.10.031; pmid: 24267887
5. A. J. Willsey *et al.*, Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013). doi: 10.1016/j.cell.2013.10.020; pmid: 24267886
6. J. Mariani *et al.*, FOXG1-Dependent Dysregulation of GABA/Glutamate Neuron Differentiation in Autism Spectrum Disorders. *Cell* **162**, 375–390 (2015). doi: 10.1016/j.cell.2015.06.034; pmid: 26186191
7. S. Gulsuner *et al.*, Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013). doi: 10.1016/j.cell.2013.06.049; pmid: 23911319
8. See supplementary materials.
9. M. Li *et al.*, Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* **362**, eaat7615 (2018).
10. S. Akbarian *et al.*, The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015). doi: 10.1038/nn.4156; pmid: 26605881
11. A. Kundaje *et al.*, Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015). doi: 10.1038/nature14248; pmid: 25693563
12. C. Colantuoni *et al.*, Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature* **478**, 519–523 (2011). doi: 10.1038/nature10524; pmid: 22031444
13. A. T. Tebbenkamp, A. J. Willsey, M. W. State, N. Sestan, The developmental transcriptome of the human brain: Implications for neurodevelopmental disorders. *Curr. Opin. Neurol.* **27**, 149–156 (2014). doi: 10.1097/WCO.0000000000000069; pmid: 24565942
14. S. J. Liu *et al.*, Single-cell analysis of long non-coding RNAs in the developing human neocortex. *Genome Biol.* **17**, 67 (2016). doi: 10.1186/s13059-016-0932-1; pmid: 27081004
15. 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
16. H. Won *et al.*, Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016). doi: 10.1038/nature19847; pmid: 27760116
17. J. Harrow *et al.*, GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012). doi: 10.1101/gr.135350.111; pmid: 22955987
18. The ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: 10.1038/nature11247; pmid: 22955616
19. Y. Nie, H. Liu, X. Sun, The patterns of histone modifications in the vicinity of transcription factor binding sites in human lymphoblastoid cell lines. *PLOS ONE* **8**, e60002 (2013). doi: 10.1371/journal.pone.0060002; pmid: 23527292
20. D. L. Daniels, W. I. Weis, Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat. Struct. Mol. Biol.* **12**, 364–371 (2005). doi: 10.1038/nsmb912; pmid: 15768032
21. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008). doi: 10.1186/1471-2105-9-559; pmid: 19114008
22. E. Larsen *et al.*, A systematic variant annotation approach for ranking genes associated with autism spectrum disorders. *Mol. Autism* **7**, 44 (2016). doi: 10.1186/s13229-016-0103-y; pmid: 27790361
23. 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
24. L. de la Torre-Ubieta *et al.*, The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289–304.e18 (2018). doi: 10.1016/j.cell.2017.12.014; pmid: 29307494
25. N. N. Parikshak *et al.*, Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016). doi: 10.1038/nature20612; pmid: 27919067

26. S. K. Reilly *et al*., Evolutionary changes in promoter and enhancer activity during human corticogenesis. *Science* **347**, 1155–1159 (2015). doi: 10.1126/science.1260943; pmid: 25745175
27. Y. Zhu *et al*., Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science* **362**, eaat8077 (2018).
28. T. J. Nowakowski, A. A. Pollen, C. Sandoval-Espinosa, A. R. Kriegstein, Transformation of the Radial Glia Scaffold Demarcates Two Stages of Human Cerebral Cortex Development. *Neuron* **91**, 1219–1227 (2016). doi: 10.1016/j.neuron.2016.09.005; pmid: 27657449
29. R. Anney *et al*., Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum. Mol. Genet.* **21**, 4781–4792 (2012). doi: 10.1093/hmg/dds301; pmid: 22843504
30. T. Gaugler *et al*., Most genetic risk for autism resides with common variation. *Nat. Genet.* **46**, 881–885 (2014). doi: 10.1038/ng.3039; pmid: 25038753
31. L. Klei *et al*., Common genetic variants, acting additively, are a major source of risk for autism. *Mol. Autism* **3**, 9 (2012). doi: 10.1186/2040-2392-3-9; pmid: 23067556
32. D. J. Weiner *et al*., Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat. Genet.* **49**, 978–985 (2017). doi: 10.1038/ng.3863; pmid: 28504703
33. D. M. Werling *et al*., An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nat. Genet.* **50**, 727–736 (2018). doi: 10.1038/s41588-018-0107-y; pmid: 29700473
34. G. Stelzer *et al*., The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Curr. Protoc. Bioinformatics* **54**, 1.30.1–1.30.33 (2016). pmid: 27322403
35. V. M. Leppa *et al*., Rare Inherited and De Novo CNVs Reveal Complex Contributions to ASD Risk in Multiplex Families. *Am. J. Hum. Genet.* **99**, 540–554 (2016). doi: 10.1016/j.ajhg.2016.06.036; pmid: 27569545
36. T. N. Turner *et al*., Genomic Patterns of De Novo Mutation in Simplex Autism. *Cell* **171**, 710–722.e12 (2017). doi: 10.1016/j.cell.2017.08.047; pmid: 28965761
37. A. F. Pardiñas *et al*., Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018). doi: 10.1038/s41588-018-0059-2; pmid: 29483656
38. S. P. Pașca, The rise of three-dimensional human brain cultures. *Nature* **553**, 437–445 (2018). pmid: 29364288
39. A. M. Pașca *et al*., Functional cortical neurons and astrocytes from human pluripotent stem cells in 3D culture. *Nat. Methods* **12**, 671–678 (2015). doi: 10.1038/nmeth.3415; pmid: 26005811
40. X. Qian *et al*., Brain-Region-Specific Organoids Using Mini-bioreactors for Modeling ZIKV Exposure. *Cell* **165**, 1238–1254 (2016). doi: 10.1016/j.cell.2016.04.032; pmid: 27118425
41. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010). doi: 10.1093/bioinformatics/btp616; pmid: 19910308
42. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018). doi: 10.1038/nbt.4096; pmid: 29608179
43. C. Trapnell *et al*., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014). doi: 10.1038/nbt.2859; pmid: 24658644
44. https://www.doi.org/10.1093/nar/gkn698
45. J. Chen, E. E. Bardes, B. J. Aronow, A. G. Jegga, ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009). doi: 10.1093/nar/gkp427; pmid: 19465376
46. Y. Zhang *et al*., Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008). doi: 10.1186/gb-2008-9-9-r137; pmid: 18798982
47. J. Ernst, M. Kellis, ChromHMM: Automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012). doi: 10.1038/nmeth.1906; pmid: 22373907
48. A. Khan *et al*., JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46** (D1), D260–D266 (2018). doi: 10.1093/nar/gkx1126; pmid: 29140473
49. PsychENCODE Consortium, Integrative Analyses of iPSC-derived Brain Organoids, Synapse (2018). doi: 10.7303/syn12033248

ACKNOWLEDGMENTS

We are grateful to the National Institute of Mental Health (NIMH), and in particular to T. Lerner and G. Senthil, for their support and guidance for this project. We thank L. Tomasini for excellent technical assistance and members of the Program in Neurodevelopment and regeneration, particularly S. Weissman, A. Szekely, J. Mariani, S. Tomasi, and others for useful comments and discussions. We thank S. Akbarian for advice on ChIP-seq. We thank G. Wang, K. Bilguvar, S. Mane, C. Castaldi, and the Yale Center for Genome Analysis for library preparation and deep sequencing. A.Ab. is also a Visiting Professor at the Child Study Center, Yale University, New Haven, CT. **Funding:** Data were generated as part of the PsychENCODE Consortium, supported by U01MH103392, U01MH103365, U01MH103346, U01MH103340, U01MH103339, R21MH109956, R21MH105881, R21MH105853, R21MH103877, R21MH102791, R01MH111721, R01MH110928, R01MH110927, R01MH110926, R01MH110921, R01MH110920, R01MH110905, R01MH109715, R01MH109677, R01MH105898, R01MH105898, R01MH094714, P50MH106934 awarded to Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke University), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (University of Southern California), Mark Gerstein (Yale University), Daniel Geschwind (University of California, Los Angeles), Fernando Goes (Johns Hopkins University), Thomas M. Hyde (Lieber Institute for Brain Development), Andrew Jaffe (Lieber Institute for Brain Development), James A. Knowles (University of Southern California), Chunyu Liu (SUNY Upstate Medical University), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Panos Roussos (Icahn School of Medicine at Mount Sinai), Stephan Sanders (University of California, San Francisco), Nenad Sestan (Yale University), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (University of California, San Francisco), Patrick Sullivan (University of North Carolina), Flora Vaccarino (Yale University), Daniel Weinberger (Lieber Institute for Brain Development), Sherman Weissman (Yale University), Kevin White (University of Chicago and Tempus Labs, Inc.), Jeremy Willsey (University of California, San Francisco), and Peter Zandi (Johns Hopkins University). Data were generated as part of the CommonMind Consortium supported by funding from Takeda Pharmaceuticals Company Limited, F. Hoffman-La Roche Ltd, and NIH grants R01MH085542, R01MH093725, P50MH066392, P50MH080405, R01MH097276, RO1-MH-075916, P50M096891, P50MH084053S1, R37MH057881 and R37MH057881S1, HHSN271201300031C, AG02219, AG05138, and MH06692. Brain tissue for the study was obtained from the following brain bank collections: the Mount Sinai NIH Brain and Tissue Repository, the University of Pennsylvania Alzheimer's Disease Core Center, the University of Pittsburgh NeuroBioBank and Brain and Tissue Repositories and the NIMH Human Brain Collection Core. CMC Leadership: Pamela Sklar, Joseph Buxbaum (Icahn School of Medicine at Mount Sinai), Bernie Devlin, David Lewis (University of Pittsburgh), Raquel Gur, Chang-Gyu Hahn (University of Pennsylvania), Keisuke Hirai, Hiroyoshi Toyoshiba (Takeda Pharmaceuticals Company Limited), Enrico Domenici, Laurent Essioux (F. Hoffman-La Roche Ltd), Lara Mangravite, Mette Peters (Sage Bionetworks), Thomas Lehner, Barbara Lipska (NIMH). **Author contributions:** The authors contributed to this study at different levels, as described in the following. Study conception: F.M.V. and A.Ab. General experimental design: F.M.V., A.Ab., G.C., and F.W. Human brain tissue procurement and processing: F.L. Human brain tissue procurement and dissection and overall project management for the PsychENCODE developmental dataset: N.S. ChIP-seq data generation from the PsychENCODE developmental dataset: S.P. and Y.S. Data generation and experimental design for iPSCs, organoid differentiation, ChIP-seq and RNA-seq assays: A.Am. and S.S. Processing of RNA-seq and ChIP-seq data: F.W., G.C., and M.G. Chromatin segmentation, Hi-C analysis and genomic variant analyses: T.R. Analysis of RNA-seq data, network analyses: G.C. Analysis of ChIP-seq data, network analyses: F.W. qPCR validation: S.S. and A.Am. Coordination of analyses, data mining and integrative analyses across datasets: F.M.V., A.Ab., F.W., and S.S. ATAC-seq assay and data analysis: A.S., L.S., and G.E.C. Rhesus interspecies DGE: A.M.M.S. and Y.Z. Display item preparation: F.M.V., A.Ab., A.Am., S.S., F.W., G.C., and T.R. Manuscript writing: F.M.V., A.Ab., A.Am., S.S., F.W., G.C., and T.R. The following authors contributed equally to the study: A.Am., S.S., F.W., G.C., T.R. All authors participated in discussion of results and manuscript editing. **Competing interests:** We declare no competing financial interests related to this article. **Data and materials availability:** Data for this manuscript will be available from Synapse (*49*).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/eaat6720/suppl/DC1
Materials and Methods
Figs. S1 to S18
Tables S1 to S17
PsychENCODE Consortium Authors and Affiliations
References (*50*–*75*)

30 March 2018; accepted 16 November 2018
10.1126/science.aat6720

RESEARCH ARTICLE SUMMARY

PSYCHIATRIC GENOMICS

Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk

Prashanth Rajarajan*, Tyler Borrman*, Will Liao*, Nadine Schrode, Erin Flaherty, Charlize Casiño, Samuel Powell, Chittampalli Yashaswini, Elizabeth A. LaMarca, Bibi Kassim, Behnam Javidfar, Sergio Espeso-Gil, Aiqun Li, Hyejung Won, Daniel H. Geschwind, Seok-Man Ho, Matthew MacDonald, Gabriel E. Hoffman, Panos Roussos, Bin Zhang, Chang-Gyu Hahn, Zhiping Weng†, Kristen J. Brennand†, Schahram Akbarian†‡

INTRODUCTION: Chromosomal conformations, topologically associated chromatin domains (TADs) assembling in nested fashion across hundreds of kilobases, and other "three-dimensional genome" (3DG) structures bypass the linear genome on a kilo- or megabase scale and play an important role in transcriptional regulation. Most of the genetic variants associated with risk for schizophrenia (SZ) are common and could be located in enhancers, repressors, and other regulatory elements that influence gene expression; however, the role of the brain's 3DG for SZ genetic risk architecture, including developmental and cell type–specific regulation, remains poorly understood.

RATIONALE: We monitored changes in 3DG after isogenic differentiation of human induced pluripotent stem cell–derived neural progenitor cells (NPCs) into neurons or astrocyte-like glial cells on a genome-wide scale using Hi-C. With this in vitro model of brain development, we mapped cell type–specific chromosomal conformations associated with SZ risk loci and defined a risk-associated expanded genome space.

RESULTS: Neural differentiation was associated with genome-wide 3DG remodeling, including pruning and de novo formations of chromosomal loopings. The NPC-to-neuron transition was defined by the pruning of loops involving regulators of cell proliferation, morphogenesis, and neurogenesis, which is consistent with a departure from a precursor stage toward postmitotic neuronal identity. Loops lost during NPC-to-glia transition included many genes associated with neuron-specific functions, which is consistent with non-neuronal lineage commitment. However, neurons together with NPCs, as compared with glia, harbored a much larger number of chromosomal interactions anchored in common variant sequences associated with SZ risk. Because spatial 3DG proximity of genes is an indicator for potential coregulation, we tested whether the neural cell type–specific SZ-related "chromosomal connectome" showed evidence of coordinated transcriptional regulation and proteomic interaction of the participating genes.

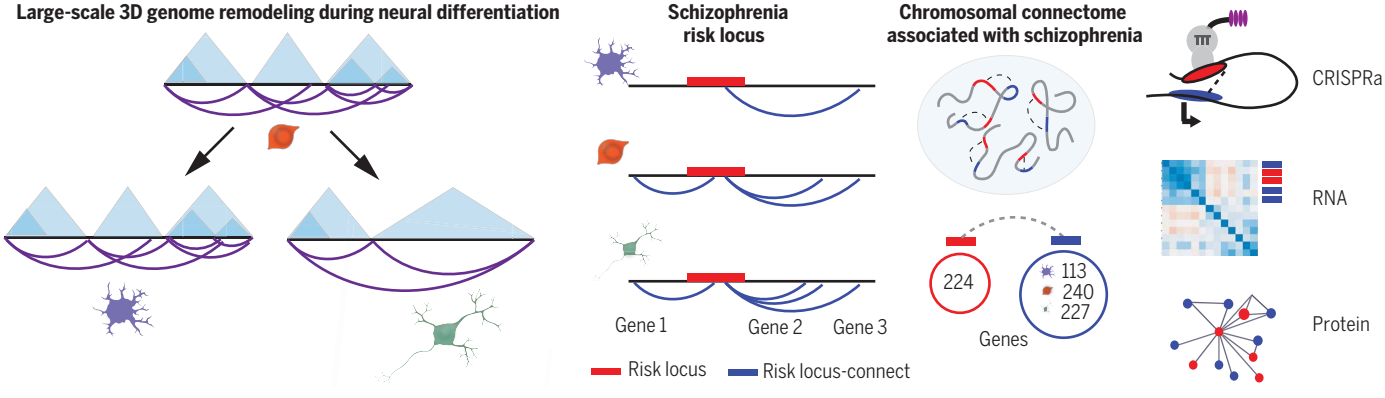
To this end, we generated lists of genes anchored in cell type–specific SZ risk-associated interactions. Thus, for the NPC-specific interactions, we counted 386 genes, including 146 within the risk loci and another 240 genes positioned elsewhere in the linear genome but connected via intrachromosomal contacts to risk locus sequences. Similarly, for the neuron-specific interactions, we identified 385 genes: 158 within risk loci and 227 outside of risk loci. Last, for glia-specific interactions, we identified 201 genes: 88 within and 113 outside of risk loci. We labeled the genes located outside of schizophrenia risk loci as "risk locus–connect," which we define as a collection of genes identified only through Hi-C interaction data, expanding—depending on cell type—by 50 to 150% the current network of known genes overlapping risk sequences that is informed only by genome-wide association studies. This disease-related chromosomal connectome was associated with "clusters" of coordinated gene expression and protein interactions, with at least one cluster strongly enriched for regulators of neuronal connectivity and synaptic plasticity and another cluster for chromatin-associated proteins, including transcriptional regulators.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat4311

CONCLUSION: Our study shows that neural differentiation is associated with highly cell type–specific 3DG remodeling. This process is paralleled by an expansion of 3DG space associated with SZ risk. Specifically, developmentally regulated chromosomal conformation changes at SZ-relevant sequences disproportionally occurred in neurons, highlighting the existence of cell type–specific disease risk vulnerabilities in spatial genome organization. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†These authors contributed equally to this work.
‡Corresponding author. Email: schahram.akbarian@mssm.edu
Cite this article as P. Rajarajan *et al.*, *Science* **362**, eaat4311 (2018). DOI: 10.1126/science.aat4311



3DG remodeling across neuronal differentiation with parallel expansion of SZ risk space. (Left) Chromatin conformation assays reveal pruning of short-range loops in neurons along with widening of TADs upon differentiation from NPCs. (Right) Cell type–specific chromatin interactions, functionally validated with CRISPR assays, expand the network of known risk-associated genes (blue circle), which show evidence for coregulation at the transcriptomic and proteomic levels.

RESEARCH ARTICLE

PSYCHIATRIC GENOMICS

Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk

Prashanth Rajarajan[1,2,3,4]*, Tyler Borrman[5]*, Will Liao[6]*, Nadine Schrode[2,3,4], Erin Flaherty[1,3,4,7], Charlize Casiño[2], Samuel Powell[1,2,3,4], Chittampalli Yashaswini[1], Elizabeth A. LaMarca[1,2,3,4], Bibi Kassim[2,4], Behnam Javidfar[2,4], Sergio Espeso-Gil[2,4], Aiqun Li[3,4], Hyejung Won[8]†‡, Daniel H. Geschwind[8], Seok-Man Ho[1,3,4], Matthew MacDonald[9], Gabriel E. Hoffman[3], Panos Roussos[2,3], Bin Zhang[3], Chang-Gyu Hahn[9], Zhiping Weng[5]§, Kristen J. Brennand[2,3,4,7]§, Schahram Akbarian[2,4]§¶

To explore the developmental reorganization of the three-dimensional genome of the brain in the context of neuropsychiatric disease, we monitored chromosomal conformations in differentiating neural progenitor cells. Neuronal and glial differentiation was associated with widespread developmental remodeling of the chromosomal contact map and included interactions anchored in common variant sequences that confer heritable risk for schizophrenia. We describe cell type–specific chromosomal connectomes composed of schizophrenia risk variants and their distal targets, which altogether show enrichment for genes that regulate neuronal connectivity and chromatin remodeling, and evidence for coordinated transcriptional regulation and proteomic interaction of the participating genes. Developmentally regulated chromosomal conformation changes at schizophrenia-relevant sequences disproportionally occurred in neurons, highlighting the existence of cell type–specific disease risk vulnerabilities in spatial genome organization.

Spatial genome organization is highly regulated and critically important for normal brain development and function (*1*). Many of the risk variants contributing to the heritability of complex genetic psychiatric disorders are located in noncoding sequences (*2*), presumably embedded in "three-dimensional genome" (3DG) structures important for transcriptional regulation, such as chromosomal loop formations that bypass linear genome on a kilobase (or megabase) scale and topologically associated domains (TADs) (*3*) that assemble in nested fashion across hundreds of kilobases (*4*–*7*). By linking noncoding schizophrenia-associated genetic variants with distal gene targets, 3DG mapping with Hi-C (*3*, *8*) and other genome-scale approaches could inform how higher-order chromatin organization affects genetic risk for psychiatric disease. To date, only a very limited number of Hi-C datasets exist for the human brain: two generated from bulk tissue of developing forebrain structures (*7*) and adult brain (*9*) and one from neural stem cells (*10*). Although such datasets have advanced our understanding of the genetic risk architecture of psychiatric disease (*7*, *11*), 3DG mapping from postmortem tissue lacks cell type–specific resolution and may not capture higher-order chromatin structures sensitive to the autolytic process (*12*). We monitored developmentally regulated changes in chromosomal conformations during the course of isogenic neuronal and glial differentiation, describing large-scale pruning of chromosomal contacts during the transition from neural progenitor cells (NPCs) to neurons. Furthermore, we uncovered an expanded 3DG risk space for schizophrenia—with a functional network of disease-relevant regulators of neuronal connectivity, synaptic signaling, and chromatin remodeling—and demonstrate neural cell type–specific coordination at the level of the chromosomal connectome, transcriptome, and proteome.

[1]Icahn School of Medicine M.D./Ph.D. Program, Icahn School of Medicine at Mount Sinai, New York, NY 10027, USA. [2]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10027, USA. [3]Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, New York, NY 10027, USA. [4]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10027, USA. [5]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA. [6]New York Genome Center, New York, NY 10013, USA. [7]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, New York, NY 10027, USA. [8]Neurogenetics Program, Department of Neurology, Center for Autism Research and Treatment, Semel Institute, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA 90095, USA. [9]Neuropsychiatric Signaling Program, Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA.
*These authors contributed equally to this work. †Present address: Department of Genetics, University of North Carolina, Chapel Hill, NC 27599, USA. ‡Present address: UNC Neuroscience Center, University of North Carolina, Chapel Hill, NC 27599, USA. §These authors contributed equally to this work.
¶Corresponding author. Email: schahram.akbarian@mssm.edu

Results

Neural progenitor differentiation is associated with dynamic 3DG remodeling

We applied in situ Hi-C to map the 3DG of two male human induced pluripotent stem cell (hiPSC)–derived neural progenitor cells (NPCs) (*13*), together with isogenic populations of induced excitatory neurons ("neuron") generated through viral overexpression of the transcription factor *NGN2* (*14*) and differentiations of astrocyte-like glial cells ("glia") (Fig. 1, A and B, and table S1) (*15*). Transcriptome RNA sequencing (RNA-seq) comparison with published datasets (*16*) confirmed that the NPCs, but not glia, from subjects S1 and S2 clustered together with NPCs from independent donors, whereas S1 and S2 *NGN2* neurons closely aligned with directed differentiation forebrain neurons (*17*) and prenatal brain datasets (fig. S1, A and B). As with our transcriptomic datasets, hierarchical clustering of our Hi-C datasets after initial processing (fig. S2A) also showed clear separation by cell type (Fig. 1A and fig. S2B). Genome-scale interaction matrices were enriched for intrachromosomal conformations (fig. S2C), with the exception of the negative control ("No Ligase") NPC library, in which we omitted the ligase step (Materials and methods) and observed an interaction map with no signal due to the loss of chimeric fragments (fig. S2D). Given the observed correlation between technical replicates of Hi-C assays from the same donor and cell type, and the correlation between cell type–specific Hi-C from the two donors (Pearson correlation of PC1, $R_{\text{technical replicates}}$, range = 0.970 to 0.979; $R_{\text{subject1-subject 2 by cell type}}$, range = 0.962 to 0.970), we pooled by cell type for subsequent analyses (fig. S2E).

We first focused on intrachromosomal loop formations, which are conservatively defined as distinct contacts between two loci in the absence of similar interactions in the surrounding sequences (*3*). Our comparative analyses included published (*3*) in situ Hi-C data from the B lymphocyte–derived cell line GM12878 (table S1). When analyzed with the HiCCUPS pipeline (5- and 10-kb loop resolutions combined, subsampled to 372 million valid-intrachromosomal read pairs to reflect the library with the fewest reads after filtration) (*3*), 17,767 distinct loops were called: n = 3118 (17.5%) were shared among all four cell types, whereas n = 5068 (28.5%) were specific to only one of the four cell types (Fig. 1C). Biologically relevant terms such as "central nervous system development," "forebrain development," and "neuron differentiation" were among the top gene ontology (GO) enrichments from genes overlapping loops shared between NPCs, glia, and neurons (brain-specific) but not identified in lymphocytes (Fig. 1D and table S2), indicating strong tissue-specific loop signatures that were also confirmed in individual cell types (fig. S3A and tables S3 to S6).

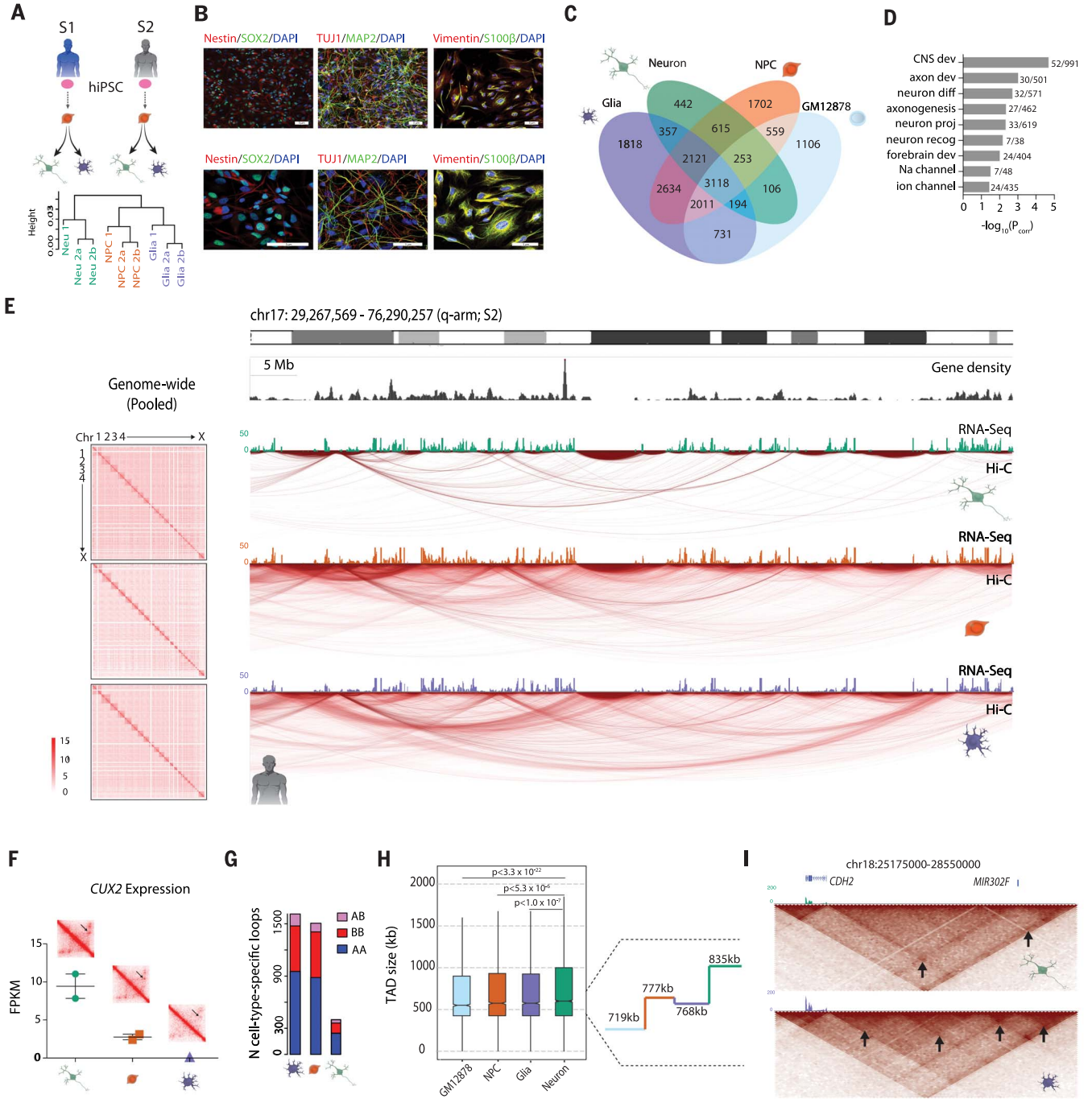


Fig. 1. Neural differentiation is associated with large-scale remodeling of the 3D genome. (**A**) (Top) Derivation scheme of isogenic cell types from two male control cell lines. Pink oval, donor hiPSC; orange, NPC; green, neuron; purple, glia. (Bottom) Hierarchical clustering of intrachromosomal interactions (Materials and methods) from six in situ Hi-C libraries. a and b are technical replicates of the same library; height corresponds to the distance between libraries (Materials and methods) (fig. S2B). (**B**) Immunofluorescent staining of characteristic cell markers for NPCs (Nestin and SOX2), neurons (TUJ1 and MAP2), and glia (Vimentin and S100β). (**C**) Venn diagram of loop calls specific to and shared by different subsets of cells, including previously published GM12878 lymphoblastoid Hi-C data. (**D**) Gene ontology (GO) enrichment (significant terms only) of genes overlapping anchors of loops shared by NPCs, neurons, and glia but absent in GM12878. (**E**) (Left) Cell-type pooled whole-genome heatmaps at 500-kb resolution (fig. S2C). (Right) "Arc map" showing intrachromosomal interactions at 40-kb resolution of the q-arm of chr17 for isogenic neurons, NPCs, and glia, as indicated, from subject 2. RNA-seq tracks for each cell type shown on top of arc maps. Green, neuron; orange, NPC; purple, glia. (**F**) FPKM gene expression of *CUX2* across three cell types with heatmap zoomed in on *CUX2* loop (black arrow) (fig. S3). (**G**) Number of loops specific to each cell type (not shared with other cell types) with one anchor in an A compartment and another in a B compartment (pink), both in B compartments (red), or both in A compartments (blue). (**H**) (Left) Box-and-whisker distribution plot of TAD size across four cell types. (Right) Median TAD length for each of the four cell types. (**I**) Heatmaps at 40-kb resolution for a 3-Mb window at the *CDH2* locus on chr18. (Bottom) Nested TAD landscape in glia with multiple subTADs (black arrows) called, which (top) is absent from neuronal Hi-C. RNA-seq tracks: green, neuron; purple, glia (figs. S1 to S5).

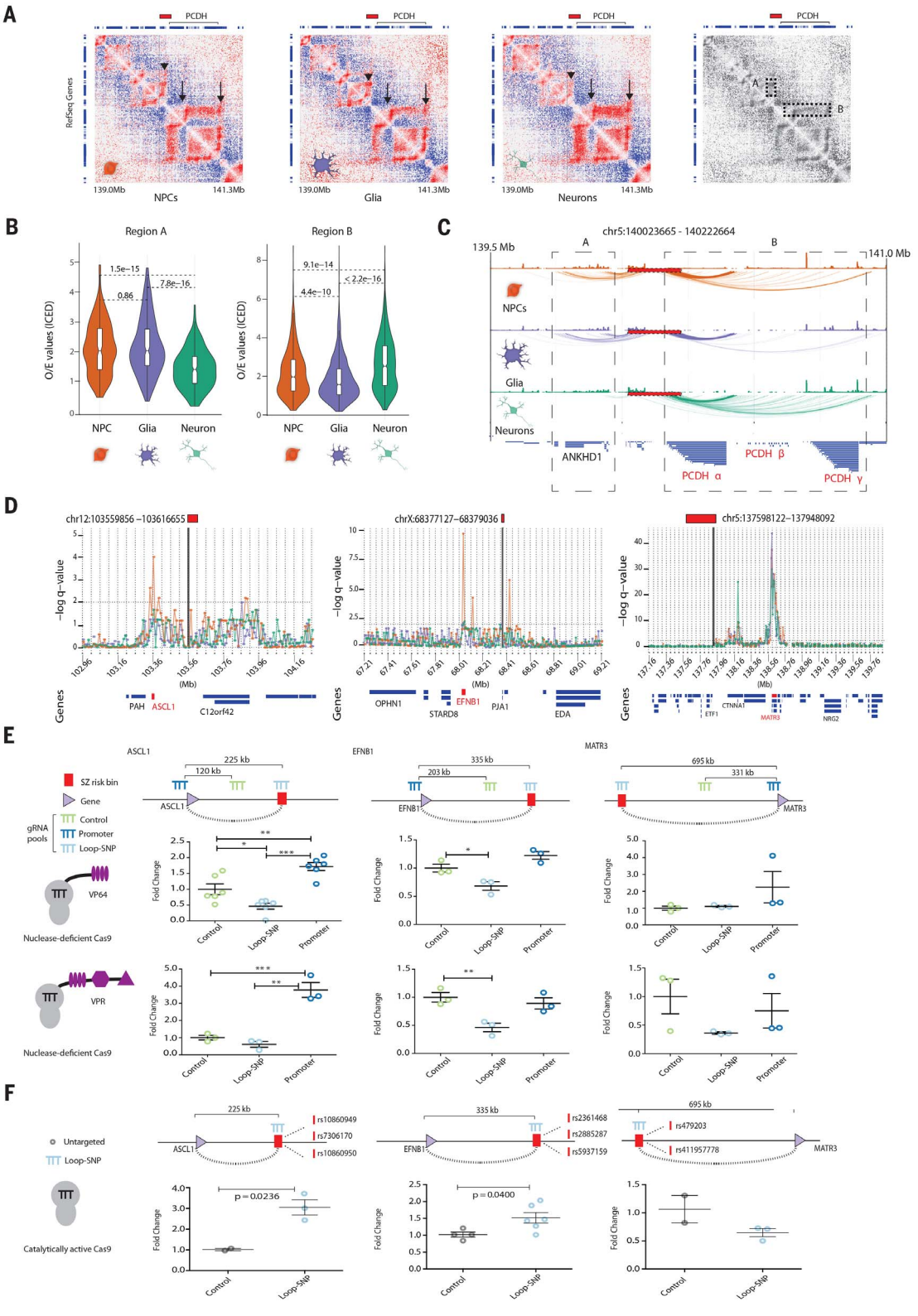


Fig. 2. Cell type–specific chromosomal contact maps at schizophrenia risk loci. (**A**) Juicebox observed/expected interaction heatmaps at 10-kb resolution for the risk-associated clustered *PCDH* locus chr5:140023665–140222664 for NPC, glia, and neurons as indicated. (Far right) Grayscale heatmap depicts areas of highly cell-specific contact enrichments: upstream genes including *ANKHD1* (dotted rectangle "A" and arrowhead) and downstream *PCDH* gene clusters (dotted rectangle "B" and arrows). Clustered *PCDH* gene expression patterns are available in fig. S6A. (**B**) Violin plots of observed/expected interaction values in the regions A and B highlighted in (A). (**C**) Map of contacts identified by binomial statistics. Red box with dashed black line represents the schizophrenia risk locus, dotted boxes regions "A" and "B" in heatmaps. (**D**) Cell-type resolved contact map of 10-kb bins (bold, black vertical lines) within risk sequences on chr12 (left), chrX (middle), and chr5 (right); NPC (orange), neuron (green), glia (purple); −log *q* value, significance of contact between schizophrenia risk locus and each 10-kb bin; gene models ("Genes") below with SNP-loop target gene highlighted in red. (**E**) Epigenomic editing (CRISPRa with nuclease-deficient dCas9 in NPCs) for three risk SNP-target gene pairs and their respective control sequences (top), measured with quantitative reverse transcription polymerase chain reaction (RT-PCR) (fold change from baseline) for VP64 (middle) and VPR (bottom) transcriptional activators. (**F**) Quantitative PCR gene expression changes upon directing catalytically active Cas9 to schizophrenia risk-associated credible SNPs (vertical red dashes with rsIDs) interacting via chromosomal contacts with promoters of *ASCL1*, *EFNB1*, and *MATR3* in NPCs. Targeting strategy and contact distances depicted above; **P* < 0.05, ***P* < 0.01, ****P* < 0.0001 (figs. S6 and S7).

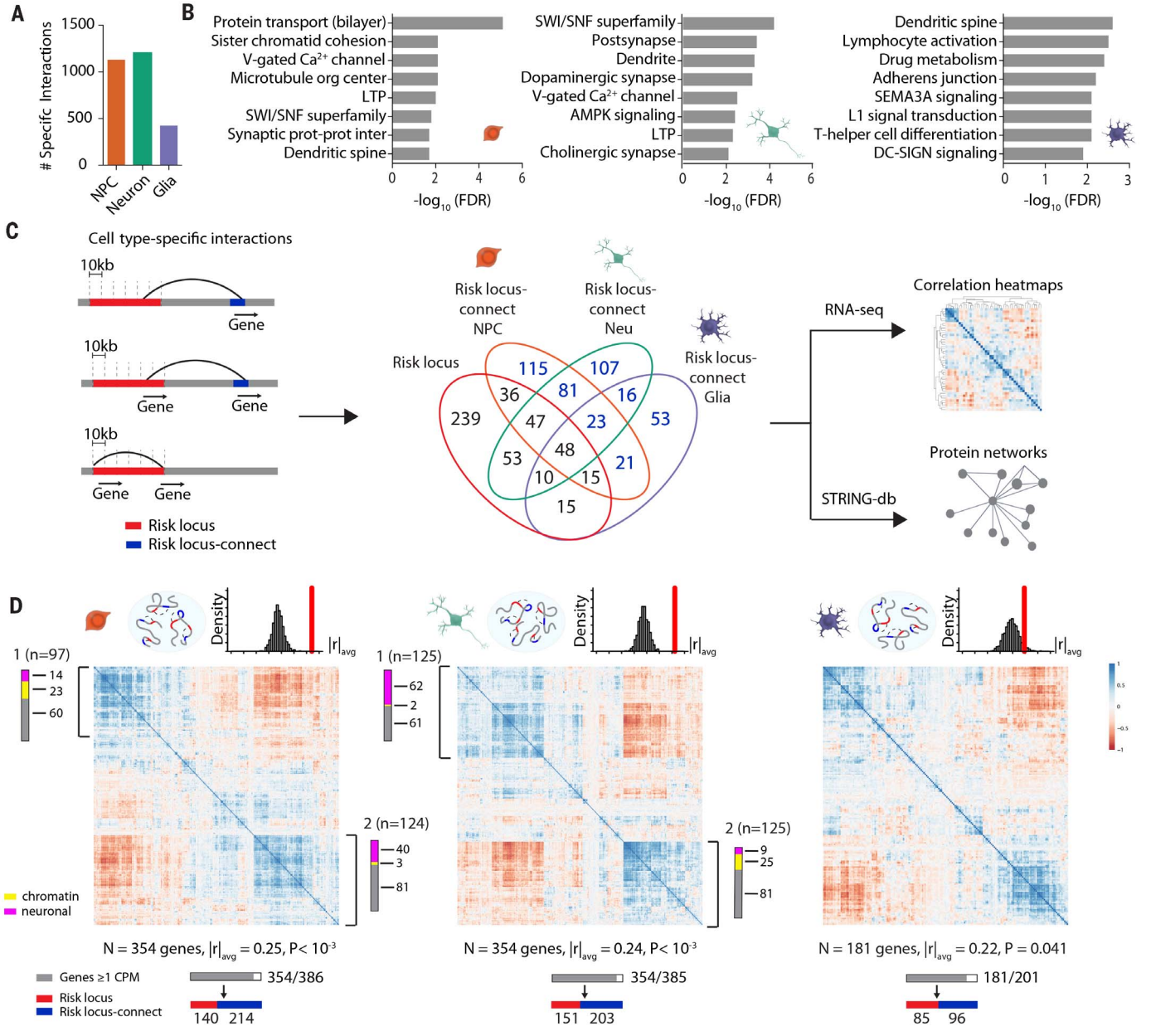


Fig. 3. Expanded GWAS risk connectome is associated with gene coregulation. (**A**) Counts of highly cell type–specific contacts associated with schizophrenia risk in each of the three hiPSC-derived cell types. (**B**) GO enrichment of genes located in schizophrenia risk contacts in NPCs (left), neurons (middle), and glia (right). (**C**) (Left) Schematic workflow of analyses performed with cell type–specific contact genes, distinguished as "risk locus" and "risk locus–connect" genes. (Middle) Venn diagram of genes located in the 145 loci and those found in cell type–specific contacts, with numbers in blue indicating "risk locus–connect" genes. (Right) Schematic workflow of analyses performed with combined set of "risk locus" and "risk locus–connect" genes. (**D**) RNA Pearson transcriptomic correlation heatmaps consisting of risk locus and risk locus–connect genes derived from the cell type–specific contacts of NPCs (left), neurons (middle), and glia (right). Organization scores ($|r|_{avg}$) for each heatmap are reported with *P* values from sampling analysis. Schematics above heatmaps are representations of each cell type's particular connectome (blue oval) and frequency distribution of organization scores from permutation analyses of randomly generated heatmaps (red, observed organization score of heatmap being tested). The gray bar corresponds to *n* genes that have at least 1 count per million in RNA-seq dataset out of the total number of genes and are used to construct the heatmap; red and blue bars indicate how many of the genes in the heatmap are in a risk locus (red) and are risk locus–connect (blue). Fuchsia, neuron connectivity/synaptic function genes; yellow, chromatin remodeling genes as determined from gene ontology analysis in (**E**). Additional information on coexpression clusters is provided in tables S22 and S23 (figs. S8 and S9).

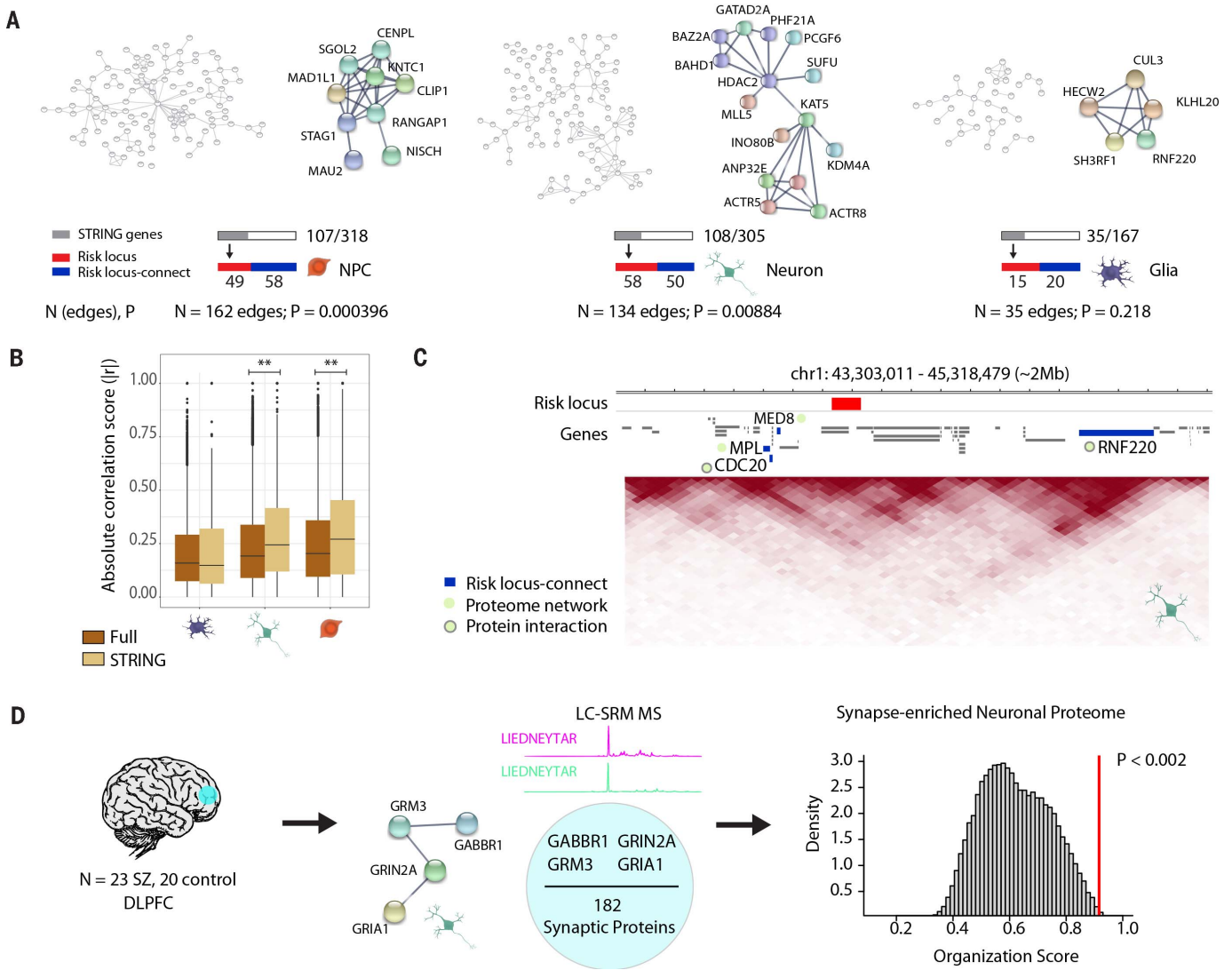


Fig. 4. Expanded GWAS risk connectome is linked to protein-protein association networks. (**A**) Overview and representative examples (zoomed in) of protein-protein association networks in NPCs (left), neurons (middle), and glia (right). Numbers of edges connecting the proteins in each network and STRING-computed *P* values are reported below. Gray bar indicates the subset of these genes whose proteins are involved in the network out of the total number of genes from cell type–specific interactions; red and blue bars indicate how many of the genes in the network are in a risk locus (red) and are risk locus–connect (blue). (**B**) Comparison of organization scores between the full RNA transcriptomic correlation heatmaps (brown) (Fig. 3D) and the "STRING" heatmaps (tan) (figs. S13 to S15), consisting of only those genes in protein networks for each cell type. Permutation test, $^{**}P < 0.01$. (**C**) Representative neuronal TAD landscape (chr1, ~2 Mb) depicting a schizophrenia risk–associated locus (red) with its risk locus–connect genes (blue), *MED8*, *MPL*, *CDC20*, and *RNF220*, which are members of the neuronal schizophrenia protein network (green circle). *CDC20* and *RNF220* interact at the protein level (green circle with gray border). (**D**) (Left) Liquid chromotography–selected reaction monitoring (LC-SRM) mass spectrometry (MS) was performed on dorsolateral prefrontal cortex (DLPFC) tissue from 43 adult postmortem brains (23 schizophrenia, 20 control). (Middle) 182 neuronal proteins were reliably quantified, and four of them were observed to have associations in the neuron protein network in (A). (Right) GABBR1, GRM3, GRIN2A, and GRIA1 proteins were found to have significantly more correlated expression than expected by random permutation analysis. Additional information on protein-protein interactions is provided in figs. S9 to S15.

Unexpectedly, there was a reduction (~40 to 50% decrease) in the total number of chromosomal loops in neurons relative to isogenic glia and NPCs (fig. S3, B and C). Reduced densities of chromosomal conformations were also evident in genome browser visualization of chromosomal arms, including chr17q (Fig. 1E). Although both glia and NPCs harbored ~13,000 loop formations, only 7206 were identified in neurons (Fig. 1C; fig. S3, B and C; and table S1), including 442 neuron-specific loop formations. One such neuron-specific loop was at *CUX2*, a transcription factor whose expression marks a subset of cortical projection neurons (*18*) and that is highly expressed in our *NGN2*-induced neurons (Fig. 1F and fig. S3, D and E). Examples of loops lost in neurons include one spanning the Ca^{2+} channel and dystonia-risk gene, *ANO3* (fig. S3F) (*19*). Furthermore, NPCs, neurons, and glia had similar proportions of loops anchored in solely active (A) compartments, solely inactive (B) compartments, or in both, indicating no preferential loss of either active or inactive loops in neurons (Fig. 1G). However, among the genes overlapping anchors of loops that underwent pruning during the course of the NPC-to-neuron transition, regulators of cell proliferation, morphogenesis, and neurogenesis ranked prominently in the top 25 GO terms with significant enrichment (Benjamini-Hochberg corrected $P < 10^{-6} - 10^{-12}$) (fig. S3G and table S4B), which is consistent with a departure from precursor stage toward postmitotic neuronal identity (*20*). Likewise, loops lost during

NPC-to-glia transition were significantly enriched (Benjamini-Hochberg corrected $P < 10^{-3} - 10^{-6}$) for neuron-specific functions, including "transmission across chemical synapse," "γ-aminobutyric acid (GABA) receptor activation," and "postsynapse" (fig. S3G and table S4C), which is consistent with non-neuronal lineage commitment.

We defined "loop genes" as genes that either have gene body or transcription start site (TSS) overlap with a loop anchor (5- or 10-kb bins forming the points of contact in a chromatin loop). Genes with loop-bound gene bodies (one-tailed Z test, $Z_{range} = 42.1$ to 59.2, $P < 10^{-324}$ for all) or loop-bound TSS (one-tail Z-test, $Z_{range} = 15.2$ to 28.8, $P_{range} < 2.32 \times 10^{-52}$ to 4.40×10^{-182}) both showed significantly greater expression [mean $\log_{10}$(FPKM + 1); FPKM, fragments per kilobase of exon per million fragments mapped] than that of background (all genes for all brain cell types) (fig. S4A), suggesting that looping architecture was associated with increased gene expression. Furthermore, 3% of loops shared by NPCs, neurons, and glia (brain-specific loops) interconnected a brain expression quantitative trait locus (eQTL) single-nucleotide polymorphism (SNP) with its destined target gene(s), representing significant enrichment over background as determined with 1000 random distance- and functional annotation–matched loop samplings, (random sampling, one-sided empirical $P = 0.012$) (Materials and methods) (fig. S4B).

We aimed to confirm that the observed net loss of loop formations during the NPC-to-neuron transition could be replicated across a variety of independent cell culture and in vivo approaches and was not specific to our methodological choice of *NGN2*-induction. We conducted an additional Hi-C experiment on cells differentiated from hiPSC-NPCs by means of a non-*NGN2* protocol that used only differentiation medium and yielded a heterogeneous population of hiPSC-forebrain-neurons in addition to a small subset of glia (*17*). In addition, we reanalyzed Hi-C datasets generated from a mouse model of neural differentiation, consisting of mouse embryonic stem cell (mESCs), mESC-derived NPCs (mNPC), and cortical neurons (mCN) differentiated from the mNPCs via inhibition of the Sonic Hedgehog (SHH) pathway (*21*). To examine whether such genome-wide chromosomal loop remodeling also occurred in the developing brain in vivo, we reanalyzed Hi-C data from human fetal cortical plate (CP), mostly composed of young neurons, and forebrain germinal zone (GZ), primarily harboring dividing neural precursor cells in addition to a smaller subset of newly generated neurons (*7*). Across both the hiPSC-NPC-to-forebrain neuron and mESC-mNPC-mCN differentiation, in vitro neurons showed a 20% decrease in loops compared with their neural progenitors (fig. S4, C and D). Consistent with this, in vivo CP (neuron) compared with GZ (progenitor) showed a 13% decrease in loops genome-wide (fig. S4E). The highly replicative cell types included here, mouse ESCs and human lymphoblastoid GM12878 cells, exhibited loop numbers very similar to their neuronal counterparts (fig. S4, D and E), suggesting that the changes in 3DG architecture from NPC to neurons do not simply reflect a generalized effect explained by mitotic potential.

Along with having fewer total loops, neurons exhibited a greater proportion of longer-range (>100 kb) loops than did NPCs or glia (two-sample two-tailed Kolmogorov-Smirnov test, $KS_{range} = 0.1269$ to 0.2317, $P < 2.2 \times 10^{-16}$ for three comparisons: Neu versus NPC/Glia/GM) (fig. S5A). Likewise, in each of the alternative in vitro and in vivo analyses considered above, neurons exhibited a greater proportion of longer-range (>100 kb) loops than did NPCs or glia [two-sample two-tailed Kolmogorov-Smirnov test, $KS = 0.0427$, $P = 1.5 \times 10^{-5}$ for hiPSC-NPC versus forebrain neuron; $KS = 0.0936$, $P = 1.1 \times 10^{-16}$ for mESC-NPC versus mCN; $KS = 0.0663$, $P = 2.04 \times 10^{-8}$ for fetal CP (neuron) compared with GZ (progenitor)] (fig. S5, B, C, D, and E). Therefore, multiple in vitro and in vivo approaches comparing, in human and mouse, neural precursors to young neurons consistently show a reduced number of loops in neuron-enriched cultures and tissues, primarily affecting shorter-range loops.

Consistent with studies in peripheral tissues reporting conservation of the overall loop-independent TAD landscape across developmental stages, tissues, and species (when considering syntenic loci) (*10*, *22*), overall TAD landscapes (*3*) remained similar between neurons, glia, and NPCs. Nonetheless, TADs also showed a subtle (~10%) increase in average size in neurons compared with isogenic NPCs, independent of the differentiation protocol applied (Wilcoxon-Mann-Whitney test, $P < 5.3 \times 10^{-6}$) (Fig. 1H and fig. S5, F and G), as highlighted here at a 3.4-Mb TAD at the *CDH2* cell adhesion gene locus (Fig. 1I). TAD remodeling may therefore reflect restructuring of nested subdomains within larger neuronal TADs (tables S7 and S8). To examine whether such developmental reorganization of the brain's spatial genomes was associated with a generalized shift in chromatin structure, we applied the assay for transposase accessible chromatin with high-throughput sequencing (ATAC-seq) to map open chromatin sequences before and after *NGN2*-neuronal induction (table S1). Genome-wide distribution profiles for transposase-accessible chromatin were only minimally different between NPCs and neurons (fig. S5H) and further revealed that both NPCs and neurons showed low to moderate chromatin accessibility [$-2.5 < \log_2(ATAC\ signal) < 1$] for ≥89% of the anchor sequences comprising cell type–specific and shared "brain" loops in our cell culture system (fig. S5I). These findings, taken together, point to widespread 3DG changes during the NPC-to-neuron transition and NPC-to-glia transition in human and mouse brain that are unlikely attributable to global chromatin accessibility differences. This includes highly cell type–specific signatures in gene ontologies of differentiation-induced loop prunings, reflecting neuronal and glial (non-neuronal) lineage commitment (fig. S3, A and G, and table S4, B and C), and a subtle widening of average loop and TAD length in young neurons (Fig. 1H and fig. S5, A to G).

Chromosomal contacts associated with schizophrenia risk sequences

Because many schizophrenia risk variants lie in noncoding regions in proximity to several genes, we predicted that chromosomal contact mapping could resolve putative regulatory elements capable of conferring schizophrenia risk via their physical proximity (bypassing linear genome) to the target gene, as has been demonstrated in tissue in vivo (*7*, *11*). We overlaid our cell type–specific interactions onto the 145 risk loci associated with schizophrenia risk (*2*, *23*). Because only very few loops (defined as distinct pixels with greater contact frequency than neighboring pixels on a contact map) (*3*) were associated with schizophrenia risk loci (n = 212, 81, and 17 loops in NPC, glia, and neurons, respectively) (table S9), we applied an established alternative approach to more comprehensively explore the 3DG in context of disease-relevant sequences (*7*). This approach defines interactions as those filtered contacts that stand out over the global background and applies binomial statistics to identify chromosomal contacts anchored at disease-relevant loci (*7*). To begin, we examined the 40 loci with strongest statistical evidence for colocalization of an adult postmortem brain eQTL and schizophrenia genome-wide association study (GWAS) signal (*24*). Chromosomal contacts were called for 29 of the 46 eQTLs present in the 40 loci, with 8 of 29 (28%) of the loci showing significant interactions (binomial test, $-\log q$ value range = 1.33 to 11.0) between the eQTL-SNPs (eSNPs) in the one contact anchor and the transcription start site of the associated gene(s) in the other anchor (table S10). We conclude that ~30% of risk locus–associated eQTLs with strong evidence for colocalization with GWAS signal bypass the linear genome and are in physical proximity to the proximal promoter and transcription start site of the target gene, resonating with previous findings in fetal brain tissue that used a similar contact mapping strategy (*7*).

Cell type–specific contact maps with 10-kb-wide bins, queried for the schizophrenia-associated loci, frequently revealed differential chromosomal conformations in NPCs, glia, and neurons. For example, the risk locus upstream of the *PROTOCADHERIN* cell adhesion molecule gene clusters (chromosome 5), which is critically relevant for neuronal connectivity in developing and adult brain (fig. S6A) (*25*, *26*), showed through both observed/expected interaction matrix (*27*) and global background-filtered contact mapping (*7*) a bifurcated bundle of interactions in NPCs, with one bundle emanating to sequences 5′ and the other bundle to sequences 3′ from the locus. In neurons, the 3′ bundle was maintained, but the 5′ bundle was "pruned," whereas glia showed the opposite pattern; these differences between the three cell types were highly significant (observed/expected Wilcoxon

rank sum $P < 10^{-9}$ to 10^{-15}) (Fig. 2, A to C). Dosage of the noncoding schizophrenia risk-SNP (rs111896713) at the *PCDH* locus significantly increased the expression of multiple *PROTOCADHERIN* genes (*PCDHA2*, *PCDHA4*, *PCDHA7*, *PCDHA8*, *PCDHA9*, *PCDHA10*, and *PCDHA13*) in adult frontal cortex of a large cohort of 579 individuals, including cases with schizophrenia and controls (fig. S6B and table S11) (*28*). The affected genes were interconnected to the disease-relevant noncoding sequence in neurons and NPCs but not in glia (fig. S6C). Therefore, cell type–specific Hi-C identified chromosomal contacts anchored in schizophrenia-associated risk sequences that affected expression of the target gene(s). On the basis of earlier chromosome conformation capture assays at the site of candidate genes, the underlying mechanisms may include alterations in transcription factor and other nucleoprotein binding at loop-bound cis-regulatory elements (*5*) or even local disruption of chromosomal conformations (*6*).

Transcriptional profiles of hiPSC-derived NPCs and neurons most closely resemble those of the human fetus in the first trimester (*29*); moreover, a portion of the genetic risk architecture of schizophrenia matches to regulatory elements that are highly active during prenatal development (*30*). We surveyed in our Hi-C datasets seven loci encompassing 36 "credible" (potentially causal) schizophrenia-risk SNPs with known chromosomal interactions in fetal brain to genes important for neuron development and function (*7*). We found that risk-associated chromosomal contacts were conserved between our hiPSC-NPCs and the published human fetal CP and germinal zone Hi-C datasets (*7*) for five of the seven loci (71%) tested (*CHRNA2*, *EFNB1*, *MATR3*, *PCDH*, and *SOX2*, but not *ASCL1* or *DRD2*) (table S12). To test the regulatory function of these conserved risk sequence-bound conformations, we performed single-guide RNA (sgRNA)–based epigenomic editing experiments on isogenic antibiotic-selected NPCs that stably express nuclease-deficient dCas9-VP64 (*31*, *32*) or dCas9-VPR (*33*, *34*) transactivators (table S13). Previous studies in peripheral cell lines succeeded in inducing gene expression changes by placing dCas9-repressor fusion proteins at the site of chromosomal contacts separated by up to 2 Mb of linear genome from the promoter target (*35*). We tested *ASCL1*-, *EFNB1*-, *MATR-3*, and *SOX2*-bound chromosomal contacts separated by 200- to 700-kb interspersed sequences (Fig. 2, D and E; fig. S7A; and table S14). Pools of five individual sgRNAs directed against a risk-associated noncoding sequence bypassing 225 and 355 kb of genome consistently resulted in significantly decreased expression of *ASCL1* [one-way analysis of variance (ANOVA), $F_{VP64}(2, 15) = 22.20$, $P < 0.0001$; Dunnett's $P_{VP64} = 0.023$] and *EFNB1* target genes [one-way ANOVA, $F_{VP64}(2, 6) = 14.47$, $P = 0.0051$, Dunnett's $P_{VP64} = 0.0356$; $F_{VPR}(2, 6) = 1.46$, $P = 0.0111$, Dunnett's $P_{VPR} = 0.0088$], in comparison with positive (promoter-bound) and negative (linear genome) control sgRNAs. Epigenomic editing of risk sequence 500 to 600 kb distant from the *SOX2* and *MATR3* loci did not alter target gene expression (Fig. 2, D and E, and fig. S7, A and B), which could reflect practical limitations in nonintegrative transfection-based (as opposed to viral) methods, impact of epigenetic landscape, or suboptimal guide RNA positioning (*34*), further limited by the 10-kb contact map resolution. Because portions of the *MATR3*-bound risk sequences are embedded in repressive chromatin, we directed five sgRNAs for Cas9 nuclease mutagenesis toward a 138–base pair (bp) sequence within a *MATR3* long-range contact that was enriched with trimethyl-histone H3K27me3, commonly associated with *Polycomb* repressive chromatin remodeling, in order to disrupt it (fig. S7, C to E). This strategy produced a significant increase in *MATR3* expression upon ablation of the putative repressor sequence, whereas targeting *MATR3* (linear genome) control sequence remained ineffective (fig. S7, D and E). We conducted additional genomic mutagenesis assays, with sgRNAs directly overlapping credible SNPs participating in chromatin contacts with *ASCL1*, *EFNB1*, *EP300*, *MATR3*, *PCDHA7*, *PCDHA8*, and *PCDHA10* (table S10). Cas9 nuclease deletion of interacting credible SNPs significantly increased gene expression of *ASCL1*, *EFNB1*, and *EP300* ($P_{range} = 0.0053$ to 0.04, $t_{range} = 2.449$ to 4.265) (Fig. 2F and fig. S7F). Similar targeting of four credible SNPs upstream of the clustered *PCDH* locus significantly decreased levels, by ~50 to 60%, of *PCDHA8* and *PCDHA10* ($P_{range} = 0.0122$ to 0.0124, $t_{range} = 4.326$ to 4.343), two of the genes whose expression increased with dosage of the risk SNP rs111896713 in adult postmortem brain (figs. S6C and S7G). Taken together, our (epi)genomic editing assays (fig. S7H) demonstrate that chromosomal contacts anchored in schizophrenia risk loci potentially affect target gene expression across hundreds of kilobases, which is consistent with predictions from chromosomal conformation maps from hiPSC-derived brain cells described here, and from developing (*7*, *11*) and adult (*5*) human brain tissue.

Cell type–specific schizophrenia-related chromosomal connectomes are associated with gene co-regulation and protein-protein association networks

Having shown that the chromosomal contact maps anchored in sequences associated with schizophrenia heritability undergo cell type–specific regulation (Fig. 2, A to C), are reproducible in neural cell culture and fetal brain (table S12), frequently harbor risk-associated eQTLs (table S10), and bypass extensive stretches of linear genome to affect target gene expression in genomic and epigenomic editing assays (Fig. 2, D to F, and fig. S7), we investigated chromosomal contacts for all 145 GWAS-defined schizophrenia risk loci together (*23*) (tables S15 to S17). We refer to the resulting "network" of risk loci and their 3D proximal genes as the "schizophrenia-related chromosomal connectome."

Earlier studies in adult brain had shown that open chromatin-associated histone modification and other "linear epigenome" mappings strongly link the genetic risk architecture of schizophrenia specifically with neuronal, as opposed to non-neuronal, chromatin (*36*), which would suggest that similar cell-specific signatures may emerge in the risk-associated 3DG. Neurons and NPCs, but not the isogenic glia, showed a high preponderance of chromosomal contacts with schizophrenia-associated risk loci (Fig. 3A). There were 1203 contacts involving schizophrenia risk sequences that were highly specific to neurons (median distance between risk and target bins = 510 kb), 1100 highly specific for NPCs (median distance between risk and target bins = 520 kb), whereas only 425 highly specific for glia (median distance between risk and target bins = 580 kb) (Fig. 3A; figs. S8, A and B; and tables S15 to S17). There were also unexpectedly robust cell type– and gene ontology–specific signatures, including genes associated with neuronal connectivity and synaptic signaling (Fig. 3B and tables S18 and S19). Separate analysis of the Psychiatric Genomics Consortium "PGC2" 108 risk loci (*2*) yielded similar results (fig. S9, A and B).

Because spatial 3DG proximity of genes is an indicator for potential coregulation (*37*), we tested whether the neural cell type–specific schizophrenia-related chromosomal connectome showed evidence of coordinated transcriptional regulation and proteomic interaction of the participating genes. To this end, we generated lists of genes anchored in the most highly cell type–specific schizophrenia risk–associated contacts (Materials and methods) (Fig. 3C, fig. S8B, and table S18). Thus, for the NPC-specific contacts, we counted 386 genes, including 146 within the risk loci and another 240 genes positioned elsewhere in the linear genome but connected via an intrachromosomal contact to within-risk-locus sequences. Similarly, for the neuron-specific contacts, we identified 385 genes, including 158 within risk loci and 227 outside of risk loci (Fig. 3C). Last, for glia-specific contacts, we identified 201 genes, including 88 within and 113 outside of risk loci. We labeled the intrachromosomal contact genes located outside of schizophrenia risk loci as "risk locus-connect," which we define as a collection of genes identified only through Hi-C interaction data, expanding—depending on cell type—by 50 to 150% the current network of known genes overlapping risk sequences that is informed only by GWAS (Fig. 3C).

To examine whether such types of disease-associated, cell type–specific chromosomal connectomes were linked to a coordinated program of gene expression, we analyzed a merged transcriptome dataset (comprised of 47 hiPSC-NPC and 47 hiPSC-forebrain-neuron RNA-seq libraries from 22 schizophrenia and control donors not related to the those of our Hi-C datasets) (*16*). We examined pair-wise correlations of the collective sets of the 386 NPC, 385 neuron, and 201 glia genes representing "risk locus" and "risk locus–connect" genes (cell type–specific "risk connectomes"). The risk connectome for each cell type showed extremely strong pair-wise

correlations, with two of the largest clusters visualized on the neuron and NPC correlation matrices involving an admixture of 354 "risk locus" and "risk locus–connect" genes each, and similarly 181 genes from the glia matrix (Fig. 3D and table S20). The averaged gene-by-gene transcript correlation index for each matrix overall, defined here as "organization score" ($|r|_{avg}$), for the NPCs, neurons, and glia were 0.22 to 0.25. Such levels of organized gene expression were robustly significant for NPC and neurons, after controlling for linear genomic distance (1000 random samplings, $|r|_{avg}$, $P < 0.001$ for NPC and for neuron; $P = 0.041$ for glia) (Fig. 3D, fig. S9E, and table S21). There were four large clusters in the correlation matrices of the neuronal and NPC risk connectome: neuronal connectivity and synaptic signaling proteins (neuron cluster 1 and NPC cluster 2) and epigenetic regulators (neuron cluster 2 and NPC cluster 1). For example, within neuron cluster 1 (Fig. 3D, middle), 62 of 125 genes encoded neural cell adhesion and synaptic molecules, voltage-gated ion channels, and other neuron-specific genes (Fig. 3E and tables S22 and S23). We thus conclude that the chromosomal connectomes associated with schizophrenia risk are cell type–specific, with the neuronal risk connectome particularly enriched for genes pertaining to neuronal connectivity, synaptic signaling, and chromatin remodeling (Fig. 3, D and E). Analyses of the subset of PGC2 risk loci (108 and 145) provided similar results (fig. S9, C to F). Additionally, organization scores for neuron cluster 1 and cluster 2 genes were similar between hiPSC-derived NPCs and forebrain neurons from schizophrenia cases ($n = 47$) and control ($n = 47$), suggesting that many risk locus–connect and risk locus genes are coregulated across individuals (fig. S9H).

Numerous proteins encoded by risk locus and risk locus–connect genes were associated with synaptic signaling (table S24). The cell type–specific risk locus–connect and risk locus genes show significant protein-protein interaction network effects for NPCs ($P = 0.0004$) and neurons ($P = 0.009$) but not glia (Fig. 4A, figs. S10 to S12, and table S24) when examined by using the STRING database v10.5 (*38*, *39*). We observed many proteomic clusters, including large groups of epigenomic regulators associated with the SWI/SNF (SWItch/Sucrose NonFermentable) chromatin remodeling complex and histone lysine methyltransferases and demethylases (Fig. 4A and figs. S10 and S11), many of which were the genes identified in NPC cluster 1 and neuron cluster 2 of the transcriptome analysis (Fig. 3, D and E). The transcriptomic correlation heatmaps for these protein networks ("STRING" genes), when compared with randomly generated subset heatmaps from the overall ("Full") schizophrenia-related chromosomal connectome (Fig. 3D), had higher organization scores in NPCs and neurons (NPC $|r|_{avg} = 0.2963$, $P = 0.007$; neuron $|r|_{avg} = 0.2877$, $P = 0.008$, glia $|r|_{avg} = 0.2225$, $P = 0.595$, STRING versus full permutation test) (Materials and methods) (Fig. 4B, figs. S13 to S15, and table S21). Because the transcriptomic correlation heatmap for the schizophrenia-related chromosomal connectome was significantly decreased by the removal specifically of the NPC STRING protein network genes ($P < 10^{-3}$) (table S24), this subset of STRING-interacting proteins may drive the observed orchestrated coregulation. Within these transcriptome- and proteome-based regulatory networks were numerous occasions of coregulated (RNA) and interacting (protein) risk locus and risk locus–connect genes that share the same TAD, including *CDC20*, which regulates dendrite development (*40*, *41*) and is associated at the protein level with *RNF220*, an E3 ubiquitin-ligase and β-catenin stabilizer (Fig. 4C) (*42*).

To examine whether such coregulation could be representative of the prefrontal cortex proteome of the adult brain, we screened a newly generated mass spectrometry–based dataset of 182 neuronal proteins, the majority of which were synaptic, quantified from prefrontal cortex of $n = 23$ adult schizophrenia and $n = 20$ control subjects (table S25) (*43*). Among the 182 proteins, there were four from the risk-associated neuronal protein network (Fig. 4D): $GABA_B$ receptor subunit GABBR1 and ionotropic (GRIA1 and GRIN2A) and metabotropic glutamate receptor subunits (GRM3). Protein-protein correlation scores were significantly higher for these four risk-associated proteins than expected from random permutation analysis from the pool of 182 proteins ($P < 0.002$) across patients and controls. We conclude that the schizophrenia-related chromosomal connectome, tethering other portions of the genome to the sequences associated with schizophrenia heritability, provides a structural foundation for a functional connectome that reflects coordinated regulation of gene expression and interactions within the proteome.

Discussion

Neural progenitor differentiation into neurons and glia is associated with dynamic remodeling of chromosomal conformations, including loss of many NPC-specific chromosomal contacts, with differentiation-induced loop pruning primarily affecting a subset of genes important for neurogenesis (NPC-to-neuron loss) and neuronal function (NPC-to-glia loss). These findings broadly resonate with a recent report linking neural differentiation to multiple scales of 3DG folding, governed by multiple mechanisms, including CTCF-dependent loop alterations, repressive chromatin remodeling, and cell- and lineage-specific transcription factor networks (*21*). Our results suggest that developmental 3DG remodeling affects a substantial portion of sequences that confer liability for schizophrenia; furthermore, these genes in 3D physical proximity with schizophrenia-risk variants show a surprisingly strong correlation at the level of the transcriptome and proteome. How might the disease-relevant reorganization of the spatial genome (the "chromosomal connectome") provide a structural foundation for coordinated regulation of expression? Recent Hi-C studies in mouse brain showed that chromosomal contacts preferentially occurred between loci targeted by the same transcription factors (*21*), and likewise, multiple schizophrenia risk loci could converge on intra- and interchromosomal hubs sharing a similar regulatory architecture including specific enhancers as well as transcription and splicing factors (*44*–*46*). Intriguingly, the three major functional categories associated with the genetic risk architecture of schizophrenia—neuronal connectivity, synaptic signaling, and chromatin remodeling (*47*, *48*)—were heavily represented within the cell type–specific chromosomal connectomes of neurons and NPCs described here (Fig. 3, B and E) and in whole tissue in vivo (*7*, *11*). Cell type–specific 3DG reorganization during the course of neural progenitor differentiation, as shown here, could therefore have profound implications for our understanding of the genetic underpinnings of psychiatric disease. For example, inclusion of the cell type–specific risk (sequence)–associated chromosomal connectome may lead to refinements of cumulative schizophrenia risk allele burden estimates, including "polygenic risk score" (PRS) or "biologically informed multilocus profile scores" (BIMPS), which currently only explain a small portion of disease risk (*49*). Cell type–specific intersection of 3DG and genetic risk maps are of clinical interest beyond psychiatric disorders; for example, risk variants that confer susceptibility to autoimmune disease were embedded in physically interacting chromosomal loci in lymphoblastoid cells (*50*). Our 3DG maps from neural progenitors and their isogenic neurons and glia are accessible through the PsychENCODE Knowledge Portal (https://synapse.org) and more than double the number of currently available Hi-C datasets from human brain (*7*, *9*, *10*), providing investigators with a resource to chart the expanded genome space associated with cognitive and neuropsychiatric disease in context of cell type–specific remodeling of chromosomal conformations during early development.

Materials and methods

In situ Hi-C from hiPSC-derived cells

In situ Hi-C libraries were generated from 2 million to 5 million cultured hiPSC-derived NPCs, glia, and neurons as described in (*3*) without modifications in the protocol. Briefly, in situ Hi-C consists of 7 steps: (i) crosslinking cells with formaldehyde, (ii) digesting the DNA using a 4-cutter restriction enzyme (e.g., MboI) within intact permeabilized nuclei, (iii) filling in and biotinylating the resulting 5′-overhangs, (iv) ligating the blunt ends, (v) shearing the DNA, (vi) pulling down the biotinylated ligation junctions with streptavidin beads, and (vii) analyzing these fragments using paired end sequencing. As quality control (QC) steps, we checked for efficient restriction with an agarose DNA gel and for appropriate size selection in using the Agilent Bioanalyzer after steps (v) and (vi). For the final QC, we performed superficial sequencing on the Illumina MiSeq (~2-3M reads/sample)

to assess quality of the libraries using metrics such as percent of reads passing filter, percent of chimeric reads, and percent of forward-reverse pairs (supplementary materials, table S1). For the forebrain directed differentiation neuronal library from subject S1, the Arima Hi-C kit (Arima Genomics, San Diego) was used according to the manufacturer's instructions.

Hi-C read mapping and matrix generation

The Hi-C libraries were sequenced on the Illumina HiSeq1000 platform (125bp paired-end) (New York Genome Center). Technical replicates of subject S2 NPCs, neurons, and glia were also sequenced to enhance resolution. Initial processing of the raw 2 ×125 bp read pair FASTQ files was performed using the HiC-Pro analysis pipeline (*51*). In brief, HiC-Pro performs four major tasks: aligning short reads, filtering for valid pairs, binning, and normalizing contact matrices. HiC-Pro implements the truncation-based alignment strategy using Bowtie v2.2.3 (*52*), mapping full reads end-to-end or the 5′ portion of reads preceding a GATCGATC ligation site that results from restriction enzyme digestion with MboI followed by end ligation. Invalid interactions such as same-strand, dangling-end, self-cycle, and single-end pairs are not retained. Binning was performed in 10kb, 40 kb and 100 kb non-overlapping, adjacent windows across the genome and resulting contact matrices were normalized using iterative correction and eigenvector decomposition (ICE) as previously described (*53*), using HiC-Pro's default settings of 100 maximum iterations, filtering of the sparse bins (lowest 2%), and a relative result increment of 0.1 before declaring convergence (http://nservant.github.io/HiC-Pro/MANUAL.html). Data are reported in browser-extensible-data-like (BED) format and visualized in the Washington University Epigenome Browser (http://epigenomegateway.wustl.edu). Hierarchical clustering was performed on the ICE-corrected intrachromosomal contact matrices after the bins with the 1% most extreme interaction values were excluded as largely artifactual. Clustering was performed using Ward's method on the 1, 5, 10, 25, 50, and 100% most variable remaining bins using (1-correlation) as a distance metric. The results using the 10% most variable interaction bins, shown here in a cluster dendrogram and a Pearson correlation matrix, are representative of these results.

Hi-C loop calls using Juicer

Loop calling was performed using the software HiCCUPS (*3*). To format data for HiCCUPS input, we remapped reads from Hi-C libraries using the Juicer pipeline (*54*). Similar to HiC-Pro, the Juicer pipeline performs read alignment, filtering, binning, and matrix normalization. Samples were pooled for each cell type (S1 and 2 technical replicates from S2) to generate the maximum amount of coverage required for accurate loop calling. The resulting .hic matrix files (MAPQ > 0) were then used as input to HiCCUPS. The following parameters were set for HiCCUPS following the analysis in (*3*): FDR threshold (f) = 0.10, 0.10; peak width (p) = 4, 2; window width (i) = 7, 5; merge distance (d) = 20 kb, 20 kb. Values for parameters correspond to calls made at 5kb and 10kb, respectively. Representative neuronal and non-neuronal loops are presented in fig. S3. As the number of loops called is dependent upon the number of Hi-C contacts in the matrix (*55*), we also generated matrices with equivalent total Hi-C contacts via subsampling. hiPSC-derived Hi-C interaction matrices were randomly subsampled to 372,787,143 cis only contacts (the lowest number of cis contacts across all cell types) and HiCCUPS was rerun on the subsampled matrices. After loops were called for each cell type, we performed a reevaluation on this union set of loop loci. HiCCUPS was rerun using the union set of loop loci as input to produce q-values for each loop in the union set for every cell type. By default, HiCCUPS does not output a q-value for every pixel. Hence, this reevaluation produced q-values for pixels in cells that did not pass the significance threshold. We then defined any pixel from the union set with a q-value < 0.10 with respect to the donut neighborhood surrounding the pixel to be a loop and defined the loop to be shared with any cell types having a q-value < 0.10 for the same pixel.

These loop calls were used for comparing loop calls between cell types. Loops were also called and subsampled as above for the GM12878 cell line using the processed data from (*3*) found here: www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63525. Loop calls were overlapped with compartment calls (supplementary materials, materials and methods), such that AA, BB, and AB refer to loops with both anchors in A, both anchors in B, and one anchor in A and other anchor in B, respectively. Loops in chromosomes 4, 18, 19, and X were removed from this compartment analysis since the first principle component most likely corresponded to p versus q arm distinctions and not A versus B compartments.

Hi-C interactions at risk loci

To approach 3DG conformation in context of the disease-relevant sequences, we adapted the binomial statistics based mapping strategy previously described by Won *et al.* (*7*). The set of schizophrenia risk loci used in this study included the original (PGC2, Psychiatric Genomics Consortium) (*2*) risk sequences, or 108 physically distinct association loci defined by 128 index SNPs (corrected P 10^{-8}) and an additional 37 loci from the CLOZUK (a series of UK cases registered for clozapine treatment with a clinical diagnosis of schizophrenia) study for a total of 145 loci defined by 179 independent genome-wide significant SNPs (corrected $P < 5 \times 10^{-8}$), determined by GWAS in 40,675 cases and 64,643 controls (*23*). A risk locus is defined as a collection of (SNPs) existing in linkage disequilibrium, ranging from 1bp to 8.9Mb (average 256.2 kb) in length and in total equivalent to approximately 0.012% of human genomic sequence.

To identify significantly enriched interactions involving a bin of interest with another bin, our principal approach was to first estimate the expected interaction counts for each interaction distance by calculating the mean of all intrachromosomal bin-bin interactions of the same separation distance throughout the raw intrachromosomal contact matrix. We used the R package, HiTC (*56*), to facilitate manipulation of our HiC-Pro-produced raw contact matrices and estimation of the expected counts at various interaction distances. The probability of observing an interaction between a bin-of-interest and another bin was then defined as the expected interaction between those two bins divided by the sum of all expected interactions between the bin-of-interest and all other intrachromosomal bins. A P value was then calculated as binomial probability of observing the number of interaction counts or more between the bin-of-interest and some other bin where the number of successes was defined as the observed interaction count, the number of tries as the total number of observed interactions between the bin-of-interest and all other intrachromosomal bins, and the success probability as the probability of observing the bin-bin interaction estimated from the expected mean interaction counts. The Benjamini-Hochberg method was used to control false discovery rate (FDR) for P values determined for all interactions with a bin-of-interest (includes all bins 1Mb up and downstream in our tests).

Generation of stable selected dCas9-VP64/VPR and Cas9 NPCs

All CRISPR-based epigenomic editing assays were performed on antibiotic-selected dCas9-VP64 (VP64 as the tetrameric VP16 activator domain) and dCas9-VPR (VPR as the tripartite activator, VP64-p65-Rta) NPCs derived as described in (*34*). For generation of Cas9 stable, selected NPCs, we used a plasmid of lentiCRISPR v2 gifted by Feng Zhang (Addgene plasmid # 52961). DNA sequencing with a U6 primer confirmed the identity. Lentiviral production and titration were performed as described previously (*14*). Control S1 and S2 NPCs were spinfected with lentiCRISPR v2 virus as described (*34*). 48 hours post-transduction, cells were selected by exposure to puromycin at 0.3 μg/mL. Without transduction, all control cells died within around 5 days after the antibiotic addition. The puromycin-selected NPCs were subject to Western blot analysis of Cas9 expression. 30 μg of proteins were electrophoresed in NuPAGE 4-12% Bis-Tris Protein Gels (NP0323PK2, Life Technologies) in 1× MES running buffer, 200 V constant, 35 min. Proteins were transferred onto nitrocellulose membrane (IB23002, Life Technologies) on the iBlot® 2 Dry Blotting System (program P3, 7:00 min). The membranes were incubated with primary antibodies against Cas9 (1:250, monoclonal, clone 7A9, Millipore) and β-Actin (1:10,000, mouse, 1406030, Ambion) overnight at 4°C. Then, membranes were incubated with the IRDye-labeled secondary antibodies for 45 min at RT in the dark on the rocker. Fluorescence was visualized using a Li-Cor Odyssey Imaging System.

In vitro transcription and transfection of gRNAs

Guide RNAs (gRNAs) were designed on Benchling (www.benchling.com) using the CRISPR tool. gRNAs were generated via in vitro transcription (IVT) with the GeneArt Precision gRNA Synthesis Kit (Thermo Fisher Scientific, A29377) as per manufacturer instructions. Five gRNAs were designed per condition (i.e., "loop-SNP", negative control, and positive control) and pooled for transfection. The genomic ranges within which loop-SNP gRNAs were designed (i.e., region spanning the SNP of interest and all gRNAs in the condition) were roughly 600 bp for *ASCL1*, 550 bp for *MATR3*, 460 bp for *EFNB1* (with 2/5 gRNAs directly overlapping the SNP), 300 bp for *SOX2*. Puromycin-selected (1μg/mL in NPC media; Sigma, #P7255) dCas9-VP64 and dCas9-VPR NPCs (*34*) were seeded at a density of ~400,000 per well on Matrigel-coated (BD Biosciences) 24-well plates. Pooled IVT gRNAs (500 ng total RNA/well) and 2 μL EditPro Stem lipofectamine (MTI-GlobalStem, #GST-2174; now, ThermoFisher, STEM00003) were diluted in 50 μL Opti-MEM (Thermo Fisher Scientific, #31985062) and added dropwise to each well. Cells were harvested with TRIzol for total RNA extraction 48 hours later. All experiments were conducted with 3 to 6 biological replicates from 1 donor (subject S1), generated in parallel, with the donor contributing isogenic dCas9-VP64 and dCas9-VPR effector cells. Each data point in Fig. 2, D to F, represents one biological replicate within each condition. For each target gene promoter and candidate loop, control gRNAs were strategically placed into the middle third of the (linear) genome portion bypassed by the candidate loop. CRISPRa results were analyzed on PRISM with a one-way ANOVA across 3 conditions with a Dunnett's test for multiple comparisons. Cas9 mutagenesis was also performed as described above with the exception of the negative control, which in these experiments consisted of an empty transfection (i.e., lipofectamine + Opti-MEM without any gRNA). Cas9 results were analyzed with an unpaired *t* test comparing the loop-SNP and negative control conditions.

RNA transcriptomic correlation heatmaps

Pearson correlation coefficient matrices were calculated for gene expression in the childhood onset schizophrenia data set (*16*) using R from lists of genes that are located in cell-type-specific loops anchored at schizophrenia risk loci and, as a subset of this list, sets of genes whose proteins participate in an association network for each of the three cell types (see below). Significance was computed calculating the absolute mean correlation coefficient of each correlation matrix ("organization score") as a test statistic against a null distribution generated by random gene sampling. Randomized gene lists were drawn only from the pool of genes with over 1 count per million (CPM) in at least 30% of the experiments described in (*16*). To generate a null distribution of organization scores for a given cell type that accounted for genomic distance and neighborhood effects, we began by randomly selecting a significant PGC interaction for that cell type (e.g., random selection from table S12). Using the bp genomic distance of this interaction we randomly selected two 10kb bins from the genome separated by the same distance. All genes overlapping these bins were then added to the list of genes with which to calculate the organization score. This process was iterated until enough genes were added to the list to match the number of genes used in the original cell-type-specific organization score. Finally, this protocol was repeated 1000 times to generate the null distribution of random organization scores. This distribution was then used to calculate significance of co-regulation (i.e., P = number of times $|r|_{avg}$ of the null exceeded that of the test heatmap / 1000). Note that STRING gene network transcriptomic analyses (Fig. 4B) were performed with 1000 random permutations of genes sampled from the full schizophrenia risk connectome (i.e., risk locus + risk locus-connect genes) for each cell type.

REFERENCES AND NOTES

1. P. Rajarajan, S. E. Gil, K. J. Brennand, S. Akbarian, Spatial genome organization and cognition. *Nat. Rev. Neurosci.* **17**, 681–691 (2016). doi: 10.1038/nrn.2016.124; pmid: 27708356
2. Schizophrenia Working Group of the Psychiatric Genomics Consortium, Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014). doi: 10.1038/nature13595; pmid: 25056061
3. S. S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014). doi: 10.1016/j.cell.2014.11.021; pmid: 25497547
4. Y. Jiang *et al.*, The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain. *Nat. Genet.* **49**, 1239–1250 (2017). doi: 10.1038/ng.3906; pmid: 28671686
5. R. Bharadwaj *et al.*, Conserved higher-order chromatin regulates NMDA receptor gene expression and cognition. *Neuron* **84**, 997–1008 (2014). doi: 10.1016/j.neuron.2014.10.032; pmid: 25467983
6. R. Bharadwaj *et al.*, Conserved chromosome 2q31 conformations are associated with transcriptional regulation of GAD1 GABA synthesis enzyme and altered in prefrontal cortex of subjects with schizophrenia. *J. Neurosci.* **33**, 11839–11851 (2013). doi: 10.1523/JNEUROSCI.1252-13.2013; pmid: 23864674
7. H. Won *et al.*, Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016). doi: 10.1038/nature19847; pmid: 27760116
8. E. Lieberman-Aiden *et al.*, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009). doi: 10.1126/science.1181369; pmid: 19815776
9. A. D. Schmitt *et al.*, A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Reports* **17**, 2042–2059 (2016). doi: 10.1016/j.celrep.2016.10.061; pmid: 27851967
10. J. R. Dixon *et al.*, Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015). doi: 10.1038/nature14222; pmid: 25693564
11. L. de la Torre-Ubieta *et al.*, The dynamic landscape of open chromatin during human cortical neurogenesis. *Cell* **172**, 289–304.e18 (2018). doi: 10.1016/j.cell.2017.12.014; pmid: 29307494
12. A. C. Mitchell *et al.*, The genome in three dimensions: A new frontier in human brain research. *Biol. Psychiatry* **75**, 961–969 (2014). doi: 10.1016/j.biopsych.2013.07.015; pmid: 23958183
13. A. Topol *et al.*, Dysregulation of miRNA-9 in a subset of schizophrenia patient-derived neural progenitor cells. *Cell Reports* **15**, 1024–1036 (2016). doi: 10.1016/j.celrep.2016.03.090; pmid: 27117414
14. S. M. Ho *et al.*, Rapid Ngn2-induction of excitatory neurons from hiPSC-derived neural progenitor cells. *Methods* **101**, 113–124 (2016). doi: 10.1016/j.ymeth.2015.11.019; pmid: 26626326
15. J. Tcw *et al.*, An efficient platform for astrocyte differentiation from human induced pluripotent stem cells. *Stem Cell Reports* **9**, 600–614 (2017). doi: 10.1016/j.stemcr.2017.06.018; pmid: 28757165
16. G. E. Hoffman *et al.*, Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains. *Nat. Commun.* **8**, 2225 (2017). doi: 10.1038/s41467-017-02330-5; pmid: 29263384
17. K. J. Brennand *et al.*, Modelling schizophrenia using human induced pluripotent stem cells. *Nature* **473**, 221–225 (2011). doi: 10.1038/nature09915; pmid: 21490598
18. C. Gil-Sanz *et al.*, Lineage tracing using Cux2-Cre and Cux2-CreERT2 mice. *Neuron* **86**, 1091–1099 (2015). doi: 10.1016/j.neuron.2015.04.019; pmid: 25996136
19. G. Charlesworth *et al.*, Mutations in ANO3 cause dominant craniocervical dystonia: Ion channel implicated in pathogenesis. *Am. J. Hum. Genet.* **91**, 1041–1050 (2012). doi: 10.1016/j.ajhg.2012.10.024; pmid: 23200863
20. J. Wang *et al.*, Single-cell gene expression analysis reveals regulators of distinct cell subpopulations among developing human neurons. *Genome Res.* **27**, 1783–1794 (2017). doi: 10.1101/gr.223313.117; pmid: 29030469
21. B. Bonev *et al.*, Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572.e24 (2017). doi: 10.1016/j.cell.2017.09.043; pmid: 29053968
22. J. R. Dixon *et al.*, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012). doi: 10.1038/nature11082; pmid: 22495300
23. A. F. Pardiñas *et al.*, Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat. Genet.* **50**, 381–389 (2018). doi: 10.1038/s41588-018-0059-2; pmid: 29483656
24. A. Dobbyn *et al.*, Landscape of conditional eQTL in dorsolateral prefrontal cortex and co-localization with schizophrenia GWAS. *Am. J. Hum. Genet.* **102**, 1169–1184 (2018). doi: 10.1016/j.ajhg.2018.04.011; pmid: 29805045
25. W. V. Chen, T. Maniatis, Clustered protocadherins. *Development* **140**, 3297–3302 (2013). doi: 10.1242/dev.090621; pmid: 23900538
26. T. Yagi, Molecular codes for neuronal individuality and cell assembly in the brain. *Front. Mol. Neurosci.* **5**, 45 (2012). doi: 10.3389/fnmol.2012.00045; pmid: 22518100
27. N. C. Durand *et al.*, Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016). doi: 10.1016/j.cels.2015.07.012; pmid: 27467250
28. M. Fromer *et al.*, Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016). doi: 10.1038/nn.4399; pmid: 27668389
29. K. Brennand *et al.*, Phenotypic differences in hiPSC NPCs derived from patients with schizophrenia. *Mol. Psychiatry* **20**, 361–368 (2015). doi: 10.1038/mp.2014.22; pmid: 24686136
30. S. Gulsuner *et al.*, Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013). doi: 10.1016/j.cell.2013.06.049; pmid: 23911319
31. P. Perez-Pinera *et al.*, RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973–976 (2013). doi: 10.1038/nmeth.2600; pmid: 23892895
32. M. L. Maeder *et al.*, CRISPR RNA-guided activation of endogenous human genes. *Nat. Methods* **10**, 977–979 (2013). doi: 10.1038/nmeth.2598; pmid: 23892898
33. A. Chavez *et al.*, Comparison of Cas9 activators in multiple species. *Nat. Methods* **13**, 563–567 (2016). doi: 10.1038/nmeth.3871; pmid: 27214048
34. S. M. Ho *et al.*, Evaluating synthetic activation and repression of neuropsychiatric-related genes in hiPSC-derived NPCs, neurons, and astrocytes. *Stem Cell Reports* **9**, 615–628 (2017). doi: 10.1016/j.stemcr.2017.06.012; pmid: 28757163
35. C. P. Fulco *et al.*, Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016). doi: 10.1126/science.aag2445; pmid: 27708057
36. K. Girdhar *et al.*, Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nat. Neurosci.* **21**, 1126–1136 (2018). doi: 10.1038/s41593-018-0187-0; pmid: 30038276
37. G. Kustatscher, P. Grabowski, J. Rappsilber, Pervasive coexpression of spatially proximal genes is buffered at the

protein level. *Mol. Syst. Biol.* **13**, 937 (2017). doi: 10.15252/msb.20177548; pmid: 28835372
38. D. Szklarczyk *et al*., STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43** (D1), D447–D452 (2015). doi: 10.1093/nar/gku1003; pmid: 25352553
39. D. Szklarczyk *et al*., The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45** (D1), D362–D368 (2017). doi: 10.1093/nar/gkw937; pmid: 27924014
40. Y. Watanabe, K. Khodosevich, H. Monyer, Dendrite development regulated by the schizophrenia-associated gene FEZ1 involves the ubiquitin proteasome system. *Cell Reports* **7**, 552–564 (2014). doi: 10.1016/j.celrep.2014.03.022; pmid: 24726361
41. S. V. Puram *et al*., A CaMKIIβ signaling pathway at the centrosome regulates dendrite patterning in the brain. *Nat. Neurosci.* **14**, 973–983 (2011). doi: 10.1038/nn.2857; pmid: 21725312
42. P. Ma *et al*., The ubiquitin ligase RNF220 enhances canonical Wnt signaling through USP7-mediated deubiquitination of β-catenin. *Mol. Cell. Biol.* **34**, 4355–4366 (2014). doi: 10.1128/MCB.00731-14; pmid: 25266658
43. C. G. Hahn *et al*., The post-synaptic density of human postmortem brain tissues: An experimental study paradigm for neuropsychiatric illnesses. *PLOS ONE* **4**, e5251 (2009). doi: 10.1371/journal.pone.0005251; pmid: 19370153
44. S. Lomvardas *et al*., Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403–413 (2006). doi: 10.1016/j.cell.2006.06.035; pmid: 16873069
45. S. A. Quinodoz *et al*., Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757.e24 (2018). doi: 10.1016/j.cell.2018.05.024; pmid: 29887377
46. N. Khanna, Y. Hu, A. S. Belmont, HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation. *Curr. Biol.* **24**, 1138–1144 (2014). doi: 10.1016/j.cub.2014.03.053; pmid: 24794297
47. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium, Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199–209 (2015). doi: 10.1038/nn.3922; pmid: 25599223
48. S. R. Gilman *et al*., Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat. Neurosci.* **15**, 1723–1728 (2012). doi: 10.1038/nn.3261; pmid: 23143521
49. R. Bogdan, D. A. A. Baranger, A. Agrawal, Polygenic Risk Scores in Clinical Psychology: Bridging Genomic Risk to Individual Differences. *Annu. Rev. Clin. Psychol.* **14**, 119–157 (2018). doi: 10.1146/annurev-clinpsy-050817-084847; pmid: 29579395
50. F. Grubert *et al*., Genetic control of chromatin states in humans involves local and distal chromosomal interactions. *Cell* **162**, 1051–1065 (2015). doi: 10.1016/j.cell.2015.07.048; pmid: 26300125
51. N. Servant *et al*., HiC-Pro: An optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015). doi: 10.1186/s13059-015-0831-x; pmid: 26619908
52. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). doi: 10.1186/gb-2009-10-3-r25; pmid: 19261174
53. M. Imakaev *et al*., Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012). doi: 10.1038/nmeth.2148; pmid: 22941365
54. N. C. Durand *et al*., Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016). doi: 10.1016/j.cels.2016.07.002; pmid: 27467249
55. M. Forcato *et al*., Comparison of computational methods for Hi-C data analysis. *Nat. Methods* **14**, 679–685 (2017). doi: 10.1038/nmeth.4325; pmid: 28604721
56. N. Servant *et al*., HiTC: Exploration of high-throughput 'C' experiments. *Bioinformatics* **28**, 2843–2844 (2012). doi: 10.1093/bioinformatics/bts521; pmid: 22923296

ACKNOWLEDGMENTS

The authors thank all members of the Brennand and Akbarian laboratories for constructive comments and discussions; C. O'Shea, S. R. Shannon, I. Silverberg, and A. Austin for valuable assistance with PCR assays; E. Xia, C. Rosenbluh, and A. Chess for generous access to sequencing equipment; L. Winterkorn and staff at the New York Genome Center for logistical support; and M. Peters for help and support as it pertains to data accessibility. **Funding:** This work was supported by National Institute of Mental Health PsychENCODE grants U01MH103392 (S.A.) and R01MH106056 (S.A. and K.J.B.) and a predoctoral Ruth L. Kirschstein fellowship 1F30MH113330 (P.R.). **Author contributions:** Cell culture work including Hi-C, 3C, RNA-seq, ATAC-seq, Cas9, and dCas9 (epi)genome editing: P.Ra., C.C., C.Y., B.K., B.J., S.P., E.A.L., B.Z., S.-M.H., and A.L. Biocomputing, informatics, and genomic analyses: T.B., W.L., N.S., S.E.-G., E.F., G.E.H., and Z.W. Contributed materials: S.-M.H., E.F., H.W., and D.H.G. Supervised research: Z.W., P.Ro., K.J.B., and S.A. Conceived and designed experiments: P.Ra., K.J.B., and S.A. Wrote the paper: P.Ra., T.B., W.L., K.J.B., and S.A., with contributions from all coauthors. **Competing interests:** The authors declare no conflicts of interest. **Data and materials accessibility:** Scripts used for bioinformatics analyses can be found at https://github.com/tborrman/Schahram-project and https://github.com/will-NYGC/sigHicInteractions. Hi-C data files and metadata are accessible through an open source platform: www.synapse.org/#!Synapse:syn12979101 (registration required; Data Download—Study "iPSC-HiC"). The code for this paper has been submitted to Zenodo, and the respective links are https://zenodo.org/badge/latestdoi/68036139 and https://zenodo.org/badge/latestdoi/151779353.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/eaat4311/suppl/DC1
Materials and Methods
Figs. S1 to S15
References (57–67)
Tables S1 to S25

27 February 2018; accepted 7 November 2018
10.1126/science.aat4311

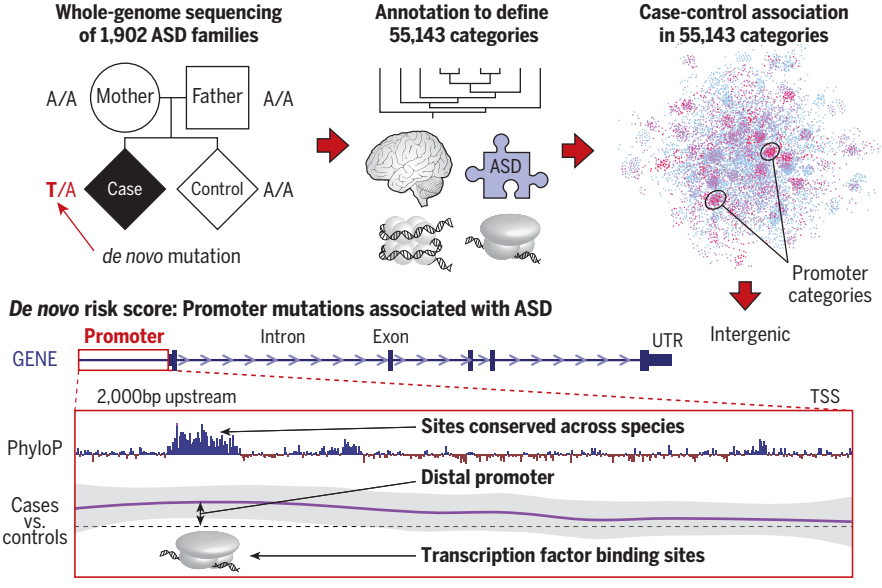
RESEARCH ARTICLE SUMMARY

PSYCHIATRIC GENOMICS

Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder

Joon-Yong An*, Kevin Lin*, Lingxue Zhu*, Donna M. Werling*, Shan Dong, Harrison Brand, Harold Z. Wang, Xuefang Zhao, Grace B. Schwartz, Ryan L. Collins, Benjamin B. Currall, Claudia Dastmalchi, Jeanselle Dea, Clif Duhn, Michael C. Gilson, Lambertus Klei, Lindsay Liang, Eirene Markenscoff-Papadimitriou, Sirisha Pochareddy, Nadav Ahituv, Joseph D. Buxbaum, Hilary Coon, Mark J. Daly, Young Shin Kim, Gabor T. Marth, Benjamin M. Neale, Aaron R. Quinlan, John L. Rubenstein, Nenad Sestan, Matthew W. State, A. Jeremy Willsey, Michael E. Talkowski†, Bernie Devlin†, Kathryn Roeder†, Stephan J. Sanders†

INTRODUCTION: The DNA of protein-coding genes is transcribed into mRNA, which is translated into proteins. The "coding genome" describes the DNA that contains the information to make these proteins and represents ~1.5% of the human genome. Newly arising de novo mutations (variants observed in a child but not in either parent) in the coding genome contribute to numerous childhood developmental disorders, including autism spectrum disorder (ASD). Discovery of these effects is aided by the triplet code that enables the functional impact of many mutations to be readily deciphered. In contrast, the "noncoding genome" covers the remaining ~98.5% and includes elements that regulate when, where, and to what degree protein-coding genes are transcribed. Understanding this noncoding sequence could provide insights into human disorders and refined control of emerging genetic therapies. Yet little is known about the role of mutations in noncoding regions, including whether they contribute to childhood developmental disorders, which noncoding elements are most vulnerable to disruption, and the manner in which information is encoded in the noncoding genome.



Promoter regions in autism. De novo mutations from 1902 quartet families are assigned to 55,143 annotation categories, which are each assessed for autism spectrum disorder (ASD) association by comparing mutation counts in cases and sibling controls. A de novo risk score demonstrated a noncoding contribution to ASD driven by promoter mutations, especially at sites conserved across species, in the distal promoter or targeted by transcription factors.

RATIONALE: Whole-genome sequencing (WGS) provides the opportunity to identify the majority of genetic variation in each individual. By performing WGS on 1902 quartet families including a child affected with ASD, one unaffected sibling control, and their parents, we identified ~67 de novo mutations across each child's genome. To characterize the functional role of these mutations, we integrated multiple datasets relating to gene function, genes implicated in neurodevelopmental disorders, conservation across species, and epigenetic markers, thereby combinatorially defining 55,143 categories. The scope of the problem—testing for an excess of de novo mutations in cases relative to controls for each category—is challenging because there are more categories than families.

ON OUR WEBSITE

Read the full article at http://dx.doi.org/10.1126/science.aat6576

RESULTS: Comparing cases to controls, we observed an excess of de novo mutations in cases in individual categories in the coding genome but not in the noncoding genome. To overcome the challenge of detecting noncoding association, we used machine learning tools to develop a de novo risk score to look for an excess of de novo mutations across multiple categories. This score demonstrated a contribution to ASD risk from coding mutations and a weaker, but significant, contribution from noncoding mutations. This noncoding signal was driven by mutations in the promoter region, defined as the 2000 nucleotides upstream of the transcription start site (TSS) where mRNA synthesis starts. The strongest promoter signals were defined by conservation across species and transcription factor binding sites. Well-defined promoter elements (e.g., TATA-box) are usually observed within 80 nucleotides of the TSS; however, the strongest ASD association was observed distally, 750 to 2000 nucleotides upstream of the TSS.

CONCLUSION: We conclude that de novo mutations in the noncoding genome contribute to ASD. The clearest evidence of noncoding ASD association came from mutations at evolutionarily conserved nucleotides in the promoter region. The enrichment for transcription factor binding sites, primarily in the distal promoter, suggests that these mutations may disrupt gene transcription via their interaction with enhancer elements in the promoter region, rather than interfering with transcriptional initiation directly. ■

The list of author affiliations is available in the full article online.
*These authors contributed equally to this work.
†Corresponding author. Email: talkowski@chgr.mgh.harvard.edu (M.E.T.); devlinbj@upmc.edu (B.D.); roeder@andrew.cmu.edu (K.R.); stephan.sanders@ucsf.edu (S.J.S.)
Cite this article as J.-Y. An *et al.*, *Science* **362**, eaat6576 (2018). DOI: 10.1126/science.aat6576

RESEARCH ARTICLE

PSYCHIATRIC GENOMICS

Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder

Joon-Yong An[1*], Kevin Lin[2*], Lingxue Zhu[2*], Donna M. Werling[1*], Shan Dong[1], Harrison Brand[3,4,5], Harold Z. Wang[3], Xuefang Zhao[3,4,5], Grace B. Schwartz[1], Ryan L. Collins[3,4,6], Benjamin B. Currall[3,4,5], Claudia Dastmalchi[1], Jeanselle Dea[1], Clif Duhn[1], Michael C. Gilson[1], Lambertus Klei[7], Lindsay Liang[1], Eirene Markenscoff-Papadimitriou[1], Sirisha Pochareddy[8], Nadav Ahituv[9,10], Joseph D. Buxbaum[11,12,13,14], Hilary Coon[15,16], Mark J. Daly[5,17,18], Young Shin Kim[1], Gabor T. Marth[19,20], Benjamin M. Neale[5,17,18], Aaron R. Quinlan[16,19,20], John L. Rubenstein[1], Nenad Sestan[8], Matthew W. State[1,10], A. Jeremy Willsey[1,21,22], Michael E. Talkowski[3,4,5,23]†, Bernie Devlin[7]†, Kathryn Roeder[2,24]†, Stephan J. Sanders[1,10]†

Whole-genome sequencing (WGS) has facilitated the first genome-wide evaluations of the contribution of de novo noncoding mutations to complex disorders. Using WGS, we identified 255,106 de novo mutations among sample genomes from members of 1902 quartet families in which one child, but not a sibling or their parents, was affected by autism spectrum disorder (ASD). In contrast to coding mutations, no noncoding functional annotation category, analyzed in isolation, was significantly associated with ASD. Casting noncoding variation in the context of a de novo risk score across multiple annotation categories, however, did demonstrate association with mutations localized to promoter regions. We found that the strongest driver of this promoter signal emanates from evolutionarily conserved transcription factor binding sites distal to the transcription start site. These data suggest that de novo mutations in promoter regions, characterized by evolutionary and functional signatures, contribute to ASD.

De novo mutations play an important role in human disorders that impair reproductive fitness, including autism spectrum disorder (ASD) (*1*), severe developmental delay (*2*), epileptic encephalopathy (*3*), and a spectrum of congenital anomalies (*4*, *5*). Analysis of de novo mutations in the 1.5% of the genome that encodes proteins has identified numerous genes associated with ASD (*1*), and these findings have provided a foundation from which to interrogate ASD etiology (*6*–*9*). The contribution of de novo variation in the 98.5% of sequence that constitutes the noncoding genome remains largely unknown (*10*, *11*). Identifying noncoding variants that regulate gene function could provide important insights into when, where, and in which cell type ASD pathology occurs. Such knowledge could have broad implications for targeted therapeutics (*10*).

Targeted sequencing of highly evolutionarily conserved loci in 7930 families with a child affected by severe developmental delay identified a modest contribution from de novo mutations at loci that are active in the fetal brain (*12*). Whole-genome sequencing (WGS) represents the next critical step in such explorations, enabling the contribution of noncoding de novo mutations to be evaluated systematically across the genome; however, the multiplicity of hypotheses that can be tested in an unbiased screen requires careful consideration of statistical interpretation. To date, WGS analyses of as many as 519 families with a child affected by ASD have yet to identify a significant noncoding contribution from de novo mutations, after appropriate correction for the multiple comparisons necessary in genome-wide analyses (*13*–*16*).

WGS analyses are complicated by the sheer scale of the noncoding genome and by limited methods to predict functional regions and disruptive variants. The category-wide association study (CWAS) framework applies multiple annotation methods to define thousands of annotation categories, each of which is tested for association with ASD. This CWAS approach is similar to that used in a genome-wide association study, with single-nucleotide polymorphisms (SNPs) substituted for annotation categories, and uses similar correction for multiple comparisons (*15*, *17*). The CWAS-defined categories can also be used to build a de novo risk score, akin to a polygenic risk score, by selecting multiple annotation categories in a training cohort for assessment in a testing cohort (*15*). This model is generated once, so it does not incur a multiple testing penalty. In the present study, our results demonstrate an association between de novo noncoding mutations and ASD that is driven by mutations in conserved promoter regions.

Identification of de novo mutations in 1902 families

We analyzed the results of WGS in 7608 samples from 1902 quartet families from the Simons Simplex Collection (*18*), each composed of a mother and father, a child affected by ASD, and an unaffected sibling (table S1). This family-based design enables the detection of newly arising de novo mutations that are rare but can have drastic effects, and allows a direct comparison between ASD cases and their unaffected siblings as controls. By comparing each affected and unaffected child to their parents, we identified 255,106 de novo mutations in 1902 families (Fig. 1A and table S2), with 61.5 de novo single-nucleotide variants (SNVs) and 5.6 de novo insertions or deletions [indels; ≤50 base pairs (bp)] per child, using a high-quality variant filter defined in our previous study (*15*). These mutation rates are similar to those reported previously (fig. S1). Independent experimental validation confirmed

[1]Department of Psychiatry, UCSF Weill Institute for Neurosciences, University of California, San Francisco, CA, USA. [2]Department of Statistics and Data Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA. [3]Center for Genomic Medicine and Department of Neurology, Massachusetts General Hospital, Boston, MA, USA. [4]Department of Neurology, Harvard Medical School, Boston, MA, USA. [5]Program in Medical and Population Genetics and the Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA. [6]Program in Bioinformatics and Integrative Genomics, Division of Medical Sciences, Harvard Medical School, Boston, MA, USA. [7]Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA. [8]Department of Neuroscience and Kavli Institute for Neuroscience, Yale School of Medicine, New Haven, CT 06510, USA. [9]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA, USA. [10]Institute for Human Genetics, University of California, San Francisco, CA, USA. [11]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [12]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. [13]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [14]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [15]Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT, USA. [16]Department of Biomedical Informatics, University of Utah School of Medicine, Salt Lake City, UT, USA. [17]Analytical and Translational Genetics Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [18]Department of Medicine, Harvard Medical School, Boston, MA, USA. [19]Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT, USA. [20]USTAR Center for Genetic Discovery, University of Utah School of Medicine, Salt Lake City, UT, USA. [21]Institute for Neurodegenerative Diseases, UCSF Weill Institute for Neurosciences, University of California, San Francisco, CA, USA. [22]Quantitative Biosciences Institute, University of California, San Francisco, CA, USA. [23]Departments of Pathology and Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [24]Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA.
*These authors contributed equally to this work.
†Corresponding author. Email: talkowski@chgr.mgh.harvard.edu (M.E.T.); devlinbj@upmc.edu (B.D.); roeder@andrew.cmu.edu (K.R.); stephan.sanders@ucsf.edu (S.J.S.)

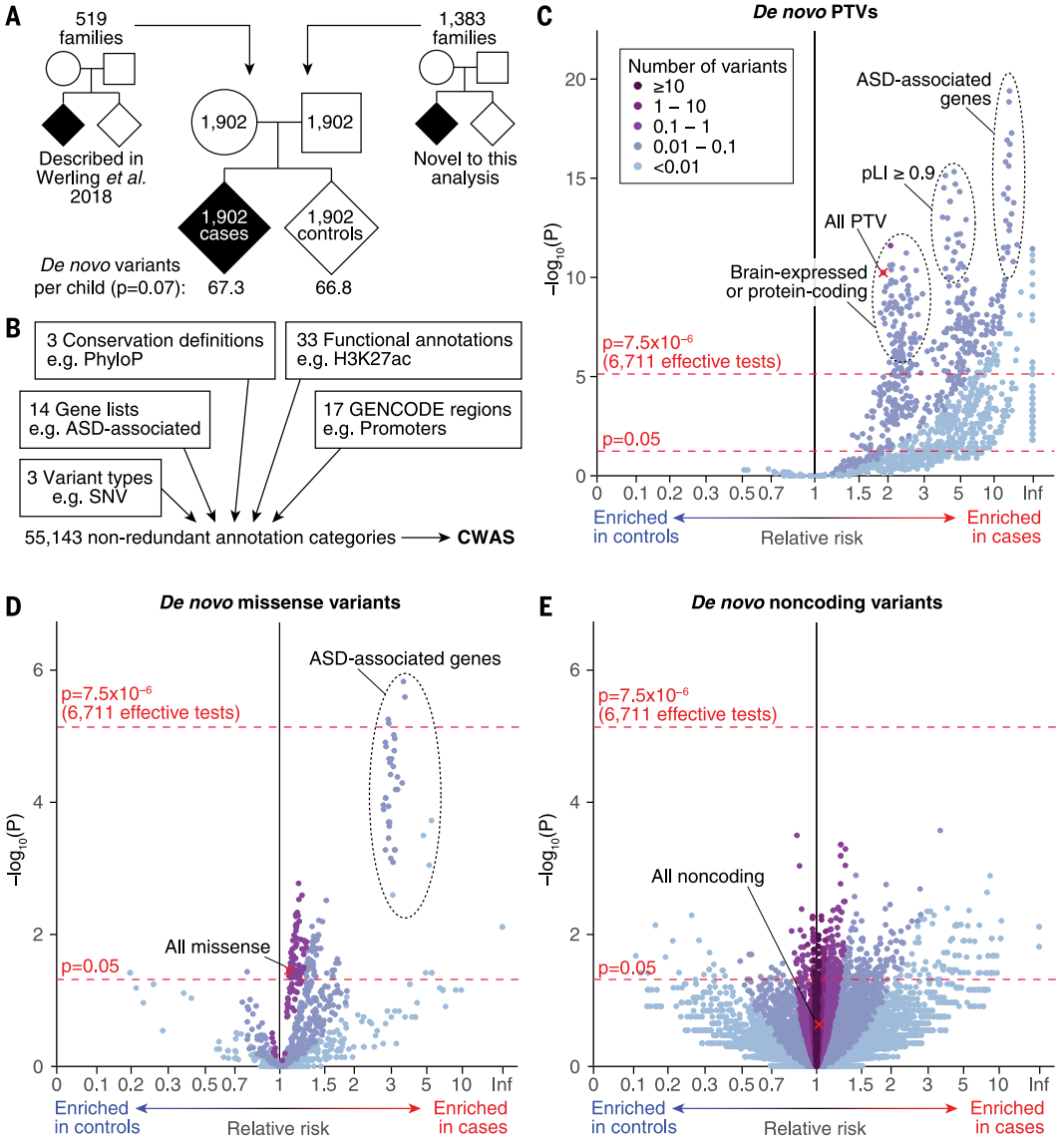


Fig. 1. Category-wide association study on 1902 ASD families. (**A**) De novo mutations were identified in 7608 samples from 1902 quartet families, each including an ASD case and an unaffected sibling control. The mean genome-wide mutation rate, corrected for paternal age, is shown for cases and controls. (**B**) Each mutation was annotated against 70 annotation terms in five groups, combinations of which defined 55,143 annotation categories (table S3 and fig. S5). (**C**) A category-wide association study (CWAS) shows the degree to which de novo protein-truncating variants (PTVs) in each category (points) are enriched in cases (right *x* axis) or controls (left *x* axis) against the statistical evidence for this enrichment (*y* axis). Red lines show the threshold for nominal significance ($P = 0.05$) and significance after correction for 6711 effective tests (*19*). The red X shows the category of all PTVs without other annotations. (**D** and **E**) The equivalent CWAS is shown for de novo missense (D) and de novo noncoding (E) variants. Statistical tests: binomial exact test, two-tailed [(C) to (E)].

97.1% of SNVs (238/245) and 82.7% of indels (148/179) (*19*). No difference in noncoding de novo rate was observed between cases and controls after correcting for the established correlation between parental age and de novo frequency (*20*) [corrected relative risk (cRR) = 1.005; $P = 0.15$ by permutation of case-control labels; table S3 and fig. S2]. Ancestry was not a significant predictor of de novo mutation rate; thus, it was not included in this correction (figs. S3 and S4).

Only protein-coding categories show genome-wide enrichment in cases

In coding regions, ASD-associated mutations are found at a small number of critical loci—for example, protein-truncating variants (PTVs) in ~5% of genes (*21*). In the absence of an equivalent definition for critical noncoding loci, we annotated the mutations against gene definitions, ASD-associated gene lists, species conservation, types of mutation, and functional annotations (e.g., ChIP-seq, ATAC-seq, DNase-seq) to define 55,143 annotation categories (Fig. 1B, fig. S5, and table S3). Considering each category separately in a CWAS, 579 categories reached our correction threshold of 7.5×10^{-6}, generated by Eigen decomposition of 10,000 simulated datasets (*15*). All 579 categories were enriched in cases rather than controls; 575 of these included de novo PTV mutations (cRR = 1.92; $P = 2.9 \times 10^{-11}$, binomial; Fig. 1C), and the remaining four categories were subsets of missense mutations in genes previously associated with ASD (cRR = 2.90; $P = 5.7 \times 10^{-6}$; Fig. 1D and fig. S6). No noncoding categories reached the correction threshold (Fig. 1E). We note that many of the ASD-associated genes were identified by de novo PTVs, and to a lesser extent de novo missense mutations, in these same cases (*1*). To focus on classes of variation with more subtle impacts on ASD risk, we excluded all annotation categories that included PTVs from further analysis.

Previous analyses have used WGS data to screen the genome, but those analyses were restricted to "candidate" noncoding categories selected on the basis of assumptions about functional impact as opposed to unbiased genome-wide analyses, in cohorts ranging from 39 to 516 ASD families (*13*, *14*, *22*). Although these candidate categories were enriched at nominal significance in ASD cases in those initial discovery cohorts, no candidate categories reached nominal significance in this larger cohort, despite similar mutation rates (table S4). Similarly, we did not observe enrichment of mutations in ASD cases in the conserved noncoding elements described with targeted sequencing of 6239 families with severe developmental delay (*12*), although we note that our replication cohort is substantially smaller than the discovery cohort and of a different phenotype.

Analysis across multiple noncoding categories highlights the role of promoters

No single noncoding annotation category passed our threshold of significance (Fig. 1E), so we

further explored the data by building a de novo risk score (*15*) to identify groups of categories in an unsupervised genome-wide analysis. To generate the score, we first restricted the analysis to annotation categories with a relatively small number of de novo mutations (*19*). This thresholding step is critical because the presence of numerous de novo mutations in an annotation category could represent false negatives in parents (i.e., apparent de novo mutations that were actually inherited variants), highly mutable regions, regions with limited impact on natural selection, or categories covering large swaths of the genome; none of these possibilities are likely to enrich for ASD risk at a small number of critical loci. Next, to select annotations likely to be important for risk from the remaining annotations, we generated a risk score using a Lasso regression from 519 families, described in (*15*), to identify annotation categories with rates of mutations that distinguish cases from controls. The resulting risk score was composed of 238 annotation categories, each with a coefficient reflecting the contribution of the category to the score (table S5). Applying the risk score to 1383 new families revealed it to be a significant predictor of case status (R^2 = 1.67%, $P = 5 \times 10^{-12}$; Fig. 2A). Of the 238 annotation categories, 75 were in coding regions (R^2 = 1.08%, $P = 4 \times 10^{-9}$; table S5) and 163 were noncoding (R^2 = 0.54%, P = 0.02; table S5); this finding demonstrates a noncoding contribution of de novo mutations to ASD risk.

To understand the nature of this noncoding contribution, we assessed the relative frequencies of the individual annotation terms from which the 163 noncoding categories are composed. The three annotation terms most frequently selected were PhastCons-defined (*23*) evolutionarily conserved regions (68 of 163 categories), PhyloP-defined (*24*) evolutionarily conserved nucleotides (49 of 163 categories), and promoter regions, defined as 2 kb upstream of the transcription start site (TSS) (45 of 163 categories). The inclusion of 45 promoter categories in the model is enriched by a factor of 2.45 over expectation ($P = 6 \times 10^{-7}$ after correcting for 62 noncoding annotation terms; Fig. 2A and table S5). The risk score remained a significant predictor of case status with only these promoter categories included and accounted for the majority of the noncoding signal (R^2 = 0.50%, P = 0.01; Fig. 2A and table S5). In contrast, the remaining 118 noncoding categories, without promoters, were not significant predictors of case status (R^2 = 0.22%, P = 0.25; Fig. 2A). The 45 promoter categories selected in the risk score encompassed 150 independent mutations, 112 in cases and 38 in controls (table S6).

To examine whether this promoter signal was detectable beyond these 150 mutations, we considered the pattern of de novo mutation enrichment across all 1855 promoter-defined annotation categories with ≥7 mutations. Of these, 112 were enriched in cases at nominal significance, which is more than expected (cross-category burden P = 0.03; Fig. 2, B and C), unlike the six categories enriched at nominal significance in controls (cross-category burden P = 0.94; Fig. 2, B and C). Ten of the 112 case-enriched categories were also selected for inclusion in the de novo risk score; no control-enriched categories were selected.

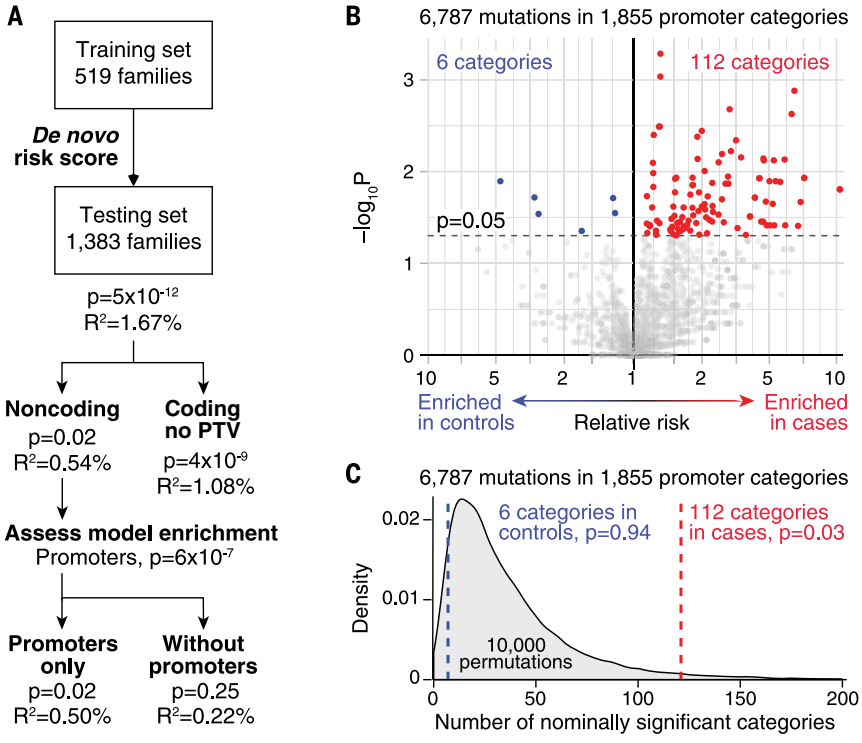


Fig. 2. Enrichment of conserved promoters in cases. (**A**) After excluding categories with PTVs, which are known to have a strong contribution to ASD, a de novo risk score was generated using Lasso regression to distinguish cases and controls in the first 519 families and tested on 1383 new families. The same risk score was tested considering 163 noncoding categories only and, based on the enrichment of promoter categories in the risk score, for 45 promoter categories and 118 noncoding categories without promoters (table S5). (**B**) Considering 1855 promoter annotation categories with ≥7 mutations, 118 reached nominal significance, 112 of which had an excess of mutations in cases. (**C**) The observation of 112 nominally significant case-enriched categories (red line) and six control-enriched categories (blue line) in (B) is compared to permuted expectation (gray distribution). Statistical tests: Lasso regression with permutation testing (A); binomial, two-sided (B); permutation testing (C).

Promoter association is driven by evolutionary conservation

To understand the types of variants and genes that account for this association between promoter mutations and ASD, we performed an exploratory analysis of the 6787 promoter region mutations and the 1310 promoter annotation categories with at least 20 mutations. Considering the correlation of P values across annotation categories, on the basis of 10,000 simulations (*19*), we identified 47 clusters, each composed of multiple highly correlated categories (Fig. 3A and table S7). Using the DAWN hidden Markov random field model (*25*) to refine the evidence for association based on the strength of association in neighboring clusters, nine of the 47 clusters were identified at a Bayesian false discovery rate of 0.01 (Fig. 3A and Table 1).

Assessment of the overlap of mutations between clusters and annotation terms identified two large groups of promoter mutations (Fig. 3, B and C): an "Active Transcription Start Site (TSS)" group (RR = 1.03; P = 0.32, binomial test; Fig. 3D), distinguished by correlated epigenetic markers (C18 and C28; Fig. 3B), and a "Conserved Loci" group (RR = 1.28; P = 0.0002, binomial test; Fig. 3D), distinguished by PhastCons and/or PhyloP scores (C12, C20, C49, C63; Fig. 3B). Of the 931 de novo mutations in the Conserved Loci group, 557 (60%) are also in the Active TSS group (Fig. 3C) and removing these conserved loci from the Active TSS group removes almost all of the signal (RR = 1.00).

The three remaining small clusters show limited overlap with the Active TSS and Conserved Loci groups (Fig. 3B and Table 1): C7, defined by long noncoding RNAs (lncRNAs) at active TSSs (RR = 1.19); C42, defined by developmental delay genes (*2*) (RR = 1.51); and C26, defined by processed transcripts (RR = 2.00).

When we consider all mutations in promoters as a single category, we see a nonsignificant trend toward weak enrichment in cases (3458 in cases versus 3329 in controls; cRR = 1.03; P = 0.16, permutation test). Because the cluster analysis highlighted the role of evolutionary conservation

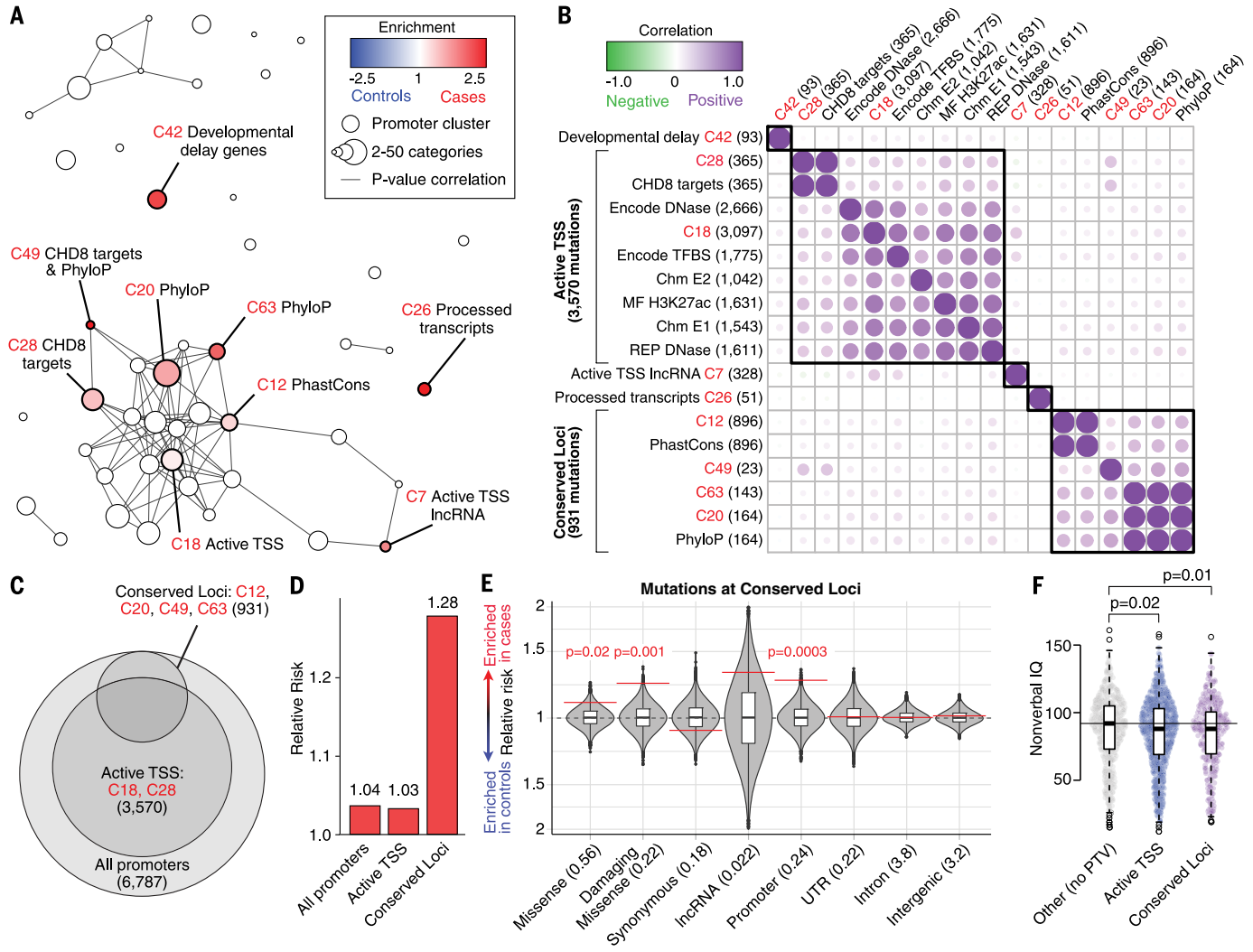


Fig. 3. Mapping ASD association within promoter regions by annotation terms. (**A**) DAWN uses *P*-value correlations between 1310 promoter categories with ≥20 mutations to define 47 clusters (nodes, with size representing the number of categories in the cluster). Evidence for ASD association is evaluated in the context of the local *P*-value correlation network (edges) to estimate false discovery rate (FDR). Enrichment is shown by color for the nine clusters with FDR ≤ 0.01 (Table 1). (**B**) The number of de novo mutations shared between these nine clusters and the annotation terms enriched in these clusters is shown as a correlation with hierarchical clustering. The black boxes show the first five divisions based on hierarchical clustering with two large groups: Active TSS and Conserved Loci. The numbers of de novo mutations in each group are shown in parentheses. (**C**) The size and relationship of the groups of promoter mutations identified in (A) and (B), based on de novo mutation counts. The number of mutations in each group is shown in parentheses. (**D**) Estimates of relative risk based on the number of de novo mutations in cases and controls within each group. (**E**) Considering mutations at Conserved Loci, the degree of enrichment of mutations in cases versus controls (red line) is shown in relation to permuted expectation (gray distributions). The mean number of mutations per child is shown in parentheses. Nominally significant uncorrected *P* values are shown in red. (**F**) Distribution of nonverbal IQ in cases with mutations at Active TSS (blue) and Conserved Loci (purple) promoters versus cases with neither (gray). Cases with de novo PTVs were excluded from all groups. Statistical tests: DAWN (A); permutation testing (E); Wilcoxon signed rank, two-sided (F). Box plot in (E) and (F) shows the median (black line), interquartile range (white box), and a further 1.5 times the interquartile range (whiskers). DD, developmental delay; MF, midfetal; REP, Roadmap Epigenome; UTR, untranslated region.

(Fig. 3D), we assessed case-control burden for all 30,891 conserved mutations, split by GENCODE-defined (*26*) genic regions (Fig. 3E). We observed an excess of mutations in cases at conserved loci in promoters (522 versus 409; cRR = 1.26; P = 0.0003, permutation test), but not for mutations in other noncoding regions (Fig. 3E and fig. S7). In coding regions, de novo mutations that are not observed in the general population according to the Genome Aggregation Database (gnomAD) (*27*) are more likely to be associated with ASD (*28*). Similarly, we observe stronger ASD association at promoter regions if mutations seen in gnomAD are excluded (470 versus 350; cRR = 1.34; $P = 3 \times 10^{-5}$, permutation test). Given the rarity and high effect sizes of protein-disrupting de novo mutations, we might expect a marginally higher rate of risk-mediating mutations in the 1759 ASD cases without previously identified ASD-associated mutations (*1*) relative to the 143 families with prior findings (table S1). However, no such difference was observed between these two groups in conserved promoters (P = 0.61, permutation test; fig. S8) or for conserved missense mutations (P = 0.20, permutation test; fig. S8).

Gene set enrichment and phenotype in the Conserved Loci group

The Conserved Loci group includes the promoters of 886 unique genes, of which 53% are protein-coding, 15% are processed pseudogenes, and 14% are lncRNAs (table S6) with similar

Table 1. Groups and clusters of categories within promoter regions.

Cluster	Description	Active TSS	Conserved loci	CHD8 binding targets	Total mutations (case/control)	Absolute RR	Binomial *P* value	DAWN RR	DAWN *P* value
C7	Active TSS lncRNAs	98%	18%	0%	328 (178/150)	1.19	0.14	1.66	0.03
C12	PhastCons	59%	100%	8%	896 (495/401)	1.23	0.002	1.22	0.003
C18	Active TSS	100%	16%	10%	3097 (1600/1497)	1.07	0.07	1.1	0.03
C20	PhyloP	82%	100%	14%	164 (100/64)	1.56	0.006	1.48	0.03
C26	Processed transcripts	57%	20%	0%	51 (34/17)	2	0.02	2.39	0.009
C28	CHD8 targets	100%	21%	100%	365 (183/182)	1.01	1	1.34	0.03
C42	Developmental delay genes	77%	11%	10%	93 (56/37)	1.51	0.06	2.06	0.02
C49	CHD8 targets and PhyloP	100%	100%	100%	23 (16/7)	2.29	0.09	2.43	0.01
C63	PhyloP	79%	100%	12%	143 (91/52)	1.75	0.001	1.87	0.03
	All promoters	53%	14%	5%	6787 (3458/3329)	1.04	0.12	—	—
	Active TSS group	100%	16%	10%	3570 (1815/1755)	1.03	0.32	—	—
	Conserved Loci group	60%	100%	8%	931 (522/409)	1.28	0.0002	—	—

distributions in cases and controls except for processed transcripts (17 in cases, 0 in controls). In cases, genes with promoter mutations in the Conserved Loci group are enriched for "regulation of cell differentiation" (GO:0045595, FDR = 0.02), "transcription, DNA-templated" (GO:0006351, FDR = 0.04), and "regulation of transcription by RNA polymerase II" (GO:0006357, FDR = 0.04), whereas no biological processes are enriched in controls (table S8). Comparing cases to controls, there are nonsignificant trends toward enrichment in cases for ASD-associated genes (5 in cases, 2 in controls) and several ASD-related gene lists: brain-expressed (*29*), constrained (*27*), or CHD8 targets (*8*, *9*, *30*) (fig. S9 and table S8).

In coding regions, ASD-associated genes can be identified by the presence of multiple independent PTVs in different cases disrupting the same gene (*1*). In the WGS data, this approach did not yield specific promoters, because similar numbers of promoters had multiple Conserved Loci mutations in cases and controls (11 promoters in cases versus 7 in controls; $P = 0.81$, Fisher exact test). An equivalent analysis of damaging missense mutations, split into 2000-bp blocks to simulate promoters, suggests that we lack the power to detect specific promoters in a cohort of this size (22 in cases, 17 in controls; $P = 1.00$).

Prior analyses of coding mutations have found large comorbid effects on nonverbal IQ, with ASD cases that carry ASD-associated mutations having a lower nonverbal IQ, on average (*1*). Excluding cases with de novo PTVs, we observed a 4-point reduction in median nonverbal IQ for cases with mutations in either the Active TSS [$P = 0.02$, Wilcoxon signed-rank test (WSRT)] and/or Conserved Loci ($P = 0.01$, WSRT) groups, relative to cases without such mutations (Fig. 3F). Furthermore, individuals with Conserved Loci promoter mutations show a trend toward a higher rate of mutations in female ASD cases (OR = 1.13; 95% CI = 0.74 to 1.73; $P = 0.31$, Fisher exact test) and increased incidence of nonfebrile seizures (OR = 1.46; 95% CI = 0.90 to 2.36; $P = 0.07$, Fisher exact test); both trends are consistent with results seen in coding mutations.

The distal promoter shows the strongest evidence of association, especially at transcription factor binding sites

Because promoters are defined by their relationship to the TSS (*31*), we considered how ASD association varied by TSS distance, with the expectation that association would diminish with distance from the TSS. We first examined four bins: the core promoter (≤80 bp), which we would expect to contain the TATA box, initiator element, and/or downstream promoter element; the proximal promoter (81 to 250 bp); and two divisions of distal promoters (251 to 1000 bp, 1001 to 2000 bp). In contrast to this expectation, mutations in the Conserved Loci group are most strongly enriched in the distal region (RR = 1.32; $P = 0.005$, binomial test; Fig. 4A). This distal association prompted us to consider only mutations at experimentally defined transcription factor binding sites (JASPAR CORE) (*32*), which enhanced the association (RR = 2.05; $P = 0.0003$, binomial test; Fig. 4B). Although a trend toward enrichment in cases is observed in the core promoter (Fig. 4, A and B), we do not see enrichment for motifs associated with RNA polymerase II (e.g., TATA; table S6). Looking at the enrichment in cases across the promoter in 200-bp sliding windows (Fig. 4, C and D), the strongest enrichment is observed between 750 and 2000 bp.

Discussion

These analyses used WGS from 7608 individuals with an unbiased genome-wide association framework to demonstrate that de novo noncoding mutations alter risk for a complex neurodevelopmental disorder (Fig. 2). In a recent study (*15*), we highlighted the importance of genome-wide analyses with appropriate correction for multiple testing to identify noncoding regions robustly associated with ASD. Following this principle, no single noncoding annotation category was significant after conservative correction for multiple testing (Fig. 1E). Similarly, we could not replicate candidate noncoding hypotheses described in previous analyses of ASD and developmental delay cohorts (table S4) (*12*–*14*, *22*, *33*). However, a "de novo risk score," developed from a genome-wide Lasso analysis of multiple noncoding annotation categories, was a significant predictor of ASD risk (Fig. 2A). Such scores are routinely used in genomic analyses, including polygenic risk scores of common variants and, recently, a rare variant risk score for coding mutations in schizophrenia (*34*). Consistent with expectations, the magnitude of the contribution from noncoding mutations is smaller than that of the coding region, even having excluded de novo PTVs (Fig. 2A). Yet this early iteration of a de novo risk score could underestimate the true risk conferred by all noncoding mutations, as has been seen for polygenic risk score from common variants in successively larger cohorts (*35*).

Enrichment of annotation terms in the de novo risk score reveals that it is mutations in promoter regions (defined as 2000 bp upstream of the TSS) that underlie this noncoding association with ASD (Fig. 2A); the risk score continues to demonstrate ASD association when considering only promoter categories (45 of 163 categories; Fig. 2A). A consistent association signal can be observed across all 1855 promoter categories (Fig. 2B) and for 931 mutations at conserved loci (Fig. 3E). Notably, ASD cases with conserved promoter mutations have lower nonverbal IQ scores than ASD cases without these mutations (Fig. 3F)—an effect also observed in children with ASD-associated PTV mutations and missense mutations (*1*). Within promoters, the most robust association is observed for promoter mutations at Conserved Loci (Table 1), particularly at known transcription factor binding sites (Fig. 4B) (*32*). At Conserved Loci, the relative risk is similar to that observed for de novo damaging missense mutations (Fig. 3E). It is possible that the true relative risk is somewhat smaller, a phenomenon seen many times when the genome is searched for loci of relatively small effect and often called the winner's curse. Surprisingly, the

strongest signal was not at the TSS and core promoter, but rather in the distal promoter, 750 to 2000 bp away from the TSS (Fig. 4). As expected for the distal promoter, the mutations in cases are frequently at experimentally defined transcription factor binding sites (Fig. 4D).

A key question is whether the de novo variation found in promoter regions is targeting the same set of genes implicated in ASD by de novo variants in protein-coding regions or a distinct set of genes not yet known to play a role in ASD. We favor the former possibility, although we cannot definitively exclude the latter, on the basis of (i) the enrichment for GO terms relating to transcriptional regulation and cell differentiation in the genes targeted by Conserved Loci mutations, terms that are also enriched in ASD-associated genes (*1*); (ii) the trend toward enrichment for ASD-associated genes and several other gene sets previously implicated in ASD (fig. S9); and (iii) the detection of clusters defined by developmental delay genes and CHD8 binding targets (Fig. 3A and Table 1), both of which are enriched for ASD risk genes.

Our analysis establishes a specific hypothesis that can be tested for replication in future ASD cohorts and assessed in developmental and neuropsychiatric disorder cohorts: De novo mutations at conserved loci (46 vertebrate species PhastCons ≥ 0.2 and/or 46 vertebrate species PhyloP ≥ 2) in promoter regions (2000 bp upstream of the TSS based on GENCODEv27 annotation with VEP) are associated with risk. To facilitate such analyses by others, we have generated a file of loci that meet these criteria (table S9). Despite these promising insights, we cannot yet identify which of the 522 conserved promoter mutations in cases truly confer risk, nor can we be confident which of the remaining 126,031 noncoding case mutations do not. Instead, our results demonstrate that elucidation of the contribution of de novo noncoding mutations to human disorders is feasible, and that the yields are likely to improve substantially with increases in cohort size (*10*, *15*).

That conserved loci are one of the major factors underlying the promoter association could be interpreted to mean that nonhuman models can be used to assay noncoding function in humans, although parallel work in humans will be required to show that the specific regulatory effects are also conserved. Enrichment at transcription factor binding sites is also promising. If ASD association can be detected for specific transcription factors or loci, it raises the prospect of high-resolution neurobiological insights into spatiotemporal development, especially when, where, and in which cell type typical development is disrupted in ASD. Such insights will require detailed functional data on transcription factors and how they relate to mutations found in ASD.

The association that we observe from these data represents the integration of work from multiple fields, including human cohort collections (*2*, *18*), gene definitions (*26*), comparative genomics (*23*, *24*), and functional genomics (*32*, *36*). Methods and infrastructure are being developed to replicate and refine this association, identify specific loci, or extend beyond promoters. These include larger cohorts with consistently analyzed WGS data [e.g., the WGSPD consortium (*10*)], refined annotation of noncoding regions in the human brain [e.g., the PsychENCODE consortium (*36*)], WGS-tailored analytical methods (*15*, *25*), and large-scale functional assays [e.g., massively parallel reporter assays (*37*)]. The evolving results from these fields provide a path to improving diagnosis and novel therapeutic strategies that could benefit a wide range of human disorders.

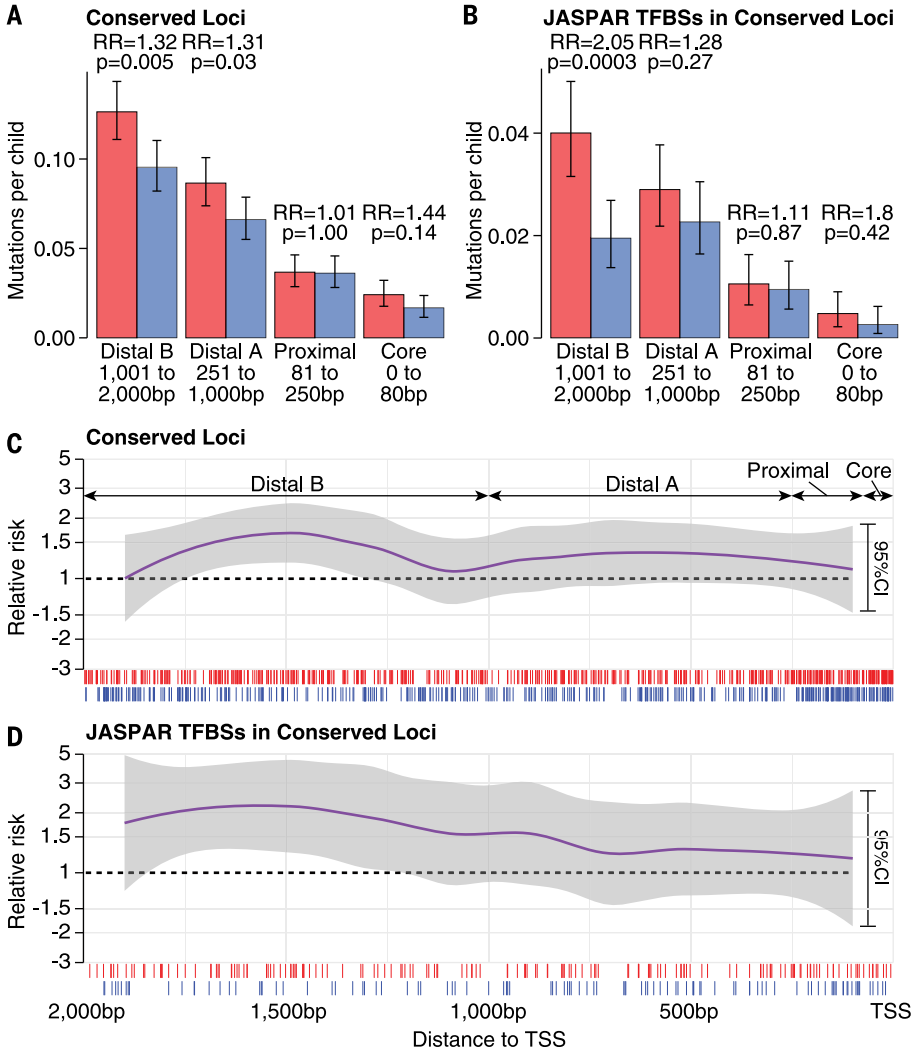


Fig. 4. Relationship of conserved promoter mutations to the TSS. (**A**) Frequency of Conserved Loci promoter mutations in cases and controls across the promoter region. (**B**) Frequency of Conserved Loci promoter mutations in cases and controls at JASPAR transcription factor binding sites (TFBSs) across the promoter region. (**C**) Enrichment of Conserved Loci promoter mutations in cases, shown as relative risk, in sliding windows of 200 bp across the promoter region. The purple line is the generalized additive model fit for relative risk and the 95% confidence interval is in gray. Ticks under the plot show individual mutations in cases (red) and controls (blue). (**D**) The plot in (C) is repeated for Conserved Loci promoter mutations at JASPAR TFBS. Statistical tests: binomial, two-sided [(A) and (B)]. Error bars show the 95% confidence interval (95% CI).

Materials and methods

See (*19*) for additional details.

Detection and annotation of de novo mutations

WGS data were generated by the New York Genome Center with a mean coverage of 35.5 in 1902 ASD quartet families. Previously described variant filtering criteria were applied (*15*) to identify 255,106 high-quality de novo mutations. These mutations were annotated using the Ensembl Variant Effect Predictor (VEP; version 90.4a44397) with GENCODE v27 gene definitions. Nucleotide sequence conservation across 46 vertebrate species (PhyloP, PhastCons), and regulatory regions (e.g., transcription factor binding sites, chromatin states) were annotated using VEP. In addition to 424 previously validated loci, 45 de novo mutations in promoter regions with two or more mutations in different samples

were validated as de novo by analyzing all four members of each family with PCR and Sanger sequencing.

Category-wide association study (CWAS)

To assess multiple hypotheses, we implemented the CWAS method, described in (*15*). Considering 70 annotation terms from five groups in combination defined 55,143 nonredundant categories for downstream analysis. ASD association was tested for each category by comparing the burden of case and control mutations with a two-sided binomial test, having corrected the rate of de novo mutations for paternal age. To estimate the penalty of multiple comparisons, the number of effective tests was estimated using Eigen decomposition of P values in 10,000 simulated datasets. Each simulated dataset contained 255,106 random variants and maintained the GC bias and proportion of SNVs to indels observed in the original data.

De novo risk score analysis

To build a de novo risk score, we excluded all categories that could contain de novo PTVs, then selected 8418 rare annotation categories with ≤3 mutations in controls. From the training dataset of 519 families described previously (*15*), we used a Lasso regression with five-fold cross-validation to estimate the regularization parameter, and then applied this fitted prediction model to the remaining 1383 new families to estimate the predictive power of the risk score. The significance of the prediction was calculated from 1000 permutations with case-control status swapped in 50% of families selected at random. The frequency of the 62 noncoding annotation terms was compared between the 36,828 nonredundant noncoding categories and the 163 noncoding categories in the de novo risk score. A binomial test was used to assess the enrichment of these terms, corrected for 62 comparisons.

DAWN clustering analysis of promoter categories

The DAWN hidden Markov random field model (*25*) was used to assess the risk factors underlying ASD association of promoters. Clusters of individual promoter categories were defined by K-means (K = 70) based on the P-value correlation network generated from 10,000 simulated datasets. Of these 70 clusters, 47 had at least 20 mutations and 2 categories and were considered further. Observed P values were transformed to z-scores and sparse PCA analysis was used to estimate the P value and relative risk per cluster. Using a hidden Markov random field model, these estimates were modified to yield a posterior probability based on enrichment in neighboring clusters in the simulated P-value correlation network.

REFERENCES AND NOTES

1. S. J. Sanders *et al*., Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* **87**, 1215–1233 (2015). doi: 10.1016/j.neuron.2015.09.016
2. Deciphering Developmental Disorders Study, Prevalence and architecture of de novo mutations in developmental disorders. *Nature* **542**, 433–438 (2017). doi: 10.1038/nature21062
3. H. O. Heyne *et al*., De novo variants in neurodevelopmental disorders with epilepsy. *Nat. Genet.* **50**, 1048–1053 (2018). doi: 10.1038/s41588-018-0143-7
4. S. C. Jin *et al*., Contribution of rare inherited and de novo variants in 2,871 congenital heart disease probands. *Nat. Genet.* **49**, 1593–1601 (2017). doi: 10.1038/ng.3970
5. C. Redin *et al*., The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. *Nat. Genet.* **49**, 36–45 (2016). doi: 10.1038/ng.3720
6. A. J. Willsey *et al*., Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013). doi: 10.1016/j.cell.2013.10.020
7. R. Ben-Shalom *et al*., Opposing effects on NaV1.2 function underlie differences between SCN2A variants observed in individuals with autism spectrum disorder or infantile seizures. *Biol. Psychiatry* **82**, 224–232 (2017). doi: 10.1016/j.biopsych.2017.01.009
8. A. Sugathan *et al*., CHD8 regulates neurodevelopmental pathways associated with autism spectrum disorder in neural progenitors. *Proc. Natl. Acad. Sci. U.S.A.* **111**, E4468–E4477 (2014). doi: 10.1073/pnas.1405266111
9. J. Cotney *et al*., The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat. Commun.* **6**, 6404 (2015). doi: 10.1038/ncomms7404
10. S. J. Sanders *et al*., Whole Genome Sequencing in Psychiatric Disorders: The WGSPD Consortium. *Nat. Neurosci.* **20**, 1661–1668 (2017). doi: 10.1038/s41593-017-0017-9
11. F. Zhang, J. R. Lupski, Non-coding genetic variants in human disease. *Hum. Mol. Genet.* **24**, R102–R110 (2015). doi: 10.1093/hmg/ddv259
12. P. J. Short *et al*., De novo mutations in regulatory elements in neurodevelopmental disorders. *Nature* **555**, 611–616 (2018). doi: 10.1038/nature25983
13. T. N. N. Turner *et al*., Genome Sequencing of Autism-Affected Families Reveals Disruption of Putative Noncoding Regulatory DNA. *Am. J. Hum. Genet.* **98**, 58–74 (2016). doi: 10.1016/j.ajhg.2015.11.023
14. T. N. Turner *et al*., Genomic Patterns of De Novo Mutation in Simplex Autism. *Cell* **171**, 710–722.e12 (2017). doi: 10.1016/j.cell.2017.08.047
15. D. M. Werling *et al*., An analytical framework for whole-genome sequence association studies and its implications for autism spectrum disorder. *Nat. Genet.* **50**, 727–736 (2018). doi: 10.1038/s41588-018-0107-y
16. W. M. Brandler *et al*., Paternally inherited cis-regulatory structural variants are associated with autism. *Science* **360**, 327–331 (2018). doi: 10.1126/science.aan2261
17. F. Dudbridge, A. Gusnanto, Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* **32**, 227–234 (2008). doi: 10.1002/gepi.20297
18. G. D. Fischbach, C. Lord, The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron* **68**, 192–195 (2010). doi: 10.1016/j.neuron.2010.10.006
19. See supplementary materials.
20. H. Jónsson *et al*., Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* **549**, 519–522 (2017). doi: 10.1038/nature24018
21. S. De Rubeis *et al*., Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014). doi: 10.1038/nature13772
22. A. Munoz *et al*., De novo indels within introns contribute to ASD incidence. BioRxiv 137471 [Preprint]. 24 May 2017.
23. A. Siepel *et al*., Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005). doi: 10.1101/gr.3715005
24. K. S. Pollard, M. J. Hubisz, K. R. Rosenbloom, A. Siepel, Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010). doi: 10.1101/gr.097857.109
25. L. Liu *et al*., DAWN: A framework to identify autism genes and subnetworks using gene expression and genetics. *Mol. Autism* **5**, 22 (2014). doi: 10.1186/2040-2392-5-22
26. J. Harrow *et al*., GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res.* **22**, 1760–1774 (2012). doi: 10.1101/gr.135350.111
27. M. Lek *et al*., Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016). doi: 10.1038/nature19057
28. J. A. Kosmicki *et al*., Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. *Nat. Genet.* **49**, 504–510 (2017). doi: 10.1038/ng.3789
29. H. J. Kang *et al*., Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011). doi: 10.1038/nature10523
30. J. C. Darnell *et al*., FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011). doi: 10.1016/j.cell.2011.06.013
31. H. Kwak, N. J. Fuda, L. J. Core, J. T. Lis, Precise Maps of RNA Polymerase Reveal How Promoters Direct Initiation and Pausing. *Science* **339**, 950–953 (2013). doi: 10.1126/science.1229386
32. A. Khan *et al*., JASPAR 2018: Update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D260–D266 (2018). doi: 10.1093/nar/gkx1126
33. R. K. C. Yuen *et al*., Genome-wide characteristics of de novo mutations in autism. *npj Genomic Med.* **1**, 16027 (2016). doi: 10.1038/npjgenmed.2016.27
34. S. M. Purcell *et al*., A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014). doi: 10.1038/nature12975
35. N. Chatterjee *et al*., Projecting the performance of risk prediction based on polygenic analyses of genome-wide association studies. *Nat. Genet.* **45**, 400–405 (2013). doi: 10.1038/ng.2579
36. S. Akbarian *et al*., The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015). doi: 10.1038/nn.4156
37. F. Inoue *et al*., A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity. *Genome Res.* **27**, 38–52 (2017). doi: 10.1101/gr.212092.116

ACKNOWLEDGMENTS

We are grateful to all the families participating in this research, including the Simons Foundation Autism Research Initiative (SFARI) Simplex Collection (SSC) and Korean cohort. We thank the SSC principal investigators (A. L. Beaudet, R. Bernier, J. Constantino, E. H. Cook Jr., E. Fombonne, D. Geschwind, D. E. Grice, A. Klin, D. H. Ledbetter, C. Lord, C. L. Martin, D. M. Martin, R. Maxim, J. Miles, O. Ousley, B. Peterson, J. Piggot, C. Saulnier, M. W. State, W. Stone, J. S. Sutcliffe, C. A. Walsh, and E. Wijsman) and the coordinators and staff at the SSC clinical sites; the SFARI staff, in particular N. Volfovsky; D. B. Goldstein and K. C. Eggan for contributing to the experimental design; the Rutgers University Cell and DNA repository for accessing biomaterials; and the New York Genome Center for generating the WGS data. Annotation data was generated as part of the PsychENCODE Consortium, supported by NIH grants U01MH103392, U01MH103365, U01MH103346, U01MH103340, U01MH103339, R21MH109956, R21MH105881, R21MH105853, R21MH103877, R21MH102791, R01MH111721, R01MH110928, R01MH110927, R01MH110926, R01MH110921, R01MH110920, R01MH110905, R01MH109715, R01MH109677, R01MH105898, R01MH105898, R01MH094714, R01MH109901, P50MH106934, 5R24HD000836 and SFARI #307705 awarded to S. Akbarian (Icahn School of Medicine at Mount Sinai), G. Crawford (Duke University), S. Dracheva (Icahn School of Medicine at Mount Sinai), P. Farnham (University of Southern California), M. Gerstein (Yale University), D. Geschwind (University of California, Los Angeles), I. Glass (Washington University), F. Goes (Johns Hopkins University), T. M. Hyde (Lieber Institute for Brain Development), A. Jaffe (Lieber Institute for Brain Development), J. A. Knowles (University of Southern California), C. Liu (SUNY Upstate Medical University), D. Pinto (Icahn School of Medicine at Mount Sinai), P. Roussos (Icahn School of Medicine at Mount Sinai), P. Sklar (Icahn School of Medicine at Mount Sinai), P. Sullivan (University of North Carolina), F. Vaccarino (Yale University), D. Weinberger (Lieber Institute for Brain Development), S. Weissman (Yale University), K. White (University of Chicago), P. Zandi (Johns Hopkins University), S.J.S., N.S., M.W.S., and A.J.W. **Funding:** Supported by Simons Foundation for Autism Research Initiative (SFARI) grants 402281 (S.J.S., M.W.S., B.D., and K.R.), 385110 (S.J.S., M.W.S., A.J.W., and N.S.), 574598 (S.J.S.), 385027 (M.E.T., B.D., K.R., J.B., and M.J.D.), 346042 (M.E.T.), 575097 (B.D. and K.R.), 573206 (M.E.T.), 513631 (G.T.M.), and 388196 (H.C. and G.T.M.); NIH grants U01 MH105575 (M.W.S.), U01 MH100239-03S1 (M.W.S., S.J.S., and A.J.W.), R01 MH110928 (S.J.S., M.W.S., and A.J.W.), R01 MH109901 (S.J.S., M.W.S., and A.J.W.), U01 MH111662 (S.J.S. and M.W.S.), U01 MH111658 (B.D.), U01 MH111660 (M.J.D.), U01 MH111661 (J.D.B.), R37 MH057881 (B.D. and K.R.), R01 HD081256 (M.E.T.), R01 MH115957 (M.E.T.), R01 MH049428 (J.L.R.), R01 MH107649-03 (B.M.N.), and R01 MH094400 (H.C.); and the Stanley Center for Psychiatric Genetics.

Author contributions: Experimental design, J.-Y.A., K.L., L.Z., D.M.W., H.B., N.A., J.D.B., H.C., M.J.D., Y.S.K., G.T.M., B.M.N., A.R.Q., J.L.R., N.S., M.W.S., A.J.W., M.E.T., B.D., K.R., and S.J.S.; data generation, G.B.S., J.D., C.D., Y.S.K., and S.J.S.; data processing, J.-Y.A., D.M.W., S.D., C.D., M.C.G., L.L., and S.J.S.; annotation of functional regions, J.-Y.A., D.M.W., E.M.-P., S.P., J.L.R., N.S., and S.J.S.; data analysis, J.-Y.A., K.L., L.Z., D.M.W., S.D., H.B., H.Z.W., X.Z., G.B.S., R.L.C., B.B.C., L.K., M.E.T., B.D., K.R., and S.J.S.; statistical analysis, J.-Y.A., K.L., L.Z., D.M.W., L.K., M.E.T., B.D., K.R., and S.J.S.; manuscript preparation, J.-Y.A., K.L., L.Z., D.M.W., S.D., N.A., H.C., G.T.M., M.E.T., B.D., K.R., and S.J.S. **Competing interests:** G.T.M. is co-founder and chief scientific officer of Frameshift Labs Inc. B.M.N. is a member of the Deep Genomics scientific advisory board and a consultant for Camp4 Therapeutics Corporation, Merck & Co., and Avanir Pharmaceuticals. M.W.S. is on the scientific advisory boards for ArRett Pharmaceuticals and BlackThorn Therapeutics and holds stock options in ArRett Pharmaceuticals. **Data and code availability:** All sequencing and phenotype data are hosted by the Simons Foundation for Autism Research Initiative (SFARI) and are available for approved researchers at SFARIbase (https://base.sfari.org/, accession ID: SFARI_SSC_WGS_p, SFARI_SSC_WGS_1, and SFARI_SSC_WGS_2). The code for running a category-wide association study can be accessed at http://doi.org/10.5281/zenodo.1489239. For methods for estimating the de novo risk score, clustering of annotation categories, and estimating the significance of clusters of annotation categories, see http://doi.org/10.5281/zenodo.1489250.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/eaat6576/suppl/DC1
Figs. S1 to S9
Tables S1 to S9
References (*38*–*52*)

21 March 2018; accepted 15 November 2018
10.1126/science.aat6576

RESEARCH ARTICLE

SUPERCONDUCTIVITY

Revealing hidden spin-momentum locking in a high-temperature cuprate superconductor

Kenneth Gotlieb[1,2]*, Chiu-Yun Lin[2,3]*, Maksym Serbyn[4], Wentao Zhang[2,5], Christopher L. Smallwood[2,3]†, Christopher Jozwiak[6], Hiroshi Eisaki[7], Zahid Hussain[6], Ashvin Vishwanath[8], Alessandra Lanzara[2,3]‡

Cuprate superconductors have long been thought of as having strong electronic correlations but negligible spin-orbit coupling. Using spin- and angle-resolved photoemission spectroscopy, we discovered that one of the most studied cuprate superconductors, Bi2212, has a nontrivial spin texture with a spin-momentum locking that circles the Brillouin zone center and a spin-layer locking that allows states of opposite spin to be localized in different parts of the unit cell. Our findings pose challenges for the vast majority of models of cuprates, such as the Hubbard model and its variants, where spin-orbit interaction has been mostly neglected, and open the intriguing question of how the high-temperature superconducting state emerges in the presence of this nontrivial spin texture.

Many of the exotic properties of quantum materials stem from the strength of spin-orbit coupling or electron-electron correlations. At one end of the spectrum are topological insulators, which have weak electron correlations but strong spin-orbit coupling (*1*, *2*); at the other end are cuprate superconductors, where electron correlations are the dominant interaction. Although unusual forms of spin response in the cuprates have been reported previously (*3*, *4*), the spin-orbit interaction has been mostly neglected or treated as a small perturbation to the Hubbard Hamiltonian and mean field theory in the context of the Dzyaloshinskii-Moriya interaction, leading to negligible changes to the electronic ground state of cuprates (*5*–*9*).

Recently, there has been an upsurge of interest in materials in which both spin-orbit coupling and strong correlations are important because of their potential to induce exotic quantum states (*10*–*13*). In the presence of superconductivity, for example, spin-orbit interaction can have fundamental consequences for the symmetry of the order parameter (*14*), driving unusual pairing mechanisms (*11*, *15*), creating Ising pairs (*16*), and even realizing the conditions for the existence of previously unobserved particles (*17*–*19*).

Spin- and angle-resolved photoemission spectroscopy (SARPES) has been instrumental in studying the consequences of such interplay for the electronic structure of a variety of materials, from heavy fermions to iridates (*20*, *21*), thanks to its ability to simultaneously probe the energy, momentum, and spin structure of quasiparticles. However, because of earlier predictions of negligible spin-orbit interaction in cuprates (*6*), the full spin character of quasiparticles has not been probed experimentally. Here, we report such a study, revealing unexpected consequences of the spin-orbit interaction for the electronic structure of cuprates.

SARPES measurements of overdoped Bi2212

We studied the spin-dependent character of overdoped $Bi_2Sr_2CaCu_2O_{8+\delta}$ (Bi2212) samples (with the superconducting transition temperature T_c = 58 K) with SARPES over a wide range of energies, momenta, temperatures, and photon energies. We performed 10 distinct measurements by coupling our efficient spectrometer (*22*) to a 6-eV pulsed laser source and synchrotron light of different photon energies. The in-plane components of the quasiparticle's spin polarization (P_x, P_y) were mapped as a function of energy and momentum over the entire Brillouin zone, in both the normal and superconducting states [for comparison, see (*23*)]. The spin spectrometer used in this study (*24*) more readily measures in-plane components of spin than the out-of-plane component (P_z). However, as we discuss later, we expect the latter to be negligible and found it to be zero within experimental uncertainty. Figure 1 shows the low-temperature spin-integrated (Fig. 1, B and E) and spin-resolved (Fig. 1, C and F) maps of energy ($E - E_F$) versus momentum (k) of the quasiparticle spectrum, where E_F is the Fermi energy. Data are shown for two different momentum cuts: along the nodal direction (Γ-Y) (Fig. 1, B and C), where the superconducting gap is zero, and along an off-nodal direction (Fig. 1, E and F), where the superconducting gap is ~10 meV. The location of the cuts (thick black line) and the photoelectron spin components (blue and red arrows) are shown in the insets of Fig. 1, B to F. In Fig. 1 and the rest of the figures, we use blue and red to indicate the two opposite spin components along a given direction, and we hereafter refer to these components as spin-up and spin-down, respectively. The spin polarimeter we used is not subject to the instrumental asymmetries typical of Mott-type detectors that require calibration or renormalization (*24*). The spin polarization measured in this study is therefore intrinsic to the photoelectrons.

Figure 1 summarizes the most surprising findings of this work: the presence of a nonzero spin polarization in Bi2212 and its strong dependence on momentum. Along the nodal direction, we find that the photoelectron spin component perpendicular to Γ-Y is strongly polarized up, as shown by the spin-resolved intensity map in Fig. 1C, which is primarily blue. The corresponding spin polarization P, defined as the relative difference between the numbers of spin-up and spin-down photoelectrons according to $P = (I_\uparrow - I_\downarrow)/(I_\uparrow + I_\downarrow)$, is positive along this entire cut (Fig. 1D). The polarization shows an overall increase as a function of momentum (or energy) from roughly +20% at the Fermi momentum, k_F (Fermi energy, E_F), to as much as +40% for smaller momenta (or higher binding energies), i.e., closer to the Brillouin zone center, Γ.

Notably, when we move away from the nodal direction, the perpendicular photoelectron spin component reverses and is strongly polarized downward, as seen in the spin-resolved intensity map in Fig. 1F, which is primarily red. The reversal of the intensity peak from primarily spin-up to primarily spin-down can be clearly seen in Fig. 1H, where the SARPES spectra at k_F as a function of energy [energy distribution curves (EDCs)] are directly compared for both the nodal and off-nodal cuts.

A closer look reveals a similar increase of the value of spin polarization for the off-nodal cut (Fig. 1G) toward smaller $|k|$ or higher binding energy. In this case, the polarization is negative (P = −15%) at k_F but eventually turns slightly positive (P = +5%) at higher binding energy. In summary, along both of these cuts, we observed an unexpected nonzero spin polarization that

[1]Graduate Group in Applied Science and Technology, University of California, Berkeley, CA 94720, USA. [2]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [3]Department of Physics, University of California, Berkeley, CA 94720, USA. [4]Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria. [5]School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai 200240, China. [6]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [7]Electronics and Photonics Research Institute, National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki 305-8568, Japan. [8]Department of Physics, Harvard University, Cambridge, MA 02138, USA.

*These authors contributed equally to this work. †Present address: Department of Physics and Astronomy, San José State University, San José, CA 95192, USA.

‡Corresponding author. Email: alanzara@lbl.gov

becomes more positive as one goes toward higher binding energies (i.e., deeper inside the Fermi surface). The observed nonzero spin polarization has been reproduced under different experimental conditions, with different samples and sample surfaces [(Figs. 1 to 4) and (*23*)], different geometry (*25*), and several photon energies. The effect also persists after sample surface exposure to a vacuum of $\approx 5 \times 10^{-11}$ torr over several days, the time scale over which some of the experiments described herein were conducted.

Figure 2 shows the evolution of the photoelectron spin polarization along the Fermi surface and at a binding energy of 160 meV (see, e.g., vertical lines in Fig. 1, C and F). The spin-resolved EDCs at k_F and at smaller momenta k_{HBE} (where HBE indicates high binding energy) are shown in Fig. 2, A and B, respectively; the location of each spectrum is shown in Fig. 2C. In both cases, we observe a net spin polarization that decreases away from the node ($\phi = 0°$), eventually reaches zero at an intermediate angle, and for the spectra at $k = k_F$, even switches sign far away from the node. These results are summarized quantitatively Fig. 2D for both $k = k_F$ and $k = k_{HBE}$ [for the full energy dependence of the spin polarization, see (*23*)]. The spin polarization is approximately even about the nodal line, where it reaches its maximum with values as high as +40%. Notably, it is higher at k_{HBE} than at k_F over the entire angular range. On the Fermi surface, the two spin channels $I_\uparrow$ and $I_\downarrow$ are each stronger in different parts of momentum space. By contrast, at higher binding energy (Fig. 2F), the dominant spin channel is spin-up, yielding an overall positive spin polarization.

The presence of any spin polarization in photoemission from Bi2212, let alone a momentum-dependent spin texture, is unexpected. It is therefore imperative, before proceeding to discuss the total spin texture, to assess whether the observed spin polarization is the result of a final state effect or represents physics intrinsic to the spin state of itinerant carriers in the material.

Figure 3 shows the evolution of the spin polarization across the Brillouin zone boundary (M point) (Fig. 3B) and Brillouin zone center (Γ point) (Fig. 3D). Spin-resolved EDCs are shown in Fig. 3B adjacent to the two opposite M points within the first Brillouin zone (points β and γ) and for a point just across the Brillouin zone boundary (α) that is separated by a reciprocal lattice vector from γ. The locations of these measurements are represented by vertical arrows in Fig. 3A. To access this momentum window, we used higher-energy photons: 33 eV. The experimental geometry is shown in fig. S3A, and the measured spin component is perpendicular to the Γ-M direction.

The data show a clear reversal of this component of spin polarization at the two opposite zone boundaries (curves β and γ) and across the zone boundary (curves α and β). The observation of a reversal of the spin polarization at two points very near in emission angle (curves α and β) but on opposite sides of the zone boundary, as well as similar polarizations for points separated by a reciprocal lattice vector and hence having similar momenta (curves α and γ) but nearly opposite emission angles, confirms the intrinsic nature of the effect and its dependence on quasiparticle momentum rather than photoemission angle. Moreover, the presence of a nonzero spin polarization at different photon energies (fig. S3) contributes to the evidence that the observed effect is a property of the quasiparticle initial state rather than being a final state effect.

Final state versus intrinsic effect

We can learn more about the pattern of spin polarization across momentum space by using a well-known property of Bi2212: the presence of an incommensurate superstructure along the *b* axis caused by the modulation of Bi-O layers. This structural distortion creates umklapp bands that are replicas of the main band on the Fermi surface (dotted lines in Fig. 3A), shifted by the superstructure vector along the Γ-Y direction (*26*, *27*). Therefore, the second-order superstructures of the main band, labeled SS1 and SS2, lie near Γ.

These replica bands are clearly visible in the $h\nu = 6$ eV angle-resolved photoemission spectroscopy (ARPES) intensity maps (where *h* is Planck's constant and ν is frequency) (Fig. 3C) at the two opposite sides of the Γ point and disperse up toward Γ. The spin-resolved EDCs at k_F, measured along the dashed lines in Fig. 3C, are shown in Fig. 3D and measure

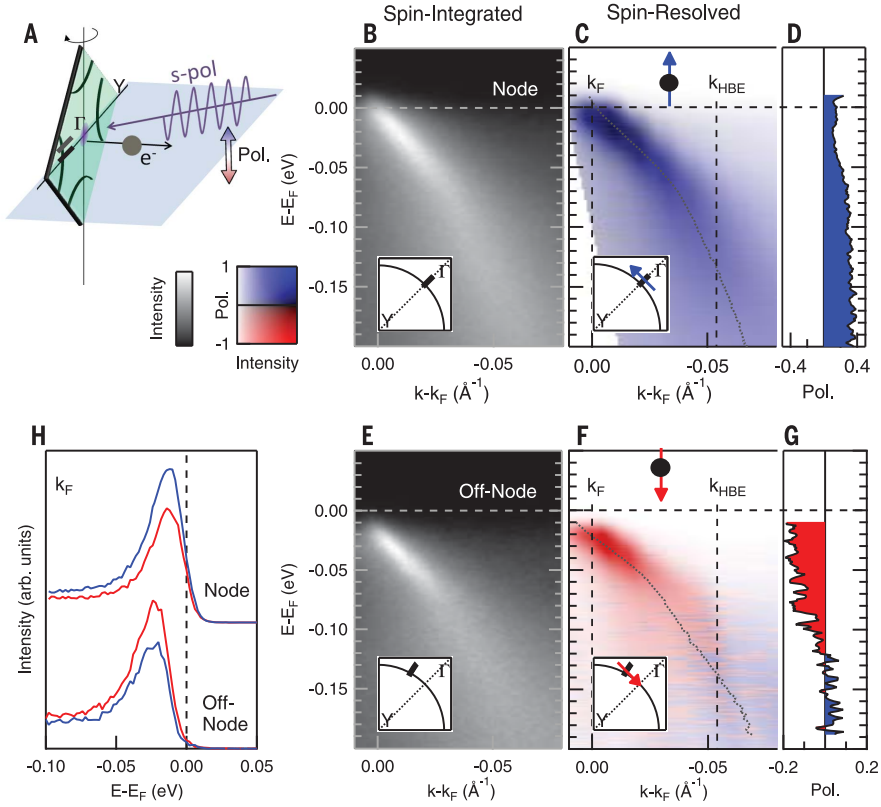


Fig. 1. Spin-resolved measurements along nodal (Γ-Y) and off-nodal cuts. (**A**) Experimental geometry. Pol., polarization; s-pol, s-polarized photons; e⁻, electron. (**B**) Spin-integrated map of the band near E_F along the nodal direction. (**C**) Spin-resolved map taken along the same cut as in (B), with darkness representing photoemission intensity $I_\uparrow + I_\downarrow$ and color representing spin polarization *P* [see the color scale in (A)]. Momenta k_F and k_{HBE} are the positions of measurements in Fig. 2 where the band is at the Fermi level and high binding energy, respectively. (**D**) Plot of the spin polarization along the band dispersion [dotted gray line in (C)]. (**E** to **G**) Same as (B) to (D) but measured along a cut parallel to the nodal direction that intersects the Fermi surface 14° away from the node, as measured from the zone corner. The same spin component was measured in (B) to (D) and (E) to (G). Insets in (B) to (F) show the location of the cuts (thick black line) and the photoelectron spin components (arrows). In this and subsequent figures, blue and red represent spin-up and spin-down, respectively. (**H**) Spin-resolved EDCs taken at the node, as well as at the Fermi momentum away from the node. arb., arbitrary.

the component of the photoelectron spin perpendicular to the Γ-Y direction. Two clear observations can be made from the data. The first one is that the superstructure bands on the two sides of the Γ point have opposite spin polarization, as seen in the EDCs for SS1 and SS2 in Fig. 3D. This reversal of the spin component through a small angle across the Brillouin zone center (SS1 versus SS2) corroborates the reversal seen at opposite momenta in EDCs β and γ, pointing to a spin polarization that not only is a function of k but also respects time reversal symmetry by switching sign across the Γ point.

The second observation is that at the node, the superstructure bands show opposite spin polarization with respect to the main bands of which they represent a second-order replica. That is, they match the spin of the main band in the same quadrant of momentum space. Though the superstructure band SS2 at $+k$ is the second-order replica of the main band MB2 at $-k$, the spin direction is opposite to that of MB2 (see MB2 in Figs. 3A and 1C for the relative spin polarization). It is the superstructure band SS1 at $-k$ that matches the positive spin polarization of the main band MB2 at $-k$. A more detailed explanation for the opposite value of spin polarization in the replica band relative to that of its "parent band" is found in (*23*); these results provide additional evidence that the observed spin polarization reflects the spin structure of the material bands.

In summary, the dependence of the spin polarization on quasiparticle momentum; the changes in the sign of polarization across the Brillouin zone center and boundaries; the observation of nonzero spin polarizations for different photon energies and geometries with spin alternately parallel and perpendicular to the electric field of light (see Fig. 1 and fig. S3A for more details); and the large values of spin polarization, up to 40%, strongly suggest that the observed effect is intrinsic and cannot be explained solely by an interference between photoemission pathways, as recently proposed (*25*). These findings point to an initial state with a well-defined spin texture in momentum space.

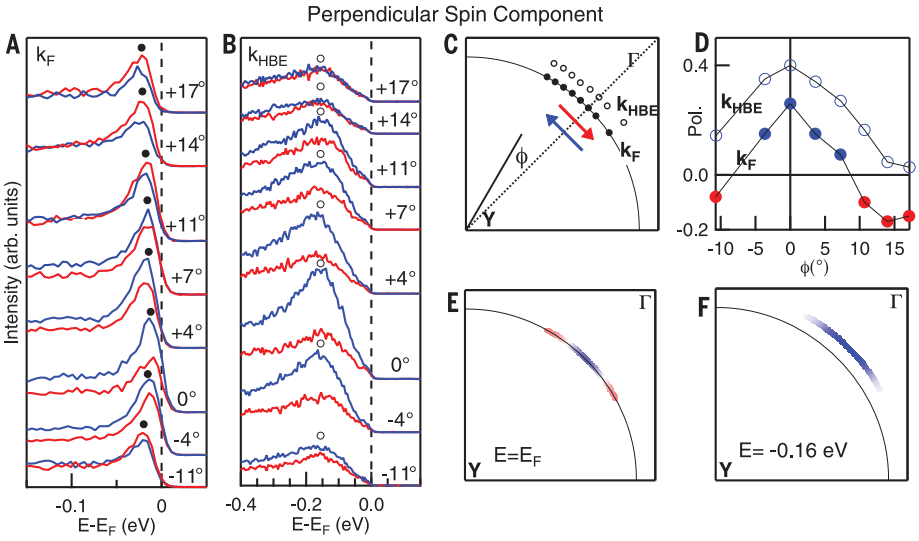


Fig. 2. Spin-resolved measurements along the Fermi surface and at higher binding energy. (**A**) Spin-resolved EDCs taken at momenta along the Fermi surface, as well as (**B**) inside the Fermi surface where the dispersion is at $E_B \approx 160$ meV (where E_B is binding energy). EDCs are marked by ϕ, the angle from the zone corner (Y point) to k_F, and are taken at momenta indicated in (**C**) one quadrant of the Brillouin zone. The spin component measured was perpendicular to the Γ-Y direction and within the plane of the sample surface. (**D**) Spin polarization as a function of the Fermi surface angle, ϕ, at k_F (solid circles) and at higher binding energy (hollow circles). (**E** and **F**) Schematics of the texture of this spin component.

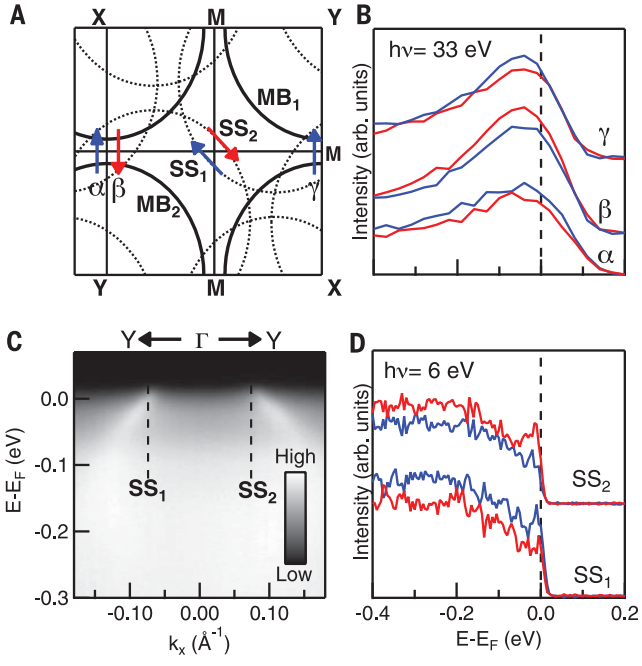


Fig. 3. Measured spin polarization near M points and spin polarization of the superstructures on either side of Γ. (**A**) Spin textures from the two distinct experiments in (**B**) and (**D**) plotted in the Bi2212 Brillouin zone. The main band is shown with thick lines, and its superstructure replicas are shown as thin dotted lines. (B) Spin-resolved EDCs taken with $h\nu$ = 33 eV at momenta shown in (A) near the M points. (**C**) Spin-integrated map of the superstructure taken with $h\nu$ = 6 eV, showing bands that replicate the main band dispersing up as they approach Γ. The dashed lines indicate approximate positions of spin-resolved measurements. (D) Spin-resolved EDCs on either side of Γ.

Full spin texture

Figure 4, A and B, shows the measured momentum-dependent spin polarization parallel to the Γ-Y direction, orthogonal to the spin component presented in Figs. 1 and 2 for several momenta. Spin-resolved EDCs for several momentum cuts are shown in Fig. 4A. For the nodal cut (ε), the intensity peaks are quite similar for the two spin components (Fig. 4A), resulting in nearly zero orthogonal spin polarization (Fig. 4B). At the same time, we see opposite spin polarization at cuts that are displaced by the same angle but in opposite directions from the node (δ and ζ), implying a reversal of the spin polarization component parallel to Γ-Y across the nodal point. Such a reversal is in contrast with the perpendicular spin component (see Fig. 2), which remains the same across the nodal direction.

The full spin texture across the Brillouin zone, obtained from the trends about the nodal line of the parallel and perpendicular spin components, is shown in Fig. 4C. The reversal of the spin polarization across the Γ-Y symmetry line (Fig. 4, A and B) and across the Brillouin zone quadrants (Figs. 2, A and B, and 3D), together with the spin polarization of replica bands, is consistent with a spin texture circling the Brillouin zone center (Γ) clockwise. Meanwhile, at larger k, the larger angle (ϕ)

measurements in Fig. 2B with spin pointing in the direction opposite that at small ϕ indicate that the texture has a more complex momentum dependence. One possibility is a change in the rotation direction of the spin pattern upon approaching boundaries of the Brillouin zone, sketched in gray in Fig. 4C.

The spin-momentum locking inferred in Fig. 4C is reminiscent of a Rashba-type effect. In typical observations of the Rashba effect, however, two bands of opposite spin polarization are split in energy. In this study, we observed only a single spin polarization at any particular momentum, regardless of the band's binding energy at that point. This leads to a single spin texture in k space.

Local inversion symmetry breaking

We now present a possible explanation for the observed spin polarization and its momentum dependence and discuss possible implications for superconductivity. Perhaps the most studied spin texture is the Dresselhaus-Rashba effect (*28*, *29*), which is manifested in noncentrosymmetric materials (i.e., materials lacking inversion symmetry) and gives rise to spin-dependent effects, inducing a momentum spin-splitting of the energy bands. Recently, it has been pointed out that even in centrosymmetric materials, a local electric field within the unit cell can lead to spin-split bands (*30*) whereas the net spin polarization remains zero as the electric field averages to zero within the unit cell. This local field can originate from specific structural characteristics that break local inversion symmetry centered on Cu atoms, such as layered structures or some types of lattice distortions that are present in the cuprates (*31*–*36*). In the case of a layered structure, the local field is perpendicular to the planes and the spin-split bands are spatially segregated in real space on top and bottom layers (*30*). In the case of a structural distortion, the spin-split bands are segregated within different parts of the unit cell. The model in (*30*) has been successfully applied to account for the nontrivial spin polarization observed in layered dichalcogenides (*37*, *38*) and a BiS_2-based superconductor (*39*), as well as to explain the nonzero nodal energy splitting between bonding and antibonding bands in a $YBa_2Cu_3O_{6+\delta}$ cuprate superconductor (*9*).

We extend this model to the case of bilayer Bi2212 by using a tight-binding model in the presence of a local electric field, treated via Rashba-type spin-orbit coupling, as in (*30*); the details of the calculations are shown in (*23*). The field is induced by the local breaking of inversion symmetry in Bi2212. Although the crystallographic space group of Bi2212 is often regarded as centrosymmetric (*40*), the local environment of Cu is noncentrosymmetric: The Ca layer separating two Cu-O planes removes the inversion center from Cu. Each Cu-O layer is now subject to a different environment: One Cu-O layer has Bi-O ions above and Ca ions below, whereas this is reversed for the other layer in the unit cell, allowing for a nonzero electric field within the unit cell (see the schematic in Fig. 5A).

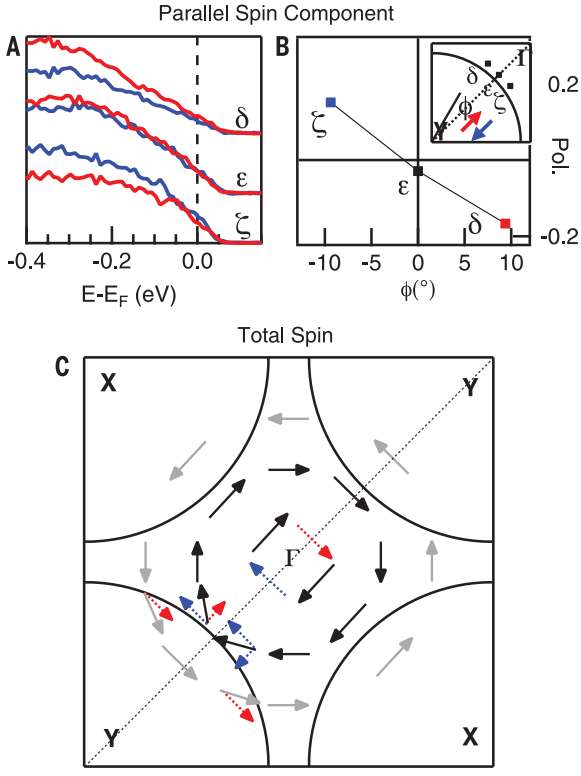


Fig. 4. Total in-plane spin texture. (**A**) Spin-resolved EDCs, acquired with sensitivity to the component of spin parallel to Γ-Y. (**B**) Spin polarization as a function of the Fermi surface angle, ϕ. The inset shows the positions in one quadrant of the Brillouin zone where EDCs were taken. (**C**) Schematic of the addition of the spin textures parallel [from (B)] and perpendicular (from Fig. 2D) to the Γ-Y direction. The counterclockwise circle of gray arrows is consistent with the one component of spin we were able to measure at high k [see (*23*) for further discussion of possible complex spin textures].

Although one would expect both Rashba and Dresselhaus contributions to spin-orbit coupling [R2 and D2 according to the notations in (*30*)], it appears that the dominant components in our experiments come from the Rashba order. This is likely a consequence of the strong anisotropy between *ab* and *c* axes in Bi2212, making the Dresselhaus component subleading. Upon the addition of such spin-orbit coupling, the former bonding (antibonding) band loses its purely antisymmetric (symmetric) character under mirror symmetry. However, we retain this naming convention herein. Both bonding and antibonding bands remain doubly degenerate at any momentum in the Brillouin zone as the crystal retains unbroken inversion and time reversal symmetries. However, these bands acquire spin-momentum locking with opposite spin polarization on each individual Cu-O layer. The spin textures for the antibonding orbital in the two Cu-O layers that result from this model are shown in Fig. 5B. Photoemission measures the interference pattern of contributions from several near-surface layers (*41*) and in this case has different intensity from bonding and antibonding bands (*42*, *43*). Therefore, a nonzero spin signal is expected, despite inversion symmetry and the lack of resolved band splitting. This spin texture stems from differences in photoemission matrix elements for different components of the wave function, as well as the surface sensitivity of the measurement and interference effects. We find that the spin polarization alternates as a function of photon energy, as discussed in (*23*), similarly to the change in the relative strength of photoemission intensity from bonding and antibonding bands (*44*). However, this could also be the result of a more complex dependence of the spin-orbit entanglement on photon energy, as shown extensively in other spin-orbit–coupled materials, such as topological insulators (*41*, *45*), where the sign of spin polarization can change with photon energy and even be zero; more detailed studies and calculations are needed.

By extending our tight-binding model to incorporate interference effects, we remove the perfect cancellation of spin polarizations between bonding and antibonding bands and get a spin texture that reverses sign across the Fermi surface (fig. S6). In addition, the interference effects can also explain the opposite direction of spin polarization between the original bands and their superstructure replicas shown in Fig. 3, as discussed in detail in (*23*).

Although our model can reproduce qualitative aspects of the spin polarization observed in our experiment, it does not capture the magnitude and precise momentum dependence of the spin, which require more involved calculations. Reports in favor of a noncentrosymmetric space group for Bi2212 (*31*, *32*, *46*) might simply argue that it is the absence of any inversion center that allows for the reported nonzero spin texture, as in a standard Rashba system, rather than the creation of a local field. Such a scenario, however, would imply the presence of spin-split bands that have not yet been observed. Moreover, some of the structural

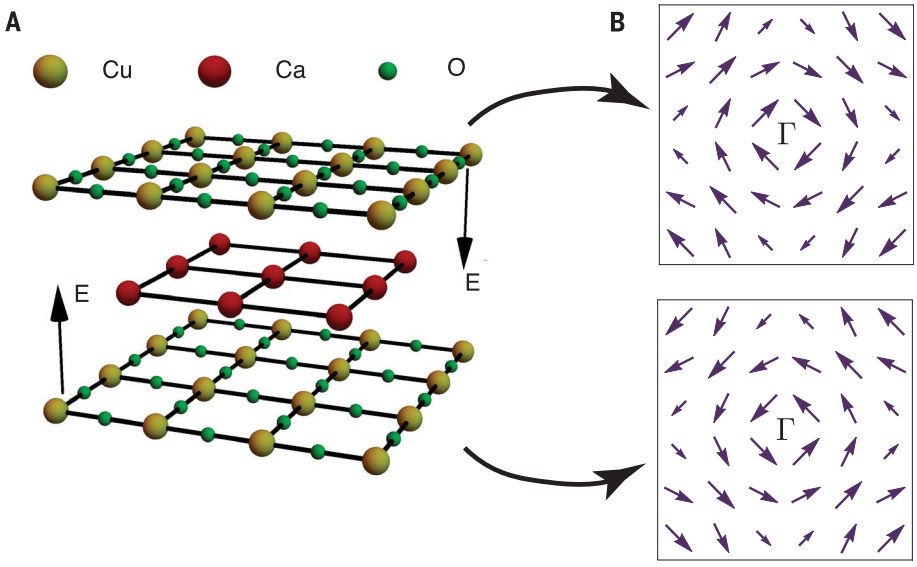


Fig. 5. Spin structure within the unit cell. (**A**) Schematic view of the two-CuO_2 bilayer structure in $Bi_2Sr_2CaCu_2O_{8+\delta}$, where we omit layers of Bi-O and Sr-O which separate bilayers. Green atoms correspond to oxygen, yellow to copper, and red atoms in between are Ca. Arrows schematically depict the possible direction of the electric field, which leads to the spin-orbit coupling of the opposite sign on different layers. (**B**) Expected spin pattern of the antibonding band for two adjacent CuO_2 layers within the unit cell.

distortions typical of cuprates, such as local Jahn-Teller distortions (*32*–*34*), modulations of the oxygens in the BiO slabs, and buckling of the CuO_2 planes (*47*), could break the local inversion symmetry and give rise to a nonzero electric field. The latter effects along with the presence of other atoms in a polar environment within the unit cell could also potentially contribute to the spin texture reported here and could be responsible for the nonzero spin polarization observed in single-layer Bi2201 (*23*, *25*).

Regardless of the origin of the observed spin-orbit interaction, it is clear that its effect on the symmetry of the Hamiltonian and on the ground state properties cannot be neglected. In the case of weak correlations, the interplay between spin-orbit coupling and superconductivity can affect spin susceptibility (*48*), alter the structure of the gap nodes, and allow for additional Ampereanlike attraction channels coming from spin fluctuations (*15*, *49*). In the case of strong correlations, spin-orbit coupling could enhance a charge density wave–type of order (*50*, *51*), as observed in cuprates, and ultimately could affect the superconducting gap and the phase diagram (*52*). Our observation of spin-orbit coupling with a magnitude comparable to that of the interlayer tunneling and superconducting gap [see discussion in (*23*)] and the persistence of a nonzero spin polarization above T_c (fig. S2) suggest that a complex correlation between superconductivity, spin-orbit coupling, and layer degrees of freedom might be at play in cuprates (*52*). As the effects of the coexistence of spin-orbit coupling, strong correlations, and superconductivity are still poorly understood, we hope that our results will stimulate further experimental and theoretical research exploring the physics in this emergent field.

REFERENCES AND NOTES

1. X.-L. Qi, S.-C. Zhang, *Phys. Today* **63**, 33–38 (2010).
2. J. E. Moore, *Nature* **464**, 194–198 (2010).
3. R. E. Walstedt, W. W. Warren Jr., *Science* **248**, 1082–1087 (1990).
4. D. J. Scalapino, *Rev. Mod. Phys.* **84**, 1383–1417 (2012).
5. D. Coffey, T. M. Rice, F. C. Zhang, *Phys. Rev. B* **44**, 10112–10116 (1991).
6. W. Koshibae, Y. Ohta, S. Maekawa, *Phys. Rev. B* **47**, 3391–3400 (1993).
7. V. M. Edelstein, *Phys. Rev. Lett.* **75**, 2004–2007 (1995).
8. C. Wu, J. Zaanen, S.-C. Zhang, *Phys. Rev. Lett.* **95**, 247007 (2005).
9. N. Harrison, B. J. Ramshaw, A. Shekhter, *Sci. Rep.* **5**, 10914 (2015).
10. D. Pesin, L. Balents, *Nat. Phys.* **6**, 376–381 (2010).
11. S. Nakosai, Y. Tanaka, N. Nagaosa, *Phys. Rev. Lett.* **108**, 147003 (2012).
12. W. Witczak-Krempa, G. Chen, Y. B. Kim, L. Balents, *Annu. Rev. Condens. Matter Phys.* **5**, 57–82 (2014).
13. J. G. Rau, E. K.-H. Lee, H.-Y. Kee, *Annu. Rev. Condens. Matter Phys.* **7**, 195–221 (2016).
14. L. P. Gor'kov, E. I. Rashba, *Phys. Rev. Lett.* **87**, 037004 (2001).
15. M. Kargarian, D. K. Efimkin, V. Galitski, *Phys. Rev. Lett.* **117**, 076806 (2016).
16. X. Xi *et al.*, *Nat. Phys.* **12**, 139–143 (2016).
17. L. Fu, C. L. Kane, *Phys. Rev. Lett.* **100**, 096407 (2008).
18. S. Fujimoto, *Phys. Rev. B* **77**, 220501 (2008).
19. J. D. Sau, R. M. Lutchyn, S. Tewari, S. Das Sarma, *Phys. Rev. Lett.* **104**, 040502 (2010).
20. C. Liu *et al.*, *Phys. Rev. B* **90**, 045127 (2014).
21. N. Xu *et al.*, *Nat. Commun.* **5**, 4566 (2014).
22. K. Gotlieb, Z. Hussain, A. Bostwick, A. Lanzara, C. Jozwiak, *Rev. Sci. Instrum.* **84**, 093904 (2013).
23. Materials and methods are available as supplementary materials.
24. C. Jozwiak *et al.*, *Rev. Sci. Instrum.* **81**, 053904 (2010).
25. M. Fanciulli, S. Muff, A. P. Weber, J. H. Dil, *Phys. Rev. B* **95**, 245125 (2017).
26. N. L. Saini *et al.*, *Phys. Rev. Lett.* **79**, 3467–3470 (1997).
27. A. Damascelli, Z. Hussain, Z.-X. Shen, *Rev. Mod. Phys.* **75**, 473–541 (2003).
28. G. Dresselhaus, *Phys. Rev.* **100**, 580–586 (1955).
29. Y. A. Bychkov, E. I. Rashba, *J. Phys. C* **17**, 6039–6045 (1984).
30. X. Zhang, Q. Liu, J.-W. Luo, A. J. Freeman, A. Zunger, *Nat. Phys.* **10**, 387–393 (2014).
31. P. Bordet *et al.*, *Physica C* **156**, 189–192 (1988).
32. V. Petricek, Y. Gao, P. Lee, P. Coppens, *Phys. Rev. B* **42**, 387–392 (1990).
33. X. B. Kan, S. C. Moss, *Acta Crystallogr. B* **48**, 122–134 (1992).
34. A. Bianconi *et al.*, *Phys. Rev. B* **54**, 4310–4314 (1996).
35. B. J. Suh, P. C. Hammel, M. Hücker, B. Büchner, *Phys. Rev. B* **59**, R3952–R3955 (1999).
36. G. M. De Luca *et al.*, *Phys. Rev. B* **82**, 214504 (2010).
37. J. M. Riley *et al.*, *Nat. Phys.* **10**, 835–839 (2014).
38. W. Yao *et al.*, *Nat. Commun.* **8**, 14216 (2017).
39. S.-L. Wu *et al.*, *Nat. Commun.* **8**, 1919 (2017).
40. P. Miles *et al.*, *Physica C* **294**, 275–288 (1998).
41. Z.-H. Zhu *et al.*, *Phys. Rev. Lett.* **110**, 216401 (2013).
42. J. D. Koralek *et al.*, *Phys. Rev. Lett.* **96**, 017005 (2006).
43. T. Yamasaki *et al.*, *Phys. Rev. B* **75**, 140513 (2007).
44. Y. D. Chuang *et al.*, *Phys. Rev. B* **69**, 094515 (2004).
45. Z.-H. Zhu *et al.*, *Phys. Rev. Lett.* **112**, 076802 (2014).
46. A. A. Ivanov *et al.*, *J. Supercond. Nov. Magn.* **31**, 663–670 (2018).
47. M. Opel *et al.*, *Phys. Rev. B* **60**, 9836–9844 (1999).
48. D. Maruyama, M. Sigrist, Y. Yanase, *J. Phys. Soc. Jpn.* **81**, 034702 (2012).
49. P. A. Lee, *Phys. Rev. X* **4**, 031017 (2014).
50. A. Tagliacozzo, E. Tosatti, *J. Phys. C* **12**, L555–L558 (1979).
51. T. Morisaki, H. Wakaura, H. Koizumi, *J. Phys. Soc. Jpn.* **86**, 104710 (2017).
52. C. O. Dias, H. O. Frota, A. Ghosh, *Phys. Status Solidi* **253**, 1824–1829 (2016).

ACKNOWLEDGMENTS

We thank D. H. Lee, C. Varma, E. Altman, and T. L. Miller for fruitful discussion. We thank K. Kurashima for sample preparation. **Funding:** This work was supported primarily by Berkeley Lab's program on Ultrafast Materials Sciences, funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under contract DE-AC02-05CH11231. A.L. acknowledges partial support for this research from the Gordon and Betty Moore Foundation's EPiQS Initiative through grant GBMF4859. The theory component of this paper was supported by the Quantum Materials Program at Lawrence Berkeley National Laboratory, funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under contract DE-AC02-05CH11231. This research used resources of the Advanced Light Source, which is a DOE Office of Science User Facility under contract DE-AC02-05CH11231. M.S. was supported by the Gordon and Betty Moore Foundation's EPiQS Initiative through grant GBMF4307. W.Z. acknowledges support from the Ministry of Science and Technology of China (2016YFA0300501) and from NSF China (11674224). **Author contributions:** K.G. and A.L. were responsible for experimental design. K.G. and C.-Y.L carried out the experiments. Calculations were performed by M.S. and A.V. Samples were prepared by H.E. A.L. was responsible for experiment planning and infrastructure. All authors contributed to the interpretation and writing of the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/1271/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
References (*53*–*58*)
Data S1

13 June 2017; resubmitted 24 February 2018
Accepted 7 November 2018
10.1126/science.aao0980

REPORT

ELECTROCATALYSIS

Ultralow-loading platinum-cobalt fuel cell catalysts derived from imidazolate frameworks

Lina Chong[1], Jianguo Wen[2], Joseph Kubal[2,3], Fatih G. Sen[2], Jianxin Zou[4], Jeffery Greeley[3], Maria Chan[2], Heather Barkholtz[1], Wenjiang Ding[4], Di-Jia Liu[1]*

Achieving high catalytic performance with the lowest possible amount of platinum is critical for fuel cell cost reduction. Here we describe a method of preparing highly active yet stable electrocatalysts containing ultralow-loading platinum content by using cobalt or bimetallic cobalt and zinc zeolitic imidazolate frameworks as precursors. Synergistic catalysis between strained platinum-cobalt core-shell nanoparticles over a platinum-group metal (PGM)–free catalytic substrate led to excellent fuel cell performance under 1 atmosphere of O_2 or air at both high-voltage and high-current domains. Two catalysts achieved oxygen reduction reaction (ORR) mass activities of 1.08 amperes per milligram of platinum (A mg_{Pt}^{-1}) and 1.77 A mg_{Pt}^{-1} and retained 64% and 15% of initial values after 30,000 voltage cycles in a fuel cell. Computational modeling reveals that the interaction between platinum-cobalt nanoparticles and PGM-free sites improves ORR activity and durability.

The oxygen reduction reaction (ORR) is more sluggish in proton-exchange membrane fuel cells (PEMFCs) than hydrogen oxidation and requires three to five times as much platinum (*1*–*3*). The high cost and scarcity of Pt have driven efforts to reduce Pt usage. Recent examples include Pt–transition metal (TM) alloys with distinctive three-dimensional (3D) structures (*4*–*8*). Excellent ORR activity and durability were demonstrated by the rotating disk electrode (RDE) method in oxygen-saturated aqueous electrolyte. Although the RDE approach provides important information about catalytically active sites, it does not fully reflect how the catalysts would perform in operating fuel cell environments of different mass and charge transport limitations (*9*, *10*).

In fuel cells, catalysts in the membrane electrode need to be easily accessible by the reactants, particularly under low fuel cell polarization voltage where a large influx of reactant (O_2) must be converted to produce high current density. For a small number of shaped but large crystallites prepared within ultralow Pt loading limitation, there will not be enough crystallites to spread over the electrode surface to encounter all of the O_2 before they exit the electrode, resulting in a drop in the fuel cell current. The opposite approach, dispersing Pt to the atomic level, can result in fast Pt dissolution and poor catalytic activity (*11*). A third approach is to use a platinum-group metal (PGM)–free catalyst, which could eliminate the Pt usage altogether. Such catalysts, generally prepared from earth-abundant elements such as TMs (mostly Fe and Co) embedded in nitrogen-carbon composites (TM-N_x-C_y), have demonstrated promising ORR activity approaching that of Pt (*12*–*15*).

When prepared from metal-organic frameworks (MOFs) or porous organic polymers as precursors, these catalysts possess densely and uniformly populated active sites throughout the electrode, easily accessible by O_2 fluxes (*16*, *17*). The key drawback, however, is their poor stability under PEMFC operations. Unlike Pt catalysts, of which the activity degradation is mainly caused by crystallite dissolution and agglomeration (*18*), the origin of the PGM-free catalyst deactivation is poorly understood because the nature of the active site is still under debate (*19*, *20*). One possible cause is the oxidative degradation by hydrogen peroxide produced during ORR (*21*).

If the shortcomings of ultralow-loading Pt and PGM-free catalysts were mutually compensated through a synergistic interaction, Pt usage could be substantially reduced while maintaining excellent activity and durability. We report the design and synthesis of synergistic ORR catalysts containing an ultralow concentration of Pt alloy supported over PGM-free materials, denoted as LP@PF. We used a Co-containing and a Co- and Zn-containing zeolitic imidazolate framework (ZIF, a subgroup of MOFs) as the precursors, which we then thermally activated and catalyzed with Pt to form alloy. The resulting catalysts had very high mass activities (MAs) of 8.64 ± 0.25 A mg_{Pt}^{-1} and 12.36 ± 0.53 A mg_{Pt}^{-1} measured by RDE or 1.08 ± 0.17 A mg_{Pt}^{-1} and 1.77 ± 0.39 A mg_{Pt}^{-1} measured in fuel cells at an internal resistance–corrected (iR-free) voltage of 0.9 V. Both values exceed the U.S. Department of Energy (DOE) target of 0.44 A mg_{Pt}^{-1} (*22*). The catalysts showed excellent activity in both high-voltage and high–current density domains and good durability in a 30,000 voltage-cycle accelerated stress test (AST) in a fuel cell.

Our catalyst design is based on the following rationales. Pt-Co alloy represents one of the most active ORR catalysts and is currently used in commercial fuel cell vehicles, whereas Co-ZIF–derived PGM-free catalysts have also shown high specific surface areas, densely distributed active sites, and excellent ORR activities in PEMFCs (*23*). During thermal activation of Co-ZIF, a fraction of Co^{2+} is reduced to metallic nanocrystallites, whereas other Co ions are converted to atomically dispersed Co-N_x-C_y sites situated nearby. The Co nanocrystallites could serve as the seeds to amalgamate with subsequently added Pt to form alloy nanoparticles (NPs) with a core-shell structure. Close proximity between Pt-Co NPs and Co-N_x-C_y sites could promote synergistic catalysis.

We prepared a monometallic cobalt zeolitic methylimidazolate framework, $Co(mIm)_2$ (also called ZIF-67), and a bimetallic ZIF containing zinc zeolitic methylimidazolate framework, $Zn(mIm)_2$ (also called ZIF-8), coated by ZIF-67 (ZIF-8@ZIF-67). Both ZIF-67 and ZIF-8@ZIF-67 were then thermally activated. A subsequent controlled acid wash formed PGM-free catalyst supports PF-1 and PF-2, respectively. These supports were ORR active by themselves and retained a fraction of metallic cobalt nanocrystallites.

A Pt precursor was subsequently applied to PF-1 and PF-2, followed by in situ reduction in oleylamine and high-temperature annealing under ammonia (NH_3) to obtain the final catalysts LP@PF-1 and LP@PF-2. Figure 1A schematically illustrates the characteristics of these catalysts. First, a majority of Pt was converted to Pt-Co NPs that were uniformly dispersed over a substrate of densely populated Co-N_x-C_y sites. Cobalt, however, was found in three different forms. In addition to Pt-Co alloy and Co-N_x-C_y, it also existed as a metal crystallite encapsulated by onion-like graphitic layers [Co@graphene (fig. S1)], which is also often considered catalytically active (*13*). High-angle annular dark-field scanning transmission electron microscopy (HAADF-STEM) images show that Pt-Co NPs are surrounded by an amorphous "particle-free" region, in which individual Co atoms and a trace amount of Pt atoms can be distinguished (Fig. 1, B and C). The energy-dispersive x-ray spectroscopy (EDS) and the electron energy-loss spectroscopy (EELS) analyses identified primarily C, N, and Co^{2+} in these regions (Fig. 1D and table S1). These compositions

[1]Chemical Sciences and Engineering Division, Argonne National Laboratory, Lemont, IL 60439, USA. [2]Center for Nanoscale Materials, Argonne National Laboratory, Lemont, IL 60439, USA. [3]Davidson School of Chemical Engineering, Purdue University, 480 Stadium Mall Drive, West Lafayette, IN 47907, USA. [4]National Engineering Research Center of Light Alloys Net Forming and State Key Laboratory of Metal Matrix Composite, Shanghai Jiao Tong University, Shanghai 200240, China.
*Corresponding author. Email: djliu@anl.gov

represent a typical makeup of PGM-free catalysts (*15*) with good ORR activity (*23*). High-resolution transmission electron microscopy (HRTEM) revealed that the Pt-Co NPs had a Pt-Co core and a Pt shell (Fig. 1E and fig. S2 and S3). Apparent ordering of Co and Pt in Pt-Co core with face-centered cubic crystal structures was also observed along the <100> and <110> directions, further supporting the existing of superstructures known to be highly active in catalysis (*24*).

Lattice contraction led to surface segregation and a highly strained Pt skin of three to four monolayers (fig. S2B), which enhances the ORR activity (*25*). In many cases, the Pt shell was partially covered by multilayered terraces composed of Co, N, and C and identified as CoN or CoC from their interlayer spacing (Fig. 1E and figs. S2 and S4). The terraces could slow down the dissolution of Pt-Co NPs while keeping the active surface exposed. The Pt:Co ratios of the overall catalysts and the Pt-Co alloy NPs were analyzed by EDS (fig. S5A). The Pt:Co ratios of NP were consistent with alloy compositions of 1:1 in LP@PF-1 and 3:1 in LP@PF-2, respectively, which were further confirmed by x-ray diffraction (XRD) (fig. S6 and table S2). This ratio was substantially lower in bulk catalyst after averaging the contributions from Co-N_x-C_y and Co@graphene sites. The NP sizes were narrowly distributed around average diameters of 5.6 ± 1.6 nm and 5.7± 1.7 nm (fig. S7), and the overall Pt loadings were 2.72 weight % (wt %) and 2.81 wt % for LP@PF-1 and LP@PF-2, respectively. The Brunauer-Emmett-Teller specific surface areas were 343 m^2/g for LP@PF-1 and 807 m^2/g for LP@PF-2 (fig. S8).

We also investigated the electronic structures of the Pt-Co alloys and the PGM-free catalyst support using x-ray photoelectron spectroscopy (XPS), x-ray absorption near-edge structure (XANES) spectroscopy, and extended x-ray absorption fine structure (EXAFS) spectroscopy. Electron transfer with Co causes a shift in the Pt d-band center energy in Pt-Co alloys, which weakens OH_{ad} binding on the Pt surface and thus improved ORR catalytic properties (*26*). As expected, the Pt XPS shows a ~0.2 eV positive energy shift in LP@PF-1 upon annealing in NH_3 (Fig. 2A). Co XPS also showed redistribution to more ionic Co^{2+} from Co^0 (Fig. 2B), with the Co^{+2}:Co peak ratio changing from 1.8 to 2.9 after NH_3 treatment, forming additional Co-N_x. The N 1s spectra demonstrated a high content of pyridinic and pyrrolic N embedded in the graphitic matrix with little change after the NH_3 treatment (Fig. 2C).

XANES analysis showed reduction of white line intensity (gray arrow) at the Pt L_3 edge, which corroborates electron transfer from Co 3d to Pt 5d orbitals in Pt-Co alloy (*27*) (Fig. 2D). Alloy formation was further confirmed by a characteristic Pt-Co interaction peak (red arrow) at 11,576 eV (*28*). Similar changes in XPS and XANES were also observed in LP@PF-2 (fig. S9). The transformation from Pt to Pt-Co alloy was further corroborated by EXAFS (fig. S10A) and XRD (fig. S6). XANES at the Co K-edge was more convoluted because it included the contributions from Pt-Co alloy, metallic Co@graphene clusters, and Co^{2+} ion embedded in N-decorated C support. After the NH_3 treatment, the intensity of the pure Co metal peak at 7013 eV (green arrow) reduced substantially, whereas the peak at 7227 eV (red arrow) grew substantially (Fig. 2E), reflecting hybridized Co 4s and 4p orbitals by Pt in the alloy (*28*) and conversion of some Co (0) to Co(II)-N_x as corroborated by XPS. EXAFS analysis revealed the loss of Co-Co peak intensity due to a decrease of metallic Co and an increase of alloy formation (fig. S10B). More importantly, it showed an enhancement of peak intensity at Co–N bond distance, indicating the increase of the N-ligated Co^{2+} population. The atomically dispersed TM ligated by four N atoms in a C matrix has been associated with the active sites for ORR in PGM-free catalysts (*15*, *20*, *29*).

We first measured the electrocatalytic ORR activities of LP@PF-1 and LP@PF-2 by the rotating ring-disk electrode (RRDE) at room temperature in an O_2-saturated 0.1 M $HClO_4$ solution. For comparison, the PGM-free catalytic substrate PF-2, a commercial Pt/C catalyst (TKK, 46.7 wt % Pt), and an in-house prepared 3 wt % Pt_3Co/ZC catalyst were also tested. Pt_3Co/ZC was prepared by adding Pt_3Co alloy NPs over ZIF-8–derived carbon (ZC). This catalyst is similar in composition and surface property to LP@PF-2, except it lacks Co-N_x-C_y sites. Figure 3A displays the linear sweep voltammetry (LSV) from the kinetic to the diffusion-limiting regions. The halfwave potential $E_{1/2}$, a gauge of electrocatalytic activity, increased in the order of PF-2 < Pt_3Co/ZC ≤ commercial Pt/C < LP@PF-1 < LP@PF-2, with LP@PF-2 at 0.96 V (table S3). Meanwhile, the electron-transfer number *n*, calculated from the ring-to-disk current ratios, was 3.99 for both LP@PF-1 and LP@PF-2, suggesting a nearly completed conversion from O_2 to H_2O instead of H_2O_2. The Pt MA Tafel plot derived from LSV demonstrated substantially higher values for LP@PF-1 and LP@PF-2 than those of the reference catalysts (Fig. 3B), whereas the specific current density Tafel plot exhibited higher onset potentials and lower slopes (fig. S11). LP@PF-1 and LP@PF-2 delivered high Pt MAs of 8.64 A mg_{Pt}^{-1} and 12.36 A mg_{Pt}^{-1} at 0.9 V versus reversible hydrogen electrode (RHE), respectively, and outperformed the commercial catalyst (Fig. 3C) and some recently reported nanostructured 3D Pt alloy catalysts (table S4) (*6*, *7*).

We further incorporated the LP@PF catalysts in the cathode of the membrane electrode assembly (MEA) and tested their performances in a PEMFC single cell with O_2 or air as the cathodic gas feed. The cathodic Pt loading were 0.033 mg_{Pt} cm^{-2} for LP@PF-1 and 0.035 mg_{Pt} cm^{-2} for LP@PF-2, respectively. Figure 3D shows their current-voltage (*i-V*) polarizations and power density distributions measured under 1 bar of fully humidified O_2. For benchmarking, we also tested a MEA with Pt_3Co/ZC catalyst with cathodic loading of 0.043 mg_{Pt} cm^{-2}, a MEA with PF-2 cathode catalyst, and commercial MEAs with much higher cathodic Pt loadings (Fig. 3D and fig. S12).

Both MEAs with LP@PF-1 and LP@PF-2 displayed higher catalytic activities than the comparative MEAs in the high-voltage region (>0.7 V) in an H_2-O_2 cell. The MEA with LP@PF-2 cathode

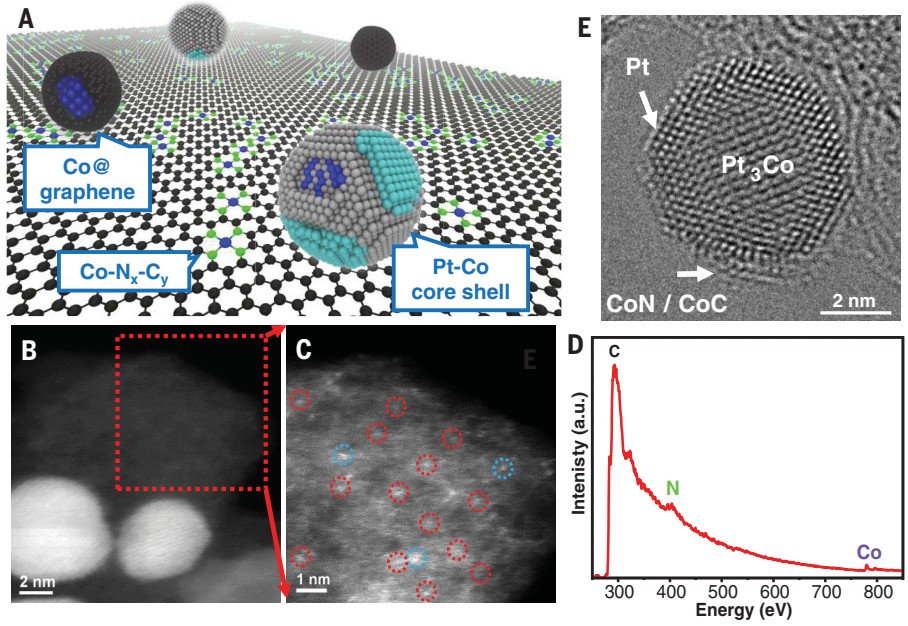


Fig. 1. LP@PF catalyst structure. (**A**) Schematics of LP@PF catalysts, showing coexistence of Pt-Co NPs, Co@graphene, and Co-N_x-C_y PGM-free active sites. (**B**) A HAADF-STEM image of Pt-Co NPs in LP@PF-1 situated over (**C**) PGM-free support containing atomically dispersed Co (circled in red) and trace Pt (circled in blue). (**D**) EELS analysis of the elemental composition of (C). a.u., arbitrary units. (**E**) HRTEM image of a representative Pt-Co alloy NP with Pt_3Co superlattice core and Pt skin partially covered by CoN and CoC terraces.

catalyst demonstrated higher current density than the commercial MEA through the entire polarization scan, even at 1/10th of the cathodic Pt loading. Its current density continued to increase nearly linearly with polarization voltage, a feature commonly observed in PGM-free fuel cells and characteristically different from conventional MEAs with PGM-only catalysts. Figure 3E shows the Pt MA Tafel plots derived from the internal resistance corrected *i-V* polarizations (fig. S13) and Pt loading. Again, the LP@PF-1 and LP@PF-2 MEAs showed higher MAs than those of comparative MEAs. The fuel cell–based Pt MAs measured at 0.9 $V_{iR\text{-}free}$ are 1.08 A mg_{Pt}^{-1} for LP@PF-1 and 1.77 A mg_{Pt}^{-1} for LP@PF-2, respectively, representing an order of magnitude improvement compared with the commercial MEAs (Fig. 3F and table S5). These values exceed the 2025 target set by DOE (0.44 A mg_{Pt}^{-1} at 0.9 $V_{iR\text{-}free}$ for MEA) by factors of approximately two and four and represent record-high ORR activities measured in a PEMFC (*22*).

The MEAs were also subjected to AST under repeated cell voltage sweeps from 0.6 to 1.0 V according to DOE catalyst stability evaluation protocols (*22*). Fuel cell polarizations and MAs were measured periodically after designated voltage cycles up to 30,000 (fig. S14). Figure 3G shows fuel cell *i-V* polarizations and power-density distributions after 30,000 voltage cycles. Although AST caused a substantial activity loss for the commercial MEA, the MEAs with LP@PF-1 and LP@PF-2 cathode catalysts showed improved durability (fig. S15). Especially, the MEA with LP@PF-1 demonstrated the highest durability with its MA retained at 0.672 A mg_{Pt}^{-1} at 0.9 $V_{iR\text{-}free}$ (Fig. 3F), or 64% of its initial value. This value surpassed the catalyst durability goal of <40% MA loss after AST set by DOE (*22*). The MA stability of LP@PF-1 was compared to a state-of-the-art dealloyed PtNi catalyst (*30*) and showed higher values at both beginning and end of life, although the PtNi MEA demonstrated higher retention of MA at the end of AST. The drop of the fuel cell voltage at current density of 0.8 A cm^{-2} after 30,000 cycles was 6 mV, well within the DOE target of <30 mV loss. For MEA with LP@PF-2, the MA was reduced to 0.263 A mg_{Pt}^{-1} at 0.9 $V_{iR\text{-}free}$, which is still comparable to the DOE target of 0.264 A mg_{Pt}^{-1} after AST based on 40% loss of the initial activity of 0.44 A mg_{Pt}^{-1}. The voltage drop at current density of 0.8 A cm^{-2} was 47 mV. In addition to voltage cycling, we also tested the MEA durability at constant voltage and found excellent performances with lower decay rates for LP@PF catalysts under both O_2 and air compared with the benchmarks (figs. S16 to S18).

Excellent MEA performances by LP@PF-1 and LP@PF-2 were also observed when the fuel cells were tested in H_2-air under different stoichiometries (flow rates) and pressures (Fig. 3H and figs. S19 and S20). Both MEAs outperformed commercial MEAs at $V > 0.6$ V, reaching a current density of 300 ± 10 mA cm^{-2} at 0.8 V, meeting the DOE target. LP@PF-1 showed slightly better fuel cell performances compared with LP@PF-2 at higher cell voltage but lower current density at low cell potential. We attribute this mainly to the difference in PGM-free substrate structure. PF-1 has a lower surface area but higher PGM-free active site area density and level of graphitization. In addition to high stability, such structure promotes robust synergistic catalysis. PF-2 has higher porosity and surface area, which facilitates the interaction with airflow and, therefore, higher current density near the mass-transport–limited region (*9*, *23*). For comparisons, the MAs and H_2-air fuel cell performances of LP@PF MEAs along with representative published reports are provided in table S6.

The morphology, composition, and electronic state of the LP@PF catalysts after AST were analyzed. For example, TEM analysis showed only minor changes in NP size distribution, with the average particle dimension remaining the same within one standard deviation, from 5.6 ± 1.6 nm to 5.7 ± 1.6 nm for LP@PF-1 and from 5.7 ± 1.7 nm to 6.0 ± 1.5 nm for LP@PF-2, respectively (figs. S21 and S22). The HRTEM images also confirmed the retention of Pt-Co core-shell structure beneath the Co-N-C terraces, which likely played an important role in preserving Pt-Co NPs during AST. EDS analysis averaged from multiple samplings showed that the Pt:Co ratios within single NPs after AST were also nearly unchanged for PtCo (44:56) in LP@PF-1 and Pt_3Co (74:26) in LP@PF-2, respectively. The Pt:Co ratios in the bulk catalysts, however, increased from 7:93 to 19:81 for LP@PF-1 and 11:89 to 14:86 for LP@PF-2, respectively, presumably owing to the dissolution of a small amount of unalloyed cobalt (fig. S5B). The preservation of Pt-Co alloy structures in both catalysts was further confirmed by XRD of the cathode layers peeled from MEAs after AST (fig. S23). The peeled cathode

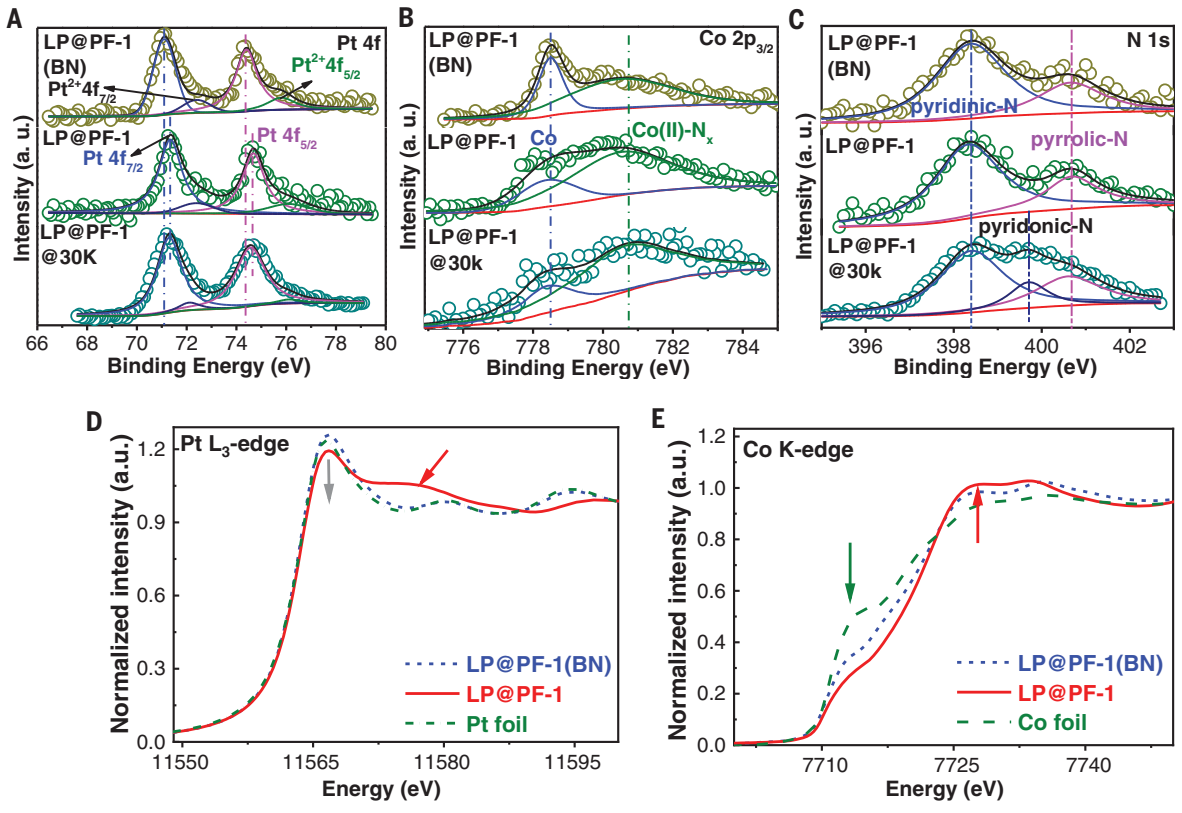


Fig. 2. LP@PF electronic structures. XPS spectra of LP@PF-1 taken at (**A**) Pt 4f, (**B**) Co $2p_{3/2}$, and (**C**) N 1s transitions before (BN) and after NH_3 treatment, as well as after a 30,000 voltage cycle AST in a fuel cell. XANES spectra taken at (**D**) Pt L_3-edge and (**E**) Co K-edge before and after NH_3 treatment with metal foils as references.

layers were also investigated by XPS, which revealed an overall Pt:Co ratio of 21:79 in LP@PF-1 and 20:80 in LP@PF-2, in agreement with the EDS measurements (Fig. 2 and fig. S9). Compared to the fresh catalyst, the Pt electronic state in the aged catalyst remained nearly unchanged in the form of alloy. The Co^{+2}:Co ratio also remained unchanged in LP@PF-1 at 2.9 after AST. The most substantial change came from the carbonaceous nitrogens. The pyridinic-N to pyrrolic-N ratio was reduced from 2.7 to 2.2, possibly due to partial conversion of pyridinic- to pyridonic-N shown by the new peak in Fig. 2C. The pyridonic-N was formed by attachment of OH to the carbon atom next to pyridinic-N, which was previously observed in a PGM-free catalyst after ORR (*19*).

To quantify the synergistic interaction between Pt-Co NPs and PGM-free active sites, we compared the specific current density of a fuel cell containing LP@PF-2 to that from Pt_3Co/ZC and PF-2. Figure 3I shows that the specific current density of LP@PF-2 at any given voltage was about twice of the sum of the contributions from Pt_3Co/ZC and PF-2. This indicates that the synergistic ORR rate in LP@PF is substantially higher than the simple sum of that from Pt_3Co NPs and PGM-free sites. The synergistic catalysis also exhibited improved catalyst stability of LP@PF versus the commercial Pt/C, Pt_3Co/ZC, and PGM-free catalysts (Fig. 3G and fig. S14) (*23*). Such effects were only observed when the Pt-Co alloy NPs were annealed by NH_3 over the PGM-free substrate. Because CoN and CoC adlayers were formed during the in situ reduction in NH_3, we speculate that they not only protect Pt-Co NPs but also serve as "bridges" in transferring the reaction intermediate H_2O_2 from PGM-free site to the Pt-Co NPs through a reverse spillover during synergistic catalysis.

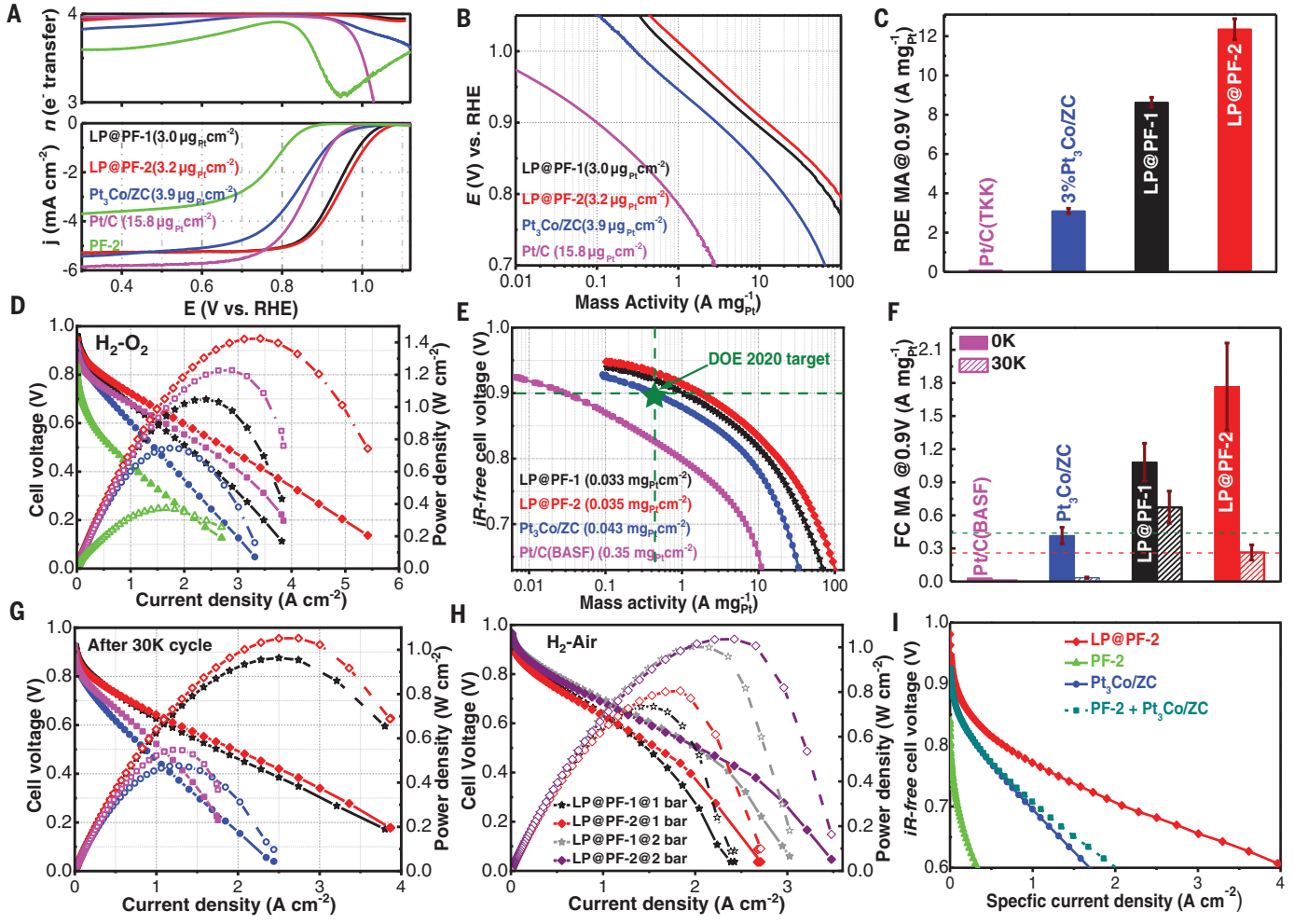


Fig. 3. Electrocatalytic activity and durability evaluations. (**A**) (Bottom) LSVs of different catalysts recorded at a rate of 10 mV s^{-1} and 1600 rotations per minute in O_2-saturated 0.1 M $HClO_4$. j, current density. (Top) Number of transferred electrons (e^-), *n*, at different potentials (E). (**B**) MA Tafel plots derived from (A). (**C**) Comparison of MAs at 0.9 V versus RHE from (B). (**D**) H_2-O_2 fuel cell *i*-*V* polarization (solid symbols and lines) and power density (hollow symbols and dashed lines) plots recorded under 1 bar of O_2 pressure with cathode Pt loading of 0.033 mg_{Pt} cm^{-2} for LP@PF-1, 0.035 mg_{Pt} cm^{-2} for LP@PF-2, 0.043 mg_{Pt} cm^{-2} for Pt_3Co/ZC, and 0.35 mg_{Pt} cm^{-2} for commercial MEA. LP@PF-1, black stars; LP@PF-2, red diamonds; PF-2, green triangles; 3% Pt_3Co/ZC, blue spheres; commercial 47% Pt/C from TKK or MEA from BASF, magenta squares. (**E**) Cathodic MA Tafel plots derived from fuel cell measurement. The green star denotes the U.S. DOE 2025 target. (**F**) Fuel cell (FC) MAs at 0.9 $V_{iR\text{-}free}$ before (solid) and after (hatched) 30,000 voltage cycles, showing that LP@PF catalysts meet or exceed DOE's 2025 MA targets for before (green dashed line, 0.44 A mg_{Pt}^{-1}) and after (red dashed line, 0.264 A mg_{Pt}^{-1} or 40% of the initial value) AST. (**G**) H_2-O_2 fuel cell *i*-*V* polarizations and power densities after 30 K voltage cycles. (**H**) H_2-air fuel cell performances for the same MEAs containing LP@PF-1 and LP@PF-2 under 1 or 2 bars of pressure. (**I**) Specific current densities of PF-2, Pt_3Co/ZC, LP@PF-2, and the sum of PF-2 and Pt_3Co/ZC as a function of iR-free fuel cell voltage measured under 1 bar of H_2-O_2 pressure. For all fuel cell tests, membrane = Nafion 211, temperature = 80°C, and anode loading = 0.35 mg_{Pt} cm^{-2}. For H_2-O_2 cell $P_{H2} = P_{O2}$ = 100 kPa at 100% relative humidity (RH) (back pressure = 50 kPa, absolute pressure = 150 kPa), flow rate = 200 ml min^{-1}. For H_2-air cell $P_{H2} = P_{air}$ = 100 kPa or 200 kPa at 100% RH, H_2 flow rate = 200 ml min^{-1} and airflow rate = 520 ml min^{-1} (equivalent of stoichiometries of 1.5/1.8 at 3.5 A cm^{-2} of the end of polarization).

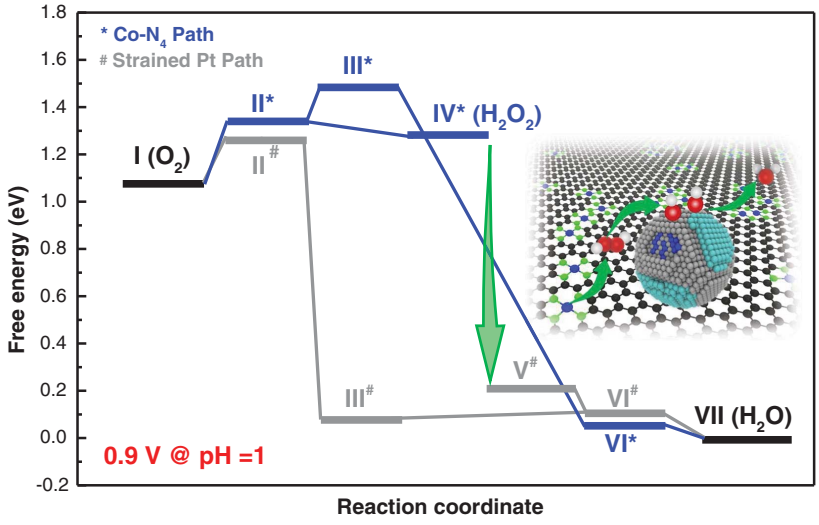


Fig. 4. Free-energy diagram of the ORR pathways. The proposed associative reaction coordinates represent the following states: (I) * or # + O_2 + $4H^+$ + $4e^-$, (II) OOH* or $OOH^{\#}$ + $3H^+$ + $3e^-$, (III) O* or $O^{\#}$ + H_2O + $2H^+$ + $2e^-$, (IV) * + H_2O_2 + $2H^+$ + $2e^-$, (V) $2OH^{\#}$ + $2H^+$ + $2e^-$, (VI) OH* or $OH^{\#}$ + H_2O + H^+ + e^-, and (VII) * or # + $2H_2O$, where * (blue) denotes the binding site on Co-N_4 embedded in graphene and # (gray) denotes the binding site on a strained Pt (111) facet. (Inset) Schematics of H_2O_2 generated over Co-N_4 migrating to the strained Pt (111) surface (green arrows), followed by dissociation to $OH^{\#}$ and water formation. (Computation was performed at 0.9 V relative to the hydrogen electrode at pH = 1.)

To better understand the improved durability of the LP@PF catalyst, we performed density function theory (DFT) calculations for the interface between a Pt_3Co NP [represented by strained Pt (111)] and Co-N_4 decorated graphene. The calculations determined that the strong interaction of the Pt surface with Co-N_4-C sites enhances binding that helps to impede the segregation of the Pt-Co NPs from the support (fig. S24). The simulation also revealed that two or three CoN and CoC adlayers grow preferentially on Pt (100) instead of (111) facets, with formation energies that are more stable than that of the bulk CoN (fig. S25 and table S8). The presence of these adlayers optimizes the exposure of the catalytically more active (111) facet yet reduces Pt dissolution through less stable (100) facets. This result may explain why most alloy particles remain intact after AST. Strong binding of Pt-Co NPs with PGM-free site–mediated surface also generates intimate contact between the two with better charge and reaction intermediate transfers, which are further facilitated through improved hydrophilicity by the adlayer over the Pt surface.

DFT calculations were also carried out to understand the enhanced activity of LP@PF catalysts. We calculated the thermodynamic barriers along two parallel ORR reaction pathways, one over Pt (111) (with 4% strain, near the calculated value for PtCo with a multilayer Pt skin) and another over a Co-N_4 active site, through multistep sequential combination of protons and electrons (Fig. 4 and fig. S26). Pt (111) has lower but non-negligible barriers for the reaction steps, including the stabilization of $OOH^{\#}$ in step II# and the formation of $OH^{\#}$ in step VI#, respectively. Over a Co-N_4 site, the kinetic barrier for OOH* formation (step II*) is only <0.1 eV higher than that over the Pt site and is highly facile. The reactions after step II* branch into two concurrent paths, formation of O* and water (step III*) and production of H_2O_2 (step IV*), with the reaction barrier of the former being less than 0.2 eV. Because H_2O_2 does not bind to the PGM-free site, it can be released after step IV* and migrate to strained Pt (111) sites in the vicinity, as denoted by the green arrow in Fig. 4. The two pathways over Pt and PGM-free catalytic sites intersect, and the subsequent decomposition of H_2O_2 over the strained Pt (111) surface is rapid, as it has no thermodynamic barrier. More details on DFT calculations and mechanistic discussion are provided in the supplementary materials.

This analysis provides an explanation for our experimentally observed synergistic catalysis over LP@PF with improved activity and durability at both high-voltage (kinetics-limited) and high-current (mass transport–limited) regions of fuel cell polarization. The Pt-Co alloy increases its utilization efficiency by not only performing direct ORR but also facilitating reduction of H_2O_2 generated from nearby PGM-free sites, leading to the nearly four-electron transfer measured by RRDE and improved catalyst activity observed in the fuel cell test. Because H_2O_2 is known to corrode TM-based PGM-free sites and porous carbon substrate, its breakdown also helps to preserve the catalyst integrity and durability. Our LP@PF catalysts exhibit improved catalytic activity and durability with lower Pt usage in fuel cells. Remaining challenges include further reducing Pt loading while maintaining synergistic interaction at different fuel cell voltages and catalytic turnover frequencies, better humidity management to ensure effective proton and peroxide transfers over the catalyst surface, as well as improved operation with air.

REFERENCES AND NOTES

1. H. A. Gasteiger, N. M. Marković, *Science* **324**, 48–49 (2009).
2. J. Greeley *et al.*, *Nat. Chem.* **1**, 552–556 (2009).
3. H. A. Gasteiger, S. S. Kocha, B. Sompalli, F. T. Wagner, *Appl. Catal. B* **56**, 9–35 (2005).
4. K. Sasaki *et al.*, *Angew. Chem.* **49**, 8602–8607 (2010).
5. M. Li *et al.*, *Science* **354**, 1414–1419 (2016).
6. C. Chen *et al.*, *Science* **343**, 1339–1343 (2014).
7. X. Huang *et al.*, *Science* **348**, 1230–1234 (2015).
8. B. Lim *et al.*, *Science* **324**, 1302–1305 (2009).
9. A. Kongkanand, M. F. Mathias, *J. Phys. Chem. Lett.* **7**, 1127–1137 (2016).
10. I. E. L. Stephens, J. Rossmeisl, I. Chorkendorff, *Science* **354**, 1378–1379 (2016).
11. R. Makharia *et al.*, *ECS Trans.* **1**, 3–18 (2006).
12. E. Proietti *et al.*, *Nat. Commun.* **2**, 416 (2011).
13. G. Wu, K. L. More, C. M. Johnston, P. Zelenay, *Science* **332**, 443–447 (2011).
14. J. Shui, C. Chen, L. Grabstanowicz, D. Zhao, D. J. Liu, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10629–10634 (2015).
15. H. T. Chung *et al.*, *Science* **357**, 479–484 (2017).
16. S. Yuan *et al.*, *Angew. Chem. Int. Ed.* **52**, 8349–8353 (2013).
17. D. Zhao *et al.*, *Adv. Mater.* **26**, 1093–1097 (2014).
18. R. Borup *et al.*, *Chem. Rev.* **107**, 3904–3951 (2007).
19. D. Guo *et al.*, *Science* **351**, 361–365 (2016).
20. A. Zitolo *et al.*, *Nat. Mater.* **14**, 937–942 (2015).
21. V. Prabhakaran, C. G. Arges, V. Ramani, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1029–1034 (2012).
22. U.S. DOE Fuel Cell Technologies Office, "Fuel Cells Multi-Year Research Development and Demonstration Plan" (2016); www.energy.gov/sites/prod/files/2016/06/f32/fcto_myrdd_fuel_cells_0.pdf
23. L. Chong *et al.*, *ChemElectroChem* **3**, 1541–1545 (2016).
24. D. Wang *et al.*, *Nat. Mater.* **12**, 81–87 (2013).
25. Y. Xu, A. V. Ruban, M. Mavrikakis, *J. Am. Chem. Soc.* **126**, 4717–4725 (2004).
26. B. Y. Xia *et al.*, *Angew. Chem.* **54**, 3797–3801 (2015).
27. N. Jung *et al.*, *NPG Asia Mater.* **8**, e237 (2016).
28. E. K. Hlil, R. BaudoingSavois, B. Moraweck, A. J. Renouprez, *J. Phys. Chem.* **100**, 3102–3107 (1996).
29. S. Ma, G. A. Goenaga, A. V. Call, D.-J. Liu, *Chemistry* **17**, 2063–2067 (2011).
30. B. Han *et al.*, *Energy Environ. Sci.* **8**, 258–266 (2015).

ACKNOWLEDGMENTS

We thank M. S. Ferrandon and D. J. Myers for experimental assistance. **Funding:** This work was supported by U.S. DOE Fuel Cell Technologies Office through Office of Energy Efficiency and Renewable Energy. The works performed at Argonne National Laboratory's Center for Nanoscale Materials and Advanced Photo Source, U.S. DOE Office of Science User Facilities, are supported by Office of Science, U.S. DOE under contract DE-AC02-06CH11357. L.C. wishes to thank Argonne National Laboratory for the Maria Goeppert Mayer Fellowship. J.K. and J.G. acknowledge funding through the U.S. DOE, Office of Basic Energy Sciences, Chemical, Biological, and Geosciences Division under DE-SC0010379. J.K. wishes to thank U.S. DOE, Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education for the DOE under contract number DE-SC0014664. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. DOE Office of Science User Facility supported under contract no. DE-AC02-05CH11231. J.Z. and W.D. wish to thank the National Key Research and Development Program of China (2016YFB0701200) and Chinese National Nature Science Foundation (51771112) for support. The views and opinions

of the authors expressed herein do not necessarily state or reflect those of the U.S. government or any agency thereof. Neither the U.S. government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. **Authors contributions:** D.-J.L. and L.C. conceived of the idea and designed the experiments. L.C., J.Z., H.B., W.D., and D.-J.L. participated in the experiments and data analysis. J.W. acquired high-resolution images and performed analysis. J.K., F.G.S., M.C., and J.G. performed computational modeling. L.C., J.K., J.G., M.C., and D.-J.L. prepared the manuscript. **Competing interests:** The authors declare no competing financial interests. Argonne National Laboratory received U.S. Patent 9,825,308 related to this work. **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/362/6420/1276/suppl/DC1
Materials and Methods
Figs. S1 to S26
Tables S1 to S8
References (*31*–*50*)

10 May 2018; accepted 26 October 2018
Published online 8 November 2018
10.1126/science.aau0630

3D PRINTING

3D nanofabrication by volumetric deposition and controlled shrinkage of patterned scaffolds

Daniel Oran[1]*, Samuel G. Rodriques[1,2]*, Ruixuan Gao[1], Shoh Asano[1,3], Mark A. Skylar-Scott[4,5], Fei Chen[1,6], Paul W. Tillberg[1,7]†, Adam H. Marblestone[1]‡, Edward S. Boyden[1,6,8,9,10]‡§

Lithographic nanofabrication is often limited to successive fabrication of two-dimensional (2D) layers. We present a strategy for the direct assembly of 3D nanomaterials consisting of metals, semiconductors, and biomolecules arranged in virtually any 3D geometry. We used hydrogels as scaffolds for volumetric deposition of materials at defined points in space. We then optically patterned these scaffolds in three dimensions, attached one or more functional materials, and then shrank and dehydrated them in a controlled way to achieve nanoscale feature sizes in a solid substrate. We demonstrate that our process, Implosion Fabrication (ImpFab), can directly write highly conductive, 3D silver nanostructures within an acrylic scaffold via volumetric silver deposition. Using ImpFab, we achieve resolutions in the tens of nanometers and complex, non–self-supporting 3D geometries of interest for optical metamaterials.

Most nanofabrication techniques currently rely on two-dimensional (2D) and 2.5D patterning strategies. Although popular direct laser writing methods allow for the single-step fabrication of self-supporting, polymeric 3D nanostructures (*1*–*8*), arbitrary 3D nanostructures (e.g., solid spheres of metal or metallic wires arranged in discontinuous patterns) are not possible (*9*, *10*). This raises the question of whether a versatile 3D nanofabrication strategy can be developed that would allow independent control over the geometry, feature size, and chemical composition of the final material.

A hallmark of 2D nanofabrication strategies is that materials are deposited in a planar fashion onto a patterned surface. By analogy, we reasoned that a general 3D nanofabrication strategy could involve deposition of materials in a volumetric fashion into a patterned scaffold. However, such scaffolds face a fundamental tension: They should be porous and solvated, to allow for introduction of reagents to their interior, while also being dense, to allow material placement with nanoscale precision. To resolve this contradiction, we reasoned that an ideal scaffold could be patterned in a solvated state and then collapsed and desiccated in a controlled way, densifying the patterned materials to obtain nanoscale feature sizes. Although several groups have experimented with shrinking materials, the shrinking process typically requires harsh conditions and chemical changes that may destroy functional materials (*11*–*13*). We use polyacrylate/polyacrylamide hydrogels for the scaffold material, as they have pore sizes in the range of 10 to 100 nm (*14*), they are known for their ability to expand and shrink up to ~10-fold in linear dimension (*15*–*18*), and methods for optically patterning hydrogels are well established (*19*–*23*).

Our implementation took place in three phases (*24*). It was previously found that two-photon excitation of fluorescein within acrylate hydrogels causes the fluorescein to react with the hydrogel (*21*–*23*). We took advantage of this phenomenon to attach fluorescein molecules carrying reactive groups to the expanded gel in defined 3D patterns (Fig. 1, A and B). In the second phase, after removal of the fluorescein patterning solution, the gel was functionalized by depositing materials onto the patterned reactive groups (Fig. 1, C and D) by using one of several available conjugation chemistries. This volumetric deposition step defines the composition of the material and may be followed by additional deposition chemistries ("intensification") to boost the functionality of the deposited molecules or nanomaterials (Fig. 1, E and F). Importantly, the functional molecules or nanoparticles are not present during the patterning process, so the specific physical properties of the molecules or nanoparticles used will not affect the patterning. In the final phase, the patterned and functionalized gel scaffold was shrunken by a factor of 10 to 20 in each dimension by using acid or divalent cations over a period of hours, and then it was dehydrated to achieve the desired nanoscale resolution (Fig. 1, G and H). The scaffold was not removed, as it supports the nanofabricated material and allows for the creation of disconnected or high–aspect-ratio structures that would otherwise collapse outside of the scaffold.

We found the polyacrylate gel to be a suitable substrate for patterning and deposition. The gel readily accommodated a wide variety of hydrophilic reagents, including small molecules, biomolecules, semiconductor nanoparticles, and metal nanoparticles (fig. S1, A to C). For laser powers below a critical threshold, the density of the deposited functional material was controllable (Fig. 1I and fig. S2). We estimated, based on the maximum pattern fluorescence (fig. S2A), that binding sites are patterned into the gel at concentrations of at least 79.2 μM in the expanded state, leading to a final concentration in the shrunken state of greater than 272.0 mM, or 1.64×10^{20} sites/cm^3 for a 10× gel (see below). By repeating our patterning and deposition process, we were able to deposit multiple materials in different patterns in the same substrate, such as gold nanoparticles and cadmium telluride nanoparticles (Fig. 1J). We observed by using fluorescence that the deposition of the second material onto the first pattern was at most 18.5% of the deposition of the second material onto the second pattern, confirming that multiple materials may be independently patterned and deposited using this process (fig. S3).

The shrinking process is performed either by exposing the expanded gel to hydrochloric acid or to divalent cations (e.g., magnesium chloride) (fig. S1, A to C). The latter may be useful if the patterned materials are sensitive to acid, although we found that both streptavidin and DNA remained functionally intact after acid shrinking (fig. S1D). Gels that are shrunken in hydrochloric acid can subsequently be dehydrated, resulting in additional shrinking, and this process preserved the patterned geometry (Fig. 1K). The final dehydrated gel was transparent (fig. S4A), and atomic force microscopy (AFM) characterization measured the surface roughness over a 1- by 1-μm window to be ~0.19 nm (root mean square) (fig. S4B). Except where stated otherwise, all samples described as "shrunken" here were shrunken and dehydrated. We tested two different gel formulations that differed only in cross-linker concentration: 10× (0.075% cross-linker) and 20× (0.0172% cross-linker) (*24*). The 10× gels, and the patterns within, shrank consistently by a linear factor of 10.6 ± 0.8 in the lateral dimension (mean ± SD, $n = 5$ gels) and 34.8 ± 1.8 in the axial dimension ($n = 6$ gels) (Fig. 1L), with the disproportionate axial shrink occurring during dehydration, possibly due to surface interactions between the shrinking polymer and the surface of the glass container. For the 20× gels, we observed 20.1 ± 2.9-fold shrink

[1]MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [2]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Pfizer Internal Medicine Research Unit, Cambridge, MA 02139, USA. [4]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. [5]Wyss Institute for Biologically Inspired Engineering, Cambridge, MA 02138, USA. [6]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [7]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [8]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [9]McGovern Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [10]Koch Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*These authors contributed equally to this work.
†Present address: Janelia Research Campus, Ashburn, VA 20147, USA. ‡These authors contributed equally to this work.
§Corresponding author. Email: esb@media.mit.edu

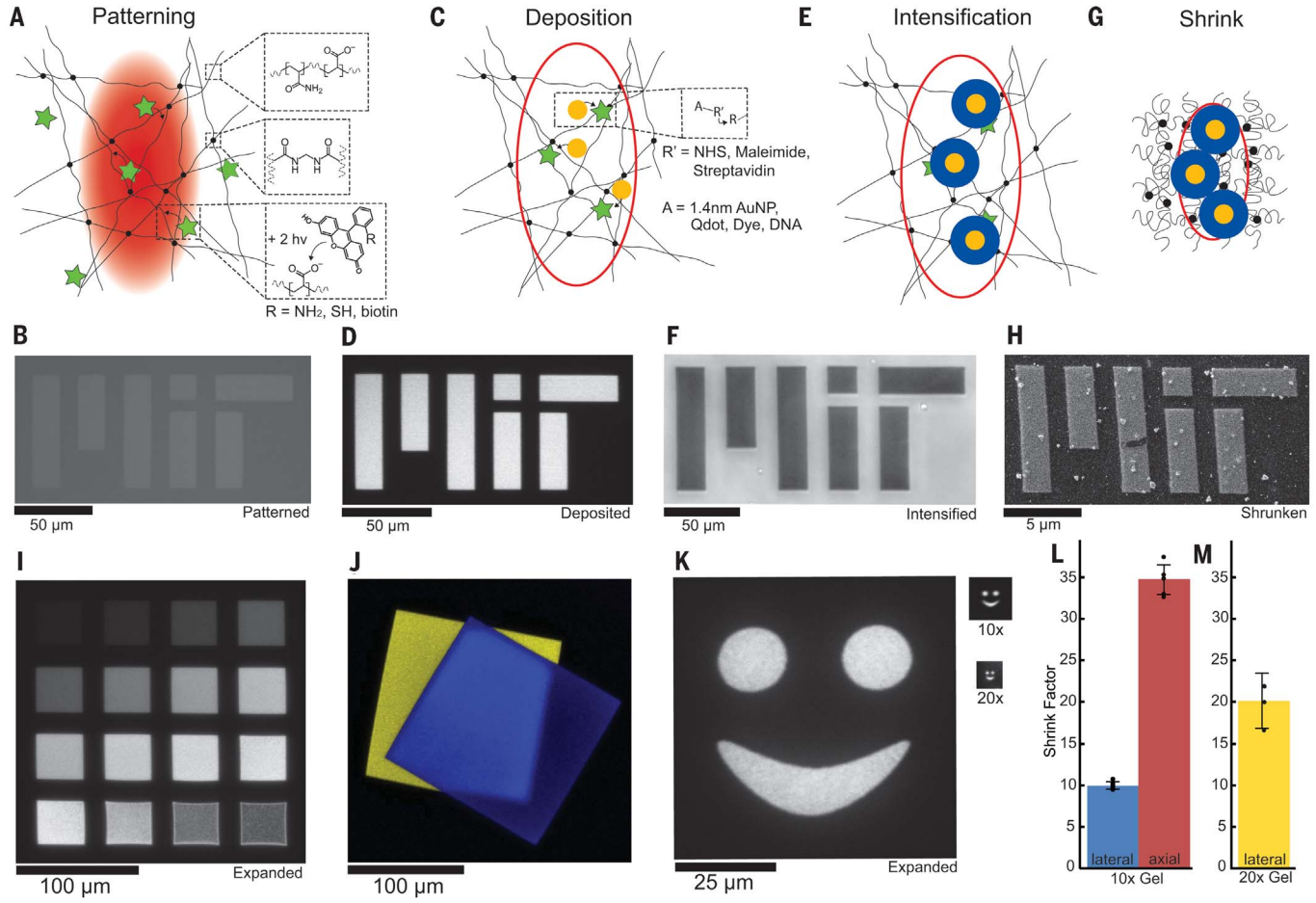


Fig. 1. The ImpFab process. (**A**) Schematic of the patterning process, showing the expanded polyelectrolyte gel (black lines and dots, top insets) and fluorescein (green star, bottom inset) binding covalently to the polymer matrix upon multiphoton excitation (red volume). Not to scale. Fluorescein bears a reactive group, R. h, Planck's constant; ν, frequency. (**B**) Residual fluorescence of patterned fluorescein immediately following patterning. (**C**) Schematic of functionalization of patterned gel by attaching small molecules, proteins, DNA, or nanoparticles to reactive R groups from (A). Red outline indicates patterned volume in (A). (**D**) Image of fluorescent streptavidin nanoparticle conjugates attached to the pattern in (B). (**E**) Schematic of the volumetric deposition process, showing growth of silver (blue) on top of gold nanoparticles within the hydrogel matrix. (**F**) Image of silver deposited onto the pattern in (D) by transmission optical microscopy. Following silver growth, the pattern has high optical density. (**G**) Schematic of the shrinking and dehydration process. (**H**) SEM image of the silverized pattern from (F) following shrinking and dehydration. (**I**) Fluorescent patterns created with different laser powers (*24*). (**J**) Image of a gel patterned with both metal nanoparticles (yellow) and CdTe quantum dots (blue) in different locations. (**K**) Images of fluorescent patterns before shrinking (left, 10× gel), after shrinking and dehydration in a 10× gel (top right) and after shrinking and dehydration in a 20× gel (bottom right). (**L**) The mean lateral (blue) and axial (red) shrink factors (initial size/final size) for 10× gels (n = 6), including dehydration. (**M**) The mean lateral shrink factor for 20× gels (yellow; n = 3). Error bars show SD.

in the lateral dimension (n = 4 gels) (Fig. 1M). The 20× gel formulation is challenging to handle manually due to its delicacy, and so the axial shrink factor was not measured and they were not used further, except for distortion measurements.

To validate the minimum feature size of Implosion Fabrication (ImpFab), we designed a test pattern containing pairs of single-voxel–wide lines (Fig. 2, A to D). Because such postshrink features are necessarily below the optical diffraction limit, we deposited gold nanoparticles and employed scanning electron microscopy (SEM) to assess the resolution after shrinking. We estimated the resolution by measuring the line width [full width at half maximum (FWHM)] (Fig. 2, E to G) and obtained a value of 59.6 ± 3.8 nm (mean ± SD across samples, n = 5) (Fig. 2H) for 10× gels. The mean within-sample standard deviation of the line width was 8.3 nm. We estimated the isotropy of the shrinking process by calculating the ratio of the longest diameter of the patterned circle to the orthogonal diameter (Fig. 2, C and D). The percent distortion thus calculated was 6.8 ± 6.9% for 10× gels (mean ± SD, n = 6 gels) and 8.2 ± 4.3% for 20× gels (n = 4 gels). We found that the ratio of axial to lateral shrink was on average within 3.1 ± 2.5% of the mean of this ratio (n = 6 10× gels), indicating that the disproportionate axial shrink is highly reproducible. Thus, it is possible to account for the disproportionate axial shrink in the design of the pattern. To illustrate this point with the fabrication of a cube, we patterned a rectangular prism and imaged it before and after dehydration (fig. S5). As expected, the rectangular prism contracted in the axial dimension during the dehydration step and turned into a cube.

Because nanoscale metal structures are broadly important in fields such as nanophotonics, metamaterials, and plasmonics, we applied ImpFab to create conductive silver structures. We anchored gold nanoparticles to patterned amines via a biotin-streptavidin linkage (*24*). We were initially unable to deposit gold nanoparticles at high enough concentrations to form conductive structures. We thus developed an intensification process based on photographic intensification chemistries, in which silver was deposited onto the surface of gel-anchored gold nanoparticles in aqueous phase while the gel was in the expanded

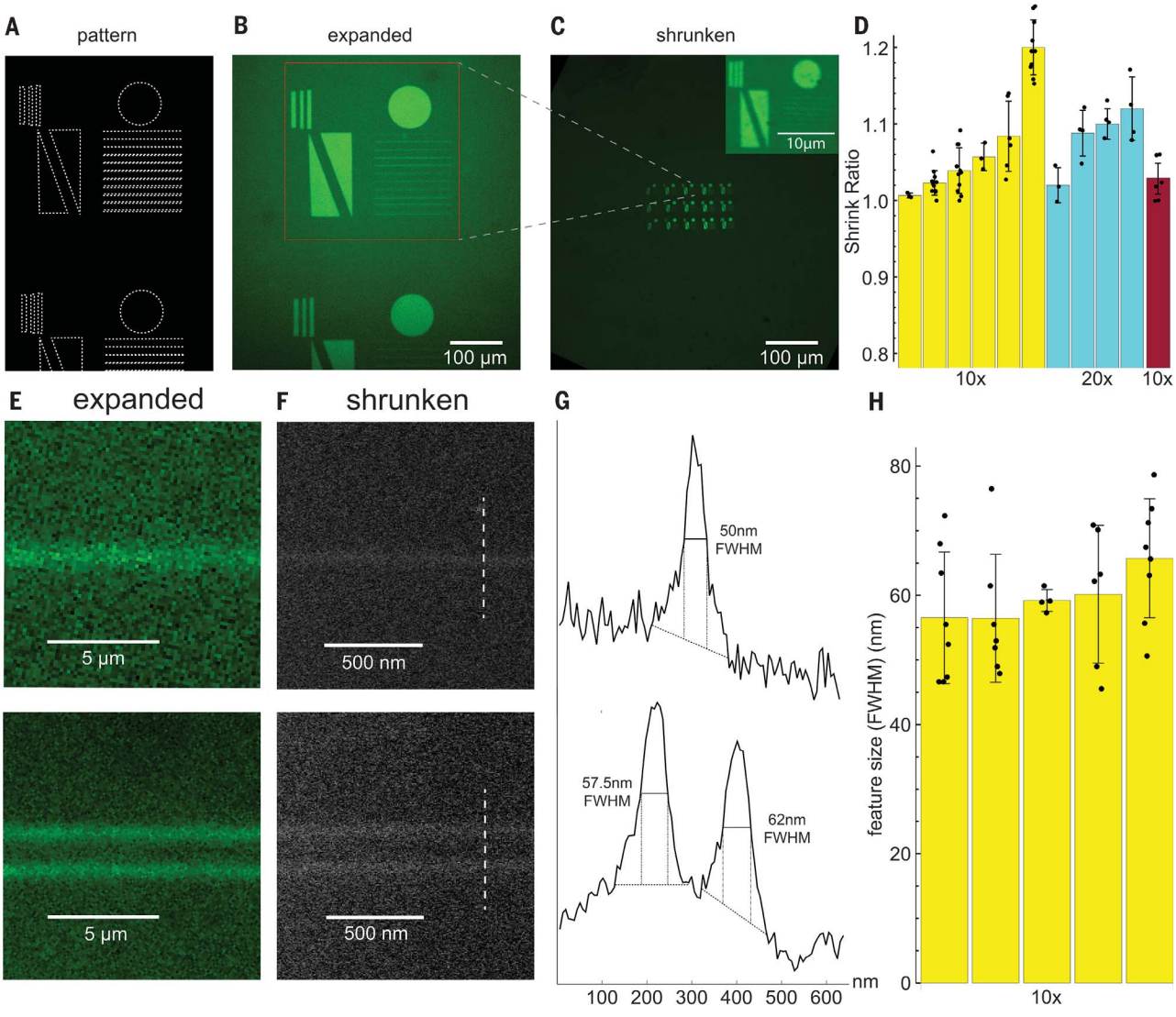


Fig. 2. Resolution of implosion fabrication. (**A**) Design of the resolution test pattern, including pairs of single-voxel-thick lines (bottom right). (**B**) Fluorescence image of the patterns from (A). (**C**) Fluorescence image of the pattern from (B) after shrinking. (**D**) Measures of isotropy in lateral and axial dimensions. Yellow and blue bars represent lateral isotropy for 10× gels and 20× gels, respectively, and the red bar represents axial isotropy for 10× gels. (**E**) Fluorescence images of single-voxel lines before shrinking. (**F**) SEM images of single-voxel lines after 10× shrinking. The gel was functionalized with gold nanoparticles for contrast. (**G**) Cross-sectional intensity profiles of the lines imaged by SEM [dashed lines in (F)], showing how the FWHM of single voxel lines were measured. (**H**) Line widths, measured in (G), for five different gel samples. Dots are measurements for individual lines; bars indicate means ± SD, across individual lines within a single gel.

state (Fig. 1, E and F). Finally, the gel was treated with a chelating agent to remove any remaining dissolved silver and was then shrunken via exposure to hydrochloric acid and subsequent dehydration.

Even with the silver intensification process, wire structures fabricated using the method above (Fig. 3A) were not reliably conductive, or they had resistances on the order of hundreds of ohms. We tested several different methods of sintering, including treatment with oxygen plasma, electrical discharge, and heating the sample to ~500°C in an oven. However, none of these methods resulted in well-preserved and evenly sintered silver structures. Instead, we found that the silver patterns could be sintered effectively when we used the same two-photon setup used for the initial photopatterning step. We found that samples irradiated at relatively low power levels (*24*) showed a distinct change in the morphology of the embedded silver nanoparticles that was consistent with sintering (Fig. 3, B and C). We measured the conductivity of three patterned silver squares both before and after sintering and found that the resistance of each square decreased by 20- to 200-fold (Fig. 3D). Sintered wires were measured in a four-point probe system and were found to have linear IV curves (fig. S6A). Wires sintered in this way had an average resistance of 2.85 ± 1.68 ohms (mean ± SD, n = 10), with the resistance depending on the density of the patterned silver (fig. S6B). By contrast, an ideal silver wire with the same geometry would have a resistance of 0.38 ohms, suggesting that our sintered structures achieved a mean conductivity 13.3% that of bulk silver, with individual samples obtaining conductivities as high as 30% that of bulk silver (Fig. 3E).

To verify that our method is compatible with a wide range of 3D geometries, we fabricated structures with dimensions ranging from hundreds of nanometers to several micrometers (Fig. 4, A to C). We found that these structures retained their morphology following sintering (Fig. 4B). We fabricated a nonlayered, nonconnected 3D structure comprised of many 2D substructures arranged at different angles relative to each other

in space, which would not lend itself to fabrication by other means (Fig. 4D). Whereas our previous experiments had only observed the fabrication of 2D silver structures, we used confocal reflection microscopy to confirm that silver was deposited throughout the volume of the 3D pattern (Fig. 4E). Finally, using confocal microscopy, we were able to validate that the structure retained its shape after shrinking (Fig. 4F).

Due to the modular nature of ImpFab, the extension of the ImpFab strategy to other kinds of materials, such as other semiconductors or metals, only requires the development of an aqueous deposition chemistry that is compatible

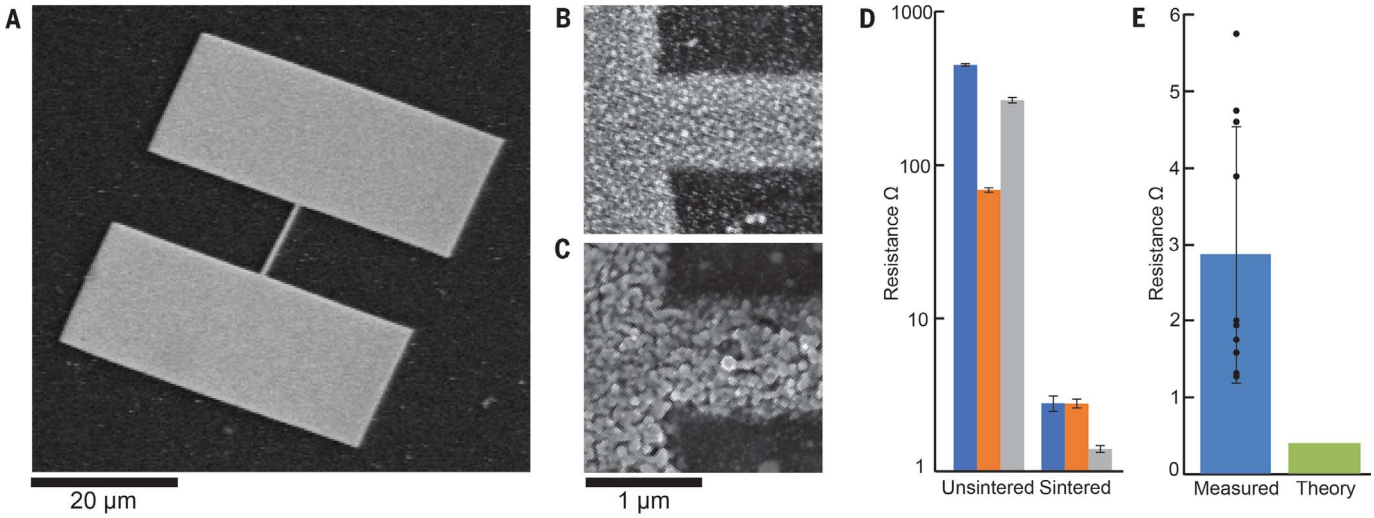


Fig. 3. Characterization of silver conductivity. (**A**) SEM overview of a shrunken silver wire between two landing pads, prior to sintering. (**B** and **C**) SEM images of wires before (B) and after (C) sintering. (**D**) Resistance of three separate conductive pads, each with dimensions of 35 μm by 35 μm, measured before and after sintering. Each color represents a single conductive pad. Error bars show standard errors in a four-point conductivity measurement. (**E**) Resistance of individual sintered wires (black dots) and the means (blue) and standard deviations, compared to the theoretical conductivity of a similar structure made of bulk silver (green).

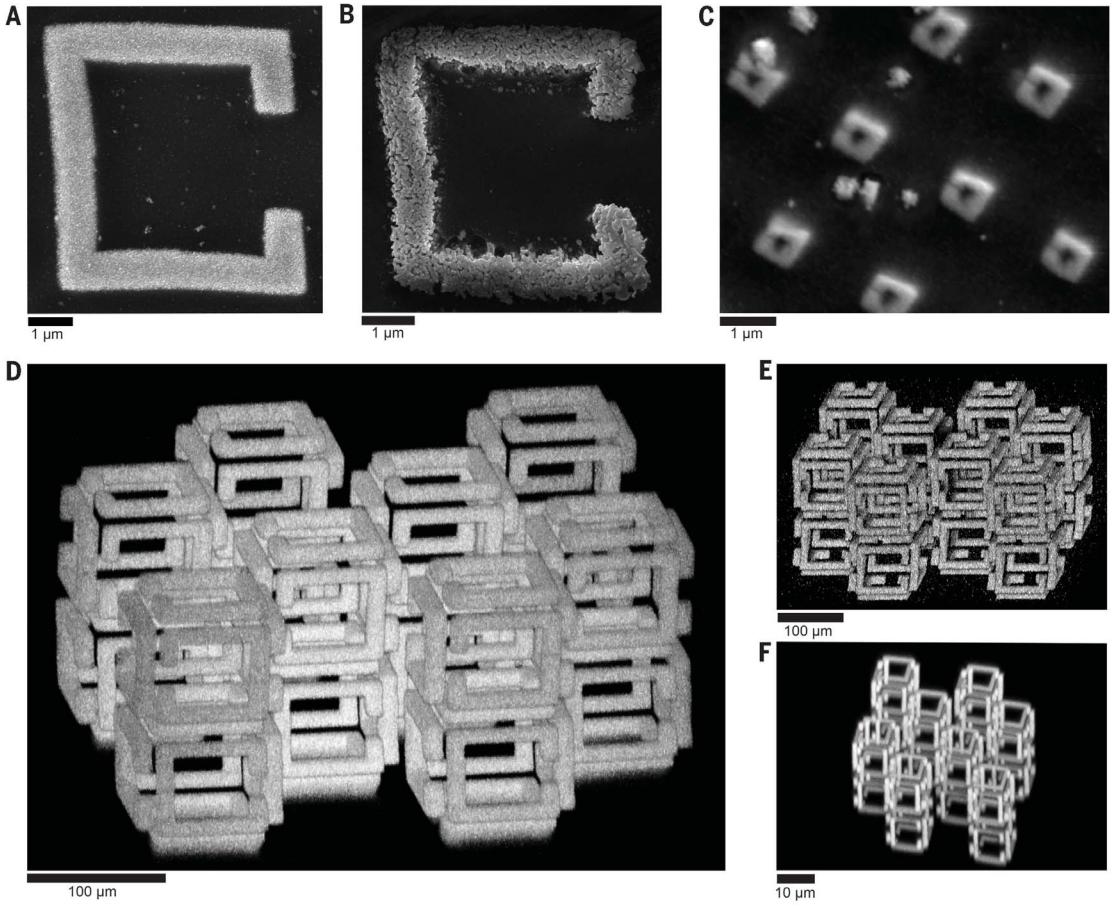


Fig. 4. Fabrication of 3D metal nanostructures. (**A** and **B**) 2D structures fabricated with ImpFab with micrometer-scale resolution, before (A) and after (B) sintering (visualized via SEM). (**C**) Similar structures fabricated with a 100-nm feature size, after shrinking and dehydration but before sintering. (**D**) Maximum-intensity projection of a fluorescent image of a 3D structure before shrinking (*2*, *28*). (**E**) Maximum-intensity projection of a reflected light image from the same structure following volumetric silver deposition, prior to shrinking. (**F**) Maximum-intensity projection of a fluorescent image of the same structure shrunken but not dehydrated.

with the gel substrate. Future iterations may use alternative chemistries, such as dendrimeric complexes for direct deposition of metals or semiconductors within the hydrogel (*25*, *26*), or DNA-addressed material deposition (*27*). Finally, we note that although we used a conventional microscope that was not optimized for patterning and that was limited to a 4-cm/s scan speed (in postshrink dimensions), we were able to create objects spanning hundreds of microns to millimeters (fig. S7). With the use of faster patterning systems (*23*), ImpFab could ultimately enable the creation of centimeter-scale nanomaterials.

REFERENCES AND NOTES

1. M. Deubel *et al*., *Nat. Mater.* **3**, 444–447 (2004).
2. C. M. Soukoulis, M. Wegener, *Nat. Photonics* **5**, 523–530 (2011).
3. C. A. Ross, K. K. Berggren, J. Y. Cheng, Y. S. Jung, J.-B. Chang, *Adv. Mater.* **26**, 4386–4396 (2014).
4. J.-B. Chang *et al*., *Nat. Commun.* **5**, 3305 (2014).
5. I. Wathuthanthri, Y. Liu, K. Du, W. Xu, C.-H. Choi, *Adv. Funct. Mater.* **23**, 608–618 (2013).
6. S. Matsui, T. Kaito, J. Fujita, M. Komuro, K. Kanda, *J. Vac. Sci. Technol. B* **3181**, 3–7 (2000).
7. S. Kawata, H. B. Sun, T. Tanaka, K. Takada, *Nature* **412**, 697–698 (2001).
8. L. R. Meza, S. Das, J. R. Greer, *Science* **345**, 1322–1326 (2014).
9. A. Vyatskikh *et al*., *Nat. Commun.* **9**, 593 (2018).
10. Y. Y. Cao, N. Takeyasu, T. Tanaka, X. M. Duan, S. Kawata, *Small* **5**, 1144–1148 (2009).
11. J. Bauer, A. Schroer, R. Schwaiger, O. Kraft, *Nat. Mater.* **15**, 438–443 (2016).
12. M. Hegde *et al*., *Adv. Mater.* **29**, 1701240 (2017).
13. X.-M. Zhao, Y. Xia, O. J. A. Schueller, D. Qin, G. M. Whitesides, *Sens. Actuators A Phys.* **65**, 209–217 (1998).
14. D. L. Holmes, N. C. Stellwagen, *Electrophoresis* **12**, 612–619 (1991).
15. F. Chen, P. W. Tillberg, E. S. Boyden, *Science* **347**, 543–548 (2015).
16. F. Ilmain, T. Tanaka, E. Kokufuta, *Nature* **349**, 400–401 (1991).
17. Y. Hirokawa, T. Tanaka, *AIP Conf. Proc.* **107**, 203–208 (1984).
18. A. Suzuki, T. Tanaka, *Nature* **346**, 345–347 (1990).
19. C. A. DeForest, K. S. Anseth, *Nat. Chem.* **3**, 925–931 (2011).
20. C. A. DeForest, K. S. Anseth, *Angew. Chem.* **124**, 1852–1855 (2012).
21. M. A. Skylar-Scott, M.-C. Liu, Y. Wu, M. F. Yanik, in *Proc. SPIE 10115, Advanced Fabrication Technologies for Micro/Nano Optics and Photonics X* (International Society for Optics and Photonics, 2017), abstr. 101150L.
22. M. A. Skylar-Scott, M.-C. Liu, Y. Wu, A. Dixit, M. F. Yanik, *Adv. Healthc. Mater.* **5**, 1233–1243 (2016).
23. M. A. Scott, Z. D. Wissner-Gross, M. F. Yanik, *Lab Chip* **12**, 2265–2276 (2012).
24. Materials and methods are available as supplementary materials.
25. A. Miura *et al*., *Opt. Mater. Express* **7**, 2224 (2017).
26. K. Esumi, A. Suzuki, N. Aihara, K. Usui, K. Torigoe, *Langmuir* **14**, 3157–3159 (1998).
27. P. W. K. Rothemund, *Nature* **440**, 297–302 (2006).
28. D. B. Burckel *et al*., *Adv. Mater.* **22**, 5053–5057 (2010).

ACKNOWLEDGMENTS

We thank, for helpful discussions, T. Swager, M. Soljacic, J. Bewersdorf, I. Gupta, and N. Barry. We also thank W. Salmon, N. Watson, N. Savidis, S. Terclavers, S. E. Kooi, E. B. Kromann, and M. Lessard for useful suggestions and technical assistance with imaging. We thank the anonymous referees for helpful comments that improved the manuscript. **Funding:** E.S.B. acknowledges funding by the Kavli Dream Team program, the HHMI-Simons Faculty Scholars Program, the Open Philanthropy Project, John Doerr, ONR N00014-17-1-2977, NIH 1R01EB024261, NIH 1U01MH106011, the New York Stem Cell Foundation-Robertson Award, NIH Director's Pioneer Award 1DP1NS087724, the U.S. Army Research Laboratory and the U.S. Army Research Office under contract number W911NF1510548, NIH 1R01MH103910, NIH 2R01DA029639, the MIT Media Lab, NIH 1RM1HG008525, and NIH 1R24MH106075. S.G.R. acknowledges funding through the Hertz Graduate Fellowship and the National Science Foundation Graduate Research Fellowship Program (award 1122374). **Author contributions:** S.G.R., D.O., M.A.S.-S., F.C., P.W.T., A.H.M., and E.S.B. conceived of strategies for implosion fabrication. S.G.R. and D.O. conceived of and developed the implosion fabrication gel, patterning, deposition, intensification, and shrink chemistries. S.G.R., D.O., R.G., A.H.M., and E.S.B. conceived of validation strategies. R.G. performed imaging for the results shown in Fig. 2F and fig. S4. S.A. wrote the program for generating arbitrary 3D patterns using the Zen software. S.G.R., D.O., A.H.M., and E.S.B. wrote the paper, with contributions and edits from all authors. **Competing interests:** S.G.R., D.O., S.A., R.G., F.C., P.W.T., A.H.M., M.A.S.-S., and E.S.B. are inventors on a patent filed on implosion fabrication. A.H.M. is also affiliated with Deepmind Technologies Ltd. F.C. is also affiliated with the Broad Institute of Harvard and MIT. **Data and materials availability:** Analyzed image data used to produce Fig. 1L and Fig. 2D (red bar) are provided as CSV files in the supplementary materials. Conductivity data used to produce Fig. 3, D and E, and fig. S6 are provided as CSV files in the supplementary materials. Raw image data used to produce fig. S2 and Fig. 2D (yellow and blue bars) and Fig. 2, G and H, are provided in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/1281/suppl/DC1
Materials and Methods
Figs. S1 to S7
Tables S1 to S3
Data S1

18 June 2018; accepted 2 November 2018
10.1126/science.aau5119

ENZYMOLOGY

Evolution of a highly active and enantiospecific metalloenzyme from short peptides

Sabine Studer[1], Douglas A. Hansen[1]*, Zbigniew L. Pianowski[1]†, Peer R. E. Mittl[2], Aaron Debon[1], Sharon L. Guffy[3], Bryan S. Der[3]‡, Brian Kuhlman[3,4], Donald Hilvert[1]§

Primordial sequence signatures in modern proteins imply ancestral origins tracing back to simple peptides. Although short peptides seldom adopt unique folds, metal ions might have templated their assembly into higher-order structures in early evolution and imparted useful chemical reactivity. Recapitulating such a biogenetic scenario, we have combined design and laboratory evolution to transform a zinc-binding peptide into a globular enzyme capable of accelerating ester cleavage with exacting enantiospecificity and high catalytic efficiency (k_{cat}/K_M ~ 10^6 M^{-1} s^{-1}). The simultaneous optimization of structure and function in a naïve peptide scaffold not only illustrates a plausible enzyme evolutionary pathway from the distant past to the present but also proffers exciting future opportunities for enzyme design and engineering.

Metal ions are ubiquitous in nature, playing structural and/or catalytic roles in nearly half of all proteins. This dual functionality conceivably fostered the emergence of primordial metalloenzymes from simpler peptidic precursors by an evolutionary pathway involving metal-mediated assembly, followed by polypeptide fusion and diversification (Fig. 1A) (*1*–*6*). In mimicry of this process, protein designers have successfully used metal ions to template binding of weakly interacting peptides and generate supramolecular structures that display modest catalytic activities at their interfaces (*7*–*14*). Here, such complexes are shown to be excellent starting points for the design and evolution of highly active, globular metalloenzymes.

To explore metalloprotein biogenesis, we chose MID1, a homodimeric peptide containing two interfacial Zn(II)His_3 sites that was computationally designed from a monomeric, 46–amino acid–long, helix-turn-helix fragment (*11*). The zinc ions originally served as prostheses for peptide assembly but also provided serendipitous activity for ester bond hydrolysis thanks to a small hydrophobic binding pocket adjacent to an open metal coordination site (*12*). Adopting nature's fusion and diversification strategy, we connected adjacent N and C termini of the dimer subunits via a short Gly-Ser-Gly linker and removed the zinc site farthest from the linker by replacing metal-binding residues with noncoordinating amino acids suggested by computation. The resulting single-chain MID1 variant, MID1sc, binds a single Zn(II) ion and hydrolyzes *p*-nitrophenyl acetate at similar rates as MID1 (fig. S1).

For protein evolution, we developed a robust and sensitive screening assay based on the racemic fluorogenic ester **1** (Fig. 2A). MID1sc hydrolyzes ester (±)-**1** with a turnover number (k_{cat}) of 0.011 ± 0.001 s^{-1} (mean ± SD) and an apparent second-order rate constant (k_{cat}/K_M) of 18 ± 2 M^{-1} s^{-1}. It also exhibits a twofold preference for cleavage of the (*R*)-configured substrate enantiomer (fig. S2 and table S1). We optimized this initial catalytic activity over nine rounds of laboratory evolution, exploiting both focused and random mutagenesis (Fig. 1B and fig. S3). Single residues close to the zinc center, lining the primitive binding pocket, and around the former zinc site were targeted by cassette mutagenesis, and the most promising mutations were shuffled. In addition, the full-length gene was randomized by error-prone polymerase chain reaction (PCR) to identify beneficial mutations distant from the active site. Over the course of evolution, self-acylating residues were replaced by arginine (Lys^{68} and Lys^{78}) or targeted for randomization (Arg^{80}) to prevent catalyst inactivation by covalent modification with the substrate (figs. S4 and S5). An average of one to two mutations were introduced per round of evolution to afford MID1sc9, which has a total of 20 substitutions distributed nearly equally over the N- and C-terminal helix-turn-helix fragments (Fig. 2B and fig. S6).

Because 21% of the protein was mutated by design and directed evolution, possible changes in Zn(II) coordination were probed by sequential replacement of each histidine in the original zinc binding site (His^{39}, His^{61}, and His^{65}) by alanine (fig. S7 and table S2). Surprisingly, substitution of His^{39} had little effect on catalytic activity, whereas replacement of His^{61} and His^{65} led to a greater than fivefold decrease in turnover number. Based on the sequence of the evolved protein, we identified another histidine (His^{35}) and two glutamates (Glu^{32} and Glu^{58}) as possible alternative metal-binding residues in close proximity to the original zinc site. Whereas substitution of Glu^{32} and Glu^{58} with glutamine had little effect on catalytic efficiency, alanine substitution of His^{35} reduced activity >1000 fold, strongly suggesting that His^{35}, together with His^{61} and His^{65}, binds the catalytic zinc ion. This change in coordination sphere occurred midway along the evolutionary trajectory because His^{35} could still be replaced after the third round of mutagenesis without loss in esterase activity (table S1). With the goal of eliminating potentially competitive zinc binding modes, we incorporated the E32Q (Glu^{32}→Gln), H39A (His^{39}→Ala), and E58Q (Glu^{58}→Gln) mutations to give the final optimized construct, MID1sc10, which had >10,000-fold higher activity than its MID1sc progenitor

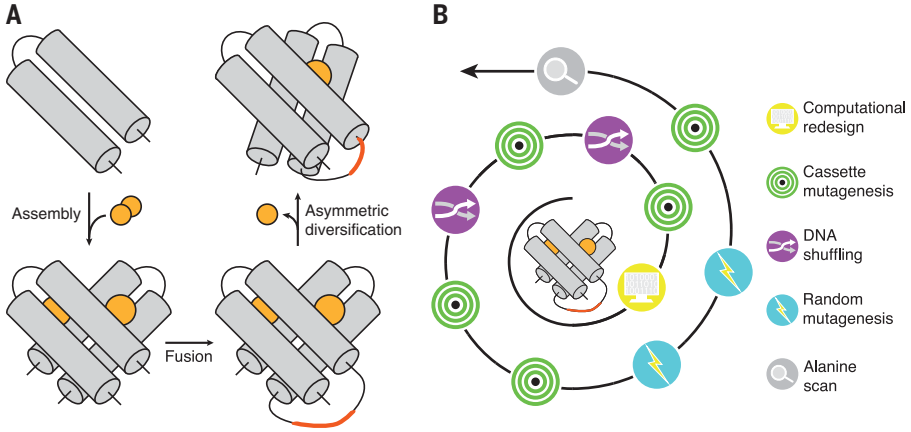


Fig. 1. Emulating metalloenzyme biogenesis from peptides. (**A**) Zinc-mediated assembly of helix-turn-helix fragments, followed by fusion and asymmetric diversification, afforded MID1sc10, an efficient metalloesterase. (**B**) Simplified schematic showing the specific steps performed in the diversification process.

[1]Laboratory of Organic Chemistry, ETH Zürich, 8093 Zürich, Switzerland. [2]Department of Biochemistry, University of Zürich, 8057 Zürich, Switzerland. [3]Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, NC 27599-7365, USA. [4]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599, USA.
*Present address: Antheia Inc., Menlo Park, CA 94025, USA.
†Present address: Institut für Organische Chemie, Karlsruher Institut für Technologie KIT, 76131 Karlsruhe, Germany.
‡Present address: Capital One, Henrico, VA 23238, USA.
§Corresponding author. Email: hilvert@org.chem.ethz.ch

at subsaturating concentrations of racemic ester **1**.

MID1sc10 is a highly active esterase. It preferentially catalyzes the hydrolysis of (*S*)-**1** with a k_{cat} of 1.64 ± 0.04 s^{-1} and a k_{cat}/K_M of 980,000 ± 110,000 M^{-1} s^{-1} (Fig. 2C and table S1). These steady-state parameters attest to notable catalytic proficiency [$1/K_{TS} = (k_{cat}/K_M)/k_{uncat} = 9.3 \times 10^{10}$ M^{-1}, where K_{TS} is the apparent transition-state binding affinity and k_{uncat} is the rate constant for the uncatalyzed reaction (*15*)]. In this respect, MID1sc10 is similar to typical natural enzymes (*16*) and outperforms other artificial esterases, including catalytic antibodies (*17*), computationally designed enzymes (*18*–*21*), and engineered zinc metalloproteins (*9*, *12*, *13*, *22*), by two to five orders of magnitude (table S3). It is also superior to the natural zinc metalloenzyme human carbonic anhydrase (hCAII), which promiscuously hydrolyzes similarly activated *p*-nitrophenyl acetate with a k_{cat}/K_M of 2500 M^{-1} s^{-1} (*23*). Even for its natural activity, the mechanistically related hydration of carbon dioxide, hCAII, a nearly perfect zinc enzyme, has a catalytic proficiency that is 100-fold lower than that of MID1sc10 (*15*).

Given the importance of stereochemical control for industrial biocatalysis, the high enantiospecificity achieved by MID1sc10, manifest in a 990-fold kinetic preference for cleavage of the (*S*)-configured ester (Fig. 2D and table S1), is particularly notable. As the entire screen was performed with racemic substrate, this property was never subject to direct selection pressure. However, active site mutations introduced in the third round of evolution fostered the initial switch from the (*R*)-specific starting scaffold, and the new stereochemical preference was subsequently enhanced in step with specific activity (table S1).

Zinc is absolutely required for MID1sc10 catalysis; removal inactivates the enzyme. Nevertheless, it binds relatively weakly with an apparent dissociation constant (K_d) of 26 μM (fig. S8). Consistent with weak binding, zinc does not stabilize the evolved protein. Its thermal denaturation is unaffected by Zn(II) (Fig. 3A), whereas the metal ion increases MID1's melting temperature by 24°C (*11*). When Zn(II) is added to apo-MID1sc10, signal broadening is observed in the ^{1}H-^{15}N-heteronuclear single-quantum coherence (HSQC) nuclear magnetic resonance (NMR) spectrum (Fig. 3B and fig. S9), suggesting the presence of several states that interconvert on an intermediate time scale. Together, these results indicate that design and evolution converted the zinc ion from an essential structural element into a dedicated catalytic cofactor.

To elucidate the origins of these effects, we cocrystallized MID1sc10 with racemic phosphonate **4**, a structural mimic of the esterolytic transition state that competitively inhibits the enzyme with an inhibition constant (K_i) of 1.1 ± 0.1 μM (fig. S10) and increases the enzyme's affinity for Zn(II) more than 100-fold (fig. S8). The crystal structure, solved at 1.34-Å resolution (Figs. 2B and 3, C to F, and table S4), confirmed that the protein adopts a helical bundle fold,

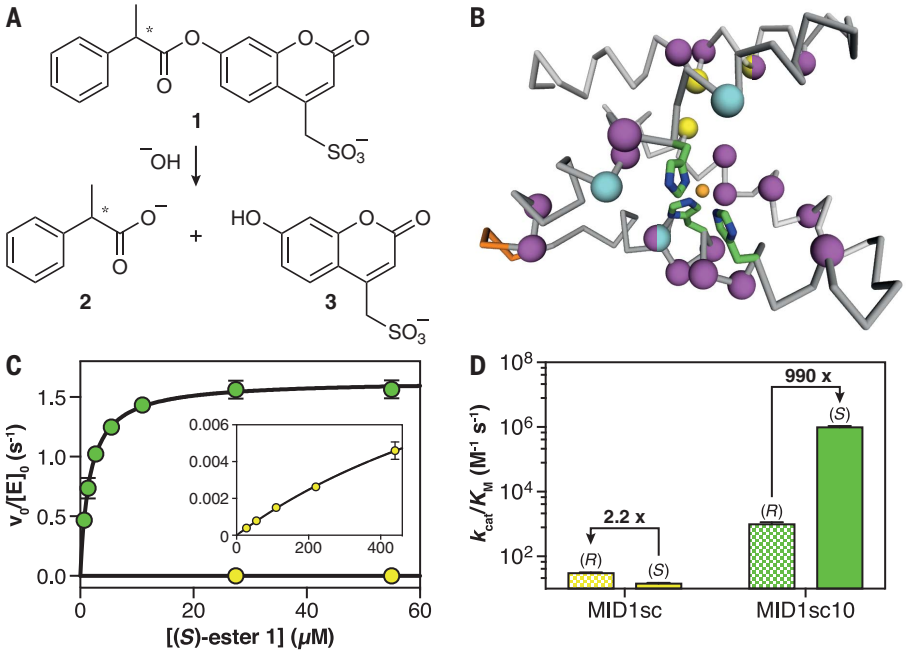


Fig. 2. Directed evolution of MID1sc. (**A**) MID1sc was evolved for the hydrolysis of fluorogenic ester **1** to give 2-phenylpropionate **2** and coumarin **3**. The * indicates the chiral center. (**B**) Crystal structure of MID1sc10, showing the zinc ion as an orange sphere and the coordinating histidines as green sticks. Linkage of two polypeptides via a Gly-Ser-Gly sequence (orange) and removal of a second zinc site present in the original MID1 design (yellow spheres) afforded MID1sc, which was subsequently optimized by mutagenesis and screening. The locations of beneficial mutations (magenta spheres) and residues replaced to prevent competitive zinc binding modes (cyan spheres) are highlighted. (**C**) Michaelis-Menten plots for MID1sc (yellow and inset) and MID1sc10 (green) show a 70,000-fold improvement in hydrolysis efficiency for (*S*)-configured **1** after optimization. $v_0/[E]_0$, initial rate divided by total enzyme concentration. (**D**) The evolved variant MID1sc10 is highly enantioselective as a consequence of a 2200-fold specificity switch from the modestly (*R*)-selective starting catalyst MID1sc. All error bars represent the SD of at least three independent measurements.

albeit with substantial structural changes compared with MID1 (*11*). In addition to the altered Zn(II) coordination sphere identified by mutagenesis, the crossover angle of the two helix-turn-helix fragments decreased to 47°, which is >30° tighter than in MID1 (Fig. 3C, fig. S11, and table S5) but still considerably larger compared with canonical four-helix bundles (typically 20°) (*24*). This dramatic conformational change was brought about by extensive remodeling of the protein interior to accommodate the large ester substrate. Five out of 13 residues lining the substrate binding pocket were mutated [M38W (Met^{38}→Trp), K68R (Lys^{68}→Arg), Q80S (Gln^{80}→Ser), L83T (Leu^{83}→Thr), and H84L (His^{84}→Leu)], substantially deepening and reconfiguring the active site for shape-complementary transition-state recognition (Fig. 3, D and E).

Another early mutation, Q36P (Gln^{36}→Pro), introduced a kink in the second helix, helping to form a tighter binding pocket for the substrate and facilitating replacement of His^{39} by His^{35} as a Zn(II) ligand (fig. S12 and movie S1). The resulting metal environment (Fig. 3F and fig. S7) resembles the zinc site in carbonic anhydrase (*25*). Introduction of a second-shell hydrogen-bonding interaction between Gln^{58} and the backside nitrogen of His^{61} is intriguing in this context, because natural zinc enzymes utilize similar interactions to tune metal ion reactivity (*25*, *26*). Like carbonic anhydrase, MID1sc10 presumably exploits the Lewis acidity of Zn(II) to acidify a coordinated water molecule and generate a high local concentration of hydroxide for substrate cleavage. Fitting the pH-rate data for ester hydrolysis afforded a kinetic pK_a of ~8 (fig. S13), which is higher than the value of 6.8 determined for ionization of zinc-bound water in carbonic anhydrase (*25*) but falls in the range of pK_a's observed for other peptides and model complexes (*26*), including MID1 (*12*).

Consistent with MID1sc10's high enantiospecificity, only the (*S*)-enantiomer of phosphonate **4** is bound in the crystallized complex (fig. S14). The inhibitor adopts an extended conformation with the 2-phenylpropionyl group sitting snugly at the bottom of the hydrophobic pocket and the charged leaving group near the entrance of the active site (Fig. 3E). This orientation allows the phosphonate to coordinate Zn(II) via one of its oxygen atoms, as expected for a mechanism involving nucleophilic attack of a zinc-bound hydroxide on the ester substrate (Fig. 3F). The other phosphonyl oxygen forms a bidentate hydrogen bond with the side chain of Arg^{68}, a residue introduced during evolution. Similar interactions have been observed in zinc enzymes like carboxypeptidase A (*27*) and contribute

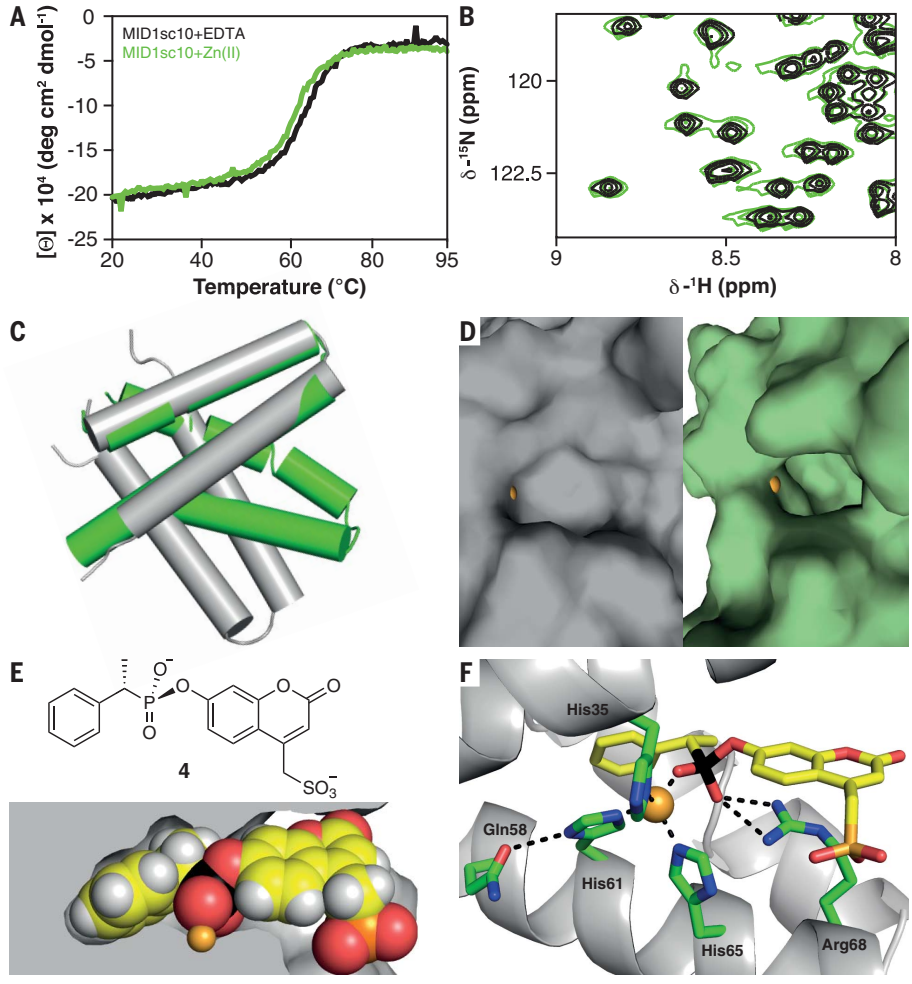


Fig. 3. Biophysical characterization and crystal structure of MID1sc10. (**A**) The thermal stability of MID1sc10 is similar in the presence (green) and absence (black) of zinc. [Θ], mean residue ellipticity. (**B**) Overlay of the ^{1}H-^{15}N-HSQC spectra of MID1sc10 in the presence (green) and absence (black) of zinc. For the full spectrum, see fig. S9. δ, chemical shift; ppm, parts per million. (**C**) Structural alignment of MID1sc10 (green) and MID1 (gray), illustrating the >30° tighter crossover angle. (**D**) The observed structural changes transformed the shallow binding site of MID1 (gray) into a deep, hydrophobic pocket in MID1sc10 (green). (**E**) Cut-away view of the active site, showing the snug fit of phosphonate **4** in the binding pocket. The zinc ion is shown as an orange sphere and the ligand is shown in space-filling representation (carbon, yellow; oxygen, red; phosphorus, black; sulfur, orange). (**F**) View of the MID1sc10 active site with phosphonate **4** (yellow) coordinating to the Zn(II)His_3 complex (orange sphere and green sticks). Arg^{68} and Gln^{58} form mechanistically relevant hydrogen bonds to phosphonate **4** and the backside nitrogen of His^{61}, respectively.

to electrostatic stabilization of the anionic transition state. In MID1sc10, the guanidinium group of Arg^{68} additionally makes productive cation-π interactions with the coumarin, which may assist departure of the leaving group.

Although the evolved catalyst shows good activity with *p*-nitrophenol and coumarin esters of 2-phenylpropionate, catalytic efficiency drops substantially for esters of simple aliphatic acids (table S6). For example, the k_{cat}/K_M for *p*-nitrophenyl acetate is similar to that observed for the starting catalyst. As for natural hydrolases, shape-complementary binding interactions between the enzyme and portions of the substrate distant from the scissile bond contribute substantially to catalytic efficiency (*28*), presumably by properly positioning the ester for effective reaction.

The extraordinary activities, efficiency, and specificities of modern-day metalloenzymes are the products of eons of evolution. The bottom-up construction of a zinc-dependent esterase by end-to-end doubling of the MID1 peptide and subsequent directed evolution shows that the putative historical roads taken by these natural catalysts are also fruitful avenues for producing new enzymes. The de novo generation of a highly active metalloesterase in this way compares favorably with computational enzyme design, which uses sophisticated software algorithms to equip the binding pockets of natural protein scaffolds with the catalytic functionality needed to accelerate a chosen target reaction and is one of the most promising approaches to tailored catalysts to emerge in the past few years (*29*, *30*). Although computationally designed enzymes have been evolved to high activities for several reactions (*31*–*34*), creation of efficient catalysts for the hydrolysis of esters like **1** has proved challenging (*18*–*21*). Instead of a metal ion cofactor, computational designs have relied on a single nucleophile (*18*, *20*) or embedded catalytic dyads (*19*) and triads (*21*) to cleave the substrate via a transient acyl-enzyme intermediate. However, even after laboratory evolution, the apparent second-order rate constants for protein acylation (k_2/K_S) have never exceeded 2000 M^{-1} s^{-1} (table S3), and slow deacylation limits overall turnover (k_{cat}/K_M < ~100 M^{-1} s^{-1}).

The comparative ease of evolving a 10,000-fold more efficient zinc-dependent esterase is thus striking and speaks to the efficacy of metal ion catalysis. Even though no reaction-relevant chemical information was provided by design, the optimized MID1sc10 active site recapitulates the natural mechanisms of native zinc enzymes, suggesting that the intrinsic chemical potential of such systems is readily realizable once the metal ion is installed in an appropriate binding pocket. The flexibility of the helical bundle fold may have been advantageous in this respect, expediting the evolutionary search for a chemically and sterically complementary binding pocket that could effectively align substrate and metal-ion-bound water and lower the transition state barrier for ester hydrolysis.

MID1sc10 embodies the structural and functional properties that metals likely imparted to proteins long ago. Promiscuous binding of different substrate molecules and metal ions by primordial scaffolds would have been a potentially rich source of novel activities. Looking forward, our simple metalloprotein may similarly constitute an excellent system for exploring divergent evolution and functional diversification. By elucidating how sophisticated enzymatic functions emerge from naïve peptide scaffolds, such experiments have the potential to inform ongoing efforts to create new metal-dependent protein catalysts for chemical transformations unknown in nature (*8*, *35*–*38*).

REFERENCES AND NOTES

1. R. V. Eck, M. O. Dayhoff, *Science* **152**, 363–366 (1966).
2. A. Lombardi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6298–6305 (2000).
3. J. R. Calhoun *et al.*, *J. Mol. Biol.* **334**, 1101–1115 (2003).
4. E. N. Salgado *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1827–1832 (2010).
5. M. L. Romero Romero, A. Rabin, D. S. Tawfik, *Angew. Chem. Int. Ed.* **55**, 15966–15971 (2016).
6. V. Alva, A. N. Lupas, *Curr. Opin. Struct. Biol.* **48**, 103–109 (2018).
7. J. Kaplan, W. F. DeGrado, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11566–11570 (2004).
8. M. Chino *et al.*, *Eur. J. Inorg. Chem.* **2015**, 3371–3390 (2015).
9. M. L. Zastrow, A. F. A. Peacock, J. A. Stuckey, V. L. Pecoraro, *Nat. Chem.* **4**, 118–123 (2012).
10. M. Tegoni, F. Yu, M. Bersellini, J. E. Penner-Hahn, V. L. Pecoraro, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21234–21239 (2012).
11. B. S. Der *et al.*, *J. Am. Chem. Soc.* **134**, 375–385 (2012).

12. B. S. Der, D. R. Edwards, B. Kuhlman, *Biochemistry* **51**, 3933–3940 (2012).
13. W. J. Song, F. A. Tezcan, *Science* **346**, 1525–1528 (2014).
14. O. Zozulia, M. A. Dolan, I. V. Korendovych, *Chem. Soc. Rev.* **47**, 3621–3639 (2018).
15. A. Radzicka, R. Wolfenden, *Science* **267**, 90–93 (1995).
16. A. Bar-Even *et al*., *Biochemistry* **50**, 4402–4410 (2011).
17. G. MacBeath, D. Hilvert, *Chem. Biol.* **3**, 433–445 (1996).
18. D. N. Bolon, S. L. Mayo, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14274–14279 (2001).
19. F. Richter *et al*., *J. Am. Chem. Soc.* **134**, 16197–16206 (2012).
20. Y. S. Moroz *et al*., *J. Am. Chem. Soc.* **137**, 14905–14911 (2015).
21. A. J. Burton, A. R. Thomson, W. M. Dawson, R. L. Brady, D. N. Woolfson, *Nat. Chem.* **8**, 837–844 (2016).
22. C. M. Rufo *et al*., *Nat. Chem.* **6**, 303–309 (2014).
23. J. A. Verpoorte, S. Mehta, J. T. Edsall, *J. Biol. Chem.* **242**, 4221–4229 (1967).
24. W. F. DeGrado, Z. R. Wasserman, J. D. Lear, *Science* **243**, 622–628 (1989).
25. D. W. Christianson, C. A. Fierke, *Acc. Chem. Res.* **29**, 331–339 (1996).
26. M. L. Zastrow, V. L. Pecoraro, *Biochemistry* **53**, 957–978 (2014).
27. M. A. Phillips, R. Fletterick, W. J. Rutter, *J. Biol. Chem.* **265**, 20692–20698 (1990).
28. A. Fersht, *Structure and Mechansim in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman and Company, 1999).
29. G. Kiss, N. Çelebi-Ölçüm, R. Moretti, D. Baker, K. N. Houk, *Angew. Chem. Int. Ed.* **52**, 5700–5725 (2013).
30. H. Kries, R. Blomberg, D. Hilvert, *Curr. Opin. Chem. Biol.* **17**, 221–228 (2013).
31. O. Khersonsky *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10358–10363 (2012).
32. R. Blomberg *et al*., *Nature* **503**, 418–421 (2013).
33. N. Preiswerk *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **111**, 8013–8018 (2014).
34. R. Obexer *et al*., *Nat. Chem.* **9**, 50–56 (2017).
35. Y. Lu, N. Yeung, N. Sieracki, N. M. Marshall, *Nature* **460**, 855–862 (2009).
36. F. Yu *et al*., *Chem. Rev.* **114**, 3495–3578 (2014).
37. J. Bos, G. Roelfes, *Curr. Opin. Chem. Biol.* **19**, 135–143 (2014).
38. F. Schwizer *et al*., *Chem. Rev.* **118**, 142–231 (2018).

ACKNOWLEDGMENTS

The authors thank C. Stutz-Ducommun and B. Blattmann from the Protein Crystallization Core Facility at the University of Zürich, R. Arnold from the NMR Service of the Laboratory of Organic Chemistry at ETH Zürich, the Mass Spectrometry Service team of the Laboratory of Organic Chemistry at ETH Zürich, the Protein Analysis Group at the Functional Genomics Center of the University of Zürich, and the staff at the Swiss Light Source (Paul Scherrer Institute) for technical support. The authors also thank Y. Azuma, M.-O. Ebert, T. Edwardson, A. Vandemeulebroucke, and D. L. Niquille for helpful discussions. **Funding:** The work was supported by the Lithuanian-Swiss Research and Development cooperation program (CH-3-SMM-01/03), the Swiss National Science Foundation (31003A-156276), the NIH (R01GM073960 and T32GM008570), and a postdoctoral fellowship from the ETH (to D.A.H.). **Author contributions:** S.S., B.K., and D.H. designed the research; B.S.D. and S.L.G carried out the design simulations; S.S. performed enzyme evolution, characterization, and crystallization; D.A.H. and Z.L.P. performed chemical synthesis; A.D. performed acylation experiments and enzyme characterization; P.R.E.M. solved the crystal structure; and S.S., D.A.H., and D.H. wrote the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the main text or the supplementary materials. Plasmids encoding the enzymes reported in this study are available for research purposes from D.H. under a material transfer agreement with the ETH Zürich. Coordinates and structure factors have been deposited in the Protein Data Bank with the accession code 5OD1.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/1285/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S7
References (*39*–*62*)
Movie S1

4 June 2018; accepted 31 October 2018
10.1126/science.aau3744

CHEMICAL PHYSICS

Observation of the geometric phase effect in the H + HD → H_2 + D reaction

Daofu Yuan[1]*, Yafu Guan[2]*, Wentao Chen[1], Hailin Zhao[2], Shengrui Yu[3], Chang Luo[1], Yuxin Tan[1], Ting Xie[1], Xingan Wang[1]†, Zhigang Sun[2]†, Dong H. Zhang[2]†, Xueming Yang[2,4]†

Theory has established the importance of geometric phase (GP) effects in the adiabatic dynamics of molecular systems with a conical intersection connecting the ground- and excited-state potential energy surfaces, but direct observation of their manifestation in chemical reactions remains a major challenge. Here, we report a high-resolution crossed molecular beams study of the H + HD → H_2 + D reaction at a collision energy slightly above the conical intersection. Velocity map ion imaging revealed fast angular oscillations in product quantum state–resolved differential cross sections in the forward scattering direction for H_2 products at specific rovibrational levels. The experimental results agree with adiabatic quantum dynamical calculations only when the GP effect is included.

In a system of potential energy surfaces (PESs) connected through a conical intersection (CI), a geometric phase (GP) must be introduced that pertains to adiabatic motions encircling the CI for the system to be treated properly in the adiabatic quantum mechanical framework. The GP effect was discovered independently by Pancharatnam in 1956 in crystal optics (*1*) and by Longuet-Higgins *et al.* in 1958 in molecular systems (*2*). In 1984, Berry (*3*) generalized the GP (also known as Berry phase) effect to all adiabatic processes, after which it became a widely studied topic in physics. Over the past three decades, the potentially profound influence of the GP on material properties such as polarization, orbital magnetism, piezoelectric and ferroelectric properties, and quantum Hall effects has become clear (*4*–*6*). The concept of GP is now essential for a coherent understanding of many basic phenomena in physics.

CIs appear in the PESs of many molecular systems and chemical reaction coordinates (*7*). Near a CI, electronic motion and nuclear motion are strongly coupled in contravention of the Born-Oppenheimer approximation. When a molecular system with a CI is treated theoretically in the adiabatic framework, i.e., only considering the lower energy electronic surface, the GP must be introduced to ensure, in accord with quantum mechanics, that the total wave function is single-valued at each nuclear geometry. GP effects have been investigated in detail in isolated molecules such as the sodium trimer (*8*), as well as in the phenol photodissociation process (*9*, *10*).

The most important chemical reaction for the study of the GP effect is the hydrogen exchange reaction, H + H_2 → H_2 + H, because it has a well-defined CI and can be treated most accurately by theory. In the associated set of PESs for this reaction, the CI between the ground electronic state and the first excited state lies at about 2.75 eV (in total energy) (*11*), at which three hydrogen nuclei form an equilateral triangular geometry of D_{3h} symmetry. In pioneering work on the role of GP in the H + H_2 → H_2 + H reaction, Mead and Truhlar showed that the GP would affect observables only if the nuclear wave function encircled the CI, and the effect could be included by introducing a vector potential (*12*). In 1988, Zhang and Miller performed full-dimensional quantum dynamics calculations on the hydrogen exchange reaction without considering the GP effect, which agreed with the relevant experimental observation, suggesting the GP effect is not important in this reaction at low collision energy (*13*, *14*). Kuppermann and co-workers studied the GP effect on the H + H_2 reaction using the multivalued basis functions approach (*15*, *16*) and predicted strong GP effects in the differential cross sections (DCSs). Their predictions, however, were not reproduced by later dynamics calculations (*17*–*19*) and by experiments (*20*, *21*). Quantum reactive scattering studies by Kendrick and co-workers and by Althorpe and co-workers established that the GP effect should be negligible at total energy below 1.8 eV (*19*, *22*–*25*), becoming significant only at total energy above 3.5 eV. Theoretical studies also pointed out that a clear signature of the GP effect on this reaction would be a shift of the fast angular oscillation in DCSs in the sideways scattering direction (*19*, *26*).

Over the past two decades, high-resolution crossed beam studies using the H atom tagging method have probed many important elementary reactions (*27*–*30*), including the H + D_2 and H + HD reactions at various collision energies (*20*, *21*, *31*–*33*). No fast angular oscillations in DCSs for these latter reactions have been observed, most likely because the angular resolution

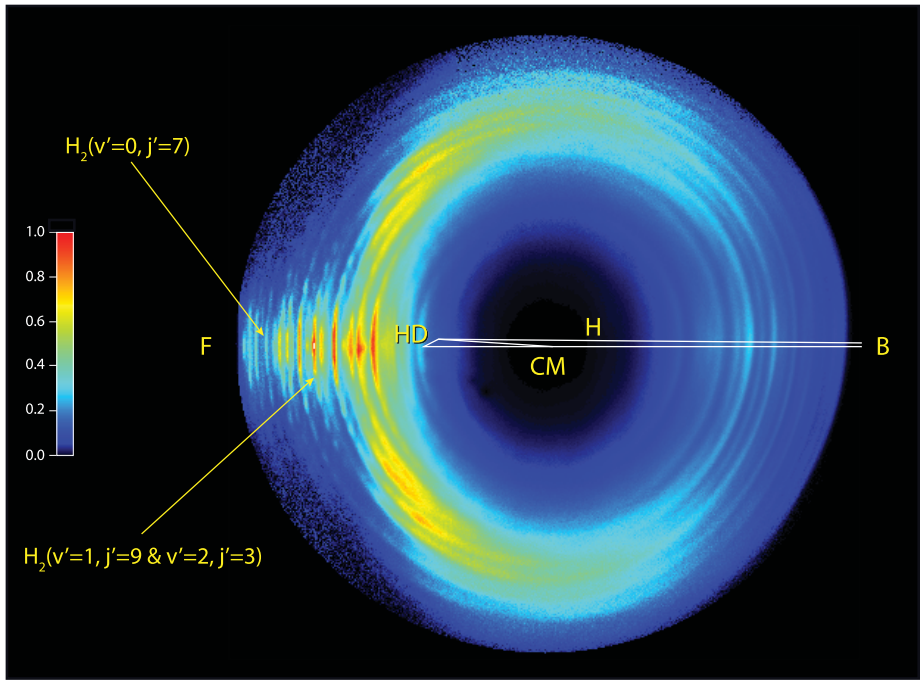


Fig. 1. Experimental images of the D atom product from the H + HD → H_2 + D reaction at a collision energy of 2.77 eV. The crossing angle of the two beams is 160°. F and B denote the forward (0°) and the backward scattering direction (180°) for the H_2 coproduct in the center-of-mass frame (CM) relative to the H atom beam direction, respectively.

[1]Hefei National Laboratory for Physical Sciences at the Microscale and Department of Chemical Physics, University of Science and Technology of China, Hefei, 230026, China. [2]State Key Laboratory of Molecular Reaction Dynamics, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, Dalian, 116023, China. [3]Hangzhou Institute of Advanced Studies, Zhejiang Normal University, Hangzhou, 311231, China. [4]Department of Chemistry, School of Science, Southern University of Science and Technology, Shenzhen, 518055, China.
*These authors contributed equally to this work.
†Corresponding author. Email: xawang@ustc.edu.cn (X.W.); zsun@dicp.ac.cn (Z.S.); zhangdh@dicp.ac.cn (D.H.Z.); xmyang@dicp.ac.cn or yangxm@sustc.edu.cn (X.Y.)

of the experimental method was limited. More recently, the PHOTOLOC (photoinitiated reaction analyzed by the law of cosines) technique has been applied to this search but with a similarly negative outcome (*34*–*36*).

We have developed a high-resolution time-sliced velocity map ion imaging (VMI) apparatus for H(D) atom product detection using the threshold ionization technique for crossed beams scattering studies (*37*). The VMI technique has proven to be a powerful technique for accurately measuring angular distributions of scattering products (*38*). The application of the threshold ionization scheme in this apparatus for D atom product detection in the H + HD → H_2 + D reaction substantially reduced the recoil of the electrons and consequently improved the velocity resolution for the D atom product significantly. Because of the high angular and velocity resolution, fast forward angular oscillations in this reaction at the collision energy of 1.35 eV have been observed and were attributed to corona scatterings in the reaction (*37*). At this collision energy, the reaction appears to occur with a simple direct abstraction mechanism. Through this study, we concluded that the GP effect plays a negligible role in the dynamics of this reaction at this collision energy, which is far below the CI total energy of 2.75 eV (2.53 eV in collision energy).

Here, we report a high-resolution crossed beams study on the H + HD → H_2 + D reaction at a collision energy of 2.77 eV, corresponding to 2.99 eV in total energy relative to the equilibrium energy of an H_2 molecule, or 0.24 eV above the CI. In addition, we have carried out accurate adiabatic quantum mechanical calculations with and without considering the GP effect, as well as diabatic quantum dynamics calculations, to investigate the GP effect on this reaction.

In this experiment, the H atom beam was generated by 193-nm laser photolysis of HI molecules in a pure HI beam at the nozzle tip. The fast H atom beam produced from the H + I($^2P_{3/2}$) channel was selected to react with HD. The HD beam was produced by supersonic expansion through a second pulsed valve (Even Lavie valve). The HD gas sample was cooled to liquid nitrogen temperature before expanding to the source chamber vacuum by means of a pulsed nozzle. About 97% of the HD molecules in the beam were in the ground vibrational and rotational level (v = 0, j = 0). Both the pulsed H atom beam and the HD beam were collimated by skimmers before entering the scattering chamber. The two beams were spatially and temporally overlapped. Differential pumping was used to reduce the residual HI background in the scattering chamber. The D atom product was ionized by means of a two-color [vacuum ultraviolet (VUV) + ultraviolet] threshold ionization scheme and subsequently detected using a VMI detector. During the experiment, the VUV laser wavelength was scanned back and forth to cover the entire Doppler profile of the D atom product to achieve uniform detection efficiency for the D atom products with different velocities. For more details about the experimental setup, refer to the materials and methods in the supplementary materials (SM).

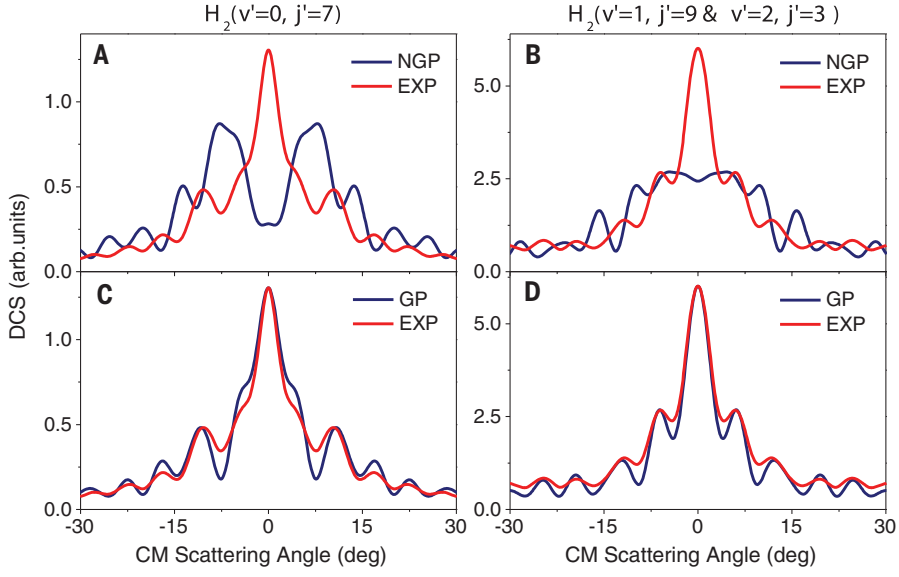


Fig. 2. Comparisons of the experimental (EXP) and theoretical product angular distributions of the H_2 product from the H + HD (v = 0, j = 0) ⟶ H_2 + D reaction at a collision energy of 2.77 eV. (**A** and **C**) Product rovibrational state is v' = 0, j' = 7. (**B** and **D**) Product rovibrational states are v' = 1, j' = 9 and v' = 2, j' = 3, which appear in the measured image as a merged ring. The theoretical results (dark blue lines) do not include the GP (NGP) in panels A and B but do include it (GP) in panels C and D. arb., arbitrary; deg, degree.

The experimental velocity map image of the D atom product from the H + HD → H_2 + D reaction at the collision energy of 2.77 eV (Fig. 1) shows rings that are well resolved in the forward scattering direction. These ring structures correspond to different rovibrational state structures of the H_2 product and are assignable (see fig. S4). Certain ring structures arise from a single rovibrational state, whereas most encompass combined rovibrational states of H_2. For each ring, there are fine oscillations in the angular distribution in the forward direction as observed in the study at the collision energy of 1.35 eV. We then acquired the experimental angular distributions for the H_2 product at the rovibrational level (v' = 0, j' = 7) and at the combined levels (v' = 1, j' = 9 and v' = 2, j' = 3) in the forward scattering direction by extracting the signals at a set of different scattering angles (in 1° intervals) for the corresponding rings (Fig. 2, A and B).

To ascertain whether the GP effect markedly influenced this reaction at this high collision energy, we first carried out adiabatic quantum dynamics calculations on the accurate adiabatic Boothroyd-Keogh-Martin-Peterson-2 (BKMP2) PES with the GP effect not included (Fig. 2, A and B). The angular distribution patterns from the calculations with no GP (NGP) are not in agreement with the corresponding experimental results: The oscillation patterns in the calculated NGP DCS are almost completely out of phase with the experimental results, with theoretical peaks located at the experimental valley positions. In particular, the experimental angular distribution for the H_2 (v' = 0, j' = 7) product state shows a pronounced peak in the exact forward direction (0°), whereas the theoretically calculated NGP distribution exhibits a deep valley there. The same calculations have also been performed on the complete configuration interaction (CCI) PES, which is considered the most accurate adiabatic PES for the reaction system (*39*). The calculated results on the CCI PES are essentially the same as those obtained on the BKMP2 PES (see fig. S5), indicating that the disagreement between the experiment and the NGP calculation is not due to inaccuracies associated with a particular adiabatic PES. Similar comparisons were made for additional H_2 product rovibrational states (see fig. S6), and the NGP-calculated angular distributions similarly disagreed with the experimental results.

We then carried out time-dependent adiabatic quantum dynamics calculations for the reaction on the BKMP2 PES with inclusion of the GP as a vector potential, as Althorpe and co-workers had done for the H + H_2 reaction (*23*). The application of the vector potential for the H + HD reaction is slightly more complicated than for the H + H_2 reaction because of the asymmetric masses. In the present calculations, this vector potential was first derived in the mass-scaled hyperspherical coordinates and then was expressed in the reactant Jacobi coordinates for the subsequent quantum reactive scattering calculations. For more details about the reactive scattering theory including the GP in reactant Jacobi coordinates (*40*), refer to section VI in the SM. The calculated angular distributions with the GP effect included are shown in Fig. 2, C and D. In marked contrast to the NGP results,

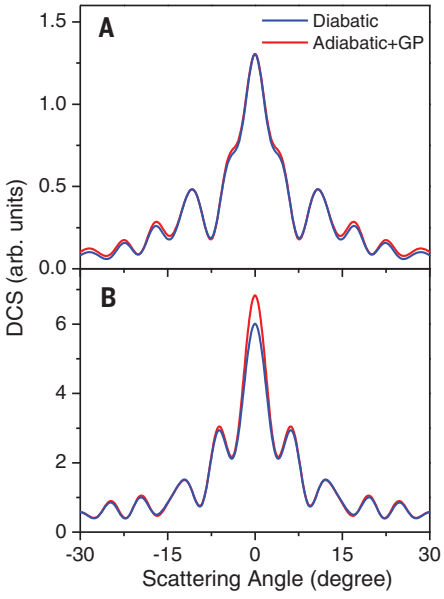


Fig. 3. Comparison between diabatic and adiabatic with GP calculations for H_2 product in specific quantum states. (**A**) H_2 (v' = 0, j' = 7); (**B**) H_2 (v' = 1, j' = 9 and v' = 2, j' = 3).

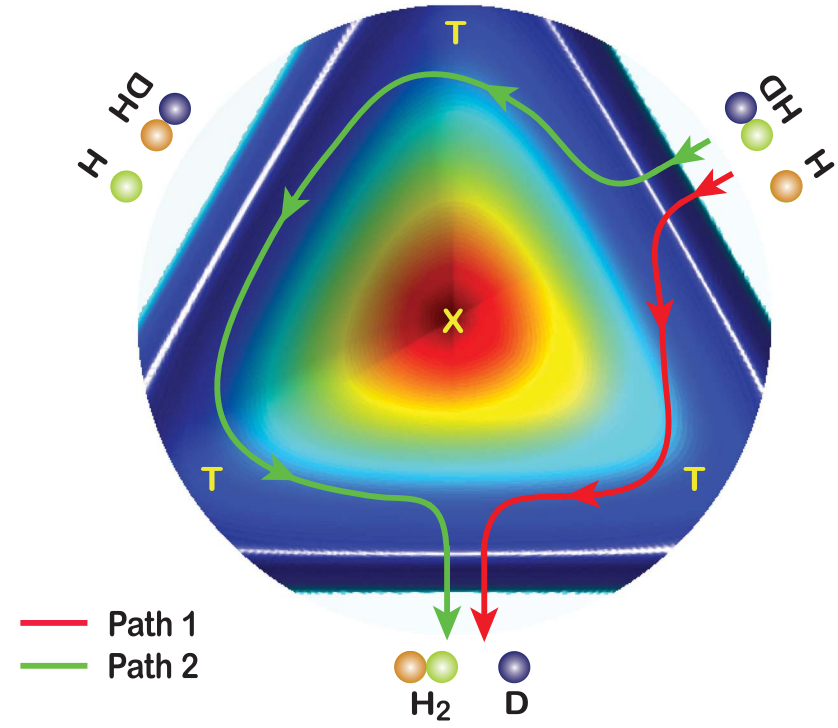


Fig. 4. A cut view through the H + HD PES. The positions of the three H + HD geometric arrangements, transition states (T), and CI (×) are shown. On the surface, representative one–transition state (path 1) and two–transition state (path 2) reaction paths are shown. The cut was calculated using hyperspherical coordinates (*45*) at a given overall separation ρ of 3.60 bohr without consideration of the mass difference between H and D atoms.

the theoretical angular distributions obtained with the GP effect included agree well with the experimental results, with the calculated angular oscillations exactly in phase with the experimental results. This agreement suggests strongly that the GP effect can be seen in the adiabatic picture for this benchmark reaction at this high collision energy. Similar comparisons were made for additional H_2 product levels (fig. S6), and results were consistent with the above conclusion.

Because the collision energy of this experiment is 0.24 eV above the CI, the question arises whether the adiabatic excited state (or the upper cone of the CI) has a significant effect on the reaction dynamics. We therefore developed accurate diabatic PESs for the H_3 system and used them to carry out state-to-state quantum dynamics calculations. To construct the diabatic PESs, we obtained the derivative coupling between the two lowest $^2A'$ states by performing MR-CISD (multireference configuration interaction, with all single and double excitations) calculations using the COLUMBUS program (*41*) with active space comprising three electrons distributed in nine a' and two a'' orbitals and basis of standard aug-cc-pVQZ (*42*). The derivative couplings were then fitted using an artificial neural network method (*43*). The ground adiabatic PES of H_3 was taken as the well-known BKMP2 PES, but the energy difference between the ground and excited states was calculated using the MOLPRO package (*44*) and fitted using the artificial neural network method. See the SM for more details. The DCSs for the title reaction were then calculated using the diabatic PESs for the products H_2 (v' = 0, j' = 7) and H_2 (v' = 1, j' = 9 and v' = 2, j' = 3) and are compared with the corresponding adiabatic GP results in Fig. 3. The calculated DCS using the adiabatic ground-state PES including the GP effect agrees well with the DCS calculated using the diabatic coupled PESs, and the calculated diabatic DCS is also in good agreement with the experimental result, demonstrating that the dynamics of the reaction can be accurately described using the diabatic theory without considering the GP effect, as expected. Therefore, the GP effect associated with the CI in a molecular system exists only in the adiabatic picture. The present results also verify that the adiabatic theory including the GP can be used to describe the detailed dynamics of this chemical reaction at this collision energy as precisely as the diabatic theory does. This, we believe, has important implications for dynamics studies of complicated quantum systems with CIs using adiabatic theory when diabatic treatment is very difficult or not possible.

There are some small differences in the forward scattering peak between the diabatic and the adiabatic GP results for the H_2 product (v' = 1, j' = 9 and v' = 2, j' = 3) (Fig. 3B), implying that the excited state might play some small role at this collision energy. To assess quantitatively the effect of the excited state, we have calculated the time-dependent population of the adiabatic ground (V_1) and excited (V_2) states for H + HD at the collision energy of 2.77 eV for differential partial waves J = 0, 10, 20, 30, and 40. The calculated results show that the J = 0 population on the adiabatic excited state V_2 reaches its maximum at ~46 fs, which is still less than 0.09% of that on the adiabatic ground state (see fig. S7A). In addition, we have also computed the time-independent wave function as a function of hyperradius ρ in hyperspherical coordinates for J = 0 with the two hyperangular coordinates integrated out (*45*). The results show that the wave function of the adiabatic excited V_2 is distributed in a very narrow region around the CI, with peak value less than 1% of that on the adiabatic ground state V_1 (see fig. S7B). By integrating the $|\psi|^2$ distribution in fig. S7B, we estimated that the population on the excited state is only about 0.053% of that on the ground state for the J = 0 partial wave. For partial waves with larger J value, the excited-state contribution becomes even smaller. The excited-state dynamics are different from those on the ground state, thus likely causing the small difference between the adiabatic + GP and the diabatic calculations. These quantitative analyses confirm that the excited state plays a very minor role in the H + HD → H_2 + D reaction at the collision energy of 2.77 eV, suggesting the reaction process occurs predominantly on the ground state and thus ensuring that the reaction at this collision energy can be adequately treated using adiabatic calculations on the ground state PES with GP.

It is intriguing that the GP effect on the H + HD → H_2 + D reaction can be seen so clearly in the forward scattering direction. According to the topological argument proposed by Althorpe and co-workers (*19*, *23*) for the H + H_2 reaction,

the GP effect should not be important for a reaction that occurs through a single pathway, because the GP only introduces a constant phase change to the wave functions of the pathway and thus will not influence the dynamics. In that context, there should be a second reaction pathway at this high collision energy that is markedly different from the normal reaction pathway. Using the topological argument, the GP effect can then change the DCS through interference between the two reaction pathways. By quasi-classical trajectory analysis, Althorpe and co-workers posited that one of the two pathways of the reaction proceeds through a single transition state (path 1), whereas the other proceeds through two transition states (path 2) (*19*). In the case of H + HD → H_2 + D, the GP effect is expected to manifest through the same interference between the two analogous reaction pathways (Fig. 4), and such an effect is more pronounced in the forward scattering direction of certain specific product quantum states at this collision energy.

Althorpe and co-workers also developed an approach to extract the contributions of the two reaction pathways on the basis of the topological argument (*19*, *23*, *24*, *26*). In this approach, the nuclear wave functions for path 1 and path 2 can be calculated by $\psi_1 = (\psi_{NGP} + \psi_{GP})/\sqrt{2}$ and $\psi_2 = (\psi_{NGP} - \psi_{GP})/\sqrt{2}$, respectively, where ψ_{NGP} and ψ_{GP} are the calculated wave functions without and with the GP effect, respectively. The scattering amplitudes from path 1 and path 2 can be expressed as $f_1(\theta) = [f_{NGP}(\theta) + f_{GP}(\theta)]/\sqrt{2}$ and $f_2(\theta) = [f_{NGP}(\theta) - f_{GP}(\theta)]/\sqrt{2}$, respectively. The square moduli of $f_1(\theta)$ and $f_2(\theta)$, $|f_1(\theta)|^2$ and $|f_2(\theta)|^2$, give the angular distributions of the product, i.e., the DCSs, for the individual paths. The total product DCS for the whole reaction can be described as

$$\sigma(\theta) = |f_1(\theta) + f_2(\theta)|^2 = |f_1(\theta)|^2 + |f_2(\theta)|^2 + f_1^*(\theta)f_2(\theta) + f_1(\theta)f_2^*(\theta)$$

whereas the interference between two pathways comes from the last two crossing terms. If the GPs introduced are different for the two pathways, then a difference in the DCS ensues. This explains the GP effect in the present case. The integral cross sections (ICSs) for the reaction via path 1 and path 2 can thus be calculated by integrating the corresponding DCS for all reaction product states.

Using this approach, we computed the total ICS for the reaction via path 1 and path 2 for collision energy up to 4 eV (Fig. 5A). At collision energies below 1.5 eV, the H + HD → H_2 + D reaction proceeds almost completely through path 1, which is the typical abstraction reaction pathway. As a result, the interference between products from path 1 and path 2 is negligible at low collision energy, and thus, the GP does not influence the dynamics of the reaction. However, as shown in Fig. 5A, at collision energies above 1.5 eV, the contribution from path 2 becomes increasingly important as the collision energy increases, even though the overall contribution from path 2 is still small at the collision energy

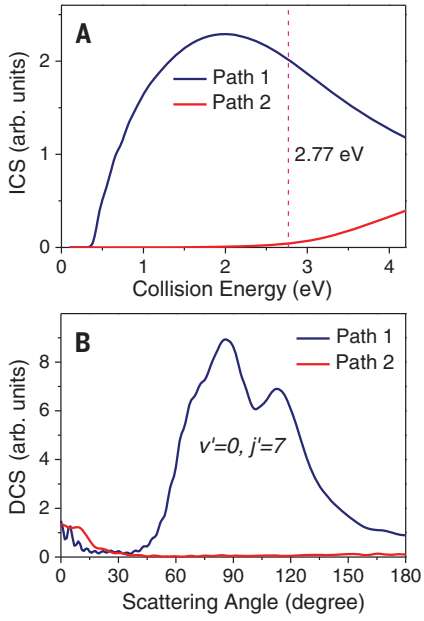


Fig. 5. Relative ICSs and DCSs from path 1 and path 2. (**A**) Calculated reactive ICSs as a function of collision energy for the H + HD ⟶ H_2 + D reaction proceeding through either path 1 or path 2; the ICS of path 2 is only 2.3% of that of path 1 at a collision energy of 2.77 eV. (**B**) Calculated DCS for the H_2 (v' = 0, j' = 7) product at a collision energy of 2.77 eV from path 1 and path 2. In the forward scattering direction, the DCSs from path 1 and 2 have comparable amplitudes, thus causing strong interference between the two paths.

of 2.77 eV, accounting for only ~2.3% of the total product.

To explore why the GP effect is so pronounced in the forward scattering direction, we calculated the DCS for the H_2 (v' = 0, j' = 7) product from path 1 and path 2 at the collision energy of 2.77 eV. The calculations show that the two reaction paths exhibit very different angular distributions (Fig. 5B), as in the H + H_2 reaction (*23*). Path 1 leads to predominantly sideways-scattered products with relatively small amplitude in the forward and backward scattering directions, whereas path 2 leads mainly to forward scattering. Coincidentally, the forward scattering amplitude for the two paths of this reaction are comparable (see Fig. 5B). With different phases introduced by the GP effect to the two paths, their comparable scattering amplitudes make the GP effect more pronounced in the forward scattering direction. In the backward and sideways scattering direction, it would be much harder to see the GP effect because of the dominance of path 1 over path 2. The detailed mechanism of path 2 through two transition states should be very similar to that of the H + H_2 → H_2 + H reaction (*46*). Here, we want to emphasize that the GP is introduced theoretically for accurate treatment of the molecular system in the adiabatic picture; thus, the GP effect on the dynamics and its observation should be discussed strictly in the context of the adiabatic theory.

This work demonstrates that fine angularly resolved scattering structure in the forward direction for reaction products in specific quantum states is an extremely sensitive probe of the GP effect in quantum dynamics of chemical reactions in the adiabatic picture.

REFERENCES AND NOTES

1. S. Pancharatnam, *Proc. Indian Acad. Sci. A* **44**, 247–262 (1956).
2. H. C. Longuet-Higgins, U. Öpik, M. H. L. Pryce, R. A. Sack, *Proc. R. Soc. London Ser. A* **244**, 1–16 (1958).
3. M. V. Berry, *Proc. R. Soc. London Ser. A* **392**, 45–57 (1984).
4. D. Xiao, M. C. Chang, Q. Niu, *Rev. Mod. Phys.* **82**, 1959–2007 (2010).
5. N. Nagaosa, J. Sinova, S. Onoda, A. H. MacDonald, N. P. Ong, *Rev. Mod. Phys.* **82**, 1539–1592 (2010).
6. J. Sinova, S. O. Valenzuela, J. Wunderlich, C. H. Back, T. Jungwirth, *Rev. Mod. Phys.* **87**, 1213–1260 (2015).
7. W. Domcke, D. R. Yarkony, H. Köppel, Eds., *Conical Intersections: Electronic Structure, Dynamics and Spectroscopy* (World Scientific, 2003).
8. H. von Busch *et al.*, *Phys. Rev. Lett.* **81**, 4584–4587 (1998).
9. M. G. D. Nix, A. L. Devine, R. N. Dixon, M. N. R. Ashfold, *Chem. Phys. Lett.* **463**, 305–308 (2008).
10. C. Xie *et al.*, *J. Am. Chem. Soc.* **138**, 7828–7831 (2016).
11. A. J. C. Varandas, F. B. Brown, C. A. Mead, D. G. Truhlar, N. C. Blais, *J. Chem. Phys.* **86**, 6258–6269 (1987).
12. C. A. Mead, D. G. Truhlar, *J. Chem. Phys.* **70**, 2284–2296 (1979).
13. J. Z. H. Zhang, S.-I. Chu, W. H. Miller, *J. Chem. Phys.* **88**, 6233–6239 (1988).
14. J. Z. H. Zhang, W. H. Miller, *Chem. Phys. Lett.* **153**, 465–470 (1988).
15. B. Lepetit, A. Kuppermann, *Chem. Phys. Lett.* **166**, 581–588 (1990).
16. Y.-S. M. Wu, A. Kuppermann, B. Lepetit, *Chem. Phys. Lett.* **186**, 319–328 (1991).
17. M. P. de Miranda, D. C. Clary, J. F. Castillo, D. E. Manolopoulos, *J. Chem. Phys.* **108**, 3142–3153 (1998).
18. B. K. Kendrick, *J. Chem. Phys.* **112**, 5679–5704 (2000).
19. J. C. Juanes-Marcos, S. C. Althorpe, E. Wrede, *Science* **309**, 1227–1230 (2005).
20. E. Wrede, L. Schnieder, *J. Chem. Phys.* **107**, 786–790 (1997).
21. E. Wrede *et al.*, *J. Chem. Phys.* **110**, 9971–9981 (1999).
22. B. Kendrick, *J. Phys. Chem. A* **107**, 6739–6756 (2003).
23. J. C. Juanes-Marcos, S. C. Althorpe, *J. Chem. Phys.* **122**, 204324 (2005).
24. F. Bouakline, S. C. Althorpe, D. Peláez Ruiz, *J. Chem. Phys.* **128**, 124322 (2008).
25. B. K. Kendrick, J. Hazra, N. Balakrishnan, *Phys. Rev. Lett.* **115**, 153201 (2015).
26. S. C. Althorpe, *J. Chem. Phys.* **124**, 084105 (2006).
27. B. R. Strazisar, C. Lin, H. F. Davis, *Science* **290**, 958–961 (2000).
28. M. Qiu *et al.*, *Science* **311**, 1440–1443 (2006).
29. T. Wang *et al.*, *Science* **342**, 1499–1502 (2013).
30. T. Yang *et al.*, *Science* **347**, 60–63 (2015).
31. L. Schnieder, K. Seekamp-Rahn, E. Wrede, K. H. Welge, *J. Chem. Phys.* **107**, 6175–6195 (1997).
32. S. A. Harich *et al.*, *Nature* **419**, 281–284 (2002).
33. D. Dai *et al.*, *Science* **300**, 1730–1734 (2003).
34. J. Jankunas, M. Sneha, R. N. Zare, F. Bouakline, S. C. Althorpe, *J. Chem. Phys.* **139**, 144316 (2013).
35. H. Gao, M. Sneha, F. Bouakline, S. C. Althorpe, R. N. Zare, *J. Phys. Chem. A* **119**, 12036–12042 (2015).
36. W. Hu, G. C. Schatz, *J. Chem. Phys.* **125**, 132301 (2006).
37. D. Yuan *et al.*, *Nat. Chem.* **10**, 653–658 (2018).
38. A. T. J. B. Eppink, D. H. Parker, *Rev. Sci. Instrum.* **68**, 3477–3484 (1997).
39. S. L. Mielke, B. C. Garrett, K. A. Peterson, *J. Chem. Phys.* **116**, 4142–4161 (2002).
40. Z. Sun, H. Guo, D. H. Zhang, *J. Chem. Phys.* **132**, 084112 (2010).
41. H. Lischka *et al.*, *Phys. Chem. Chem. Phys.* **3**, 664–673 (2001).
42. T. H. Dunning Jr., *J. Chem. Phys.* **90**, 1007–1023 (1989).

43. I. E. Lagaris, A. Likas, D. I. Fotiadis, *IEEE Trans. Neural Netw.* **9**, 987–1000 (1998).
44. H. J. Werner *et al*., MOLPRO, version 2012.1. A package of ab initio programs (2012).
45. R. T. Pack, G. A. Parker, *J. Chem. Phys.* **87**, 3888–3921 (1987).
46. F. Bouakline, S. C. Althorpe, P. Larregaray, L. Bonnet, *Mol. Phys.* **108**, 969–980 (2010).

ACKNOWLEDGMENTS

We thank J. Sang and S. Wang for their help with the experiments. **Funding:** This work was supported by National Natural Science Foundation of China (no. 21688102, no. 21127902, no. 21433009), Chinese Academy of Sciences (grant no. XDB 17010000), and Ministry of Science and Technology. **Author contributions:** D.Y., W.C., C.L., Y.T., T.X., X.W., and X.Y. performed the crossed beams experiments and data analysis. Y.G., H.Z., Z.S., and D.H.Z. performed the quantum dynamics calculations and data analysis. X.W., Z.S., D.H.Z., and X.Y. designed the research. X.W., Z.S., D.H.Z., and X.Y. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the supplementary data file.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/1289/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S9
Tables S1 and S2
References (*47*–*55*)
Data S1

18 August 2018; accepted 2 November 2018
10.1126/science.aav1356

RADIOCARBON

Atmospheric $^{14}C/^{12}C$ changes during the last glacial period from Hulu Cave

Hai Cheng[1,2]*, R. Lawrence Edwards[2]*, John Southon[3], Katsumi Matsumoto[2], Joshua M. Feinberg[2,4], Ashish Sinha[5], Weijian Zhou[6], Hanying Li[1], Xianglei Li[1], Yao Xu[1], Shitao Chen[7], Ming Tan[8], Quan Wang[7], Yongjin Wang[7], Youfeng Ning[1]

Paired measurements of $^{14}C/^{12}C$ and ^{230}Th ages from two Hulu Cave stalagmites complete a precise record of atmospheric ^{14}C covering the full range of the ^{14}C dating method (~54,000 years). Over the last glacial period, atmospheric $^{14}C/^{12}C$ ranges from values similar to modern values to values 1.70 times higher (42,000 to 39,000 years ago). The latter correspond to ^{14}C ages 5200 years less than calibrated ages and correlate with the Laschamp geomagnetic excursion followed by Heinrich Stadial 4. Millennial-scale variations are largely attributable to Earth's magnetic field changes and in part to climate-related changes in the oceanic carbon cycle. A progressive shift to lower $^{14}C/^{12}C$ values between 25,000 and 11,000 years ago is likely related, in part, to progressively increasing ocean ventilation rates.

Libby pioneered the ^{14}C dating method (*1*), which revolutionized a number of scientific disciplines, most notably archeology and climatology. However, variations in atmospheric ^{14}C, likely caused by changes in the shielding of cosmic rays induced by the Earth's and Sun's magnetic fields and/or the redistribution of ^{14}C among different carbon reservoirs, were soon recognized (*2*). These changes necessitate the calibration of ^{14}C ages against a calendar time scale. A precise and accurate ^{14}C calibration is considered the Holy Grail of radiocarbon dating.

Our ability to calibrate the ^{14}C time scale has been limited by our ability to establish the absolute age of a material that contains information about atmospheric $^{14}C/^{12}C$. By the late 1980s, the most recent portion of the ^{14}C time scale [last ~10 thousand years (ka)] was calibrated extremely precisely using dendrochronology. The development of mass spectrometric ^{230}Th dating methods (*3*) and their continued refinement (*4*) opened up the possibility of extending the calibration much deeper in time, led to the first large extension of the calibration well back into the Pleistocene (*5*), and ultimately has led to the current contribution. However, the ^{230}Th dating approach has its own constraints. Corals, which are good materials for ^{230}Th dating, do not accumulate continuously over thousands of years and are difficult to collect since those in the time range of interest are now largely submerged. Stalagmites, which can be excellent choices for ^{230}Th dating, typically contain a significant fraction of carbon ultimately derived from limestone bedrock, which is essentially ^{14}C-free. Stalagmite-based calibrations must therefore correct for a dead carbon fraction (DCF), which can be large and variable and is typically the main hurdle in such efforts (*6*, *7*).

Southon *et al.* (*8*) demonstrated that the DCF in one Hulu Cave (32°30′N, 119°10′E) stalagmite, H82, was unusually small and stable, allowing a precise and accurate ^{14}C calibration in the 26.8 to 10.6 ka B.P. (before the present; "present" is 1950 CE) interval (fig. S1). Here, we show that older Hulu Cave stalagmites, MSD and MSL, have similarly low and stable DCFs (Figs. 1 and 2), which allow for precise and accurate ^{14}C calibration for the remainder of the ^{14}C time scale back to ~54 ka B.P.

All three Hulu stalagmites record climatic conditions in their oxygen isotopic compositions (*9*, *10*), including Asian monsoon equivalents of the stadial and interstadial events recorded in Greenland and the Heinrich Stadials recorded in North Atlantic sediments. Thus, we are able to compare our final $^{14}C/^{12}C$ record to the major climate events of the last glacial period, with negligible stratigraphic uncertainty.

Here, we present ~300 pairs of ^{14}C and ^{230}Th dates from MSD (51 to 18.5 ka B.P.) and MSL (analyzed between 54 and 36 ka B.P.), extending the ^{14}C record back to 54 ka B.P. (Fig. 1, figs. S2 to S5, and tables S1 and S2). Temporal resolution per pair is ~170 years. We drilled sequential powders for ^{230}Th dating, leaving a ridge of solid calcite behind for ^{14}C dating (figs. S2 and S3). This procedure avoids use of a powdered sample for ^{14}C analysis, which can lead to ^{14}C contamination (*8*). Methods are described in the supplementary materials (*11*). The large overlaps in ages between MSD and MSL (15 ka) and between H82 (*8*) and MSD (8 ka) (Figs. 1 and 2 and figs. S4 and S5) allow us to test for precision, accuracy, differential contamination/diagenesis, and differential changes in the DCF.

Through comparison to the dendrochronology record, DCF in H82 is low and constant within tight bounds, even across major climate boundaries, equivalent to a ^{14}C age offset of merely 450 ± 70 years (*8*) (fig. S1). With the same DCF correction for MSD and MSL, we observe strong agreement between the overlapping portions of $\Delta^{14}C$ records from MSD and H82 as well as for MSD and MSL (Figs. 1 and 2 and fig. S6). Although we cannot rule out scenarios where, for example, the DFC shifts similarly in pairs of stalagmites, the replication among stalagmites is consistent with small DCF for all three speleothems and DCF stability within tight bounds over the period of our extended record (fig. S6). We therefore adopt the H82 DCF correction of 450 ±70 years for the entire record. ^{14}C data from modern dripwaters (figs. S7 to S9) suggest that the soil above portions of the cave is characterized by open system conditions, which together with an unusual sandstone ceiling above the three samples provide a possible explanation for the low DCF that we infer for the three stalagmites (*11, 12*).

A number of arguments support the accuracy of the record. The younger portion of the H82 record agrees with the dendrochronology record (*8*). The overlapping portions of the three stalagmite records are internally consistent. There is agreement between one of the highest values in our record ($\Delta^{14}C$ = 700‰, at ~39.85 ka B.P.) with a precisely and carefully determined independent data point based upon wood associated with the Campanian Ignimbrite and precise Ar-Ar dating (*13*). Finally, two floating dendrochronology sections (*14*, *15*) can be placed on the Hulu calibration in such a way that overall trends and finer-scale features match the Hulu curve (*11*) (Figs. 1 and 2 and fig. S10). We should point out, however, that others have previously proposed a placement later by ~1 ka for one of these floating chronologies (*16*); see the supplementary materials for a discussion of this issue.

Considering the full record, there is a general correspondence with the latest IntCal compilation (*17*) (Figs. 1 and 2) within fairly large uncertainties, confirming the general validity of the compilation. However, for the portion older than 30 ka B.P., clear differences emerge. The Hulu record has less uncertainty and resolves previously unknown fine-scale structure. Between 33.5 and 42.5 ka B.P., the Hulu record indicates larger offsets between ^{230}Th ages and ^{14}C ages than IntCal13, with offsets between the records as high as 1 ka, corresponding to a higher $\Delta^{14}C$ by as much as 170‰ as recorded at Hulu. Conversely, from 42.5 ka B.P. to the end of the IntCal curve at 50 ka B.P., the Hulu record indicates smaller offsets between ^{230}Th and ^{14}C ages, by ~1 ka, which corresponds to ~140‰ lower $\Delta^{14}C$. From 50 to 54 ka B.P., the Hulu curve indicates similar though nominally higher $\Delta^{14}C$ than during the subsequent few millennia. Another notable difference is the sharper and higher amplitude increase in $\Delta^{14}C$ around 42.5 ka B.P. A notable similarity is the lack of a prominent low $\Delta^{14}C$

[1]Institute of Global Environmental Change, Xi'an Jiaotong University, China. [2]Department of Earth Sciences, University of Minnesota, Minneapolis, MN, USA. [3]Department of Earth System Science, University of California, Irvine, CA, USA. [4]Institute for Rock Magnetism, University of Minnesota, Minneapolis, MN, USA. [5]Department of Earth Science, California State University Dominguez Hills, Carson, CA, USA. [6]Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, China. [7]College of Geography Science, Nanjing Normal University, Nanjing, China. [8]Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing, China.
*Corresponding author. Email: cheng021@xjtu.edu.cn (H.C.); edwar001@umn.edu (R.L.E.)

excursion around 31 ka B.P. This low, present in Cariaco sediment and Bahamas speleothem datasets (*7*, *18*), was omitted from IntCal13 because of its absence from the Lake Suigetsu record (*19*). The Hulu data support this omission.

^{14}C ages are generally less than calendar ages throughout the full record, reaching a maximum offset of ~5200 years between ~39.3 and ~40.8 ka B.P. (Fig. 1). The offset is largely due to higher atmospheric Δ^{14}C, although there is also a progressive offset of 2.83% of the age due to the use of the Libby half-life in calculating the ^{14}C age. Between 54 and 43 ka B.P., Δ^{14}C values range between 0 and 300‰, then increase sharply to values exceeding 600‰ by 42 ka B.P. (Fig. 2). High values continue until 38.8 ka B.P., reaching the highest values in the full record of 700‰ at 40.8 and 39.3 ka B.P. Between 38.8 and 38.0 ka B.P., Δ^{14}C decreases sharply to values around 500%. Between 38.0 and 25.0 ka B.P., Δ^{14}C values exhibit millennial-scale variability with highs around 600% and lows around 400%. Notable is a relative high of about 600% at 33.8 ka B.P. From 25.0 ka B.P. to the mid-19th century (as previously known), Δ^{14}C values gradually diminish from around 500% to 0, with significant changes in slope between 16 and 11 ka B.P.

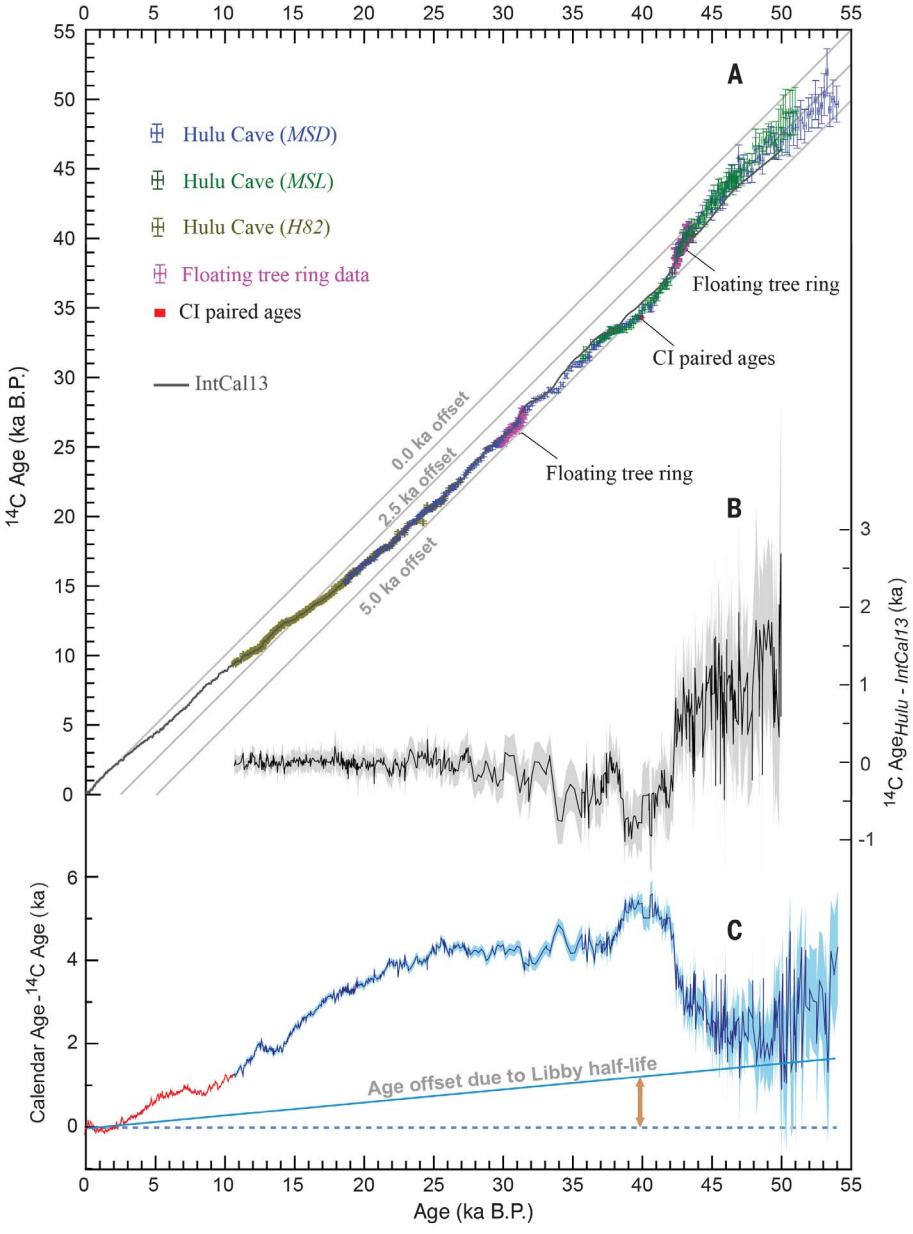


Fig. 1. Hulu speleothem ^{14}C versus ^{230}Th ages and comparison between Hulu and IntCal13 ^{14}C ages. (**A**) Hulu [olive-brown, H82 (*8*); blue, MSD, and green, MSL (this study), and IntCal13 ^{14}C (*17*)] vs. ^{230}Th ages. ^{14}C error bars are 1σ. For clarity, uncertainties in IntCal13 are not shown. The floating tree ring Δ^{14}C datasets (purple) (*14*, *15*) are tuned to the Hulu ^{14}C record (*11*). The red square (1σ) is the independent data point based on ^{14}C measurements on wood associated with the Ar-Ar dated Campanian Ignimbrite (*13*). (**B**) ^{14}C age difference (black) between Hulu dataset and IntCal13 (*17*). The gray envelope shows the uncertainty (1σ). Hulu ^{14}C ages are corrected for the DCF (450 ± 70 years) (*8*). (**C**) Calendar age minus IntCal13 (red)/Hulu (blue) ^{14}C age. The light blue envelope shows the uncertainty (1σ). The three Hulu sample datasets replicate over contemporary growth periods. Hulu Cave ^{14}C data are consistent with IntCal13 between ~10.6 and 33.3 ka B.P. but lower in ^{14}C ages between ~33.3 and 42 ka B.P. and higher between 42 and 53 ka B.P.

The new data provide critical constraints on the causes of changes in Δ^{14}C during the last 54 ka. The millennial-scale pattern of Δ^{14}C variations (Fig. 3) has similarities to the geomagnetic record (Virtual Axial Dipole Moment data) (*20*), suggesting that changes in shielding of cosmic rays by the geomagnetic field are responsible for much of the millennial-scale variation in Δ^{14}C. Of note is the coincidence within tight age uncertainties between the abrupt increase in Hulu Δ^{14}C and the onset of the Laschamp magnetic excursion at ~42.3 ka B.P. (*21*), as well as between the period of weakest geomagnetic field during the Laschamp (~41.1 ka B.P.) (*21*), which correlates with the highest Δ^{14}C values over the past 54 ka. This suggests that the Laschamp is responsible for both of these features. Additionally, a second prominent peak in the Hulu record at ~34 ka B.P. is consistent with the timing of the Mono Lake excursion (*22*), suggesting that this excursion is responsible for the Δ^{14}C peak (Fig. 3).

We estimated the component of Δ^{14}C variability caused by geomagnetic field changes by using a magnetic record (*20*), a cosmogenic production model (*23*), and the MESMO-2 Earth system model (*24*). The output simulates that component of atmospheric Δ^{14}C variability caused by geomagnetic field changes alone (*11*) (Fig. 3C). We subtracted this model curve from the observed Hulu Δ^{14}C record to obtain a model–observation residual curve ($\Delta\Delta^{14}$C), which shows the component of the observed variability not captured by our model, likely due to some combination of uncertainties in the input magnetic field data, inaccuracies in the model itself, solar modulation of production, and changes in the carbon cycle (Fig. 3E). We cannot use this residual as a quantitative target curve for, say, a model with a changing carbon cycle, as there are nonlinearities in the overall problem (*25*). Nevertheless, we consider the residual curve useful for the remaining discussion, because it guides us to the magnitude and direction of observation–model differences.

The residual is characterized by a series of millennial-scale events during the last glacial period (Fig. 4B). Given uncertainties, we have not attempted to assign a one-to-one correspondence between climate events and features in the residual trace. However, we highlight two cases where temporal constraints are robust and

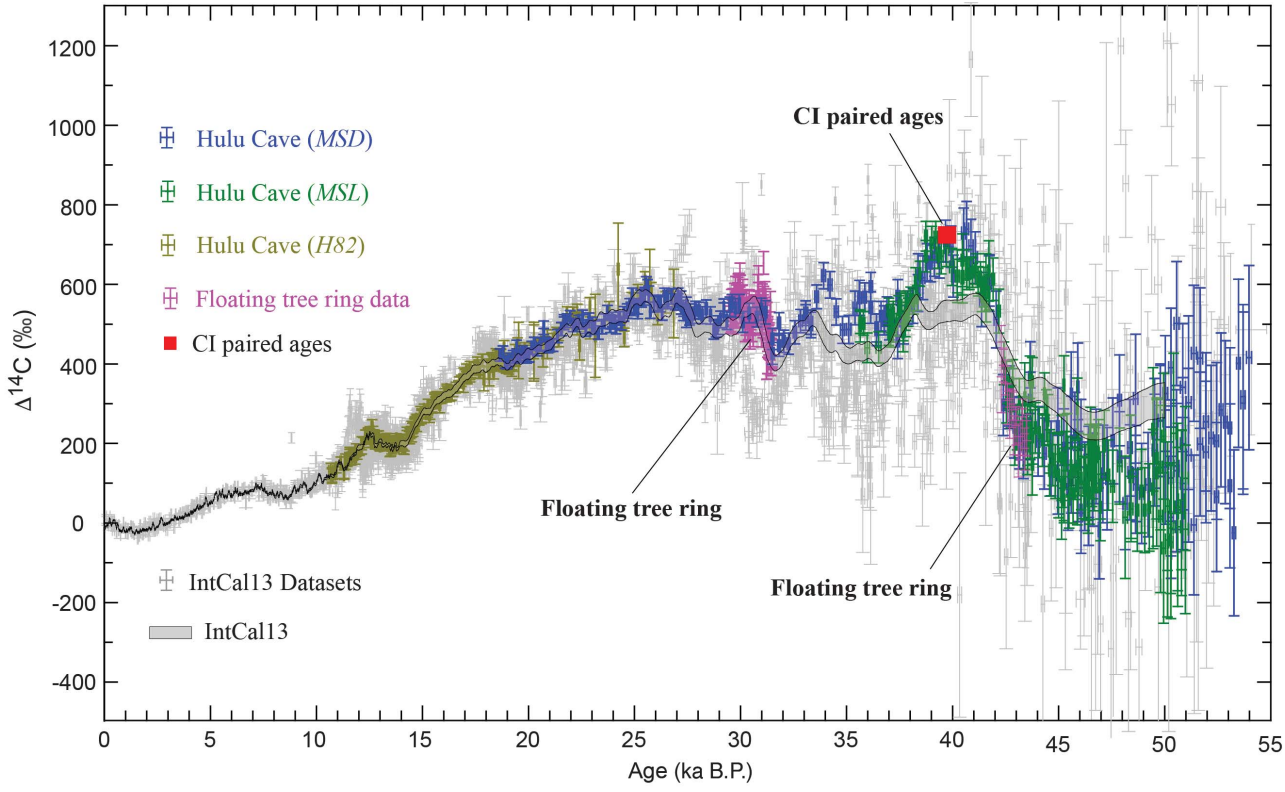


Fig. 2. Comparison of Hulu $\Delta^{14}C$ data with IntCal13. Hulu $\Delta^{14}C$ data are shown with error bars with the same color codes as in Fig. 1. IntCal13 and its dataset (*17*) are shown in the gray envelope and gray bars. ^{14}C error bars are 1σ. Hulu data overlap with IntCal13 between ~10.6 and 33.3 ka B.P.; however, there are substantial offsets, particularly before 30 ka B.P., and the Hulu record exhibits substantial previously unknown millennial-scale structure. The purple error bars and red square are the floating tree ring series and Campanian Ignimbrite data, as in Fig. 1.

where the trace shows a prominent feature, the Younger Dryas (YD) and Heinrich Stadial 4 (HS 4). In both cases, residual highs correlate with cold anomalies in the North Atlantic region. For the YD, this observation confirms earlier work (*26*–*29*). These studies all explained the relatively high $\Delta^{14}C$ by invoking carbon cycle changes associated with climate change with, in one case (*29*), an additional contribution from solar modulation during the early YD. For HS 4, temporal constraints place the end of the Laschamp (*16*, *21*) ~1 ka well before the prominent residual peak that correlates with HS 4. Even that long after the end of the Laschamp, one would expect high atmospheric ^{14}C, because the e-folding time for reaching isotopic steady state after a production change is on the order of thousands of years, the time scale of deep ocean ventilation. However, the time scale for the initial significant diminution of atmospheric ^{14}C following a production drop is a few hundred years, a time scale tied to reaching isotopic steady state with the upper portion of the ocean. Our model captures this, as evidenced by the few-hundred-year difference between production shift (Fig. 3B) and $\Delta^{14}C$ response (Fig. 3C) for numerous production changes. Since $\Delta^{14}C$ does not fall in the centuries after the Laschamp but instead rises slightly to a high value that correlates with HS 4, we conclude that another factor besides magnetic field change has contributed to these high values, likely carbon cycle changes associated with climate change.

Given the general character of the millennial-scale variability in the residual trace, it is plausible that the relationships that we observe for the YD and HS 4 are more general features of the last glacial period climate and carbon cycle. The YD, HSs, and Greenland stadials (GSs) correspond to weak modes in the Atlantic Meridional Overturning Circulation (AMOC), as inferred from the $^{231}Pa/^{230}Th$ record (*30*). A weak mode may increase atmospheric ^{14}C due to diminished flux of ^{14}C to the intermediate and/or deep ocean, as supported by observed increases in radiocarbon-based ventilation ages during HS 1 and the YD in the western equatorial Atlantic (*31*). Regardless of the specific mechanisms, there is clear evidence at the millennial scale for elevated $\Delta^{14}C$ at specific cold times in the North Atlantic, perhaps associated with AMOC slowdown.

We now consider the long-term gradual lowering of $\Delta^{14}C$, from ~500‰ 25 ka B.P. to ~150‰ 11 ka B.P. Bard *et al.* (*5*) attributed much of the decline to steady increase in Earth's magnetic field, with some (100 to 150‰) plausibly caused by carbon cycle changes. Köhler *et al.* (*25*) reached similar conclusions. Notable was their use of the ice core ^{10}Be record to predict production-related changes in $\Delta^{14}C$. This strategy takes into account production changes caused both by the terrestrial magnetic field and by solar modulation. They reached a similar conclusion as Bard *et al.* (*5*), i.e., that production changes could not explain the full $\Delta^{14}C$ shift over this interval and that carbon cycle changes could account for up to 100‰ of the shift. Our work confirms some of these conclusions, as our residual trace shows a significant decline after accounting for magnetic field-related production changes.

The broad lowering of $\Delta^{14}C$ throughout this interval could plausibly result from progressively increasing ocean ventilation. All other factors being equal, the shorter the mixing time, the less time for ^{14}C to decay, the more ^{14}C in deep waters and, by mass balance, the lower the $\Delta^{14}C$ of the atmosphere. Presuming an average deep water age of 1000 years at 11 ka B.P. and a 60:1 ratio of deep water to atmospheric carbon, the lowering of atmospheric $\Delta^{14}C$ over this time period can be explained by a progressive shift in deep water age from about 3000 years at 25 ka B.P. to the assumed 1000-year value at 11 ka B.P.

There is some support for the inference of increasing ventilation with time, as observations indicate that the deep Southern Ocean and South Pacific were poorly ventilated at the last glacial maximum (*32*–*34*). Deep ocean $\Delta^{14}C$ data for times since the last glacial maximum (*35*) do not clearly resolve pre-Holocene from Holocene ventilation ages, but they also do not preclude large pre-Holocene ventilation ages. Thus, while it is likely that deep ocean ventilation

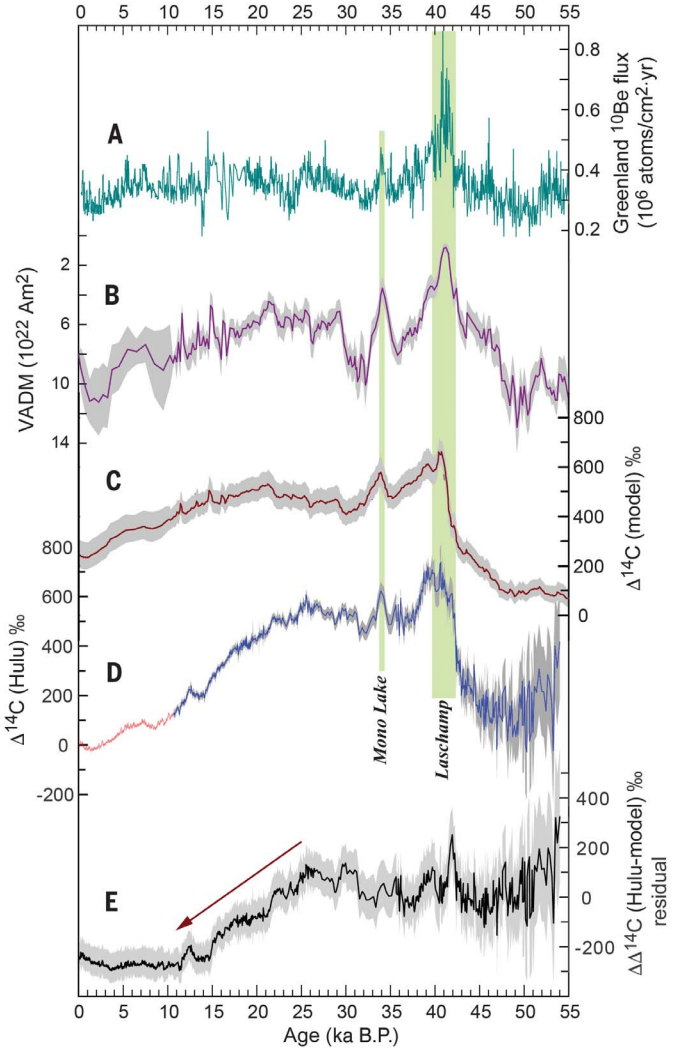


Fig. 3. Comparison of ^{10}Be flux, geomagnetic field, model and Hulu ^{14}C data. (**A**) Greenland ^{10}Be flux (*36*). (**B**) Stacked geomagnetic field (gray, 1σ envelope) (*20*). (**C**) The model Δ^{14}C record (*11*) (gray, 1σ envelope) based on ^{14}C production inferred from the geomagnetic field (*20*). (**D**) Blue and red envelopes (1σ) are composite Hulu (10.6 to 54.0 ka B.P.) and IntCal13 (0 to 10.6 ka B.P.) Δ^{14}C data, respectively. (**E**) The $\Delta\Delta^{14}$C is the residual obtained by subtracting the model Δ^{14}C result from the Hulu/Intcal13 Δ^{14}C data. The gray envelope shows the uncertainty from Hulu data and model uncertainties (1σ). Two vertical bars show the Laschamp and Mono Lake excursions. The arrow indicates the large decline in Δ^{14}C from ~25 to 11 ka B.P. See also fig. S10.

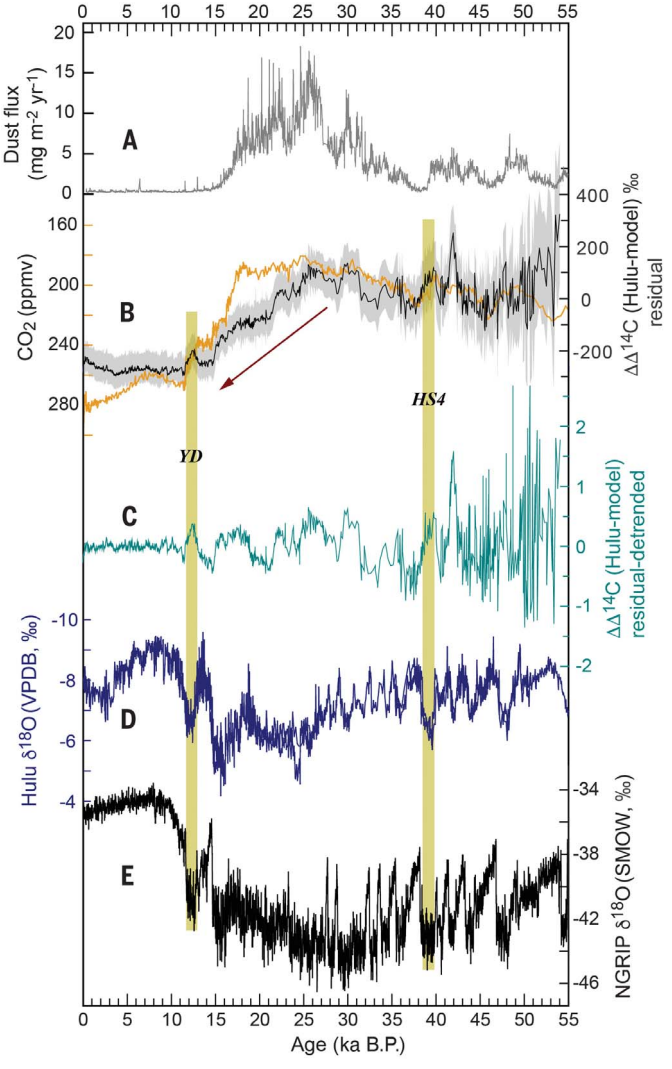


Fig. 4. Comparison of the $\Delta\Delta^{14}$C record with other climate proxy records. (**A**) Antarctic ice core dust flux record (EDC) (*37*). (**B**) The $\Delta\Delta^{14}$C record (the residual as determined for Fig. 3E) and composite atmospheric CO_2 record (yellow) (*38*). (**C**) Detrended $\Delta\Delta^{14}$C record (*11*). (**D**) The Hulu δ^{18}O record (*10*). (**E**) Greenland ice core δ^{18}O record (NGRIP) (*39*). Vertical light yellow bars indicate HS 4 and YD. The arrow shows the Δ^{14}C trend, as in Fig. 3.

change accounts for a portion of the residual 25 to 11 ka B.P. Δ^{14}C drop, it is still not clear whether it can account for the full drop. Further work is needed to close the loop on this critical issue.

REFERENCES AND NOTES

1. W. F. Libby, *Radiocarbon Dating* (Univ. Chicago Press, 1952).
2. H. de Vries, *Proc. Koninkl. Nederl. Akad. Wetenschappen B* **61**, 1 (1958).
3. R. Lawrence Edwards, J. H. Chen, G. J. Wasserburg, *Earth Planet. Sci. Lett.* **81**, 175–192 (1987).
4. H. Cheng *et al.*, *Earth Planet. Sci. Lett.* **371–372**, 82–91 (2013).
5. E. Bard, B. Hamelin, R. G. Fairbanks, A. Zindler, *Nature* **345**, 405–410 (1990).
6. J. W. Beck *et al.*, *Science* **292**, 2453–2458 (2001).
7. D. L. Hoffmann *et al.*, *Earth Planet. Sci. Lett.* **289**, 1–10 (2010).
8. J. Southon, A. L. Noronha, H. Cheng, R. Lawrence Edwards, Y. J. Wang, *Quat. Sci. Rev.* **33**, 32–41 (2012).
9. Y. J. Wang *et al.*, *Science* **294**, 2345–2348 (2001).
10. H. Cheng *et al.*, *Nature* **534**, 640–646 (2016).
11. Materials and methods are available as supplementary materials.
12. A. L. Noronha *et al.*, *Quat. Sci. Rev.* **127**, 37–47 (2015).
13. B. Giaccio, I. Hajdas, R. Isaia, A. Deino, S. Nomade, *Sci. Rep.* **7**, 45940 (2017).
14. C. S. M. Turney *et al.*, *Quat. Sci. Rev.* **29**, 3677–3682 (2010).
15. C. S. M. Turney *et al.*, *Quat. Sci. Rev.* **137**, 126–134 (2016).
16. R. Muscheler, F. Adolphi, A. Svensson, *Earth Planet. Sci. Lett.* **394**, 209–215 (2014).
17. P. J. Reimer *et al.*, *Radiocarbon* **55**, 1869–1887 (2013).
18. K. A. Hughen, J. Southon, S. Lehman, C. Bertrand, J. Turnbull, *Quat. Sci. Rev.* **25**, 3216–3227 (2006).
19. C. Bronk Ramsey *et al.*, *Science* **338**, 370–374 (2012).
20. C. Laj, H. Guillou, C. Kissel, *Earth Planet. Sci. Lett.* **387**, 184–197 (2014).
21. I. Lascu, J. M. Feinberg, J. A. Dorale, H. Cheng, R. Lawrence Edwards, *Geology* **44**, 139–142 (2016).
22. S. Lund, L. Benson, R. Negrini, J. Liddicoat, S. Mensing, *Earth Planet. Sci. Lett.* **458**, 120–129 (2017).
23. G. A. Kovaltsov, A. Mishev, I. G. Usoskin, *Earth Planet. Sci. Lett.* **337–338**, 114–120 (2012).
24. K. Matsumoto, K. Tokos, A. Huston, H. Joy-Warren, *Geosci. Model Dev.* **6**, 477–494 (2013).
25. P. Köhler, R. Muscheler, H. Fischer, *Geochem. Geophys. Geosyst.* **7**, 2005GC001228 (2006).
26. R. Lawrence Edwards *et al.*, *Science* **260**, 962–968 (1993).

27. K. A. Hughen, J. R. Southon, S. J. Lehman, J. T. Overpeck, *Science* **290**, 1951–1955 (2000).
28. Y. T. Goslar *et al*., *Nature* **377**, 414–417 (1995).
29. R. Muscheler *et al*., *Nat. Geosci.* **1**, 263–267 (2008).
30. L. G. Henry *et al*., *Science* **353**, 470–474 (2016).
31. E. Freeman *et al*., *Earth Planet. Sci. Lett.* **424**, 237–244 (2015).
32. L. C. Skinner, S. Fallon, C. Waelbroeck, E. Michel, S. Barker, *Science* **328**, 1147–1151 (2010).
33. T. A. Ronge *et al*., *Nat. Commun.* **7**, 11487 (2016).
34. E. Freeman, L. C. Skinner, C. Waelbroeck, D. Hodell, *Nat. Commun.* **7**, 11998 (2016).
35. N. Zhao, O. Marchal, L. Keigwin, D. Amrhein, G. Gebbie, *Paleoceanogr. Paleoclimatol.* **33**, 128–151 (2018).
36. R. Muscheler *et al*., *Earth Planet. Sci. Lett.* **219**, 325–340 (2004).
37. F. Lambert *et al*., *Nature* **452**, 616–619 (2008).
38. B. Bereiter *et al*., *Geophys. Res. Lett.* **42**, 542–549 (2015).
39. A. Svensson *et al*., *Clim. Past* **4**, 47–57 (2008).

ACKNOWLEDGMENTS

We thank W. Broecker for inspiring and supporting this work over decades. K. Tokos ran numerical model experiments using the facilities of the UM Minnesota Supercomputing Institute following the experiment design by K.M. **Funding:** This work was supported by China NSFC grants 41888101, 41731174, 4157020432, and 41561144003; U.S. NSF grants 0502535, 1103404, 0823554, 1003690, 1137693, 1337693, and 1702816; and the Comer Science and Education Foundation. **Author contributions:** H.C., R.L.E., and J.S. proposed and directed the research. R.L.E. led the writing and revision of the manuscript. H.C. and J.S. performed ^{230}Th dating and ^{14}C analysis of the speleothem samples, respectively. K.M. designed and performed the model simulation. W.J.Z. performed the ^{14}C analysis of drip-water. H.Y.L., X.L.L., Y.X., Y.F.N., A.S., J.M.F., and H.C. performed subsampling work and data analyses. S.T.C., M.T., Q.W., Y.J.W., and H.C. collected samples and performed cave monitoring and fieldwork. All authors discussed the results and commented on the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/1293/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S10
Tables S1 and S2
References (*40*–*55*)

4 May 2018; accepted 2 November 2018
10.1126/science.aau0747

ARCHAEOLOGY

1.9-million- and 2.4-million-year-old artifacts and stone tool–cutmarked bones from Ain Boucherit, Algeria

Mohamed Sahnouni[1,2,3]*, Josep M. Parés[1], Mathieu Duval[4,1], Isabel Cáceres[5,6], Zoheir Harichane[2,7], Jan van der Made[8], Alfredo Pérez-González[1], Salah Abdessadok[9,2], Nadia Kandi[10], Abdelkader Derradji[2,11], Mohamed Medig[11], Kamel Boulaghraif[2,12], Sileshi Semaw[1,3]

East Africa has provided the earliest known evidence for Oldowan stone artifacts and hominin-induced stone tool cutmarks dated to ~2.6 million years (Ma) ago. The ~1.8-million-year-old stone artifacts from Ain Hanech (Algeria) were considered to represent the oldest archaeological materials in North Africa. Here we report older stone artifacts and cutmarked bones excavated from two nearby deposits at Ain Boucherit estimated to ~1.9 Ma ago, and the older to ~2.4 Ma ago. Hence, the Ain Boucherit evidence shows that ancestral hominins inhabited the Mediterranean fringe in northern Africa much earlier than previously thought. The evidence strongly argues for early dispersal of stone tool manufacture and use from East Africa or a possible multiple-origin scenario of stone technology in both East and North Africa.

The earliest archaeological evidence for the Oldowan and associated fossil bones with evidence of butchery is within the 2.6 million to 1.9 million years (Ma) ago time interval, primarily from East Africa (*1*–*7*). Most paleoanthropologists believe that early hominins dispersed into northern Africa much later (*8*). Continued research at Ain Hanech and El Kherba (Algeria) over the past two decades has expanded the geographic range and pushed back the evidence for hominin stone tool use and carnivory to ~1.8 Ma ago (*9*–*11*). We recently explored the nearby deposits at Ain Boucherit (Algeria) and report evidence of Oldowan stone tools and associated hominin-modified fossil bones from two distinct strata estimated to ~2.4 and ~1.9 Ma ago, respectively.

Ain Boucherit is an archaeological locality in the Ain Hanech research area in northeastern Algeria. The research area is in the Beni Fouda basin, one of the several intramontane sedimentary basins in the High Plateaus of eastern Algeria. The stone tools and associated fossil bones at Ain Boucherit come from two distinct strata situated in a sedimentary outcrop cut by a deep ravine. The archeological strata belong to the Ain Hanech Formation (Fm), which rests on an erosive disconformity atop the Oued Laatach Fm [supplementary text S2, see (*12*)]. The Ain Hanech Fm contains six stratigraphic members (Mb), bottom to top, from P to U (Fig. 1), consisting of fluvial deposits made of alternating gravels and sandstone with mudstone. The lowermost artifact-bearing stratum (AB-Lw) is located in the sequence near the top of Mb P. Within this stratum, presence of fossil fauna was known (*13*, *14*), and we excavated in situ Oldowan artifacts in association with a sizable faunal assemblage, some with evidence of stone tool cutmarks. The lithic artifacts were overall fresh, but the bones were subjected to minor alterations (fig. S4). The materials were sealed in fine-grained sediments consisting primarily of silt, fine sand, and clay (fig. S6).

The second artifact-bearing stratum (AB-Up), 9 m higher in the sequence, is sealed by the overlying 3.5-m-thick Mb R deposits. A 38-m^2 excavation yielded a faunal assemblage associated with Oldowan artifacts encased in a 0.40-m-thick silty clay and fine sand, underlain by gravels. The fine-grained sediment context (fig. S6), the fresh quality of the artifacts with a large amount of debitage, and the absence of preferred orientation or high dip of the remains suggest a low-energy depositional environment (figs. S12 and S13). Microscopic observations show some taphonomic alterations related to water activities, but sorting of skeletal parts is entirely absent (fig. S4).

The age of the Ain Boucherit archaeological materials is constrained by means of magnetostratigraphy, electron spin resonance (ESR), and mammalian biochronology. The magnetostratigraphic study was carried out on two sections, totaling a 50-m-thick profile (Fig. 1) [materials and methods 1, see (*12*)]. The results indicate a vertical succession of both normal and reversed magnetozones. The independent age control provided by numerical dating (ESR method) enabled us to anchor the local magnetic polarity stratigraphy to the global polarity time scale (GPTS) (*15*). ESR dating was performed on optically bleached quartz grains from Mb P, located ~1 m below AB-Lw (Fig. 1). The ESR age calculations, using the multiple centers approach (*16*), yielded highly consistent dates for the Al and Ti-Li centers. A final combined Al-Ti age is 1.92 ± 0.18 Ma ago (1σ) (fig. S3 and table S4). Although the uncertainty associated with the dose-rate evaluation may affect this result [materials and methods 2, see (*12*)], this numerical chronology unambiguously indicates that the reverse magnetozone in the lower part of the Ain Hanech Fm corresponds to the early Matuyama chron (C2r), which is chronologically constrained between 1.94 and 2.58 Ma ago. Subsequent magnetostratigraphic interpretations indicate that the bottom of the sequence begins with the Gilbert reversed polarity (C2Ar), followed by the Gauss (C2An) normal polarity, ending with the Matuyama above the Olduvai subchron (C2n). Level AB-Lw in Mb P falls within the lower Matuyama chron (C2r), whereas level AB-Up in Mb R correlates to the bottom of C2n (*9*). The Ain Hanech and El Kherba artifact-bearing layers, located higher up in Mb T, are near the top of Olduvai, thus dating to ~1.78 Ma ago (*9*). The calcrete deposits in Mb U, which preserve Acheulean artifacts, are in the reverse chron C1r postdating Olduvai.

This chronostratigraphic framework is supported by mammalian taxa (table S8), several of which are of biochronological relevance. *Kolpochoerus heseloni* (equivalent to *K. limnetes*) (*17*) is present at Ain Hanech (fig. S7) and El Kherba (*18*), and its last appearance is ~1.7 Ma ago (*19*, *20*). *Anancus* is present at AB-Lw (Mb P) (fig. S7, 1a and 1b) and at Ain Hanech (*13*), with the youngest occurrence in East, South, and North Africa and Europe, dating to around 3.8 to 3.5, <3.1 to 2.5, and 2.3 to 2.2 Ma ago, respectively (*21*, *22*). In the Indian subcontinent at Pinjor, and in China in the Nihewan Fm (*23*, *24*), the latest record for *Anancus* dates to the earliest Pleistocene. *Equus numidicus* from AB-Lw and the smaller *E. tabeti* from Ain Hanech and El Kherba have extremely gracile metapodials, whereas African species younger than ~1.2 Ma ago are more robust (fig. S8), that is, until the appearance of the Late Pleistocene *E. melkiensis* [supplementary text S4, see (*12*)]. These taxa support an early post-Olduvai age for Ain Hanech and El Kherba (~1.8 Ma ago) (*9*) and the correlation of AB-Up and AB-Lw to Olduvai and early Matuyama (C2r.2r) subchrons, respectively.

Therefore, the magnetostratigraphic and biochronological data combined with the ESR age lead to the following interpretations: (i) AB-Lw is chronostratigraphically positioned between the beginning of the Olduvai subchron and the top of the Gauss chron, and thus, it is chronologically constrained between 1.94 and 2.58 Ma ago; and (ii) AB-Up has been deposited during the Olduvai

[1]Centro Nacional de Investigación sobre la Evolución Humana (CENIEH), Burgos, Spain. [2]Centre National de Recherches Préhistoriques, Anthropologiques et Historiques (CNRPAH), Algiers, Algeria. [3]Stone Age Institute and Anthropology Department, Indiana University, Bloomington, IN, USA. [4]Australian Research Centre for Human Evolution, Griffith University, Brisbane, Queensland, Australia. [5]Àrea de Prehistòria, Universitat Rovira i Virgili, Tarragona, Spain. [6]Institut Català de Paleoecologia Humana i Evolució Social (IPHES), Tarragona, Spain. [7]Musée National du Bardo, Algiers, Algeria. [8]Museo Nacional de Ciencias Naturales and Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain. [9]Département Homme et Environnement, Museum National d'Histoire Naturelle (MNHN), Paris, France. [10]Département d'Archéologie, Université Lamine Debaghine Sétif 2, Sétif, Algeria. [11]Institut d'Archéologie, Université Alger 2, Algiers, Algeria. [12]Dipartimento di Studi Umanistici, Università Degli Studi di Ferrara, Ferrara, Italy.
*Corresponding author. Email: mohamed.sahnouni@cenieh.es

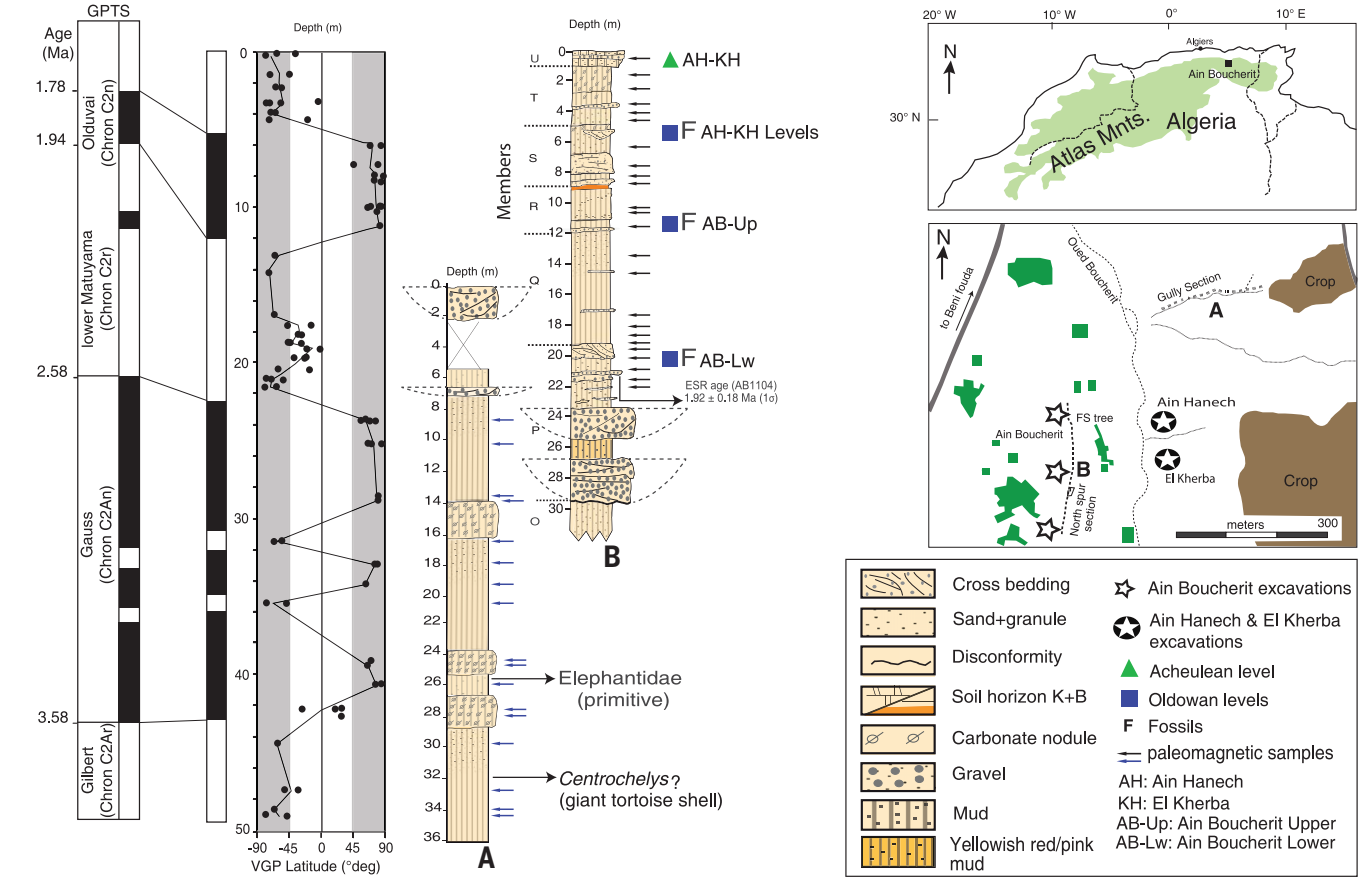


Fig. 1. Location of Ain Boucherit, stratigraphy, and magnetostratigraphic data of the site. The locations of sections A and B (labeled) are shown in the maps on the right. Magnetostratigraphy is expressed with the virtual geomagnetic pole (VGP) latitudinal position. The solid line connects the averaged VGP latitude when several specimens (dots) are used. Data from the upper 22 m of section B are modified from (*9*).

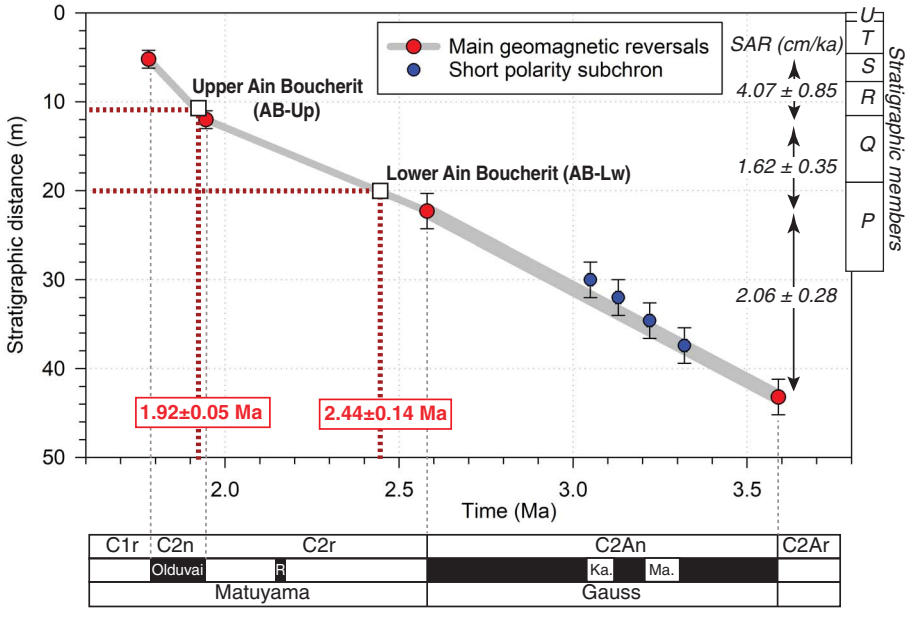


Fig. 2. Sediment accumulation rate values for the Ain Boucherit section and interpolated numerical ages obtained for AB-Lw and AB-Up. AB-Lw and AB-Up are indicated with open squares. The thickness of the gray line and the vertical error bar on the individual points display the depth uncertainty (about 1 m from 0 to 22 m and about 2 m below). See further explanations in supplementary text S1 (*12*). SAR, sediment accumulation rate; cm/ka, cm per thousand years; R, Réunion; Ma., Mamoth; Ka., Kaena.

subchron and therefore has an age between 1.94 and 1.78 Ma ago. Thus, the age of the Olduvai and the Gauss chrons (*15*) and sediment accumulation rates allowed further age estimation [supplementary text S1, see (*12*)], which could not be achieved with the ESR result alone, owing to current limitations of the method for long chronologies. Assuming constant rates during the Olduvai and the Matuyama C2r and neglecting compaction effects, we estimate the age of AB-Up and AB-Lw to 1.92 ± 0.05 Ma ago and 2.44 ± 0.14 Ma ago, respectively (Fig. 2). The latter is, in our opinion, the most reasonable age estimate for AB-Lw, although we do acknowledge a slightly younger age given the possibility of uncertainty on the position of the Gauss-Matuyama boundary [supplementary text S1, see (*12*)].

The lithic assemblages from AB-Lw and AB-Up are made on limestone and flint and consist of 17 and 236 specimens, respectively (Fig. 3, fig. S11, and table S10). The probable sources of the limestone and flint raw materials were the nearby channel beds [supplementary text S5, see (*12*); fig. S10]. The technological and typological features of the Ain Boucherit stone assemblages are similar to the Oldowan from the Early Pleistocene sites in East Africa. The artifact assemblage from AB-Lw includes seven cores, nine flakes, and one retouched piece (Fig. 3). The AB-Lw cores are

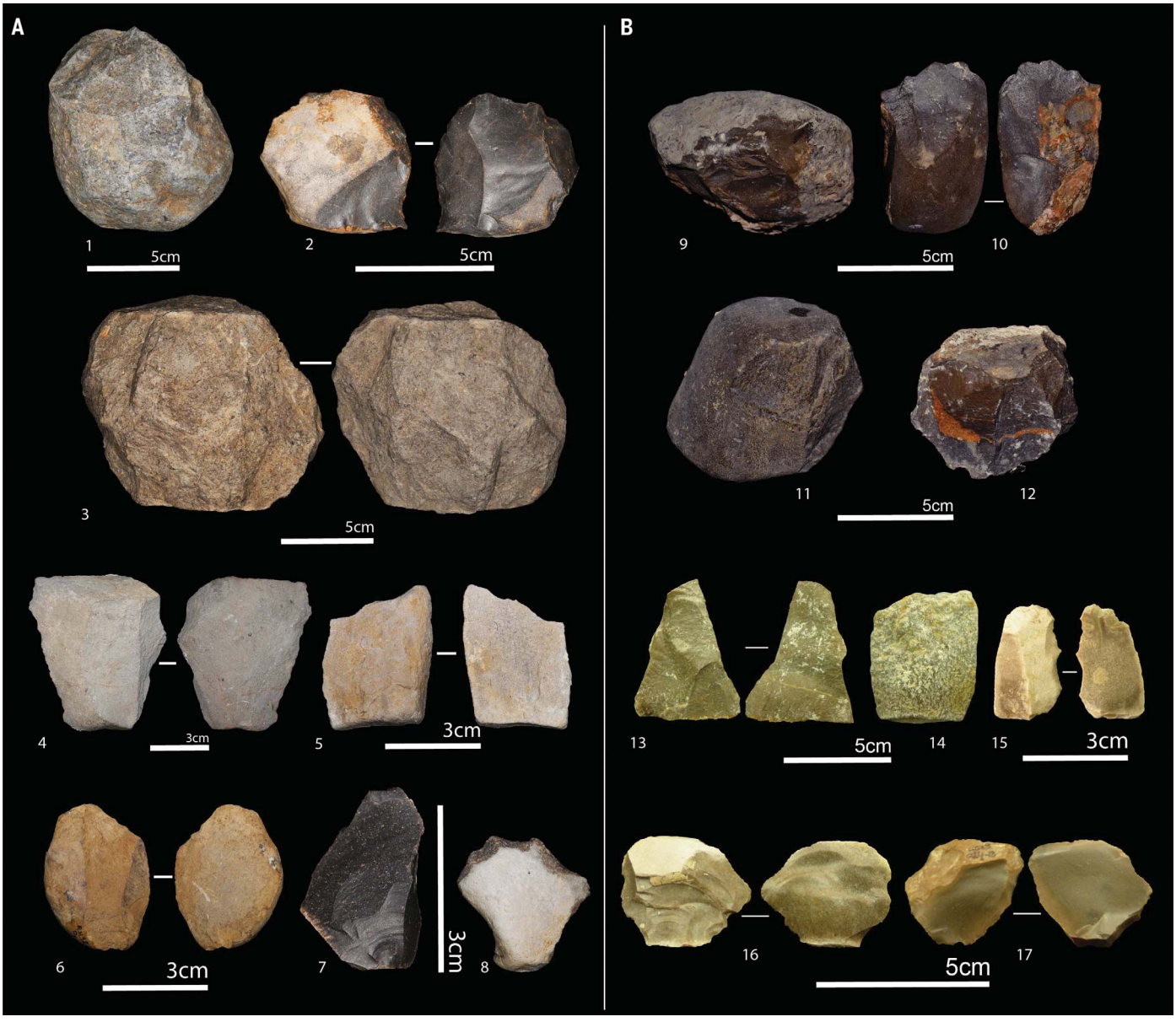


Fig. 3. Oldowan artifacts. (**A** and **B**) Oldowan artifacts from AB-Lw [(A), images 1 to 8] and AB-Up [(B), images 9 to 17], including unifacial cores on limestone (1 and 9); bifacial core made of limestone (10) and on flint (2); polyhedral cores on limestone (11 and 12); subspherical core on limestone (3); whole flakes on flint (7, 16, and 17) and on limestone (4, 5, 6, 13, and 14); and retouched pieces on flint (8 and 15).

variably flaked, with most retaining residual cortical areas, ranging from lightly flaked with two to eight scars to heavily flaked, with one specimen bearing 29 scars. Despite marked technological similarities, some of the cores are predominantly polyhedral and subspherical. The flakes range between 30 and 58 mm in length, and most retain cortex. The retouched specimen is a notched scraper on a cortical flake made of flint.

Abundant stone artifacts were recovered from AB-Up: 121 cores, 65 whole flakes (>2 cm), 3 retouched flakes, and 47 fragments (Fig. 3). The cores are primarily made on limestone (95.8%), with a few made on flint (4.13%). The cores include unifacial choppers (16.94%), bifacial choppers (8.05%), polyhedrons (23.05%), subspheroids (1.69%), and spheroids (0.84%). They were variably flaked, from light to heavy; more than half still retain cortex. Specimens with high scar counts (15 to 30) represent 11.5% of the assemblage. There are also facetted subspheroids with pitting marks suggestive of possible pounding activities. The flakes are predominantly made on limestone, and nearly half of the specimens retain cortex on dorsal faces and platforms. The retouched pieces, chiefly in flint, are small and can be typologically characterized as scrapers and notched scrapers.

The faunal assemblages of AB-Lw and AB-Up include 296 [minimum number of individuals (MNI) = 19] and 277 (MNI = 14) fossil bones, respectively. They are primarily composed of small and medium-sized bovids and equids (tables S5 to S7), also with the best skeletal representations; the appendicular parts in both levels are the most abundant, followed by cranial and axial elements. Evidence of cutmarked and hammerstone-percussed bones is present in both assemblages (Fig. 4). The cutmarks are characterized by isolated or grouped striae with straight trajectory and oblique or transversal orientations. Although variable in depth, many of the specimens have narrow V-shaped cutmarks in cross section with clear internal microstriation and Hertzian cones. In AB-Lw, cutmarks are recognized on 17 bones (5.7% of the assemblage), half of which belong to very small or small-sized animals. The cutmarks are located primarily on limb bones, on ribs, and

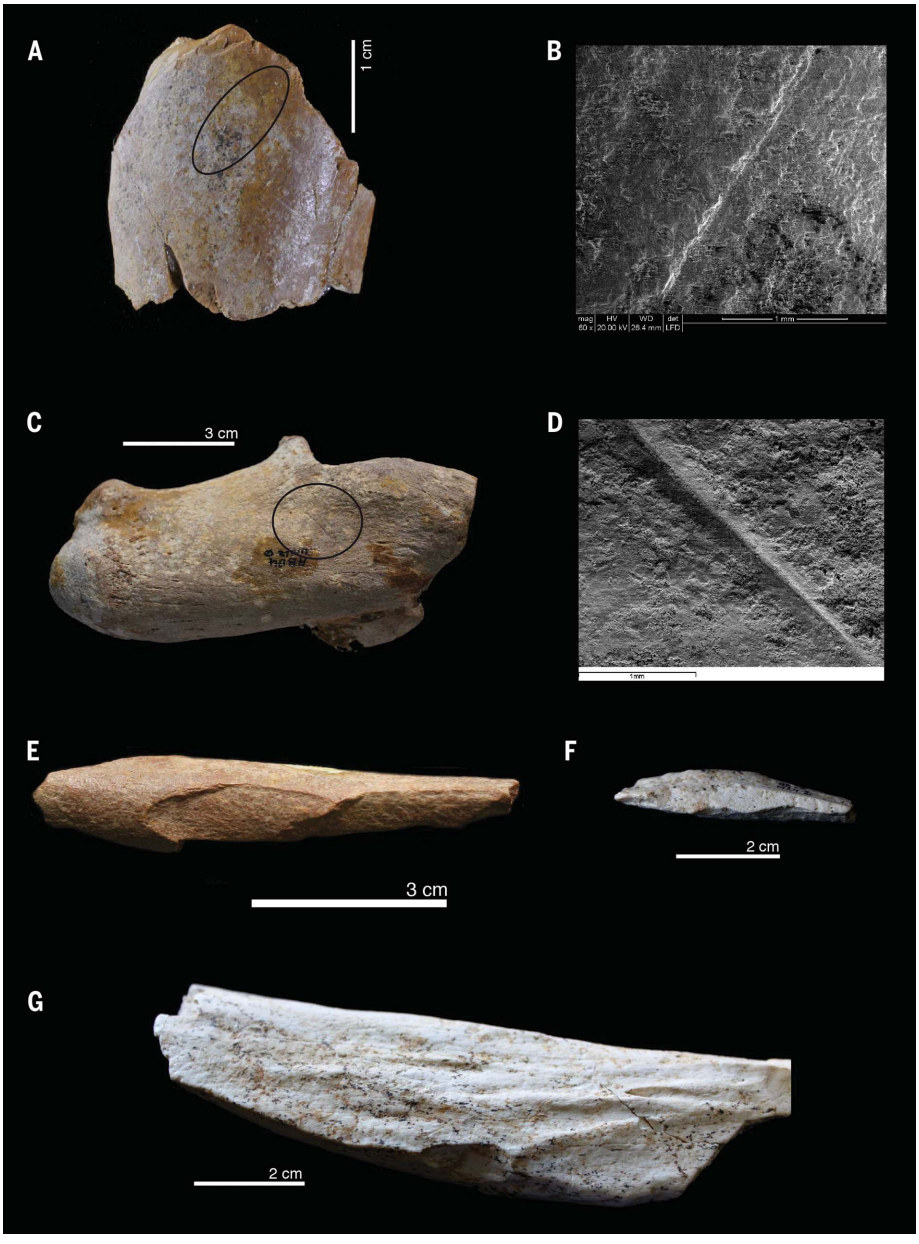


Fig. 4. Evidence of hominin activity from Ain Boucherit faunal assemblages. (**A** and **B**) Slicing mark on a medium-sized bovid humerus shaft from AB-Lw (A), with scanning electron microscopy (SEM) micrograph detail (B). (**C** and **D**) Cutmarked equid calcaneum from AB-Lw (C), with SEM micrograph detail (D). (**E**) Hammerstone-percussed medium-sized long bone from AB-Lw. (**F**) Bone flake from AB-Up. (**G**) Equid tibia from AB-Up, showing cortical percussion notch.

on cranial remains, suggesting skinning, evisceration, and defleshing activities (*25*) (table S7). Four of the bones show hominin-induced percussion marks, including percussion pits, medullary or cortical percussion notches, and a bone flake, implying marrow extraction. The AB-Up bone assemblage yielded two cutmarked bones (an equid tibia and a medium-sized long bone) and seven hammerstone-percussed long bones, which include large (equid) and medium-sized animals and a tibia of a small-sized animal.

The Ain Boucherit stone assemblages are typical of the Oldowan technology, though with subtle typological variations compared to the near-contemporary East African assemblages dated to 2.6 to 1.9 Ma ago, such as Gona, Omo, Hadar (Ethiopia), West Turkana, and Kanjera (Kenya) (*1*, *3*–*6*). In addition to the ubiquitous mode I core and flake stone assemblages, Ain Boucherit also yielded facetted subspheroids. In East Africa, variable mode I artifact assemblages are documented with the early Oldowan (2.6 to 2.0 Ma ago), but facetted spheroids are unknown at these early sites. The observed variability between East and North Africa may be a result of differences in the type and qualities of raw materials used or attributable to function-related factors that we have yet to identify. Moreover, except for Gona and Kanjera, Ain Boucherit stands alone in Africa as the only site with evidence of cutmarked and hammerstone-percussed bones associated with in situ stone tools dated to 2.4 Ma ago. In addition to Kanjera, the Ain Boucherit materials represent a larger sample excavated from a single site, allowing us to make stronger inferences on how hominins butchered carcasses. The Ain Boucherit data unambiguously show hominin exploitation of meat and marrow from all animal size categories and skeletal parts involving skinning, evisceration, and defleshing of upper and intermediate limbs. These activities suggest early access to animal carcasses by hominins (*25*, *26*).

For decades, East Africa has been considered the place of origin of the earliest hominins and lithic technology. Surprisingly, the earliest currently known hominin dated to ~7.0 Ma ago, and the ~3.3-million-year-old *Australopithecus bahrelghazali* have been discovered in Chad, located in the Sahara thousands of kilometers away from the East African Rift (*27*, *28*). Now that Ain Boucherit has yielded Oldowan archaeology estimated to 2.4 Ma ago, northern Africa and the Sahara may be a repository of further archaeological materials. Despite its distance from East Africa, the evidence from Ain Boucherit implies either rapid expansion of stone tool manufacture from East Africa to other parts of the continent or a possible multiple-origin scenario of ancestral hominins and stone technology in both East and North Africa. On the basis of the potential of Ain Boucherit and the adjacent sedimentary basins, we suggest that hominin fossils and Oldowan artifacts as old as those documented in East Africa could be discovered in North Africa as well.

REFERENCES AND NOTES

1. S. Semaw, *J. Archaeol. Sci.* **27**, 1197–1214 (2000).
2. S. Semaw *et al.*, *J. Hum. Evol.* **45**, 169–177 (2003).
3. H. Roche *et al.*, *Nature* **399**, 57–60 (1999).
4. A. Delagnes *et al.*, *J. Hum. Evol.* **61**, 215–222 (2011).
5. W. H. Kimbel *et al.*, *J. Hum. Evol.* **31**, 549–561 (1996).
6. T. Plummer, L. C. Bishop, P. Ditchfield, J. Hicks, *J. Hum. Evol.* **36**, 151–170 (1999).
7. J. V. Ferraro *et al.*, *PLOS ONE* **8**, e62174 (2013).
8. R. Klein, *The Human Career* (Univ. of Chicago Press, ed. 3, 2009).
9. J. M. Parés *et al.*, *Quat. Sci. Rev.* **99**, 203–209 (2014).
10. M. Sahnouni *et al.*, *J. Hum. Evol.* **43**, 925–937 (2002).
11. M. Sahnouni *et al.*, *J. Hum. Evol.* **64**, 137–150 (2013).
12. Materials and methods 1 to 3 and supplementary text S1 to S5 are provided as supplementary materials.
13. C. Arambourg, *Les Vertébrés du Pléistocène de l'Afrique du Nord, Volume 1* (Archives du Muséum national d'histoire naturelle Series, vol. 10, Kapp & Lahure, Paris, 1970), pp. 1–127.
14. C. Arambourg, *Les Vertébrés villafranchiens d'Afrique du Nord* (Singer-Polignac, 1979).
15. F. M. Gradstein, J. G. Ogg, M. D. Schmitz, G. M. Ogg, Eds., *The Geologic Time Scale* (Elsevier, ed. 1, 2012).
16. S. Toyoda, P. Voinchet, C. Falguères, J. M. Dolo, M. Laurent, *Appl. Radiat. Isot.* **52**, 1357–1362 (2000).
17. H. S. B. Cooke, *Geobios* **30**, 121–126 (1997).
18. M. Sahnouni, J. van der Made, in *The Cutting Edge: New Approaches to the Archaeology of Human Origins*, K. Schick, N. Toth, Eds. (Stone Age Institute Press, 2009), pp. 179–210.

19. T. D. White, in *Paleoclimate and Evolution, With Emphasis on Human Origins*, E. S. Vrba, G. H. Denton, T. C. Partridge, L. H. Burckle, Eds. (Yale Univ. Press, 1995), pp. 369–384.
20. H. S. B. Cooke, in *Hominin Environments in the East African Pliocene: An Assessment of the Faunal Evidence*, R. Bobe, Z. Alemseged, A. K. Behrensmeyer, Eds. (Springer, 2007), pp. 107–127.
21. W. J. E. Sanders, E. Gheerbrant, J. M. Harris, H. Saegusa, C. Delmer, in *Cenozoic Mammals of Africa*, L. Werdelin, W. J. Sanders, Eds. (Univ. California Press, Berkeley, 2010), pp. 161–251.
22. G. Garrido, G. A. Arribas, *Palaeontol. Electronica* **17**, 1–16 (2014).
23. A. C. Nanda, *Quat. Int.* **192**, 6–13 (2008).
24. G.-F. Chen, *Vertebrata PalAsiatica* **37**, 175–189 (1999).
25. P. J. Nilssen, "An actualistic butchery study in South Africa and its implications for reconstructing hominid strategies of carcass acquisition and butchery in the Upper Pleistocene and Plio-Pleistocene," thesis, University of Cape Town, South Africa (2000).
26. M. Domínguez-Rodrigo, T. R. Pickering, S. Semaw, M. J. Rogers, *J. Hum. Evol.* **48**, 109–121 (2005).
27. M. Brunet *et al*., *Nature* **378**, 273–275 (1995).
28. M. Brunet *et al*., *Nature* **418**, 145–151 (2002).

ACKNOWLEDGMENTS

We would like to thank the Algerian Ministry of Culture for the research permit; S. Hachi, director of CNRPAH; the Wilaya of Sétif; the municipality of Guelta Zergua; K. Guechi (president of the University of Sétif 2) and Y. Aibeche (vice president of the same) for administrative and logistic support during fieldwork at Ain Boucherit; and the CENIEH (Spain) staff, especially M. J. de Miguel del Barrio and B. de Santiago Salinas for administrative support. M.D. is grateful to V. Guilarte and D. Martínez Asturias for their invaluable contribution in the ESR dating analytical procedure. J.M.P. thanks C. Álvarez for fieldwork assistance and C. Saiz for sample preparation in the laboratory. **Funding**: Support by grants from CNRPAH, MINECO (HAR2013-41351-P), The L.S.B. Leakey Foundation, the National Science Foundation (NSF-BCS-0517984), the Wenner-Gren Foundation (grants 7815 and 8323), the European Research Council (FP7-People-CIG2993581), and the Stone Age Institute (Bloomington, IN) to M.S.; MINECO (CGL2010-16821) to J.M.P. and M.D.; the Australian Research Council (Future Fellowship 150100215) to M.D.; and the European Science Foundation (Synthesys GB-TAF-4119 and DE-TAF-668) to J.v.d.M., CGL2015-65387-C3-1-P (MINECO/FEDER) to J.v.d.M. and I.C., and AGAUR (2017SGR1040) and URV (2017PFR-URV-B2-91) to I.C. **Author contributions**: M.S. was principal author and directed the project; J.M.P. and M.D. led the geochronological research; A.P.-G. and S.A. led geological research; M.S., Z.H., A.D., and M.M. conducted the excavation and recovery of archaeological and fossil materials; M.S., Z.H., and S.S. conducted the lithic analysis; J.v.d.M. and K.B. conducted the paleontological analysis; and I.C. and N.K. conducted the taphonomical analysis. All authors participated in the writing of the manuscript. **Competing interests**: The authors have no competing interests. **Data and materials availability**: All data are available in the paper and supplementary materials, and the archaeological and paleontological materials are deposited in CNRPAH, Musée du Bardo, and Musée de Sétif in Algeria.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/1297/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 to S10
References (*29*–*74*)

26 April 2018; accepted 6 November 2018
Published online 29 November 2018
10.1126/science.aau0008

PLANT SCIENCE

A femtomolar-range suicide germination stimulant for the parasitic plant *Striga hermonthica*

Daisuke Uraguchi[1]*, Keiko Kuwata[2], Yuh Hijikata[2,3], Rie Yamaguchi[2], Hanae Imaizumi[2], Sathiyanarayanan AM[2], Christin Rakers[3]†, Narumi Mori[4], Kohki Akiyama[4], Stephan Irle[2,3]‡, Peter McCourt[5], Toshinori Kinoshita[2,3], Takashi Ooi[1,2,6]*, Yuichiro Tsuchiya[2]*

The parasitic plant *Striga hermonthica* has been causing devastating damage to the crop production in Africa. Because *Striga* requires host-generated strigolactones to germinate, the identification of selective and potent strigolactone agonists could help control these noxious weeds. We developed a selective agonist, sphynolactone-7, a hybrid molecule originated from chemical screening, that contains two functional modules derived from a synthetic scaffold and a core component of strigolactones. Cooperative action of these modules in the activation of a high-affinity strigolactone receptor ShHTL7 allows sphynolactone-7 to provoke *Striga* germination with potency in the femtomolar range. We demonstrate that sphynolactone-7 is effective for reducing *Striga* parasitism without impinging on host strigolactone-related processes.

Striga hermonthica (*Striga*) parasitizes crops widely across various parts of sub-Saharan Africa, causing loss in crop yields that result in economic pressure on millions of smallholder farmers and lead to annual losses of billions of dollars (*1*). Protecting crops from the numerous tiny *Striga* seeds buried in the soil requires integration of various approaches to suppress infestation (*1*). A group of host-generated small-molecule hormones, called strigolactones (SLs), provoke germination of *Striga* seeds. Because *Striga* is an obligate parasite, germination in the absence of a host is lethal, and this has prompted researchers to develop SL agonists as inducers of suicidal germination to purge the soil of viable *Striga* seeds (*2*). This approach requires the development of potent and accessible compounds that only act on *Striga* and do not impede normal crop development. For example, SLs are also plant chemical cues that attract root symbiotic arbuscular mycorrhizal fungi (AM fungi) that supply host plants with nutrients (*3*, *4*). Here, we report the development of a *Striga*-selective SL agonist acting in the femtomolar range.

SLs are a group of plant-derived molecules whose structures consist of butenolide rings (D-rings), which are connected to cyclic moieties, usually three-ring systems (ABC-rings), through an enol-ether bridge (Fig. 1A). In vascular plants, SLs are plant hormones that optimize plant body architectures through the *DWARF14* (*D14*) family of α/β hydrolase-fold receptors (*5*). D14 defines a noncanonical receptor because it initiates signal transduction by using enzymatic activity. Upon binding, SLs undergo cleavage of the enol-ether bridge through hydrolysis to leave the D-ring as a covalently linked intermediate molecule (CLIM) at the catalytic histidine residue in the receptor (*6*–*8*). Previous studies suggest that the ABC-portion of the SL is released from the D14 pocket, and the receptor–CLIM complex alters D14 conformation to recruit downstream negative regulators such as the SCF^{MAX2} protein (*7*). In *Striga*, it is thought that SLs trigger seed germination through 11 members of an independently diverged α/β hydrolase-fold receptors called *Striga HYPOSENSITIVE TO LIGHT/KARRIKIN INSENSITIVE2* (*ShHTL/KAI2*, here called "ShHTLs") (*9*–*11*). The hydrolytic activity of ShHTLs was exploited in the development of fluorogenic SL probes to uncover an ethylene-mediated amplification of a wave-like pattern of SL perception initiated during *Striga* germination (*10*). Moreover, in vitro binding suggests that the divergence of ligand preferences in ShHTLs is beneficial for *Striga* seeds to detect the blend of SLs exuding from preferred host species (*10*). Among these ShHTL isoforms, we have focused on ShHTL7 because this receptor is sensitive to picomolar levels of SLs when heterologously expressed in *Arabidopsis*, and its large binding pocket ensures a response to structurally diverse molecules (*11*, *12*). These characteristics make ShHTL7 a suitable target for the development of agonists for stimulating *Striga* germination.

Chemical analysis on SLs over the past 40 years suggests that the structure of the D-ring is essential to SL activity (*2*, *3*). By contrast, structural flexibility in the ABC-portion has led to the development of various synthetic SLs or SL mimics, including GR24 or simplified phenol–D-ring derivatives called debranones (*2*, *13*). However, the structural element of the ABC-portion that would contribute to both potency and specificity to *Striga* remains elusive. To explore the chemical characteristics that define species selectivity toward *Striga*, we performed a small-molecule screen for compounds that germinate *Striga* seeds (harvests from sorghum fields in Sudan). This screening of 12,000 synthetic molecules was followed by additional synthesis of 60 analogs of hit compounds that were found from the initial screening. On the basis of median inhibitory concentration (IC_{50}) using the fluorogenic SL-mimic Yoshimulactone Green (YLG), the binding assay resulted in the identification of *N*-arylsulfonylpiperazine as a molecular scaffold that selectively bound to ShHTL7 (Fig. 1, A and B, fig. S1, and table S1). A representative molecule, SAM690, which contains the arylsulfonylpiperazine moiety, exhibited potency toward *Striga* germination at the micromolar level. The mode of action of SAM690 was similar to that of (+)-GR24, in that germination activity was suppressed by inhibition of ethylene production (Fig. 1C). However, unlike (+)-GR24, SAM690 was not hydrolyzed by ShHTL7 (fig. S2) (*10*). These observations indicate that SAM690 stimulates *Striga* germination with selective activation of ShHTL7 through a mechanism independent of hydrolysis.

During a series of above assays, we noticed inconsistency in stimulant activities of several SAM690 derivatives depending on the purification method due to active impurity. This byproduct, although only 0.01% of the total product, appeared to be an unusually oxidized molecule that has a hybrid structure resembling SAM690 with a D-ring–like butenolide moiety (Fig. 1A and fig. S3). In order to verify the structure and potency of this derivative, we established a three-step synthetic procedure, and the resulting oxidized SAM690 exhibited potency comparable with that of (+)-GR24, as evident from its minimum effective concentration (MEC) of 10 pM (Fig. 1D). As expected from its structure, oxidized SAM690 was hydrolyzed by

[1]Graduate School of Engineering, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8603, Japan. [2]Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan. [3]Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8602, Japan. [4]Graduate School of Life and Environmental Sciences, Osaka Prefecture University, 1-1 Gakuen-cho, Naka-ku, Sakai, Osaka 599-8531, Japan. [5]Department of Cell and Systems Biology, University of Toronto, 25 Willcocks Street, Toronto M5S 3B2, Canada. [6]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency (CREST), Japan Science and Technology Agency (JST), Nagoya University, Nagoya 464-8601, Japan.
***Corresponding author. Email: uraguchi@chembio.nagoya-u.ac.jp (D.U.); tooi@chembio.nagoya-u.ac.jp (T.O.); yuichiro@itbm.nagoya-u.ac.jp (Y.T.)** †Present address: Graduate School of Pharmaceutical Sciences, Kyoto University, 46-29 Yoshida-Shimo-Adachi-cho, Sakyo-ku, Kyoto 606-8501, Japan. ‡Present address: Computational Sciences and Engineering Division & Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.

ShHTL7 (fig. S2). The structural similarity of this compound to SLs led us to hypothesize that attaching a methyl group to the C4′ position may enhance the potency of the molecules. Indeed, this modification improved MEC from 10 pM to 10 fM (Fig. 1, A and D). We named the D-ring/sulfonylpiperazine-hybrid molecule sphynolactone-7 (SPL7) and named its demethylated analog H-SPL7 (sulfonylpiperazine hybrid strigolactone mimic of ShHTL7) (the stability and toxicology of SPL7 are summarized in fig. S5). The name is derived from the sphinx, a mythical creature with the head of a human and the body of a lion, to represent the hybrid nature of the

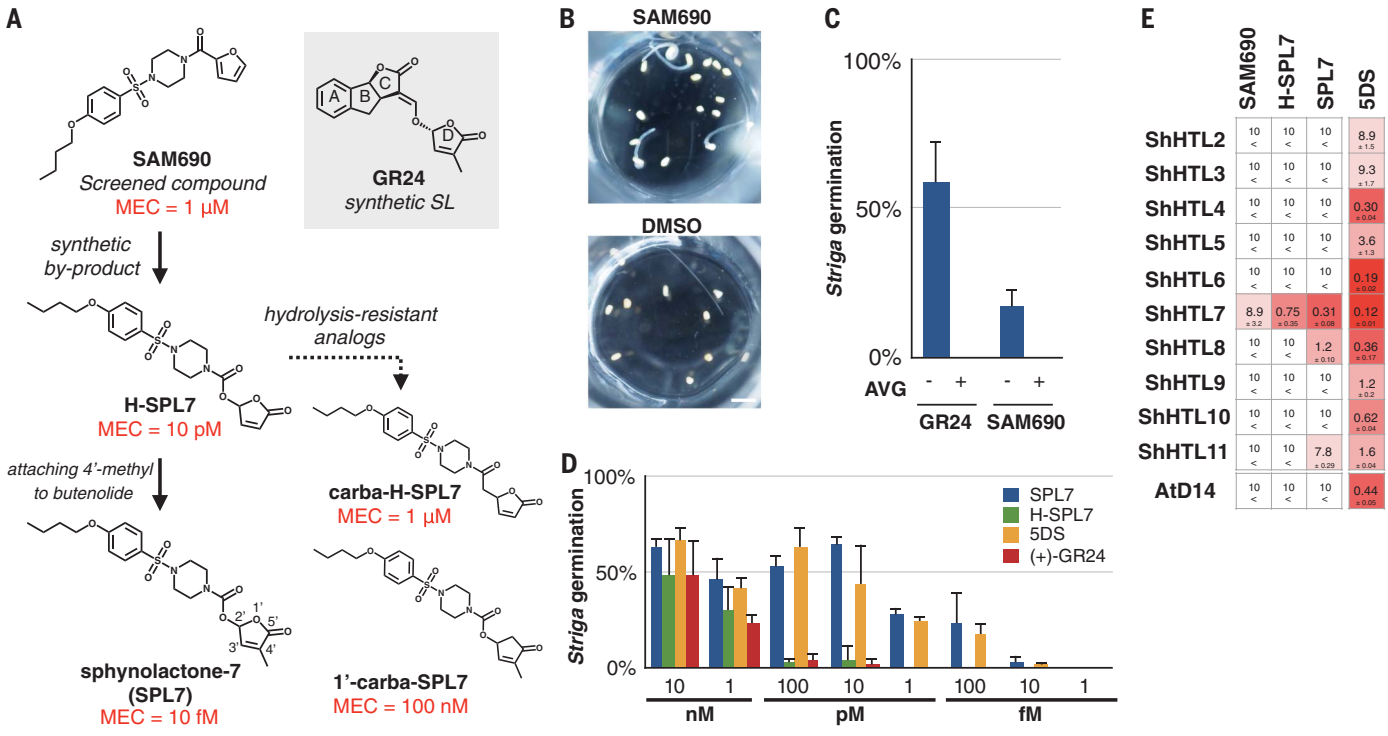


Fig. 1. Development of a femtomolar-range germination stimulant for *Striga*. (**A**) Scheme of structure development. MEC represents the lowest concentration of the compound that produces any seed germination. (**B**) SAM690 induces *Striga* seed germination at 10 µM. Scale bar, 1 mm. (**C**) 10 µM aminoethoxyvinyl glycine (AVG) suppresses (+)-GR24 and SAM690. (**D**) *Striga* germination in dilution series of SPL7, H-SPL7, 5DS, and (+)-GR24. (**E**) Competitive bindings to ShHTLs and AtD14. IC_{50} value (in micromolar) in the YLG assay is presented as a heat map with SD (n = 3 technical replicates). Data for 5DS was obtained from (*10*). Error bars in (C) and (D) indicate SD (n = 3 biological replicates).

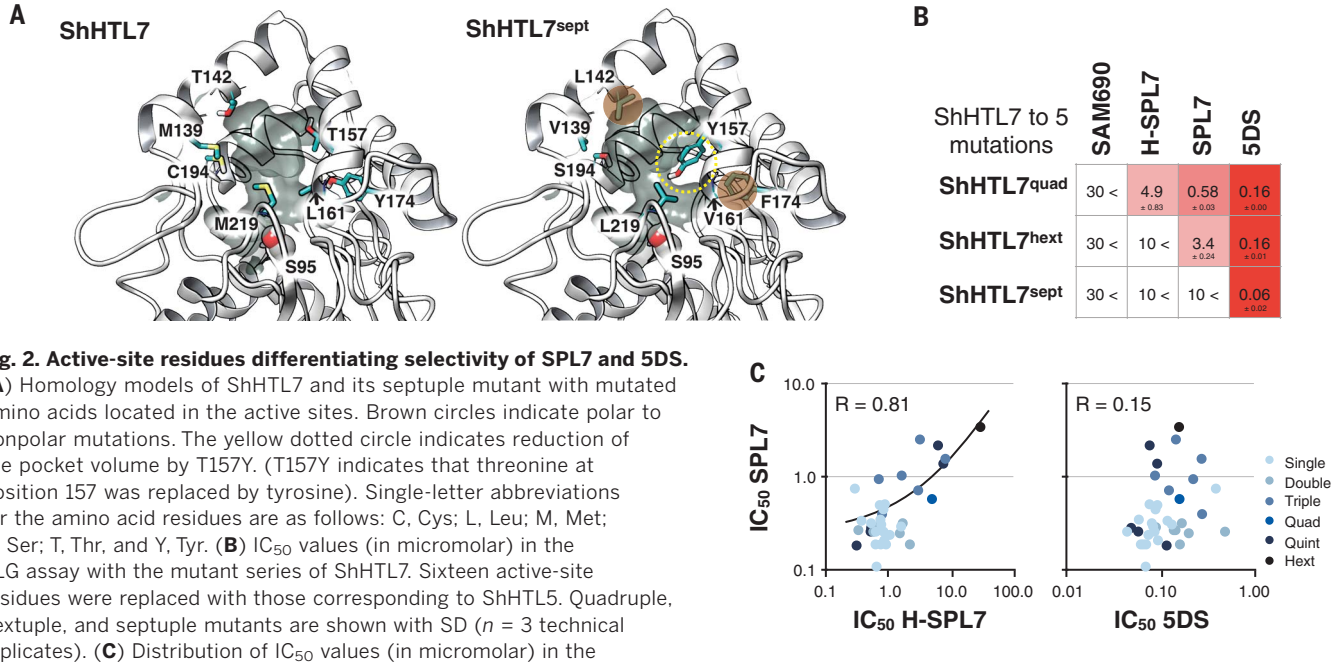


Fig. 2. Active-site residues differentiating selectivity of SPL7 and 5DS. (**A**) Homology models of ShHTL7 and its septuple mutant with mutated amino acids located in the active sites. Brown circles indicate polar to nonpolar mutations. The yellow dotted circle indicates reduction of the pocket volume by T157Y. (T157Y indicates that threonine at position 157 was replaced by tyrosine). Single-letter abbreviations for the amino acid residues are as follows: C, Cys; L, Leu; M, Met; S, Ser; T, Thr, and Y, Tyr. (**B**) IC_{50} values (in micromolar) in the YLG assay with the mutant series of ShHTL7. Sixteen active-site residues were replaced with those corresponding to ShHTL5. Quadruple, hextuple, and septuple mutants are shown with SD (n = 3 technical replicates). (**C**) Distribution of IC_{50} values (in micromolar) in the series of ShHTL7 mutants.

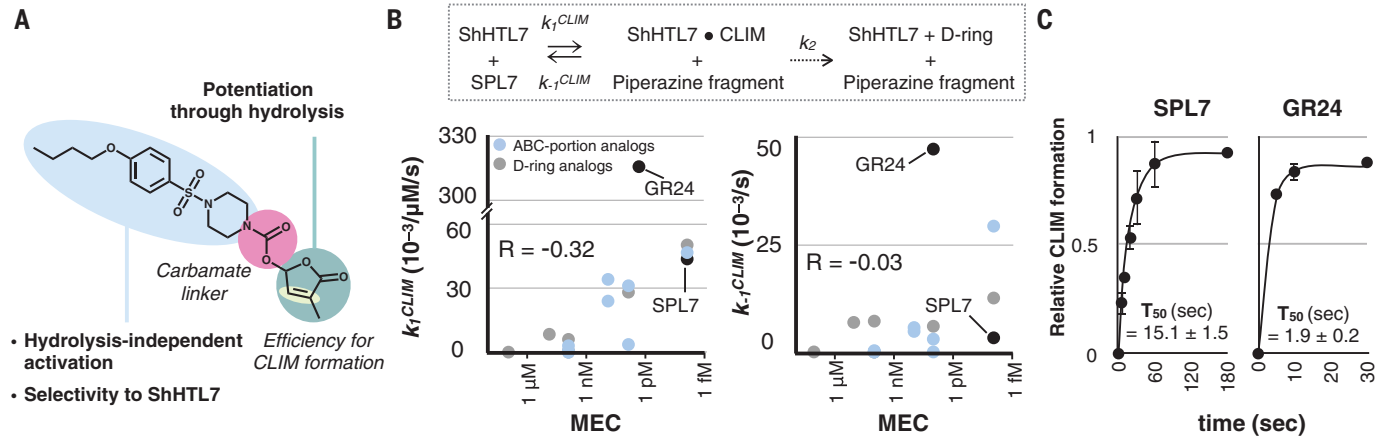


Fig. 3. Mode of action of SPL7. (**A**) Annotation of structural modules identified from the structure–activity–relationship study. (**B**) Relationship between reaction rate constants and MEC among SPL7 analogs. (Top) Reaction scheme and (Bottom) scatter plot of k_1^{CLIM} or k_{-1}^{CLIM} against MEC of *Striga* germination are presented. (**C**) Time-dependent CLIM formation quantified by LC-MS. T_{50} indicates the half-maximal time. Error bar indicates SD (n = 3 technical replicates).

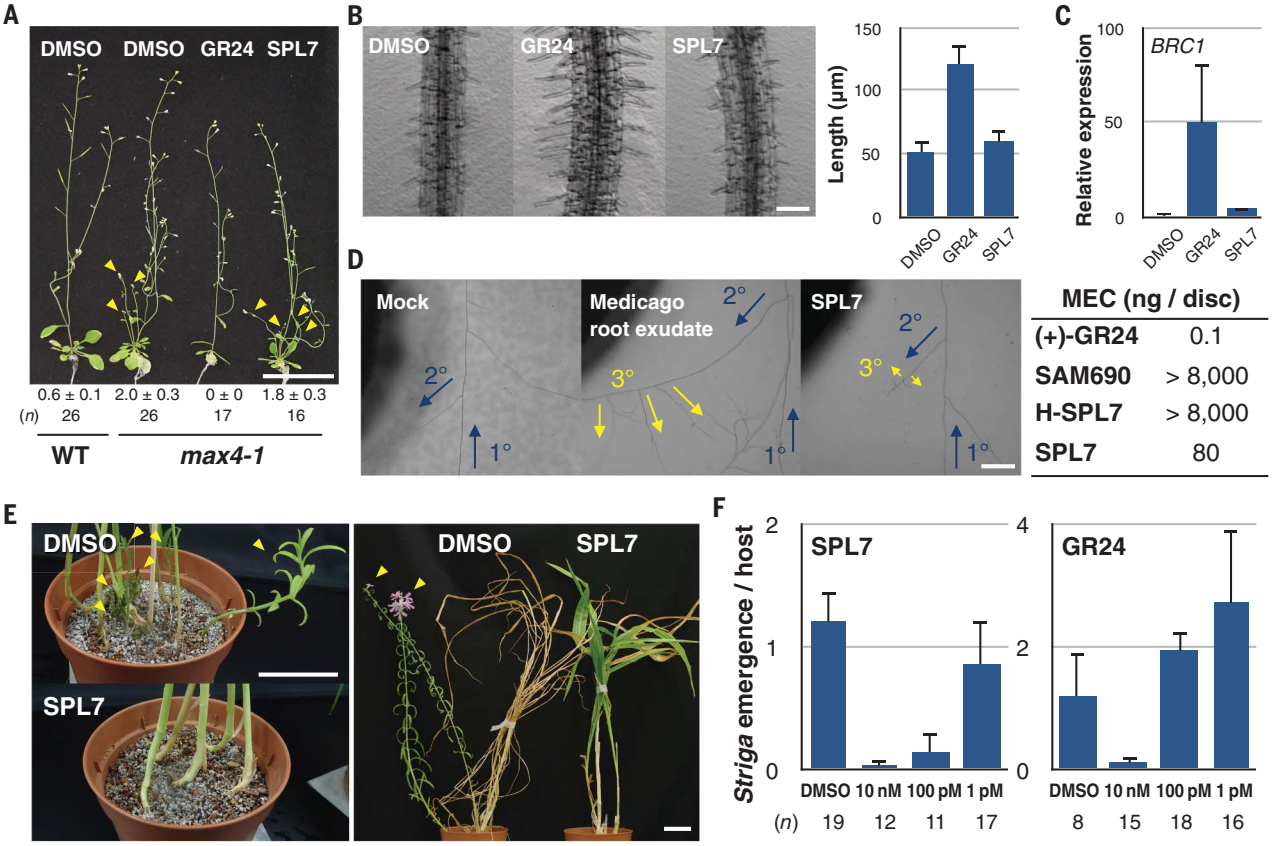


Fig. 4. Bioassays with SPL7. (**A**) SPL7 does not suppress shoot-branching phenotype of *Arabidopsis* SL biosynthetic mutant, *max4-1*, at 10 μM. Arrowheads indicate axillary branches. Average numbers of axillary branches are indicated with SE; n indicates number of plants tested. Scale bar, 5 cm. (**B**) SPL7 fails to enhance root hair elongation in *Arabidopsis* wild-type at 10 μM. Average length of root hair is presented with SD (n = 7 biological replicates). Scale bar, 100 μm. (**C**) SPL7 fails to induce SL-inducible *BRANCHED1* (*BRC1*) expression in *Arabidopsis* wild-type at 10 μM. Average expression obtained from quantitative reverse transcription polymerase chain reaction (RT-PCR) analysis is presented as relative value to DMSO control with SD (n = 3, biological replicates). (**D**) SPL7 shows 800 times less potency for AM fungi than that of (+)-GR24. MEC represents the lowest concentration of compound that induces multiple 3° hyphae. Data for (+)-GR24 were obtained from (*19*). Scale bar, 1 mm. (**E**) Suicide germination assay. Representative pictures taken after 2 months (left) or 3 months (right) of cocultivation of maize with *Striga*. The soil was pretreated with DMSO or 10 nM of SPL7. Arrowheads indicate emerged *Striga*. Scale bar, 5 cm. (**F**) Number of emerged *Striga* after 2 months of cocultivation. n indicates number of hosts tested. Error bar indicates SE.

molecule. The IC_{50} values of SPL7 improved from SAM690 (0.31 versus 8.9 μM), and our liquid chromatography–mass spectrometry (LC-MS) analysis revealed that SPL7 was hydrolyzed by ShHTL7 to form CLIM at the catalytic histidine residue (Fig. 1E and figs. S2 and S4) (*7*, *14*). The potency of SPL7 is comparable with that of (+)-5-deoxystrigol (5DS), a natural SL that is currently the most potent commercially available germination stimulant for *Striga*.

Despite their high potencies, the presence of the *N*-arylsulfonylpiperazine scaffold allows SPL7 to retain selectivity toward ShHTL7, whereas 5DS binds to all the SL receptors with different ranges of IC_{50} values (Fig. 1E) (*10*). To gain insight into this difference in selectivity, we replaced 16 active-site residues of ShHTL7 with those of ShHTL5 (*11*). Using the YLG binding assay, we identified seven residues that are essential for the binding with SPL7 (M139, T142, T157, L161, Y174, C194, and M219) (Fig. 2, A and B, and fig. S6). The combination of these mutations led to a distribution of IC_{50} values of SPL7, which was correlated with that of H-SPL7 [correlation coefficient (r = 0.81)] but not with that of 5DS (r = 0.15) (Fig. 2C). These results indicate that SPL molecules use a different subset of residues for binding compared with those of natural SLs, displaying selectivity. Our computational investigation supports the hypothesis that SPL7 could fit to the active site of the homology model of ShHTL7, whereas changes in polarity and volume through active-site mutations may impair its fit (Fig. 2A and fig. S7). These seven amino acids as a combination are specific to *ShHTL7* among known *HTL*/*KAI2* homologs, including those from a parasitic plant *Orobanche minor*, which also uses SLs as germination stimulants (fig. S8) (*3*, *9*). Consistently, SPL7 exhibits nanomolar-level potency to *O. minor* and is effective at femtomolar range for several *S. hermonthica* ecotypes that parasitize to different hosts (fig. S8).

Because SPL7 and GR24 have identical D-ring structures, the selectivity to ShHTL7 and the femtomolar-range potency must be encoded in the ABC-portion of SPL7 (Fig. 3A). In light of an activation model solely dependent on CLIM formation as proposed in D14, the ABC-portion of SPL7 possibly contributes to efficient CLIM formation on the receptor (*7*, *14*). Alternatively, the ABC-portion may have additional functions other than accelerating CLIM formation. We assessed these possibilities through investigation of the relationship between potencies and D-ring hydrolysis using various SPL7 analogs. The potencies of two hydrolysis-resistant analogs, carba-H-SPL7 and 1′-carba-SPL7, were ≥100 nM, implying that the hydrolysis of D-ring is dispensable for activity yet essential to gain the femtomolar-level potency (Fig. 1A and fig. S9). Next, to investigate the quantitative relationship between potencies and the hydrolysis reaction rate, we performed a kinetic analysis similar to that involving surface plasmon resonance, which allows estimation of reaction rate constants k_1 and k_{-1} independently (*15*). Briefly, we obtained the parameter k_1^{CLIM} and (k_{-1}^{CLIM} + k_2) by fitting an equation formularized from a reaction scheme in Fig. 3B to experimentally obtained time-dependent CLIM-formation curves (supplementary materials, materials and methods) (*8*). We assumed (k_{-1}^{CLIM} + k_2) ≈ k_{-1}^{CLIM} because observed stability of CLIM-ShHTL7 complex over 30 min theoretically limited k_2 to <1% fraction of (k_{-1}^{CLIM} + k_2) in our analysis. The kinetic analysis with SPL7 analogs allowed us to observe only a vague trend between potency and k_1^{CLIM} (r = −0.32), indicating that the rate of CLIM formation, although important, was not a sole factor for determining potency (Fig. 3B and figs. S10 and S11). This interpretation was supported by the observation with GR24, in which the reaction rate of the CLIM formation was higher (k_1^{CLIM} = 316 × 10^{-3}/μM/s) than that of SPL7 (k_1^{CLIM} = 43.5 × 10^{-3}/μM/s) despite a potency 1000 times lower than that of SPL7 (Figs. 1D and 3, B and C). These results are contradictory to the model proposed for D14, thus indicating that the ABC-portion of SPL7 has additional functions other than accelerating CLIM formation for delivering the difference in potency (*7*, *14*). Although difference in the uptake or stability in *Striga* seeds could account for differences in potency, we obtained no positive results supporting this assumption (fig. S12). On the basis of these observations, we hypothesized that the function of the ABC-portion after the hydrolysis is essential to deliver femtomolar-level potency (fig. S13). Verification of this model will require detailed studies on the metabolic fate of SPL7 and crystallization of SPL7-ShHTL7 complex.

We next tested the utility of SPL7 as a *Striga*-selective suicide germination stimulant, using three organism-based bioassays. First, we applied 10 μM SPL7 to a SL biosynthetic mutant, *more axillary growth4-1* (*max4-1*), to see whether SPL7 restores the increased branching phenotype (*16*). SPL7 failed to rescue *max4-1* branching defects, although a similar concentration of GR24 did suppress axillary branch emergence (Fig. 4A). SPL7 also failed to induce root hair elongation or induce SL-inducible gene expressions in wild-type *Arabidopsis* (Fig. 4, B and C) (*17*, *18*). Thus, SPL7 exhibits no hormonal SL activity in *Arabidopsis* assays. Second, we evaluated the effect of SPL7 on AM fungi, which are agronomically important microbes that support the growth of crops. Whereas SLs induced multiple 3° hyphal branches as in *Medicago* root exudate, SPL7 exhibited only a mild effect at the highest concentration, showing 800 times less activity than that of (+)-GR24 (Fig. 4D) (*19*). Last, we evaluated the ability of SPL7 to induce suicide germination of *Striga* in a pot infestation assay (Fig. 4, E and F). In the dimethyl sulfoxide (DMSO) control, *Striga* seeds parasitized maize and emerged from the soil at an average of one seedling per host. Soil treatment with SPL7 at a concentration of 100 pM or higher for a week before planting maize reduced the emergence of *Striga* and protected the host plants from senescence caused by parasitism. By contrast, GR24 requires 10 nM to obtain similar effect. Taken together, we concluded that SPL7 is effective as a *Striga*-selective suicide-germination stimulant, at least in laboratory experiments.

The discovery of SPL7 reinforced the design principle of SL mimics as a hybrid of two functional modules, a modifiable synthetic scaffold responsible for both receptor selectivity and potency as the ABC-portion and the D-ring component of natural SLs. Implications of the strategy for basic science includes direct dissection of the roles of specific SL receptors in experimentally intractable organisms such as *Striga*. For practical purpose, the strategy appears applicable to other noxious parasitic weeds, including *Orobanche* or *Phelipanche* species.

REFERENCES AND NOTES

1. G. Ejeta, *Integrating New Technologies for Striga Control* (World Scientific Publishing, 2007), pp. 3–16.
2. S. C. Wigchert *et al.*, *J. Agric. Food Chem.* **47**, 1705–1710 (1999).
3. X. Xie, K. Yoneyama, K. Yoneyama, *Annu. Rev. Phytopathol.* **48**, 93–117 (2010).
4. K. Akiyama, K. Matsuzaki, H. Hayashi, *Nature* **435**, 824–827 (2005).
5. T. Arite *et al.*, *Plant Cell Physiol.* **50**, 1416–1424 (2009).
6. C. Hamiaux *et al.*, *Curr. Biol.* **22**, 2032–2036 (2012).
7. R. Yao *et al.*, *Nature* **536**, 469–473 (2016).
8. A. de Saint Germain *et al.*, *Nat. Chem. Biol.* **12**, 787–794 (2016).
9. C. E. Conn *et al.*, *Science* **349**, 540–543 (2015).
10. Y. Tsuchiya *et al.*, *Science* **349**, 864–868 (2015).
11. S. Toh *et al.*, *Science* **350**, 203–207 (2015).
12. D. Holbrook-Smith, S. Toh, Y. Tsuchiya, P. McCourt, *Nat. Chem. Biol.* **12**, 724–729 (2016).
13. K. Fukui, D. Yamagami, S. Ito, T. Asami, *Front. Plant Sci.* **8**, 936 (2017).
14. R. Yao *et al.*, *Cell Res.* **27**, 838–841 (2017).
15. R. Karlsson, A. Michaelsson, L. Mattsson, *J. Immunol. Methods* **145**, 229–240 (1991).
16. K. Sorefan *et al.*, *Genes Dev.* **17**, 1469–1474 (2003).
17. Y. Kapulnik *et al.*, *Planta* **233**, 209–216 (2011).
18. D. C. Nelson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8897–8902 (2011).
19. K. Akiyama, S. Ogasawara, S. Ito, H. Hayashi, *Plant Cell Physiol.* **51**, 1104–1117 (2010).

ACKNOWLEDGMENTS

We thank A. Babiker, S. Runo, and P. Matana for providing the *S. hermonthica* seeds; K. Yoneyama for providing *O. minor* seeds; and S. Hagihara and M. Yoshimura for providing YLG. We thank N. Nakamichi for instructing quantitative RT-PCR analysis and J. X. Yap for supporting biochemistry works. We thank A. Miyazaki for proofreading. **Authors contributions:** The chemical aspect of the research was managed by D.U. and T.O. Conceptualization of the project and the management of biological aspect of the research was performed by Y.T. Chemical screening was performed by Y.T. under the supervision of P.M. and T.K.; Y.T. and H.I. performed *Striga* germination assays and YLG assays. S.AM. synthesized analog series of initial

hits, and R.Y. synthesized SPL7 analogs under the supervision of D.U. and T.O. *Arabidopsis* assays, quantitative RT-PCR, and suicide germination assay were performed by H.I. under supervision of Y.T.; N.M. performed hyphal branching assay with AM fungi under the supervision of K.A.; K.K. performed LC- MS analyses for small molecules and proteins. Mathematical characterization of CLIM formation was performed by Y.H. Homology model and docking simulations were performed by C.R. under supervision of S.I.; Y.T. wrote the overall story of the manuscript. The manuscript was edited by D.U., P.M., T.K., and T.O. All the authors discussed the manuscript. **Funding:** This work was supported by a Grant in Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology (15KT0031 and 15K07102 to Y.T. and 15H059556 to T.K.) and a grant from the Advanced Low Carbon Technology Research and Development Program from the Japan Science and Technology Agency (T.K.). Support for C.R. from the Japan Society for the Promotion of Science (JSPS) and the Alexander von Humboldt Foundation (AvH) is gratefully acknowledged. P.M. was funded by the Natural Sciences and Engineering Research Council of Canada (NSERC). ITbM is supported by the World Premier International Research Center Initiative (WPI), Japan. **Competing interests:** Nagoya University has filed for patents regarding the following topics: "Regulators for germination in *Striga* species," inventors Y.T., D.U., S.AM., S.H., M.Y., T.K., T.O., and K.I. (patent publication nos. WO 2017/002898 and JP 2017-014149); "Regulators for germination in parasitic plants," inventors Y.T., D.U., T.O., T.K., and K.K. (patent application nos. PCT/JP2018/36785 and JP 2017-193773). We declare no financial conflicts of interest in relation to this work. **Data and materials availability:** All data are available in the manuscript or the supplementary materials. The complete sets of raw data underlying all figures in the main text and supplement can be found in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/1301/suppl/DC1
Materials and Methods
Figs. S1 to S13
Table S1
References (*20*–*30*)

28 June 2018; accepted 25 October 2018
10.1126/science.aau5445

IMMUNOGENOMICS

High-affinity allergen-specific human antibodies cloned from single IgE B cell transcriptomes

Derek Croote[1], Spyros Darmanis[2], Kari C. Nadeau[3,4,5], Stephen R. Quake[1,2,6]*

Immunoglobulin E (IgE) antibodies protect against helminth infections but can also cause life-threatening allergic reactions. Despite their role in human health, the cells that produce these antibodies are rarely observed and remain enigmatic. We isolated single IgE B cells from individuals with food allergies and used single-cell RNA sequencing to elucidate the gene expression and splicing patterns unique to these cells. We identified a surprising example of convergent evolution in which IgE antibodies underwent identical gene rearrangements in unrelated individuals. Through the acquisition of variable region mutations, these IgE antibodies gained high affinity and unexpected cross-reactivity to the clinically important peanut allergens Ara h 2 and Ara h 3. These findings provide insight into IgE B cell transcriptomics and enable biochemical dissection of this antibody class.

Although the immunoglobulin E (IgE) antibody class is the least abundant of all isotypes in humans, it plays an important role in host defense against parasitic worm infections (*1*). It can also become misdirected toward otherwise harmless antigens, as in the case of food allergies, where the recognition of allergenic food proteins by IgE antibodies can lead to symptoms ranging from urticaria to potentially fatal anaphylaxis. Despite their central role in immunity and allergic disease, human IgE antibodies are scarce and remain poorly characterized (*2*). Recent studies have inferred IgE B cell characteristics and origins (*3*, *4*) and have described clonal families to which IgE antibodies belong (*5*). However, none have successfully isolated single IgE-producing cells or the paired heavy and light chain sequences that constitute individual IgE antibodies, leaving unanswered questions regarding the functional properties of such antibodies, the transcriptional programs of these cells, and the degree to which these features are shared across individuals. Here, we report the successful isolation and transcriptomic characterization of single IgE and IgG4 B cells from humans.

We performed plate-based single-cell RNA sequencing (scRNA-seq) on B cells isolated from peripheral blood of six food-allergic individuals (Fig. 1A). We used a simple fluorescence-activated cell sorting (FACS) strategy (fig. S2 and supplementary materials) that prioritized the capture of single B cells with surface IgE; we also included B cells of other isotypes for comparison. The isotype identity of each B cell was determined post hoc using the bioinformatic assembly of its heavy chain sequence from scRNA-seq reads. This allowed us to sacrifice specificity and capture IgE B cells with high sensitivity while avoiding stringent FACS gate purity requirements or the need for complex gating schemes. In total, 973 B cells were analyzed, of which 89 were IgE. We were unable to purify useful numbers of such cells from nonallergic controls.

Principal components analysis of normalized gene expression (fig. S3 and supplementary materials) separated B cells into two distinct clusters identifiable as plasmablasts (PBs) and naïve/memory B cells (Fig. 1, B and C). PBs expressed *PRDM1*, *XBP1*, and *IRF4*, which encode the triad of transcription factors that drive plasma cell differentiation (*6*). In contrast, naïve/memory B cells expressed *IRF8*, which encodes a transcription factor that antagonizes the PB fate (*7*), as well as *MS4A1*, which encodes the canonical mature B cell surface marker CD20. Additional FACS and gene expression data corroborated these B cell subsets (fig. S4).

Circulating IgE B cells overwhelmingly belonged to the PB subset (Fig. 1D and fig. S5A), which is in contrast to the other isotypes but consistent with the preferential differentiation of IgE B cells into PBs observed in mice (*8*). Notably, we found that the number of circulating IgE B cells for each individual correlated with total plasma IgE levels (fig. S1C). A similar phenomenon has been noted in atopic individuals and individuals with hyper-IgE syndrome (*9*).

Across all individuals, the 89 IgE antibodies we found varied widely in antibody heavy chain variable region (VH) gene usage as well as mutation frequency (Fig. 2A). They also varied in VH and light chain variable region (VL) complementarity-determining region 3 (CDR3) lengths (fig. S6A). There was moderate correlation between the VH and VL mutation frequency within single cells (fig. S6B), with evidence of selection via an enrichment of replacement mutations relative to silent mutations in VH and VL CDRs (fig. S6C). Relative to other isotypes, IgE B cells had a similar distribution of VH mutation frequency, use of λ versus κ light chains, and VH V and J gene usage (fig. S6, D to F).

A host of major histocompatibility complex (MHC) genes were robustly up-regulated in IgE PBs relative to PBs of other isotypes (Fig. 2B),

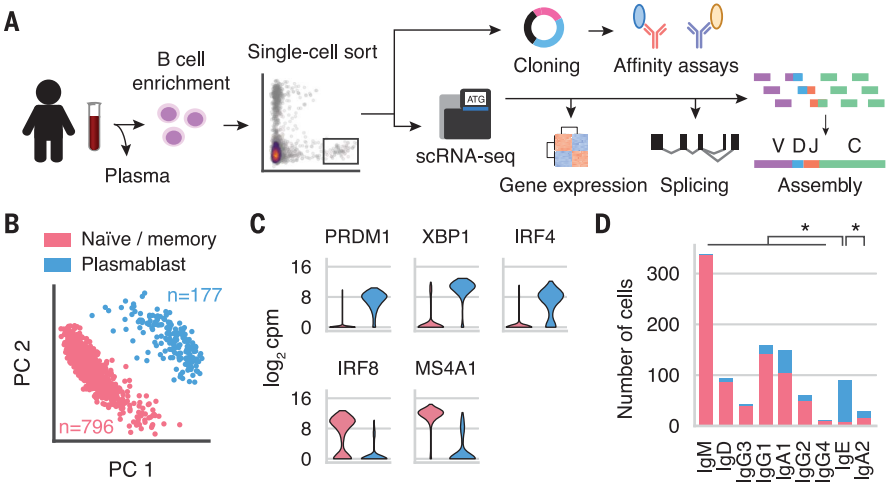


Fig. 1. Characterization of single B cells isolated from peripheral blood of food-allergic individuals. (**A**) Study overview. (**B** to **D**) Analysis of single cells pooled from all six individuals, n = 973. Cells colored by B cell subset. (B) Principal components analysis of single-cell gene expression separates B cells into two distinct clusters. (C) Gene expression distributions [$\log_2$ counts per million (cpm)] of established transcription factors and marker genes identify the clusters in (B) as naïve/memory (pink) and plasmablast (blue) B cell subsets. (D) Number of cells belonging to each subset by isotype. *$P < 10^{-5}$ between IgE and each other isotype (Fisher exact test).

[1]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA. [2]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA. [3]Sean N. Parker Center for Allergy and Asthma Research, Stanford University, Stanford, CA 94305, USA. [4]Department of Medicine, Stanford University, Stanford, CA 94305, USA. [5]Department of Pediatrics, Stanford University, Stanford, CA 94305, USA. [6]Department of Applied Physics, Stanford University, Stanford, CA 94305, USA.
*Corresponding author. E-mail: quake@stanford.edu

suggesting a more immature transcriptional program given the loss of MHC class II during the maturation of PBs to plasma cells (*10*). *FCER2*, which encodes the low-affinity IgE receptor CD23, was also highly up-regulated and coexpressed with *ADAM10* in 30% of IgE PBs, indicating that a subset of IgE PBs may secrete soluble CD23 (*11*). *LAPTM5*, which encodes a negative regulator of B cell activation and antibody production (*12*), was also up-regulated. Down-regulated genes included *LGALS1*, which supports plasma cell survival (*13*), and those encoding the S100 proteins S100A4, S100A6, and S100A10, which may indicate reduced proliferative and survival signaling (*14*, *15*). One of the most significantly down-regulated genes in IgE PBs encodes spleen-associated tyrosine kinase (SYK), which plays an essential role in B cell development, activation, survival, and differentiation (*16*). Thus, the IgE PB cell state is immature relative to other PBs with weakened activation, proliferation, and survival capacity. This suggests a potential mechanism for the short-lived IgE PB phenotype described in murine models of allergy (*17*).

Human IgE B cells belonging to the naïve/memory subset were deficient in immunoglobulin heavy chain membrane IgE (mIgE) transcripts, as evidenced by a lack of membrane exon splicing relative to other common isotypes. Furthermore, membrane exon splicing was detected in significantly fewer IgE PBs than non-IgE PBs (Fig. 2, C and D). The lack of mature mIgE transcripts, which could be explained by atypical polyadenylation signals that lead to poor processing of pre-mRNA (*18*), is consistent with low IgE B cell receptor levels measured by others (*3*) and low relative IgE surface protein levels we observed by FACS. Indeed, mIgE surface protein levels on IgE B cells did not exceed those of some non-IgE B cells, which presumably display surface IgE as a result of CD23-mediated capture (fig. S2B).

By clustering cells into clonal families (CFs) according to the similarity of their antibody VH sequences (*19*), we were able to observe elements of classical germinal center phenomena such as somatic hypermutation, class switching, and fate determination (Fig. 3). Only 49 cells formed CFs with multiple members (fig. S5B), which was unsurprising given the vast diversity of potential immunoglobulin gene rearrangements. Overall, these CFs contained two to six sequences, had variable isotype membership, and had a comprehensive distribution of VH mutation frequency. Four CFs illustrated the two possible B cell differentiation pathways in that they contained both PBs and memory B cells, whereas other CFs contained cells belonging to multiple isotypes. Notably, we also found that in contrast to other isotypes, IgE and IgG4 showed higher proportional membership in CFs (fig. S5C).

Surprisingly, we identified one CF (CF1) comprising cells belonging to multiple individuals: Three were IgE PBs from individual PA12 and three were IgE PBs from individual PA13 (Fig. 3). The antibodies produced by these six cells were highly similar in VH and VL sequences (Fig. 4A and fig. S7, A and B), and all used the IGHV3-30*18 and IGHJ6*02 VH genes as well as the IGKV3-20*01 and IGKJ2*01 VL genes. These antibodies were also among the most mutated of all class-switched antibodies in our dataset and were enriched in replacement mutations within the VH and VL CDRs (fig. S7, C and D).

We cloned and expressed the six IgE antibodies belonging to this convergent CF in order to assess whether they bind the natural forms of the major allergenic peanut (*Arachis hypogaea*) proteins Ara h 1, Ara h 2, or Ara h 3. Surprisingly, all six antibodies were cross-reactive: They bound strongly to Ara h 2, moderately to Ara h 3, and very weakly to Ara h 1 (Fig. 4B). Furthermore, these antibodies have high affinity; dissociation constants determined by biolayer interferometry for Ara h 2 and Ara h 3 were as low as picomolar and subnanomolar, respectively (Fig. 4, C and D, and fig. S8). These affinities are comparable to some of the highest-affinity native human antibodies against pathogens such as HIV, influenza, and malaria (*20*–*22*).

We also cloned and expressed eight engineered variants of IgE antibody PA13P1H08 to assess the effects of VH and VL mutations on allergen binding. Retaining the actual VH while swapping the VL with another κ VL from an antibody without peanut allergen specificity abrogated binding to both allergenic proteins, whereas reverting both VH and VL to the inferred naïve sequences (fig. S7, E to G) largely eliminated Ara h 3 binding and markedly reduced Ara h 2 affinity (Fig. 4D and fig. S8C). Reverting only the VH or VL reduced the affinity to Ara h 2 and Ara h 3, but disproportionately. We also found a synergistic contribution of VH mutations to affinity through independent reversion of the VH CDR1, CDR2,

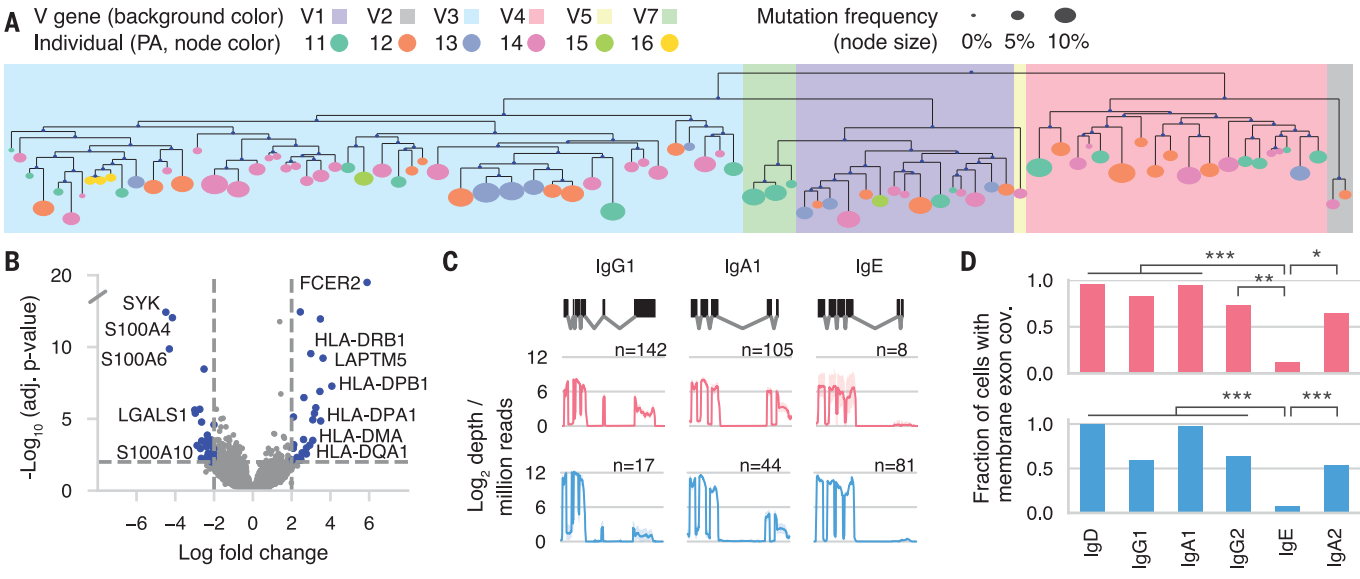


Fig. 2. Characterization of 89 IgE antibodies and the single B cells that produce them. (**A**) Phylogenetic depiction of antibody heavy chain variable region (VH) sequences arranged by VH V gene (background color), individual of origin (node color), and VH mutation frequency (node size). (**B**) Differential gene expression between IgE PBs (n = 81) and PBs of other isotypes (n = 96). Positive log fold change indicates genes enriched in IgE PBs. (**C**) Heavy chain constant region gene coverage histograms for naïve/memory B cells (top) and PBs (bottom) for select isotypes. Mean normalized read depth and 95% confidence interval are indicated by solid lines and shaded area, respectively, for the number of cells (n) inscribed. Heavy chains are oriented in the 5′ to 3′ direction and membrane exons are the two most 3′ exons of each isotype. (**D**) Summary of (C), but depicting the fraction of cells of each isotype with any membrane exon coverage for naïve/memory B cells (top) and PBs (bottom) and isotypes with at least five cells of each subset. *P < 0.05, **P < 0.005, ***P < 0.0005 between IgE and other isotypes (Fisher exact test).

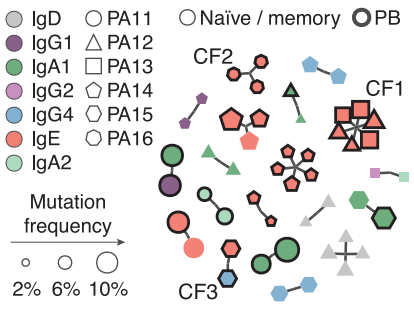


Fig. 3. Clonal families (CFs) capture B cell phenomena relevant to allergic disease. For each cell (node), the isotype (color), B cell subset (outline thickness), individual of origin (shape), and VH mutation frequency (size) are illustrated. CFs referred to in the text are labeled.

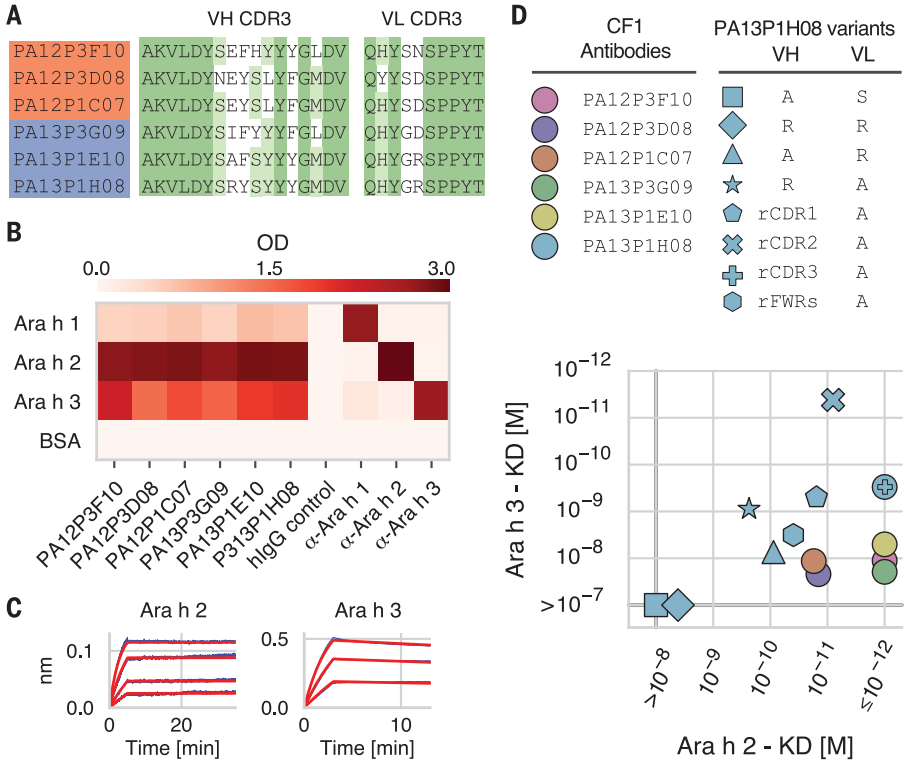


Fig. 4. High-affinity cross-reactive human IgE antibodies belonging to CF1. (**A**) Highly similar VH and VL CDR3s depict convergent evolution in two unrelated individuals (PA12 and PA13). Positions with >50% conservation are shaded. Amino acid abbreviations: A, Ala; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; Y, Tyr. (**B**) Indirect enzyme-linked immunosorbent assay depicting antibody cross-reactivity to multiple peanut allergens. Commercially available mouse monoclonal α-Ara h antibodies served as positive controls. OD, optical density; hIgG, human IgG. (**C**) Biolayer interferometry was used to determine antibody dissociation constants (K_Ds). Shown are binding curves for PA13P1H08 against Ara h 2 and Ara h 3. (**D**) Ara h 2 and Ara h 3 K_Ds for each CF1 antibody as well as eight engineered variants of PA13P1H08. For each variant, the VH and/or VL was either the actual sequence from PA13P1H08 ("A"), reverted to the inferred naïve sequence ("R"), swapped with another non–peanut-specific sequence ("S"), or had only specific region(s) of the sequence reverted ("r"). FWRs, framework regions.

CDR3, and framework regions. Interestingly, reversion of the VH CDR2 increased Ara h 3 affinity while only marginally decreasing Ara h 2 affinity. Thus, although the inferred naïve antibody is capable of binding the most clinically relevant peanut allergen Ara h 2 (*23*), mutations in both VH and VL are necessary to produce the high-affinity and cross-reactive antibodies that we found in circulating IgE PBs of unrelated individuals.

We also cloned and expressed antibodies from two other CFs. CF2 contained three IgE PBs from individual PA16 (two of which were identical), but these antibodies did not bind Ara h 1, 2, or 3, which was unsurprising given that this individual had low plasma peanut-specific IgE levels as well as IgE specific to other allergens (fig. S1). In contrast, CF3 contained an IgE PB (PA15P1D05) and IgG4 PB (PA15P1D12) from individual PA15. These antibodies did not bind Ara h 1 appreciably, but bound Ara h 3 with nanomolar affinity and Ara h 2 with subnanomolar affinity (fig. S8). Notably, these two antibodies used the same VL V gene and a highly similar VH V gene (IGHV3-30-3*01) as the six convergent antibodies of CF1.

Our transcriptomic characterization of circulating human IgE B cells suggests that an immature IgE PB gene expression program indicative of weakened activation, proliferation, and survival capacity contributes to the short-lived phenotype of these cells. Additionally, the absence of mIgE transcript expression supports the hypothesis that impaired membrane IgE expression compromises IgE B cell entry into the memory compartment and/or memory B cell survival, therefore causing the scarcity of circulating memory IgE B cells in vivo. These results show that the human IgE system shares many important features with that of the mouse. Studies of mIgE signaling and IgE memory in murine models of allergy (*24*, *25*) are therefore likely relevant for human disease.

Isolating single IgE and IgG4 B cells also provides insight into the antibodies they produce. We discovered a striking case of antibody convergence, where two unrelated individuals produced high-affinity cross-reactive peanut-specific IgE antibodies comprising identical gene rearrangements within respective VHs and VLs. Convergent antibody evolution is believed to occur in response to a number of pathogens such as influenza (*26*) and HIV (*22*). Although our results offer a single additional example, another study of peanut-allergic individuals (*27*) reported IgE VH sequences that used identical V and J genes and shared at least 70% CDR3 identity with one or more of the six convergent antibodies in our dataset (fig. S9).

We discovered high-affinity IgE antibodies with cross-reactivity to two major peanut allergens and demonstrated that these properties originated from the acquisition of mutations within the VH and VL. Interestingly, although Ara h 2 and Ara h 3 belong to two distinct protein families, cross-inhibition experiments with purified allergens and plasma IgE have shown that this cross-reactivity may be common within peanut-allergic individuals (*28*). We also found an example within one individual of in vivo competition between peanut-specific IgE and IgG4 antibodies. Further study of such processes has the potential to increase our understanding of the contribution of IgG4 to the reduced clinical allergen reactivity that accompanies immunotherapy and early allergen exposure (*29*). Lastly, we anticipate that either these antibodies or engineered variants could be used as therapeutic agents. Recent clinical results have shown that engineered allergen-specific IgG antibodies provide effective treatment for cat allergies, perhaps by outcompeting native IgE for antigen (*30*).

REFERENCES AND NOTES

1. C. M. Fitzsimmons, F. H. Falcone, D. W. Dunne, *Front. Immunol.* **5**, 61 (2014).
2. J. Eckl-Dorna, V. Niederberger, *Curr. Allergy Asthma Rep.* **13**, 281–287 (2013).
3. M. A. Berkowska *et al.*, *J. Allergy Clin. Immunol.* **134**, 688–697.e6 (2014).
4. T. J. Looney *et al.*, *J. Allergy Clin. Immunol.* **137**, 579–586.e7 (2016).
5. R. A. Hoh *et al.*, *J. Allergy Clin. Immunol.* **137**, 157–167 (2016).
6. S. L. Nutt, P. D. Hodgkin, D. M. Tarlinton, L. M. Corcoran, *Nat. Rev. Immunol.* **15**, 160–171 (2015).

7. H. Xu *et al.*, *Nat. Immunol.* **16**, 1274–1281 (2015).
8. A. Erazo *et al.*, *Immunity* **26**, 191–203 (2007).
9. A. Horst *et al.*, *Clin. Exp. Immunol.* **130**, 370–378 (2002).
10. K. L. Calame, K.-I. Lin, C. Tunyaplin, *Annu. Rev. Immunol.* **21**, 205–230 (2003).
11. G. Weskamp *et al.*, *Nat. Immunol.* **7**, 1293–1298 (2006).
12. R. Ouchida, T. Kurosaki, J.-Y. Wang, *J. Immunol.* **185**, 294–301 (2010).
13. A. Anginot, M. Espeli, L. Chasson, S. J. C. Mancini, C. Schiff, *J. Immunol.* **190**, 5526–5533 (2013).
14. W. Leśniak, Ł. P. Słomnicki, A. Filipek, *Biochem. Biophys. Res. Commun.* **390**, 1087–1092 (2009).
15. K. Boye, G. M. Maelandsmo, *Am. J. Pathol.* **176**, 528–535 (2010).
16. J. A. Ackermann *et al.*, *J. Immunol.* **194**, 4650–4656 (2015).
17. Z. Yang, B. M. Sullivan, C. D. C. Allen, *Immunity* **36**, 857–872 (2012).
18. A. Karnowski, G. Achatz-Straussberger, C. Klockenbusch, G. Achatz, M. C. Lamers, *Eur. J. Immunol.* **36**, 1917–1925 (2006).
19. G. Georgiou *et al.*, *Nat. Biotechnol.* **32**, 158–168 (2014).
20. J. Wrammert *et al.*, *Nature* **453**, 667–671 (2008).
21. R. Murugan *et al.*, *Sci. Immunol.* **3**, eaap8029 (2018).
22. J. F. Scheid *et al.*, *Science* **333**, 1633–1637 (2011).
23. S. J. Koppelman, M. Wensing, M. Ertmann, A. C. Knulst, E. F. Knol, *Clin. Exp. Allergy* **34**, 583–590 (2004).
24. B. Laffleur, O. Debeaupuis, Z. Dalloul, M. Cogné, *Front. Immunol.* **8**, 1277 (2017).
25. G. Achatz, L. Nitschke, M. C. Lamers, *Science* **276**, 409–411 (1997).
26. K. J. L. Jackson *et al.*, *Cell Host Microbe* **16**, 105–114 (2014).
27. K. Kiyotani *et al.*, *J. Hum. Genet.* **63**, 239–248 (2018).
28. M. Bublin *et al.*, *J. Allergy Clin. Immunol.* **132**, 118–124.e12 (2013).
29. L. Tordesillas, M. C. Berin, H. A. Sampson, *Immunity* **47**, 32–50 (2017).
30. J. M. Orengo *et al.*, *Nat. Commun.* **9**, 1421 (2018).

ACKNOWLEDGMENTS

We thank F. Horns and L. Penland for helpful discussions, and J. Okamoto and N. Neff for assistance with sequencing. **Funding:** Supported by the Simons Foundation (SFLIFE #288992 to S.R.Q.), the Chan Zuckerberg Biohub, and the Sean N. Parker Center for Allergy and Asthma Research at Stanford University. D.C. is supported by an NSF Graduate Research Fellowship and the Kou-I Yeh Stanford Graduate Fellowship. **Author contributions:** D.C., S.D., K.C.N., and S.R.Q. designed the research; K.C.N. contributed to sample collection; D.C. performed the research; D.C., K.C.N., and S.R.Q. analyzed the data; and D.C. and S.R.Q. wrote the manuscript with input from all authors. **Competing interests:** The authors are inventors on a patent application (number 62/673,713) covering aspects of this work. The authors declare no other competing interests. **Data and materials availability:** Processed data including antibody assembly information and the gene expression count matrix are available in the supplementary materials. Raw sequencing data are available from the Sequence Read Archive (SRA) under BioProject accession PRJNA472098. Code associated with this manuscript is available at https://github.com/dcroote/singlecell-ige. All other data needed to evaluate the conclusions in this paper are present either in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/1306/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S3
References (*31*–*44*)

25 May 2018; accepted 19 October 2018
10.1126/science.aau2599

MAIZE DOMESTICATION

Multiproxy evidence highlights a complex evolutionary legacy of maize in South America

Logan Kistler[1,2]*, S. Yoshi Maezumi[3,4], Jonas Gregorio de Souza[3], Natalia A. S. Przelomska[1,5], Flaviane Malaquias Costa[6], Oliver Smith[7], Hope Loiselle[1,8], Jazmín Ramos-Madrigal[7], Nathan Wales[9], Eduardo Rivail Ribeiro[1], Ryan R. Morrison[2], Claudia Grimaldo[10], Andre P. Prous[11], Bernardo Arriaza[12], M. Thomas P. Gilbert[7,13], Fabio de Oliveira Freitas[14]*, Robin G. Allaby[2]*

Domesticated maize evolved from wild teosinte under human influences in Mexico beginning around 9000 years before the present (yr B.P.), traversed Central America by ~7500 yr B.P., and spread into South America by ~6500 yr B.P. Landrace and archaeological maize genomes from South America suggest that the ancestral population to South American maize was brought out of the domestication center in Mexico and became isolated from the wild teosinte gene pool before traits of domesticated maize were fixed. Deeply structured lineages then evolved within South America out of this partially domesticated progenitor population. Genomic, linguistic, archaeological, and paleoecological data suggest that the southwestern Amazon was a secondary improvement center for partially domesticated maize. Multiple waves of human-mediated dispersal are responsible for the diversity and biogeography of modern South American maize.

Maize (*Zea mays* ssp. *mays*) evolved from wild Balsas teosinte (*Z. mays* ssp. *parviglumis*, hereafter *parviglumis*) in modern-day lowland Mexico beginning around 9000 years ago (*1*) and spread to dominate food production systems throughout much of the Americas by the beginning of European colonization in the 15th century. Archaeological and genetic data from ancient DNA studies have highlighted aspects of maize natural history, including the evolution and fixation of agricultural traits and adaptation of maize to diverse new environments (*2*–*6*). Archaeological remains establish that maize was brought to the southwestern United States and the Colorado Plateau by ~4000 years before the present (yr B.P.) (*7*), traversing Panama by ~7500 yr B.P. (*8*) and arriving in Coastal Peru (*9*), the Andes (*10*), and lowland Bolivian Amazon (*11*) between ~6500 and 6300 yr B.P. (Fig. 1 and table S1). Today, maize is a staple food species, yielding over 6% of all food calories for humans, plus more in livestock feed and processed foods (*12*).

Maize domestication is thought to have occurred once, with little subsequent gene flow from *parviglumis* (*13*, *14*). However, archaeogenomic evidence reveals maize was only partially domesticated in Mexico by ~5300 yr B.P. (*2*, *3*), carrying a mixture of wild-type and maize-like alleles at loci involved in the domestication syndrome. For example, the domestic-type *TGA1* gene variant responsible for eliminating the tough teosinte fruitcase was already present by this time period (*2*), whereas other loci associated with changes to seed dispersal and starch production during domestication still carried wild-type variants (*2*, *3*). The state of partial domestication sets these archaeogenomes apart from modern fully domesticated maize, which carries a complete, stable set of domestication alleles conferring the domesticated phenotype. This partially domesticated maize was grown in Mexico well after maize had become established in South America, which raises the question of how South American maize came to possess the full complement of fixed domestication traits. To reconcile archaeobotanical and genomic data concerning the domestication and dispersal history of maize in South America, we sequenced maize genomes from 40 indigenous landraces and 9 archaeological samples from South America (Fig. 1 and tables S2 and S3) and analyzed them alongside published modern (n = 68) and ancient (n = 2) maize and teosinte genomes (*15*).

Model-based clustering highlights extensive admixture and population overlap between maize populations, but we observe several robust lineages (*15*) (Fig. 1): (i) the Andes and the Pacific coast of South America; (ii) lowland South America, including the Amazon and Brazilian Savanna; (iii) North America north of the domestication center; and (iv) highland Mexico and Central America, previously observed to contain introgression from wild *Z. mays* ssp. *mexicana* (*14*, *16*). We also observe a widespread "Pan-American" lineage spanning from northern Mexico into lowland South America. In a previous analysis based on multiple nuclear microsatellites, maize formed a monophyletic subset of teosinte, with South American lineages as the most derived elements in a phylogenetic tree (*13*). This pattern has been interpreted as evidence for a single episode of domestication followed by dispersal culminating in the Andes after maize became established throughout the rest of the range of cultivation (*13*). However, archaeological evidence for persistent maize cultivation indicates it was established in numerous locations throughout South America by ~6500 to 4000 yr B.P. regionally. On the basis of this information, we propose that South American maize was carried away from the Mesoamerican domestication center soon after initial stages of domestication and may have been one of several partially domesticated maize lineages that independently fissioned from the primary gene pool after the onset of domestication in Mexico (Fig. 2).

Using f_4 statistics (*17*), we observe asymmetry in *parviglumis* ancestry among modern maize populations (Fig. 2). This reveals that maize-*parviglumis* gene flow was ongoing in some lineages after others became reproductively isolated. Whereas later gene flow from *Z. mays* ssp. *mexicana*, a highland subspecies of teosinte, is well documented in some maize (*6*, *14*, *16*), this finding contradicts the assumption that dispersal and diversification throughout the Americas happened only after the severance of gene flow from *parviglumis* (*13*, *14*). Thus, while South American maize became reproductively isolated from the wild progenitor when it was carried away from the domestication center, maize lineages remaining in Mexico underwent continued crop-wild gene flow before diversifying into extant landraces over subsequent millennia. The Pan-American lineage shows excess shared ancestry with *parviglumis* relative to all other major groups (Fig. 2B), suggesting that this group emerged from the domestication center and dispersed after other maize lineages became regionally established. Because the Pan-American lineage carries excess *parviglumis* ancestry relative to the strictly South American lineages, it appears to represent a second episode of maize dispersal from Mesoamerica, reinforcing two major waves of maize movement into South America as previously suggested (*5*).

[1]Department of Anthropology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA. [2]Department of Life Science, University of Warwick, Coventry CV4 7AL, UK. [3]Department of Archaeology, College of Humanities, University of Exeter, Laver Building, North Park Road, Exeter EX4 4QE, UK. [4]Department of Geography and Geology, The University of the West Indies, Mona Campus, Kingston, Jamaica. [5]Center for Conservation Genomics, Smithsonian Conservation Biology Institute, National Zoo, Washington, DC 20008, USA. [6]University of São Paulo, Escola Superior de Agricultura Luis de Queiroz, Piracicaba, SP 13418-900, Brazil. [7]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark. [8]Department of Anthropology, University of Washington, Denny Hall 314, Seattle, WA 98195, USA. [9]Department of Archaeology, University of York, King's Manor, York YO1 7EP, UK. [10]Department of Oncology, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford, OX3 7DQ, UK. [11]Museu de Historia Natural e Jardim Botânico da Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil. [12]Instituto de Alta Investigación, Universidad de Tarapacá, Arica, Chile. [13]Norwegian University of Science and Technology, University Museum, 7491 Trondheim, Norway. [14]Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, CEP 70770-901, Brazil.
***Corresponding author. Email: kistlerl@si.edu (L.K.); fabio.freitas@embrapa.br (F.O.F.); r.g.allaby@warwick.ac.uk (R.G.A.)**

The genomes of two ancient maize cobs from the Tehuacan Valley of Mexico at ~5300 yr B.P. recently revealed a state of partial domestication, a mixture of maize- and *parviglumis*-like alleles at loci involved in domestication (*2*, *3*). This is puzzling, given the sustained use of domesticated maize from ~6500 yr B.P. onward in South America (Fig. 1 and table S1) (*11*, *18*). However, principal components analysis and f_3 statistics reveal considerable genomic distance between these two Mesoamerican archaeogenomes (Fig. 1 and fig. S2), and f_3 statistics confirm that the SM10 genome (*3*) is more maize-like, whereas the Tehuacan162 genome (*2*) is more *parviglumis*-like (fig. S2). In total, the two genomes are from the same region and time period, and both are partially domesticated, but otherwise, they appear to represent independent samples out of a diverse semidomesticated population containing an array of domestic and wild-type alleles.

Given the state of partial domestication observed in the Tehuacan and San Marcos genomes (*2*, *3*), early South American maize emerging from their common ancestral population would likely also have been a partially domesticated form of maize containing an assortment of wild and domestic alleles. This ancestral population likely harbored the building blocks for fully domesticated maize but lacked the allelic fixation and linkage of the modern domesticated crop. We expect that in this ancestral semidomesticated population, domestication loci under ongoing selection would have been continually decoupled from their chromosomal neighborhood through recombination (*19*, *20*), resulting in an enrichment of the original *parviglumis* genomic background near domestication genes relative to its genome-wide retention. If the domestication syndrome was fully established in the common ancestor of all extant maize, no modern *parviglumis* genome should carry this enriched affinity to domestication loci to differing degrees in different maize lineages, because the same background would have become fixed in their common ancestor. However, if South American maize became isolated while fundamental domestication was still ongoing, as we hypothesize, then components of the *parviglumis* genomic background are expected to differ between early stratified maize lineages. Therefore in this case, modern *parviglumis* genomes would carry a specifically South American or non–South American affinity for the enriched wild-type background near domestication loci.

We compared *D*-statistics (*21*) across the whole genome (D_{WG}) and within 10 kb of 186 known domestication loci (D_{dom}) to test for these asymmetrical *parviglumis* contributions between pairs of extant South American and non–South American maize around domestication genes (*15*). We found that *parviglumis* enrichment associated with domestication is highly patterned among major ancestry groups, with several *parviglumis* genomes associated exclusively with either South American or non–South American D_{dom} enrichment and a significant association with ancestry overall (Fig. 2C; χ^2 test $P = 2.74 \times 10^{-6}$). That is,

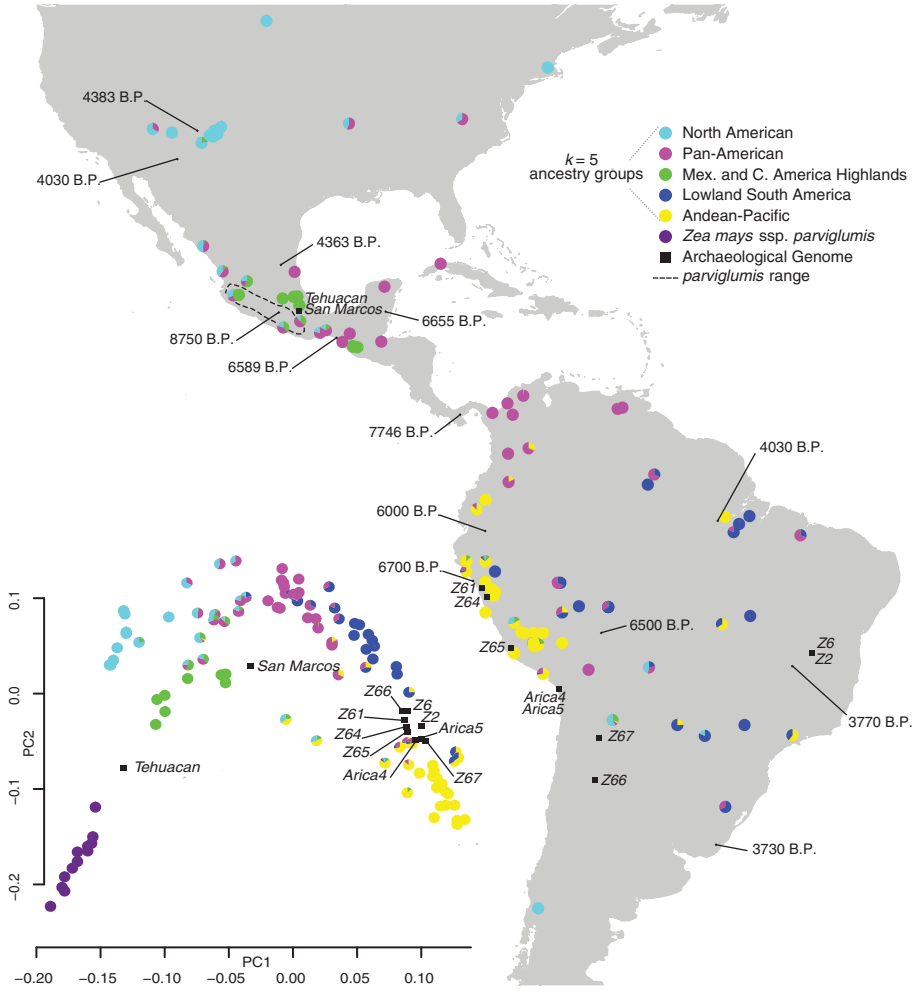


Fig. 1. Distribution and ancestry proportions of maize genomes and principal components analysis (PCA) of maize and *parviglumis* genomes. Pie colors reflect ancestral proportions estimated by means of model-based clustering (k = 5) of modern maize genomes (*15*). Archaeological genomes were projected onto the PCA to mitigate degradation biases (*15*). Dates reflect early regional maize archaeobotanical remains (table S1 and fig. S1). C., Central; Mex., Mexico; PC1, First principal component; PC2, second principal component.

we observe that *parviglumis* ancestry is enriched near domestication genes in a pattern demonstrating that domestication-associated selection was still ongoing after the stratification of the major extant lineages from their semidomesticated ancestral population. This pattern validates a model in which the ancestral population in South America was itself only partially domesticated during its dispersal away from the domestication center.

In total, we find support for a model of stratified domestication in maize (Fig. 2). The initial stages of maize domestication likely occurred only once within a diverse wild Balsas River basin gene pool, as previously suggested (*13*). However, before the domestication syndrome was fixed and stable, multiple lineages separated, and selection pressures on domestication loci continued independently outside of the primary domestication center. Some of these divergent semidomesticated populations likely led to terminal lineages lacking sufficient diversity and ecological context to continue the domestication process. Others, like ancestral South American maize, evolved into fully domesticated lineages under continuing anthropogenic pressures.

The earliest evidence places maize in the southwestern Amazon by ~6500 yr B.P. (*11*), a region serving as a geographic interface of the lowland and Andean-Pacific genetic lineages (Fig. 1). We hypothesize that the southwestern Amazon may have been a secondary improvement center for the partially domesticated crop before the divergence of the two South American groups. When maize arrived, southwestern Amazonia was a plant domestication hotspot (*22*). Additionally, microfossil assemblages (*11*, *22*) reveal the presence of polyculture (mixed cropping) from ~6500 yr B.P. onward, such that a new crop species could be integrated into existing food production systems supporting domestication activities.

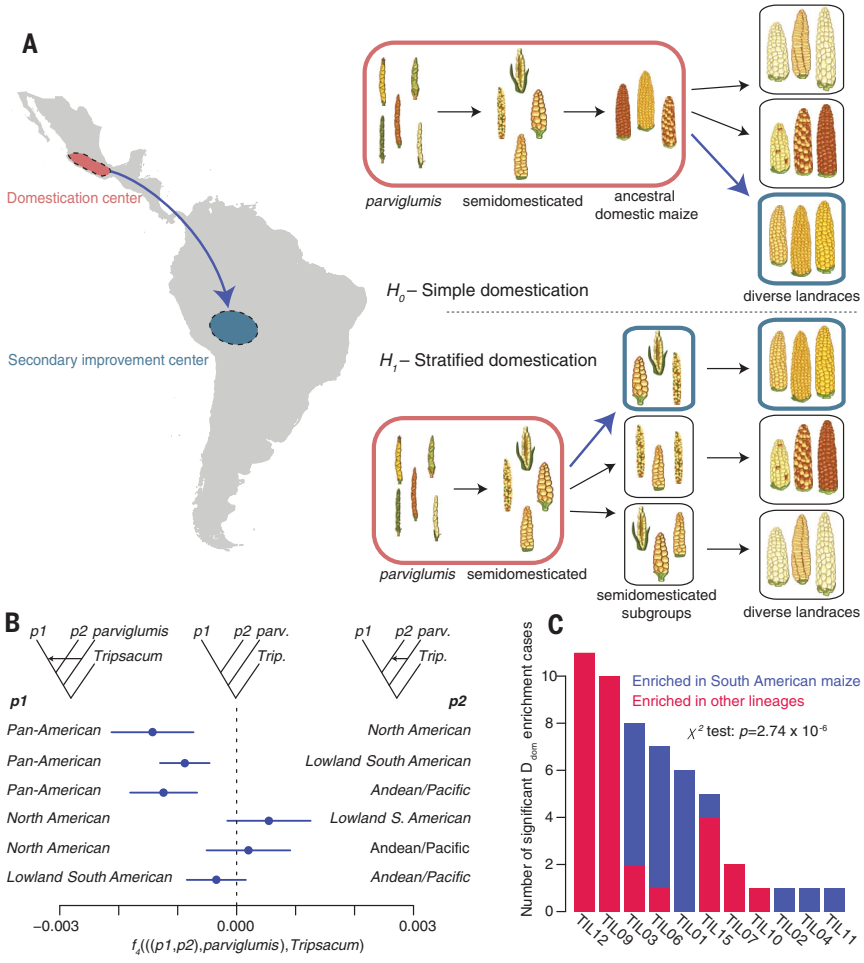


Fig. 2. A stratified domestication model for maize. (**A**) Schematic comparing the conventional domestication model under which maize became fully domesticated and then dispersed throughout the Americas, versus a stratified domestication model in which partially domesticated subpopulations became reproductively isolated before the fixation of the domestication syndrome. (**B**) f_4 statistics demonstrating excess allele sharing between the Pan-American lineage and wild *parviglumis* compared with other maize, revealing nonuniform crop-wild gene flow after initial domestication. Bars are three standard errors under a block jackknife (*15*). (**C**) Bar plot of enriched *parviglumis* contributions to ancestry near domestication genes, in which each bar is a *parviglumis* genome contributing to South American maize (blue) or other maize (red) D_{dom} enrichment. Geographic segregation in D_{dom} enrichment among *parviglumis* genomes suggests that the domestication syndrome was not yet fixed in a common domesticated ancestor of modern maize.

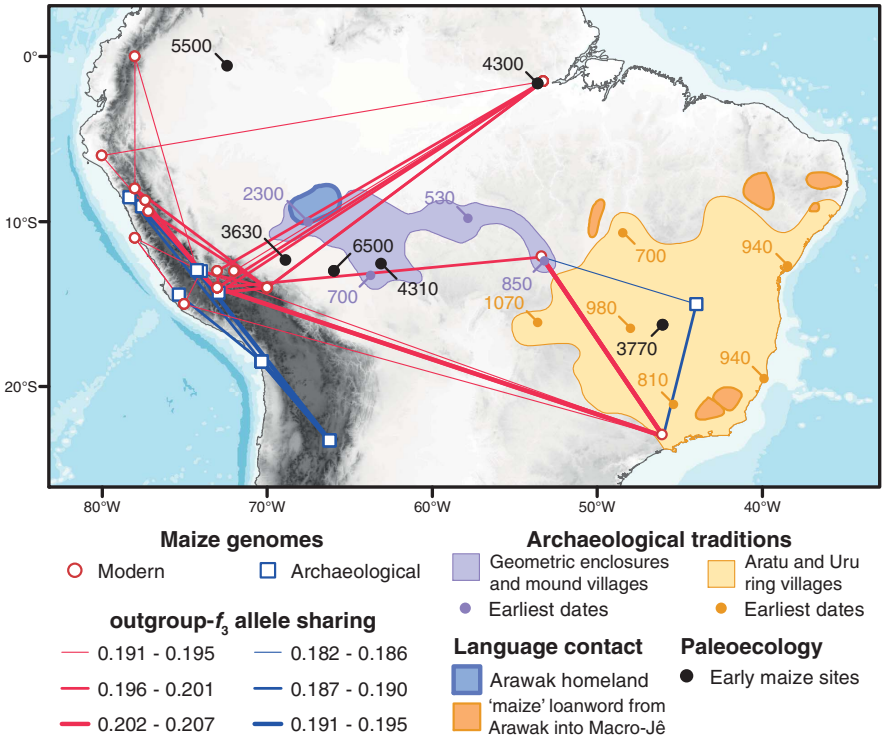


Fig. 3. Genomic relatedness overlapping linguistic and archaeological patterns in lowland South America. Maize genomes with ≥50% Andean-Pacific ancestry and ≥99% South American ancestry are connected by lines with the two other genomes with which they share the highest outgroup-f_3 value. Geometric enclosures and mound ring villages of southern Amazonia broadly coincide with the expansion of Arawak languages, whereas the Uru and Aratu ring villages coincide with the distribution of Macro-Jê languages (*15*) (figs. S3 and S4). Only the earliest regional dates for each archaeological tradition are shown (see table S4). Macro-Jê languages borrowing an Arawak loanword for "maize" are based on (*24*). Arawak homeland is shown approximately in the modern location of Apurinã, in accordance with (*29*).

Pollen and phytolith data demonstrate a west-to-east pattern of maize expansion across the Amazon and show that maize was consistently present from ~4300 yr B.P. onward in the eastern Amazon (*18*). Initially, maize in the eastern Amazon was part of a polyculture agroforestry system combining annual crop cultivation with wild resource use and low-level management through burning (*18*). Maize cultivation proceeded alongside the progressive enrichment of edible forest species and subsequent waves of new crop arrivals, including sweet potato (~3200 yr B.P.), manioc (~2250 yr B.P.), and squash (~600 yr B.P.). The development of anthropogenically enriched Amazonian Dark Earth soils ~2000 yr B.P. (*23*) enabled the expansion and intensification of maize cultivation, likely increasing carrying capacity to sustain growing populations in the eastern Amazon (*18*). The extant endemic maize lineage in lowland South America likely originated with this long-term process involving millennia of evolving land-use practices.

Several landraces and two archaeogenomes (~700 yr B.P.) in eastern Brazil also show strong genetic links to Andean maize near the southwestern Amazon (Fig. 3). This pattern closely mirrors linguistic patterns linking Andean, Amazonian, and eastern Brazilian maize cultivation and suggests a second major west-to-east cultural expansion of maize traditions. A loanword for maize with possible Andean origins was transmitted from Amazonian Arawak languages—most likely originating in southwest Amazonia (*24*)—into Macro-Jê stock languages in the Brazilian savanna and Atlantic coast (*24*) (fig. S3). Archaeological evidence suggests this expansion occurred ~1200 to 1000 yr B.P. with the spread of a cultural horizon of geometric enclosures and mound ring villages throughout southern Amazonia and ring villages in the central Brazilian savannas and the Atlantic coast (Fig. 3 and fig. S4) (*25*–*27*). This process is roughly contemporaneous with archaeological Andean-admixed genomes in the area. Thus, Arawak speakers likely brought nonlocal Andean-Pacific maize lineages into a landscape where maize was an established component of long-term land management and food production strategies.

Finally, we quantified the mutation load in maize genomes—the accumulation of potentially deleterious alleles due to drift and selection (*16*)—using a phylogenetic framework to estimate evolutionary constraint (*15*). We observe that South American lineages carry a higher mutation load than other maize lineages. Mutation load increases linearly with distance from the domestication center and is linked with ancestry, and the Andean-Pacific group carries the highest burden of potentially deleterious variants (Fig. 4) (*15*). The mutation load in the Andes has been attributed to selection for high-altitude adaptations (*16*), but the elevated mutation load in lowland maize also suggests a history of shared selection and drift effects prior to highland adaptation. These processes would likely have included a founder episode as maize was carried into South America, persistent selection pressures for regional adaptation, and the latter stages of domestication after isolation from the founding gene pool. We also find that Andean and Pacific maize from ~1000 yr B.P. to the early colonial period has a low mutation load compared with its modern Andean-Pacific counterparts (Wilcoxon $P = 0.00247$) (*15*) (Fig. 4); although still elevated compared with non–South American lineages. It is possible that Andean maize experienced a wave of deleterious allele accumulation as human and crop populations were disrupted by changes caused by the arrival of Europeans (*28*). Alternatively, the increasing mutation load in modern crops could represent the ongoing effects of burdensome allele accumulation over nine millennia of human intervention.

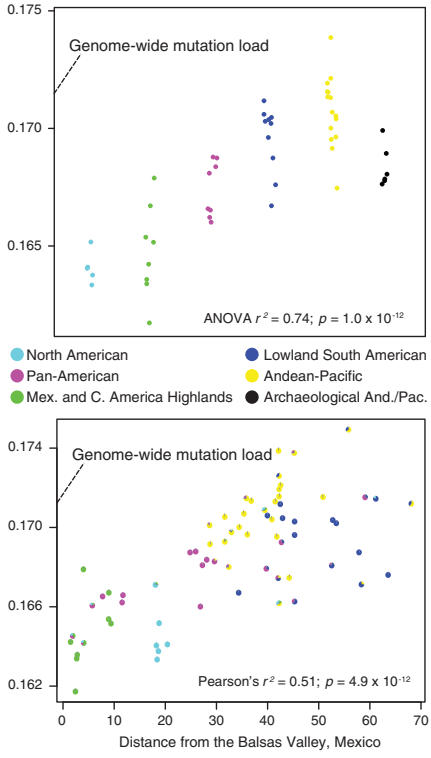


Fig. 4. Genome-wide mutation load across ancestry groups (non-admixed samples only in top panel) and load compared with distance to the domestication center. Mutation load is calculated as a proportion of the theoretical maximum load over observed single-nucleotide polymorphisms, and ancient load scores are rescaled for missingness using a Procrustes transformation (*15*). Euclidean distance in degrees to the Balsas River valley is shown. And./Pac., Andean-Pacific.

REFERENCES AND NOTES

1. D. R. Piperno, A. J. Ranere, I. Holst, J. Iriarte, R. Dickau, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 5019–5024 (2009).
2. J. Ramos-Madrigal *et al.*, *Curr. Biol.* **26**, 3195–3201 (2016).
3. M. Vallebueno-Estrada *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14151–14156 (2016).
4. V. Jaenicke-Després *et al.*, *Science* **302**, 1206–1208 (2003).
5. F. O. Freitas, G. Bendel, R. G. Allaby, T. A. Brown, *J. Archaeol. Sci.* **30**, 901–908 (2003).
6. R. R. da Fonseca *et al.*, *Nat. Plants* **1**, 1–5 (2015).
7. W. L. Merrill *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21019–21026 (2009).
8. D. R. Piperno, K. H. Clary, R. G. Cooke, A. J. Ranere, D. Weiland, *Am. Anthropol.* **87**, 871–878 (1985).
9. A. Grobman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 1755–1759 (2012).
10. M. B. Bush *et al.*, *Quat. Sci. Rev.* **141**, 52–64 (2016).
11. S. O. Brugger *et al.*, *Quat. Sci. Rev.* **132**, 114–128 (2016).
12. F. A. O. of the United Nations, FAOSTAT statistics database (2018); www.fao.org/faostat/.
13. Y. Matsuoka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6080–6084 (2002).
14. J. van Heerwaarden *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1088–1092 (2011).
15. Supplementary materials are available online.
16. L. Wang *et al.*, *Genome Biol.* **18**, 215 (2017).
17. N. Patterson *et al.*, *Genetics* **192**, 1065–1093 (2012).
18. S. Y. Maezumi *et al.*, *Nat. Plants* **4**, 540–547 (2018).
19. W. G. Hill, A. Robertson, *Genet. Res.* **8**, 269–294 (1966).
20. M. W. Feldman, S. P. Otto, F. B. Christiansen, *Annu. Rev. Genet.* **30**, 261–295 (1996).
21. R. E. Green *et al.*, *Science* **328**, 710–722 (2010).
22. J. Watling *et al.*, *PLOS ONE* **13**, e0199868 (2018).
23. B. Glaser, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 187–196 (2007).
24. E. R. Ribeiro, *Liames* **9**, 61–76 (2009).
25. J. G. de Souza *et al.*, *Nat. Commun.* **9**, 1125 (2018).
26. M. J. Heckenberger *et al.*, *Science* **321**, 1214–1217 (2008).
27. I. Wüst, C. Barreto, *Lat. Am. Antiq.* **10**, 3–23 (1999).
28. C. R. Clement, *Econ. Bot.* **53**, 203–216 (1999).
29. R. S. Walker, L. A. Ribeiro, *Proc. Biol. Sci.* **278**, 2562–2567 (2011).
30. L. Kistler *et al.*, *Dryad Digital Repository* (2018); doi:10.5061/dryad.70t85k2.

ACKNOWLEDGMENTS

We thank Admera Health for assistance with sequence data collection and D. Piperno for comments on the manuscript. **Funding:** Work was supported by Natural Environment Research Council Independent Research Fellowship NE/L012030/1 to L.K., and a sub-award from Science and Technology Facilities Council grant ST/K001760/1 (PI Thomas Meagher, co-I Peter Kille) to L.K. and R.G.A. **Author contributions:** Study conceptualization and design: L.K., F.O.F, and R.G.A.; Sample acquisition: F.O.F., A.P.P., C.G., B.A., and M.T.P.G.; Genomic data collection: L.K., F.O.F., O.S., N.W., and R.R.M.; Genomic data analysis: L.K. and N.A.S.P.; Archaeology and linguistic background and interpretation: J.G.S., S.Y.M., F.O.F., F.M.C., and E.R.R.; Interpretation and integration of results: L.K., S.Y.M., J.G.S., F.M.C., J.R.-M., N.W., F.O.F., and R.G.A.; Visualization: L.K., S.Y.M., J.G.S., H.L., and N.A.S.P.; Manuscript drafting: L.K., S.Y.M., and J.G.S., with input from N.A.S.P., F.O.F., and R.G.A. All authors reviewed and contributed to the final manuscript. **Competing interests:** We declare no competing interests. **Data and materials availability:** Raw sequence data, NCBI Sequence Read Archive accession SRP152500. In-house scripts for data handling and analysis (allele frequency estimation, *f* and *D* statistic calculation, genome alignment conformation for mutation load analysis, and exclusion amplification duplicate removal), genome-wide GERP scoring details, genomic mappability bed file, SNP calls, and mapDamage results are available in (*30*). Germplasm for newly sequenced maize landraces is curated at the Embrapa gene bank in Brasilia, Brazil, and Programa Cooperativo de Investigaciones en Maíz in Peru, which provided sample material for this study to F.O.F. and C.G. Archaeological samples from Santa, Chorrillos, Ica, and Jujuy were originally obtained from the PSUM Archaeological Project, Paurarku Archaeological Project and Samaca Archaeological Project, facilitated by archaeologists V. Pimentel, K. Lane, D. Beresford-Jones, and H. Yacobaccio.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/362/6420/1309/suppl/DC1
Materials and Methods
Figs. S1 to S4
Tables S1 to S4
References (*31*–*94*)

8 August 2018; accepted 22 October 2018
10.1126/science.aav0207

NEW ENGLAND BioLabs® Inc.

PASSION IN SCIENCE AWARDS®

What is your PASSION?

New England Biolabs® is pleased to announce our **2019 Passion In Science Awards®**, recognizing those within the scientific community working to make a difference in the world through their scientific mentorship, humanitarian efforts, environmental stewardship, or arts and creativity.

Through January 31st, 2019, we invite you to nominate yourself or a colleague for one of these awards.* Award winners will be our guests at a celebration at the NEB® campus in Ipswich MA, USA and will also receive a donation to help further their cause. Remember, no act is too big or too small. Help us celebrate those individuals who demonstrate true passion in science.

Tell us your story.

Learn more and apply at NEBPassionInScience.com

* View terms and conditions, and additional details about the awards at NEBPassionInScience.com

E-mail : info@axion-research.com
www.axionr.com/eng/

AXiR Engine® tells you the risk of future disease in your body.

Until today we' ve been ignoring small signal in our body if we feel that the pain is a split second or very short period. In Japan we see the number of deaths by cancer like 370,000 people and the number is similar in the past 6 years. Maybe the fatalities number will be similar in 2018 and this causes the next guess like that the next year candidate 370,000 are waiting for cancer death order because of invisible reason today.
The total number of individuals affected by cancer in Japan is about 1,000,000 and the number means 37% people might be going to die. Medical doctor is focusing on patients who are clearly suffering from actual diseases like cancer. AXiON Research Inc., we assume the invisible path to reach those diseases and some reason with scientific background must be there to meet the numbers repeated every year. In The "Compass to Healthy Life" Research Complex Program organized by RIKEN and supported by JST, in July 2018, AXiON Research Inc. shows the performance of Replica Generator III to increase data of healthcare on 415 people data up to 4,150.
It' s really good result and succeeded in increase of 4 initial disease risk categories up to 7 ones after the data generation with keeping the original statistical feature and accuracy. It also keeps the probability density, the dispersion and the distribution of the original data increasing the number per different age and gender. It contributes the future analysis of 10,000 people data and analyzes health indices and invisible disease risk prediction for pre-disease state.

AXiR Engine

P-HARP
PRECISION HEALTH ANALYSIS RESEARCH PLATFORM

AXiON Research Inc. is introducing new methodology to identify the health indices / position in pre-disease state group and how quickly getting worse by checking autonomic nerves activities or up / down parameters related to individual immunity. Heatmap of health condition/disease risk level helps the identification of the risk and vector-map of disease risk level works how much, how quickly getting worse or shifting the next stage of future diseases.
The challenge started a few years ago and various technologies / research results are accelerating the analysis process and its accuracy. Today medical service is often using X-ray, MRI, CT or PET and it can identify the disease itself at very high rate. However, it does usually work as an answer like you have no problem at all, at least today. Some of pre-disease state people might be waiting for new technology to make some advice or estimation of future disease risk at very early stage. We introduce AXiR Engine® for how much healthy you are or we are. We' re going to the next pre-service stage with strategic partners to run the service into the real world and commercial service.

Business Model

AXiON Research provides AXiR Engine and P-HARP customized services for the customers and partners. Healthcare Services are built in various application. We' re interested in strategic alliance partners to accelerate the market adoption.

Global Partnership

AXiON Research is building global partnership with AR/VR and robotics companies targeting healthcare industry. We' re also looking for worldwide business partners.

Fund Raising

We' re seeking the next fundraising Series A in 2019. Series A 1st will be between Jan. and Apr. 2019. Series A 2nd will be between June and Oct. 2019. Welcome, Early Entries.

AXION RESEARCH INC.

IoT for environmental conservation and sustainable agriculture

In modern day agriculture, high inputs of nitrogen are being used to gain high crop yields. Of this nitrogen, 50-70% is lost from the soil *(1)* with devastating environmental effects. An IT solutions provider, PS Solutions, is taking part in an ambitious project in Colombia to build efficient cultivation management practices. The project is using internet of things (IoT) solutions designed by PS Solutions to maximize the crop production and minimize the environmental impact of nitrogen.

Nitrogen in water contamination and greenhouse gas emission

Urea is one of the world's most commonly used nitrogen-based fertilizers. When applied to soil, nitrifying bacteria transform urea into nitrite (NO_2-) and then to nitrate (NO_3-). As NO_3-, nitrogen easily moves from soil to water *(1, 3)*, with approximately 30-50% of the nitrogen in fertilizer leaching into the ocean.

Nitrification also contributes to global warming by emitting nitrous oxide (N_2O) and other nitrogen oxides (NO_x). These molecules can destroy the ozone layer and trap heat with 300 times more efficiency than CO_2 *(1)*.

The total economic losses from environmental pollution related to nitrogen-based fertilizers are estimated to be €70~€320 billion, which more than doubles the agricultural revenues gained *(2)*.

Biological Nitrification Inhibition (BNI) to reduce the nitrogen footprint

Many native tropical grasslands are highly nitrifying ecosystems, rapidly transforming nitrogen into very mobile forms. As explained above, these forms have negative environmental effects, reducing agriculture productivity and stocking capacity of agropastoral systems. On the other hand, some tropical grasses have high BNI. These plants release nitrification inhibitors and have the potential to enhance crop productivity and improve the use of nitrogen in crop rotations *(3, 4)*. With the ultimate aim to develop productive and environmentally friendly agropastoral systems, the Ministry of Agriculture, Forestry and Fisheries (MAFF) of Japan is supporting a research project run by Dr. Manabu Ishitani, a primary investigator at the International Center for Tropical Agriculture (CIAT, Colombia), that is replacing native grass in farm fields with *Brachiaria humidicola*, a tropical grass with high BNI.

The project goals are twofold. First is to show that nitrification suppression significantly reduces the environmental footprint of farming, and second is to improve agronomic practices by implementing IoT to exploit BNI functions.

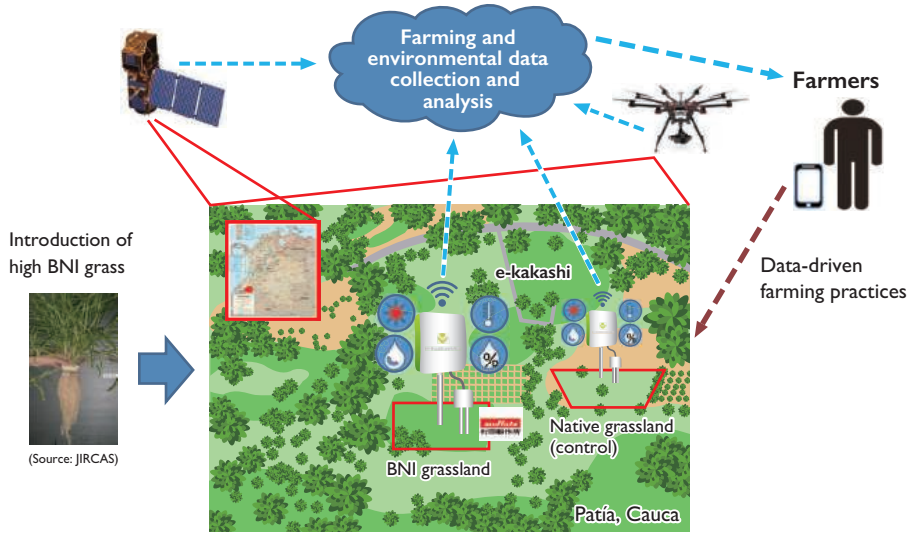
e-kakashi, an agricultural IoT

The effective use of tropical grasses with strong BNI function for efficient agricultural productivity largely depends on the field environment and farming practice. Thus, the first step in the project is to collect a large amount of cultivation and environmental data from the field. These data are collected by **e-kakashi**, a powerful IoT tool designed by PS Solutions for the purpose of enhancing farming practices.

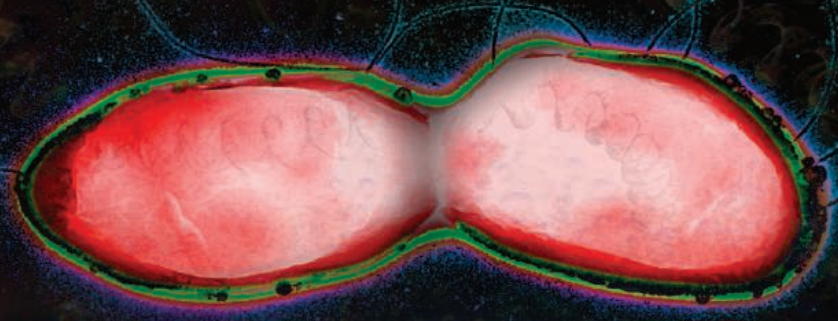
e-kakashi has been selected because of its proven data performance. Reliable data is critical for translating field observations and agricultural theories into practical farming. Connected to **e-kakashi** are new soil sensors with high sensitivity and accuracy manufactured by Murata Manufacturing Co., Ltd. **e-kakashi** will collect the sensor data, which includes soil temperature and soil moisture. Scientists will then analyze the data to estimate bacterial activity and nitrogen loss which enable them to assess the effects of BNI grasses on nitrogen retention and propose effective agricultural practice.

e-kakashi is an example of how IoT can benefit both agriculture and the environment. Dr. Takashi Togami, the developer of **e-kakashi**, sees it as a way to bring scientific solutions to everyday farming. "I look forward to a day when farmers use **e-kakashi** like we use TVs and smartphones."

1 Coskun D, et al. Nitrogen transformations in modern agriculture and the role of biological nitrification inhibition. Nat Plants. 2017;3:17074. doi: 10.1038/nplants.2017.74.
2 Sutton M, et al. Too much of a good thing. Nature. 2011;472(7342):159-61. doi: 10.1038/472159a.
3 Subbarao GV, et al. Evidence for biological nitrification inhibition in *Brachiaria* pastures. Proc Natl Acad Sci U S A. 2009;106(41):17302-7. doi: 10.1073/pnas.0903694106.
4 Karwat H, et al. Residual effect of BNI by *Brachiaria humidicola* pasture on nitrogen recovery and grain yield of subsequent maize. Plant Soil. 2017; 420(1-2):389-406. https://doi.org/10.1007/s11104-017-3381-z



MICROBES WITHIN THE HOST
IN HEALTH AND DISEASE



13th Annual Salk/Fondation Ipsen/Science Symposium on Biological Complexity

January 22 - 24, 2019 • Salk Institute for Biological Studies, La Jolla, CA

THE SYDNEY BRENNER NOBEL LECTURE JEFFREY GORDON *WASHINGTON UNIVERSITY IN ST. LOUIS*

KEYNOTE LECTURE
MARGARET MCFALL-NGAI *UNIVERSITY OF HAWAII AT MANOA*

SESSION 1: GUT-BRAIN AXIS
JANELLE AYRES (Chair) *SALK INSTITUTE FOR BIOLOGICAL STUDIES*
SARKIS MAZMANIAN *CALIFORNIA INSTITUTE OF TECHNOLOGY*
ROSA KRAJMALNIK-BROWN *ARIZONA STATE UNIVERSITY*
JOHN CRYAN *UNIVERITY COLLEGE CORK, IE*

SESSION 2: IMMUNOMETABOLISM
SUSAN PRESCOTT (Chair) *UNIVERSITY OF WESTERN AUSTRALIA, AS*
RUSLAN MEDZHITOV *YALE UNIVERSITY*
LUKE O'NEILL *TRINITY COLLEGE DUBLIN, IE*
RUTH LEY *MAX PLANCK INSTITUTE, GE*

SESSION 3: CANCER AND CANCER THERAPIES
LAURENCE ZITVOGEL (Chair) *GUSTAVE ROUSSY, FR*
CINDY SEARS *JOHNS HOPKINS MEDICAL INSTITUTE*
OMER YILMAZ *MASSACHUSETTS INSTITUTE OF TECHNOLOGY*
THOMAS GAJEWSKI *UNIVERSITY OF CHICAGO*

SESSION 4: HOST-PATHOGEN INTERACTIONS
DENISE MONAK (Chair) *STANFORD UNIVERSITY*
ANDREAS BÄUMIER *UNIVERSITY OF CALIFORNIA, DAVIS*
THAD STAPPENBECK *WASHINGTON UNIVERSITY IN ST. LOUIS*
HERBERT "SKIP" VIRGIN *WASHINGTON UNIVERSITY IN ST. LOUIS*

SESSION 5: IMMUNE-MICROBE INTERACTIONS
LORA HOOPER (Chair) *UNIVERSITY OF TEXAS SOUTHWESTERN MEDICAL CENTER*
GREGORY BARTON *UNIVERSITY OF CALIFORNIA, BERKELEY*
ANDREW MACPHERSON *UNIVERISTY OF BERN, SW*
YASMIN BELKAID *NATIONAL INSTITUTE OF HEALTH, NIAID*

SESSION 6: MICROBIAL COMMUNITIES IN HEALTH AND DISEASE
MARTY BLASER (Chair) *NEW YORK UNIVERSITY*
SUSAN LYNCH *UNIVERSITY OF CALIFORNIA, SAN FRANCISCO*
DAN LITTMAN *NEW YORK UNIVERSITY*
MICHAEL FISCHBACH *STANFORD UNIVERSITY*

ORGANIZING COMMITTEE: JANELLE AYRES · JAMES LEVINE · VALDA VINSON

SCIENTIFIC PROGRAM COMMITTEE: JANELLE AYRES (CHAIR) · RON EVANS · MICHAEL KARIN · ROB KNIGHT · MANUELA RAFFATELLU · INDER VERMA (PAST CHAIR) · VALDA VINSON · CAROLINE ASH

ONLINE REGISTRATION and INFORMATION: www.salk.edu/SFIS2019
CONTACT: events@salk.edu
#SFIS2019

ABSTRACT SUBMISSION DEADLINE: December 19, 2018 **REGISTRATION DEADLINE:** January 4, 2019

Poster design by Jamie Simon, Salk Institute

FONDATION IPSEN

salk® Where cures begin.

Science AAAS

FUJITSU Human Centric AI

Zinrai

Fujitsu's Artificial Intelligence Framework

FUJITSU

"Zinrai" is Fujitsu's approach to human-centric Artificial Intelligence (AI). Based on many years of research and development, Zinrai incorporates the latest AI technologies across both products and services to transform business and create new value for all.

Fujitsu's Zinrai has many applications, including:

City Surveillance

Real time observation using image recognition systems

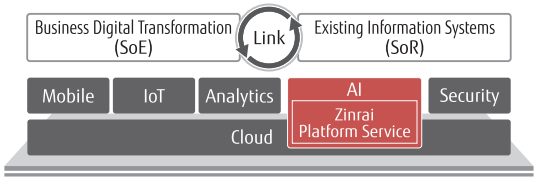
Retail

Analysis of shopper preferences by gaze tracking

Healthcare

Helping doctors make faster medical diagnosis

Digital Business Platform – MetaArc



Services developed on Zinrai technology are available through MetaArc, accelerating innovation to creat new value.

*SoE: Systems of Engagement, SoR: Systems of Record

shaping tomorrow with you

Automated Liquid-Handling System

CyBio FeliX is a flexible pipetting platform for fully automated single- to multichannel liquid-handling tasks, and can be equipped according to your individual process. It features up to 384 channels and a volume range of 0.5 µL–1,000 µL, enabling parallel transfer in 96- and 384-well formats, and pipetting with single channel, by column, or by row. Given the platform's modular design, customization to lab requirements is possible. Application-specific configurations can be adjusted and expanded according to specific needs. The high degree of automation allows CyBio FeliX to detect and change tips and pipetting tools independently within a pipetting routine. The pipetting heads are also easy to replace.

Analytik Jena
For info: +49-3641-77-7444
www.analytik-jena.de/en

Live-Cell Imaging System

The BioSpa Live-Cell Imaging System integrates the BioSpa 8 Automated Incubator and Cytation 5 Cell Imaging Multi-Mode Reader, fully automating kinetic live-cell imaging and analysis, such as 3D cell culture and phenotypic assays. Plate washing and reagent dispensing can be added to the system for complete walkaway automation, from sample prep through image analysis. BioSpa allows full, unattended workflow automation for up to eight microplates and other labware. The system also features integrated temperature and CO_2/O_2 control, plus humidity monitoring; fluorescence, high-contrast brightfield, color-brightfield, and phase-contrast imaging; and an OnDemand mode that enables the removal/replacement of labware, independent scheduling, and multiuser profiles.

BioTek
For info: 888-451-5171
www.biotek.com

DNA Sequencing Chip Kit

The Ion 520 Chip Kit contains eight barcoded chips for sample tracking and sequencing with the Ion S5 and Ion S5 XL Sequencing Systems. The kit electronically detects polymerase-driven base incorporation without the use of fluorescence. By eliminating the use of an optical detection system, this NGS technology allows for rapid sequencing times—as little as 2.5 h for 200-bp sequencing. Depending on the application, the Ion 520 generates 3–6 million reads. It is compatible with current library preparation methods, and can read lengths up to 600 bp to support numerous research protocols, including targeted gene sequencing and microbial sequencing.

Thermo Fisher Scientific
For info: 800-955-6288
www.thermofisher.com/order/catalog/product/a27762

Automated DNA/RNA Isolation

The chemagic Prime instrument offers automated nucleic acid isolation and assay setup by combining PerkinElmer's chemagic 360 instrument with the JANUS automated liquid-handling system. It uses patented magnetic bead technology [magnetic polyvinyl alcohol (M-PVA) Magnetic Beads] along with fully automated liquid handling, to provide high-quality, high-yield isolation of nucleic acids from a variety of sample types, suitable for NGS, multiplex ligation-dependent probe amplification (MLPA), genotyping, and PCR. Unlike other automated solutions, chemagic Prime uses magnetized rods instead of magnetic plates to separate nucleic acids from solutions. By transferring the beads instead of the process solutions, contamination risk is minimized, and higher-purity, more-intact DNA and RNA can be isolated. Reagent kits are available for isolating DNA and RNA from various human samples, including whole blood, saliva, plasma, tissues, FFPE samples, and feces.

PerkinElmer
For info: +49-(0)-2401-8055-00
www.perkinelmer.com

Automated Nucleic Acid Preparation System

A modular, automated liquid-handling and purification system developed by Promega Corporation for its Maxwell nucleic acid preparation offers labs newfound flexibility as compared to large, all-in-one instruments. The configurable system has two components: The Maxprep Liquid Handler provides automation for sample preparation of the Maxwell RSC (Rapid Sample Concentrator) cartridges and trays as well as postextraction sample preparation for fluorescent quantitation, sample normalization, and a variety of PCR reaction setups; and the Maxwell RSC 48 Instrument works with convenient, individual prefilled cartridges to process any number of samples from 1 to 48, without the risk of wasting reagents. Promega Portal Software allows the new and existing Maxwell RSC and Maxprep instruments to work together to transfer sample-tracking information from one device to the next to build a complete nucleic acid preparation workflow. Data can also be imported into a laboratory information management system.

Promega
For info: 608-274-4330
www.promega.com

Automated Plate Handler

The S-LAB from Peak Analysis & Automation (PAA) is an easy-to-use, compact, and affordable single-instrument loading solution, with the same reliability as a robotic arm. It easily fits standard lab benches and most safety cabinets, delivering reliable automation that accommodates varying workflows and lab space without stretching budgets. An onboard camera system enables easy instrument alignment and verification, and there is no requirement for a separate PC, thanks to an embedded controller and software. S-LAB is easily controlled via a smart, web-based application that can be run on any device, including a smartphone. It also boasts efficient, reliable handling of lidded microplates, with a capacity of up to 100 unlidded/80 lidded plates and the ability to handle most standard and deep-well microplates. There is also an optional barcode reader, compatible with industry standard formats.

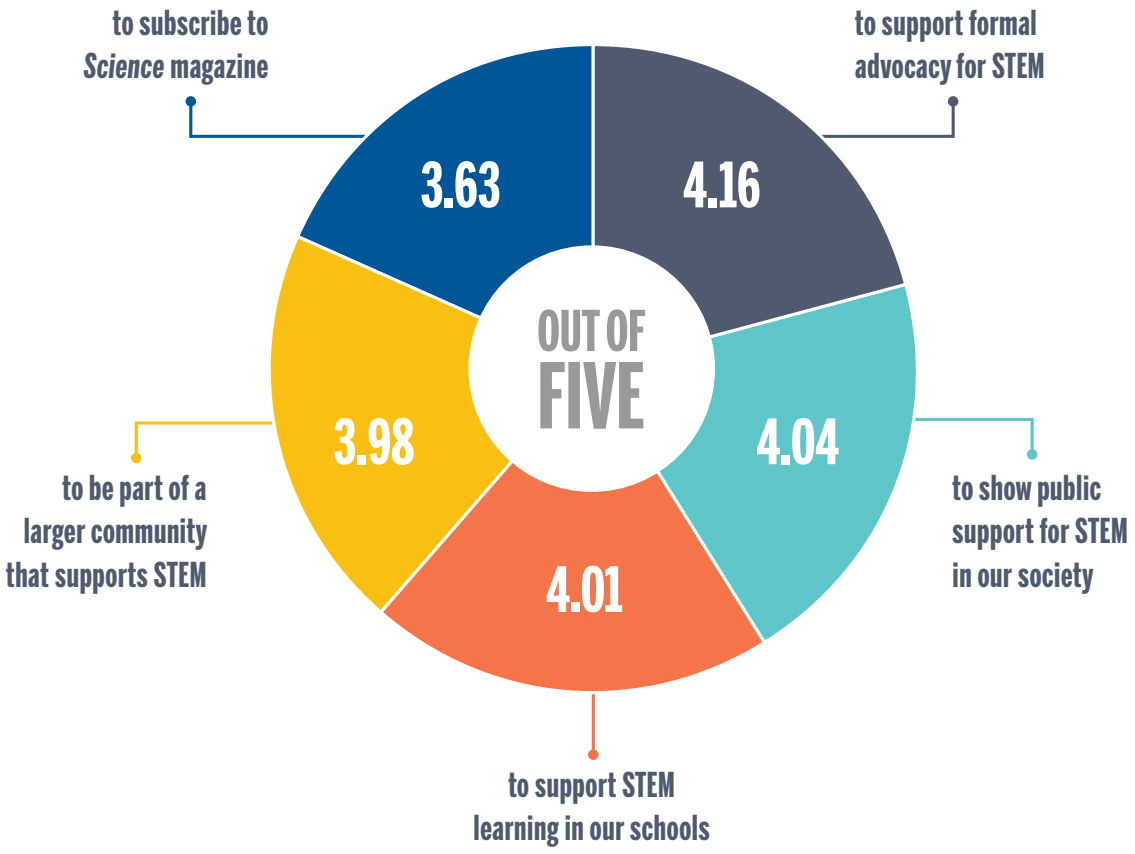
Peak Analysis & Automation
For info: +44-(0)-1252-373000
www.paa-automation.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

AAAS IS THE FORCE FOR SCIENCE

According to the 2017 Member Survey, you joined AAAS ...



TELL US WHAT'S IMPORTANT TO YOU!

The 2018 Member Survey is launching in September. Look in your inbox for a link.

Your responses help us to better serve science, scientists, and the global community. Don't miss your chance to tell us what's most important to you!

R&D SYSTEMS
a bio-techne brand

Consistency Is

N21-MAX and N-2 Neural Media Supplements from R&D Systems

Each lot of serum-free media is checked for performance consistency by our in-house quality team.

Learn more | rndsystems.com/neuralmedia

POSITIONS OPEN

UC San Diego

Four **Postdoctoral Researchers** are offered to join an interdisciplinary effort funded by the GB Moore Foundation to develop new algorithms and artificial intelligence methods for interpreting small molecule structure information from mass spectrometric and NMR data. Each postdoctoral scientist will join this team in one of four collaborating laboratories, with needed skill sets in natural products chemistry (William Gerwick, wgerwick@ucsd.edu), multidimensional mass spectrometry (Pieter Dorrestein, pdorrestein@ucsd.edu), computational and data science (Nuno Bandeira, bandeira@ucsd.edu), and deep convolutional neural networks for image recognition (Gary Cottrell, gary@ucsd.edu). These positions are available at the University of California San Diego beginning 1 January 2019 and are expected to last 2 years each, although extensions are possible. Send email letter of application to the PI's laboratory of primary interest, your full Curriculum Vitae, and names of 3 references.

INSTRUCTOR, PSYCHIATRY

The Department of Psychiatry, Yale University School of Medicine is seeking applicants to conduct NIH- supported research using Positron Emission Tomography (PET) in nonhuman primates and humans, to better understand and develop treatments for schizophrenia. Experience in schizophrenia research, the analysis of PET data is required and a history of funded research is desirable. Salary range: $100,000 - $150,000. Apply to: **Attn Dr. John Krystal, Dept. Of Psychiatry, Yale University School of Medicine, 300 George Street, Suite 901, New Haven, CT 06511**. *Yale is an Equal Opportunity, Affirmative-action employer. Women, minorities, persons with disabilities and protected veterans are encouraged to apply.*

POSITIONS OPEN

UT Southwestern Medical Center

Structural analyses of biomolecular condensates using cryoelectron tomography

Drs. Michael Rosen and Daniela Nicastro, in the Departments of Biophysics and Cell Biology at UT Southwestern Medical Center seek to jointly recruit a postdoctoral fellow to develop and apply cryoelectron tomography (cryoET) methods to understand the molecular organization of biomolecular condensates. Biomolecular condensates are compartments in eukaryotic cells that concentrate macromolecules in discrete foci without a surrounding membrane. Examples include cytoplasmic P bodies associated with RNA metabolism, promyelocytic leukemia nuclear bodies involved in transcription and anti-viral responses, signaling clusters in T cell activation, HP1 clusters in heterochromatin organization, and transcriptional assemblies involved in gene regulation. The behaviors of these compartments suggest that they may form through liquid-liquid phase separation of multivalent proteins and RNA. Our goal is to understand how molecules are organized in natural condensates in cells and in reconstituted condensates in vitro, and ultimately relate this organization to biochemical and cellular functions.

Candidates should have experience in cryoEM and/or areas relevant to condensate biology, such as cell biology or biophysics. Those without experience in cryoEM must have a strong commitment to learn all aspects of cryoET. Candidates will have access to state-of-the-art facilities, including a Titan Krios microscope and Aquilos cryoFIB mill, and will work at the leading edge of cryoEM technology and condensate biology.

Information on the UT Southwestern postdoctoral training program and benefits can be found in our Postdoc Handbook or at http://www.utsouthwestern.edu/postdocs.

Candidates should send Curriculum Vitae and a 1-page statement of research experience and future goals, as well as arrange to have three letters of reference sent to:

Dr. Michael Rosen
Department of Biophysics
UT Southwestern Medical Center
5323 Harry Hines Blvd.
Dallas, Texas 75390-8816
Michael.Rosen@utsouthwestern.edu
https://www.utsouthwestern.edu/labs/rosen/
https://www.utsouthwestern.edu/labs/nicastro/

UT Southwestern Medical Center is an Affirmative Action/Equal Opportunity Employer. Women, minorities, veterans and individuals with disabilities are encouraged to apply.

10 ways that *Science* Careers can help advance your career

1. Register for a free online account on **ScienceCareers.org**.
2. Search thousands of job postings and find your perfect job.
3. Sign up to receive e-mail alerts about job postings that match your criteria.
4. Upload your resume into our database and connect with employers.
5. Watch one of our many webinars on different career topics such as job searching, networking, and more.
6. Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
7. Complete an interactive, personalized career plan at "my IDP."
8. Visit our Career Forum and get advice from career experts and your peers.
9. Research graduate program information and find a program right for you.
10. Read relevant career advice articles from our library of thousands.

Visit **ScienceCareers.org** today — all resources are free

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

SCIENCECAREERS.ORG

By **Luke A. Schwerdtfeger**

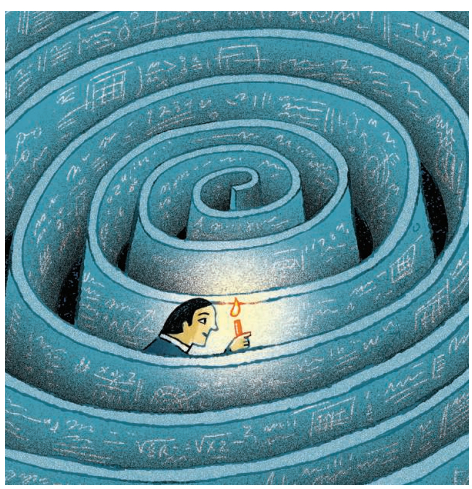
Spirals of science

The timing was perfect. A few weeks after the experimental protocol that had served me for years inexplicably stopped working, my grad school adviser approached me about writing a review paper detailing the history of our field. I was feeling hopeless about my lab work. I had seemingly tried everything to fix the broken tissue culture system, but nothing worked, crippling not only my productivity, but also my confidence. Shifting my focus to literature review and writing offered a welcome respite. And although I didn't expect it, this historical venture ended up teaching me how science proceeds across generations—and it provided the key to getting my research back on track.

While diving into the literature, I stumbled upon a string of reports from the 1910s describing experiments that were shockingly similar to the protocol I had been struggling with. The more I read, the more I questioned whether anything I was doing was actually novel. They led me to another set of intriguing papers, including one from 1934 describing a tissue culture method so complex that it seemed impossible.

Maybe there was unrealized value in this long-forgotten system, I thought. I set off to replicate the work.

After a couple of tries, I succeeded, which provided a much-needed confidence boost. I was amazed that such an advanced system had been invented more than 80 years ago. I couldn't wait to tell my adviser that I had replicated it.



"Each generation is assisted by the knowledge of the scientists who came before."

To my surprise, he wasn't as enthusiastic as I was. He asked a simple question: What's the utility? I didn't have an answer. This method was much more difficult than my current, albeit broken, protocol. So why should we care about it? Why not instead pour my time and energy into fixing my modern protocol?

I was determined to come up with some hidden utility for the rediscovered system. It was too interesting, too radical compared with what I had been doing to not be useful. But I couldn't think of anything. My surge of inspiration gave way to gloom. I was stuck with my broken method.

Returning to my adviser's office the next day, I was ready to admit that he was right and that replicating the historical method had been a waste of time. But he asked another simple question: Had it taught me anything? The answer was yes. On a technical level, it reminded me of the importance of paying attention to details, even minute and seemingly insignificant ones, which were critical for reproducing such an intricate method. Just as important, my excitement in trying a challenging, complex method got me out of my research slump. Perhaps that was the true utility.

With renewed vigor and focus on the details, I finally got my protocol working again, after nearly 5 months of troubleshooting. The problem turned out to be infuriatingly simple: a bad component in the tissue culture media. As fate would have it, I got my protocol working just as our review paper was accepted for publication.

My adviser often uses the phrase "spirals of science" to describe how science progresses. The idea—which he inherited from his postdoc adviser—is that researchers sometimes retrace paths conceptually similar to those explored by previous generations. But each generation is assisted by the knowledge of the scientists who came before, which allows the spiral to progress upward.

This notion hadn't fully resonated with me when he mentioned it after we found the tissue culture papers from the 1910s. But after troubleshooting my protocol while diving into the history of my field, I saw exactly what he meant. And I took confidence from the thought that, no matter how slowly, I was progressing up the spiral. ■

Luke A. Schwerdtfeger is a Ph.D. student at Colorado State University in Fort Collins. *He thanks his adviser and mentor Stuart Tobet for helpful comments on the piece and the inspiration to write it. Send your career story to SciCareerEditor@aaas.org.*